

The New Rules of Engagement: HD Audio Production and Payout

By Tim Carroll, Linear Acoustic

Digital television in Europe is in the process of a major and exciting transformation from standard definition to high definition video, and from stereo to multichannel surround sound. Consumer equipment is affordable and plentiful, thanks to DVD, and consumer demand for surround sound content is on the rise. Watching a sporting event such as The World Cup with exciting surround sound in the comfort of a typical living room will surely bring this demand to a fevered pitch.

How then can a broadcaster produce, distribute and finally deliver this content to consumers? Amazingly, with all of the technology developed in the interest of making things easier, the general trend has headed back to basics, and work flow can remain largely the same.

Infrastructure

The infrastructure goals for modern digital television facilities are the ability to accept content from various sources; routing, storing, and manipulating this content; distributing it to other facilities for further manipulation; and finally transmitting this content to consumers.

The chain can then be broken logically into three major logical sections: **Contribution, Distribution**, and finally **Emission**. These sections have been part of broadcasting since the beginning of television and continue today. Broadcasting of discrete 5.1-channel surround sound in multiple languages and dialects to all consumers along with high-definition video via digital terrestrial transmission (DTT), cable and satellite soon will be a reality. This is a major step forward, but it does not radically change any of the goals - the same things will have to happen regardless of any new service. A properly planned implementation of each of the three major stages can easily support these advanced services.

To satisfy the real needs of carrying many channels of audio between facilities over bandwidth-limited RF links, links with tariffs based on bandwidth (i.e. telephone company), or via VTRs with limited audio channel capacity, mezzanine, or middle-level compression systems are certainly useful and appropriate. But upon receiving such signals or playing them back from a VTR, the signals should be brought back to baseband PCM.

Baseband vs. Compressed

One very important over-riding theme should be carefully observed. It is short-sighted to base facility design exclusively on any single proprietary approach for handling multichannel audio. Fifteen years since it began, most of the leading broadcasters involved in the world-wide transition to digital television and surround sound are naturally drifting towards a common approach within each of the goals prior to transmission: PCM digital audio. It is the most basic, most supported, most flexible and least expensive format. PCM is supported by all digital audio equipment world-wide, and is an ideal goal during design. Further, thanks to large storage dropping substantially in price, it is the heart of most file based systems as well.

Situations such as a VTR or server that can only record four channels of audio, or a link over satellite, microwave, or DS3 - where bandwidth is costly and therefore restricted - all usually require the use of a mezzanine (i.e. middle-level) system such as StreamStacker-HD or Dolby E. Once these mezzanine compressed signals reach a facility, it is very wise to decode them back to baseband PCM audio and metadata and route them as such, or embed them into HD-SDI. SMPTE-standardized methods can place audio in the Horizontal Ancillary (HANC) space and metadata in the Vertical Ancillary (VANC) space of the video signal, therefore keeping all audio, video and metadata in a single, common serial format. The long-standing problems of delay between sound and picture caused by mezzanine compression will not be an issue in a baseband system.

In baseband PCM form, the signals are easily monitored, manipulated, edited, and changed if necessary using commonly available and cost-effective tools. (Note that a standard in-rack monitor device that adds mezzanine compression capabilities does so with nearly a doubling in cost per unit.)

The basic rule for facility audio is: Use mezzanine compression only where necessary, which is generally over links of many types and for storage on legacy formats. Keep audio as baseband or embedded PCM everywhere else for easy, cost effective manipulation.

Getting started - Contribution

This first stage is where content is created and accepted into the signal chain that will eventually lead to the consumer. The content can begin as a studio production, a transfer of a film via telecine, or from a live event. Similar tools are used by each of these different sources, including two-channel to 5.1-channel audio upmixing (i.e. upconversion), metadata authoring to create data describing the audio, plus storage and distribution of the multichannel audio and metadata.

As mentioned above, when audio is accepted or “ingested” into the signal chain, it should ideally be decoded back to standard PCM audio to simplify routing of this program through the facility for further production or editing.

While the goal should always be to produce discrete 5.1-channel audio, there are hundreds of thousands of hours of legacy programs that exist as two-channel stereo or matrix surround encoded (i.e. LtRt) audio. Upmixing is the process of turning these legacy two-channel programs into 5.1-channel surround sound to help maintain good consistency for all viewers.

It is very important at this stage to consider authoring, or creating audio metadata. This “data about the audio data” can be used further down stream to optimize the eventual reproduction of the audio by consumers. While the subject of metadata is outside of the scope of this article (perhaps explored in a future piece), several simple things should be considered. Metadata can be used to control loudness, limit dynamic range, and control the audio so that a single program can be reproduced by mono, stereo and surround listeners simultaneously. Importantly, the system can work both ways: metadata can be individually created to that it accurately matches each program; or static metadata can be used and all programs varied to match these standardized values.

When the audio and metadata content is finally produced, it is time for distributing the signal to the next stage outside of the facility. If a VTR is capable of storing the multichannel baseband PCM and metadata, this is ideal. If the VTR does not offer sufficient channels, or if the signals need to be transmitted via radio or telephone, then mezzanine compression must be employed; this will be discussed below.

Getting to local broadcasters - Distribution

Distribution describes how the program produced during the contribution stage is handed off to local broadcast, satellite and cable facilities for emission (i.e. transmission) to consumers. This stage will likely be accomplished using some sort of link: terrestrial microwave, satellite, DS3 via telephone company (Telco) or fiber. These services use public or government spectrum or other regulated space and each bit comes at a price. It is therefore quite common to find mezzanine compression such as Linear Acoustic StreamStacker-HD or Dolby E. Both systems allow a number of audio channels to be carried over a path that would normally accommodate only two channels.

Once the contribution material arrives at the point of distribution, again ideally it should be decoded to PCM audio and metadata for routing within the plant, most easily as an HD-SDI signal. This scheme will allow for easy and cost-effective manipulation of the content prior to the audio be distributed to the next stage.

Getting to the consumer – Emission (transmission)

Emission, or transmission as it is commonly known, describes the hand-off of programming from the broadcaster to the consumer. This can occur via digital terrestrial television (DTT) transmission, or via digital cable and satellite distribution. In all cases, the final encoding that occurs at here should be identical.

At this stage, baseband audio and metadata are processed to ensure that audio and metadata signals match each other to prevent loudness shifts or other potential problems. Mismatches can be due to metadata being created incorrectly, or to outright metadata failure. If the audio and metadata do not match, one or both of the signals will be adjusted in real time to targets that can either be chosen by the local broadcaster or mandated as standard operating practice for all broadcasters. In this manner, regardless of what may occur upstream, the broadcaster can ensure perfect compliance locally.

While many modern tools are available to support the preceding sections, emission is going through many changes, and there are several choices. This is mostly due to the bitrate of video lowering more and more, thereby placing more demands on the efficiency of the audio. With the advent of H.264/MPEG-4 video, common audio data rates such as 384 kbps start to seem high. Improvements in audio coding commensurate with those of video coding are now available. In particular, the aacPlus system from Coding Technologies is able to deliver 5.1-channel audio at rates as low as 160 kbps.

If 160 kbps sounds low, it should because it is. It is less than half the data rate of other common (but older) emission rate systems and is achievable thanks to recent advances in audio coding. While this number is near the lower end of what may be capable any time soon, the quality is amazingly good. Figure 1 below shows the results of testing completed by the IRT.

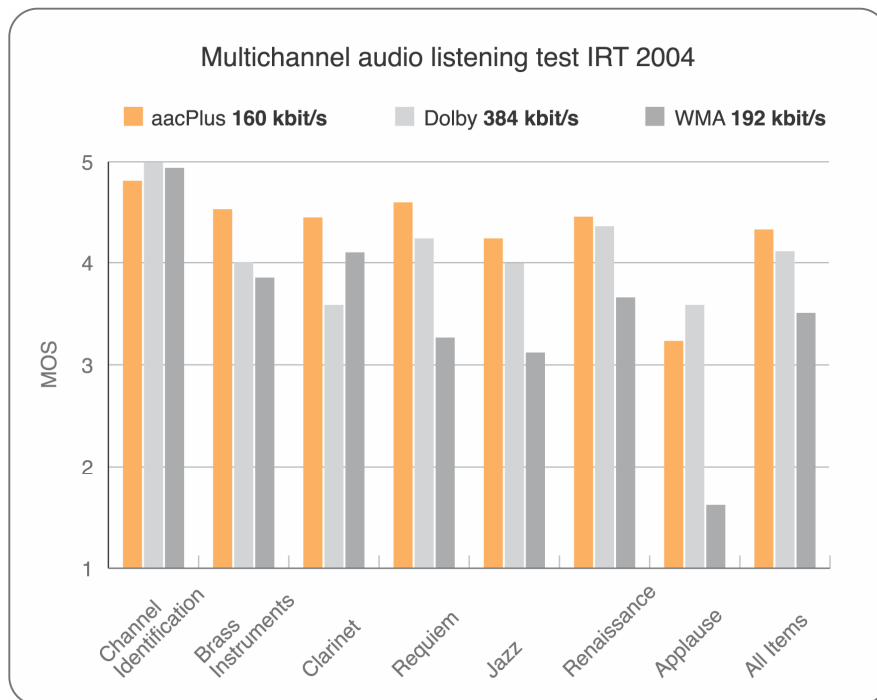


Figure 1 – Results of listening tests performed by the IRT, comparing aacPlus at 160 kbps, Dolby Digital (AC-3) at 384 kbps, and Windows Media at 192 kbps. Note that at less than half the data rate, aacPlus outperformed both other codecs when all results were averaged (Courtesy of Coding Technologies).

Of course, the data rate can also be increased as long as transmission systems can support it, and with the efficiencies provided by new video codecs, this is challenging. Broadcasters now have the ability to match data rate to service requirements. The 5.1 channel primary language version of a cooking channel could be run at 192 kbps for example, while the 5.1 channel alternate language can be set to 160 kbps.

Figure 2 below shows a comparison of video data rates as compared to data rates for two common audio codecs.

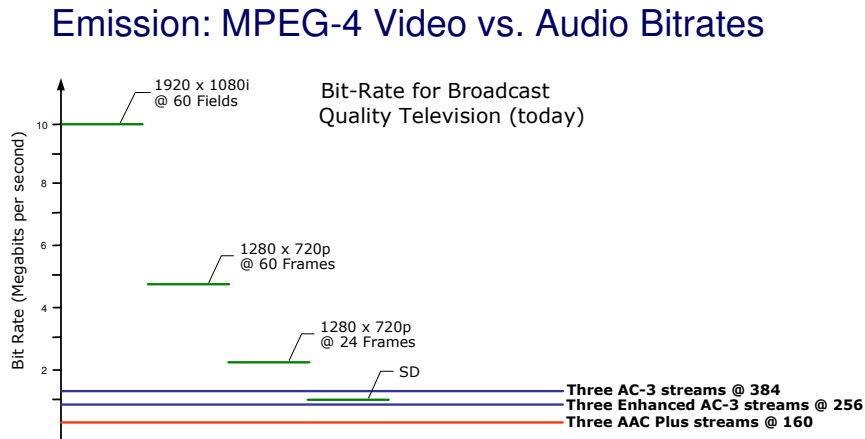


Figure 2 – MPEG-4 video data rates versus audio data rates.

Note that Figure 2 shows audio coding rates that can actually exceed the video data rates - this will not be accepted, so newer more efficient coding systems must be considered. A common question that arises is: "How do we get these newer coding systems connected into legacy A/V receivers?" The answer is simple: transcoding. Enhanced AC-3 can be transcoded to standard AC-3 at 640 kbps, while aacPlus can be transcoded to DTS Coherent Acoustics at 1.536 Mbps. Both formats are in tens of millions of A/V receivers thanks to the popularity of DVD. Figure 3 below shows one example of how this could work.

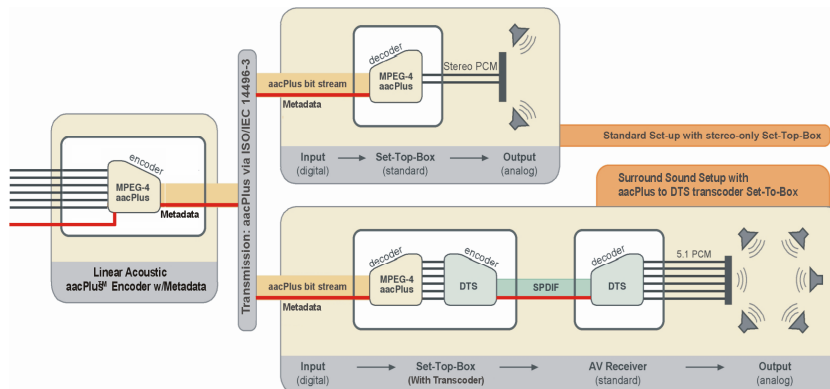


Figure 3 – aacPlus in typical set top box application with DTS transcode for compatibility (courtesy of Coding Technologies).

Summary

Consumers are increasingly demanding 5.1-channel audio. It is very possible to deliver this content using workflows that are common today. New tools and technologies allow this to be done more cost effectively, faster, and better than ever before. New coding technologies now allow audio efficiency to be leveraged to enable more sound to be carried on to consumers in a manner compatible with existing equipment.