Audio in next-generation DVB broadcast systems

Roland Vlaicu
Dolby Laboratories

Broadcasters have significant new requirements for audio delivery in next-generation broadcast systems such as High-Definition Television. These include the capability to deliver soundtracks ranging from mono to 5.1 channels and beyond – with greater efficiency than with current systems, but also to maintain compatibility with existing consumer home cinema systems.

A new audio delivery system, referred to as Enhanced AC-3 (marketing name: Dolby Digital Plus), has been developed to meet these requirements, and has been standardized in DVB and ATSC, referring to ETSI TS 102 366 V1.1.1 (2005-02) [1].

This article discusses the requirements for an audio coding scheme to address both the near- and long-term requirements of DVB broadcasting such as High-Definition Television (HDTV). A number of factors have been considered in determining the characteristics of a suitable audio coding scheme:

- requirements for new DVB services such as HDTV and services using next-generation video codecs such as H.264;
- opportunities for improving the audio performance of current DVB broadcast services;
- impact of a new audio coding scheme on broadcast production practices;
- impact of a new audio coding scheme on consumer hardware and listening environments.

The scope of this article concerns MPEG TS-based DVB services as defined in ETSI standards document TR 101 154 [2].

Requirements for existing DVB services, near and long term

The audio codecs previously specified within TS 101 154 offer solutions for many of the requirements facing DVB broadcasters – requirements for high-quality audio at low bitrates, multichannel audio services, guaranteed connectivity with consumer hardware through existing IEC 61937 interfaces (S/PDIF or Toslink™), and control by a broadcaster over the consumer listening experience through the use of audio metadata. Each codec previously specified in TS 101 154 meets some of these requirements, but there is currently no single codec that meets all of them.

For example, the AC-3 (Dolby Digital) codec can deliver up to 5.1 channels of audio, offers broadcasters full control over the listening experience for all consumer environments through comprehensive metadata control, and offers standardized connectivity via IEC 61937 to around 34.5 million

existing consumer A/V systems. However, the AC-3 codec is not optimized for low bitrate performance. In contrast, the MPEG-4 HE-AAC codec delivers excellent performance at low bitrates, but does not directly offer IEC 61937 connectivity to consumers' existing A/V systems for delivery of 5.1 channel content, and does not offer mandated implementation and interoperability testing of metadata in encoder and decoder products.

When considering the feature set used by current DVB services, a new audio codec should offer at least the following:

- support for mono to 5.1 channel capability;
- comprehensive metadata support, mandated in both the encoder and decoder – all parameters under encoder control:
  - dialogue normalization to ensure consistent listening levels between programmes;
  - downmix to ensure backward compatibility with matrix surround, stereo and mono systems;
  - control of dynamic range to ensure optimal reproduction for all consumer listening environments.
- delivery of discrete 5.1 channel audio to the current installed base of A/V receivers via IEC 61937 interfaces, and support for other emerging digital interface standards;
- improved bitrate efficiency compared with audio codecs currently used in DVB services, in line with the efficiency gains of new video codecs;
- licensing costs and terms in line with existing audio codecs;
- encoder and decoder products should be subject to interoperability testing to ensure consistent performance.

In addition to these requirements for core broadcast services, there is also the opportunity to improve the current provisions for deployment of ***audio description services*** for the visually and hearing impaired. While relative levels between audio description (AD) and main programme services can be controlled both by the broadcaster and the listener, variations in loudness and dynamic range between programmes leads to a need for regular adjustments to listening levels by the consumer.

A new audio codec should meet the following requirements to deliver improved AD services:

- metadata control of dialogue levels to ensure a consistent relative level between main and AD programmes;
- metadata control of the dynamic range of the main programme to ensure that AD services are clearly audible at all times;
- metadata to control the mixing of main programme and AD services in a broadcast receiver should be supported, to remove the need for frequent manual adjustment of levels in the broadcast receiver. Support for mixing AD services with multichannel as well as stereo programme content should be available;

Abbreviations

A/V	Audio / Video (Visual)
AAC	(MPEG) Advanced Audio Coding
AC3	Audio Coding 3, known as Dolby Digital
AD	Audio Description
AES	Audio Engineering Society http://www.aes.org/
ATSC	Advanced Television Systems Committee (USA)
DCT	Discrete Cosine Transform
DVB	Digital Video Broadcasting
E-AC3	Enhanced AC3, known as Dolby Digital Plus
ETSI	European Telecommunication Standards Institute http://pda.etsi.org/pda/queryform.asp
GAQ	Gain Adaptive Quantization
HE-AAC	High Efficiency AAC
IEC	International Electrotechnical Commission http://www.iec.ch/
MPEG	Moving Picture Experts Group http://www.chiariglione.org/mpeg/
S/PDIF	Sony/Philips Digital InterFace
VQ	Vector Quantization

- to simplify implementation in a broadcast receiver, the ability to deliver both main programme and AD services as a single stream that can be decoded and mixed using a single decoder in the broadcast receiver is desirable.

Requirements for new services

Standards and technologies are being developed for the next generation of DVB broadcast services — and a new audio codec must be flexible enough to adapt to these new requirements. Applications such as HDTV and interactive services present new opportunities for audio services. A new audio codec should satisfy at least the following requirements to meet the demands of future broadcast services:

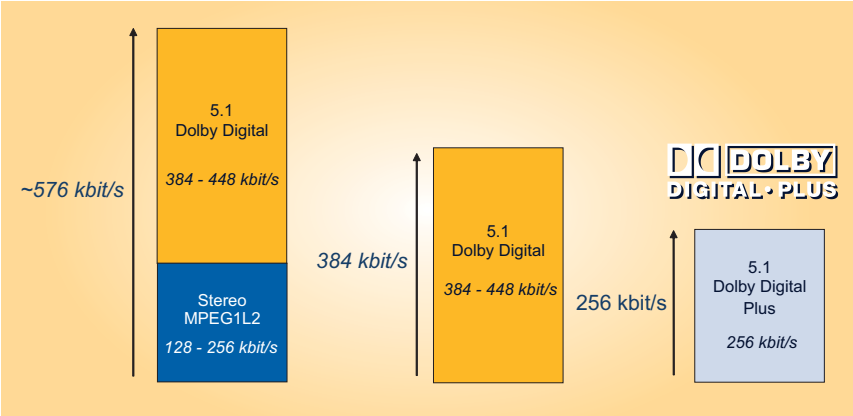

Figure 1
Improved coding efficiency with Dolby Dogital Plus

- be able to deliver audio quality improvements to match the video quality improvements of HD broadcasting;
- flexibility to deliver more than 5.1 channels of audio to match future feature-film mixing formats;
- support for mixing of interactive audio content with main programme audio, including multichannel content.
- deployment of multiple programmes in a single stream, enabling multiple languages, director's commentaries or other advanced services, all controlled by mixing metadata, to be decoded using a single decoder in the broadcast receiver.

Impact on the broadcast production environment

The adoption of a new audio codec for final broadcast should have minimal affect on a broadcaster's working methods, and should certainly not adversely affect them. In the case of stereo audio services, the process of creating audio content does not differ greatly from codec to codec – the selection of encoding settings need be done only once, based upon the target quality of a broadcaster's service and the capability of the codec selected. If the selected codec supports control of programme loudness and dynamics through metadata, this will also need to be factored into the production process.

When considering multichannel audio services, the interfacing of a new emission codec with a broadcast production environment must be carefully considered. Multichannel audio presents a number of challenges to a broadcaster: distribution of content within a broadcast facility equipped for stereo-only content, and the task of creating audio metadata to ensure that 5.1 programming can deliver optimal backward compatibility with all listening environments.

When considering the integration of a new audio codec with a broadcast production environment, it is important that the level of metadata functionality should at least match, and preferably exceed that of current codecs, to ensure that a broadcaster's ability to deliver consistent audio quality is maintained, while allowing flexibility for development of future services. Simple interfacing between multichannel production equipment and transmission encoders for both audio and metadata is desirable.

Impact on consumer products and listening environments

A new audio codec should offer performance improvements for the consumer, while ensuring simple integration into the current consumer listening environment, and offering flexibility for future developments in consumer product design and functionality. To ensure a consistent listening experience for all consumers, a new audio codec should meet the following requirements:

- decoders must maintain compatibility with existing consumer A/V receivers and IEC 61937 interfaces when delivering discrete 5.1 channel content, without introducing excessive complexity to the decoder design.
- to remove the need for audio simulcasting, all decoders must be able to receive and decode, as required, multichannel audio services, in order to deliver either a matrix surround, stereo or mono downmix as required. (The new audio codec should not remove the possibility of audio simulcasting if desired by a broadcaster).
- decoder complexity should be in line with current designs for an equivalent feature set.
- metadata created during the production process must be supported by all decoders. (If this is not the case then the benefits of using metadata are usually lost).
- the new codec should be compatible with emerging and future digital interface standards without introducing excessive complexity to the decoder design.
- licensing costs and terms should be in line with that of existing audio codecs.

A new solution

Considering these requirements, a new audio coding scheme has been developed for use in next generation applications. This coding scheme, referred to as ***Enhanced AC-3*** (E-AC-3) or Dolby Digital Plus, has already been awarded Candidate Standard status within ATSC and has been selected as mandatory technology for HD-DVD players and as optional technology for Blu-Ray players. The scheme has also been standardized by DVB for next generation broadcast services as well as by ETSI in tech. standard TS 102 366 Annex E [1].

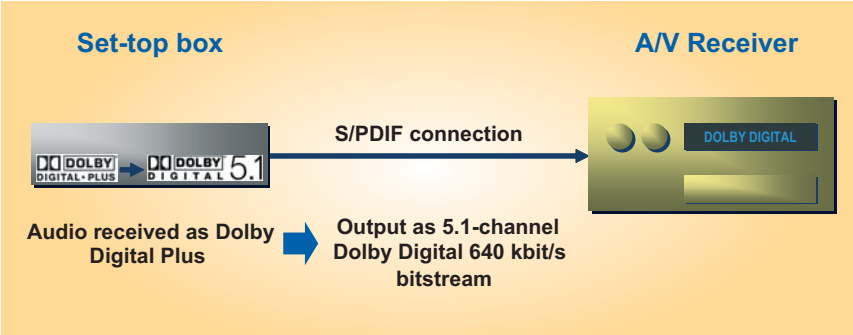

Figure 2
Connection to A/V receiver via S/PDIF

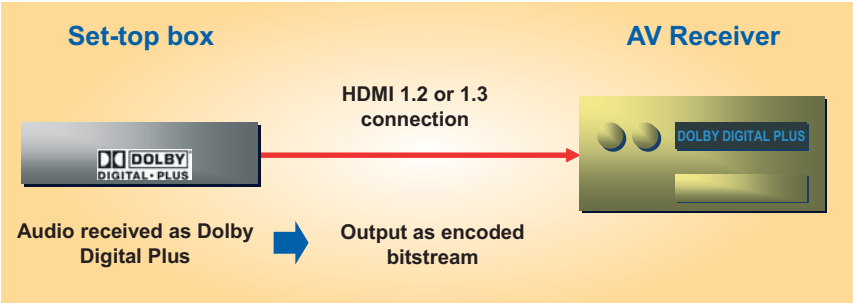

Figure 3
Connection to A/V receiver via HDMI

Applications for which E-AC-3 is well suited include:

- lower data-rate carriage of audio and its conversion to the AC-3 coding standard for playback on today's installed base of audio/video entertainment equipment;
- interactive multimedia capabilities that allow the combination of streamed content with a main audio programme;
- the reproduction of greater than 5.1 channels for the support of playback of existing and future cinema content; and
- the efficient transcoding of AC-3 programme content to lower data-rate E-AC-3 bitstreams and conversion back to AC-3 for playback on the installed base of AC-3 decoders.

Compatible lower data-rate carriage

There are a growing number of applications that require lower data rates, but also require compatibility with the existing broadcast-reception and audio/video decoding infrastructure.

The E-AC-3 system is an excellent solution for these applications because of its inherent lower tandem coding losses than AC-3 and its greater coding efficiency provided by new coding tools. Because of the very large installed base of AC-3 decoders, the E-AC-3 system has been designed to permit a very low-loss conversion to standard AC-3 over a digital audio interconnect such as S/PDIF and decoding by a standard AC-3 decoder.

The conversion stage is a special form of transcoder that minimizes quality degradations resulting from tandem coding losses. This is feasible with the use of the same filterbank, transform block alignment, bit-allocation process, and basic framing structure as conventional AC-3.

A next-generation STB application

The next-generation television application is very similar to the conventional AC-3 reception paradigm except that the need for greater channel capacity requires the transmission of audio programming at lower data rates than is typical for AC-3 applications. Traditionally, AC-3 has been employed at 128 – 192 kbit/s for stereo and 384 – 448 kbit/s for 5.1-channel applications. The use of the new coding tools in E-AC-3 allows for the practical use of lower data rates while permitting efficient conversion into a conventional AC-3 bitstream at 640 kbit/s for compatibility with existing home theatres. *Fig. 4* shows the configuration of this converter/decoder.

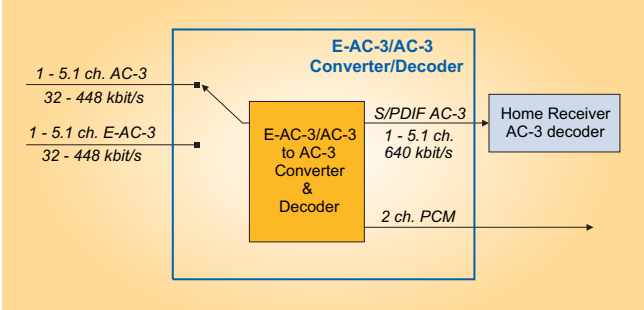

Figure 4
Television set-top box converter/decoder *(existing AC-3 devices shown in light blue)* **[3]**

The device shown in *Fig. 4* accepts both the E-AC-3 and AC-3 bitstreams but always outputs standard AC-3 as the consumer switches channels within the network. The greater efficiency of the E-AC-3 system allows for a greater number of programmes within the broadcast system while preserving full functionality for legacy receiver hardware, shown in the diagram with a light-blue infill.

New coding tools for greater efficiency

Coding efficiency has been increased to allow for the beneficial use of lower data rates. This is accomplished using an improved filterbank, improved quantization, enhanced channel coupling, spectral extension, and a technique called ***transient pre-noise processing***, designed to reduce excess noise before transients.

The improved filterbank is a result of the addition of a second-stage Discrete Cosine Transform (DCT) after the existing AC-3 filterbank, when audio with stationary characteristics is present. This converts the six 256-coefficient transform blocks into a single 1536-coefficient hybrid transform block with increased frequency resolution.

This increased frequency resolution is combined with 6-dimensional Vector Quantization (VQ) and Gain Adaptive Quantization (GAQ) to improve the coding efficiency for “hard to code” signals such as pitch pipe and harpsichord. VQ is used to efficiently code frequency bands requiring lower accuracies, while GAQ provides greater efficiency when higher accuracy quantization is required.

Improved coding efficiency is also obtained through the use of channel coupling with phase preservation. This method expands on the AC-3 method of employing a high-frequency mono composite

Roland Vlaicu graduated from the Technical University of Munich, Germany in 1997 with the German equivalent to a Master's Degree (Dipl. -Ing. univ.), specialising in MPEG-4 video compression. During his studies he spent a semester in the UK working with the University of Cambridge on an Internship project. He also worked part-time from 1990 at ProSieben Television, Germany's second largest commercial broadcaster which he then joined full-time in 1998 to work as a systems engineer in the department for distribution networks.

After having worked as a project manager on the introduction of Dolby Digital 5.1 sound to ProSieben's main service, Mr Vlaicu joined Dolby Laboratories in July 2000 as a Broadcast Consultant. His current tasks include working with broadcasters in Europe and Asia on the implementation of Dolby Digital in new and existing Digital TV services. He was also a key player in the successful launch of the Dolby E format, which won Dolby a Prime-Time Emmy Award in 2005.

channel which reconstitutes the high-frequency portion of each channel on decoding. The addition of phase information, and encoder-controlled processing of spectral amplitude information sent in the bitstream, improves the fidelity of this process so that the mono composite channel can be extended to lower frequencies than was previously possible. This decreases the effective bandwidth encoded, and thus increases the coding efficiency.

Another powerful tool added is ***spectral extension***. This method builds on the channel coupling concept by replacing the upper frequency transform coefficients with lower frequency spectral segments translated up in frequency. The spectral characteristics of the translated segments are matched to the original through spectral modulation of the transform coefficients.

The aformentioned “transient pre-noise processing” is a post-decoding process that reduces pre-noise error via time-scaling synthesis techniques which reduce pre-noise duration and therefore audibility of transient artefacts common to bitrate-reduction audio coders. The parameters used in the post-decoding time-scaling synthesis processing, which employs auditory scene analysis, are provided by metadata calculated by the encoder and transmitted in the E-AC-3 bitstream.

Conclusions

5.1 surround sound is an important component of the HDTV experience. Existing MPEG-2-based HDTV services in the USA, Europe and Australia already offer 5.1 surround sound using the AC-3 technology as documented by DVB and ATSC. With the move to next-generation video coding systems, broadcasters have many new requirements for audio delivery. The E-AC-3 system has been developed to meet these requirements whilst also maintaining compatibility with around 34.5 million existing consumer home cinema systems.

References

[1] TS 102 366, v1.1.1: **Digital Audio Compression (AC-3, Enhanced AC-3) Standard**
ETSI, February 2005.

[2] TS 101 154, v1.6.1: **DVB: Implementation Guidelines for the use of Video and Audio Coding in Broadcasting Applications**
ETSI, January 2005.

[3] Louis D. Fielder et al.: **Introduction to Dolby Digital Plus, an Enhancement to the Dolby Digital Coding System**
117th AES Convention, October 2004.