INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11, MPEG2014/W14936
October 2014, Strasbourg, France

Source **Audio**
Status **Approved**
Title **The AAC-ELD Family for High Quality Communication Services**

Introduction

Since 1994, when the development of AAC in MPEG-2 was initiated, five generations of AAC codecs have evolved. The audio codecs are designed for meeting every possible need in the fields of communications, broadcast, and streaming.

The AAC-ELD family consists of AAC-LD, AAC-ELD, and AAC-ELD v2. The state-of-the-art MPEG-4 audio codecs are designed for maximum speech and audio quality at very low coding delay, and therefore they are all excellent solutions for professional and consumer communication applications.

This paper introduces the three members of the AAC-ELD family and gives a closer look at how to tackle coding delay.

The Members of the AAC-ELD Family

AAC-LD (Low Delay AAC), AAC-ELD (Enhanced Low Delay AAC), and AAC-ELD v2 (Enhanced Low Delay AAC Version 2) are optimized for a low algorithmic delay, which is essential for natural real-time communication. In contrast to common speech codecs, they extend the application area from clean voice to a broad variety of source material, including voice and singing, music and ambient sounds. Due to their technical superiority, the three members of the AAC-ELD family are widely represented across the field of telecommunication, including Over-the-Top (OTT) services, video telephony, video conferencing, and telepresence, as well as broadcast contribution services. The highly successful Apple FaceTime is just one example of a video telephony application that relies on the quality of AAC-ELD. The codec is also natively included in the operating systems iOS, Android, and Mac OS X.

The AAC-ELD family delivers a new level of audio quality, which is called Full-HD Voice. Unlike Plain Old Telephone Services (POTS), Integrated Services for Digital Network (ISDN), and mobile phone calls, Full-HD Voice offers an unsurpassed level of quality, resulting in calls that sound as clear as talking to someone in person.

In addition to the millions of calls already being made with AAC-ELD, this technology is set to enable many new Full-HD Voice applications, including telepresence at home, and mobile rich media telephony.

The three members of the AAC-ELD family can be regarded as a superset of each other, as they share the same coding core, but each adds new coding tools [Figure 1]. The software of the AAC-ELD family can be expected to be fully backward compatible. The codecs can handle mono, as well as stereo and multi-channel signals and run at a wide variety of sampling rates and bit rates (down to 16 kbit/s) -

all with latencies as low as 15 ms.

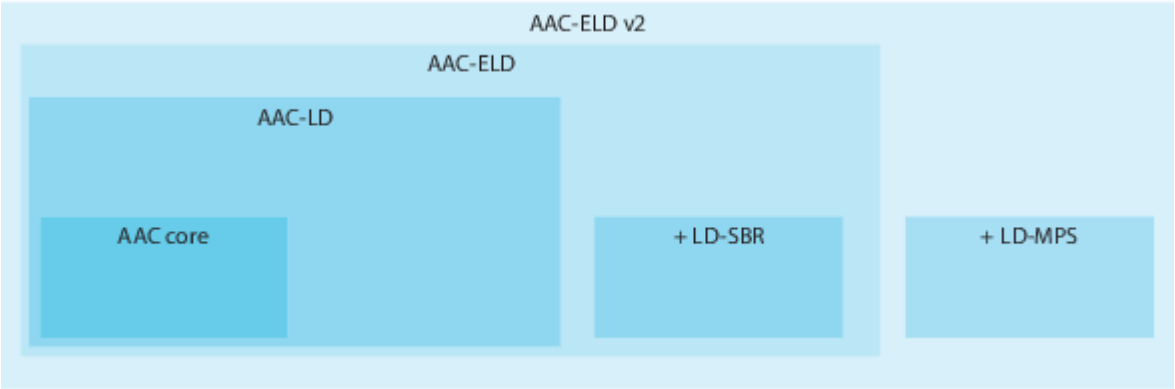

Figure 1: AAC-ELD family. AAC-LD and AAC-ELD share all major tools, only minor differences of the bit stream syntax and filter bank separate AAD-LD from AAC-ELD core. (Image source: Fraunhofer IIS)

The audio quality and operating point of the members of the AAC-ELD family is described in Figure 2 for stereo audio. While AAC-LD is a good choice for stereo bit rates above 96 kbit/s, AAC-ELD improves the audio quality down to 48 kbit/s. Below this bit rate, AAC-ELD v2 is the best choice to keep the stereo audio quality high. For mono applications, a similar relationship between AAC-ELD and AAC-LD at half the bit rate can be expected, whereas AAC-ELD v2 delivers identical audio quality to AAC-ELD.

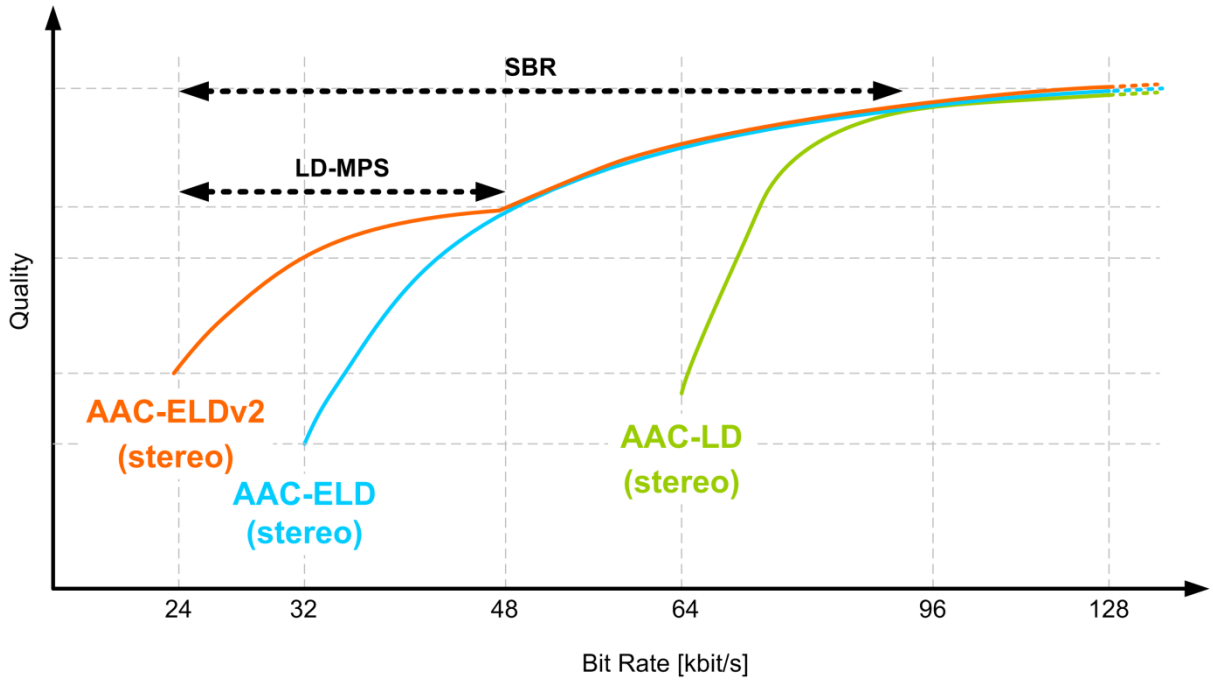

Figure 2: AAC-ELD stereo operating points (image source: Fraunhofer IIS)

A Closer Look at AAC-LD

The core structure of AAC-LD is directly derived from AAC and was standardized in ISO/IEC 14496-3:1999/Amd 1:2000. The time domain input samples are transformed into a frequency domain representation by an MDCT (or Low Delay MDCT, in the case of AAC-ELD) filter bank. The 960 (or 1024) sample size of the MDCT analysis window utilizes a frequency resolution of 50 Hz and a time

resolution of 10 ms. These parameters are chosen to efficiently exploit psychoacoustic effects of frequency and time domain masking.

As natural audio signals show diverse signal characteristics, specialized tools take care of them:

- Temporal Noise Shaping allows the AAC-LD coder to improve the time resolution and to exercise control over the temporal fine structure of the audio signal.
- Intensity Coupling and Mid/Side Stereo increase the coding gain for a stereo channel pair compared to encoding two mono channels separately.
- Perceptual Noise Substitution (PNS) uses a parametric representation of noise-like frequency bands for an efficient transmission.

The codec can operate in a fixed frame length mode, where every packet is an equal size, or in a fixed bit rate mode, where the average bit rate within a limited time frame is constant. The detailed technical description of AAC-LD can be found in ISO/IEC 14496-3:2009, section "4.6.17 Low delay codec".

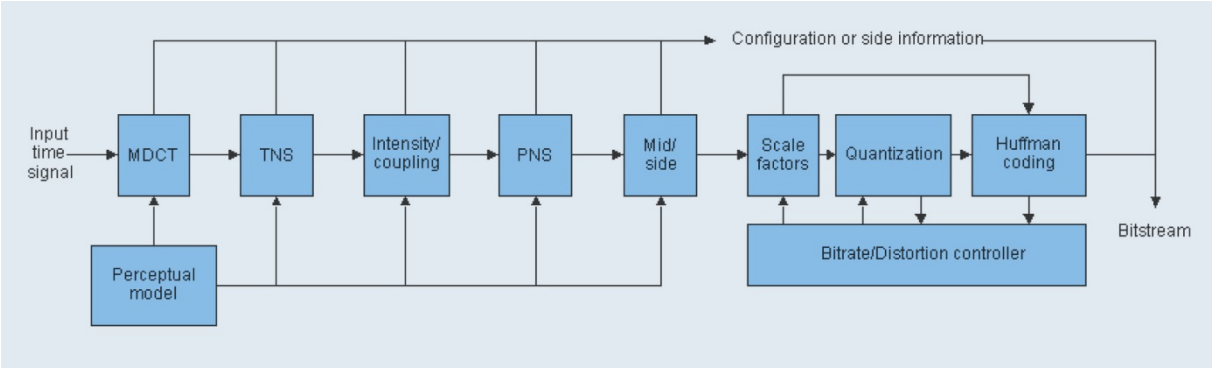

Figure 3: AAC core encoder (image source: Fraunhofer IIS)

A Closer Look at AAC-ELD

AAC-ELD, which was standardized in ISO/IEC 14496-3:2005/Amd 9:2008, is the most flexible codec to suit the different needs of all possible Full-HD Voice applications. To achieve this level of flexibility, AAC-ELD can be used in three different operation modes – all of them completely compatible with standard compliant decoders:

1. **AAC-ELD core:** This mode can be used in all applications where high bit rates are available, for example 96 kbit/s and more for a stereo signal. A Low Delay MDCT filter bank replaces the MDCT filter bank used in AAC-LD. With this delay-optimized filter bank, AAC-ELD operates with a lower delay compared to AAC-LD.
2. **AAC-ELD with SBR:** This mode is the most flexible mode of AAC-ELD as it covers a very wide range of bit rates (approximately 24 to 64 kbit/s per channel) and sampling rates; therefore this mode is the preferred mode for video telephony applications such as Apple FaceTime. The delay stays constant over a wide range of bit rates enabling dynamically switching of bit rates without causing delay variances. In MPEG documents, this mode is typically called “down sampled mode”. It incorporates a delay-optimized version of Spectral Bandwidth Replication (LD-SBR) technology into the AAC-ELD core. LD-SBR allows the reduction of overall bit rate while maintaining excellent audio quality. The lower part of the audio spectrum is coded with AAC-ELD core, while the LD-SBR tool encodes the upper part of the

spectrum. LD-SBR is a parametric approach that exploits the harmonic structure of natural audio signals. It uses the relationship of the lower and upper part of the spectrum for a guided recreation of the whole audio spectrum of the signal.

3. **AAC-ELD with Dual Rate SBR:** For applications that are demanding for even lower data rates, like those in live broadcast contribution, the "Dual Rate SBR" mode can be used. This mode is the most bit rate efficient mode and enables bit rates as low as 16 kbit/s per channel with an increased delay compared to the other two modes. In this mode, again the LD-SBR tool is added to AAC-ELD; however, the AAC-ELD core is coded with half the sampling frequency of the overall signal, instead of coding at the full sampling rate. This results in the best possible audio quality at very low bit rates. The structure of an AAC-ELD codec with Dual Rate SBR is shown in Figure 4.

Every AAC-ELD standard-compliant decoder can operate in any of the three modes, which allows the designer of the encoder to freely choose the mode that best fits the application scenario.

The audio quality of AAC-ELD has been confirmed in several independent listening tests. In 2010, Deutsche Telekom Laboratories investigated the bit rate demands of state-of-the-art, super-wideband communication codecs (see Figure 5). The study showed AAC-ELD as the only codec that delivers an excellent overall quality at 32 kbit/s. Other codecs employed in the test procedure, e.g. G.722.1-C and CELT, required a minimum of 48 kbit/s to reach the same quality level, while speech codecs such as Speex and Skype's SILK failed to deliver excellent quality at any of the selected bit rates. The detailed technical description of AAC-ELD can be found in ISO/IEC 14496-3:2009, section "4.6.20 Enhanced Low Delay Codec". The Low Delay SBR tool is described in ISO/IEC 14496-3:2009, section "4.6.19 Low Delay SBR".

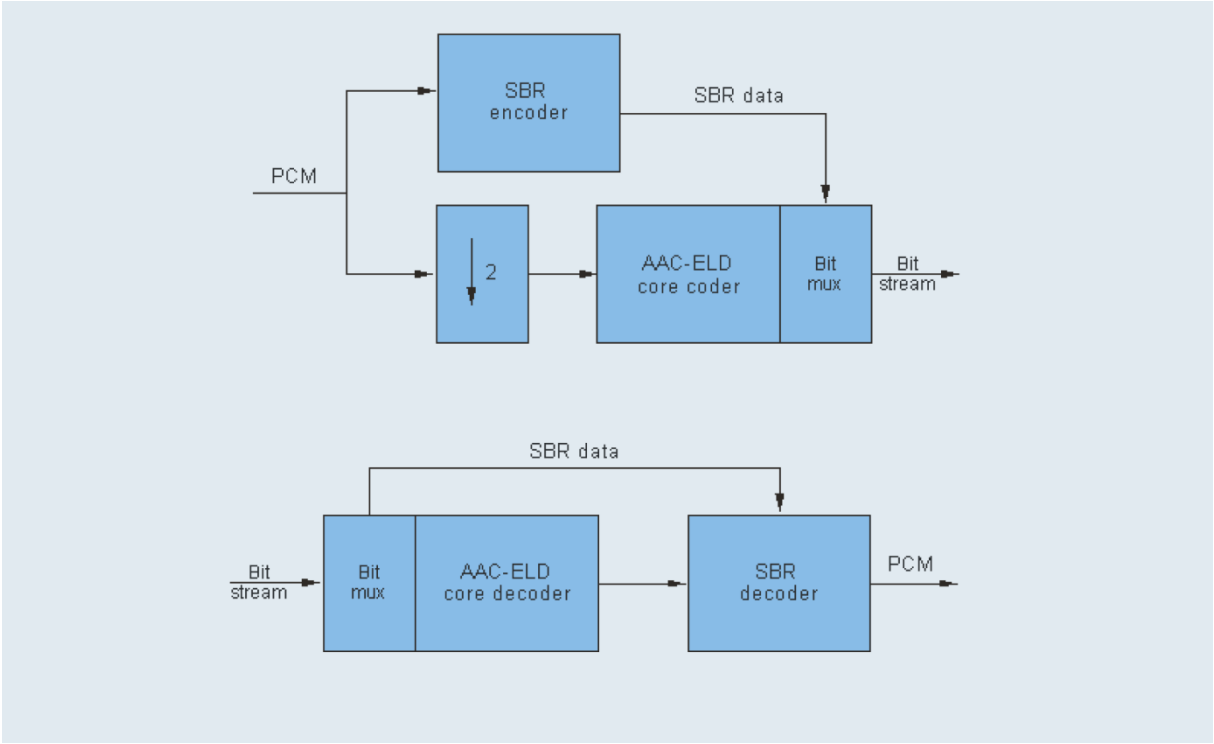

Figure 4: AAC-ELD codec with Dual Rate SBR (image source: Fraunhofer IIS)

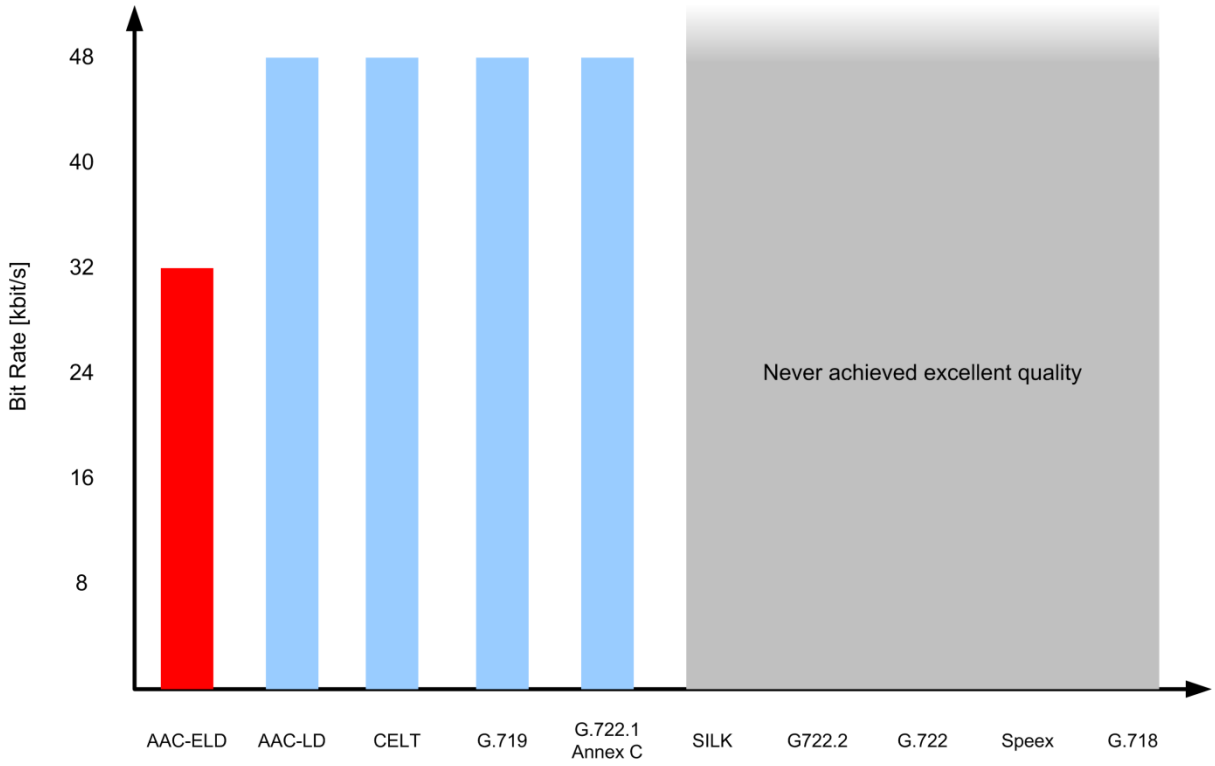

Figure 5: Minimum bit rate for excellent audio quality (image source: Deutsche Telekom, 2010, AES 129, Ulf Wüstenhagen et. all, "Evaluation of Super-Wideband Speech and Audio Codecs".)

A Closer Look at AAC-ELD v2

To achieve stereo performance at bit rates close to monophonic operation, a parametric stereo extension has been integrated into AAC-ELD v2. This parametric extension is based on a 2-channel version of Low Delay MPEG Surround (LD-MPS, standardized in ISO/IEC 23003-2:2010) that further reduces the bit rate. Instead of transmitting two channels, the LD-MPS encoder extracts spatial parameters to enable reconstruction of the stereo signal at the decoder side and the remaining mono down mix is AAC-ELD encoded. The LD-MPS data is transmitted together with the SBR data in the AAC-ELD bit stream. The AAC-ELD decoder reconstructs the mono signal and the LD-MPS decoder recreates the stereo image.

Typically, the bit rate overhead for the stereo parameters is around 3 kbit/s at 48 kHz. This allows AAC-ELD v2 to code stereo signals at bit rates significantly lower than those coded with discrete stereo coding. The detailed technical description of the LD-MPS tool can be found in ISO/IEC 23003-2:2010, "Annex B Low Delay MPEG Surround".

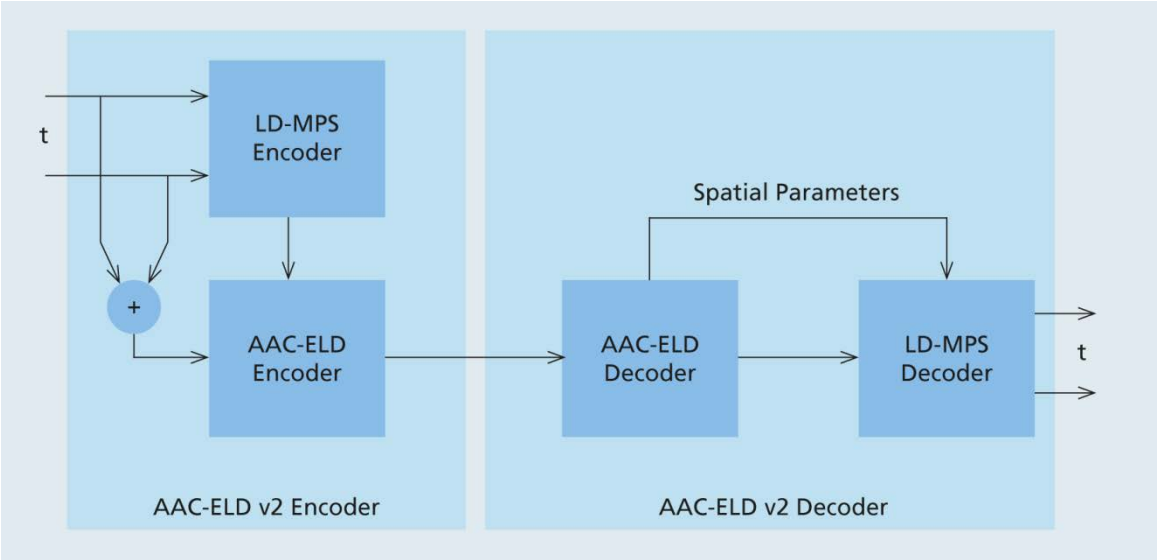

Figure 6: AAC-ELD v2 (image source: Fraunhofer IIS)

Tackling the Delay Issue

In a face-to-face conversation, delays in response can be interpreted in a variety of ways, including hesitation, requiring time to think, or not wanting to give an answer. When conversations take place via technology, and not face-to-face, these important intentional pauses may be misinterpreted as technical glitches, and technical glitches on the other hand can be misinterpreted as intentional, meaningful pauses. Therefore, it is very important to keep technical delays, also called latencies, to a minimum of 150 to 200 ms of end –to-end delay.

The end-to-end delay of a VoIP call is aggregated by several processing steps and components, such as echo cancellation, noise suppression, automatic gain control, routers, jitter buffer, and speech/audio coding. Since it is very important to maintain a low total latency, it becomes crucial that every component uses this resource responsibly. AAC-ELD is ideally suited in this regard as it contributes only 15 to 32 ms, depending on what bit rate and sampling rate are used.

Delay of AAC-LD

The only sources of AAC-LD algorithmic delay for an IP-based transmission are the overlap-add delay of the MDCT filter bank, which generates a delay of 480 samples and the framing (audio input buffering), which adds another 480 samples. This corresponds to a minimum algorithmic delay of 20 ms at a sampling rate of 48 kHz.

Delay of AAC-ELD

In AAC-ELD core mode, the overlap-add delay of the filter bank is cut in half, from 480 samples to 240 samples, resulting in a very low delay of 15 ms. In the AAC-ELD with SBR mode, the SBR tool adds only a small delay of 64 (or 32) samples, which leads to a very low delay of 15.7 ms. Finally, the dual rate SBR mode achieves the best coding efficiency and ends up with a delay of only 31.3 ms.

Delay of AAC-ELD v2

With AAC-ELD v2, the Low Delay MPEG Surround tool is incorporated in a way that it only causes a small filter-bank delay on the decoder end. If the core coder operates in the AAC-ELD core mode, the additional delay is 5.3 ms (sampling rate 48 kHz). In case the core codec operates in a mode with LD-SBR, the additional delay can be reduced to 4 ms. This results in a typical algorithmic delay of 35 ms.

MPEG Reference Documents

Since the AAC-ELD family has grown over many years, there is no single document in the MPEG standard that describes all members of the AAC-ELD family, thus several documents are needed to describe all the members of the AAC-ELD family. The following list gives an overview of all relevant MPEG standard documents.

- **ISO/IEC 14496-3:2009.** This MPEG-4 standard defines the basic AAC-LD and AAC-ELD coding schemes, including the low delay SBR tool and signaling of AAC-LD and AAC-ELD.

- **ISO/IEC 14496-3:2009/Amd 2:2010.** This amendment to the 14496-3 standard contains signaling of Low Delay MPEG Surround.

- **ISO/IEC 14496-3:2009/Amd 3:2012.** This amendment to the 14496-3 standard defines the Low Delay AAC v2 Profile.

- **ISO/IEC 23003-1:2007.** This standard is part of MPEG-D and defines basic tools and bit stream elements of MPEG Surround that are also valid for Low Delay MPEG Surround.

- **ISO/IEC 23003-2:2010.** This standard is part of MPEG-D and describes the configuration and payload of Low Delay MPEG Surround.