ATSC Technology Group Report: ATSC 3.0 Initial AC-4 Implementation

Doc. TG3-181r0
24 May 2019

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 and recommended practices for digital television. ATSC member organizations represent the broadcast, broadcast equipment, motion picture, consumer electronics, computer, cable, satellite, and semiconductor industries. ATSC also develops digital television implementation strategies and supports educational activities on ATSC standards. ATSC was formed in 1983 by the member organizations of the Joint Committee on Inter-society Coordination (JCIC): the Electronic Industries Association (EIA), the Institute of Electrical and Electronic Engineers (IEEE), the National Association of Broadcasters (NAB), the National Cable Telecommunications Association (NCTA), and the Society of Motion Picture and Television Engineers (SMPTE). For more information visit www.atsc.org.

Note: The user's attention is called to the possibility that compliance with this document may require use of an invention covered by patent rights. By publication of this document, 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.

Implementers with feedback, comments, or potential bug reports relating to this document may contact ATSC at https://www.atsc.org/feedback/.

Revision History

Version	Date
Initial version approved	24 May 2019
Revision 1 approved	Insert date here

Table of Contents

Index of Figures

ATSC Technology Group Report: ATSC 3.0 Initial AC-4 Implementation

1. SCOPE AND INTRODUCTION

This is an ATSC Technology Group Report prepared by TG3. This is a document that incorporates consensus on information regarding ATSC Standards and related industry activities.

This document provides both general information and recommendations for the early implementation of the AC-4 audio system used with the ATSC 3.0 Digital Television System. It describes:

- An AC-4 feature overview.
- Recommendations for how encoding should be performed when transitioning from AC-3 to AC-4.
- Requirements for content production and formatting that enable new features for viewers.
- The characteristics and features of the AC-4 audio system that when signaled correctly, can enhance the media experience.
- How AC-4 can be utilized in a wide range of viewing and listening environments.
- The need to deliver a satisfying experience for consumers, which means content needs to be correctly produced and formatted, and new controls need to be presented to viewers in a consistent, user-friendly manner.

Review of a list of AC-4 features that resonated well with test audiences during recent consumer surveys indicates that the availability and a simple and clear presentation of audio features is important to viewers. This report describes and recommends the features that can be delivered by broadcasters in the initial phase of the ATSC 3.0 transition. Concerns about conversion to and from legacy streams during this phase are also addressed.

2. REFERENCES

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

2.1 Informative References

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

[1] IEEE: "Use of the International Systems of Units (SI): The Modern Metric System," Doc. SI 10, Institute of Electrical and Electronics Engineers, New York, NY.

[2] ATSC: "ATSC Standard: A/342 Part 1, Audio Common Elements," Doc. A/342-1:2017, Advanced Television Systems Committee, Washington D.C., 24 January 2017.

[3] ATSC: ATSC A/52:2018, "ATSC Standard: Digital Audio Compression (AC-3, E-AC-3)", Advanced Television Systems Committee, Washington, D.C., 25 January 2018.

[4] ATSC: ATSC A/85:2013, "ATSC Recommended Practice: Techniques for Establishing and Maintaining Audio Loudness for Digital Television", Advanced Television Systems Committee, Washington, D.C., 12 March 2013.

[5] ATSC: ATSC A/331:2017, "ATSC Standard: Signaling, Delivery, Synchronization, and Error Protection," Advanced Television Systems Committee, Washington, D.C., 6 December 2017.

[6] [5] FCC: "Consumer Guide: Video Description," Federal Communication Commission, Washington, D.C., 28 June 2018.

[7] AES: aes-71:2018, "Recommended Practice Loudness Guidelines for Over the Top Television and Online Video Distribution", Audio Engineering Society, New York, NY, 21 July 2018.

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 are used [1]. Where an abbreviation is not covered by IEEE practice or industry practice differs from IEEE practice, the abbreviation in question will be described in Section 3.1 of this document.

3.1 Acronyms and Abbreviations

The following acronyms and abbreviations are used within this document:

A/V – Audio/Video
AC – Audio Codec
AES – Audio Engineering Society
ATSC – Advanced Television Systems Committee
AVR – Audio/Video Receiver
CRC – Cyclic Redundancy Check
CTA – Consumer Technology Association
DASH – Dynamic Adaptive Streaming over HTTP
DE – Dialog Enhancement
DRC – Dynamic Range Control
DSI – Decoder Specific Information
EMDF – Extensible Metadata Delivery Format
FCC – Federal Communications Commission
fps – frames per second
HD-SDI – High Definition Serial Digital Interface
HEVC – High Efficiency Video Coding
IPTV – Internet Protocol Television
kb – kilo bits
kHz – kilo Hertz
LFE – Low Frequency Effects
LKFS – Loudness, K-weighted, relative to Full Scale
M&E – Music & Effects
M&E+D – Music & Effects with separate Dialog
MMT – MPEG Media Transport
MPD – Media Presentation Description
MVPD – Multichannel Video Programming Distributor
OTA – Over the Air
OTT – Over the Top
PCM – Pulse-Code Modulation
RAP – Random Access Point
SAP – Secondary Audio Programming
SDI – Serial Digital Interface

SMPTE – Society of Motion Picture and Television Engineers
STB – Set Top Box
TOC – (AC-4 bitstream) Table of Contents
TV – Television
VDS – Video Description Service

3.2 Terms

The following terms are used within this document:

Adaptation Set – A set of interchangeable encoded versions of one or several media content components.

Agile Metadata – Audio metadata values, including dialnorm, which can change at content boundaries. See ATSC A/85 [4].

Alternate Complete Main – An additional Complete Main service selectable by the viewer offering languages or Video Description that is different than the primary soundtrack's.

Alternate Dialog Commentary – A Dialog-only element selectable by the viewer offering commentary that is different than the primary Presentation's soundtrack.

Alternate Languages –Dialog-only elements selectable by the viewer offering languages that are different than the primary Presentation's soundtrack.

Associated Audio – An audio element that can be used in addition to or substituted for elements of the main audio soundtrack.

Audio Element – The smallest addressable unit of an audio program.

Audio Program Component – Logical group of one or more Audio Elements that is used to define an audio Presentation; e.g., Complete Main, Music & Effects, Dialog, etc.

Audio Stream Properties Descriptor – Signals the properties of an audio stream, such as language, speaker configuration and accessibility support.

Audio/Video Receiver – A consumer electronics device capable of decoding, switching and amplifying electronic media sources.

Clean Production – A production with a primary Audio Element that is isolated from any other Audio Element.

Codec – A system consisting of an encoder and decoder.

Complete Main – An audio program format consisting of an audio mix that includes all elements of a soundtrack.

Composite Mix – An audio mix that includes all elements of a soundtrack.

Controllable Dynamic Range – The process of continually adjusting audio signal level to control the level difference between loud and soft passages according to some desired objective (see AES-71 [7]).

dialnorm – An AC-3 metadata parameter, defined in A/52 [3], that is carried in the AC-3 bit stream. This is used to indicate how far the average Dialog level is below 0 LKFS. Valid values are 1 to 31. Loudness values outside this range cannot be expressed by dialnorm. The value of dialnorm is numerically equal to the absolute value of the Dialog Level. The value of 0 is reserved.

Dialog – The primary vocal element of a soundtrack.

Dialog Enhancement – The ability to vary the loudness of Dialog above background sounds.

Dynamic Range – The difference between the quietest and loudest sounds.

MUSHRA score – A measurement value resulting from a Multiple Stimuli with Hidden Reference and Anchor Audio quality test.

Music & Effects – Discrete elements of a soundtrack that when combined with Dialog, make up a Composite Mix.

Presentation – A set of one or more AC-4 Substreams intended to be decoded and presented to the user simultaneously. In the context of this document, the term Presentation and the term Audio Presentation as defined in Part 1 of A/342 [2] refer to identical concepts.

Representation – A collection and encapsulation of one or more media streams in a delivery format and associated with descriptive metadata.

Substream – A part of an AC-4 elementary stream that contains audio data and/or corresponding metadata. A Substream is referenced by one or more Presentations.

Video Description – Audio-narrated descriptions of a television program's key visual elements. These descriptions are inserted into natural pauses in the program's Dialog. See FCC Consumer Guide [6].

4. AC-4 SYSTEM FEATURE OVERVIEW

AC-4 is a highly efficient audio Codec that supports legacy channel-based content, object-based and channel-based immersive content, personalized content and advanced metadata.

4.1 AC-4 System Features

AC-4 provides all the features required in ATSC 3.0 for its next generation audio system, as well as certain unique capabilities and implementations.

4.1.1 A/V Sync and Frame Alignment

The A/V frame alignment feature of AC-4 avoids complex problems that can occur when trying to keep content in sync at segment boundaries, without compromising the audio at switching points. When enabled, this feature simplifies splicing workflows. It also simplifies transcoding from or to formats that use video-based data frame alignment, such as HD-SDI.

AC-4 audio frames may be coded at the same time interval as the associated video frame. With this frame alignment, audio can be passed transparently through a cable, satellite, or Internet Protocol Television (IPTV) turnaround facility, eliminating the need to decode and re-encode audio to correct time-base differences or to perform switching/splicing. All common integer and fractional video frame rates are supported. To select the correct frame rate and time-align the audio frame boundaries to video, the AC-4 encoder is provided with reference timing information. There is a direct whole-frame relationship between video and audio frames from a reference time, such as at the start of a coded video sequence or at a random-access point (RAP).

This A/V sync time alignment is a critical feature that facilitates accurate metadata delivery, switching and splicing throughout the content delivery workflow. The delivery of metadata is crucial for viewer's discovery and selection of many of AC-4's most advanced features. See Figure 4.1.

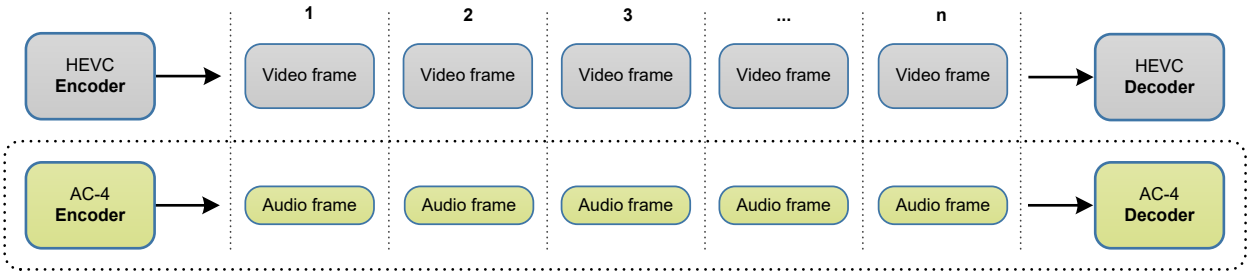

Figure 4.1 Timing relationship when using AC-4.

4.1.2 Dialog Enhancement

AC-4 provides user-controllable enhancement of Dialog intelligibility during decoding. The AC-4 system generates Dialog Enhancement information in the bit stream for all content, including content that has the Dialog pre-mixed with other elements. The ability for a receiver to control Dialog Enhancement can be configured at or signaled to the AC-4 encoder based on broadcaster or programmer guidelines. Receivers also need to be provisioned to access this feature. Recent consumer polling indicates that AC-4 Dialog Enhancement is one of the most desirable features of AC-4.

4.1.3 Extensible Metadata Delivery Format Support

The AC-4 bit-stream includes support for the carriage of Extensible Metadata Delivery Format (EMDF). EMDF provides a structured and extensible container for additional user data (for example, third-party metadata and third-party application data) to be carried in AC-4 bit-streams. This metadata is critical for accurate representations of the streams in packagers as well as viewer-facing content control and feature labeling in receivers.

4.1.4 Loudness and Dynamic Range Control

The AC-4 bit-stream carries an extensive set of loudness and Dynamic Range Control metadata as described below.

4.1.4.1 Loudness

The loudness metadata included in the AC-4 elementary stream allows for a full range of parameter description, including:

- True peak and maximum true peak
- Relative gated loudness values
- Speech gated loudness values
- Dialog gating type
- Momentary and maximum momentary loudness
- Short-term and maximum short-term loudness

4.1.4.2 Dynamic Range Control (DRC)

The Dynamic Range Control (DRC) elements of AC-4 provide a flexible set of options to address a wide variety of device profiles and user applications. This includes legacy AC-3 DRC profiles as well as custom DRC profiles that can be defined for an output mode (e.g., home theater, flat panel TV, portable speaker, and portable headphones). Control of the Dynamic Range can be exposed to the viewer by activation in the encoder and a user-facing control in the receiver.

4.1.5 Intelligent Loudness Management

In addition to the carriage of an enhanced set of loudness metadata beyond dialnorm, AC-4 incorporates a means to verify that the loudness information carried in the AC-4 bit-stream correctly describes the accompanying content. The system can use this means to signal to devices after decoding that the loudness metadata is accurate and no further loudness processing is needed. This behavior protects the audio from additional and unnecessary processing that can degrade audio quality. The AC-4 encoder incorporates a real-time loudness normalizer that can be dynamically enabled when the incoming loudness metadata cannot be validated.

4.1.6 Target Devices

AC-4 supports device-specific metadata to optimize rendering based on output-device characteristics. Target-device metadata can optionally enable conditional authoring and rendering based on output speaker configuration. This feature gives content creators artistic flexibility for creating an optimal sounding mix for all output speaker configurations without the compromises of downmixing.

4.1.7 Alternative Metadata

Alternative metadata supplements existing object metadata to allow different renditions of the same object to be created for each Presentation. Alternative metadata can also be defined for each target device.

4.1.8 Advanced Single-Stream and Multi-Stream (Hybrid) Presentations

AC-4 enables advanced single-stream Presentations by carrying multiple Audio Program Components in a single AC-4 bit-stream. This allows all Audio Program Components of a single Presentation, as well as components of multiple Presentations, to be carried within a single AC-4 bit-stream. Hybrid delivery uses transport of one Audio Program Component over a first path, such as a broadcast path, and one or more Audio Elements over a second path, such as broadband (Internet) or an alternate ATSC 3.0 physical layer pipe. AC-4 supports advanced multi-stream Presentations to enable hybrid-delivery use cases.

4.1.9 Core and Full Decoding

The AC-4 decoder has two decoding modes: core decoding and full decoding. Core decoding enables a low-complexity decode of a complete audio Presentation for devices with limited output capabilities (e.g., mobile devices, tablets, televisions, etc.). Full decoding enables a complete audio Presentation for devices with expanded output capabilities (e.g., Audio/Video Receiver). The choice of decoding mode enables a single bit stream to be compatible with a wide range of device types and applications.

4.1.10 High Sampling Frequencies

AC-4 supports high audio sampling frequencies of 96 kHz and 192 kHz. However, the ATSC 3.0 standard constrains the sampling frequency to 48 kHz. The AC-4 bit-stream is structured such that bit streams with high sampling frequencies can be decoded to PCM at 48 kHz without any penalties.

This feature minimizes the complexity of decoders that do not need to support high sampling frequencies.

4.1.11 Seamless Audio Switching

DASH allows transitions between Representations within the same Adaptation Set in order to optimize playback quality for changing network conditions. AC-4 enables seamless switching

between AC-4 streams of the same media content with the following types of configuration changes:

- Bit-rate changes
- Channel-mode changes
- Frame-rate changes where the higher frame rate is a factor of two or four times the lower frame rate (e.g., from 48 to 24 fps and vice versa)

4.2 High-Level AC-4 Structure

AC-4 supports the carriage of multiple audio Presentations in a single bit stream. Information for each Presentation includes instructions for selecting and mixing Audio Program Components. The AC-4 TOC contains a list of one or more Presentations that are carried in the stream. Presentations consist of Substream groups, where each Substream group has a specific role in the user experience: Music & Effects, Dialog, Associated Audio, etc. Substream groups can be shared between Presentations so that parts common to several Presentations do not need to be transmitted twice. The use of Agile Metadata to provide the ability for the viewer to select custom languages, Dialog and descriptions with minimal bandwidth impact, facilitates valuable features.

A Presentation informs the decoder which parts of an AC-4 stream are intended to be decoded simultaneously at a given point in time, describing available user experiences. Features such as loudness and Dialog Enhancement are managed by the Presentation.

Figure 4.2 shows an example AC-4 TOC with several Presentations for M&E with different language-Dialog Substreams. The selected Presentation contains the 5.1 M&E Substream, and an English Dialog Substream.

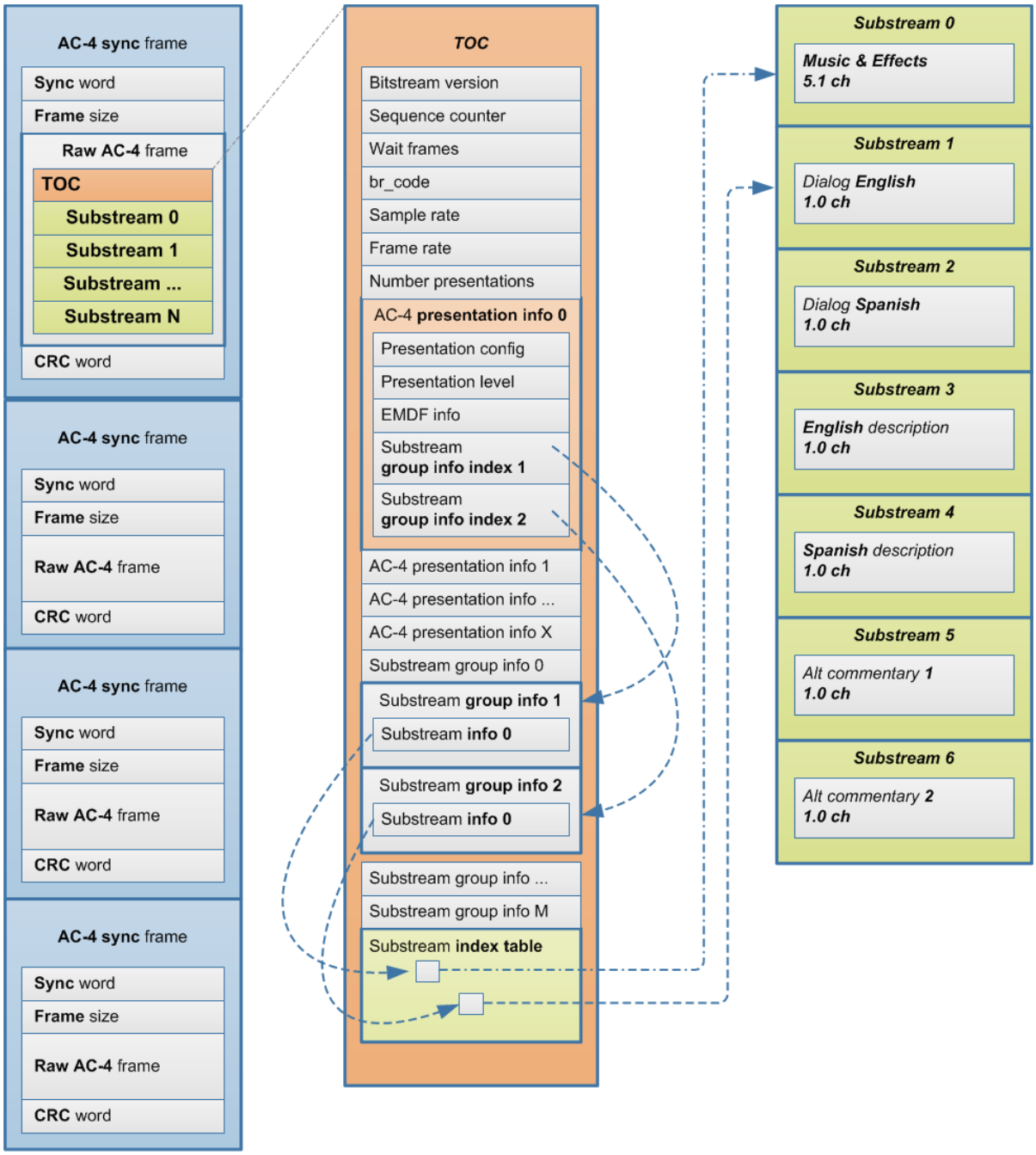

Figure 4.2 Example of complex frame with several Presentations and Substreams.

5. AC-4 OPERATING MODES

AC-4 and its predecessor AC-3 have several operating modes using separate and/or Complete Main components as shown in Figure 5.1, Figure 5.2, and Figure 5.3. There are considerable advantages in operating with Music & Effects separate from Dialog components (M&E+D). This mode, however, requires compatible content formatting and the availability and delivery of the metadata (as found in the AC-4 TOC) throughout the workflow that is needed by the AC-4 encoder and the receiver's decoder, to identify each component.

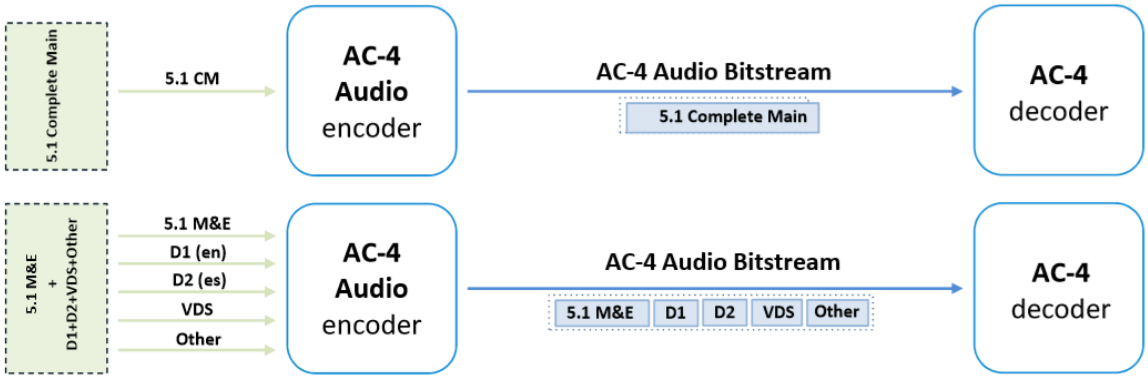

Figure 5.1 AC-4 operating modes.

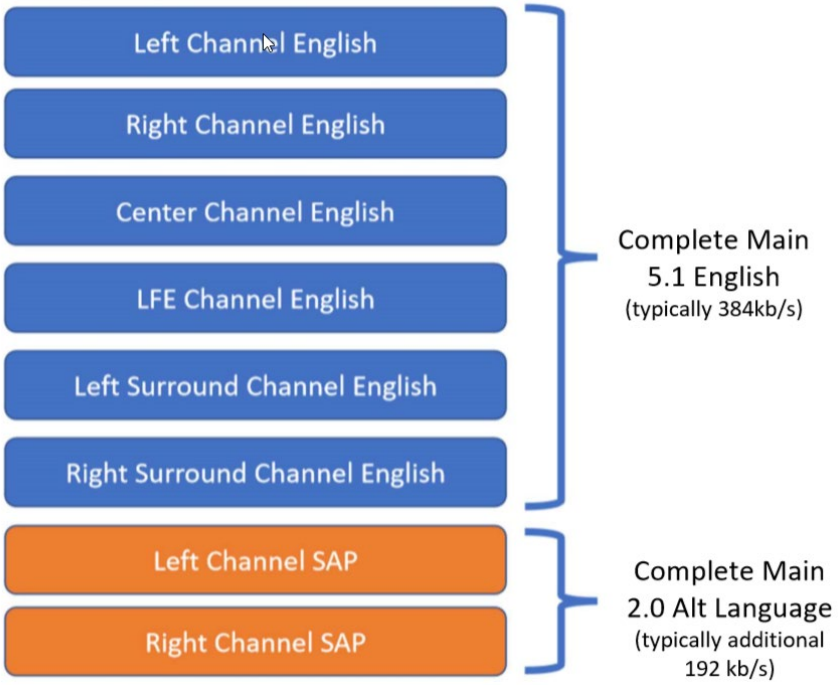

Figure 5.2 Current AC-3 Complete Main without Agile Metadata.

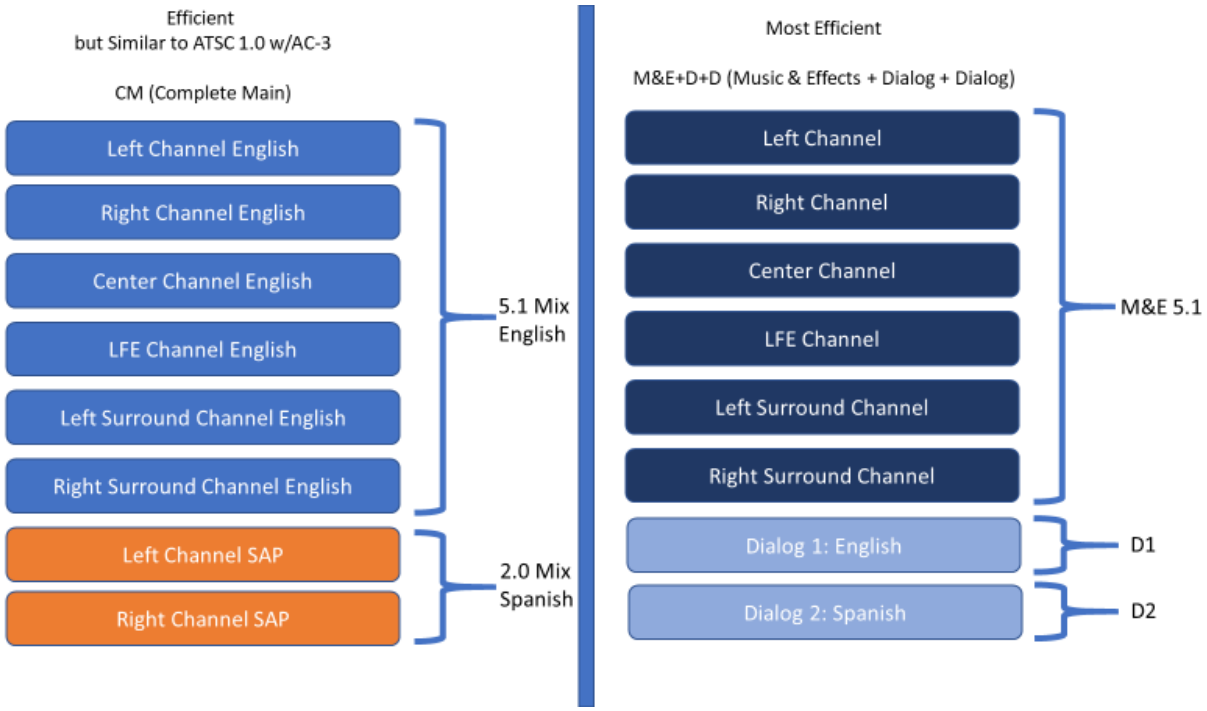

Figure 5.3 AC-4 options with and without metadata.

6. METADATA CHALLENGES

Practices for consistent delivery of agile AC-4 metadata (as found in the AC-4 TOC) from content creator through to the AC-4 encoder and ultimately on to the receiver still require development and standardization.

The practice of using static metadata for AC-3 encoding enables terrestrial television to avoid distribution complexities while improving reliability. However, advanced AC-4 features like the selection of alternative program components or advanced Dialog Enhancement require the transport of metadata throughout the workflow from content creator to the viewer's receiver. This increases infrastructure complexity. See Figure 6.1.

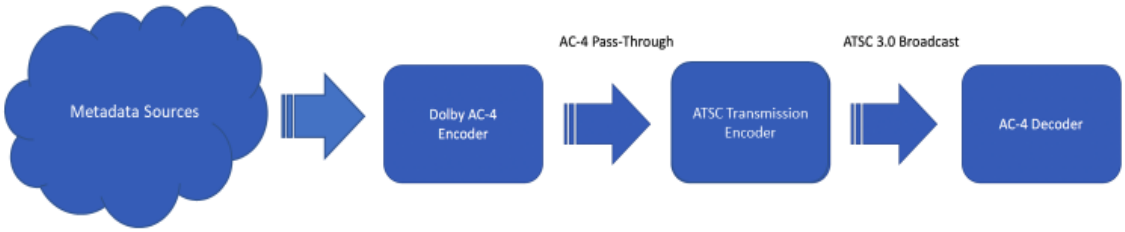

Figure 6.1 Metadata challenge.

Receivers will build Presentation streams based on the contents of the MPD (Media Presentation Description) in DASH and Audio Stream Properties Descriptor in MMT (See A/331

Section 7.2.3.4 [5]) through an interim parser of the AC-4 DSI (Decoder Specific Information) from the AC-4 TOC in the broadcaster's encoder. The AC-4 TOC is also used by the decoder in the receiver to acquire Presentation options and control, but does not provide viewer stream selection options.

Agile Metadata is required for any format other than Complete Main. It is also needed for the receiver to correctly identify, decode and annotate elements, giving the viewer access to personalized and immersive audio features. Manual or fixed entry of metadata is possible in the encoder's packager, but does not provide a practical solution for a wide variety of content.

AC-3 also has Agile Metadata capability, but the use of fixed loudness and dialnorm with a primary Complete Main and a secondary Complete Main (SAP: second language or VDS) made the use of simpler, static metadata delivery a regular practice.

To enable AC-4's most beneficial features, the system requires Agile Metadata and Mix & Effects + Dialog + Dialog formatting that:

- Offers the most efficient use of alternative languages
- Offers multiple alternative tracks such as language and Video Description
- Enables features such as home and away commentary
- Enriches the viewer experience with potentially simple, custom controls

7. AC-4 MODES OF DIALOG ENHANCEMENT

There are multiple modes of Dialog Enhancement (DE) that can be used and delivered to the viewer:

- Single-ended/Simple/Unguided Mode (a.k.a. "parametric mode") – In this mode the encoder automatically generates DE metadata parameters from a Complete Main program (works with all content, and is always available). The broadcaster can set a range of enhancement that the receiver can present to the user.
- Guided Mode (a.k.a. "guided parametric mode" or "hybrid mode") – In this mode the encoder automatically generates DE metadata parameters from a Complete Main program and a Clean Production-provided Dialog track (desirable but requires content with discrete Dialog and upstream and downstream workflow changes).
- M&E+D Mode – This mode is enabled by carriage of independent M&E and Dialog Substreams that are combined in the decoder (facilitated in post-production – desirable but requires content with discrete Dialog and upstream and downstream workflow changes).
- Fully single-ended – Receiver only, not provisioned by the encoder (limited capability and range). This is a feature provided by the TV set or STB.

The AC-4 decoder always makes the best available Dialog Enhancement mode accessible. However, full functionality requires the receiver to expose the feature to the viewer to activate and then control the Dialog Enhancement function, perhaps with a slider-like control, capable of setting low, medium, or high enhancement. Note that Dialog Enhancement is established by using a combination of increasing the gain of Dialog and reducing background sounds, thereby maintaining the intended content loudness.

8. HYBRID TRACK DELIVERY

Delivery of alternative elements or components via broadband is theoretically possible with AC-4, but broadband delivery latencies often experienced as of publication of this document can require considerable video delay compensation at the receiver. Because the additional bandwidth

required for OTA delivery of an alternative Complete Main program is relatively small, this approach provides an alternative delivery means for hybrid delivery until broadband latency is generally reduced. Hybrid delivery can also pose complex content licensing issues that will require resolution.

9. TERMINOLOGY

Current standards terminology used to describe new AC-4 features (Dialog Enhancement, Controllable Dynamic Range, Alternate Languages, Video Description or Alternate Dialog Commentary) can be difficult to understand and confusing to consumers/viewers. The industry needs to work together to create commercially viable consumer-facing terminology that is both descriptive and self-evident as to its meaning. New, common and easily-understood terms used in all consumer-facing devices would help ensure proper identification and use of these powerful features by viewers.

10. AC-4 ENCODING DATA RATES

Broadcasters have a wide choice of AC-4 Complete Main encoding data rates, as indicated by the relatively small differences shown in Figure 10.1 among MUSHRA scores for 5.1 Complete Main encoding with AC-4 at 96, 144, and 192 kb/sec. The figure indicates that performance for 5.1 at 192 kb/sec using AC-4 is like that of AC-3 at 384 kb/sec. This improved efficiency makes multiple Complete Mains a good choice for the ATSC 1.0-to-ATSC 3.0 transition due to the minimal bandwidth penalty incurred using AC-4.

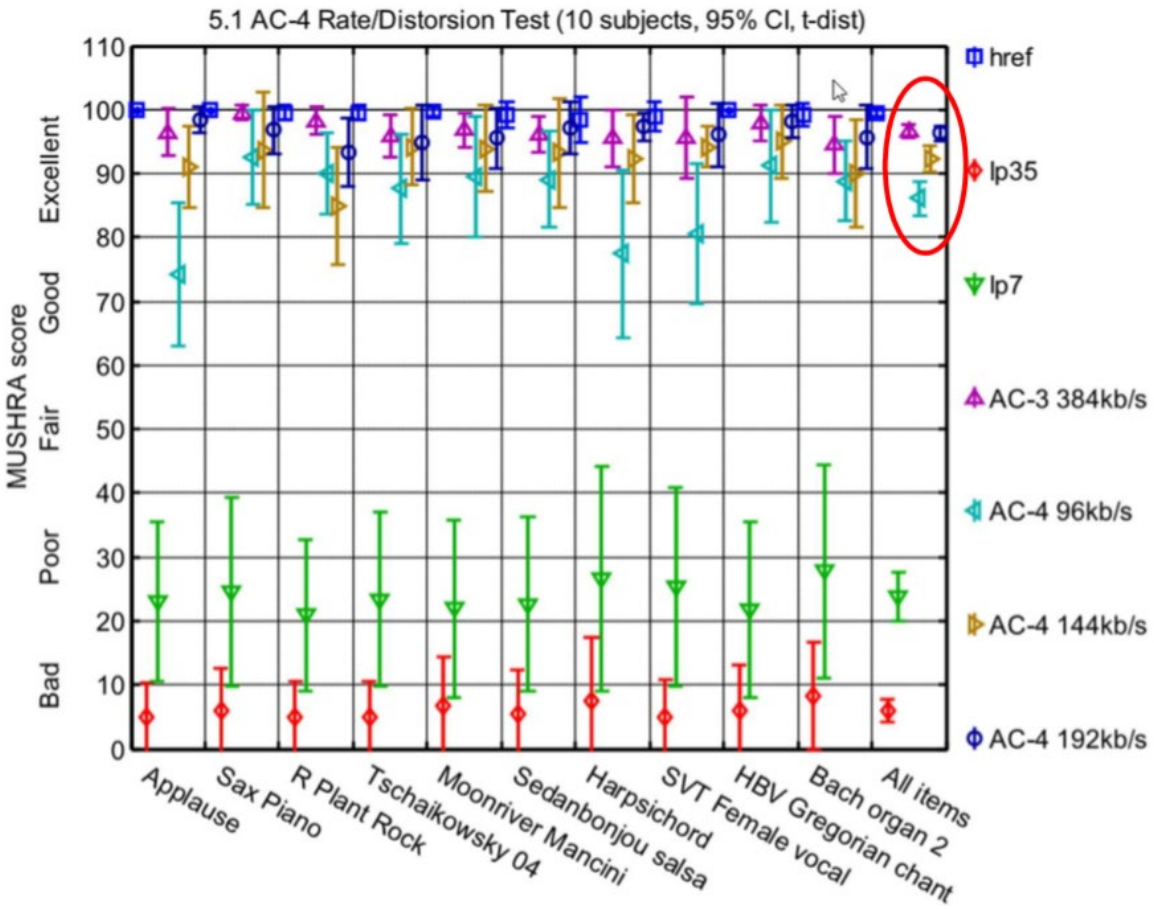

Figure 10.1 AC-4 versus AC-3 encoding data-rate performance.

11. CONVENTION FOR PCM – SDI WORKFLOW

Conversion from AC-3 to AC-4 requires full decoding of AC-3 to baseband. PCM or SDI streams can transport compatible metadata for both AC-3 and AC-4. An upstream source of metadata can provide valuable information such as guided Dialog Enhancement data, track labeling, and object placement. Alternate Complete Mains can use duplicates of the convention shown in Figure 11.1. (A 5.1 channel format is shown in the figure, but 5.1.2 and 5.1.4 can also use this paradigm.)

Figure 11.1 SDI - PCM formatting with metadata.

12. FINDINGS AND RECOMMENDATIONS

- Multiple 5.1 or stereo Complete Main AC-4 encoding (with option for languages and Video Description) will be effective for the initial ATSC 3.0 transition phase (ATSC 1.0's AC-3 to ATSC 3.0's AC-4), facilitating a smooth broadcaster rollout and providing adequate viewer optionality and reasonable bandwidth efficiency.
- Dynamic Range Control will be effective only when enabled by the broadcaster and presented to the viewer with simple and understandable receiver controls and terminology.
- Dialog Enhancement needs to be enabled by the broadcaster using Single-ended/Simple/Unguided Mode with ranges, and presented to the viewer with simple and easily-understood receiver controls and terminology. For a consumer to be able to use Dialog Enhancement, a receiver needs to access this feature at the AC-4 decoder and make control available.
- AC-4 Complete Main encoder data rates of 96 kb/sec to 192 kb/sec provide excellent results, but planning is required for overall bandwidth allocation.
- Metadata workflows need to be discussed with the intent to create guidelines to enable improved Dialog Enhancement and other features, and to provide consistent track announcement.
- Hybrid OTA/OTT track delivery is not considered viable as of the publication of this document due to potential latencies that can be encountered by consumers. Complete

alternate program encoding is an alternative, considering that the additional bandwidth required for its OTA delivery is not onerous.

- The system can operate with multiple Complete Mains without saturating receiver decoder capability (e.g., alternative languages and Video Description), but this practice minimally increases transmission bandwidth overhead.
- Metadata parameters can be manually entered into the audio encoder for simple 5.1 Complete Main or stereo operation. In the long term this does not provide a sustainable solution that offers the greatest number of viewer features and capabilities.
- Custom track naming and viewer selection options become difficult with manual metadata parameter entry. Many features will not be available without the use of Agile Metadata.
- The AC-4 audio decoder makes all AC-4 functionality possible. Some receiver implementations could limit a viewer's access to certain AC-4 features. Receiver manufacturers are encouraged to provide access to all AC-4 features in all ATSC 3.0 products.
- Consistent, simple, and easily understood terms are necessary for easy viewer discovery of audio capabilities in TVs, AVRs, and set tops.
- ATSC members should work with MVPDs and cable industry representatives to develop best practices for passing AC-4's functions and capabilities to downstream systems.
- Broadcasters should remain focused on fulfilling viewer preferences while managing risk to content delivery.
- ATSC should liaise with CTA and SMPTE on the consumer decoder and facility interface topics, respectively.
- ATSC should author a Recommended Practice to provide background and more detailed and specific recommendations based on the findings in this report.

End of Document