Doc. CS/T3-614A
8 September 2004

Candidate Standard Amendment 2 to ATSC Digital Television Standard, Doc. A/53C

Annex G: High Efficiency Audio System Characteristics

Advanced Television Systems Committee
1750 K Street, N.W.
Suite 1200
Washington, D.C. 20006
www.atsc.org

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 Television Association (NCTA), and the Society of Motion Picture and Television Engineers (SMPTE). Currently, there are approximately 140 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.

About the Candidate Standard

This specification is being put forth as a Candidate Standard by the T3/S6 Specialist Group on Video and Audio Coding. ATSC members and non-members are encouraged to review and implement this specification and return comments to cs_amend_editor@atsc.org. ATSC Members can also send comments directly to the T3/S6 Specialist Group. The ATSC believes this specification is stable. It is expected to progress to Proposed Standard within a period of time ending 10 May 2005.

This Amendment adds the following Annex G to A/53C:

Annex G:
High Efficiency Audio System Characteristics (Normative)

1. SCOPE

This Annex describes the robust mode audio system characteristics and normative specifications of the Digital Television Standard. Audio encoded per this Annex may be transmitted over a TS-E (see Annex C).

2. NORMATIVE REFERENCES

The following documents contain provisions which in whole or part, through reference in this text, constitute provisions of this standard. At the time of publication, the editions indicated were valid. All standards are subject to revision and amendment, and parties to agreement based on this standard are encouraged to investigate the possibility of applying the most recent editions of the documents listed below.

[G1] ATSC Standard A/52B (2004), "Digital Audio Compression (AC-3)."

[G2] AES3-2003, "AES Recommended practice for digital audio engineering "Serial transmission format for two-channel linearly represented digital audio data (Revision of AES3-1992)"

[G3] ANSI S1.4-1983 (R 2001) with Amd.S1.4A-1995, "Specification for Sound Level Meters."

3. COMPLIANCE NOTATION

As used in this document, "shall" or "will", denotes a mandatory provision of the standard. "Should" denotes a provision that is recommended but not mandatory. "May" denotes a feature whose presence does not preclude compliance, and that may or may not be present at the option of the implementer.

4. SYSTEM OVERVIEW

As illustrated in Figure G1, the audio subsystem comprises the audio encoding/decoding function and resides between the audio inputs/outputs and the transport subsystem. The audio encoder(s) is (are) responsible for generating the audio elementary stream(s) which are encoded representations of the baseband audio input signals. At the receiver, the audio subsystem is responsible for decoding the audio elementary stream(s) back into baseband audio.

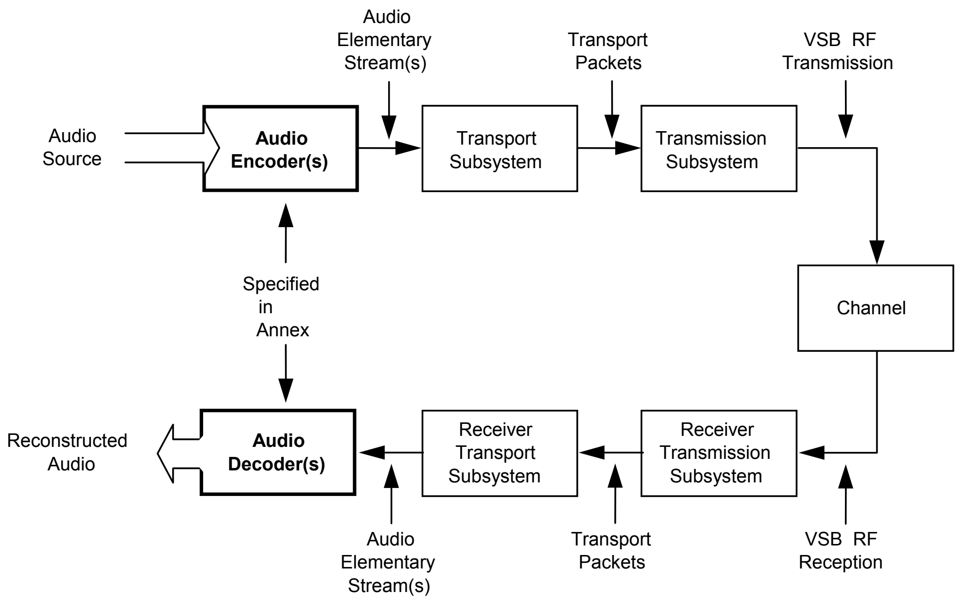

Figure G1 Audio subsystem in the digital television system.

5. SPECIFICATION

This Section forms the normative specification for the robust mode audio system that may be transmitted as part of TS-E (see Annex C). The robust mode audio compression system conforms to Annex E of the A/52B Digital Audio Compression (AC-3) Standard, subject to the constraints outlined in this Section.

5.1 Constraints With Respect to ATSC Standard A/52B Annex E

The robust mode digital television audio coding system is based on the Enhanced AC-3 Digital Audio Compression Standard specified in Annex E of ATSC Doc. A/52B. Audio bit streams encoded per that specification may be included the TS-E that is delivered by E-VSB. Constraints on the robust mode audio system are shown in Table G1, which shows permitted values of certain syntactical elements. These constraints are further described in Sections 5.2–5.4, and Section 6.

Table G1 Audio Constraints

AC-3 Syntactical Element	Comment	Allowed value
fscod	Indicates sampling rate	‘00’ (indicates 48 kHz)
frmsize	Indicates the size of the audio frame	≤ ’011 1000 0000’ (indicates a frame size ≤ 448 kb/s for a six block frame)
bstyp	Indicates an independent stream (no sub-streams)	‘00’
acmod	Indicates number of channels, prohibits 1+1 mode	≥ ‘001’
bsmod	Restricts audio service types to CM, VI, HI, C	0, 2, 3, or 5

5.2 Sampling Frequency

The system conveys digital audio sampled at a frequency of 48 kHz that shall be locked to the 27 MHz MPEG-2 system clock. The 48 kHz audio sampling clock is defined as:

$$48 \text{ kHz audio sample rate} = (2 \div 1125) \times (27 \text{ MHz MPEG-2 system clock})$$

If analog signal inputs are employed, the A/D converters shall sample at 48 kHz locked to the 27 MHz clock. If digital inputs are employed, the input sampling rate shall be 48 kHz locked to the system clock, or the audio encoder shall contain sampling rate converters which convert the sampling rate to 48 kHz locked to the system clock.

5.3 Frame Size

The audio frame size shall be less than or equal to 1792 bytes. This implies a bit-rate limitation of 448 kb/s for AC-3 frames of 1536 samples (32 msec at 48 kHz).

5.4 Audio Coding Modes

Audio services shall be encoded using any of the audio coding modes specified in A/52, with the exception of the 1+1 mode. The value of acmod in the AC-3 bit stream shall have a value in the range of 1–7, with the value 0 prohibited.

5.5 Dialogue Level

The value of the dialnorm parameter in the AC-3 elementary bit stream shall indicate the level of average spoken dialogue within the encoded audio program. Dialogue level may be measured by means of an “A” weighted integrated measurement (LAeq) [G3}. (Receivers use the value of dialnorm to adjust the reproduced audio level so as to normalize the dialogue level.) In order to enable clean switching (i.e., without level shifts) between main and fallback audio services (that might have a different number of audio channels), linked audio services shall have values of dialnorm that result in matched dialogue levels when decoded by compliant decoders.

5.6 Dynamic Range Compression - Artistic

Each encoded audio block may contain a dynamic range control word (dynrng) that is used by decoders (by default) to alter the level of the reproduced audio. The control words allow the decoded signal level to be increased or decreased by up to 24 dB. In general, elementary streams may have dynamic range control words inserted or modified without affecting the encoded audio. When it is necessary to alter the dynamic range of audio programs that are broadcast, the dynamic range control word should be used. In order to enable clean switching between main and fallback audio services (that might have a different number of audio channels), linked audio services shall have values of dynrng that result in matched audio levels when decoded by compliant decoders.

5.7 Dynamic Range Compression - Heavy

Each encoded audio frame may contain a dynamic range control word (compr) that may be optionally used by decoders to render the audio with a very narrow dynamic range. The control words allow the decoded signal level to be increased or decreased by up to 48 dB. In order to enable clean switching between main and fallback audio services (that might have a different number of audio channels), linked audio services shall have values of compr that result in matched audio levels when decoded by compliant decoders.

6. MAIN AND ASSOCIATED SERVICES

An AC-3 elementary stream contains the encoded representation of a single audio service. Multiple audio services are provided by multiple elementary streams. Each elementary stream is conveyed by the transport multiplex with a unique PID. There are a number of audio service types that may (individually) be coded into each elementary stream. Each AC-3 elementary stream is tagged as to its service type using the bsmod bit field. There is a *complete main service* and there are three types of *associated services*.

Associated services delivered in a TS-E shall contain complete program mixes containing all audio program elements (dialog, music, effects, etc.) that are intended to be presented to a listener. This is indicated by the full_svc bit in the AC-3 descriptor being set to a value of '1' (see Annex C of this document and A/52B, Annex A).

This section specifies the meaning and use of each type of service.

6.1 Summary of Service Types

The audio service types are listed in Table G2.

Table G2 Audio Service Types

bsmod	Type of Service
000 (0)	Main audio service: complete main (CM)
010 (2)	Associated service: visually impaired (VI)
011 (3)	Associated service: hearing impaired (HI)
101 (5)	Associated service: commentary (C)

6.2 Complete Main Audio Service (CM)

The CM type of main audio service contains a complete audio program (complete with dialogue, music, and effects). This is the type of audio service normally provided. The CM service may contain from 1 to 5.1 audio channels. Audio in multiple languages may be provided by supplying multiple CM services, each in a different language.

6.3 Visually Impaired (VI)

The VI associated service a complete program mix containing music, effects, dialogue, and additionally a narration that describes the picture content. The VI service may be coded using any number of channels (up to 5.1).

6.4 Hearing Impaired (HI)

The HI service is a complete program mix containing music, effects, and dialogue with enhanced intelligibility. The HI service may be coded using any number of channels (up to 5.1).

6.5 Commentary (C)

The commentary associated service is a complete program mix containing music, effects, dialogue, and additionally some special commentary. This service may be provided using any number of channels (up to 5.1).

7. AUDIO ENCODER INTERFACES

See Annex B, Section 7.