SCTE•ISBE CABLE-TEC EXPO® VIRTUAL EXPERIENCE » OCTOBER 12-15 2020

2020 Fall Technical Forum SCTE·ISBE • NCTA • CABLELABS®

The Future of Cable Television Audio is Accessible

A Technical Paper prepared for SCTE•ISBE by

Mark Francisco
Fellow
Comcast Cable
1800 Arch Street Philadelphia, PA 19103
(215)286-8959
Mark_francisco@comcast.com

Table of Contents

List of Figures

List of Tables

1. Introduction

Television sound hasn't changed much since its first appeared in the U.S. in 1941. In the 1940s, television audio was broadcast as a frequency modulated (FM) subcarrier to the video program. A stereo subcarrier was added in the 1970s. A secondary audio channel was introduced in the 1980s, primarily for dominant non-native languages such as Spanish. Digital audio introduced surround sound over cable and satellite services in the late 1990s, and over-the-air transmission with the conversion to Advanced Television Systems Committee (ATSC) broadcasting in the 2000 aughts. Twenty years later, despite the explosion of Internet Protocol (IP) and Internet delivery, television audio has made little progress, with the exception of the broadcast of Descriptive Audio or Video Description (formally named as Audio Descriptions) and the expansion of two dimensional audio to three dimensions with object-based audio, the latter of which is only available using IP-delivery and Blu-Ray. The future of broadcast television has been standardized internationally. with the recent release of ATSC 3.0. Its advances include video and audio formats by moving to IP encapsulation, adding high-dynamic range and UltraHD (4K) video, and advancing audio to a novel audio ecosystem, AC-4 [1], supporting greater efficiency, fidelity and personalization compared to current audio delivery.

Since the advent of "talkies," in the late 1800s, audio has always been an appreciated accompaniment to video. There is a long-standing truism that "sound is more than half the picture". Despite its ability to convey a story comprehensively (as in radio), however, television sound is seldom delivered on its own.

For 12 million people in the US [2], sound conveys the entire television experience. Sight impairment and blindness affects a large and growing segment of the population. Congenital conditions, disease, injuries and age fuel the reliance on audio over visual inputs, and these causes are predicted to grow over time. Conversely, the same is true of those who rely partly or completely on vision over hearing. The need for information, communications and entertainment in video- or sound-only format is increasing.

The desire to advance single sensory experiences is not just a regulatory or moral imperative -- it is good business. Inclusive products yield business opportunities for disabled and abled people alike. Disability is not necessarily a permanent condition -- it is defined by the World Health Organization [3] as a mismatch between the individual's ability at the moment and the environment with which they are in. Therefore, we may all experience disabilities (and need for accommodation) at some point in our lives.

While much has been accomplished in the pursuit of equivalent experiences for people of all abilities, many opportunities exist. Cable television is a sexagenarian industry with incredible establishment, reach and constraints preventing evolution. This fact can only serve to increase resolve, as disruptive change is possible and can offer large advances in value, equitability and experience.

2. What we must do – Regulations for MVPDs and Broadcasters

The Twenty-First Century Communications and Video Accessibility Act of 2010 [4] resulted in the creation of a set of Federal Communications Commission (FCC) regulations. Enacted in 2011, these regulations require commercial television stations that are affiliated with one of the top four commercial television broadcast networks (ABC, CBS, Fox, and NBC) and are located in the top 60 television markets to provide 50 hours of video-described programming per calendar quarter during prime time or on children's programming, as well as an additional 37.5 hours of video-described programming per calendar quarter at any time between 6 a.m. and midnight. In addition, multichannel video programming distributor (MVPD) systems that serve 50,000 or more subscribers must provide 50 hours of video description per calendar quarter during prime time or on children's programming, as well as an additional 37.5 hours of video description per calendar quarter at any time between 6 a.m. and midnight, for each of

the top five national nonbroadcast networks that they carry on those systems. The top five nonbroadcast networks currently subject to the video description requirements are USA Network, HGTV, TBS, Discovery, and History. The FCC published a Notice of Proposed Rulemaking (NPRM) on April 22, 2020 proposing an expansion from the top 60 markets by 10 annually for four years beginning in 2021. [5]

The NPRM also recommends changing the term for video-described content to "audio description" from the current convention of "video description". Additionally, the apparatus must provide a simple and easy-to-use mechanism for activating the secondary audio stream for audible emergency information.

Manufacturers must provide access to video description through a mechanism that is reasonably comparable to a button, key, or icon.

Digital apparatus must be designed so that control of appropriate built-in functions included in the digital apparatus (i.e., user interfaces) are accessible to and usable by individuals who are blind or visually impaired, if achievable. Specific requirements exist with respect to on-screen text menus and other built-in digital apparatus functions.

Manufacturers of digital apparatus must ensure that information and documentation it provides to its customers is accessible, if possible (e.g., user guides, bills, etc.). Digital apparatus manufacturers also must comply with requirements relating to the manner in which they notify consumers that digital apparatus with the required accessibility features are available.

3. Conventions in Cable Television Audio

Cable television is accessible. As a primary and often lifeline service to its customers, cable, satellite and fiber providers of multi-channel television services are responsive to regulation, social need and the business opportunity that serving the entire market can bring. Being accessible can often result in unanticipated benefits in what is not traditionally considered disability usage. An example is the use of closed captioning in health clubs and bar/restaurants to benefit patrons who are unable to hear due to the high ambient noise level. Cable television companies are subject to regulation by the FCC, state regulators and franchisers and must adhere to content carriage contracts. The result is high availability of accessible content and services. Many do not stop at the minimum bar of regulation, and add additional features that benefit people with disabilities beyond the law. Examples include the Comcast Accessible Remote and the Altice One Voice Remote, among others. The benefits are significant to the populations served, be they with disabilities or without, and to brand identity.

3.1. Retransmission

Either by convention or contract, broadcast over-the-air networks and cable channels are typically retransmitted in their as-received state, that is MPEG-2 Transport Stream with MPEG-2 video compression and AC-3 audio compression. Approximately 2/3 of cable television stations are received in 1080i (1920 x 1080 interlaced, 30 fps) video, 5.1 channel primary audio with AC-3 compression, monaural secondary (SAP) audio with AC-3 compression and CTA-608 closed captions wrapped in CTA-708 transport. The majority of the remaining 1/3 of stations are 720p (1280x720 progressive, 60 fps) video, 5.1 channel primary audio with AC-3 compression, monaural secondary audio program (SAP) with AC-3 compression and CTA-608 closed captions wrapped in CTA-708 transport. A small number of stations are received as standard definition 480i (640 x 480 interlaced, 30 fps) video, 2.0 channel primary audio with AC-3 compression, monaural secondary (SAP) audio with AC-3 compression and CTA-608 closed captions wrapped in CTA-708 transport. This is almost identical to the ATSC over-the-air

broadcast format, with the exception of the vestigial sideband (VSB) modulation required for radio transmission. Increasingly, MVPDs are converting the received video compression to a more efficient H.264 while retaining the audio bitstreams unchanged. Typical as-delivered cable channels require 8-12 Mbps for MPEG-2 video (less for standard definition), 3-6 Mbps for H.264 video and 384-512 kbps for two audio services.

Due to constraints in receivers and the inability for broadcasters to dynamically change descriptors, the primary and secondary audio streams, when both are present, have differing but no more than two language codes. This results in the typical practices of signaling the primary language as English (ISO_639_Language code ENG) and secondary audio program language of Spanish (SPA), Portuguese (POR), French (FRA) or Middle English (ENM). The secondary language codes are often not representative of the audio program's actual language, potentially resulting in viewer confusion. Furthermore, broadcasters that deliver audio descriptions (AD) can change the service provided on the secondary audio program between a secondary language such as Spanish and a complete mix of English program with English Audio Descriptions but retain a consistent but incorrect SAP language code of Spanish.

3.2. Conventions in Streaming Media Provider Audio

Streaming media providers are content creator, aggregator, and hosting entities that rely on user-supplied internet network connections. They deliver content using Hypertext Transfer Protocols (HTTP) protocols such as HTTP Live Streaming (HLS) or MPEG-DASH (dynamic adaptive streaming over HTTP). These relatively modern packaging methods are more flexible in terms of variety of content presentations, and this can include multiple audio and video formats. These providers deliver almost exclusively video on demand, rather than broadcast formats, so they use file-based rather than live workflows. This agility and lack of linear complexity results in a much greater offering of audio selections on a per-asset basis. A recent survey of popular streaming media providers discovered thousands of assets containing audio descriptions as an audio option. Many are delivered as a complete mix in surround sound format, rather than monaural. The most advanced offerings are found on Apple TV+ which includes several languages of audio descriptions in object-based audio format (Dolby ATMOS). Table 1 lists the number of titles available with audio descriptions from popular streaming media providers. [6]

Table 1 – Streaming Offerings with Audio Descriptions

Provider	Count (with AD)
Apple TV+	32
iTunes	1341
Disney+	589
Hulu	91
Netflix	1271
Amazon Prime	1831

3.3. Conventions in Cable Television Broadcast Programming

A check of a single day's (August 5, 2020) programming listed 109 shows with audio descriptions. (Table 2)

Table 2 - Broadcast Offerings with Audio Descriptions

Broadcast Network	with AD
ABC	28
CBS	13
Discovery	25
Fox	25
HGTV	35
History	28
NBC	18
Oxygen	4*
SyFy	2*
TBS	19
Telemundo	7*
CW	5*
TNT	12*
USA	20

* denotes networks without AD mandates

These numbers are substantial, and they are generally increasing. The challenge remains in finding specific programming with AD, and then selecting the proper audio service when desired, and deselecting it when not. Most households with visually impaired residents are shared with household members who lack vision challenges, and these individuals may not want audio descriptions present while viewing programming.

Broadcast television programming audio services are ambiguous for a number of reasons. Due to constraints discussed in section 3.1, cable channels are limited to two audio services and the two must be identified with differing languages. Audio descriptions are conventionally placed in the SAP, often displacing a second language such as Spanish. Broadcasters do not dynamically change language codes, which are part of the MPEG-2 transport stream that is created by the broadcast contribution encoder. Therefore, if a broadcaster occasionally transmits Spanish as SAP and occasionally transmits AD as SAP, the broadcast language (ISO_639_Language) is always set to the SPA denoting Spanish. In the U.S. a workaround is often used that signals AD as Portuguese (ISO 639 language code POR), and in Canada, Middle English (ENM) and Middle French (FRM) is used. If a cable service-provided user interface allows a user to select from the languages signaled by the network broadcaster, audio descriptions are almost always incorrectly identified. The network also supplies metadata to populate program guides, and these often have correct information about audio services, but only for time of broadcast. Shows are increasingly viewed in a time-shifted fashion through digital video recording, start-over and on demand methods, and these may differ in provided audio services. Programs that are broadcast with audio descriptions as SAP are not available with SAP when viewed on demand. MVPDs have increasingly been improving their user experiences, using logic to assist with selecting audio descriptions. Oftentimes, program metadata is considered first, followed by selecting SAP when the program data is not affirmative as the existence of AD.

The ambiguities mentioned above do not apply to streaming media providers. There, the content hosting and user experience is tightly bound, and content is accurately and affirmatively identified in the file-based metadata and presentation information.

4. Where Do We Go From Here

Technology is quick to change, standards are very slow to evolve, and consumer behaviors range the entirety of the space between. The technologies to improve the user experience around accessibility of content are with us already, and some of them have been defined in standards for some time. Let's look at what can be done now and what we can anticipate in the near future.

4.1. The United Nations Already Solved This Problem

The International Standards Organization (ISO) of the United Nations includes the Moving Picture Experts Group (MPEG). The system standard for MPEG, including the basis for ATSC and Digital cable transport streams (ISO13818-1), was released in the late 1990s [7]. Version 2, released in late 2000, contains a second field to the ISO 639 [8] language descriptor entitled **audio_type**. The values for audio type are located in Table 3.

Table 3 – MPEG-2 Transport Stream Audio Type Values

Value	Description
0x00	Undefined
0x01	Clean effects
0x02	Hearing impaired
0x03	Visual impaired commentary
0x04-0xFF	Reserved

Inserting the code 0x03 in the MPEG-2 headers will result in unambiguous signaling of the AD content. While conceptually simple, this method is challenging for broadcast workflows, as encoders do not dynamically change headers. Content selection is also challenging, as the value is embedded in the audio stream, requiring decoding of the asset to examine. Nonetheless, this method is of interest and being pursued.

Fast forward to modern Internet streaming, which is also standardized by MPEG through MPEG-DASH, which includes a similar audio type descriptor entitled role. Other streaming standards exist, such as the Internet Engineering Task Force (IETF) standardized HLS format, a segmented transport stream format that mirrors the descriptors identified in the MPEG system specification ISO13818-1.

4.2. Accessibility is About More Than Just Content Signaling

Disambiguation of accessible content audio will be a great benefit to viewers with interest and disinterest in audio descriptions. Nothing is more impactful to the interested parties than more quantity and quality of content. The tallies of content shown in Table 1and Table 2 include children's shows, movies, and episodic series. However, live content, including news, sports and events, such as awards shows and concerts, is absent. This is partially due to the lack of regulation around live or the explicit exemption of live programs and programmers, and the technical challenges of creating descriptions in a live workflow. Describing live events is possible using a commentator calling a live baseball game. Several events have been live described, including parts of the 2018 Winter Olympics in PyeongChang, and the GlobalCitizen One World Together at Home Concert in April of 2020. These events were challenging only due to their non-standard inclusion of audio descriptions. Several professional agencies, including the Media Access Group at WGBH, have demonstrated proficiency in live descriptions. An important aspect of the creation is appropriately "ducking" the program audio during narration. A depiction of ducking is shown in Figure 1.

Lower waveform - recording audio description
Upper waveform is base music and effects
Dotted white line in upper waveform is result of ducking

Figure 1. Audio Ducking Example

Narrators are proficient with the timing of their descriptions, and automatic or manual equipment is used to ramp down the main program audio and ramp it up prior to and after the audio description segment. To date, live descriptions have resulted in time-delayed versions of the events, as the live descriptions were conformed to the event after completion. What will it take to produce described live events and deliver live?

As you will learn in the next section, new technologies are on the horizon to make live event descriptions possible.

4.3. The Future of Over-the-Air Broadcasting

Today's over-the air-broadcast format is three decades old. Three decades ago, the Internet was accessed through dial-up modems, smartphones were ten years away, and compact discs were in their prime. The Americans with Disabilities Act was brand new at the time, and the 21st Century Communications and Video Accessibility Act was still twenty years away. So much has changed in thirty years. In 2019, the NextGen TV brand was created, which embodies the new ATSC 3.0 broadcasting standard, which was authorized by the FCC in 2017. ATSC 3.0's key features include higher resolution, dynamic range, IP delivery, interactivity and greater audio fidelity. The very modern HEVC video compression and a pair of audio compression standards, AC-4 and MPEG-H, have been selected. Selection of the national audio format is per country. The U.S. has selected Dolby's AC-4 compression as standard for NextGen TV. Two generations ahead of current over-the-air television audio compression, AC-4 is approximately three times more efficient, from a pure compression perspective, and when combined with its novel capabilities to deliver audio in substreams and objects, the bandwidth efficiencies are much higher. In the next section we will discuss some of these new audio features. [9]

4.4. M&E+D

While sounding like a math formula or SMS shortcut, M&E+D refers to a common Music and Effects mix and additive Dialog track. This is a key enabler for next generation accessible audio experiences. Allowing a single common Music and Effects mix to be delivered with synchronous but separate (and multiple) additive Dialog streams, M&E+D offers personalization, bandwidth savings and reduced production complexity. No longer is a complete main -- a mix of all music, effect and dialog -- required

for each audio service. The viewer/listener may select from multiple presentations, mixes of common plus one or more dialog tracks. Furthermore, these mixes can be adjusted to the listener's preference, allowing dialog to be enhanced or diminished quite easily. In the use case of audio descriptions, the narrative can be created and delivered without requiring a mix-down to the primary audio. This reduces time to deliver and offers the opportunity to lay narrative on top of a surround base. Spatializing audio is of great benefit to the vision impaired community, as it provides direction and space to the experience. Today, almost all broadcast with descriptions are delivered as monaural audio. Figure 2 illustrates how multiple audio presentations can be assembled through an AC-4 bitstream.

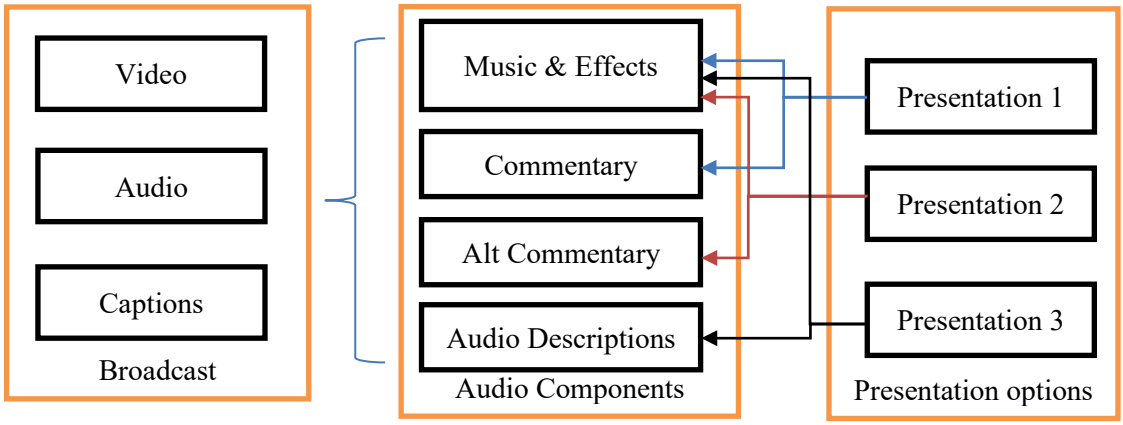

Figure 2. Multiple Audio Presentations Through AC-4 Substreams

AC-4 audio can be frame-aligned with video frames, unlike preceding audio formats. This offers seamless switching at content insertion (i.e. advertising) boundaries. Seamless switching can ease the addition of accessible advertising, ads with audio descriptions, a benefit to the viewer and provider alike.

4.5. Why is this important

To understand the impact of AD, a survey of 626 visually impaired adults was conducted in 2017 by the American Foundation for the Blind (AFB). It found that over half reported watching television 4 or more hours a day -- and 65% reporting problems looking up what is on TV. Less than half were aware of assistive technologies such as AD.

Comments filed with the FCC by the AFB in support of expanding coverage of audio descriptions included the following from Jerrell Harris of Paragould, AR, which really sums up the impact of AD:

> "I do not currently receive audio description from ANY of my current local channels or cable channels. I do get some audio description on NETFLIX and OMG, I ABSOLUTELY LOVE IT!!!! It changes the way I watch TV and now I feel like I am included and can SEE what's going on, on the screen. There is SO much that we visually [interpret] on the TV screen and audio description is AMAZING and describes in detail what is on the screen and what is going on. As a blind person, it is frustrating to me to watch TV without audio description because we miss so much of what is going on. My husband will tell me what is happening on a show, but I know that gets annoying to him at times. The audio description lets us enjoy TV again, and gives a sense of confidence that we are almost on the same playing field as the sighted now because we actually have the description of what is going on the screen."

This quotation speaks to the expansion of availability and ease of selection of audio descriptions in broadcast and cable television.

4.6. Automating Audio Description

Machine-based transcriptions of program audio to create closed captions is commonplace and steadily improving as models evolve and model training continues. Is it possible to automate the creation of audio descriptions? Image classification is mature and implemented in accessible applications, such as Microsoft's Seeing AI, benefitting the visually impaired as they navigate unfamiliar places and things such as food packaging. The Massachusetts Institute of Technology recently completed research analyzing television images to determine the point of focus where the actor's attention is placed [10]. Constantly describing each scene would be distracting, which makes the remaining challenge to describe the instrumental parts of a scene. It seems within reach to train domain-specific programs, such as sports or news, to machine-describe the segments and plays. Consider the ease with which a sports radio commentator describes the play of game.

5. Beyond Audio Description

Audio is instrumental in creating an accessible television experience beyond audio description. The FCC requires that all features of apparatus that consumers use to display television programming are accessible. Voice guidance is increasingly the means of complying with this requirement. Many features of televisions appear on video menus, which can challenge those with visual impairments. Voicing out the elements of a screen menu as they are navigated is a standard feature of screen readers applications on PCs and voice out features of smart phones. Smart televisions and cable receivers are increasingly implementing voice guidance to voice out menus and program guides. These applications typically rely on an active Internet connection to the device, which is a challenge as many smart televisions remain unconnected to networks. Embedded text-to-speech technology has evolved in performance and compactness and will increase in availability in consumer and subscriber devices.

Emergency information is new to ATSC 3.0. This information differs from Emergency Alerts, which may take over the audio and or video programming during emergency events. Emergency information may include warnings, such as weather and school closings, but doesn't automatically preempt programming. Through the use of text-to-speech technology and signaling in the AC-4 audio stream as type E, these textual alerts can be converted to audio and added as a presentation option for viewers who prefer audio narration of emergency information.

6. Conclusion

Cable television is accessible. That said, Internet streaming media providers offer more content and better fidelity of accessible audio than over-the-air networks and cable channels. Next generation broadcasting and pending regulation will offer expanded coverage of accessible audio on cable television programs and provide a path to more innovation in terms of immersive audio experiences, multi-language, multi-cultural accessible audio experiences and description availability on live and event-based programming. Improved signaling of audio content can improve the ability to find accessible audio. ATSC3.0 includes a new audio subsystem, AC-4, that offers presentation options that can improve personalization and lead to more descriptions on live programming.

Abbreviations

AC-4	audio compression 4th generation (Dolby)
AD	audio description

AFB	American Foundation for the Blind
ATSC	Advanced Television Systems Committee
CTA	Consumer Technology Association
DASH	Dynamic Adaptive Streaming over HTTP
FCC	Federal Communications Commission
FM	frequency modulation
HEVC	high efficiency video coding
HTTP	Hypertext Transfer Protocol
IETF	Internet Engineering Task Force
IP	Internet protocol
ISO	International Standards Organization
M&E	music and effects
MPEG	Moving Pictures Experts Group
MVPD	multichannel video programming distributor
NPRM	Notice of Proposed Rulemaking
SAP	secondary audio programming
VSB	vestigial sideband
WGBH	A public media producer based in Boston, MA

Bibliography & References

[1] ATSC Standard: A/342 Part 2, AC-4 System

[2] Vision Health Initiative – Centers for Disease Control and Prevention - https://www.cdc.gov/visionhealth/basics/ced/fastfacts.htm

[3] Disability – World Health Organization - https://www.who.int/health-topics/disability

[4] C.F.R. Title 47 Chapter 1 Subchapter C Part 79 – Electronic Code of Federal Regulations – Accessibility of Video Programming

[5] Notice of Proposed Rule Making Comments on Expanding Video Description Requirements to Increase Programming Accessibility to Blind and Visually Impaired Americans – FCC Filing MB 11-43 and comment filing 106223000511985

[6] The Audio Description Project – American Council for the Blind - https://acb.org/adp/appletvad.html

[7] ISO/IEC 13818-1 – Information technology – Generic coding of moving pictures and associated audio information: Systems

[8] ISO/IEC 639-1 – Codes for the representation of languages

[9] ATSC 3.0 Dolby Audio Handbook – February 2020

[10] MIT Gaze 360 Project - http://gaze360.csail.mit.edu

