Audio Monitoring

COMPUTING
COMMUNICATIONS
VIDEO

▶ Introduction

A picture may say a thousand words, but the addition of audio completes the viewing experience. If viewers experience audio problems as they watch a program, they will believe the associated video material is of lower quality compared to the same video signal with no audio problems. Therefore, it is essential to ensure the quality of the audio signal.

Audio devices use either balanced or unbalanced signals. Each format has its own physical and electrical characteristics and specific strengths and weaknesses. A good understanding of these formats will aid in the understanding and appropriate application of audio signals.

Unbalanced Signals

Unbalanced systems use a signal and ground. Shields are sometimes employed as well. Interconnection of unbalanced signals is simple using relatively inexpensive cables and connectors such as the RCA phono jack as shown in Figure 1. The outer conductor of the connector mounted onto the equipment is often in contact with the chassis. Reasonable care is required to avoid shorting active signals to chassis ground by reversing plug connections. A simple audio input stage can be used for unbalanced audio as shown in Figure 1. The lower cost and complexity of unbalanced signals means that they are used on consumer products. These applications don't enjoy the same noise and common-mode rejection properties of balanced audio signals. Thus, they are more susceptible to interfering signals and have to use short cable runs.

Tektronix
Enabling Innovation

Balanced Signals

The term balanced refers to the fact the signal has two components, equal in amplitude but opposite in polarity and the impedance characteristics of the conductors are matched. Current practices designate these as non-inverted and inverted, + and –, or positive and return. Interconnect cables usually have three conductors. Two arranged as a twisted pair, carry the non-inverted and inverted. By employing a twisted pair of conductors for the signal leads, the loop area responsible for magnetic interference is minimized. The third conductor is a shield.

The shield attenuates the effects of external electric fields. However, because no shield can completely eliminate interference, fields penetrating the shield are imposed equally on both the inverted and non-inverted signals. This signal becomes a common-mode signal to the input circuits. See Figure 2.

Balanced audio systems employ either electronic or transformer-based differential input stages. Differential inputs subtract the

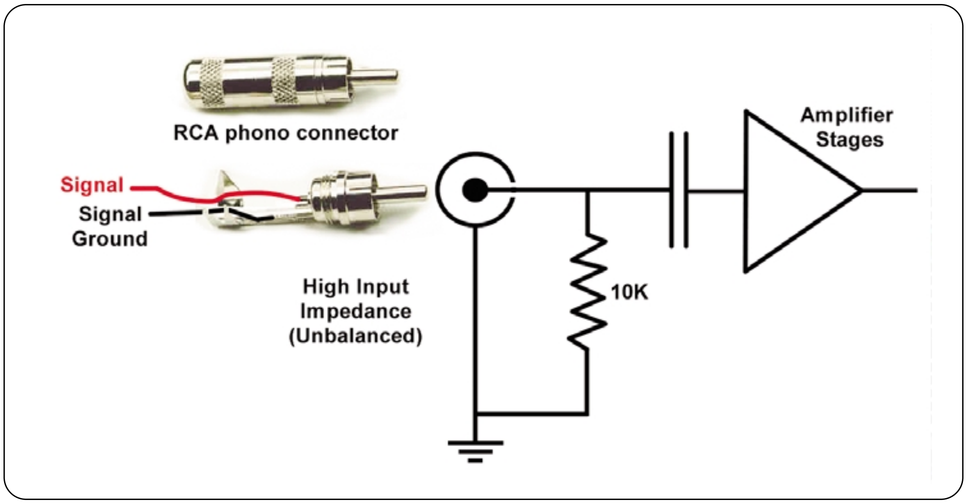


Figure 1. *Unbalanced audio connection.*

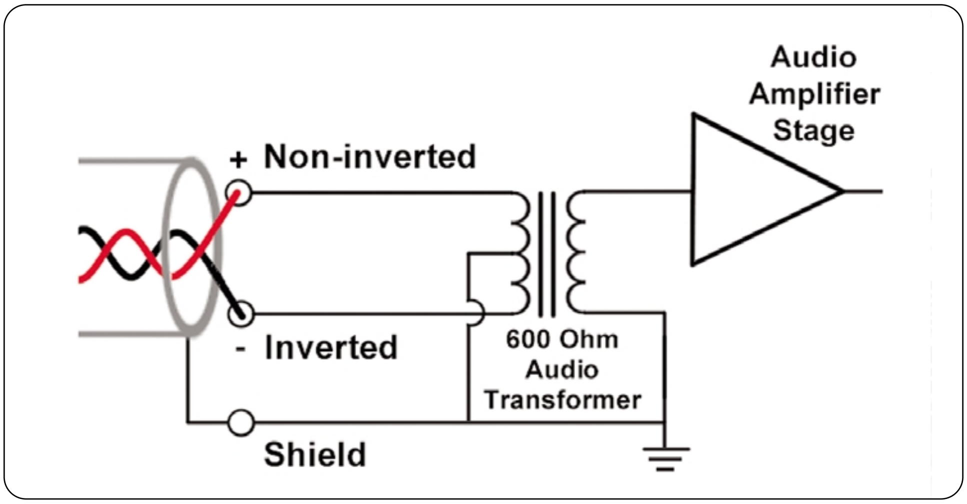


Figure 2. *Balanced audio connection.*

inverted input from the non-inverted input. Due to this behavior, common-mode signals are cancelled. The balanced format is popular in professional applications where noise rejection properties and high signal amplitude outweigh the interconnect complexity and higher cost. The typical connector used is the XLR as shown in Figures 3 and 4.

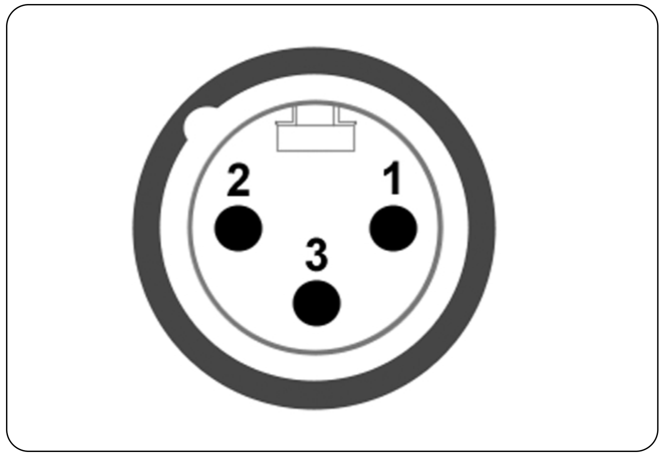
Table 1. Balanced output configuration

Non-inverted	XLR Pin	2
Inverted	XLR Pin	3
Shield	XLR Pin	1

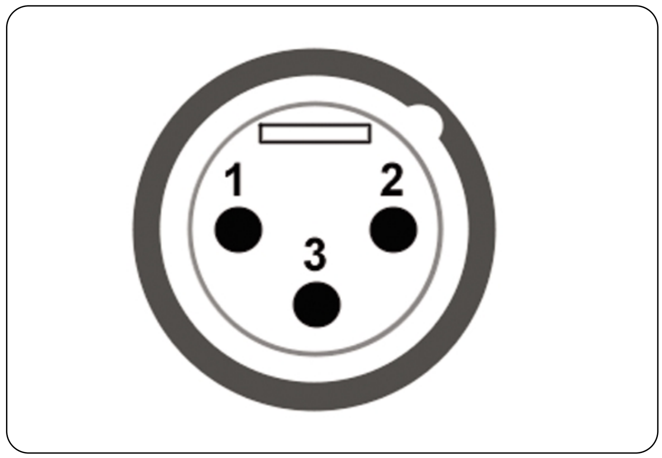
One must ensure the cables used within the facility are correctly wired to prevent problems within the facility such as polarity reversal.

The simplest form of audio monitoring is to use a level meter that displays the audio signal amplitude. There are two types of metering, a VU (Volume Unit) meter or a PPM (Peak Program Meter), and there are important differences between them. VU meters and PPMs present different responses to audio program material. The VU meter displays the average volume level of the audio signal and has symmetrical rise and fall times with a relatively long integration time (typically 300 ms). The integration time is strongly influenced by the mechanical inertia of the needle mechanism. A PPM displays the peak volume level of the audio signal with a fast rise time (10 ms), a slow fall time (2.85 s) and a 10 ms integration time. Electronic circuits compensate for the inertia of any mechanical variation in the PPM. Because of these differences, it is rare for a VU meter and PPM to have identical responses to audio program material.

When lining up a system with a test tone, the PPM must read lower than the VU meter to make them equivalent using the same audio program. Broadcast authorities have found that 8 dB (decibel) is a good average difference between peak-to-reading ratio of the PPM and VU meter. Hence, they have specified that a line-up tone reading of 0 VU on the VU meter should read –8 dB on the PPM. With this alignment, both meters will read substantially the same with audio program material, with the PPM giving more reliable control of program peak levels. Audio program material should be adjusted to have peak amplitude of 0 dB on a PPM.



▶ ***Figure 3.*** *Female XLR connector.*



▶ ***Figure 4.*** *Male XLR connector.*

The Decibel Scale

Audio measurements are often expressed in the decibel (dB) scale because of the wide dynamic range of audio signal levels. The dB scale is a logarithmic function either expressing a voltage or power measurement. Using the dB scale allows us to quantify changes in audio signal, because the human ear senses changes in amplitude on a logarithmic basis.

$$\mathrm{dB} = 20 \log \frac{V_2}{V_1} = 10 \log \frac{P_2}{P_1}$$

$$\text{Note: } P = \frac{V^2}{R}$$

V_1 = Reference Voltage Level
V_2 = Measured Voltage Level
P_1 = Reference Power Level
P_2 = Measured Power Level

P = Power
V = Voltage
R = Resistance

Typically within audio measurements a dBm value is specified. This means that a reference power of 1 mW was used with a 600 Ω termination. Therefore using the equations, 0 dBm is equivalent to a voltage of 0.775 V into a 600 Ω load. You may encounter several different types of dB measurements used within audio. The following list indicates the typically used equations:

- dBm = 10 log P_1 / 0.001 W
- dBV = 20 log V_2 / 1 V RMS
- dBv = 20 log V_2 / 775 mV RMS
- dBu = 20 log V_2 / 775 mV RMS
- dBSPL = 20 log P_1 / P_2

Setting Audio Levels

The Lissajous pattern

The left and right input signals are applied to an X-Y display similar to a vectorscope. The left channel is applied to the N-S axis and right signal content on the E-W axis, as seen in Figure 5. Many audio professionals are more familiar with the “Sound Stage” mode that simply rotates the display by 45 degrees to more easily visualize correct phasing of the channels, as shown in Figure 6. The Lissajous display provides instantaneous feedback of the overall energy distribution during a remix. The pattern orientation indicates at a glance whether the present mix is balanced or concentrated to either side. Figures 7, 8, and 9 illustrate different energy distributions.

An audio monitor can clearly indicate errors within the program material such as clipping which manifests itself on the Lissajous display by a ‘Squaring Off’ of the pattern edges. Figure 10 illustrates a severe case of clipping.

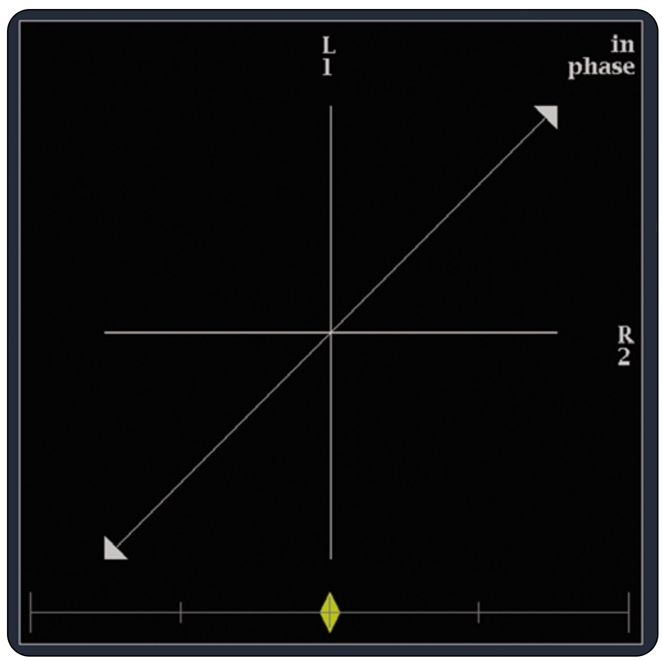


Figure 5. *Lissajous X-Y display.*

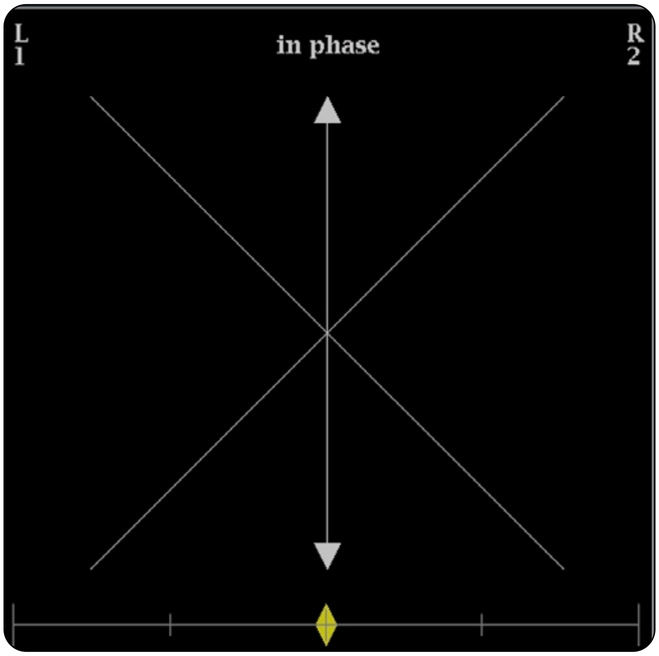


Figure 6. *Lissajous display in "Sound Stage" mode.*

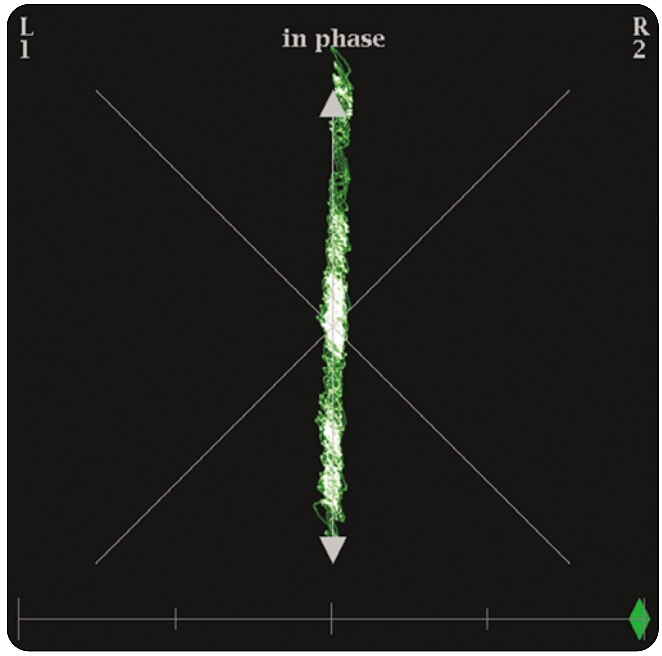


Figure 7. *Single tone from synthesizer.*

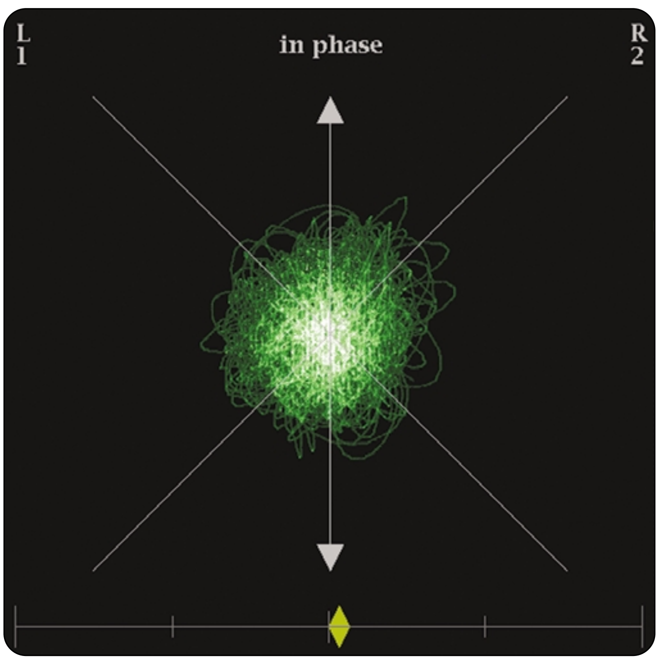


Figure 8. *Stereo signal with little correlation.*

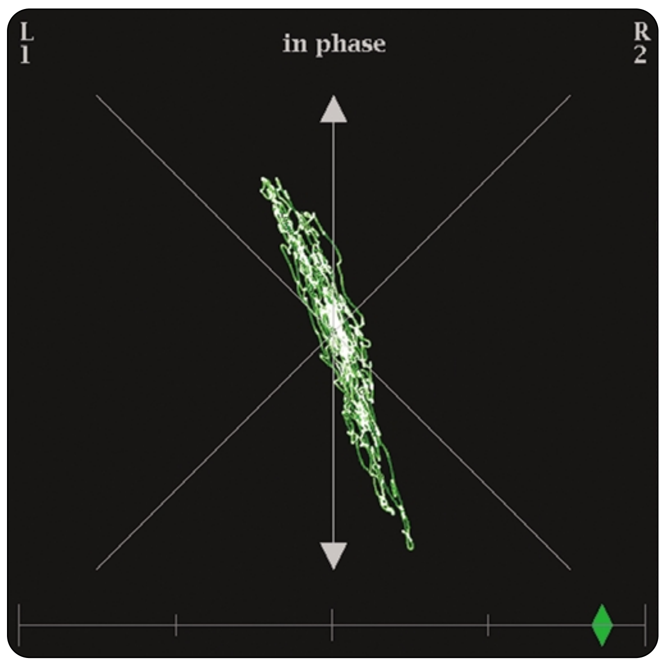


Figure 9. *Stereo signal with strong left component.*

System phase errors

Phase errors can introduce any number of undesirable effects in an audio signal. A quick check with an audio monitor can help identify and quantify any significant amount of system phase error.

Select the auto gain control to make the edges of the ellipse just touch the phase tangent lines. If a straight line coincident with the L = R axis is observed, the left and right channels of the equipment-under-test are exactly matched in phase and gain as seen in Figure 11. If a slanted line is observed, the left and right channels match in phase but do not have the same amplitude. A straight line perpendicular to the L = R axis indicates reversed phase between channels. Finally, an ellipse whose major axis falls on the L = R line indicates equal amplitude but phase mismatch.

Signal polarity reversal

Multi-microphone recordings in a complex studio present dozens of opportunities to introduce polarity reversals. Any time a polarity reversal occurs, the audio monitor can be used to quickly trace the problem back to its source. By introducing a sine wave into both

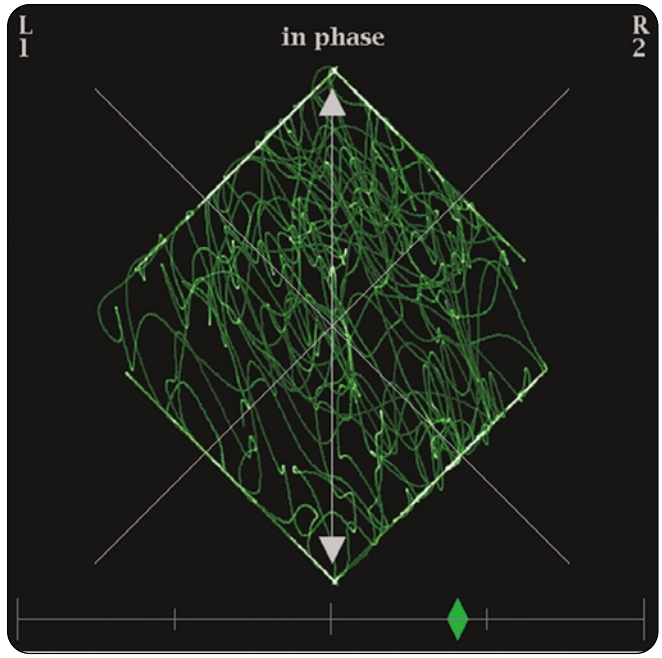


Figure 10. *Stereo signal with severe clipping.*

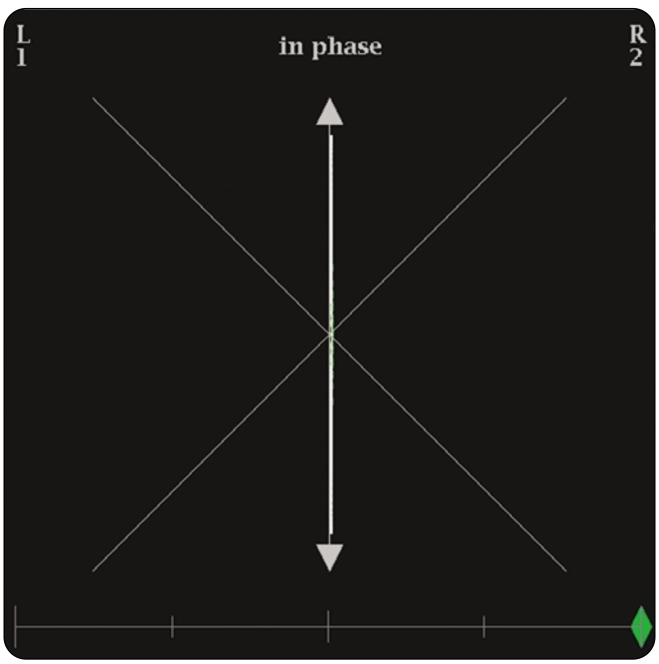


Figure 11A. *Left and right channels matched in phase and gain.*

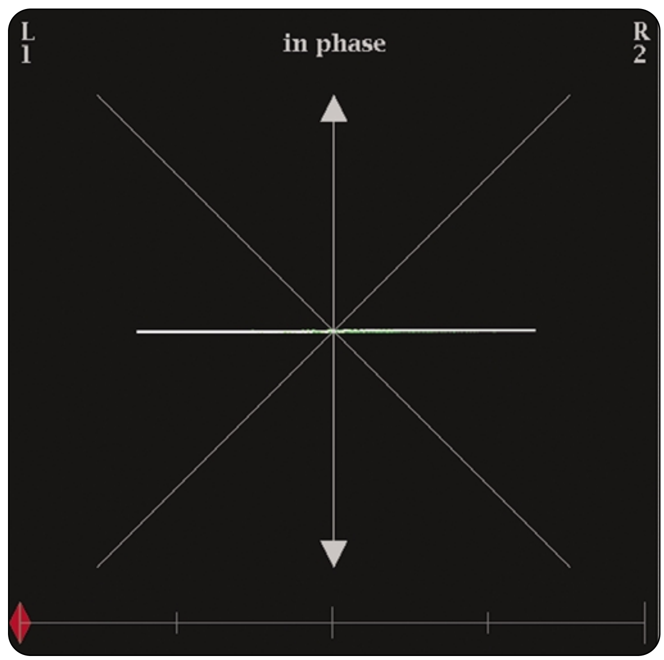


Figure 11B. *Left and right channels matched in gain but phase reversed.*

channels of the system and checking outputs stage by stage, the source of the phase reversal can be quickly identified. A correctly phased signal will produce a straight vertical line on the soundstage Lissajous display. If the signal is phase reversed, the Lissajous display will indicate a horizontal line within the display.

Digital Audio

The transition to digital audio has evolved over many years with the AES 3 standard dominating the video industry. The interface is a serial data stream, there is no separate clock signal, and in order for the receiver to recover the data, it must extract the clock from the data stream. This is achieved by using a simple coding scheme known as bi-phase mark coding, as illustrated in Figure 12. A transition occurs every bit period and when the data value is a "1", an additional transition occurs at half the bit period. This ensures easy clock extraction from the data and minimizes the DC component present within the signal. Since transitions represent the data values, the signal is also polarity insensitive.

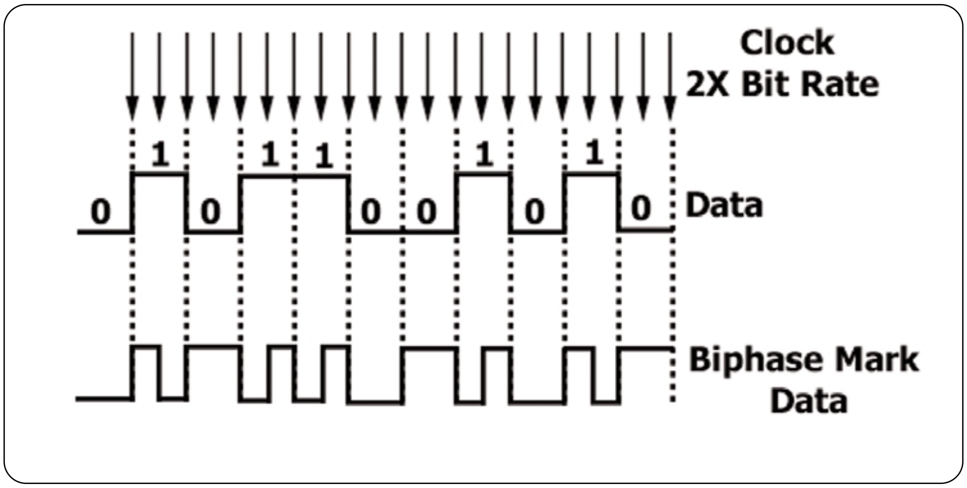


Figure 12. *Bi-phase mark coding.*

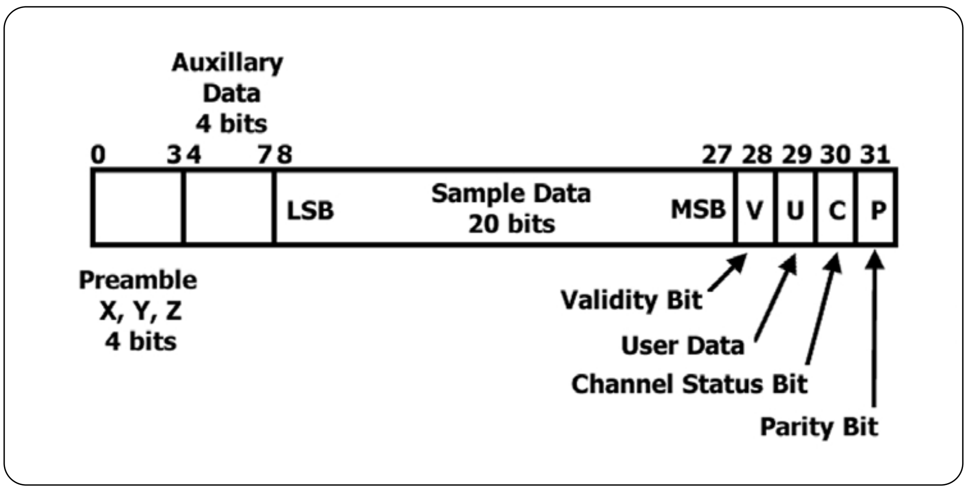


Figure 13. *Subframe structure.*

The AES 3 standard supports multiple sampling rates of 32 kHz, 44.1 kHz (CD), and 48 kHz (Professional) which is predominately used within video facilities. The analog audio signal is sampled at the clock rate and 16, 20, or 24 bits can be used to represent the amplitude of the audio signal. The greater number of bits needed for audio compared to video is because of the larger dynamic range; increasing the number of bits produces an adequate signal-to-noise ratio (SNR).

The basic formula for determining the SNR for digital audio is:

$$SNR = (6.02 * n) + 1.76$$

where "n" is the number of bits per sample

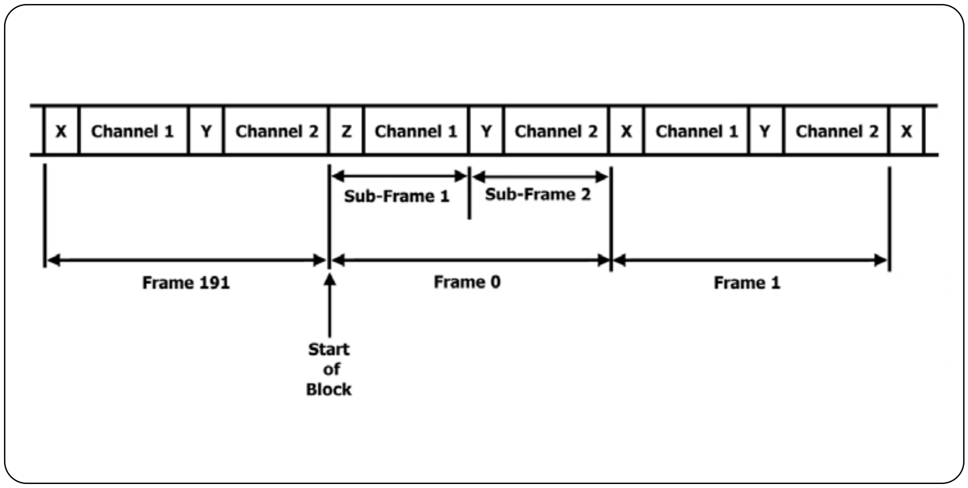
For a 16-bit system, the maximum theoretical SNR would be (6.02 * 16) + 1.76 = 98.08 dB; for a 20-bit device the SNR would be 122.16 dB; and for a 24-bit device, 146.24 dB. A well-designed 20-bit analog-to-digital converter (ADC) typically offers a value of between 100 and 110 dB SNR.

The data embedded within the serial data stream contains two audio channels, Channel 1 and Channel 2, that are multiplexed together. These channels may be separate monophonic channels, a stereo pair containing Left and Right, a single audio channel with the second channel identical to the first, or they may contain no data with the values set to a logical "0". A 4-bit sync word called a preamble identifies the channels and does not conform to the bi-phase mark-coding scheme. This helps to identify the sync words from the rest of the data. The data for each channel is grouped into a 32-bit subframe as shown in Figure 13. The multiplexing of these subframes for Channel 1 and Channel 2 forms a frame. 192 frames are further grouped into a block as shown in Figure 14.

The subframe is comprised of the following:

Preamble

The preamble is a 4-bit synchronizing word used to identify the channels and start of a block. Channel 1 is identified by a preamble X and Channel 2 is identified by a preamble Y. The 192-frame block is identified by preamble Z, as shown in Figure 15. The preamble violates the bi-phase mark-coding scheme and allow easy identification of the preamble from the rest of the data.

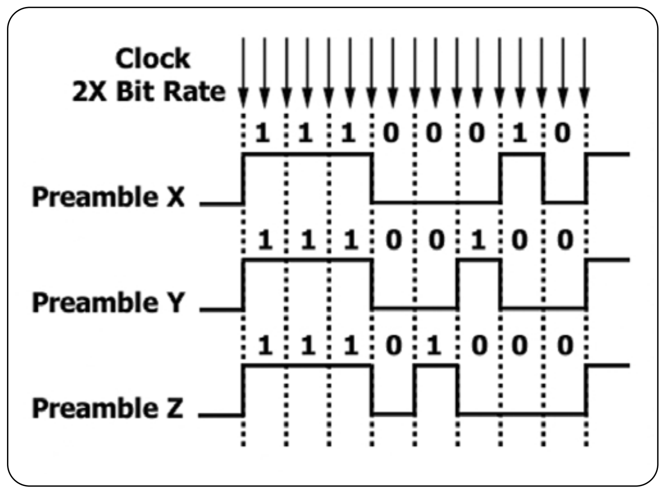


▶ ***Figure 14.*** *AES/EBU frame format.*

Table 2 shows the values for the preamble with a preceding state of "0" or "1". Normally, only one type of these values will be transmitted. However, polarity reversal within the signal path requires that either of the states be decoded.

Table 2. Preamble Definitions.

Identification	Preceding State 0	Preceding State 1	
X	11100010	00011101	Sub-Frame 1
Y	11100100	00011011	Sub-Frame 2
Z &	11101000	00010111	Sub Frame 1 Block Start



▶ ***Figure 15.*** *Preambles X, Y, and Z.*

Auxiliary data bits

When a 20-bit audio sample is used, the four least significant bits (LSB) can be used for auxiliary data. One application for these auxiliary data bits is for use as a voice-quality audio channel to provide a talkback channel. Otherwise, these bits can be used to carry the four LSBs of a 24-bit audio sample.

Audio sample data bits

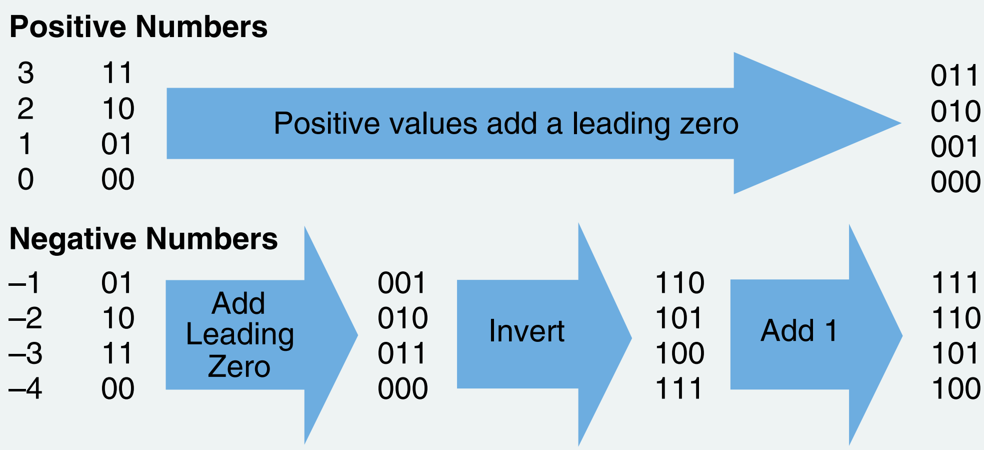
The audio sample data is placed between bits 4 to 27 with the most significant bit (MSB) placed at bit 27 and supporting a maximum sample of 24 bits. If not all the 24 bits are used for an audio data sample, the LSBs are set to "0". Typically within broadcast facilities, an audio sample of 20 bits is used. This allows for auxiliary data channel within the four LSBs from 4 to 8.

The 20-bit audio sample is used for most applications within a broadcast environment. However, a 24-bit audio sample is supported in AES/EBU by using the sample bits from 4 to 27 and not providing any auxiliary data bits. The binary audio data sample is 2's-complement encoded. Using this simple technique greatly reduces the complexity of the audio hardware design.

The subframes also carry additional data bits, which provide useful information on the audio channels.

2'S Complement Code

The 2's complement coding scheme uses the Most Significant Bit (MSB) to indicate the positive or negative nature of the signal. If the MSB is set to '0,' the value indicates a positive number. If the MSB is set to '1', the value is a negative number. The following example shows the possible range for a 2-bit number in 2's complement coding:



Validity bit (V)

When the validity bit is set to "0", the subframes audio data is suitable for decoding to analog audio. If the validity bit is set to "1", the audio sample data is not suitable for decoding to an analog audio signal. Test equipment can be set up to ignore the validity bit and continue to use the data for measurement purposes.

User data bit (U)

The user data bits can be used to carry additional information about the audio signal. Each U bit from the 192 subframes can be assembled together to produce a total of 192 bits per block. The operator can use this information for such purposes as copyright information.

Channel status bit (C)

The channel status bit provides information on various parameters associated with the audio signal. These parameters are gathered for each C bit within the 192 subframes for each audio channel. Table 3 shows the information carried within these bits.

There are three levels of implementation of the channel status data: minimum, standard, and enhanced. The standard implementation is recommended for use in professional television applications, hence the channel status data will contain information about signal emphasis, sampling frequency, channel mode (stereo, mono, etc.), use of auxiliary bits (extend audio data to 24 bits or other use), and a CRC (cyclic redundancy code) for error checking of the total channel status block. For additional information on the detailed description of the channel status information, please see Appendix A.

Parity bit (P)

The parity bit is set such that the values of bits 4 to 31 form an even parity (even number of ones) used as a simple means of error checking to detect an error within a subframe. Note that it is not necessary to include the preambles since they already have even parity.

Table 3. Channel Status Use

<table>
<tr><th></th><th colspan="8">Bit</th></tr>
<tr><th>Byte</th><th>0</th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th></tr>
<tr><td>0</td><td>Use of Channel Status Channel</td><td>Audio/non-audio use</td><td colspan="3">Audio Signal emphasis</td><td>Locking of source sampling frequency</td><td colspan="2">Sampling frequency</td></tr>
<tr><td>1</td><td colspan="4">Channel Mode</td><td colspan="4">User bit management</td></tr>
<tr><td>2</td><td colspan="3">Use of auxiliary sample bits</td><td colspan="3">Source word length & source encoding history</td><td colspan="2">Reserved</td></tr>
<tr><td>3</td><td colspan="8">Future multi-channel function description</td></tr>
<tr><td>4</td><td colspan="2">Digital Audio reference signal</td><td colspan="6">Reserved</td></tr>
<tr><td>5</td><td colspan="8">Reserved</td></tr>
<tr><td>6</td><td colspan="8" rowspan="4">Alphanumeric Channel origin data</td></tr>
<tr><td>7</td></tr>
<tr><td>8</td></tr>
<tr><td>9</td></tr>
<tr><td>10</td><td colspan="8" rowspan="4">Alphanumeric Channel destination data</td></tr>
<tr><td>11</td></tr>
<tr><td>12</td></tr>
<tr><td>13</td></tr>
<tr><td>14</td><td colspan="8" rowspan="4">Local sample address code</td></tr>
<tr><td>15</td></tr>
<tr><td>16</td></tr>
<tr><td>17</td></tr>
<tr><td>18</td><td colspan="8" rowspan="4">Time-of-day sample address code</td></tr>
<tr><td>19</td></tr>
<tr><td>20</td></tr>
<tr><td>21</td></tr>
<tr><td>22</td><td colspan="8">Reliability flags</td></tr>
<tr><td>23</td><td colspan="8">Cyclic redundancy check character</td></tr>
</table>

AES Interconnection

There are two basic connection types that can carry the AES/EBU serial digital data. A standard XLR can be used to carry the digital signal over a twisted pair cable. Pins 2 and 3 carry the balanced data signal and pin 1 is used as the shield. Note that since the signal is polarity insensitive, it does not matter which of the two wires is connected to pins 2 and 3. However convention defines pin 2 as 'positive' and pin 3 as 'negative.'

Table 4. Balanced output configuration

Signal '+'	XLR Pin	2
Signal '–'	XLR Pin	3
Shield	XLR Pin	1

▶ ***Figure 16.*** *Male XLR connector.*

Figure 17 illustrates the standard interconnection circuit. The output voltage level can be between 2 and 7 Volts with an impedance of 110 Ω and the signal is transformer-coupled and balanced. The signal can be transported across a 100-meter cable without the need for an equalizer.

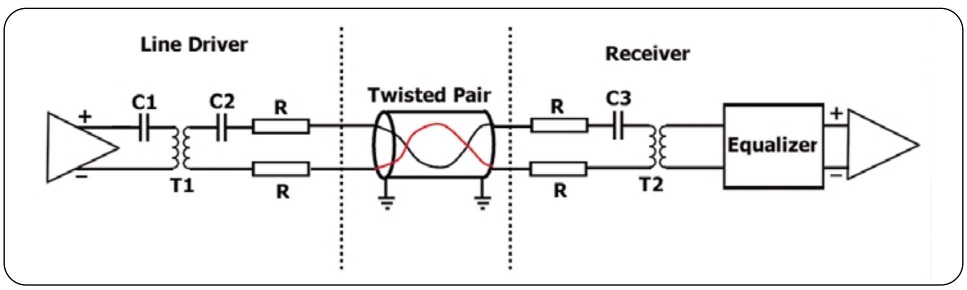
An alternative connection is to use a standard 75 Ω coaxial cable and BNC connector as defined by standard AES3-ID. This is an unbalanced interface that permits broadcast facilities to transmit AES/EBU digital audio on standard coax using the existing infrastructure of the plant.

The signal is normally transformer-coupled but this is not necessary for the unbalanced signal. Figure 18 illustrates the standard interconnection circuit. The output voltage level is 1.0 $V_{p\text{-}p}$ with an impedance of 75 Ω unbalanced. The signal can be transported across up to 1000 meters of cable. There are a variety of circuits that allow interconnection between XLR and BNC interfaces either using simple resistor networks or using a transformer and attenuator circuit.

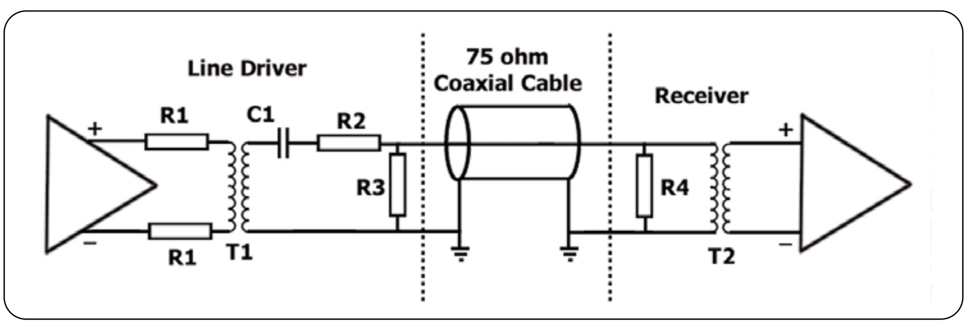
Setting Levels

There are several differences to understand when setting levels and interpreting the digital audio signal. The maximum digital audio value is represented by an audio sample of all 1's that is referred to as 0 dBFS (decibels Full Scale). Clipping may occur if the original analog audio signal exceeds this value and produces distortions in the audio signal. Additionally the digital audio signal may produce a high analog amplitude when converted back to the analog domain. This is because the low pass filter that is added to the analog output stage of the conversion process gives rise to higher analog amplitude signal level than the digital value represents. Therefore, most audio monitors offer an interpolated view of the audio signal to show these peak values and provide indication of when clipping occurs.

Besides the VU and PPM meter ballistics available on an analog audio monitor, digital monitors offer a True Peak Meter. This type of metering displays actual signal peaks regardless of their duration with an almost instantaneous response. Within any of the meter ballistics, the user can choose the appropriate reference level and peak level to meet his specific requirements. A practical implementation would be to choose a test level of –18 dBFS and peak program level of –8 dBFS within the configuration of the instrument. The level meter display provides indication of the test level by a yellow diamond shape and the peak program level by a red diamond positioned at the side of the bar display to allow the operator to easily interpret the level meter. When the audio signal is below the test level reference, a green bar is displayed, a yellow bar indicates the signal is above the test reference, and a red bar indicates



▶ ***Figure 17.*** *AES3 Interconnection Circuit.*



▶ ***Figure 18.*** *AES3-ID interconnectivity.*

a signal above peak level. Figure 19 shows a typical display from the WFM700 with digital audio option DG.

The display will also provide indication of clips, mutes, and two thresholds that can be programmed by the user.

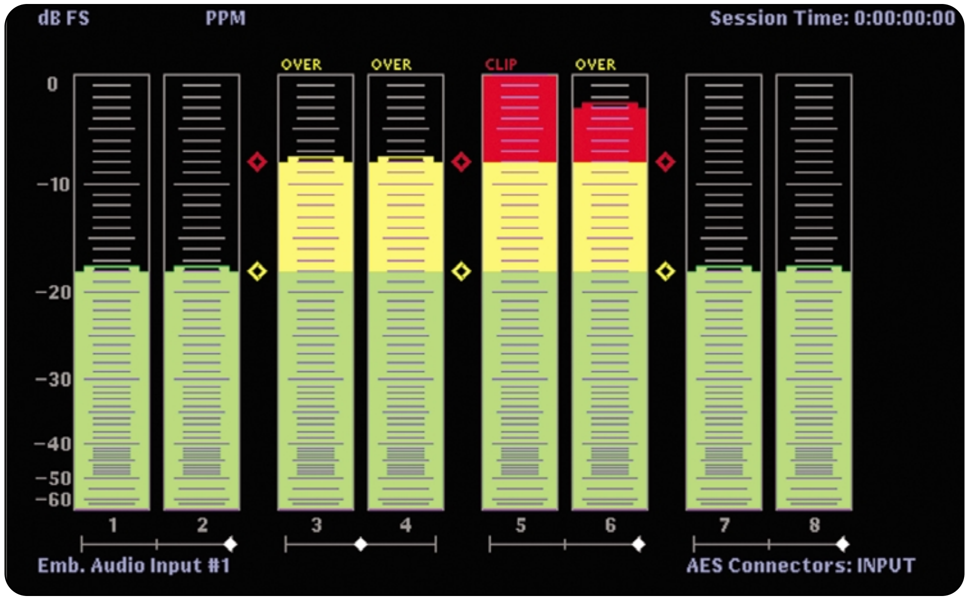
Clips, Mutes, Over Indications

A Clip condition occurs when the audio sample data is 0 dBFS for a number of consecutive samples. When this happens, an indicator will be displayed in the level display for a period of time set by the user.

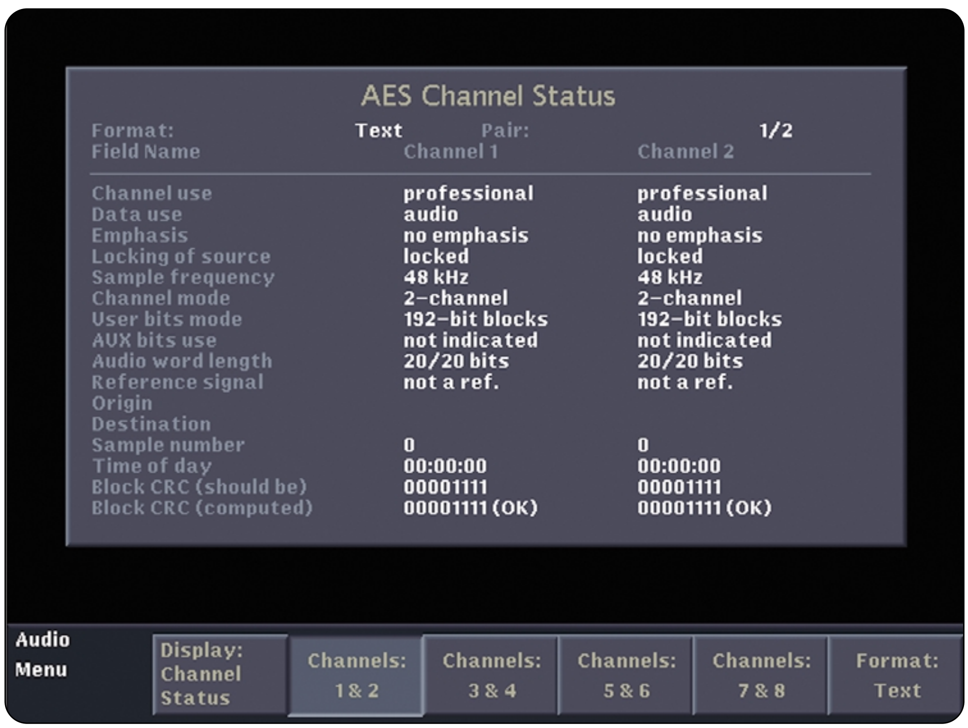
A Mute condition is indicated within the bar display and occurs when the digital audio data of the channel remains at zero-value for a number of consecutive samples. An Over condition occurs when the audio signal exceeds a level set by the user. Similarly, a Silence condition occurs when the audio signal is below the level set by the user. Each of these indications can be logged and displayed on a session screen to indicate the number and type of errors that have happened during the time interval monitored by the user.

Another failure mode in digital audio is that the audio sample data may become corrupt. If this occurs, the validity flag contained within the audio channels subframe should be set to a value of '1' to indicate the error. Audio equipment will then ignore this data and mute the audio signal for this sample. If the user wishes to see all the audio data samples, the audio monitor can be set to ignore the validity bit and use the audio data sample within the signal for its measurement.

Channel status can be displayed in a number of ways. Figure 20 shows an example of a text description generated by the WFM700 with option DG. This interprets the data as defined by the AES/EBU 3 specification. The display can also be configured to look at the specific data values individually in binary (Figure 21), hex, or XMSN binary order.



▶ ***Figure 19.*** *Audio bar level.*



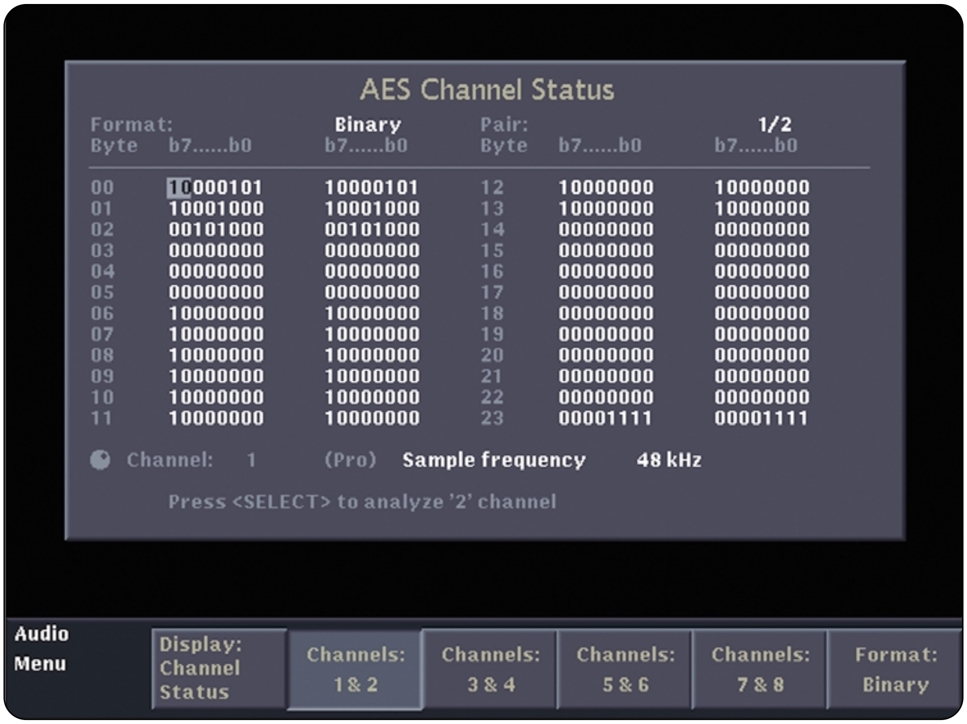
▶ ***Figure 20.*** *AES Channel Status descriptions in text format.*

Embedded Audio SD/HD

The AES/EBU audio data can be embedded into a serial digital video signal within the ancillary data space. This is particularly useful in large systems where separate routing of digital audio becomes a cost consideration and the assurance that the audio is associated with the appropriate video is an advantage. In smaller systems, such as a post-production suite, it is generally more economical to maintain separate audio, thus eliminating the need for numerous embedder and de-embedder modules. We shall concentrate our discussion on the serial digital component signal. Although the standards define methods for embedding digital audio within the digital composite, this is not discussed here. For information on embedding digital audio into a digital composite signal, please refer to ***A Guide to Digital Television Systems and Measurements*** (Literature number 25W-7203-3).

The multiplexing of audio data within the Serial Digital Interface (SDI) is defined by the following standards for standard definition (SD) and high definition (HD) video formats:

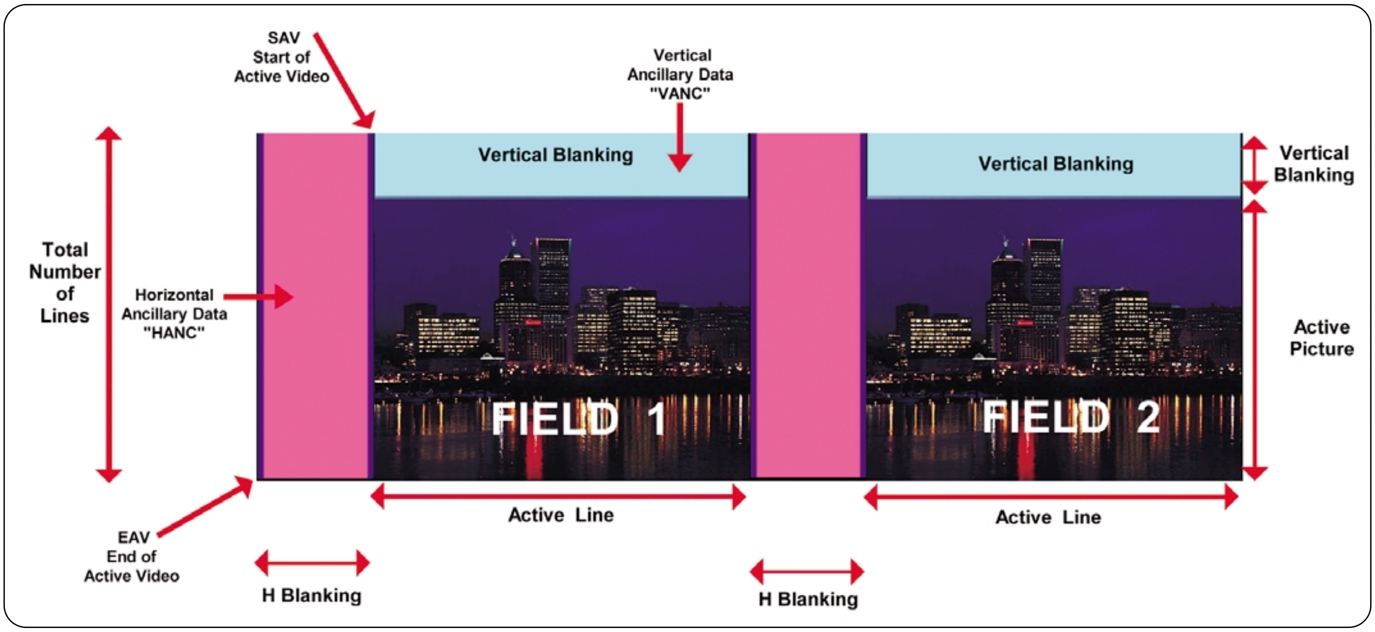
- ***SMPTE 272M Formatting AES/EBU Audio and Auxiliary Data into Digital Video Ancillary Data Space*** for standard definition to be used within SMPTE 259M
- ***SMPTE 299M 24-Bit Digital Audio Format for HDTV Bit Serial Interface*** to be used within SMPTE 292M



▶ ***Figure 21.*** AES Channel Status in binary format.

Ancillary data space

The ancillary data space available in component digital video is shown in Figure 22. All of the horizontal and vertical blanking intervals are available except for the small amount required for EAV (end of active video) and SAV (start of active video) sequences, and, in HD, the line number and CRC information. The ancillary data space has been divided into two types – HANC (horizontal ancillary data) and VANC (vertical ancillary data). In SD, the audio data is divided up into the 20-bit audio samples and the extended auxiliary four bits of data, whereas in HD, the 24 bits of audio data are carried as one packet. Audio data is located in the HANC area for both SD and HD formats. Additional extended data is also carried within HANC for SD systems. In SD, the Cb/Y/Cr/Y' (Y is the co-sited luma sample and Y' is the independent luma sample) carry the audio



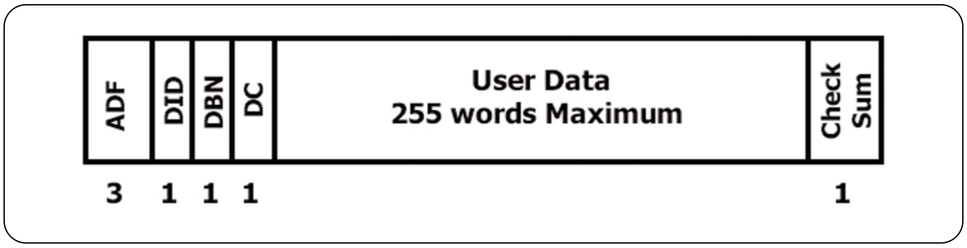
▶ ***Figure 22.*** *Component ancillary data space.*

data whereas in HD format, the Cb/Cr data words are used for audio data and the Y sample is used to carry an optional audio control packet which is transmitted once per field on the second line after the switching point. This may not be required if the audio data rate is using a 48 kHz clock. During the switching point of the appropriate video format, no audio data is present within HANC. In standard definition, no audio is present during the Error Detection Handling insertion.

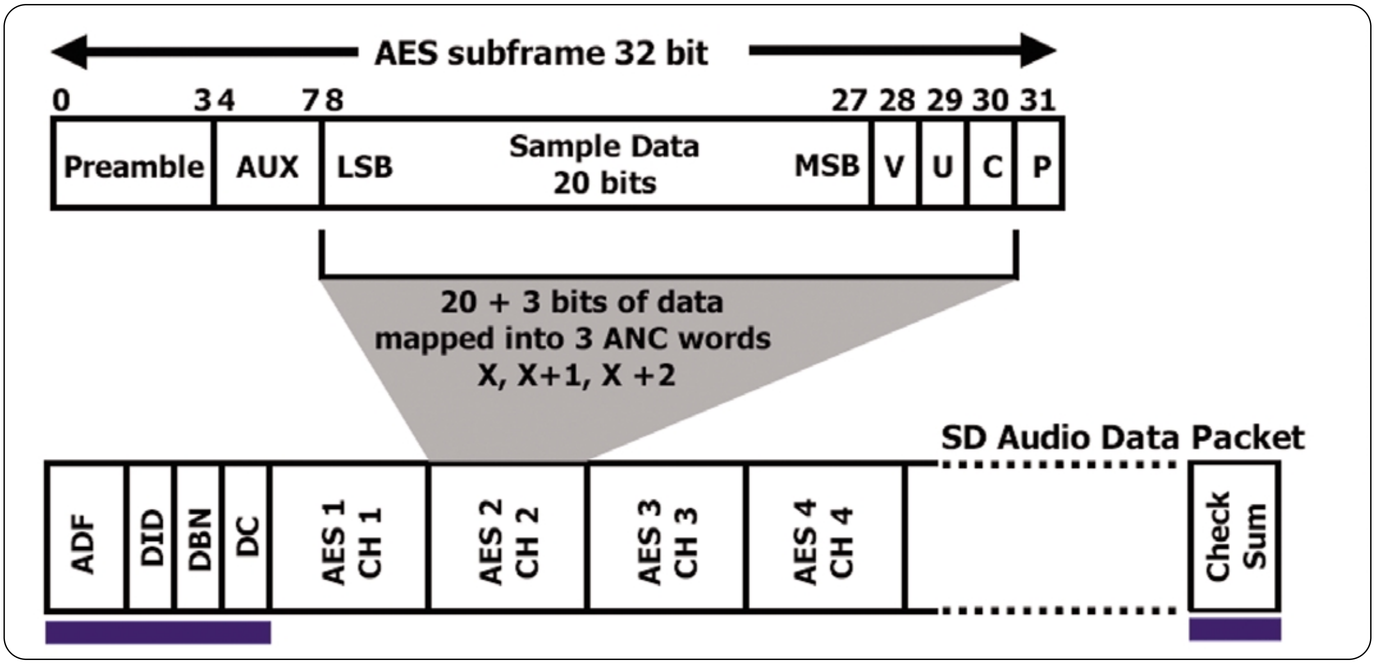
Up to 16 channels of embedded audio are specified for HANC, which is assembled into four groups, each containing four audio data channels.

Ancillary data formatting

Ancillary data is formatted into packets prior to multiplexing the data into the video data stream as shown in Figure 23. Each data block may contain up to 255 user data words and multiple data packets may be placed in individual ancillary data spaces, thus providing a rather flexible data communications channel. At the beginning of each data packet is a header using word values that are excluded for digital video data and reserved for synchronizing purposes. The Ancillary Data Flag (ADF) is a three-word header 000_h, $3FF_h$, $3FF_h$. Each type of data packet is identified with a different one-word Data Identification (DID). Several different DID words are defined to organize the various data packets used for embedded audio. The Data Block Number (DBN) is a one word optional counter that can be used to provide sequential order to ancillary data packets allowing a receiver to determine if data is missing. As an example with embedded audio, the DBN may be used to detect the occurrence of a vertical interval switch, thereby allowing the receiver to process the audio data to remove the likely transient "click" or "pop." Just prior to the data is the Data Count (DC) word indicating the amount of data in the packet. Finally, following the data is a one-word checksum that is used to detect errors in the data packet.



▶ ***Figure 23.*** *Ancillary data formatting.*



▶ ***Figure 24.*** *Basic embedded audio.*

Basic SD embedded audio

Embedded audio defined in SMPTE 272M provides up to 16 channels of 20-bit audio data sampled at 48 kHz with the sample clock locked to the television signal. Although specified in the composite digital part of the standard, the same method is also used for component digital video. This basic embedded audio corresponds to Level A in the embedded audio standard. Other levels of operation provide more channels, other sampling frequencies, and additional information about the audio data. Basic embedded audio data packet formatting is derived from AES audio as shown in Figure 24.

The Audio Data Packet contains one or more audio samples from up to four audio channels. 23 bits (20 audio bits plus the C, U, and V bits) from each AES sub-frame are mapped into three 10-bit video words (X, X+1, X+2) as shown in Table 5. Bit-9 is always not Bit-8 to ensure that none of the excluded word values ($3FF_h$-$3FC_h$ or 003_h-000_h) are used. The Z-bit is set to "1" corresponding to the first frame of the 192-frame AES block. Channels of embedded audio are essentially independent (although they are always transmitted in pairs), so the Z-bit is set to a "1" in each channel even if derived from the same AES source. C, U, and V bits are mapped from the AES signal; however the parity bit is not the AES parity bit. Bit-8 in word X+2 is even parity for bits 0-8 in all three words. There are several restrictions regarding distribution of the audio data packets although there is a "grandfather clause" in the standard to account for older equipment that may not observe all the restrictions. Audio data packets are not transmitted in the horizontal ancillary data space following the normal vertical interval switch as defined in RP 168. They are also not transmitted in the ancillary data space designated for error detection checkwords defined in RP 165. Taking into account these restrictions, "data should be distributed as evenly as possible throughout the video field." For basic Level A, this results in either three or four audio samples per channel in each audio data packet.

Table 5. Embedded Audio Bit Distribution

Bit	X	X+1	X+2
B9	NOT B8	NOT B8	NOT B8
B8	AUD 5	AUD 14	P*
B7	AUD 4	AUD 13	C
B6	AUD 3	AUD 12	U
B5	AUD 2	AUD 11	V
B4	AUD 1	AUD 10	AUD 19 (MSB)
B3	AUD 0	AUD 9	AUD 18
B2	CH 1	AUD 8	AUD 17
B1	CH 2	AUD 7	AUD 16
B0	Z-BIT	AUD 6	AUD 15

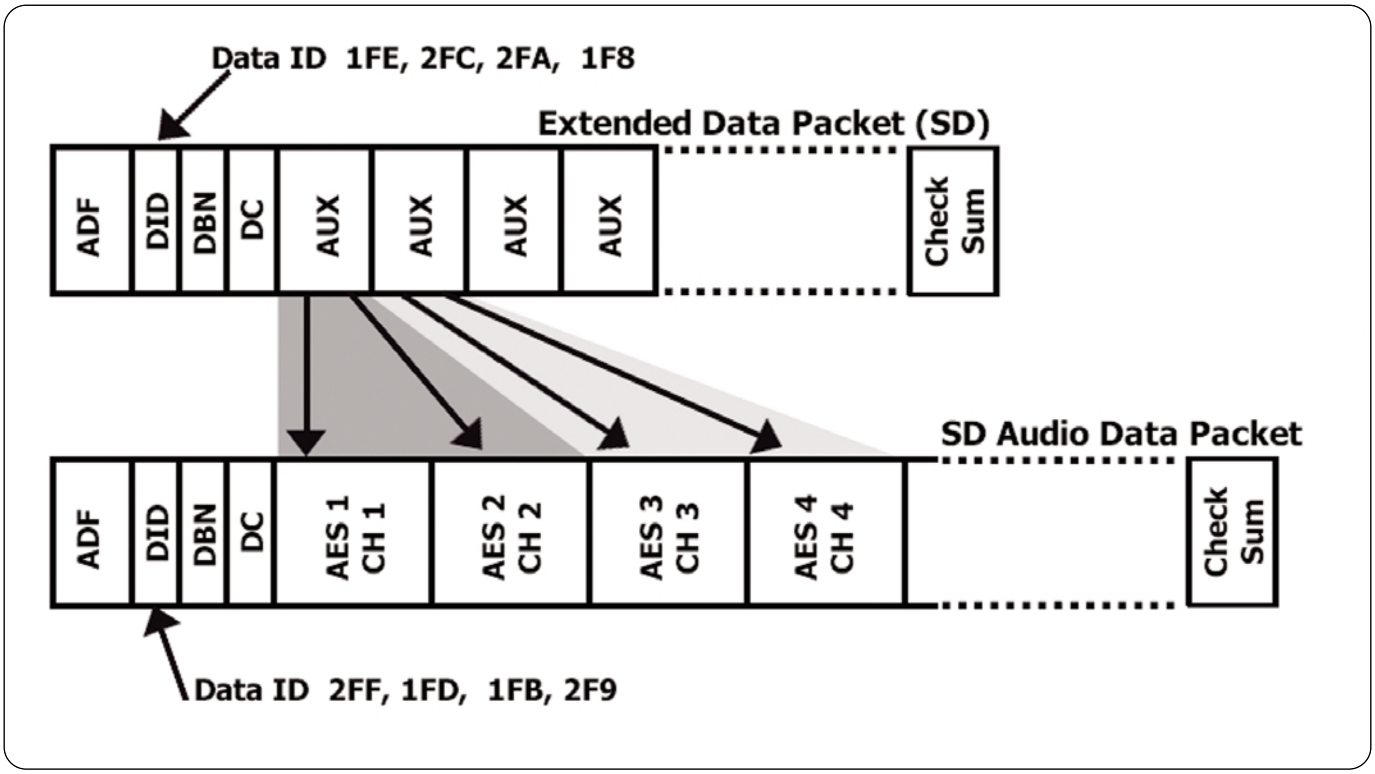
Extended embedded audio

Extended embedded audio provides the following features:

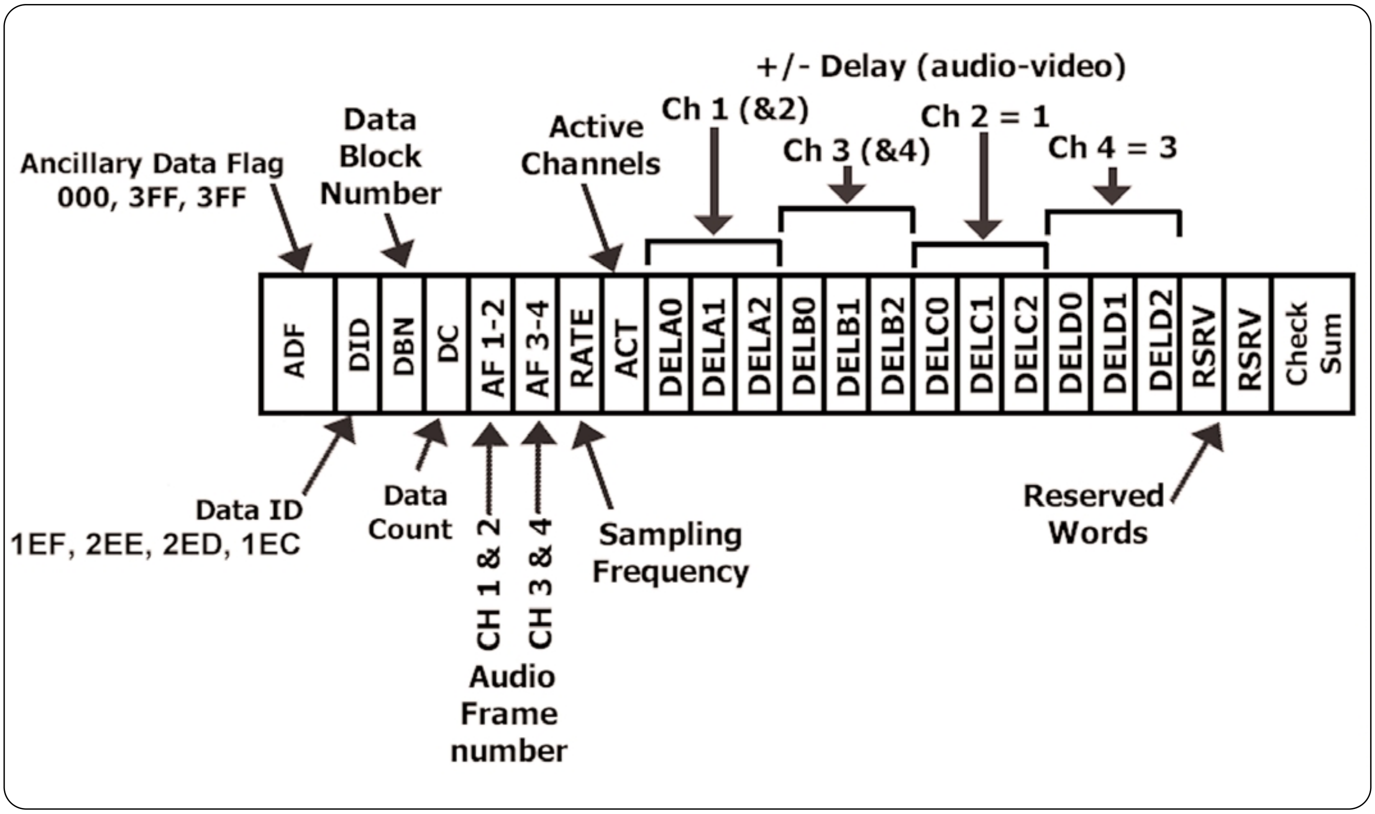
- Carrying the four AES auxiliary bits (which may be used to extend the audio samples to 24-bits)
- Allowing non-synchronous clock operation
- Allowing sampling frequencies other than 48 kHz
- Providing audio-to-video delay information for each channel
- Documenting Data IDs to allow up to 16 channels of audio in component digital systems
- Counting “audio frames” for 525-line systems

To provide these features, two additional data packets are defined. Extended Data Packets carry the four AES auxiliary bits formatted such that one video word contains the auxiliary data for two audio samples as shown in Figure 25. These packets must be located in the same ancillary data space as the associated audio data packets and must follow the audio data packets.

The Audio Control Packet shown in Figure 26 is transmitted once per field in the second horizontal ancillary data space after the vertical interval switch point. It contains information on audio frame number, sampling frequency, active channels, and relative audio-to-video delay of each channel. Transmission of audio control packets is optional for 48 kHz synchronous operation and required for all other modes of operation (because it contains information about which mode is being used). Audio frame numbers are an artifact of 525-line, 29.97 frame/second operation. In that system, there are exactly 8008 audio samples in exactly five frames, which means there are a non-integer number of samples per frame. In PAL 625-line, 25 frames/second, the process is much simpler: there are exactly 1920 audio samples per frame because of the even frame rate. An audio frame sequence is the number of frames for an integer number of audio samples. In a 525/59.94 Hz system, there are five frames and the audio frame number indicates where in the sequence a particular frame belongs. In a 625/50 Hz system, there



▶ ***Figure 25.*** *Extended embedded audio.*



▶ ***Figure 26.*** *Audio control packet format.*

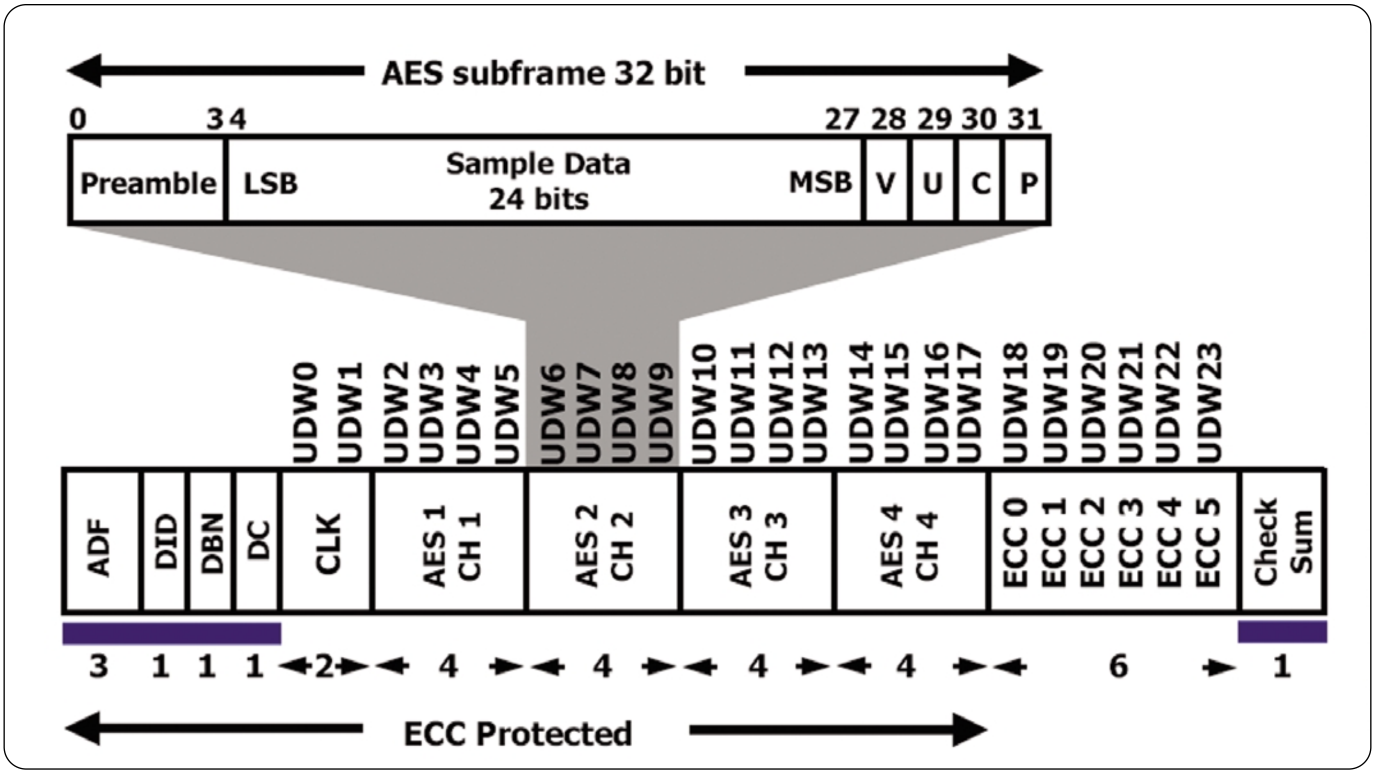
is one frame. The frame number is important when switching between sources because certain equipment, such as digital VTRs, requires consistent synchronous operation to prevent buffer over/under flow. Where frequent switching is planned, receiving equipment can be designed to add or drop a sample following a switch in the four out of five cases where the sequence is broken. The challenge in such a system is to detect that a switch has occurred. This can be facilitated by use of the data block number in the ancillary data structure, and by including an optional frame counter with the unused bits in the audio frame number word of the audio control packet. Audio delay information contained in the audio control packet uses a default channel-pair mode. That is, delay-A (DELA0-2) is for both Channel 1 and Channel 2 unless the delay for Channel 2 is not equal to Channel 1. In that case, the delay for Channel 2 is located in delay-C. Sampling frequency must be the same for each channel in a pair, hence the data in "ACT" provides only two values – one for Channels 1 and 2 and the other for Channels 3 and 4. In order to provide for up to 16 channels of audio in component digital systems, the embedded audio is divided into audio groups corresponding to the basic four-channel operation. Each of the three data packet types is assigned four Data IDs, as shown in Table 6.

Table 6. Data IDs for up to 16 channel operations

Control	Audio Channels	Audio Data Packet	Extended Data Packet	Audi Packet
Group 1	1-4	$2FF_h$	$1FE_h$	$1EF_h$
Group 2	5-8	$1FD_h$	$2FC_h$	$2EE_h$
Group 3	9-12	$1FB_h$	$2FA_h$	$2ED_h$
Group 4	13-16	$2F9_h$	$1F8_h$	$1EC_h$

Receiver buffer size

In component digital video, the receiver buffer in an audio demultiplexer is not a critical issue since there is sufficient data space available. For this reason, the standard requires a receiver buffer of 64 samples per channel with a grandfather clause of 48 samples per channel to warn designers of the limitations in older equipment. In the standard, Level A defines a sample distribution allowing use of a 48 sample-per-channel receiver buffer while other levels generally require the use of the specified 64-sample buffer.



▶ ***Figure 27.*** *Structure of HD audio data packet.*

Basic HD embedded audio

There are some similarities and several differences in the implementation of AES/EBU within an HD environment. The formatting of the ancillary data packets is the same between SD and HD. The information contained within the user data is different because the full 24 bits of audio data are sent as a group and not split-up into 20 bits of audio data and an extended packet containing the four auxiliary bits. Therefore the total number of bits used in HD is 29 bits (compared with 23 bits in SD), the 24 bits of audio data are placed in four ancillary data words along with C, V, U, and Z-bit flag. Additionally, the CLK (Audio Clock Phase Data) and ECC (Error Correction Code) words are added to the packet as shown in Figure 27. Since the full 24 bits of audio data are carried within the user data, there is no extended data packet used within HD.

Conformance to the ancillary data packet structure means that the Ancillary Data Flag (ADF) has a three-word value of 000_h, $3FF_h$, $3FF_h$, per SMPTE 291M. The one-word DID has the following values to identify the appropriate group of audio data as shown in Table 7. DBN is a one-word value for data block number and DC is a one-word data count, which is always 218_h. The User Data Words (UDW) always contain 24 words of data and are structured as shown in Figure 27. The first two words, UDW0 and UDW1, are used for audio clock phase data and provide a means to regenerate the audio sampling clock. The data within these two words provides a count of the number of video clocks between the first word of EAV and the video sample corresponding to the audio sample.

Table 7. Data IDs for up to 16 channel operations

	Audio Channels	Audio Data Packet	Audio Control Packet
Group 1	1-4	$2E7_h$	$1E3_h$
Group 2	5-8	$1E6_h$	$2E2_h$
Group 3	9-12	$1E5_h$	$2E1_h$
Group 4	13-16	$2E4_h$	$1E0_h$

Each audio data subframe is distributed across four UDW samples as described in Table 8.

Note that the full preamble data is not carried within the four words, only a reference to the start of the 192 frame by use of the Z-bit indicator. Also, the parity bit is that used within the 32-bit subframe, unlike standard definition.

The ECC is a set of six words that are used to detect errors within the first 24 words from ADF to UDW17. The value is calculated by applying the eight bits of data B0 to B7 of the 24 words through a BCH code (an error correction method) information circuit that produces the six words of the error correction code.

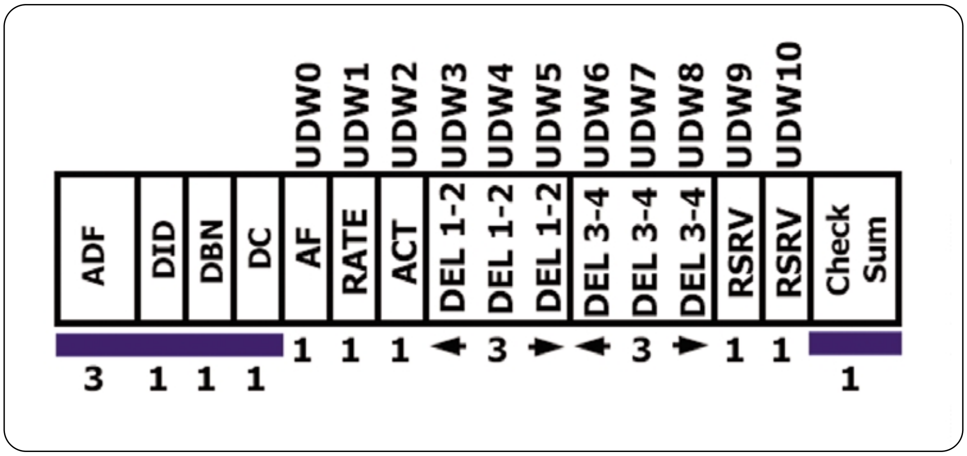


Figure 28. Structure of audio control packet.

The ancillary data information is multiplexed within the color difference Cb/Cr data space only. Unlike the standard definition structure which applies the ancillary audio data across Cb/Y/Cr/Y', the Y data space is only used for the audio control packet that occurs once per field and is placed on the second line after the switching point of the Y data. No ancillary data is placed within the signal on the line subsequent to the switching point. The switching point location is dependent on the format of the high-definition signals, for example in the 1125/60 system no ancillary data is put on line 8.

Audio Control Packet

The audio control packet carries additional information used in the process of decoding the audio data and has a similar structure to standard definition. Its structure is shown in Figure 28 and contains the following information. The Ancillary Data Flag has a three-word value of 000_h, $3FF_h$, $3FF_h$. The one-word DID has the following values to identify the appropriate group of audio data as shown in Table 7. DBN is always 200_h and DC is always $10B_h$. The UDW contains 11 words of data structured into five different types of data. The audio frame number data (AF) provides a sequential number of video frames to assist in indicating the position of the audio samples when using a non-integer number of audio samples per frame. The one word value RATE indicates the sampling rate of the audio data and whether the data is synchronous or asynchronous. The ACT word indicates the number of active channels within the group. DELm-n indicates the amount of accumulated audio processing delay relative to video measured in audio sample intervals for each channel pair 1 & 2 and 3 & 4.

This is a slightly different format than that used in standard definition. The two-word value RSRV is reserved for future use.

Systemizing AES/EBU audio

Serial digital video and audio are becoming commonplace in production and post-production facilities as well as television stations. In many cases, the video and audio are married sources; and it may be desirable to keep them together and treat them as a single serial digital data stream. This has, for one example, the advantage of being able to keep the signals in the digital domain and switch them together with a serial digital video routing switcher. In the occasional instances where it's desirable to break away some of the audio sources, the digital audio can be demultiplexed and switched separately via an AES/EBU digital audio routing switcher. At the receiving end, after the multiplexed audio has passed through a serial digital routing switcher, it may be necessary to extract the audio from the video so that editing, audio sweetening, or other processing can be accomplished. This requires a demultiplexer that strips off the AES/EBU audio from the serial digital video. The output of a typical demultiplexer has a serial digital video BNC as well as connectors for the two-stereo-pair AES/EBU digital audio signals.

Table 8. Bit Assignment of audio data

Bit	UDW2	UDW3	UDW4	UDW5
B9	NOT B8	NOT B8	NOT B8	NOT B8
B8	EVEN PARITY	EVEN PARITY	EVEN PARITY	EVEN PARITY
B7	AUD_1 3	AUD_1 11	AUD_1 19	P_1
B6	AUD_1 2	AUD_1 10	AUD_1 18	C_1
B5	AUD_1 1	AUD_1 9	AUD_1 17	U_1
B4	AUD_1 0	AUD_1 8	AUD_1 16	V_1
B3	Z	AUD_1 7	AUD_1 15	AUD_1 23 (MSB)
B2	0	AUD_1 6	AUD_1 14	AUD_1 22
B1	0	AUD_1 5	AUD_1 13	AUD_1 21
B0	0	AUD_1 4	AUD_1 12	AUD_1 20
Bit	**UDW6**	**UDW7**	**UDW8**	**UDW9**
B9	NOT B8	NOT B8	NOT B8	NOT B8
B8	EVEN PARITY	EVEN PARITY	EVEN PARITY	EVEN PARITY
B7	AUD_2 3	AUD_2 11	AUD_2 19	P_2
B6	AUD_2 2	AUD_2 10	AUD_2 18	C_2
B5	AUD_2 1	AUD_2 9	AUD_2 17	U_2
B4	AUD_2 0	AUD_2 8	AUD_2 16	V_2
B3	0	AUD_2 7	AUD_2 15	AUD_2 23 (MSB)
B2	0	AUD_2 6	AUD_2 14	AUD_2 22
B1	0	AUD_2 5	AUD_2 13	AUD_2 21
B0	0	AUD_2 4	AUD_2 12	AUD_2 20
Bit	**UDW10**	**UDW11**	**UDW12**	**UDW13**
B9	NOT B8	NOT B8	NOT B8	NOT B8
B8	EVEN PARITY	EVEN PARITY	EVEN PARITY	EVEN PARITY
B7	AUD_3 3	AUD_3 11	AUD_3 19	P_3
B6	AUD_3 2	AUD_3 10	AUD_3 18	C_3
B5	AUD_3 1	AUD_3 9	AUD_3 17	U_3
B4	AUD_3 0	AUD_3 8	$AUD_3$16	V_3
B3	Z	AUD_3 7	AUD_3 15	AUD_3 23 (MSB)
B2	0	AUD_3 6	AUD_3 14	AUD_3 22
B1	0	AUD_3 5	AUD_3 13	AUD_3 21
B0	0	AUD_3 4	AUD_3 12	AUD_3 20
Bit	**UDW14**	**UDW15**	**UDW16**	**UDW17**
B9	NOT B8	NOT B8	NOT B8	NOT B8
B8	EVEN PARITY	EVEN PARITY	EVEN PARITY	EVEN PARITY
B7	AUD_4 3	AUD_4 11	AUD_4 19	P_4
B6	AUD_4 2	AUD_4 10	AUD_4 18	C_4
B5	AUD_4 1	AUD_4 9	AUD_4 17	U_4
B4	AUD_4 0	AUD_4 8	AUD_4 16	V_4
B3	0	AUD_4 7	AUD_4 15	AUD_4 23 (MSB)
B2	0	AUD_4 6	AUD_4 14	AUD_4 22
B1	0	AUD_4 5	AUD_4 13	AUD_4 21
B0	0	AUD_4 4	AUD_4 12	AUD_4 20

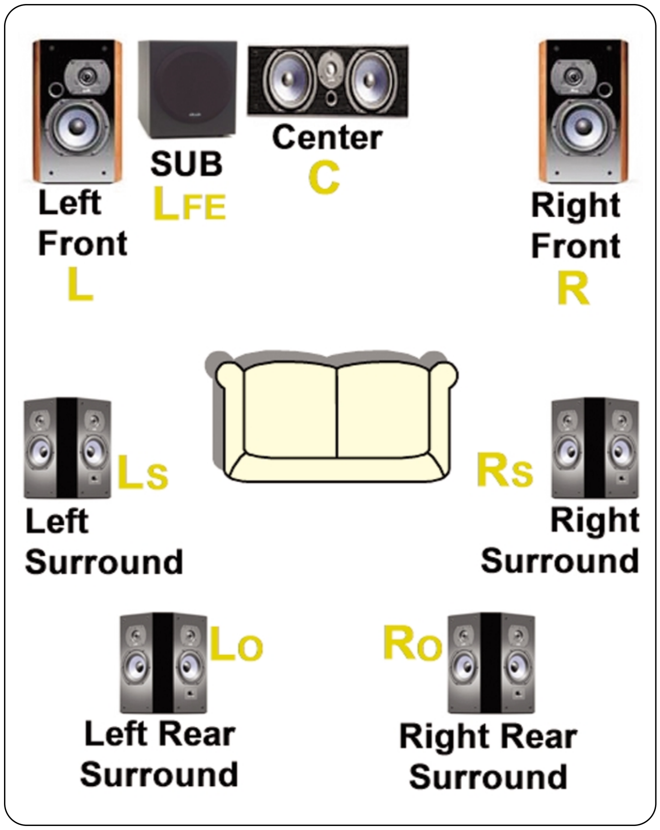
Multiple Channel Audio Monitoring

Initially, it was important to ensure the correct balance of the stereo mix and the lissajous display helped in defining the overall balance of the two audio signals. The advent of multi-channel audio requires the ability to monitor various audio pairs to ensure that they are correctly balanced, independently between each channel. This is achieved within the audio monitor by a flexible lissajous display that uses the familiar X-Y display to independently monitor any of the available channels. A surround sound system is normally configured with a center speaker in the middle of the left and right channels forming the front speaker section, and a pair of multiple rear speakers for the left and right surround channels. The addition of a subwoofer orientated in the front speaker system can carry the low frequency effects as shown in Figure 29.

Within the WFM700 and WVR600 series monitors it is possible to configure the audio bar level displays for surround mode audio. Typically the signals coming from a surround sound decoder are transmitted in AES pairs as Left-L/Right-R, Center-C/Sub-LFE, Left Surround-L_s/Right Surround-R_s and Left Back Surround-L_o/Right Back Surround-R_o as shown in Figure 30. Sometimes when a 5.1 surround system is being used, the L_o and R_o channels can be used for Left and Right stereo down-mix of the material.

Flexible Lissajous Display

A standard lissajous display allows you to compare a channel pair: for instance, left versus right, or left surround versus right surround. However no comparison can be made between the center channel and the front left or right channels since those signals are on different pairs. The flexible lissajous display allows users to select any of the available eight channels to compare against another channel, independently of channel pair location.



▶ ***Figure** 29. Typical surround sound configuration.*

The flexible lissajous display is useful in ensuring correct line-up of the various surround channels since you can independently configure the pair of channels you wish to view within the lissajous display. By applying a line-up tone of the same level and frequency to each channel, the phase of each channel can be independently compared to another channel. For instance, it is useful to compare both the front left and right channels against the center channel, or to compare the left front and left surround channel, or similarly the right front and right surround channel. Thus the phase of the channels can be individually compared to ensure correct alignment between all channels as shown in Figure 31.

In this case the custom lissajous pair is selected between the Left and Center channel. Since the lissajous display line is not vertical there is an amplitude difference between the left and center channel. Comparison can be made between each channel and another channel and preset can be configured to quickly recall each of these configurations.

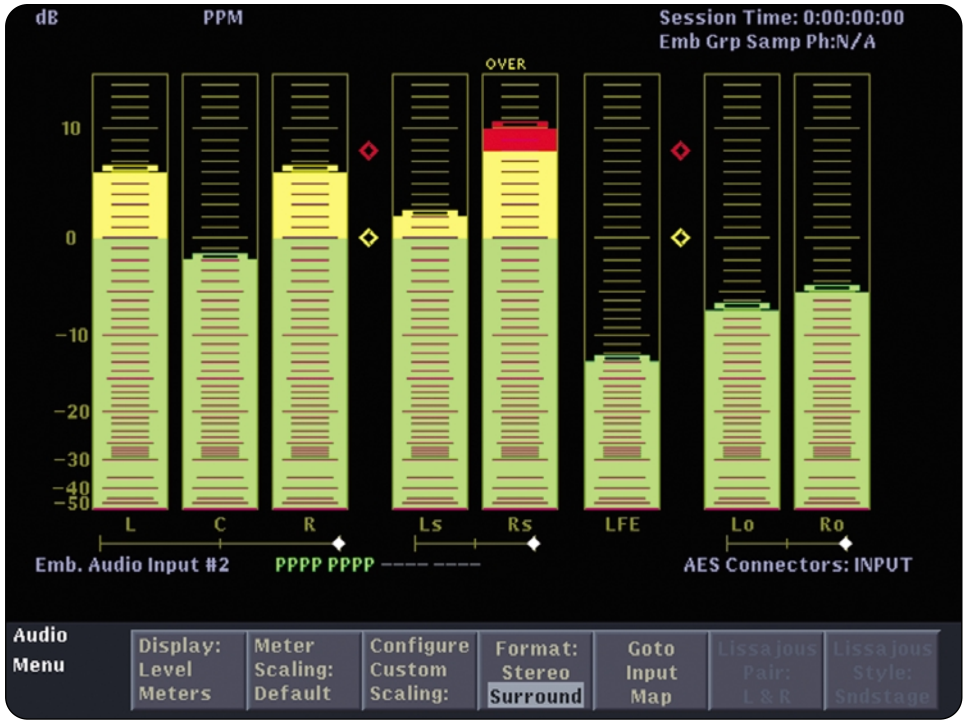


Figure 30. Surround sound level display from WFM700 with option DG.

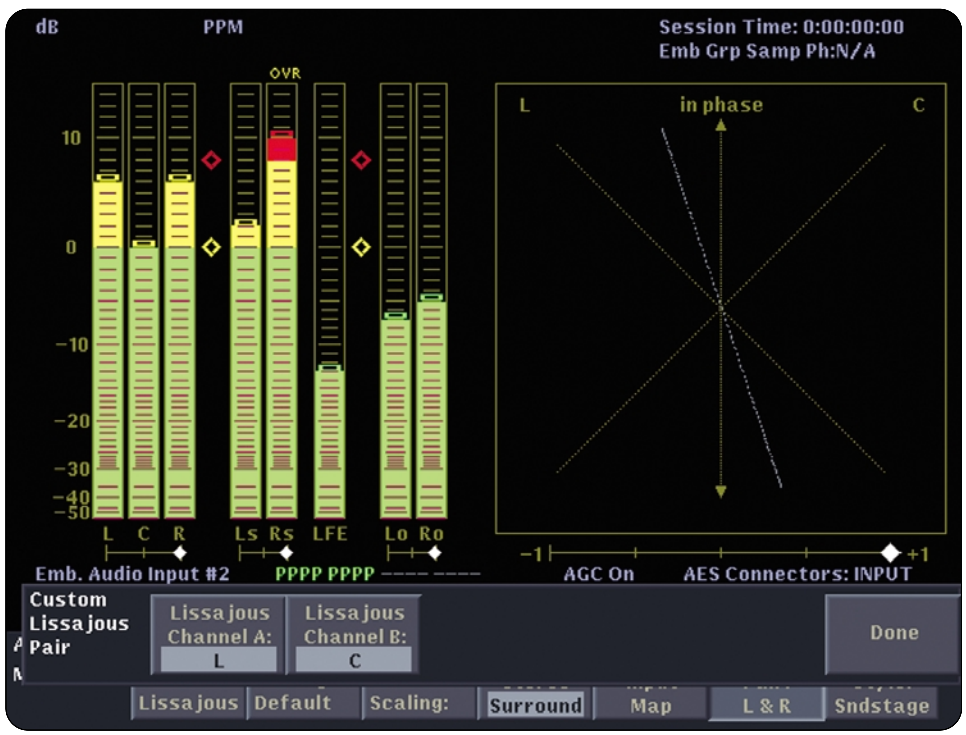


Figure 31. Flexible Lissajous display on WFM700 with option DG comparing the left and center channels.

Conclusion

Audio monitoring tends to get overlooked because manipulation of the video images gets more attention. However, audio is just as important to the quality of the production. Therefore, it is important to constantly monitor the audio signal to ensure that the signal does not suffer from distortion or loss of the audio signal.

Appendix A

Channel Status Protocol

BYTE 0

Bit 0	0	Consumer use of channel status block
	1	Professional use of channel status block
Bit 1	0	Normal Audio mode
	1	Non-Audio Mode
Bits 2-4	**2 3 4**	**Encoded audio signal emphasis**
	0 0 0	Emphasis not indicated
	1 0 0	No emphasis
	1 1 0	50/15 microseconds emphasis
	1 1 1	CCIT J.17 emphasis
		All other states are reserved for future use
Bit 5	**0**	**Default lock condition not indicated**
	1	Source sampling frequency unlocked
Bits 6-7	**6 7**	**Encoded sampling frequency**
	0 0	Sampling frequency not indicated
	0 1	48 kHz sampling frequency
	1 0	44.1 kHz sampling frequency
	1 1	32 kHz sampling frequency

BYTE 1

Bit 0-3	**0 1 2 3**	**Encoded channel mode**
	0 0 0 0	Mode not indicated
	0 0 0 1	Two channel mode
	0 0 1 0	Single channel mode
	0 0 1 1	Primary/Secondary mode
	0 1 0 0	Stereo
	0 1 0 1	Reserved used defined application
	0 1 1 0	Reserved used defined application
	0 1 1 1	Single channel double sampling frequency mode
	1 0 0 0	Single channel double sampling frequency mode stereo mode left
	1 0 0 1	Single channel double sampling frequency mode stereo mode right
	1 1 1 1	Multichannel mode
		All other states are reserved for future use

Bits 4-7	**4 5 6 7**	**Encoded user bits management**
	0 0 0 0	Default no user information indicated
	0 0 0 1	192-bit block structure
	0 0 1 0	Reserved for AES 18 standard
	0 0 1 1	User defined
	0 1 0 0	User data conforms to IEC 60958-3
		All other states are reserved for future use

BYTE 2

Bit 0-2	**0 1 2**	**Encoded use of auxiliary sample bits**	
	0 0 0	Max audio sample word is 20 bits. Use of AUX is not defined	
	0 0 1	Max audio sample word is 24 bits. AUX is used for audio sample data	
	0 1 0	Max audio sample word is 20 bits. Use of AUX channel for addition of audio data channel (e.g. Talkback)	
	0 1 1	Reserved for user defined application	
		All other states are reserved for future use	
Bits 3-5	**0 1 2**	**Encoded audio sample word length**	
	0 0 0	Word length not indicated	
		If audio sample data 24 bits	**If audio sample data 20 bits**
	0 0 1	23 bits	19 bits
	0 1 0	22 bits	18 bits
	0 1 1	21 bits	17 bits
	1 0 0	20 bits	16 bits
	1 0 1	24 bits	20 bits
		All other states are reserved for future use	
Bits 6-7	**0 1**	**Indication of alignment level**	
	0 0	Alignment level not indicated	
	0 1	Alignment level to SMPTE RP155 (20 dB alignment level)	
	1 0	Alignment level to EBU R68 (18.06 dB alignment level)	
	1 1	Reserved for future use	

BYTE 3		
Bit 7	0	Undefined multi-channel mode
	1	Defined multi-channel mode
Bits 0-6		Channel number when Bit 7 = 0
		Channel number is the value of the byte + 1
Bits 4-6	**4 5 6**	**Multi-channel mode when Bit 7 = 1**
	0 0 0	Multi-channel mode 0
	1 0 0	Multi-channel mode 4
	0 1 0	Multi-channel mode 2
	1 1 0	Multi-channel mode 3
	1 1 1	User Defined Multi-channel mode
		All other states are reserved for future use
Bits 0-3		**Channel number when Bit 7 = 1**
		The channel number is one plus the numeric value of those bits taken as a binary number

BYTE 4		
Bits 0-1	**0 1**	**Digital audio reference signal**
	0 0	Not a reference signal
	0 1	Grade 1 reference signal
	1 0	Grade 2 reference signal
	1 1	Reserved and not used until further defined
Bit 2		**Reserved**
Bits 3-6	**3 4 5 6**	**Sampling frequency**
	0 0 0 0	Not indicated
	0 0 0 1	24 kHz
	0 0 1 0	96 kHz
	0 0 1 1	192 kHz
	0 1 0 0	Reserved
	0 1 0 1	Reserved
	0 1 1 0	Reserved
	0 1 1 1	Reserved
	1 0 0 0	Reserved
	1 0 0 1	22.05 kHz
	1 0 1 0	88.2 kHz
	1 0 1 1	176.4 kHz
	1 1 0 0	Reserved
	1 1 0 1	Reserved
	1 1 1 0	Reserved
	1 1 1 1	User defined

Bit 7		**Sample frequency scaling flag**
	0	No scaling
	1	Sampling frequency is 1/1.001 times that indicated

BYTE 5	
Bits 0-7	**Reserved and are set to logic level 0 until further defined**

BYTE 6-9	
Bits 0-7	**Alphanumeric channel origin data**

BYTE 10-13	
Bits 0-7	**Alphanumeric channel destination data**

BYTE 14-17	
Bits 0-7	**Local sample address code**

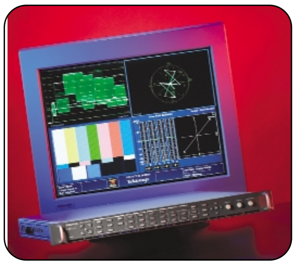
BYTE 18-21	
Bits 0-7	**Time of day sample address code**

BYTE 22	
Bits 0-3	**Reserved; set to logic level 0 until further defined**
	If following data bytes are reliable then following flag set to logic level 1
Bit 4	**Bytes 0 to 5**
Bit 5	**Bytes 6 to 13**
Bit 6	**Bytes 14 to 17**
Bit 7	**Bytes 18 to 21**

BYTE 23	
Bits 0-7	**Channel Status Data Cyclic Redundancy Check**

WFM700 Family of Waveform Monitors

- Monitors and measures HD and SD signals in a single unit
- HD and SD eye pattern measurements and jitter displays
- Configurable/modular architecture



WVR600 Series Waveform Rasterizer

- Monitors SD, SDI, and analog composite in a single unit
- Optional monitoring for analog and/or AES/EBU Audio
- FlexVu™ display capability enables ease of use and flexibility

764 Audio Monitor

- Two balanced or unbalanced AES/EBU or SPDIF loop-through inputs drive four-channel level meters
- Adjustable clip/mute indicators support a variety of technical preferences

TG700 Multiformat Video Generator

- Multiformat analog and digital test signal generation
- Modular expandable platform

Contact Tektronix:

ASEAN / Australasia / Pakistan (65) 6356 3900
Austria +43 2236 8092 262
Belgium +32 (2) 715 89 70
Brazil & South America 55 (11) 3741-8360
Canada 1 (800) 661-5625
Central Europe & Greece +43 2236 8092 301
Denmark +45 44 850 700
Finland +358 (9) 4783 400
France & North Africa +33 (0) 1 69 86 80 34
Germany +49 (221) 94 77 400
Hong Kong (852) 2585-6688
India (91) 80-2275577
Italy +39 (02) 25086 1
Japan 81 (3) 6714-3010
Mexico, Central America & Caribbean 52 (55) 56666-333
The Netherlands +31 (0) 23 569 5555
Norway +47 22 07 07 00
People's Republic of China 86 (10) 6235 1230
Poland +48 (0) 22 521 53 40
Republic of Korea 82 (2) 528-5299
Russia, CIS & The Baltics +358 (9) 4783 400
South Africa +27 11 254 8360
Spain +34 (91) 372 6055
Sweden +46 8 477 6503/4
Taiwan 886 (2) 2722-9622
United Kingdom & Eire +44 (0) 1344 392400
USA 1 (800) 426-2200
USA (Export Sales) 1 (503) 627-1916
For other areas contact Tektronix, Inc. at: 1 (503) 627-7111

For Further Information

Tektronix maintains a comprehensive, constantly expanding collection of application notes, technical briefs and other resources to help engineers working on the cutting edge of technology. Please visit **www.tektronix.com**

12/03 TD/SFI 21W-16463-2

Tektronix
Enabling Innovation