Working Draft Partitioning of ATSC Digital Television Standard (A/53), Revision E, Part 5 – AC-3 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.

Table of Contents

Index of Tables and Figures

ATSC Digital Television Standard – Part 5: AC-3 Audio System Characteristics

1. SCOPE

This Part describes the audio system characteristics and normative specifications of the Digital Television Standard.

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.

[1]   AES3-2003, "AES Recommended practice for digital audio engineering—Serial transmission format for two-channel linearly represented digital audio data," (revision of AES3-1992, including subsequent amendments).

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

[3]   ATSC A/52B: "Digital Audio Compression (AC-3, E-AC-3)," Advanced Television Systems Committee, Washington, D.C., 14 June 2005.

3. COMPLIANCE NOTATION

As used in this document, "shall" 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, that may or may not be present at the option of the implementor.

4. SYSTEM OVERVIEW (INFORMATIVE)

A basic block diagram representation of the system is shown in Figure 4.1. According to this model, the digital television system can be seen to consist of three subsystems.

- Source coding and compression
- Service multiplex and transport
- RF/transmission
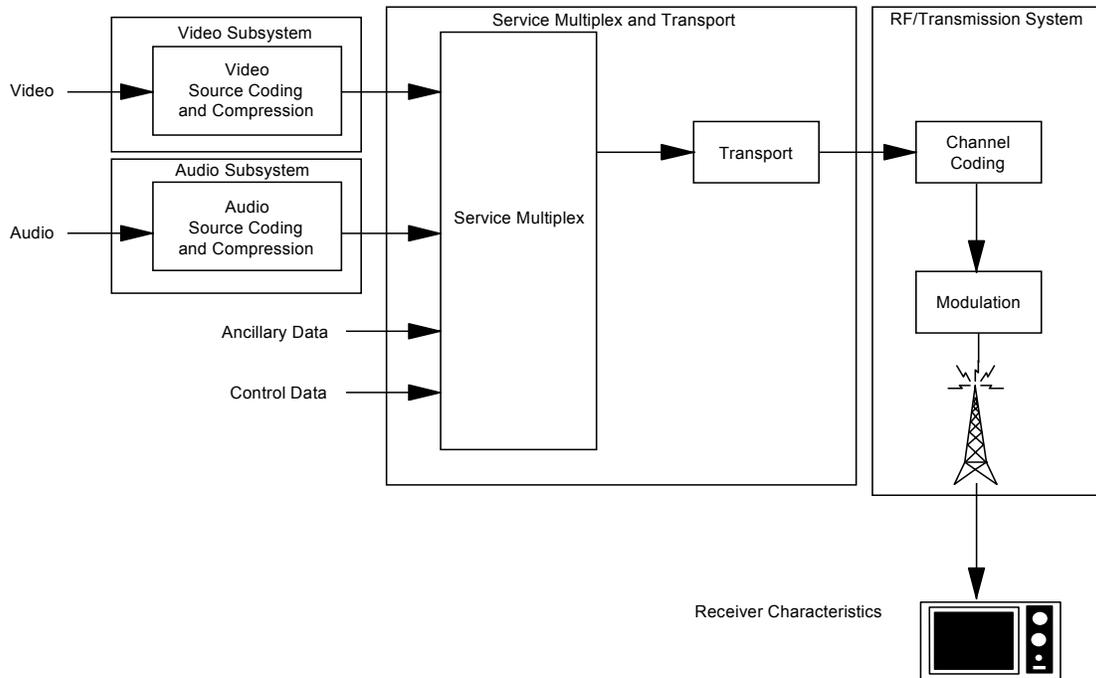
Figure 4.1 ITU-R digital terrestrial television broadcasting model.

Figure 4.2 illustrates a high level view of encoding equipment. This view is not intended to be complete, but is used to illustrate the relationship of various clock frequencies within the encoder.
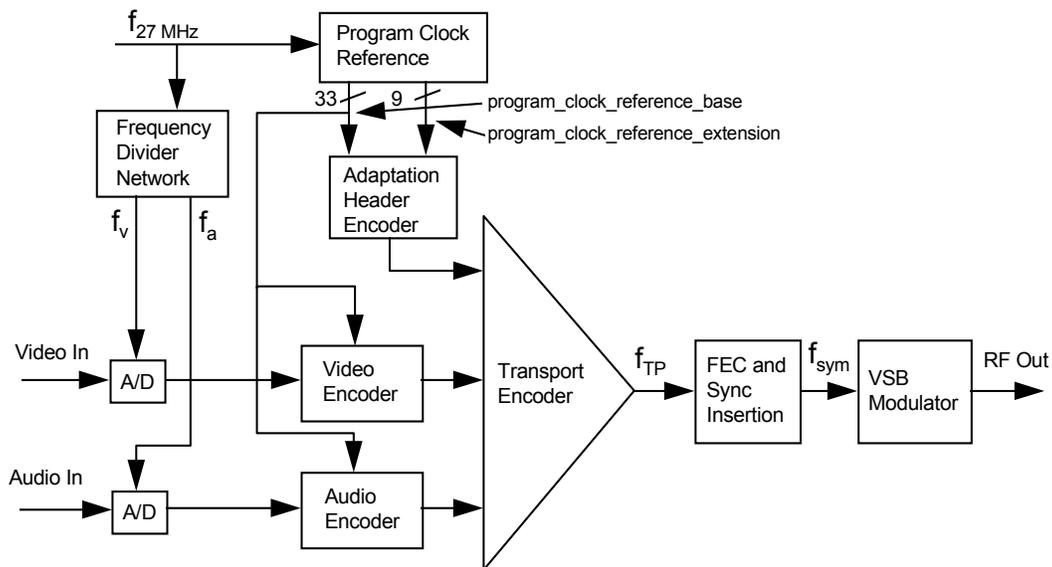
Figure 4.2 High level view of encoding equipment.

The source coding domain, represented schematically by the video, audio, and transport encoders, uses a family of frequencies which are based on a 27 MHz clock ($f_{27\text{MHz}}$). This clock is used to generate a 42-bit sample of the frequency which is partitioned into two parts defined by the MPEG-2 specification. These are the 33-bit program_clock_reference_base and the 9-bit program_clock_reference_extension. The former is equivalent to a sample of a 90 kHz clock which is locked in frequency to the 27 MHz clock, and is used by the audio and video source encoders when encoding the presentation time stamp (PTS) and the decode time stamp (DTS). The audio and video sampling clocks, f_a and f_v respectively, are frequency-locked to the 27 MHz clock. This can be expressed as the requirement that there exist two pairs of integers, (n_a, m_a) and (n_v, m_v), such that

$$f_a = \frac{n_a}{m_a} \times 27 \text{ MHz}$$

and

$$f_v = \frac{n_v}{m_v} \times 27 \text{ MHz}$$

As illustrated in Figure 4.3, 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.
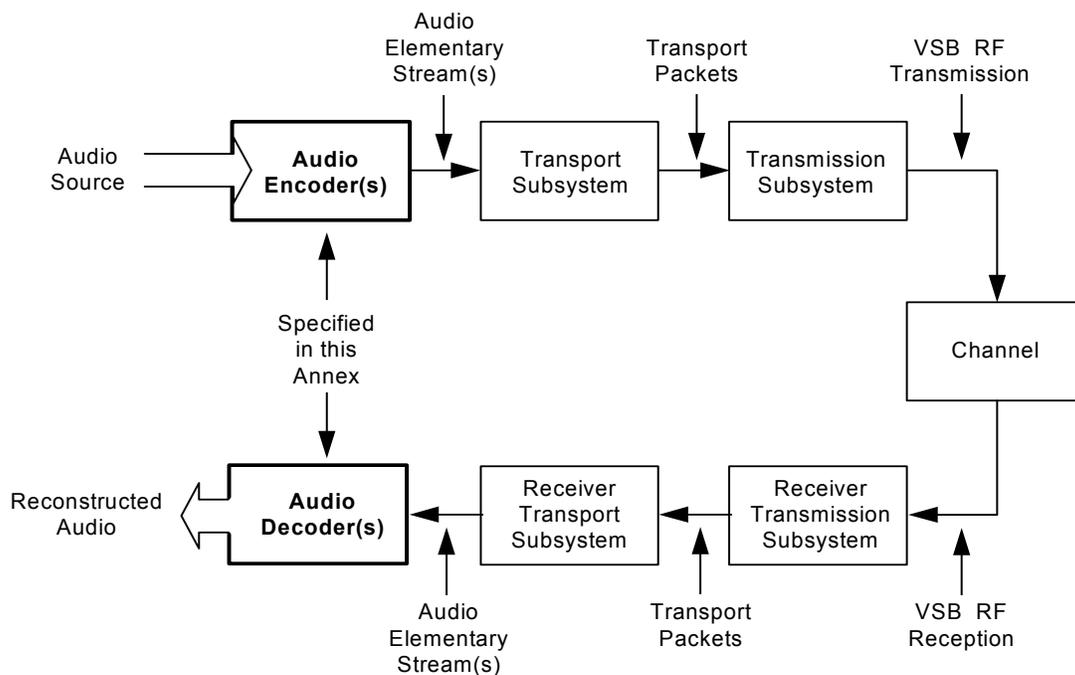
Figure 4.3 Audio subsystem in the digital television system.

5. SPECIFICATION

This Section forms the normative specification of the audio system. The audio compression system shall conform with the Digital Audio Compression (AC-3) Standard, subject to the constraints outlined in this Section.

5.1 Constraints With Respect to ATSC Standard A/52B

The digital television audio coding system is based on the Digital Audio Compression (AC-3) Standard specified in the body of ATSC Doc. A/52B (the non-normative annexes are not included). Constraints on the system are shown in Table 5.1, which shows permitted values of certain syntactical elements. These constraints are described in Sections 5.2 – 5.4. The receive audio buffer is specified in Section 5.7.

Table 5.1 Audio Constraints

AC-3 Syntactical Element	Comment	Allowed value
fscod	Indicates sampling rate	'00' (indicates 48 kHz)
frmsizecod	Main audio service or associated audio service containing all necessary program elements	≤ '011110' (indicates ≤ 448 kb/s)
frmsizecod	Single channel associated service containing a single program element	≤ '010000' (indicates ≤ 128 kbps)
frmsizecod	Two channel dialogue associated service	≤ '010100' (indicates ≤ 192 kbps)
(frmsizecod)	Combined bit rate of a main and an associated service intended to be simultaneously decoded	(total ≤ 576 kbps)
acmod	Indicates number of channels	≥ '001'

5.2 Sampling Frequency

The system conveys digital audio sampled at a frequency of 48 kHz, locked to the 27 MHz system clock. The 48 kHz audio sampling clock is defined as:

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

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

5.3 Bit Rate

A main audio service, or an associated audio service which is a complete service (containing all necessary program elements) shall be encoded at a bit rate less than or equal to 448 kbps. A single channel associated service containing a single program element shall be encoded at a bit rate less than or equal to 128 kbps. A two channel associated service containing only dialogue shall be encoded at a bit rate less than or equal to 192 kbps. The combined bit rate of a main service and an associated service which are intended to be decoded simultaneously shall be less than or equal to 576 kbps.

5.4 Audio Coding Modes

Audio services shall be encoded using any of the audio coding modes specified in A/52B [3], 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) (ANSI S1.4) [2]. (Receivers use the value of dialnorm to adjust the reproduced audio level so as to normalize the dialogue level.)

5.6 Dynamic Range Compression

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 which are broadcast, the dynamic range control word should be used.

5.7 STD Audio Buffer Size

The main audio buffer (BSn, see A/52B Annex A [3]) shall be 2592 bytes.

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 which 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 are two types of *main service* and six types of *associated service*. Each associated service may be tagged (in the AC-3 audio descriptor in the transport PSI data) as being associated with one or more main audio services. Each AC-3 elementary stream may also be tagged with a language code.

Associated services may contain complete program mixes, or may contain only a single program element. Associated services which are complete mixes may be decoded and used as is. They are identified by the full_svc bit in the AC-3 descriptor (see A/52B, Annex A [3]). Associated services which contain only a single program element are intended to be combined with the program elements from a main audio service.

This section specifies the meaning and use of each type of service. In general, a complete audio program (what is presented to the listener over the set of loudspeakers) may consist of a main audio service, an associated audio service that is a complete mix, or a main audio service combined with an associated audio service. The capability to simultaneously decode one main service and one associated service is required in order to form a complete audio program in certain service combinations described in this section. This capability may not exist in some receivers.

6.1 Summary of Service Types

The audio service types are listed in Table 6.1.

Table 6.1 Audio Service Types

bsmod	Type of Service
'000' (0)	Main audio service: complete main (CM)
'001' (1)	Main audio service: music and effects (ME)
'010' (2)	Associated service: visually impaired (VI)
'011' (3)	Associated service: hearing impaired (HI)
'100' (4)	Associated service: dialogue (D)
'101' (5)	Associated service: commentary (C)
'110' (6)	Associated service: emergency (E)
'111' (7)	Associated service: voice-over (VO)

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. The CM service may be further enhanced by means of the VI, HI, C, E, or VO associated services described below. Audio in multiple languages may be provided by supplying multiple CM services, each in a different language.

6.3 Main Audio Service, Music and Effects (ME)

The ME type of main audio service contains the music and effects of an audio program, but not the dialogue for the program. The ME service may contain from 1 to 5.1 audio channels. The primary program dialogue is missing and (if any exists) is supplied by simultaneously encoding a D associated service. Multiple D associated services in different languages may be associated with a single ME service.

6.4 Visually Impaired (VI)

The VI associated service typically contains a narrative description of the visual program content. In this case, the VI service shall be a single audio channel. The simultaneous reproduction of both the VI associated service and the CM main audio service allows the visually impaired user to enjoy the main multi-channel audio program, as well as to follow (by ear) the on-screen activity.

The dynamic range control signal in this type of VI service is intended to be used by the audio decoder to modify the level of the main audio program. Thus the level of the main audio service will be under the control of the VI service provider, and the provider may signal the decoder (by altering the dynamic range control words embedded in the VI audio elementary stream) to reduce the level of the main audio service by up to 24 dB in order to assure that the narrative description is intelligible.

Besides providing the VI service as a single narrative channel, the VI service may be provided as a complete program mix containing music, effects, dialogue, and the narration. In this case, the service may be coded using any number of channels (up to 5.1), and the dynamic

range control signal applies only to this service. The fact that the service is a complete mix shall be indicated in the AC-3 descriptor (see A/52B, Annex A [3]).

6.5 Hearing Impaired (HI)

The HI associated service typically contains only dialogue which is intended to be reproduced simultaneously with the CM service. In this case, the HI service shall be a single audio channel. This dialogue may have been processed for improved intelligibility by hearing impaired listeners. Simultaneous reproduction of both the CM and HI services allows the hearing impaired listener to hear a mix of the CM and HI services in order to emphasize the dialogue while still providing some music and effects.

Besides providing the HI service as a single dialogue channel, the HI service may be provided as a complete program mix containing music, effects, and dialogue with enhanced intelligibility. In this case, the service may be coded using any number of channels (up to 5.1). The fact that the service is a complete mix shall be indicated in the AC-3 descriptor (see A/52B, Annex A [3]).

6.6 Dialogue (D)

The D associated service contains program dialogue intended for use with an ME main audio service. The language of the D service is indicated in the AC-3 bit stream, and in the audio descriptor. A complete audio program is formed by simultaneously decoding the D service and the ME service and mixing the D service into the center channel of the ME main service (with which it is associated).

If the ME main audio service contains more than two audio channels, the D service shall be monophonic (1/0 mode). If the main audio service contains two channels, the D service may also contain two channels (2/0 mode). In this case, a complete audio program is formed by simultaneously decoding the D service and the ME service, mixing the left channel of the ME service with the left channel of the D service, and mixing the right channel of the ME service with the right channel of the D service. The result will be a two channel stereo signal containing music, effects, and dialogue.

Audio in multiple languages may be provided by supplying multiple D services (each in a different language) along with a single ME service. This is more efficient than providing multiple CM services, but, in the case of more than two audio channels in the ME service, requires that dialogue be restricted to the center channel.

Some receivers may not have the capability to simultaneously decode an ME and a D service.

6.7 Commentary (C)

The commentary associated service is similar to the D service, except that instead of conveying essential program dialogue, the C service conveys optional program commentary. The C service may be a single audio channel containing only the commentary content. In this case, simultaneous reproduction of a C service and a CM service will allow the listener to hear the added program commentary.

The dynamic range control signal in the single channel C service is intended to be used by the audio decoder to modify the level of the main audio program. Thus the level of the main

audio service will be under the control of the C service provider, and the provider may signal the decoder (by altering the dynamic range control words embedded in the C audio elementary stream) to reduce the level of the main audio service by up to 24 dB in order to assure that the commentary is intelligible.

Besides providing the C service as a single commentary channel, the C service may be provided as a complete program mix containing music, effects, dialogue, and the commentary. In this case the service may be provided using any number of channels (up to 5.1). The fact that the service is a complete mix shall be indicated in the AC-3 descriptor (see A/52B, Annex A [3]).

6.8 Emergency (E)

The E associated service is intended to allow the insertion of emergency or high priority announcements. The E service is always a single audio channel. An E service is given priority in transport and in audio decoding. Whenever the E service is present, it will be delivered to the audio decoder. Whenever the audio decoder receives an E type associated service, it will stop reproducing any main service being received and only reproduce the E service out of the center channel (or left and right channels if a center loudspeaker does not exist). The E service may also be used for non-emergency applications. It may be used whenever the broadcaster wishes to force all decoders to quit reproducing the main audio program and reproduce a higher priority single audio channel.

6.9 Voice-Over (V0)

The VO associated service is a single channel service intended to be reproduced along with the main audio service in the receiver. It allows typical voice-overs to be added to an already encoded audio elementary stream without requiring the audio to be decoded back to baseband and then re-encoded. It is always a single audio channel. It has second priority (only the E service has higher priority). It is intended to be simultaneously decoded and mixed into the center channel of the main audio service. The dynamic range control signal in the VO service is intended to be used by the audio decoder to modify the level of the main audio program. Thus the level of the main audio service may be controlled by the broadcaster, and the broadcaster may signal the decoder (by altering the dynamic range control words embedded in the VO audio elementary stream) to reduce the level of the main audio service by up to 24 dB during the voice-over.

Some receivers may not have the capability to simultaneously decode and reproduce a voice-over service along with a program audio service.

7. AUDIO ENCODER INTERFACES

7.1 Audio Encoder Input Characteristics

Audio signals which are input to the digital television system may be in analog or digital form. Audio signals should have any dc offset removed before being encoded. If the audio encoder does not include a dc blocking high-pass filter, the audio signals should be high-pass-filtered before being applied to the encoder. In general, input signals should be quantized to at least 16-bit resolution. The audio compression system can convey audio signals with up to 24-bit resolution. Physical interfaces for the audio inputs to the encoder may be defined as voluntary industry standards by the AES, SMPTE, or other standards organizations.

7.2 Audio Encoder Output Characteristics

Conceptually, the output of the audio encoder is an elementary stream which is formed into PES packets within the transport subsystem. It is possible that systems will be implemented wherein the formation of audio PES packets takes place within the audio encoder. In this case, the output(s) of the audio encoder(s) would be PES packets. Physical interfaces for these outputs (elementary streams and/or PES packets) may be defined as voluntary industry standards by SMPTE or other standards organizations.

<div align="center">End of document</div>