Jürgen Herre
and Martin Dietz

MPEG-4 High-Efficiency AAC Coding

The name MPEG-4 High-Efficiency AAC (HE-AAC) refers to a family of recent audio coders that was developed by the International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) Moving Picture Experts Group (MPEG) by subsequent extension of the established Advanced Audio Aoding (AAC) architecture. These algorithmic extensions facilitate a significant increase in coding efficiency relative to previous standards and other known systems. Thus, they provide a representation for generic audio/music signals that offers high audio quality also to applications limited in transmission bandwidth or storage capacity, such as digital audio broadcasting and wireless music access for cellular phones. This article presents a compact overview of the evolution, technology, and performance of the MPEG-4 HE-AAC coding family.

BACKGROUND

From the very beginning in 1992, MPEG audio coding formats and technology such as MPEG-1/2 Layer 3, popularly referred to as "mp3," have successfully supported and inspired new applications for audio-only and audio-visual storage/transmission. Among these coders, the MPEG-2 AAC scheme (see the document in the "HE-AAC Resources" sidebar) has emerged as the prominent "all-round coder" and evolved into the root of most subsequent MPEG audio coder developments. The AAC architecture was carried forward as the core of MPEG-4 Audio coding for generic audio signals and further developed to support a range of functionalities, such as scalability, low-delay operation (Low-Delay AAC, AAC-LD), and lossless signal representation. Two of the latest additions aim at providing improved coding efficiency at very low data rates and are referred to by the name HE-AAC.

Digital Object Identifier 10.1109/MSP.2008.918684

MOTIVATION

The rapid development of digital communication during the past decade has opened numerous opportunities for new multimedia services, some of them operating under severe restrictions regarding the available bit rate for representing audio content. These include terrestrial and satellite-based digital audio broadcasting as well as wireless music downloads to cellular phones. Since these applications cannot be served well using regular AAC, an enhanced audio format is desired that can deliver high audio quality and full audio bandwidth even at very low data rates, for instance at rates of 24 kb/s and below per audio channel.

OBJECTIVES

To enable audio and music delivery for very-low-bit-rate applications, a substantial increase of coding efficiency is required compared to the performance offered by regular AAC at such rates. As a common limitation, previous general audio coders typically have to reduce the transmitted audio bandwidth when operating at low bit rates (e.g., below 48 kb/s per audio channel) in order to avoid excessive coding artifacts from being introduced in the transmitted low-frequency region. HE-AAC technology was designed to overcome this obstacle by reproducing a wide audio bandwidth independent of the coding bit rate by using audio bandwidth extension. An enhanced version of the coder (HE-AAC v2) is designed to additionally exploit models of human spatial perception to achieve a further boost in coding efficiency. In both cases, a further objective was to achieve this goal by means of simple extensions to the existing AAC architecture that come at a limited increase in computational complexity.

ISSUING BODY AND SCHEDULE

Both the AAC and the HE-AAC specifications were created by ISO/IEC MPEG, which is the joint working group ISO/IEC JTC1/SC29/WG11 of the ISO and IEC.

MPEG AAC technology was published as the MPEG-2 AAC specification in 1997 and carried forward into MPEG-4 general audio coding, a first version of which was finalized in 1999.

In response to an MPEG call for proposals on new tools for audio coding providing increased coding efficiency at low bit rates (24 kb/s per audio channel), two standardization projects were pursued in parallel—the first on a tool for audio bandwidth extension and the second on a parametric audio coder for high-quality audio. The work on bandwidth extension was successfully concluded in 2003 and led to a first version of HE-AAC (HE-AAC v1), consisting of a combination of AAC with the newly developed bandwidth extension tool. The parallel work on high-quality parametric audio coding was finalized in 2004 and includes, among other components, a parametric stereo coding module that can be combined with the HE-AAC scheme. This resulted in HE-AAC v2, offering a further increase in coding efficiency at very low data rates. Related documents are listed in the "HE-AAC Resources" sidebar.

TARGET APPLICATIONS

Target applications for HE-AAC are mobile music, mobile TV, digital radio

 1053-5888/08/$25.00©2008IEEE

HE-AAC RESOURCES

The Standard

- *Information Technology—Coding of Audio-Visual Objects—Part 3: Audio, Third Edition*, ISO/IEC 14496-3:2005, 2005.

Overviews

- P. Ekstrand: "Bandwidth extension of audio signals by spectral band replication," in *Proc.1st IEEE Benelux Workshop on Model based Processing and Coding of Audio (MPCA-2002)*, Leuven, Belgium, Nov. 15, 2002, pp. 73–79.
- H. Purnhagen, "Low complexity parametric stereo coding in MPEG-4," in *Proc. 7th Int. Conf. Audio Effects (DAFX-04)*, Naples, Italy, Oct. 2004, pp. 163–168.
- E. Schuijers, J. Breebaart, H. Purnhagen, and J. Engdegård: "Low complexity parametric stereo coding," in *Proc. 116th AES Convention*, Berlin, Germany, 2004, preprint 6073.
- J. Breebaart, S. van de Par, A. Kohlrausch, and E. Schuijers: "Parametric coding of stereo audio," *EURASIP J. Applied Signal Processing*, Sept. 2005, pp. 1305–1322. [Online]. Available: http://www.hindawi.com/GetArticle.aspx?doi=10.1155/ASP.2005.1305 (article).

Books

- J. Herre and H. Purnhagen, "General audio coding," in *The MPEG-4 Book (Prentice Hall IMSC Multimedia Series)*, F. Pereira and T. Ebrahimi, Eds. Englewood Cliffs, NJ: Prentice-Hall, 2002.

Reference Software

- *Information Technology—Coding of Audio-Visual Objects—Part 5: Reference Software, Amendment 6: Advanced Video Coding (AVC) and High Efficiency Advanced Audio Coding (HE AAC) Reference Software*, ISO/IEC 14496-5:2001/Amd.6:2005, June 15, 2005. [Online]. Available:http://standards.iso.org/ittf/PubliclyAvailableStandards/c039593_ISO_IEC_14496-5_2001_Amd_6_2005(E)_Reference_Software.zip.

Test Sequences

- *Information Technology—Coding of Audio-Visual Objects—Part 4: Conformance Testing, Amendment 8: High Efficiency Advanced Audio Coding, Audio BIFS, and Structures Audio Conformance*, ISO/IEC 14496-4:2004/Amd.8:2005, June 1, 2005 [Online]. Available:http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_IEC%2014496-4_2004_Amd_8_2005_Bitstreams/.

General MPEG-4 Resources

(Including Discussion List, Reference Software, Bitstreams, FAQ)

- MPEG Industry Forum [Online]. Available: http://www.m4if.org; see section "About MPEG-4 -> Resources."

Other Standards Documents

- ISO/IEC 13818, Part 7, MPEG-2 AAC
- ISO/IEC 14496, Part 3, Subpart 4 MPEG-4 General Audio Coding
- ISO/IEC 14496-3:2001/AMD1:2003, Bandwidth Extension
- ISO/IEC 14496-3:2001/AMD2:2004, High-Quality Parametric Audio Coding
- ISO/IEC 14496-3:2005/AMD2:2006, HE-AAC v2
- ISO/IEC 23003-1:2006 MPEG Surround Specification

Resources for Further Development

- M. Schnell, R. Geiger, M. Schmidt, M. Jander, M. Multrus, G. Schuller, and J. Herre, "Enhanced MPEG-4 low delay AAC—Low bitrate high quality communication," in *Proc. 122nd AES Convention*, Vienna, Austria, May 2007, preprint 6998.

and TV broadcasting, Internet streaming, and consumer electronics. The codec has been adopted by various application-oriented standards/consortia worldwide.

In the mobile music and TV market, HE-AAC is used for music downloads, music streaming, ring tones, ring-back tones, and the audio part of various mobile TV broadcasting systems. In audio broadcasting, HE-AAC is a mandatory component of multiple existing and emerging systems. In TV broadcasting, the codec is of special interest in combination with the H.264/Advanced Video Coding (H.264/AVC) video codec in new systems. The first commercial service using both MPEG standards was launched in 2007 in Norway. In Internet streaming, HE-AAC is of special interest because of the significant bandwidth savings at the server side and the capability of streaming directly into the mobile environment. Spillover effects into consumer electronics are expected due to the convergence of devices and the desire to store and play content across various platforms.

STRUCTURE OF THE STANDARD

Like most MPEG audio coding standards, the MPEG-4 HE-AAC audio coding standard specifies the bit stream format and the decoding process, including conformance testing methods and reference implementations. The decoding process defines how the syntax elements present in the encoded bit stream are converted into a time domain pulse code modulated (PCM) digital audio signal. As a result, every decoder conforming to the standard will produce a well-defined output signal for any bit stream conforming to the standard. The encoding algorithm, on the other hand, is not normatively specified, thus, for example, enabling the balance of real-time execution speed and audio quality depending on the individual application demands. In 2005, the third edition of the MPEG-4 Audio standard was published, which comprises all three technologies contained in HE-AAC (i.e., MPEG-4 AAC, spectral band replication bandwidth extension, and parametric stereo) in one document (ISO/IEC 14496-3:2005 - Part 3: Audio).

TECHNOLOGY

FUNCTIONALITIES

HE-AAC supports a broad range of compression ratios and configurations ranging from highly efficient mono and stereo coding (typical operation point 32 kb/s stereo with HE-AAC v2) via high-quality multichannel coding (typical operation point 160 kb/s for 5.1 configuration) to near-transparent multichannel compression (typical operation point 320 kb/s using AAC without extensions). Because subsequent HE-AAC versions form a superset of their predecessors, HE-AAC v2 decoding is fully compatible with AAC-only and HE-AAC v1 content.

ARCHITECTURE

The basic architecture of the HE-AAC codec is shown in Figure 1. The core of the system is the AAC waveform codec. For increased compression efficiency, the spectral band replication (SBR) bandwidth enhancement tool and the parametric stereo (PS) advanced stereo compression tool are added to the system. Both SBR and PS act as preprocessing blocks at the encoder side and postprocessing blocks at the decoder side. The bit streams created by the two tools are seamlessly transmitted in specific, previously unused sections of the AAC bit stream. This allows for reusability of existing AAC implementations and full decoding compatibility with existing AAC content. An HE-AAC v2 decoder comprises all three technologies and is a superset of an AAC or AAC+SBR = HE-AAC v1 decoder.

The bit stream syntax of HE-AAC allows for up to 48 audio channels. In practice, mono, stereo, and 5.1 multichannel are the most commonly used configurations. The PS technology is defined for stereo configurations only.

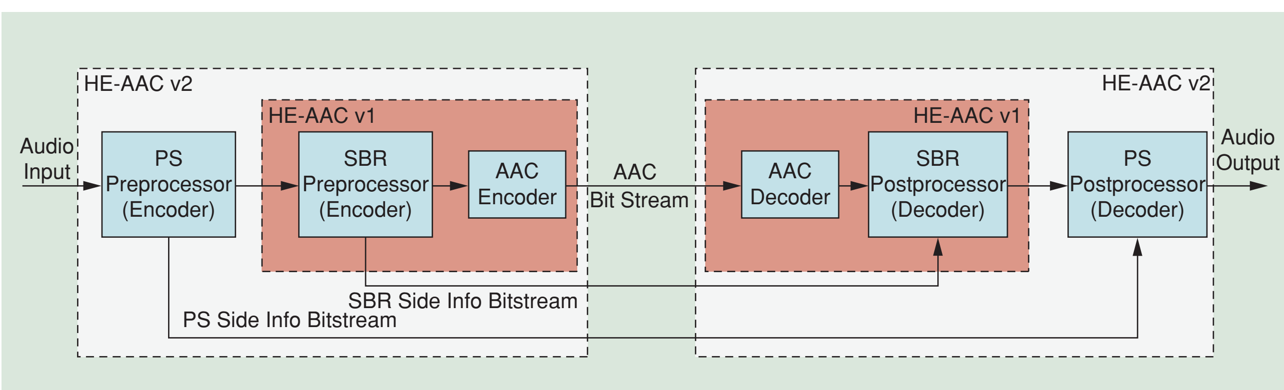
TOOLS

Here we describe the three core ingredients of HE-AAC: the AAC coding kernel, the SBR bandwidth extension, and the PS tools.

AAC

AAC belongs to the class of perceptually oriented traditional waveform codecs, meaning that it aims at reproducing the waveform of the original input audio signal with a minimum amount of data while taking into account psychoacoustic principles to minimize the audibility of coding effects. Other well-known representatives of this class of codecs are MPEG-1/2 Layer 2, MPEG-1/2 Layer 3 (better known as mp3), and Dolby AC-3 (better known as Dolby Digital). Today, AAC constitutes the most efficient waveform compression standard. Its most important tools are:

- *Modified Discrete Cosine Transform* (MDCT) filterbank using window switching: Transforming the signal into a spectral representation is the key to apply psychoacoustic principles and redundancy reduction algorithms for audio content. For this purpose, AAC employs a 1,024 spectral line MDCT filterbank, creating spectra corresponding to 1,024 PCM input samples. In the case of highly time-varying signals, the filterbank time resolution can be increased by producing a series of lower-resolution spectra each corresponding to 128 input audio samples.
- *Stereo Processing*: Intensity and midside stereo processing are available to increase the compression efficiency for stereo signals. While the former is rarely used in practice, the latter is improved over what was available in earlier codecs like mp3.
- *Temporal Noise Shaping*: The temporal noise shaping (TNS) tool allows the codec to shape quantization noise in the time domain by running a prediction across frequency on the spectral data. This avoids undesirable effects caused by the relatively coarse time resolution of the MDCT filterbank. The TNS tool has been newly developed for AAC.
- *Quantization and Coding*: The tools to quantize and code the spectrum are



[FIG1] Basic architecture of the HE-AAC codec.

similar to what is used in mp3, with significant refinements in the entropy coding stage, resulting in improved compression efficiency. The AAC bit stream syntax has been defined in a much more flexible way than in earlier codecs to support various configurations and future extensions.

SPECTRAL BAND REPLICATION

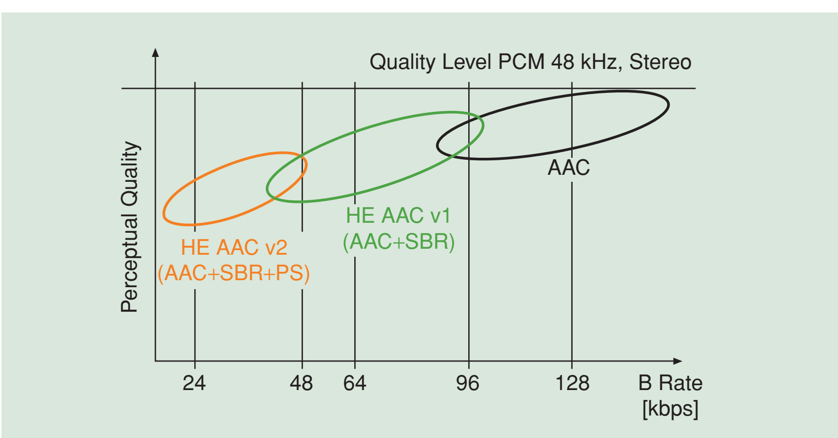
Bandwidth extension technology is based on the observation that usually the upper part of the spectrum of an audio signal contributes only marginally to the "perceptual information" contained in the signal, and that human auditory perception is less sensitive in the high frequency range. As an example, an audio signal that has been band-limited to 8 kHz is still fully recognized by humans (although it may not sound attractive without the upper part of the spectrum). SBR exploits this observation for the purpose of improved compression; instead of transmitting the upper part of the spectrum with AAC, SBR regenerates it from the lower part with the help of some low-bit-rate guidance data. For regenerating the missing high-frequency components, SBR operates in the frequency domain using a quadrature mirror filter (QMF) filterbank analysis/synthesis system. The most important building blocks of SBR are:

- *High-Frequency Reconstruction*: The so-called transposer generates a first estimate for the upper part of the spectrum by copying and shifting the lower part of the transmitted spectrum. In order to generate a high-frequency spectrum that is close to the original spectrum in its fine structure, several provisions are available including the addition of noise, the flattening of the spectral fine structure, and the addition of missing sinusoids.
- *Envelope Adjustment*: The upper spectrum generated by the transposer needs to be shaped subsequently with respect to frequency and time in order to match the original spectral envelope as closely as possible.

The SBR bit stream data controls both the operation of the high-frequency reconstruction and the envelope adjustment. Depending on the specific configuration, the SBR side information rate is typically a few (e.g., 2–3) kb/s.

PARAMETRIC STEREO

PS is an extension of a well-known principle for efficient joint coding of stereo audio. Instead of the stereo signal, just a mono-downmix is transmitted, along with a small data stream describing how to upmix the signal back to stereo in the decoder. The intensity stereo tool available in AAC and many other codecs (like mp3) is a simple implementation of this approach, whereas PS is a significantly more sophisticated variant thereof. To reproduce a high-quality stereophonic sound image, it is vital to consistently preserve the cues that determine human spatial hearing of sound, i.e., interaural level difference, interaural time/phase difference, and interaural correlation/coherence. While traditional intensity stereo can only reproduce level (intensity) differences between the stereo channels, the PS technology can also produce phase differences and decorrelation between the stereo pair to yield a convincing upmix quality. Most notably, PS includes a decorrelator tool that creates an adjustable degree of decorrelation between the two channels and is steered by coherence factors measured in the encoder and transmitted in the PS data. This is vital for modeling sound sources with a wide sound image (e.g., a choir) or room ambience. Since PS operates on the same spectral representation as SBR, both can be efficiently integrated to form an even more efficient compression algorithm for audio signals at relatively low additional computational complexity. Also, PS coding requires only a few kb/s transmitted as its side information data rate.



[FIG2] Operating range of technologies versus bit rate (idealized view).

PROFILES AND LEVELS

MPEG-4 Audio provides a large toolset comprising several technologies, of which only a subset are used in common applications. Thus, *profiles* have been introduced to define meaningful combinations of such tools that are targeted at certain application areas. Furthermore, for each profile, *levels* are defined that describe the specific decoder processing capabilities. Hence, profiles and levels ensure interoperability between devices that conform to a certain profile/level combination. Since the HE-AAC v2, HE-AAC v1, and AAC profiles form supersets of each other, the addition of each coding tool extends the range of efficient operation towards lower data rates, as illustrated in an idealized manner in Figure 2. This assumes that the HE-AAC encoder will pick the appropriate combination of technologies depending on the desired compression ratio.

A summary of the profiles and levels for the HE-AAC family is provided in Table 1. The profiles and levels have been designed in a strictly hierarchical fashion such that the HE-AAC v2 profile is a superset of the HE-AAC profile, which in turn is a superset of the AAC profile.

[TABLE 1] PROFILES AND LEVELS OF THE HE-AAC FAMILY.

	AAC PROFILE		HE-AAC V1 PROFILE				HE-AAC V2 PROFILE			
LEVEL	MAX. CHANNELS	MAX. SAMPLING RATE [kHZ]	MAX. CHANNELS	MAX. AAC SAMPLING , RATE, SBR NOT PRESENT [kHZ]	MAX. AAC SAMPLING RATE, SBR PRESENT [kHZ]	MAX. SBR SAMPLING RATE [kHZ]	MAX. CHANNELS	MAX. AAC SAMPLING RATE, SBR NOT PRESENT [kHZ]	MAX. AAC SAMPLING RATE, SBR PRESENT [kHZ]	MAX. SBR SAMPLING RATE [kHZ]
1	2	24	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
2	2	48	2	48	24	48	2	48	24	48
3	N/A	N/A	2	48	48	48	2	48	48	48
4	5	48	5	48	48	48	5	48	48	48
5	5	96	5	96	48	96	5	96	48	96

Also, within all profiles each higher level is a superset of the lower levels such that, for example, a decoder conforming to the HE-AAC v2 profile, Level 5 is capable of handling all bit streams conforming to any combination of profile and levels as given in Table 1. In practice, the most relevant levels are Level 2 for stereo devices (e.g., cell phones, broadcasting receivers) and Level 4 for multichannel systems (e.g., digital television).

COMPARISON WITH OTHER STANDARDS

HE-AAC has been the first compression standard that made use of bandwidth extension techniques and advanced parametric stereo coding. Over time, such techniques have been adopted in other specifications outside MPEG as well, although with partly differing technical approaches. Two examples are the wideband—adaptive multirate plus coder (AMR-WB+) of the Third Generation Partnership Project (3GPP), which is a hybrid of a speech coder (AMR-WB) and an audio coder for improved performance for music, and ITU-T Recommendation G.729.1, which is a wideband extension (=7kHz) of the G.729 speech coder.

PERFORMANCE

OBJECTIVE AND SUBJECTIVE QUALITY

Due to the significant impact of complex perceptual aspects on perceived audio quality, a reliable quality assessment of intermediate-quality audio signals (as they may appear in very-low-bit-rate audio coding) to date still has to be achieved by subjective assessment (listening tests) rather than by using objective measurement, especially when the signal under test was processed by parametric coding tools (such as SBR and PS) that depart from the waveform coding paradigm.

AAC technology has been shown to meet the ITU-R Recommendation BS.1115 audio quality requirement for perceptual audio codecs in broadcast applications at bit rates of 64 kb/s per audio channel for both stereo and multichannel operation (i.e., none of the test items had a mean grade below -1.00 relative to the original/uncoded item). In a comparison across a broad set of stereo test material, AAC at 96 kb/s provided an average quality comparable to that of AC-3 ("Dolby Digital") at 160 kb/s or MPEG-1 Layer 2 at 192 kb/s. At such rates, AAC still defines the state of the art for very-high-quality audio coding.

Combining this technology with SBR bandwidth extension, HE-AAC v1 offers an increase in coding efficiency by more than 25% over AAC when operated at or near 24 kb/s per audio channel (according to MPEG verification tests). With the inclusion of parametric stereo coding, a

HE-AAC HARDWARE AND SOFTWARE PRODUCTS

Standardization Organizations, Consortia, and System Providers

- Mobile: 3GPP, 3GPP2
- Mobile TV: DVB-H, T-DMB, S-DMB, MediaFlo, ISDB 1-seg
- Digital TV (SDTV and HDTV): DVB-S/-T/-C/-IPTV
- Digital Radio: DAB+, Digital Radio Mondiale
- Consumer electronics: DVD, HD-DVD, SDcard, Nero Digital
- Others: DLNA, ISMA
- Mobile phone operators: KDDI, SK Telecom, NTT DoCoMo, Vodafone, T-Mobile, Sprint, Verizon, Telefonica, Telenor, SFR, Telstra, etc.
- Radio and TV providers: BBC, Deutsche Welle, RTL, Voice of Russia, Radio Netherlands, XM Satellite Radio (Note that XM Satellite Radio uses an earlier version of AAC+SBR and is hence not compatible with HE-AAC.), Euro 1080, etc.

Hardware Products

- Mobile phones: Nokia, Motorola, Samsung, SonyEricsson, LG, Panasonic, NEC, Palm, etc.
- Set-top boxes: ADB, Amino, Scientific Atlanta, Motorola, Tilgin, Even Technologies, Thomson, etc.
- Radios: Delphi, Pioneer, Samsung, Morphy Richards, Roberts, Audiofox, Antex, Polk Audio, Himalaya, etc.
- Chipsets and hardware platforms: ARM, Analog Devices, Texas Instruments, Broadcom, Sigma Design, Conexant, NXP, STMicroelectronics, Qualcomm, AMD, Micronas, Renesas, Tensilica, LSI Logic, Equator/Pixelworks, MIPS, etc.
- Professional encoders: Tandberg, Harmonic, Orban, Mayah, Thomson/Grass Valley, Rhode & Schwartz, Telos Systems, Modulus, Linear Acoustic, Transradio, Comrex, etc.

Software Products

- Professional encoder software: Sorensen, Vidiator, Spacial Audio, Cube Tec
- Consumer software: AOL/Winamp, RealNetworks, Nero, Yahoo, Poiksoft

further increase in coding efficiency is achieved and HE-AAC v2 typically performs as well as HE-AAC v1 when the latter is operating at a 33% higher bit rate (up to 40 kb/s stereo, according to MPEG verification tests). When compared to the AMR-WB+ coder, tests within 3GPP found that HE-AAC v2 provides better subjective quality for music and "music over speech" content at and above 18 kb/s stereo and for "speech between music" at and above 24 kb/s mono.

SPEED/COMPLEXITY PERFORMANCE

The significant increase in coding efficiency of HE-AAC over MPEG-4 AAC comes at moderate additional computational complexity. While both the SBR and the PS tool consume additional calculations, this increase is partially compensated by running the AAC core at half the original sampling rate and just for one channel (in the case of PS). As a consequence, the approximate computational complexity of the decoder is increased by a factor of 1.5 and 2, when comparing HE-AAC v1 and HE-AAC v2 to AAC, respectively. The encoder complexity is roughly similar for all three variants.

FURTHER TECHNICAL DEVELOPMENTS

As a generalization of the principles behind parametric stereo coding, the recently finalized "MPEG Surround" specification enables the bit-rate-efficient and backward-compatible coding of multichannel audio (surround sound). Similarly to parametric stereo, it can be combined efficiently with HE-AAC v1 by using the same filterbank core structure as employed by the SBR bandwidth extension. The combination of both schemes enables the transmission of high-quality surround sound at extremely low bit rates (below 64 kb/s total).

Another ongoing technical development is the extension of the HE-AAC approach towards low delay operation, as is required for high-quality bidirectional communication applications (e.g. high-quality teleconferencing/videoconferencing/VoIP). This is achieved by combining delay-optimized versions of Low-Delay AAC (AAC-LD) and the SBR bandwith extension tool into a so-called "MPEG-4 Enhanced AAC Low Delay" Coder (AAC-ELD).

RESOURCES

As it is customary for MPEG, the coding specifications are provided with reference software (including both normative and nonnormative sections), a conformance specification defining conformant bit stream and decode behavior, and conformance test bit streams in order to support widespread interoperable implementation. A list of available resources is included in the "HE-AAC Resources" sidebar.

PRODUCTS

The HE-AAC specification has been embraced by numerous application standards and thus found its way into widespread use. In some applications, HE-AAC is referred to as aacPlus and HE-AAC v2 is referred to as enhanced aacPlus or aacPlus v2. Some of the adopting organizations and product uses of the standards are included in the "HE-AAC Hardware and Software Products" sidebar.

AUTHORS

Jürgen Herre (juergen.herre@iis.fraunhofer.de) is the chief scientist for the audio and multimedia activities of the Fraunhofer Institute for Integrated Circuits in Erlangen, Germany. He has been with MPEG since 1994 and served as a coeditor of several generations of MPEG Audio standards, including AAC.

Martin Dietz (martin.dietz@dolby.com) was CEO of Coding Technologies, which was acquired by Dolby in November 2007. He was involved in the development and/or standardization of the three technologies within HE-AAC and editor of the MPEG-2 AAC standard. [SP]