Fraunhofer IIS

WHITE PAPER

THE AAC AUDIO CODING FAMILY FOR BROADCAST AND CABLE TV

Over the last few years, the AAC audio codec family has played an increasingly important role as an enabling technology for state-of-the-art multimedia systems. The codecs combine high audio quality with very low bit-rates, allowing for an impressive audio experience even over channels with limited bandwidth, such as those in broadcasting or mobile multimedia streaming. The AAC audio coding family also includes special low delay versions, which allow high quality two way communication.

Fraunhofer Institute for Integrated Circuits IIS

Director
Prof. Dr.-Ing. Albert Heuberger
Am Wolfsmantel 33
91058 Erlangen
www.iis.fraunhofer.de

Contact
Matthias Rose
Phone +49 9131 776-6175
matthias.rose@iis.fraunhofer.de

Contact USA
Fraunhofer USA, Inc.
Digital Media Technologies*
Phone +1 408 573 9900
codecs@dmt.fraunhofer.org

Contact China
Toni Fiedler
Phone +86 138 1165 4675
china@iis.fraunhofer.de

* Fraunhofer USA Digital Media Technologies, a division of Fraunhofer USA, Inc., promotes and supports the products of Fraunhofer IIS in the U. S.

A BRIEF HISTORY AND OVERVIEW OF THE MPEG ADVANCED AUDIO CODING FAMILY

The first version of Advanced Audio Coding (AAC) was standardized in 1994 as part of the MPEG-2 standard. Based on the experience of the development of MPEG Layer 3 (better known as mp3) and other proprietary codecs, the main contributors, AT&T, Dolby, Fraunhofer IIS and Sony started from scratch to design a new state-of-the-art audio codec. Compared to MPEG-2, the MPEG-4 standard defines a toolbox of coding tools, which can be combined to best match the application requirements. A number of these combinations for different applications are defined in the MPEG-4 standard as profiles. The MPEG-2 AAC codec was extended in the MPEG-4 standard by the addition of Perceptual Noise Shaping (PNS), Spectral Band Replication (SBR) and the Parametric Stereo (PS) tools. The basic MPEG-4 AAC profile is the "AAC Profile", which is commonly referred to as AAC-LC (low complexity). The most prominent application of this profile is Apple iTunes. The combination of AAC-LC with the SBR tool results in the "High Efficiency AAC profile" (HE-AAC). The SBR tool especially extends the coding efficiency for low bitrates. For a further increase of coding efficiency, HE-AAC can be combined with the PS tool to form the "High Efficiency AAC v2 Profile" (HE-AAC v2). By generating a new profile through adding new tools to an underlying profile, an encoder has all tools available to generate bit streams for the underlying profiles. The same holds true for the decoder. A HE-AAC v2 decoder can decode AAC-LC, HE-AAC and HE-AAC v2 bit streams. Since the additional computational complexity for the SBR and the PS tools are negligible in today`s implementations, all broadcast standards except the Japanese ISDB standard specify the HE-AAC v2 profile to offer broadcasters full flexibility. The Japanese ISDB standard was finalized before the MPEG-4 standardization work was finished and therefore is based on the MPEG-2 AAC standard. The "Low Delay AAC Profile" (AAC-LD) and "Enhanced Low Delay AAC Profile" (AAC-ELD) were developed for communication applications, and also for some types of contribution links. With a coding delay down to 15 ms for AAC-ELD, these codecs still offer remarkable bitrate efficiency. For highest quality and archival purposes, the MPEG-4 "High Definition AAC Profile" (HD-AAC) offers scalable, lossless audio compression.

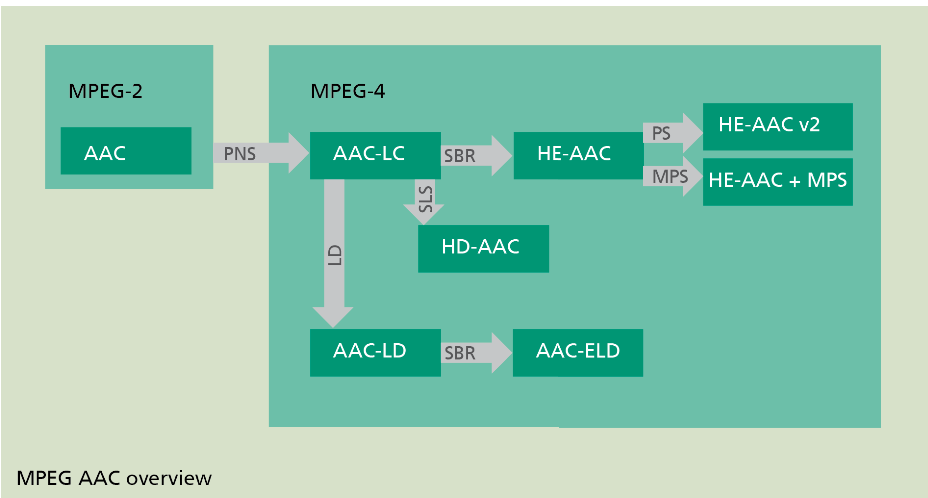

Figure 1: Overview of MPEG AAC audio coding family.

The following sections provide brief descriptions of the codecs of the AAC family that are most relevant to broadcast and cable television applications.

THE MOST COMMON PROFILES

These sections provide information on the most commonly used MPEG-4 AAC profiles. All references to AAC in the following sections are related to the MPEG-4 standard.

AAC-LC: MPEG-4 "AAC Profile"

The "AAC Profile", usually referred to simply as AAC or as AAC-LC (low complexity), provides audio quality up to transparency. In the audio coding domain, transparency stands for an audio quality that even for expert listeners, so-called "golden ears", is indistinguishable from the original, although the original and the coded signal are not mathematically equivalent. Therefore AAC-LC fulfills even the highest quality requirements for broadcasters. Typical bit rates for AAC-LC are 128-192 kbps for stereo and 320 kbps for 5.1 multichannel signals, all encoded as discrete channels. With sampling rates ranging from 8 kHz up to 192 kHz, bit-rates up to 256 kbps per channel, and support for up to 48 channels, AAC-LC is one of the most flexible audio codecs available today. The AAC-LC format includes ancillary data fields that provide a mechanism to transport additional data with the coded audio signal. This feature is used, for example, to transport the SBR data within the AAC bit stream (see next section). Another use for ancillary data is the transport of audio-specific metadata. Since an AAC decoder ignores the data part in the ancillary data fields it cannot decode, these fields allow the addition of new features in the future without compromising the existing receiver population.

HE-AAC: MPEG-4 "High Efficiency AAC Profile"

The "High Efficiency Profile" (HE-AAC) is the combination of MPEG-4 AAC-LC with the parametric SBR (Spectral Band Replication) tool, which allows a further reduction of overall bitrate while maintaining excellent audio quality. At bit rates below 128 kbps for stereo signals, HE-AAC offers a bit rate reduction of up to 30% compared to AAC-LC at equivalent audio quality. For HE-AAC, the lower part of the audio spectrum is coded with AAC-LC, while the SBR tool encodes the upper part of the spectrum. SBR is a parametric approach that uses the relationship of the lower and upper part of the spectrum for a guided recreation of the whole audio spectrum of the signal. To further reduce the bit rate, the lower, AAC-LC encoded, part is coded with half the sampling frequency of the overall signal. Typical data rates used for HE-AAC are 48-64 kbps for stereo and 160 kbps for 5.1 multichannel signals. Like AAC-LC, HE-AAC supports sampling rates from 8 to 192 kHz and up to 48 channels, as well as audio specific metadata.
HE- AAC is also known as "aacPlus" and "Dolby Pulse". Codecs with these different names are all fully compliant with the HE-AAC specification.

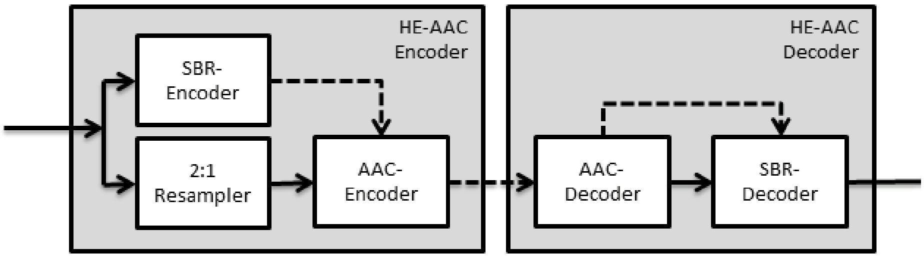

Figure 2: Block diagram of HE-AAC encoder/decoder chain.

HE-AAC v2: MPEG-4 "High Efficiency AAC v2 Profile"

The "High Efficiency AAC v2 Profile" (HE-AAC v2) adds the PS (Parametric Stereo) tool to HE-AAC. This applies a parametric approach to coding the stereo signal and achieves a further reduction in bit rate. Instead of transmitting two channels, the PS encoder extracts parameters from the stereo signal, to enable reconstruction of the stereo signal at the decoder side, and produces a mono downmix, which is HE-AAC encoded. The PS data is transmitted together with the SBR data in the ancillary data fields of the AAC bit stream. The decoder decodes the mono signal and the PS decoder recreates the stereo image. The transmission of the HE-AAC encoded mono signal with parametric data for the stereo image is more efficient than transmission of a two-channel HE-AAC encoded signal. HE-AAC v2 is the most efficient audio coding profile in the AAC audio coding family and typical bit rates used are 24 to 32 kbps for a stereo signal.

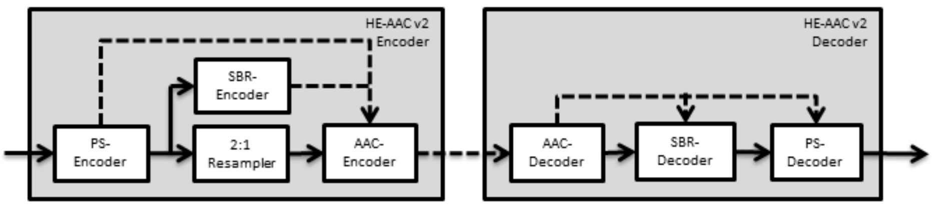

Figure 3: Block diagram of HE-AAC v2 encoder/decoder chain.

Although it would be logical that HE-AAC v2 is only specified for stereo signals, the MPEG-4 standard includes the multichannel option for HE-AAC v2. In case of a multichannel signal, the PS tool is not used and the profile definition of HE-AAC v2 is equivalent to HE-AAC. The reason to specify HE-AAC v2 for multichannel signals is to make it easier for application standards to reference all options of AAC, HE-AAC, and HE-AAC v2. By specifying HE-AAC v2 including multichannel support, all possibilities of the three above described profiles can be utilized.

PERFORMANCE

The audio quality of an audio codec at a given bit rate can only be measured by subjective listening tests. These kinds of tests require a huge effort in time and personnel and therefore are very rarely performed. The last independent multichannel audio codec test was carried out by the European Broadcasting Union (EBU) in 2006/2007. This test included all broadcasting-relevant audio codecs. It also tested the influence of subsequent transcoding, which may be required when using modern codecs like HE-AAC and Dolby Digital Plus in order to maintain backward compatibility with the installed audio-video receiver population. Such receivers may accept multichannel signals only over the S/PDIF input using legacy codecs. For HE-AAC and AAC, the transcoding was done to DTS at 1.5 Mbps, while Dolby Digital Plus was transcoded to Dolby Digital at 640 kbps. The results of the tests are summarized in the chart in Figure 4. This chart clearly shows the excellent quality of AAC at a bit rate of 320 kbps. It also confirms the high coding efficiency of HE-AAC. At a bit rate of 160 kbps the average audio quality over all signals is clearly in the excellent range of the MUSHRA scale. Even at a bit rate of 128 kbps the quality is at the border to the excellent range. As a logical consequence, the report of the EBU states:

"For bitrates equal and higher than 160kbit/s, the average of all ten test items was found to be in the region of "Excellent". This means that the mean value of HE-AAC is similar to the mean value of the above mentioned codecs operating at almost 3-times higher bit-rate!"

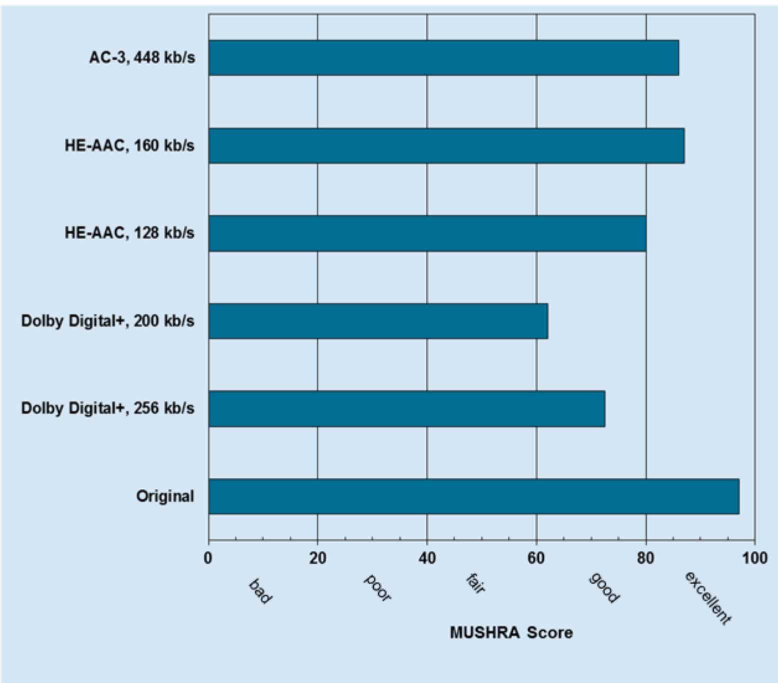

Figure 4: MUSHRA scores of various codecs in EBU evaluation of multi-channel audio codecs (EBU tech 3324; EBU, Geneva, Sept 2007).

The full report is publicly available for download from the EBU web site (http://tech.ebu.ch/docs/tech/tech3324.pdf).

The graph in Figure 5 is based on results of various published tests of the different multichannel audio codecs in the market. It gives a good overview of the performance of the different audio codecs in terms of bit rate and audio quality.

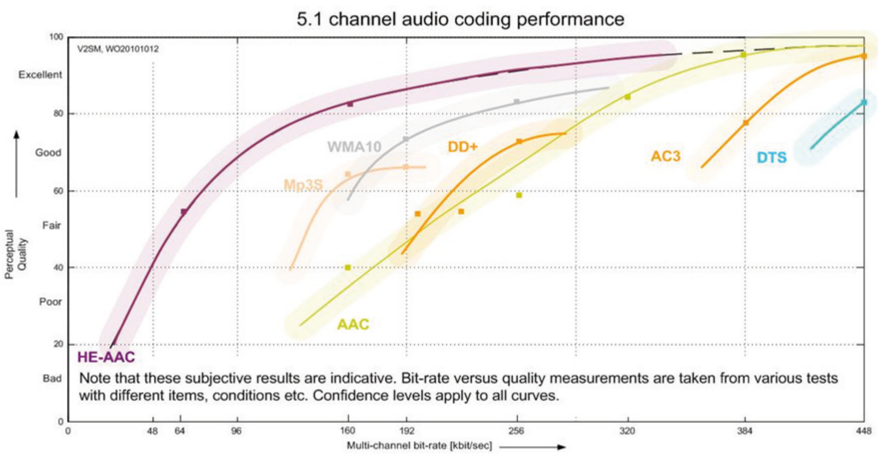

Figure 5: Audio Compression Landscape (figure courtesy of Werner Oomen, Philips)

AUDIO SPECIFIC METADATA

Audio-related metadata for broadcasting systems are typically generated at some point in the content production chain or are part of a pre-encoded delivery. The metadata can be conveyed alongside the coded audio. Three features of metadata are of particular importance and are frequently referred to as the “3 Ds”:

- Dialogue Normalization is used to adjust and achieve a constant long-term average level of the main program components across various program materials, e.g. a feature film interspersed by commercials.
- Dynamic Range Control (DRC) facilitates control of the final dynamic range of the audio and adjusts compression to suit individual listening requirements.
- Downmix maps the channels of a multi-channel signal to the user’s mono or two-channel stereo speaker configuration.

These terms come from the metadata parameters defined for the AC-3 audio codec, used for emission in some digital television systems. They also relate to the Dolby® E[1] audio codec, used in the broadcast production and contribution chain.

The AAC codec supports these same features. The naming convention is slightly different and the following table compares the Dolby® nomenclature of the parameters listed above with their equivalents specified for the AAC codec.

[1] Dolby and Dolby E are registered trademarks of Dolby Laboratories.

Table 1: Nomenclature of AAC metadata in comparison to Dolby Metadata.

	AAC	(E-)AC-3
Loudness Normalization		
	Program Reference Level	Dialnorm
Dynamic Range Control		
Light Compression	Dynamic Range Control	Line Mode
Heavy Compression	Compression Value	RF Mode
Downmix		
	matrix-mixdown / Downmixing Levels	Downmix

The metadata in the (E-) AC-3 format can be translated into the AAC metadata format and vice versa, allowing a seamless integration of the AAC codec into a production chain using Dolby-E or an environment with AC-3 pre-encoded delivery.
A detailed description of the AAC metadata and the translation of the Dolby metadata format into the AAC metadata format can be found in the Fraunhofer IIS Whitepaper "HE-AAC Metadata for Digital Broadcasting". (http://www.iis.fraunhofer.de/content/dam/iis/de/doc/ame/wp/FraunhoferIIS_White-Paper_HE-AAC-Metadata-Broadcasting.pdf)

APPLICATION STANDARDS USING CODECS OF THE AAC AUDIO CODING FAMILY

Broadcast Standards

Digital Video Broadcasting (DVB)
DVB uses a toolbox approach for audio and video coding, describing different audio and video coding algorithms for use within a DVB system, but mandating none. As a consequence each country, region or operator has to define which subset of audio and video codecs needs to be supported for its services. Most DVB-based standard definition services today are using MPEG-2 Layer-2 as the audio codec but the situation is different for the high definition and DVB-T2 services. For these services more recent codecs have been selected. Currently AAC is specified for the following regions and countries:

- United Kingdom
- France
- Spain
- Italy
- Austria
- Nordig (Norway, Sweden, Denmark, Finland and Ireland)
- Israel
- Slovenia

- Malaysia
- New Zealand

Relevant specifications for audio and video coding are:

- ETSI TS 101 154 ("Digital Video Broadcasting (DVB); Specification for the use of Video and Audio Coding in Broadcasting Applications based on the MPEG-2 Transport Stream")
- ETSI TS 102 005 ("Digital Video Broadcasting (DVB); Specification for the use of Video and Audio Coding in DVB services delivered directly over IP protocols")

Advanced Television Standards Committee (ATSC)

The ATSC standard A/153 ("ATSC-Mobile DTV Standard") Part 8 defines the use of HE-AAC v2 for mobile and handheld applications.
The candidate standard ATSC "Non-Real Time" includes, besides other codecs, also HE-AAC for stereo and multichannel sound.

ARIB ISDB

The Japanese TV standard uses MPEG-2 AAC for the fixed services and MPEG-2 AAC in combination with MPEG-2 SBR for mobile services, also known as "1-seg broadcasting". With ISDB-Tmm (mm for mobile multimedia) a new mobile service was launched in April 2012 using part of the spectrum available after the analog TV switch-off in July 2011. This new service uses MPEG-4 HE-AAC v2 together with MPEG Surround for multichannel capability (see description later in this paper). The other part of the released analog TV spectrum is assigned to a service called ISDB-Tsb (sb for sound broadcasting). This service is planned as a digital radio service to eventually replace the analog radio transmissions. Service start is expected in April 2013. Like ISD-Tmm, ISDB-Tsb will use HE-AACv2 together with MPEG Surround.

Sistema Brasileiro de Televisão Digital -SBTVD

SBTVD is based on the Japanese ISDB-T standard and was developed for the Brazilian market. In the meantime it is widely adopted in South America and is the terrestrial digital TV standard for Peru, Argentina, Chile, Venezuela, Ecuador, Costa Rica, Paraguay, Philippines, Bolivia, Nicaragua and Uruguay. The system is specified for fixed and mobile reception and for both services uses MPEG-4 HE-AAC v2.

Digital Audio Broadcasting (DAB)

In the newest version of the DAB standard, the MPEG-2 Layer-2 audio codec was replaced by HE-AAC v2, with MPEG surround multichannel extension. The latest deployment of DAB in Europe is now based on DAB+. DAB+ services are on air in Australia, Germany, Switzerland and several other European countries.

Digital Radio Mondiale (DRM)

Digital Radio Mondiale has selected HE-AAC v2 as the standard codec together with MPEG Surround as a multichannel option. While the DRM30 mode is designed for the frequency range below 30 MHz, DRM+ is the configuration applied in the VHF spectrum above 30 MHz.

Internet and IPTV Standards

Open IPTV Forum, "Release 2 Specification, Volume 2 – Media Formats"
The Open IPTV Forum specification defines HE-AAC as the only mandatory codec for stereo and multichannel audio. Additionally HE-AAC v2 and MPEG Surround are included as optional extensions for HE-AAC.

Alliance for Telecommunication Industry Solutions ATIS-0800013, "Media Formats and Protocols for IPTV Services"
The Alliance for Telecommunication Industry Solutions specifies HE-AAC v2 and MPEG Surround as audio codecs for IPTV services.

Hybrid Broadcast Broadband TV (HbbTV)
The HbbTV specification is based on the Open IPTV Forum specification and defines HE-AAC as the only mandatory codec for stereo and multichannel audio.

Digital Living Network Alliance (DLNA)
The Digital Living Network Alliance has defined AAC as a mandatory audio codec for mobile and handheld devices. For home devices AAC-LC is an optional format.

Applications

AAC and HE-AAC are used in a large number of applications today. Especially in Internet applications, AAC and HE-AAC are established as the main audio codecs besides mp3. Numerous Internet radio stations are using HE-AAC v2 as a main audio codec and an increasing number of download services are offering files in the AAC format. Some of the more prominent applications using AAC and HE-AAC are:

- iTunes
- YouTube
- Sony PS3, Nintendo Wii

ADDITIONAL MEMBERS OF THE AAC AUDIO CODING FAMILY

MPEG Surround

The MPEG Surround technology can be seen as extension of the Parametric Stereo principle from stereo to multichannel. In contrast to the Parametric Stereo tool, MPEG Surround is more scalable in terms of bitrate and quality. MPEG Surround can be combined with the codecs of the AAC family and offers a very high coding efficiency. One other advantage of MPEG Surround is that it is backwards-compatible with stereo signals. The bit stream always includes the AAC encoded core stereo signal as one element and the MPEG Surround description as the second element. As shown in Figure 6, a stereo decoder can extract the core stereo signal and decode it, while an MPEG Surround capable decoder can recreate the full multichannel audio signal. As a result, this allows the use of MPEG Surround in a mixed receiver population with inexpensive or legacy stereo-only as well as multichannel receivers, without the simulcast of a stereo and multichannel signal.

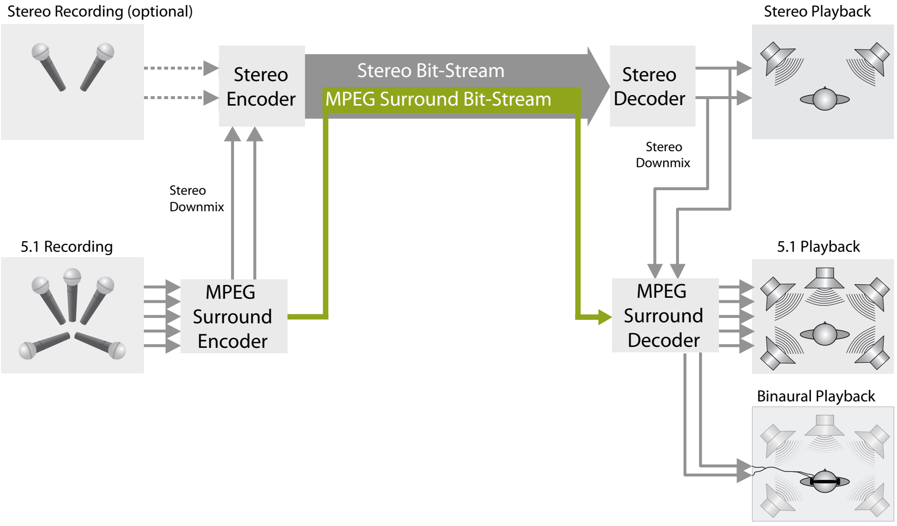

Figure 6: MPEG Surround encoding/decoding scheme.

AAC-LD

AAC-LD is the industry standard for high-quality video conferencing, offering full-bandwidth, low-delay audio coding. It features an algorithmic delay of only 20 ms while offering a good compression ratio and high sound quality for all types of audio signals. It is widely used in high-quality audio and video conferencing systems, bi-directional communication, and Audio over IP applications.

AAC-ELD

AAC-ELD is an enhanced version of AAC-LD combining MPEG-4 AAC-LD and Spectral Band Replication (SBR). AAC-ELD is the best choice for any delay-critical application that demands full audio bandwidth at data rates as low as 24 kbps.

AAC-LD and AAC-ELD are interesting candidates for OTT services in a cable network, which require two way communications. Possible services range from video conferencing to multiplayer games. AAC-LD and AAC-ELD are already in use today for professional and consumer video conferencing applications. For example, Apple's FaceTime application is based on AAC-ELD.

REFERENCES

All AAC family codecs are fully standardized in the following standards:

- MPEG-2 AAC: MPEG-2, Part 7: ISO/IEC 13818-7
- MPEG-4 Audio: MPEG-4, Part 3: ISO/IEC 14496-3
- MPEG Surround: MPEG-D, Part 1: ISO/IEC 23003-1

FURTHER READING

MPEG-4 HE-AAC v2 – audio coding for today's digital media world, Stefan Meltzer and Gerald Moser, Coding Technologies, EBU technical review, January 2006: http://tech.ebu.ch/docs/techreview/trev_305-moser.pdf

EBU Evaluations of Multichannel Audio Codecs, EBU-Tech report 3324, Geneva September 2007: http://tech.ebu.ch/docs/tech/tech3324.pdf

INFORMATION IN THIS DOCUMENT IS PROVIDED ‚AS IS' AND WITHOUT ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, WITHOUT LIMITATION, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE.
INFORMATION IN THIS DOCUMENT IS OWNED AND COPYRIGHTED BY THE FRAUNHOFER-GESELLSCHAFT AND MAY BE CHANGED AND/OR UPDATED AT ANY TIME WITHOUT FURTHER NOTICE. PERMISSION IS HEREBY NOT GRANTED FOR RESALE OR COMMERCIAL USE OF THIS SERVICE, IN WHOLE OR IN PART, NOR BY ITSELF OR INCORPORATED IN ANOTHER PRODUCT.

Copyright ©July 2013 Fraunhofer-Gesellschaft

ABOUT FRAUNHOFER IIS

When it comes to advanced audio technologies for the rapidly evolving media world, Fraunhofer IIS stands alone. For more than 25 years, digital audio technology has been the principle focus of the Audio and Multimedia division of Fraunhofer Institute for Integrated Circuits (IIS). From the creation of mp3 and the co-development of AAC to the future of audio entertainment for broadcast, Fraunhofer IIS brings innovations in sound to reality. Today, technologies such as Fraunhofer Cingo for virtual surround sound, Fraunhofer Symphoria for automotive 3D audio, AAC-ELD for telephone calls with CD-like audio quality, and Dialogue Enhancement that allows television viewers to adjust dialogue volume to suit their personal preferences are among the division's most compelling new developments.

Fraunhofer IIS technologies enable more than 7 billion devices worldwide. The audio codec software and application-specific customizations are licensed to more than 1,000 companies. The division's mp3 and AAC audio codecs are now ubiquitous in mobile multimedia systems.

Fraunhofer IIS is based in Erlangen, Germany and is an institute of Fraunhofer-Gesellschaft. With 23,000 employees worldwide, Fraunhofer-Gesellschaft is comprised of 67 institutes making it Europe's largest research organization.

For more information, contact Matthias Rose, matthias.rose@iis.fraunhofer.de, or visit www.iis.fraunhofer.de/audio.