ATSC Standard
A/153 Part 8 – HE AAC Audio System Characteristics

Doc. A/153 Part 8:2012
18 December 2012

Advanced Television Systems Committee
1776 K Street, N.W.
Washington, D.C. 20006
202-872-9160

The Advanced Television Systems Committee, Inc., is an international, non-profit organization developing voluntary standards for digital television. The ATSC member organizations represent the broadcast, broadcast equipment, motion picture, consumer electronics, computer, cable, satellite, and semiconductor industries.

Specifically, ATSC is working to coordinate television standards among different communications media focusing on digital television, interactive systems, and broadband multimedia communications. ATSC is also developing digital television implementation strategies and presenting educational seminars on the ATSC standards.

ATSC was formed in 1982 by the member organizations of the Joint Committee on InterSociety Coordination (JCIC): the Electronic Industries Association (EIA), the Institute of Electrical and Electronic Engineers (IEEE), the National Association of Broadcasters (NAB), the National Cable and Telecommunications Association (NCTA), and the Society of Motion Picture and Television Engineers (SMPTE). Currently, there are approximately 150 members representing the broadcast, broadcast equipment, motion picture, consumer electronics, computer, cable, satellite, and semiconductor industries.

ATSC Digital TV Standards include digital high definition television (HDTV), standard definition television (SDTV), data broadcasting, multichannel surround-sound audio, and satellite direct-to-home broadcasting.

Note: The user's attention is called to the possibility that compliance with this standard may require use of an invention covered by patent rights. By publication of this standard, no position is taken with respect to the validity of this claim or of any patent rights in connection therewith. One or more patent holders have, however, filed a statement regarding the terms on which such patent holder(s) may be willing to grant a license under these rights to individuals or entities desiring to obtain such a license. Details may be obtained from the ATSC Secretary and the patent holder.

Revision History

Version	Date
A/153 Part 8:2009, initial version of standard approved	15 October 2009
A/153 Part 8:2012, first revision approved	18 December 2012

Table of Contents

1. SCOPE 4
   1.1 Organization 4
2. REFERENCES 4
   2.1 Normative References 4
   2.2 Informative References 4
3. DEFINITION OF TERMS 5
   3.1 Compliance Notation 5
   3.2 Treatment of Syntactic Elements 5
      3.2.1 Reserved Elements 5
   3.3 Acronyms and Abbreviation 5
   3.4 Terms 6
4. SYSTEM OVERVIEW 6
   4.1 Use of SBR and PS 8
5. HE AAC V2 CONSTRAINTS 9
   5.1 Audio Elementary Stream Configuration 9
   5.2 RTP Packetization 10

ANNEX A: RELATIONSHIP BETWEEN MH COMPONENT DATA DESCRIPTOR AND SDP...11

Index of Figures and Tables

Figure 4.1 ATSC broadcast system with TS main and M/H services. 7
Figure 4.2 MPEG-4 Audio tools that together create HE AAC v2. 8
Figure 4.3 Simple block diagram showing how AAC and SBR work together. 8
Figure 4.4 Simple block diagram showing how AAC, SBR and PS work together. 9

Table 5.1 Valid Audio Sampling Frequencies and Maximum Bitrates 9
Table A.1 Some Example HE AAC v2 AudioSpecificConfig Strings 11
Table A.2 ATSC MH Component Descriptor Data 12

Proposed Revision of ATSC Standard A/153 Part 8 – HE AAC Audio System Characteristics (A/153 Part 8:2009)

1. SCOPE

This Part describes a set of constraints on ISO/IEC 14496-3 [1] ("Audio") HE AAC v2 when used in the ATSC Mobile DTV (mobile/handheld, or simply M/H) system. It also defines the RTP packetization for audio elementary streams.

1.1 Organization

This document is organized as follows:

- **Section 1** – Outlines the scope of this Part and provides a general introduction.
- **Section 2** – Lists references and applicable documents.
- **Section 3** – Provides a definition of terms, acronyms, and abbreviations for this Part.
- **Section 4** – System overview.
- **Section 5** – System specifications.
- **Annex A** – Sample SDP file.

2. REFERENCES

All referenced documents are subject to revision. Users of this Standard are cautioned that newer editions might or might not be compatible.

2.1 Normative References

The following documents, in whole or in part, as referenced in this document, contain specific provisions that are to be followed strictly in order to implement a provision of this Standard.

[1] ISO: "Information technology – Coding of audio-visual objects – Part 3: Audio," Doc. ISO/IEC 14496-3:2009, International Standards Organization, Geneva, Switzerland.

[2] IEEE: "Use of the International Systems of Units (SI): The Modern Metric System", Doc. IEEE/ASTM SI 10-2002, Institute of Electrical and Electronics Engineers, New York, N.Y.

[3] IETF: "RTP payload for transport of generic MPEG-4 elementary streams," Doc. IETF RFC 3640, Internet Engineering Task Force, Freemont, CA.

[4] ITU: "Algorithms to measure audio programme loudness and true-peak audio level," ITU-R Recommendation BS.1770-3, International Telecommunications Union, Geneva, 2012.

2.2 Informative References

The following documents contain information that may be helpful in applying this Part.

[5] ATSC: "ATSC Digital Television Standard, Part 2 – RF/Transmission System Characteristics," Doc. A/53 Part 2:2011, Advanced Television Systems Committee, Washington, D.C., 7 October 2011.

[6] ATSC: "ATSC Mobile/Handheld Digital Television Standard, Part 1 – Mobile/Handheld Digital Television System," Doc. A/153 Part 1:2011, Advanced Television Systems Committee, Washington, D.C., 1 June 2011.

[7] ATSC: "ATSC Mobile/Handheld Digital Television Standard, Part 3 – Service Multiplex and Transport Subsystem Characteristics," Doc. A/153 Part 3:2009, Advanced Television Systems Committee, Washington, D.C., 15 October 2009.

[8] IETF: "SDP: Session Description Protocol," Doc. RFC 4566, Internet Engineering Task Force, Freemont, CA.

[9] ATSC: "Recommended Practice – Techniques for Establishing and Maintaining Audio Loudness for Digital Television," Doc. A/85, Advanced Television Systems Committee, Washington, D.C., 25 July 2011.

3. DEFINITION OF TERMS

With respect to definition of terms, abbreviations, and units, the practice of the Institute of Electrical and Electronics Engineers (IEEE) as outlined in the Institute's published standards [2] shall be used. Where an abbreviation is not covered by IEEE practice or industry practice differs from IEEE practice, the abbreviation in question is described in Section 3.3 of this document.

3.1 Compliance Notation

This section defines compliance terms for use by this document:

shall – This word indicates specific provisions that are to be followed strictly (no deviation is permitted).

shall not – This phrase indicates specific provisions that are absolutely prohibited.

should – This word indicates that a certain course of action is preferred but not necessarily required.

should not – This phrase means a certain possibility or course of action is undesirable but not prohibited.

3.2 Treatment of Syntactic Elements

This document contains symbolic references to syntactic elements used in the audio, video, and transport coding subsystems. These references are typographically distinguished by the use of a different font (e.g., restricted), may contain the underscore character (e.g., sequence_end_code) and may consist of character strings that are not English words (e.g., dynrng).

3.2.1 Reserved Elements

One or more reserved bits, symbols, fields, or ranges of values (i.e., elements) may be present in this document. These are used primarily to enable adding new values to a syntactical structure without altering its syntax or causing a problem with backwards compatibility, but they also can be used for other reasons.

The ATSC default value for reserved bits is '1.' There is no default value for other reserved elements. Use of reserved elements except as defined in ATSC Standards or by an industry standards setting body is not permitted. See individual element semantics for mandatory settings and any additional use constraints. As currently-reserved elements may be assigned values and meanings in future versions of this Part, receiving devices built to this version are expected to ignore all values appearing in currently-reserved elements to avoid possible future failure to function as intended.

3.3 Acronyms and Abbreviation

The following acronyms and abbreviations are used within this Part.

AAC – Advanced Audio Coding
ATSC – Advanced Television Systems Committee
HE AAC – High Efficiency Advanced Audio Coding
HE AAC v2 – High Efficiency Advanced Audio Coding version 2
RTP – Real-time Transport Protocol
SBR – Spectral Band Replication
SDP – Session Description Protocol
PS – Parametric Stereo

3.4 Terms

The following terms are used within this Part.

AAC core codec – The plain AAC codec with AAC Profile (as specified in ISO/IEC 14496-3 [1] Table 1.3).

AAC core channel – The down-mixed mono audio channel within an HE AAC v2 codec.

LKFS – Loudness, K-weighted, relative to full scale, measured with equipment that implements the algorithm specified by ITU-R BS.1770 [4]. A unit of LKFS is equivalent to a decibel.

reserved – Set aside for future use by a Standard.

MPEG – Refers to standards developed by the ISO/IEC JTC1/SC29 WG11, Moving Picture Experts Group. MPEG may also refer to the Group.

4. SYSTEM OVERVIEW

Please see ATSC A/153 Part 1 [6] for an overall description of the M/H system. The ATSC Mobile/Handheld service (M/H) shares the same RF channel as a standard ATSC broadcast service described in ATSC A/53 [5]. M/H is enabled by using a portion of the total available ~19.4 Mbps bandwidth and utilizing delivery over IP transport. The overall ATSC broadcast system including standard (TS Main) and M/H systems is illustrated in Figure 4.1.

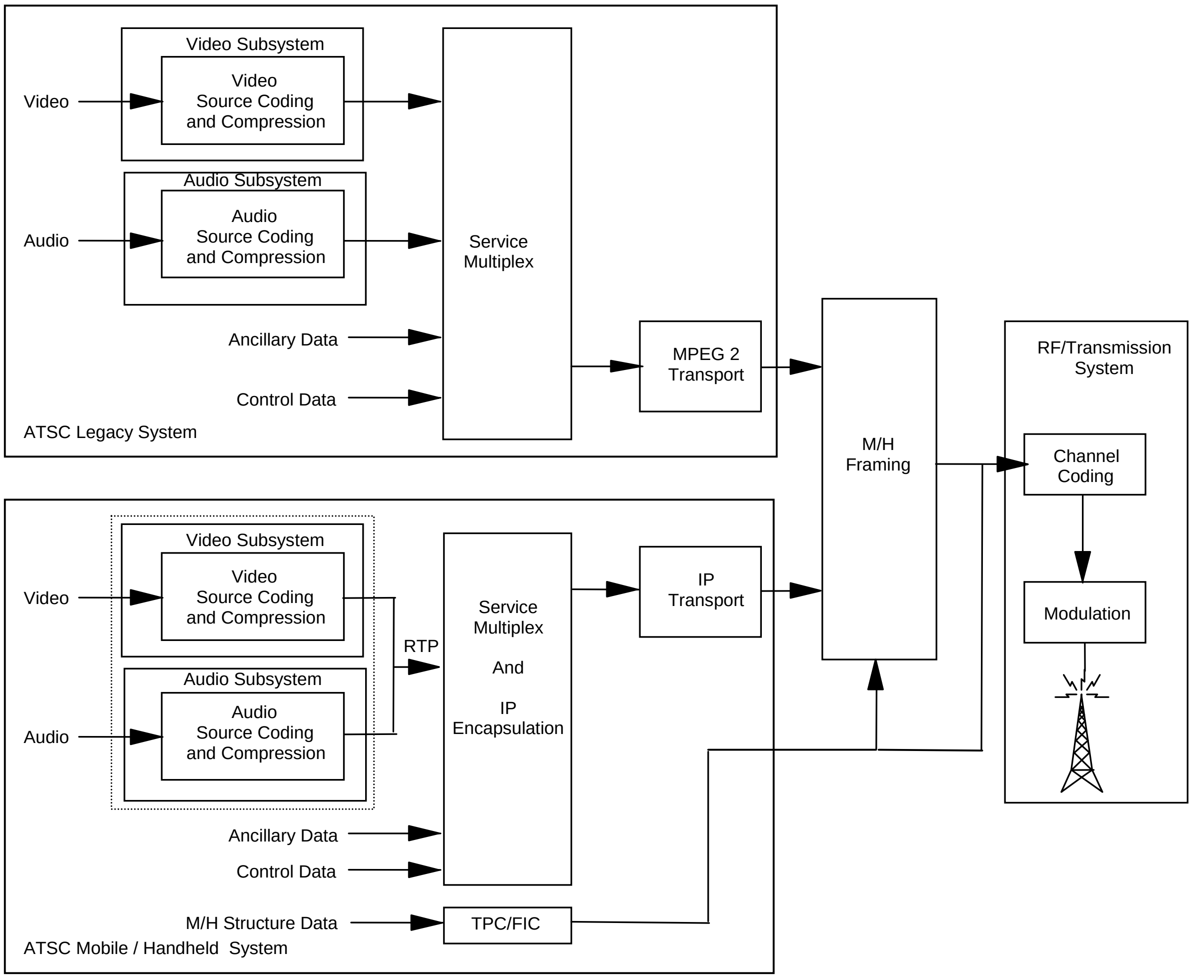

Figure 4.1 ATSC broadcast system with TS main and M/H services.

This Part relates to the Audio Source Coding and Compression block and specifies audio coding using MPEG-4 HE AAC v2 as described in ISO/IEC 14496-3 [1], with the constraints indicated herein. HE AAC v2 is used to code mono or stereo audio. HE AAC v2 is the combination of three audio coding tools, MPEG-4 AAC, Spectral Band Replication (SBR) and Parametric Stereo (PS). This furthermore means that HE AAC v2 includes both HE AAC and AAC as illustrated in Figure 4.2.

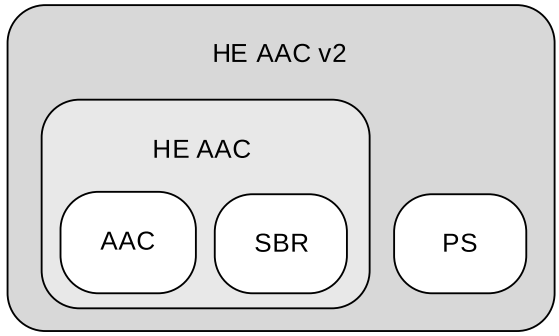

Figure 4.2 MPEG-4 Audio tools that together create HE AAC v2.

4.1 Use of SBR and PS

MPEG-4 AAC is a highly efficient traditional perceptual audio-coding algorithm. Its combination with the parametric SBR tool in HE AAC allows a further reduction of overall bitrate, while maintaining good audio quality. This is because, when using SBR, the AAC encoder may be fed with half the input sampling rate. The lower part of the audio spectrum[1], sampled at this reduced rate, is AAC-encoded, while the upper part is described by the parametric SBR data. On the decoder side, the AAC core decoder generates the lower spectrum. This lower spectrum is fed to the SBR decoder, which uses it to regenerate the full spectrum with the help of the transmitted parametric data. As only the lower part the spectrum is encoded by AAC and the SBR data is negligible, encoding with HE AAC requires about half the bitrate of AAC. This process is illustrated in Figure 4.3.

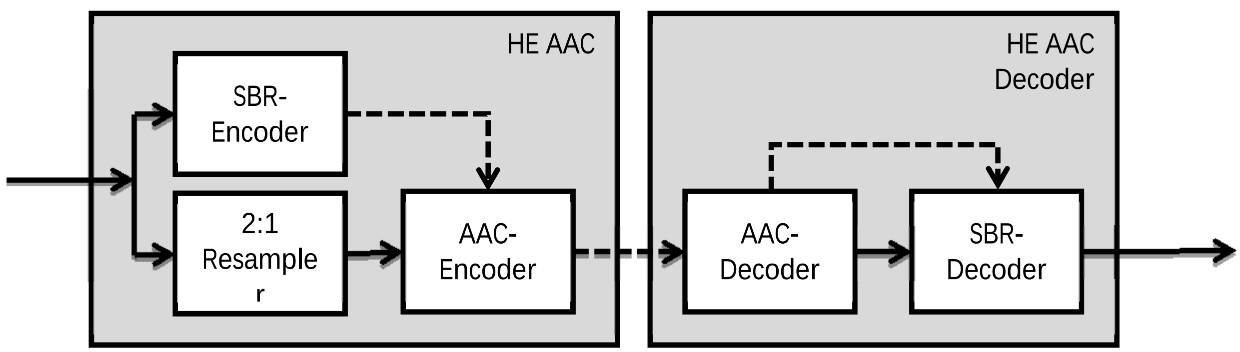

Figure 4.3 Simple block diagram showing how AAC and SBR work together.

To achieve an even further bitrate reduction, the number of discrete coded audio channels may be reduced by utilizing the Parametric Stereo tool in HE AAC v2. In this case, the two-channel input signal is down-mixed to a mono channel (AAC core channel) for coding and a parametric description of the stereo representation is added to the bit-stream payload. A HE AAC v2 decoder first creates this one-channel mono output signal and then renders the 2-channel output by utilizing the additional parametric data. This process is illustrated in Figure 4.4.

[1] "Audio spectrum" in this case is the audio bandwidth to be coded, which will vary depending on the sampling rate being used.

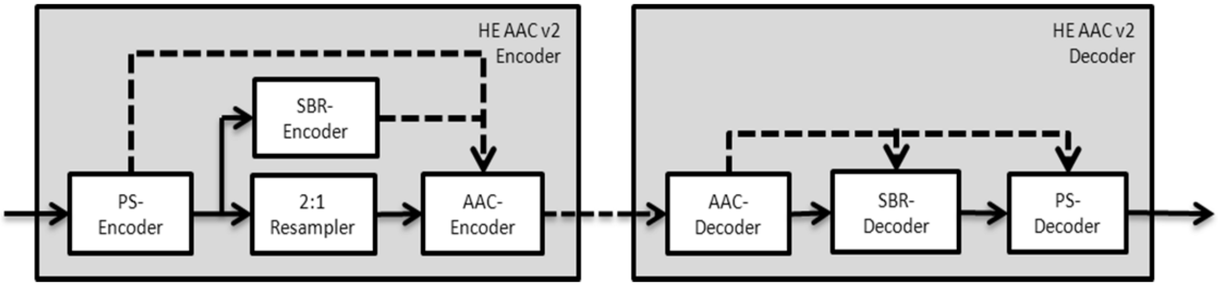

Figure 4.4 Simple block diagram showing how AAC, SBR and PS work together.

5. HE AAC V2 CONSTRAINTS

The audio content at the input to the HE AAC v2 encoder shall have a target measured loudness[2] value of –14 LKFS.

5.1 Audio Elementary Stream Configuration

The audio elementary streams shall conform to ISO/IEC 14496-3 [1] "High Efficiency AAC v2" Profile, Level 2. The definitions of Profiles and Levels for High Efficiency AAC v2 are listed in ISO/IEC 14496-3 [1] Table 1.11A. It is recommended that Dynamic Range Control (DRC) [1] information not be transmitted in the bitstream.

The AAC core codec sampling rate shall be constrained to 32, 44.1 or 48 kHz if no SBR is present, or to 16 kHz, 22.05 kHz, and 24 kHz if SBR is present (see Table 5.1).

The maximum bitrate shall meet the AAC bit buffer requirements as specified in ISO/IEC 14496-3 [1] by the equation in paragraph 4.5.3.3.

> Note: The maximum bit rate is dependent on the sampling frequency of the AAC core codec. According to the restriction on sampling frequencies made by this document for High Efficiency AAC v2 Profile – Level 2, valid AAC core codec sampling frequencies are noted above, and their resulting maximum bitrates are shown in Table 5.1. The maximum bitrates are not influenced by the usage of the HE AAC v2 profile.

Table 5.1 Valid Audio Sampling Frequencies and Maximum Bitrates

AAC Core Codec Sampling Frequency	Audio Stream Component Output Sampling Frequency	SBR Present	Maximum Bitrate / AAC Core Channel
48 kHz	48 kHz	N	288 kBit/s
44.1 kHz	44.1 kHz	N	264.6 kBit/s
32 kHz	32 kHz	N	192 kBit/s
24 kHz	48 kHz	Y	144 kBit/s
22.05 kHz	44.1 kHz	Y	132.3 kBit/s
16 kHz	32 kHz	Y	96 kBit/s

[2] Methods to measure loudness are explained in the ATSC Recommended Practice A/85, "Techniques for Establishing and Maintaining Audio Loudness for Digital Television" [9]. See particularly Section 5.2.

The presence of SBR and PS in the audio stream shall be indicated by the usage of explicit hierarchical signaling. Therefore the audio stream signaling shall be indicated as follows:

- If SBR data is not present, the audioObjectType indicated by the AudioSpecificConfig, shall be set to the value 2 (indicating AAC LC as given by Table 1.1 of ISO/IEC 14496-3 [1]).
- If SBR data is present, the first audioObjectType indicated by the AudioSpecificConfig, shall be set to the value 5 (indicating HE AAC as given by Table 1.1 of ISO/IEC 14496-3 [1]).
- If SBR data and PS data are present, the first audioObjectType indicated by the AudioSpecificConfig, shall be set to the value 29 (indicating HE AAC v2 as given by Table 1.1 of Amendment 2 of ISO/IEC 14496-3 [1]).

5.2 RTP Packetization

HE AAC v2 audio elementary streams shall be packetized in RTP packets according to IETF RFC 3640 [3]. Each individual AU-header shall only contain the required fields in AAC-hbr mode as described by RFC 3640 [3]. The AU-size field shall use 13 bits and AU-Index respectively AU-Index-delta shall use 3 bits. RFC 3640 [3] requires that the concatenated AU-headers in the AU-header-section be preceded by the 16 bit AU-headers-length field which is required to indicate the overall size of the available AU-headers within the RTP payload.

The packetization mode shall be “AAC-hbr” as defined in RFC 3640 [3]. Access units shall be transmitted in directly increasing time order. Access unit duration shall be constant, and is signaled in the field constant_duration per 7.8.1.2 of ATSC A/153 Part 3 [7].

The signaling of the RTP payload format—i.e., the relevant audio/mpeg4-generic media type parameters (as defined in RFC 3640 [3])—is defined in A/153 Part 3 [7].

The RTP_clock_rate parameter is specified in A/153 Part 3 [7] and corresponds to the rate parameter from the rtpmap attribute defined by RFC 3640 [3]. It signals the time base for RTP time stamping.

Annex A: Relationship between MH Component Data Descriptor and SDP

ATSC M/H transmits SDP messages according to RFC 4566 [8] for announcement of services. For signaling of audio codec capabilities, however, the MH_component_descriptor() with the MH_component_data() structure (for Component Type 37) as defined in Section 7.8.1.3 of A/153 Part 3 [7] is used. The MH_component_data() structure and the SDP messages carry many of the same parameters. It is strongly recommended to use the MH_component_descriptor() with the MH_component_data() structure for initialization of the audio decoder because the MH_component_descriptor() is defined to take precedence over the SDP message.

The following section explains the elements in an example SDP message to help clarify how the audio signaling in ATSC M/H is related to SDP. Consider the following SDP message:

```
1  c = IN IP4 192.0.2.1 / 127
2  m=audio 5000 RTP/AVP 37
3  a=rtpmap:37 mpeg4-generic/48000/2
4  a=fmtp:37 streamType=5; profile-level-id=48; mode=AAC-hbr; config=EB098800
   sizeLength=13; indexLength=3; indexDeltaLength=3; constantDuration=2048
```

Within this SDP message,

1) Lines 2 – 4 describe the session information for the HE AAC v2 layer.
2) Lines 2 and 3 describe the use of the audio/mpeg4-generic RTP payload format, as specified in RFC 3640 [3]. The RTP time stamp clock rate in this example is 48 kHz, and the number of audio channels is two.
3) Line 4 describes the required media format packetization parameters from RFC 3640 [3] and is in line with the requirements specified in Section 5.2 above.
4) Line 4 also describes the media format parameters for the HE AAC v2 bitstream. The bitstream is coded in HE AAC v2 Profile at Level 2 (profile-level-id=48) and the config string contains the hexadecimal representation of the HE AAC v2 AudioSpecificConfig [audioObjectType=2 (AAC LC); extensionAudioObjectType=5 (SBR); psPresentFlag = 1; samplingFrequencyIndex=0x6 (24kHz); extensionSamplingFrequencyIndex=0x3 (48kHz); channelConfiguration=1 (1.0 channels for the AAC LC part)].

Some possible config strings are listed below in Table A.1. Please note that Table A.1 is not exhaustive even for the listed set of sampling frequencies and channel modes, but just contains examples. The AudioSpecificConfig is defined in [1] Table 1.13.

Table A.1 Some Example HE AAC v2 AudioSpecificConfig Strings

Sampling Frequency	AAC LC Mono	AAC LC Stereo	HE AAC Mono [a]	HE AAC Stereo	HE AAC v2 Stereo
32 kHz	1288	1290	2C0A8800	2C128800	EC0A8800
44.1 kHz	1208	1210	2B8A0800	2B920800	EB8A0800
48 kHz	1188	1190	2B098800	2B118800	EB098800

[a] These values also apply to HE AAC v2 mono.

The MH_component_data() structure contains information which is partially present on different levels inside an SDP message as illustrated below in Table A.2, with the example values shown above as contents of an SDP file.

Data about the audio stream (e.g., language), along with the needed parameters from the SDP file are placed into a structure forming an octet string. The correspondence is shown in Table A.2 and described below the table. This results in an octet string which is the configuration record of the MH Component Data for HE AAC v2 (Type 37), as specified in A/153 Part 3 [7].

Backwards-compatible audio extensions (e.g., for multi-channel surround sound) rely on the possibility of transmitting additional config strings. This is enabled by the specified loop in the MH Component Data for HE AAC v2 (Component Type 37).

Since the length of config strings is available in ATSC M/H it is possible to either parse all config strings or skip additional ones with potentially unknown content.

Table A.2 ATSC MH Component Descriptor Data

MH_component_data()	Size	Contents	Associated with information from:
ISO_639_language_code	3 * 8	0x656E67	SDP a=lang:eng
reserved	6	‘111111’	
RTP_clock_rate	18	48000	SDP rtpmap section
constant_duration	16	2048	SDP fmtp section
sampling_rate	4	3	SDP fmtp config string
audio_service_type	4	0	Complete Main service (CM)
audio_channel_association	8	0xF8	First audio service
reserved	4	‘1111’	
num_configs	4	1	Number of available SDP fmtp config strings
for(num_configs)			
profile_level_id	8	48	SDP fmtp section
num_audio_channels	4	2	SDP fmtp config string
reserved	4	‘1111’	
config_size	8	4	SDP fmtp config string (size)
config	config_size * 8	0xEB098800	The SDP fmtp config string

The values in the fields sampling_rate and num_audio_channels may be obtained from the AudioSpecificConfig embedded in the config field of an SDP message, when an SDP message is part of the source data flow. If SBR data is not present, the samplingFrequencyIndex parameter in the MH_component_data() structure corresponds to the sampling_frequency_index parameter from the hexadecimal AudioSpecificConfig string within the descriptor. If SBR data is present, the sampling_rate parameter in the MH_component_data() structure corresponds to the extensionSamplingFrequencyIndex parameter from the hexadecimal AudioSpecificConfig string within the descriptor. The num_audio_channels parameter in the descriptor indicates the number of audio channels to be rendered. This information may also be obtained from the channelConfiguration parameter inside the AudioSpecificConfig in conjunction with the information from the profile_level_id. The base-16 encoded representation of the config string from the above example SDP message is directly placed in the config field of the descriptor.

The user should note that an ISO/IEC 14496-3 [1] compliant standard HE AAC v2 decoder is required to render two channels when a genuine mono stream is signaled and sent.