

CTA Bulletin

Recommended Practice for Selection and Presentation of DTV Audio

CTA-CEB21-A

October 2016

NOTICE

Consumer Technology Association (CTA)™ Standards, Bulletins and other technical publications are designed to serve the public interest through eliminating misunderstandings between manufacturers and purchasers, facilitating interchangeability and improvement of products, and assisting the purchaser in selecting and obtaining with minimum delay the proper product for his particular need. Existence of such Standards, Bulletins and other technical publications shall not in any respect preclude any member or nonmember of the Consumer Technology Association from manufacturing or selling products not conforming to such Standards, Bulletins or other technical publications, nor shall the existence of such Standards, Bulletins and other technical publications preclude their voluntary use by those other than Consumer Technology Association members, whether the standard is to be used either domestically or internationally.

Standards, Bulletins and other technical publications are adopted by the Consumer Technology Association in accordance with the American National Standards Institute (ANSI) patent policy. By such action, the Consumer Technology Association does not assume any liability to any patent owner, nor does it assume any obligation whatever to parties adopting the Standard, Bulletin or other technical publication.

This document does not purport to address all safety problems associated with its use or all applicable regulatory requirements. It is the responsibility of the user of this document to establish appropriate safety and health practices and to determine the applicability of regulatory limitations before its use.

This document is copyrighted by the Consumer Technology Association (CTA)™ and may not be reproduced, in whole or part, without written permission. Federal copyright law prohibits unauthorized reproduction of this document by any means. Organizations may obtain permission to reproduce a limited number of copies by entering into a license agreement. Requests to reproduce text, data, charts, figures or other material should be made to the Consumer Technology Association (CTA)™.

(Formulated under the cognizance of the CTA **R4.3 Television Data Systems Subcommittee.**)

Published by
©CONSUMER TECHNOLOGY ASSOCIATION 2016
Technology & Standards Department
www.cta.tech

All rights reserved

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

FOREWORD

This bulletin was developed under the auspices of the Consumer Technology Association (CTA) R4.3 TV Data Systems Subcommittee.

Copyright Consumer Technology Association
Provided by IHS Markit under license with CTA
No reproduction or networking permitted without license from IHS

(This page intentionally left blank.)

CONTENTS

1 Scope.....	1
2 General	1
2.1 References	1
2.1.1 Informative References	1
2.1.2 Informative Reference List.....	1
2.1.3 Informative Reference Acquisition	2
2.2 Definitions	2
2.3 Symbols and Abbreviations	3
3 Audio Metadata Elements in the Transport Stream	3
3.1 PMT Construction and Comments.....	4
3.1.1 Mandatory Elements.....	4
3.1.2 Other PMT Considerations	4
3.2 Audio Stream Considerations	4
3.3 Inconsistencies Between PSIP and PMT	5
3.4 Use of Metadata	5
4 Recommendations	7
4.1 User Set-Up	7
4.2 User Operation.....	7
4.3 User Interface Recommendations	8
4.4 Automated Audio Stream Selection Recommendations	8
5 Other Metadata	10
6 Receiver Processing Examples and Recommendations	10
6.1 Example 1 Overview	11
6.2 Example 2 Overview	11
6.3 Example 3 Overview	13
6.4 Example 4 Overview	15
Annex A Examples.....	16
A1. Example 1.....	16
A2. Example 2.....	18
A3. Example 3.....	21
A4. Example 4.....	25

Figures

Figure 1 Transport Stream Language Information	8
---	----------

(This page intentionally left blank.)

Recommended Practice for Selection and Presentation of DTV Audio

1 Scope

CTA-CEB21-A recommends the actions ATSC receivers with a single audio decoder should take in the presence or absence of AC-3 audio metadata carried in ATSC standards-based broadcasts. These services are defined in the ATSC references shown in the References section. CTA-CEB21-A includes recommendations and suggestions on how these receiver actions should interact with user selections via different implementation choices. CTA-CEB21-A is in no way intended to prohibit consumer device manufacturers from including additional features, and CTA-CEB21-A should not be interpreted as recommending an upper limit to device performance.

CTA-CEB21-A provides recommendations to manufacturers to facilitate user setup of audio features in the receiver without professional assistance. This bulletin provides guidance in the following areas for the reasons listed below:

- One purpose of the bulletin is to describe recommended receiver behavior in the presence of the minimum required audio metadata described by ATSC A/53.
- Another purpose of the bulletin is to describe recommended receiver behavior in the presence of additional optional audio metadata described by ATSC A/53.

This document is organized by first providing descriptions of the key metadata elements and the structural locations in the MPEG-2 Transport Stream (table sections, descriptors, etc.). The recommendations for using this information along with user settings to control audio stream selection are then provided.

Manufacturers are encouraged to develop and deploy updated firmware when practical to follow the recommendations of this bulletin.

2 General

2.1 References

2.1.1 Informative References

The following references contain provisions, which, through reference in this text, constitute informative provisions of this standard. At the time of publication, the editions indicated below were valid. All documents are subject to revision, 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.

2.1.2 Informative Reference List

- [1] ATSC A/52:2015, Digital Audio Compression (AC-3, E-AC3) Standard, Advanced Television Systems Committee.
- [2] ATSC A/53, ATSC Digital Television Standard, Parts 1: 2013, Part 2: 2011, Part 3: 2013, Part 4: 2009, Part 5: 2014, Part 6: 2013, Advanced Television Systems Committee.
- [3] ATSC A/54A, Guide to the Use of the ATSC Digital Television Standard, with Corrigendum No. 1, December 2003, Corrigendum No. 1 dated 20 December 2006, Advanced Television Systems Committee. (Withdrawn, 11 Jan 2015.)
- [4] ATSC A/65:2013, Program and System Information Protocol for Terrestrial Broadcast and Cable, 2013, Advanced Television Systems Committee.
- [5] ATSC A/69:2009, Program and System Information Protocol Implementation Guidelines for Broadcasters, Annex I, 2009, Advanced Television Systems Committee.
- [6] ATSC A/79, Conversion of ATSC Signals for Distribution to NTSC Viewers, December 2008, Advanced Television Systems Committee
- [7] CTA-2009-B, 2010, Receiver Performance Specification for Public Alert Receivers

- [8] CTA-CEB12-B R 2015, PSIP Recommended Practice, 2010, Consumer Technology Association.
- [9] ETSI TS-101154 V1.11.1, Digital Video Broadcasting (DVB); Specification for the use of Video and Audio Coding in Broadcasting Applications based on the MPEG-2 Transport Stream, 2012.
- [10] ANSI/SCTE 54 2015, Digital Video Service Multiplex and Transport System Standard for Cable Television, Society of Cable Telecommunications Engineers.
- [11] ISO 639-2, “Codes for the representation of Names of Languages – Part 2: Alpha-3 code, “ as maintained by the ISO 639/Joint Advisory Committee (ISO 639/JAC), <http://www.loc.gov/standards/iso639-2/iso639jac.html>
- [12] ISO/IEC IS 13818-1:2015, “Information technology – Generic coding of moving pictures and associated audio information: Systems”, 2015.
- [13] ANSI/SCTE 194-1 (2013) DTS-HD Audio System – Part 1: Coding Constraints for Cable Television
- [14] ANSI/SCTE 194-2 (2014) DTS-HD Audio System – Part 2: Constraints for Carriage over MPEG-2 Transport
- [15] SCTE 193-1 (2014) MPEG-4 AAC Family Audio System – Part 1 Coding Constraints for Cable Television
- [16] SCTE 193-2 (2014) MPEG-4 AAC Family Audio System – Part 2 Constraints for Carriage over MPEG-2 Transport

2.1.3 Informative Reference Acquisition

ATSC Standards

- Advanced Television Systems Committee (ATSC), 1766 K Street N.W., Suite 200, Washington, DC 20006; Phone 202-872-9160; Fax 202-872-9161; Internet <http://www.atsc.org/cms/index.php/standards/published-standards>

CTA Standards

- Global Engineering Documents, World Headquarters, 15 Inverness Way East, Englewood, CO USA 80112-5776; Phone 800-854-7179; Fax 303-397-2740; Internet <http://global.ihs.com>; Email: global@ihs.com

ETSI

- European Telecommunications Standards Institute, 650, route des Lucioles, 06921 Sophia-Antipolis Cedex, France; Phone +33 (0)4 92 94 42 00; Fax +33 (0)4 93 65 47 16; Internet <http://www.etsi.org>

SCTE Standards

- Global Engineering Documents, World Headquarters, 15 Inverness Way East, Englewood, CO USA 80112-5776; Phone 800-854-7179; Fax 303-397-2740; Internet <http://www.scte.org/standards>

2.2 Definitions

For the purposes of CTA-CEB21-A, the following definitions apply.

Audio Stream — The Transport Stream packets carrying the elementary stream which contains audio essence.

Complete Main — An audio service containing all audio components (dialog, music, and effects) for the program.

Described Video — Described Video is the Canadian name for audio consisting of narrated descriptions of a television program’s key visual elements. Such streams are properly signaled as Visually Impaired (VI) streams.

Hearing Impaired (HI) — The HI service is a complete program mix containing music, effects, and dialogue with enhanced intelligibility.

Visually Impaired (VI) - The VI service is a complete program mix containing music, effects, dialogue, and additionally a narrative description of the program content.

2.3 Symbols and Abbreviations

CM	Complete Main
DTV	Digital Television
EIT	Event Information Table (A/65)
HI	Hearing Impaired
MVPD	Multichannel Video Programming Distributor
PMT	Program Map Table
PSIP	Program and System Information Protocol (A/65)
SAP	Secondary Audio Program
VI	Visually Impaired

3 Audio Metadata Elements in the Transport Stream

This bulletin only addresses detailed audio stream field information and examples for the AC-3 audio system. References for developing examples for other audio systems are provided. The following list describes data structures that contain the audio metadata addressed in CTA-CEB21-A:

AC-3_audio_stream_descriptor()

This descriptor defines parameters relevant to a given AC-3 audio stream program element including the type of audio stream, language, sampling rate, surround mode (number of channels), and bit-rate. See A/52 [1]

E-AC-3 Audio Descriptor

This descriptor defines parameters relevant to a given E-AC-3 audio stream program element including the type of audio stream, language, sampling rate, surround mode (number of channels), and bit-rate. See A/52 [1]

DTS-HD Audio Descriptor

This descriptor defines parameters relevant to a given DTS-HD audio stream program element. See ANSI/SCTE 194-2[14]

MPEG AAC Descriptor

This descriptor defines parameters relevant to a given MPEG AAC audio stream program element. See ANSI/SCTE 193-2[16]

ISO_639_language_descriptor()

Until March, 2008, the **ISO_639_language_descriptor()** (MPEG-2 Systems [12]) was used to identify the language of a given audio program element. Many broadcast and cable programs continue to use this method. The language of an audio program element is now required to be signaled by the three-byte ISO-639 language field in the **AC-3_audio_stream_descriptor()**. See A/52 [1].

component_name_descriptor()

This is used to associate textual labels with an Elementary Stream component of a program. Its structure is in a tag-length-data format and the data portion is a **component_name_string** (see below). The **component_name_descriptor()** is required to be present whenever multiple audio tracks of the same type (same **bsmod** value) and

language are present in a program. Receivers are expected to use the text in the **component_name_descriptor()** in the user interface offering the user to choose which audio track to decode. See A/65 [4].

component_name_string

A Multiple String Structure giving the name of the associated Elementary Stream component. See A/65 [4].

3.1 PMT Construction and Comments

Programs delivered in the digital multiplex may include more than one audio service. The receiver should not infer anything from the order in which audio streams are listed in the PMT.

3.1.1 Mandatory Elements

ATSC A/53 Part 3:2009 Section 6.3 [2] requires at least one audio service be designated as a Complete Main (CM). All audio services are required to be constructed so that receivers can render each independently of any other (complete mixes). Some may be identified as Hearing Impaired (HI) or Visually Impaired (VI) which are authored for improved accessibility. When two or more audio streams of the same language and **bsmod** value are present, receivers can expect the **component_name_descriptor()** to be present in the audio metadata (as required in ATSC A/65 [4]). The **component_name_string** provides information describing the differences between the audio streams. In some cases, the **component_name_descriptor()** is a mandatory element that should be used to distinguish between audio streams (see Example 3 on page 13). In other cases, presence of the **component_name_descriptor()** could allow a receiver to build a more informative user interface. (See A/69 [5], Annex I)

3.1.2 Other PMT Considerations

Audio language is required to be signaled in the **AC-3_audio_stream_descriptor()** by the 3-byte **language** field. In earlier versions of the ATSC standards, audio language was signaled in the **AC-3_audio_stream_descriptor()** by fields called **langcod** and **langcod2**. It is expected that broadcasters will follow ATSC A/79 [6] to indicate which audio stream is primary or secondary; but this signaling is for MVPD equipment, and is not intended for consumer product use.

As of the date of this bulletin, some broadcasters and Multichannel Video Program Distributors manage PMT and PSIP audio metadata using systems that might not create real-time updates to transmitted tables. These systems might not alter the signaling data to exactly match the audio metadata of interstitial insertions (e.g. commercials). The result can be momentary inconsistencies in the transmitted metadata.

3.2 Audio Stream Considerations

Receivers should ignore any **langcod** and **langcod2** information present in the transport stream. These fields can appear in the audio elementary stream or in the AC-3 descriptor. See ATSC A/52 [1].

Note – Receivers are expected to gracefully handle legacy transport streams containing one or more **AC-3_audio_stream_descriptor()**s that have non-null **langcod** or **langcod2** fields.

Once an audio stream is selected for decoding, the metadata in the stream itself provides all information the decoder needs to render that stream.

Use of Audio Systems other than AC-3

The following audio system descriptor information is provided to emphasize the differences between the audio system information sets and to provide some guidance when systems other than AC-3 are used. For the use of audio systems other than AC-3, the audio descriptor associated with that audio system should be used to provide users with the ability to setup and select languages and audio services.

E-AC-3 Audio Descriptor

When an elementary stream of stream_type 0x87 (E-AC-3 audio) is present in the digital television transport stream, an E-AC-3 Audio Descriptor (E-AC-3_audio_descriptor()) is included in the descriptor loop immediately following the ES_info_length field in the TS_program_map_section() describing that elementary stream. The syntax is given in Table G.1 of Annex G of ATSC Standard A/52[1]. The value of the descriptor_tag is 0xCC. The constraints specified in sections 5.8.1.3 and 5.8.1.4 of ATSC Standard A/53 Part 3 [2] apply to the E-AC-3 Audio Descriptor.

Receivers should use the metadata included in the E-AC-3 Audio Descriptor to provide users with the ability to setup and select languages and audio services.

DTS-HD Audio Descriptor

The DTS-HD Audio Descriptor is defined in ANSI/SCTE 194-2[14]. When an Elementary Stream of stream_type 0x88 (DTS-HD audio) is present in the digital television transport stream, a DTS-HD Audio Descriptor is included in the descriptor loop immediately following the ES_info_length field in the TS_program_map_section() describing that Elementary Stream.

Receivers should use the metadata included in the DTS-HD Audio Descriptor to provide users with the ability to setup and select languages and audio services.

MPEG AAC Descriptor

The MPEG AAC Descriptor is defined in SCTE 193-2[16]. When an Elementary Stream of stream_type 0x0F or 0x11 (AAC audio) is present in the digital television transport stream, an MPEG AAC Descriptor is included in the descriptor loop immediately following the ES_info_length field in the TS_program_map_section() describing that Elementary Stream.

Receivers should use the metadata included in the MPEG AAC Descriptor to provide users with the ability to setup and select languages and audio services.

3.3 Inconsistencies Between PSIP and PMT

In the event that **AC-3_audio_stream_descriptor()** data is inconsistent between EIT-0 and PMT, receivers should use **AC-3_audio_stream_descriptor()** data from the PMT.

Situations may be encountered in which the broadcast emission does not follow established standards. Receivers may encounter broadcast streams that have only a single audio stream with no audio metadata present or broadcast streams that do not have a primary audio stream assigned. If this occurs, the receiver should decode an audio stream, and not produce an error message.

3.4 Use of Metadata

A receiver should parse the PMT and each elementary stream's fields and descriptors to build an understanding of each stream's content.

For the purposes of display to the user and audio stream selection, receivers should use the **AC-3_audio_stream_descriptor()** as the source of metadata about the audio stream. For the use of audio systems other than AC-3, the audio descriptor associated with that audio system should be used to provide users with the ability to setup and select languages and audio services.

For the purposes of rendering the audio stream, receivers should use the metadata fields from the elementary stream. While some additional metadata elements are present in this elementary stream, use of those metadata elements to select audio is not addressed in this recommended practice. Examples include **dynrng**, **dsurmod**, **fscod**, and copyright bit. See ATSC A/52 [1].

For example, a receiver may find that the metadata contained in an elementary stream will dynamically track the number of channels and bit-rate of the audio for an inserted commercial advertisement, while the corresponding information in the AC-3 descriptor may apply to the program as a whole.

The PMT has data for the current stream, and the EIT has a subset of audio metadata for current and future events. If a receiver allows selection of a future event or an event on a channel that is not currently tuned (through a timer feature e.g. DVR, timer feature, etc.), the receiver should identify the corresponding PMT at the time the program is broadcast, and use it to confirm the currently selected the audio stream.

In case the scheduled selection's metadata does not match the PMT metadata, the receiver should use the PMT instead. A receiver that can make recordings should record all available audio streams including information that will allow users to make selections during playback.

In certain cases cached audio metadata may be used to reduce channel change time. See CTA-CEB12 [8] for example strategies.

3.5 Use of the ISO 639 Language Descriptor for Signaling the Visually Impaired Audio Track

The ISO 639 Language Descriptor is used today by many deployed digital ATSC receivers to allow users to select the second AC-3 audio track. This is compatible with broadcasters and cable systems that support the delivery of only two audio tracks within a program, and with programs that carry only two audio tracks. The second audio track, when present, might actually be Spanish language audio, or it might be Video Description audio (also called Descriptive Video or visually impaired audio). The second audio track might be signaled using the language value of "spa" (Spanish) in the ISO 639 Language Descriptor, even if it is Video Description audio. It might also be signaled using the language code for Portuguese or Middle English.

ATSC A/53 Part 3 [2] has made optional the carriage in the PMT of the ISO 639 Language Descriptor [11]. Many broadcasters nonetheless transmit the ISO 639 Language Descriptor [11] to accommodate deployed receivers, in addition to transmitting the AC-3 Audio Descriptor. Many deployed receivers do not process the AC-3 Audio Descriptor.

In order to accomplish a transition from this method of signaling the Video Description audio track in use today (for example, as Spanish in the ISO 639 Language Descriptor) to the recommended method using the AC-3 Audio Descriptor, the language in the AC-3 Audio Descriptor should correctly reflect the language of the Video Description audio track. It need not agree with the language in the ISO 639 Language Descriptor during the transition period. The recommended method uses the the language and bsmod value in the AC-3 Audio Descriptor.

As an example, consider a program that includes two AC-3 audio components, an English Complete Main audio track and an English Video Description audio track. In the AC-3 Audio Descriptor, they are signaled as (language = "eng", bsmod = 0) for the Complete Main audio and (language = "eng", bsmod = 2) for the Video Description audio. For use with deployed receivers, in the ISO 639 Language Descriptor they are signaled as (language = "eng") for the Complete Main audio and (language = "spa") for the Video Description audio. Thus, the language of the Video Description audio in the AC-3 Audio Descriptor does not agree with the language of the second audio track in the ISO 639 Language Descriptor. This accommodates digital ATSC receivers that do not process the AC-3 Audio Descriptor.

3.6 Current Broadcast Practice

While all broadcasters carry the AC-3 Audio Descriptor, not all broadcasters accurately populate the values for language and audio service. Rather, some broadcasters set those values to match the information in the ISO 639 Language Descriptor. In this case, the second audio track is often labeled as Spanish and the bsmod value is 0 to indicate Complete Main, even when the second audio track is English language and Video Description audio. The values in the AC-3 Audio Descriptor are "fixed" rather than "dynamic" even when the type of service in the second audio track changes from one program to another.

Moreover, some broadcasters that deliver programming created with Spanish language audio as the dialog have indicated that they signal that audio as English and signal the second audio track (carrying dubbed English dialog) as Spanish. Some foreign language broadcasters signal additional audio track as "mul" to indicate multilingual content. Some broadcasters that deliver programs in numerous languages signal all of them as "mul" regardless of the actual language.