SCTE CABLE-TEC EXPO FAST FORWARD 2021
VIRTUAL EXPERIENCE
OCTOBER 11-14

2021 Fall Technical Forum
SCTE • NCTA • CABLELABS

Enabling Automation for Mapping Linear Channel Feeds and VOD Files into DASH Structures

A Technical Paper prepared for SCTE by

Yasser F. Syed, PhD.
Comcast Distinguished Architect
Comcast TPX Viper Architecture
1701 JFK Boulevard Philadelphia, PA 19103
215-286-1700
yasser_syed@comcast.com

Alex Giladi
Comcast Fellow
Comcast TPX Viper Architecture
1899 Wynkoop St., Denver CO 80206
215-286-1700
alex_giladi@comcast.com

Ali C. Begen, PhD , Consultant, Comcast

Table of Contents

Title **Page Number**

List of Figures

Title **Page Number**

List of Tables

Title **Page Number**

1. Introduction

Established MPEG-2 TS based video workflows have been used for distributing content in traditional consumer delivery ecosystems such as QAM-based cable, satellite, ABR Streaming, as well as over-the-air broadcast. With the advent of adaptive streaming technologies such as MPEG DASH and Apple® HLS, these same content video workflows have been adapted to these new types of video distribution. With that said, manual workarounds are often needed to fit in the new features enabled by adaptive streaming. Thus, the ABR distribution workflows may still depend on MPEG-2 TS source content delivery systems for several years due to existing equipment and network tools, and the need to provide traditional QAM linear services [4]. It is expected that backend systems of MVPDs will change as third-party content contribution and demultiplexed media component delivery becomes more widely used.

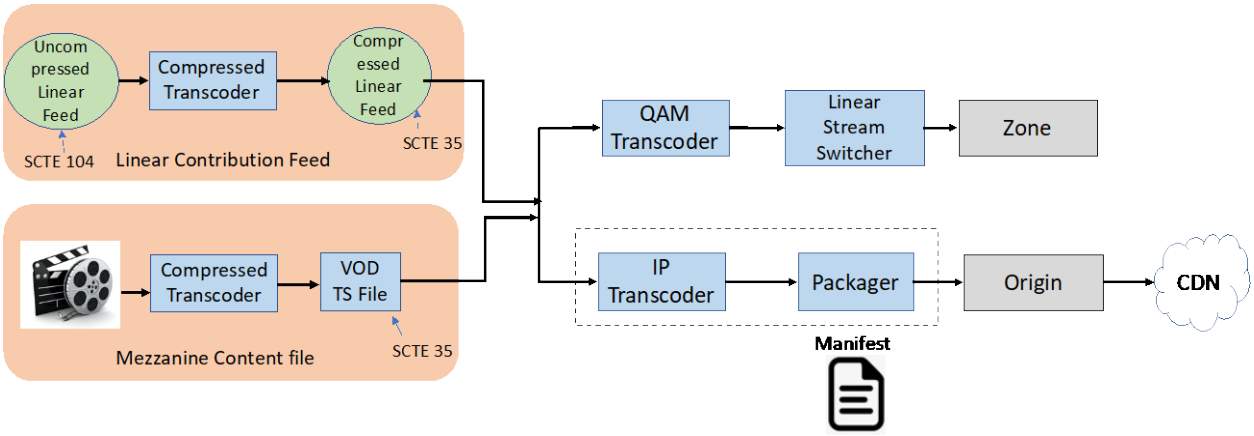

Figure 1 – Source Content workflows for contribution linear feeds and mezzanine files to distribution ABR Streaming and QAM services

This paper describes new mechanisms in existing structures added, proposed or planned in the MPEG-2 Systems (ISO/IEC 13818-1 8th ed) [9], SCTE 214 (DASH constraints) [3]. These additions to existing specifications can then be used in adapting the infrastructure to allow for better automation in mapping of source content into DASH structures for distribution of ABR streaming services.

2. Problem Statement

Between providing MPEG-2 TS source content to the MVPD through traditional source linear channels or VOD mezzanine files and then transcoding/packaging/ and distributing that content for adaptive streaming services, there is no clear mapping from source to distribution that can adjust to source content variations and automate this into adaptive streaming services while providing DASH-enabled, enhanced consumer experiences to the content.

Adaptive Streaming technologies can provide a more individualized customer experience for an asset with choices to match accessibility, language, commentary choices of the viewer. With that said, this is predicated on media content being identifiable and available for these purposes. For example, a selection of English, Spanish, French, and Chinese subtitles assumes that the player is aware of their existence, while a good user experience also depends on a sensible default client behavior. The contribution video ingest infrastructure for video feeds or VOD mezzanine is still operating on an MPEG-2 TS format, be it via fiber, satellite, or IP using TS-over-UDP, SRT, or RIST delivery protocols.

The MPEG-2 TS structure is more geared towards ingestion for QAM-based linear channels and VOD assets where language offerings are more static and limited to 1-2 choices which are determined by hard-coded PID convention ordering (e.g.,101- English, 102-Spanish). Automatically mapping these existing methods over to DASH Role and Accessibility constructs would limit players to static configuration of the channel or the VOD Asset unless a assets specific specialized workaround is done [4]. Additionally, accessibility services such as Audio Description (AD)/ Descriptive Video Services (DVS) are gaining popularity [10]. Initially accessibility tracks were represented using the same hard-coded audio PID construct and often signaling dead or archaic languages (e.g. Middle English) was needed to distinguish between these and the secondary language. Automatically mapping such a setup into DASH constructs is predominantly a manual process because the orthogonality between language and accessibility characteristics of audio streams were determined through PID numbering conventions and overloaded use of the ISO_639 language codes carried in the audio descriptor in the PMT. Often accessibility features would differ program by program on the channel which would require manual per-channel configuration which is usually limited to be a channel configuration instead of a program configuration. Moreover, this approach does not scale.

Lastly, the consumer experience from switching from a main program to an ad needs to be consistent. Achieving consistency becomes harder as the number of player options increase. Similarly, ads also need the same information for smooth playback experience. At the streaming player, another factor needed is the default playback mode set by the customer [10]. Once the player default playback information is known, the switching behavior between the main program and ad, and the playback behavior, can be deterministic and consistent even if the content experience between the main program and the ad do not overlap. But the current mechanisms cannot carry this information dynamically so it can change program to program. An understanding of the content experience offerings in the main program and the default consumer experience of the main program is needed to provide a deterministic behavior for the customer experience as the program moves into an ad or alternate content situation.

3. Latest additions to the MPEG-2 TS language descriptor

The MPEG-2 TS specification contains an ISO_639_language descriptor which has a three-character ISO 639-2 field for the language and an 8-bit audio_type field. [9][15].The descriptor resides in the PMT section. This information can change every PMT occurrence but should be limited to program boundary changes. The descriptor can add multiple table entry pairs of ISO 639-2 languages codes and audio_type codes, but current usage anticipates a single language-audio_type pair. In earlier editions of the MPEG-2 Systems standard, the audio_type codes were limited to accessibility roles -- visually impaired commentary (0x03), hearing impaired (0x02), clean effects (0x01), in addition to undefined (0x00) used for all other audio roles. In past common usage, a single occurrence of the language was used but the audio_type was limited to the undefined values [9][15]. If there was further need to define the audio, the bsmod descriptor values were used but were limited to specific types of Audio formats. These were tolerable in QAM delivery but become more of an issue with multiple different audio formats needing to be carried in adaptive streaming environments [3].

With the 8th Edition of ISO/IEC 13818-1, the structure of the ISO_639_language descriptor was not changed as it is ubiquitously used throughout the ecosystem. However, more values were added to the audio_type table [9]. The resulting efforts provide a clearer approach to map information from the MPEG-2 Stream into DASH values for audio services and accessibility associated services.

Table 1 – Mapping source audio_type values or bsmod descriptors (for (E)AC3) to Role and Accesibility values for audio services

Source Audio_type or bsmod descriptor	Role@value	Accessibility@value
Audio default (audio_type = 0x00 | bsmod [ST] = 000)	Main	N/A
Clean effects (audio_type = 0x01 | bsmod [ST] = 001)	SCTE: Music & Effects	N/A
Primary Audio (audio_type = 0x80)	main[i]	N/A
Native Audio (audio_type = 0x81)	absence of dub	N/A
Emergency (audio_type = 0x82 | bsmod [ST] = 110)	Emergency	N/A
Primary Commentary (audio_type = 0x83 | bsmod [ST] = 101)	main[ii], commentary	N/A
Alternate Commentary (audio_type = 0x84)	alternate, commentary	N/A
Bsmod [ST] = 100 or 111	TBD	N/A

Table 2 – Mapping source audio_type values or bsmod descriptors (for (E)AC3) to Role and Accesibility values for accessibility associated services

Type	Role@value	Accessibility@value
Audio description (audio_type = 0x03 | bsmod [ST] = 010)	alternate	description
Clean audio (audio_type = 0x02 | bsmod [ST] = 011)	alternate	enhanced-audio-intelligibility
Closed Captions[1]	main	captions
Sign language[2]	supplementary	sign

Tables 1 and 2, above, exemplify how one can map the signaling in the MPEG-2 TS mezzanine to the DASH Role or Accessibility values in the DASH MPD for consumer delivery. A similar mapping can be constructed for HLS which uses a different vocabulary for the purpose. It is also possible to map the bsmod values in (E-)AC-3 bitstreams to the DASH constructs if the information is not provided through the ISO_639_language descriptor. AAC provides functionally similar accessibility parameters. Given a basic pair of language code and Audio_type (set to undefined), this would allow the manifest to define a main role for each audio component using a specific language. With the addition of multiple pairs of Language-Audio_types, these same media components could also additionally define the stream as containing AD/ DVS. The tables provide a reference to automatically map what is listed on the transport stream, to role and accessibility assignments in the manifest. To be noted, this mapping in Table 1 and Table 2 is compatible with the role assignments and accessibility assignments of DVB in areas where these configurations overlap [7].

[1] Closed captioning is an accessibility component for a video or text track indicated by the caption service descriptor in the PSI. The equivalent audio_type value would be 0x00

[2] Sign language is identified through the @lang attribute (e.g. “ase” or “bfi”, as defined in ISO 639-3). The proper audio_type value would be 0x00.

In the 8th edition, the audio_type now has additional values of *primary, native, emergency, primary commentary, and alternate commentary*. With these additional values and the use of multiple language-audio_type pairs, there is even more that can be automatically transformed into DASH contexts [9]. For instance, using an audio_type value of "native" can distinguish between original and dubbed audio tracks in a manner that can be automatically processed. With the use of primary and alternate content, a sporting event could have more than one announcer, or just stadium sound, depending on the preference of the listener. In VOD assets, the number of audio experience choices can keep growing but with some direct mapping these can be automatically captured within a single manifest. But even for linear channels, these additional values for audio_type can be beneficial to indicate things like sporting event, native audio, emergency channel audio, and primary audio of the channel. Furthermore, additional information on the audio tracks can benefit DASH constructs by providing a way to have continuous audio media components in the created period, through avoiding overloading of the ISO 639 language value to additionally indicate audio streams with properties like audio description.

Figure 2 and Figure 3 show how the MPEG-2 TS language descriptor values could be set to dynamically provide several playouts, for both linear channels and mezzanine files. These can vary from program to program on a linear channel, or asset to asset on VOD services, so the customer experience offerings can be tailored to the content as well as the experience. For VOD services, there are fewer limitations on customer experiences. The number of languages allowed, for instance, can make a content asset more worldly. Distinguishing between dubbed and native sound, and additional accessibility options, can enable people to hear dialogue more clearly. This benefits both the hard of hearing, and people wanting to watch loud action movies at 2am without disturbing the rest of the family.

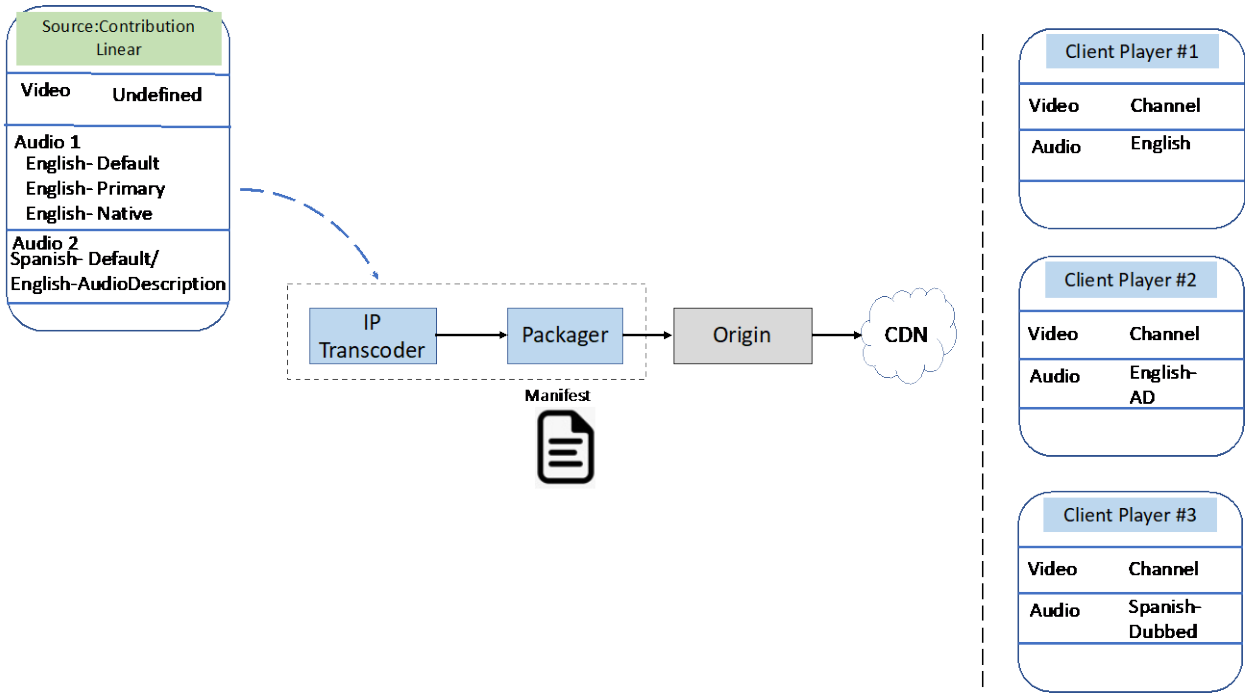

Figure 2 – Example of contribution feed configuration and resulting player options

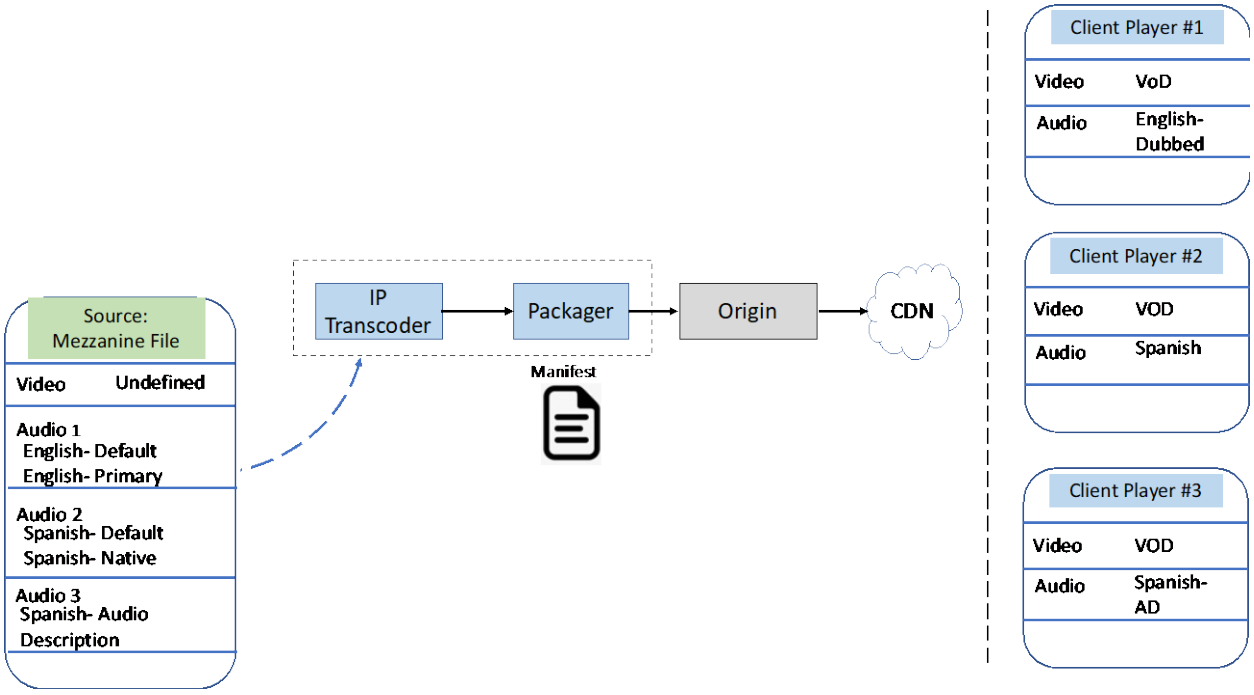

Figure 3 – Example of mezzanine file configuration and resulting player options

Figure 4 shows how the audio_type values can be set to handle different customer experiences for a sporting event. In this case, a sporting event can be offered with two different announcers, or just ambient stadium sound. This can be tailored to use cases like supporting a home announcer and an away announcer (including using different languages), or none at all, or including 3rd party commentary during a sporting event. For example, sourcing comedians for color commentary on sporting events is increasingly popular, as evidenced by the Olympic games.

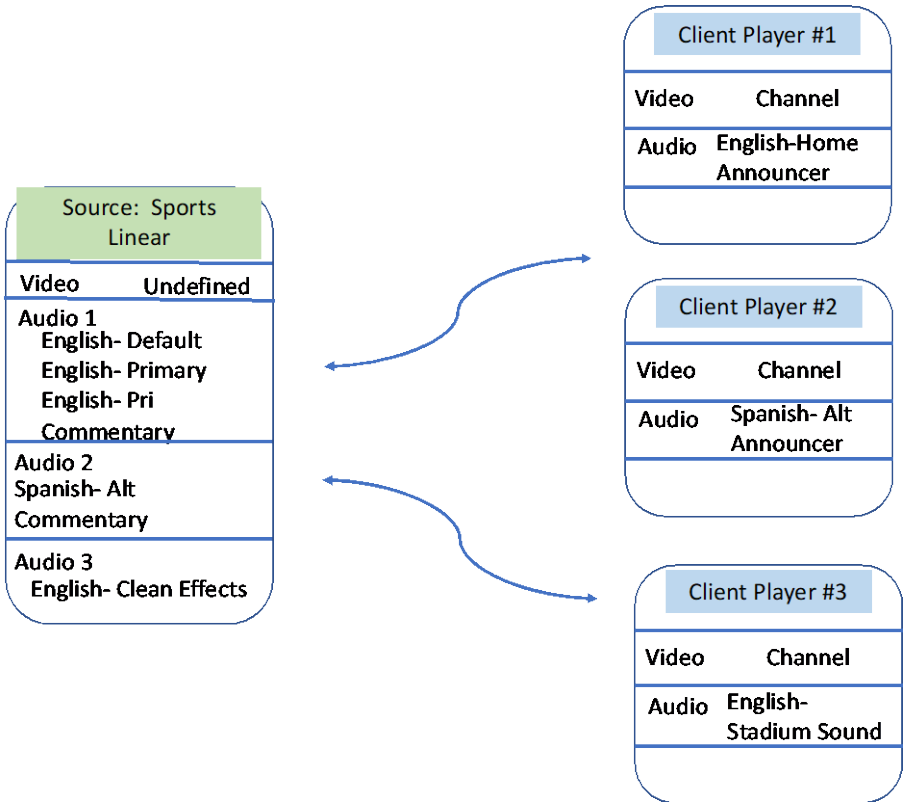

Figure 4 – Sport event contribution feed configuration and resulting player options

4. Using planned modifications to SCTE 214-1

The upcoming SCTE 214-1 revisions have proposed a supplemental property for media language defaults. From a DASH perspective, the player can set consumer preferences for media languages. The NorDig IRD specification provides an example of such logic [16]. From the content side, there was also a media language default that can be indicated in the ISO-639_language descriptor through assigning an audio_type *primary* value, but there is no way to reflect this in the DASH manifest until the media language default supplemental property was proposed. Knowledge of this information helps ad insertion where the language options may not exactly match the main content primary languages. With expanded choices in languages for the content asset, the playout experience can be disruptive between the main program and the 3rd party ad (e.g., switching from French main program to a Spanish ad) and inconsistent depending on implementation. With the primary language of the channel known, this can allow for deterministic ways to map to the customer experience at the client player. Even with an assigned language default language option at the player side, this information is useful especially when no languages overlap -- which can happen during an ad since the ad is independently prepared from the program or asset content. This supplemental property can be placed at the MPD level to indicate language default of the entire content, or provide a language default at the period level, which can override what is set at the MPD level. This supplemental property can also be indicated at a period level as well to indicate in cases of inserted ads what default language to use especially if there is no overlap with the main program [1][2]. From an automation perspective, providing this language default information in the manifest allows an

approach that provides consistent and deterministic playout behavior for the customer experience, even across main program and alternate content such as Ads. Examples of this are shown in Figure 5.

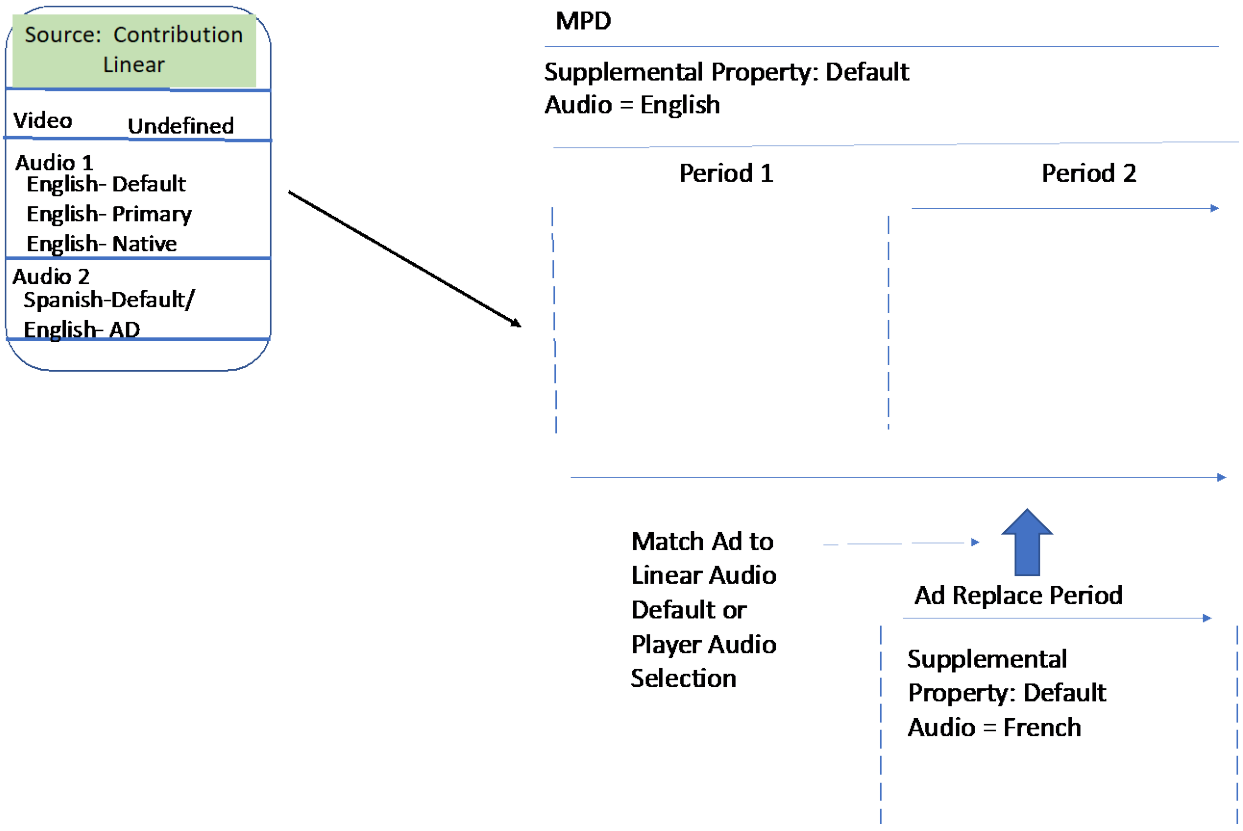

Figure 5 – Use of Media Language Defaults to Align Playout of Main Program and Ads

5. Future and International Extensions

The initial modifications described here worked by only adding additional values to the ISO_639_language structure [11][12]. But the makeup of content assets and linear feeds is expanding and evolving which can go beyond the ISO_639_language Descriptor structure on the audio stream. For instance, more accessibility features such as Video Sign Languages which is placed on the video stream cannot be well described just using the ISO_639_language descriptor. It also introduces the shift from having the content asset built around a single video stream to having several dependent video streams from 3rd parties as part of the content asset. With adaptive streaming there is also a shift to demuxing the media components which may manifest upstream delivery as demuxiplexed MPEG-2 TS streams which may originate from different parties (e.g., multi-language subtitles, dubbed audio, multi video sign languages).

In the next amendment of 13818-1, a new descriptor is proposed called the *Media_service_kind* descriptor that can co-exist with the ISO_639_language descriptor. This allows for a transition strategy from the old descriptor to the new descriptor depending on equipment software modifications. This new descriptor

expands from the ISO_639_language descriptor by applying to video and text streams as well as the audio descriptor and provides a main and dependent relationships between the different media components of the content asset. It also provides more additions to the media_type (similar to audio_type but extended to additional types of media components such as video or text) to address a more complete match between the full set of DASH roles and what is carried in the transport stream which includes values such as forced subtitles, substitution and dialogue. This also accommodates the future trends of having the media components of content being delivered on separate paths by multiple parties.

For the international requirements, the language code will be referencing ISO 639-3 to accommodate multiple sign languages and more dialects with BCP 47 extensions to handle regional dialect languages and scripts [5][6]. Furthermore, it will handle multi-language native audio tracks that may differ from a program-to-program basis but always provide the native audio of the program (ISO 639-3 “mul” code point.) In different parts of the world, this can be a differential playback mode for individuals who know multiple languages and would prefer the program to be played in the native language.

6. Conclusion

Enabling automation needs to define a clear path from mapping content source transport streams to distribution adaptive streaming manifests. With the additional values of audio_type in the ISO_639 language descriptor, adding the supplemental property media defaults in SCTE 214 [3]. We believe these enhancements will provide a clearer path from mapping information carried from the content source into its equivalent DASH roles and accessibility values.

The benefits of having this automation are clear. With DASH Role and Accessibility elements, the content is well described which benefits the customer by providing many more different types of experiences with the same asset and making it more accessible. In the past multiple assets of the same content had to be created separately and displayed differently to capture the different ways of viewing the content asset. But unless there is a way to automatically map this information from MPEG-2 TS structures, the presentation of the content would be limited to avoid the complexities of manually and correctly adding this information to each linear feed or VOD asset. The need to map from a MPEG-2 TS structure is needed due to the existing ingest system that is in place.

Furthermore, there needs to be an approach for this that works for inserting local ads, replacing national ones, or providing alternate content. The content experience can be affected by the ad playout of 3rd party content. Minimal disruption of the experience should be provided as an ad gets played out and when there is no way to avoid disruption of the experience during an ad then the behavior should be deterministic and consistent (e.g. keep the ad in Spanish even if the Spanish audio description is not available).

As content gets more internationalized, the customer experience expands, and these approaches provide a way of including these experiences while still providing a way to put this under the same manifest without changing backend operations.

Lastly in the future content assets and experiences will be expanding to include demuxed media component delivery and things like 3rd party independent signing tracks or new sporting or live event experiences. In anticipation of this, we believe the development of the media service kind descriptor in the next MPEG-2 TS systems standard edition will be needed as well as an approach to have these two descriptors co-exist for some time.

Abbreviations

ABR	Adaptive Bitrate
AD	Audio Description
bps	bits per second
CDN	Content Distribution Network
DASH	[MPEG] Dynamic Adaptive Streaming over HTTP
DVB	Digital Video Broadcasting
DVS	Descriptive Video Services
Ed	Edition
FEC	forward error correction
GOP	Group of Pictures
HD	high definition
HLS	HTTP Live Streaming
IEC	International Electrotechnical Commission
IP	Internet Protocol
ISBE	International Society of Broadband Experts
ISO	International Organization for Standardization
MPD	Media Presentation Description
MPEG-2 TS	MPEG -2 Transport Stream
MVPD	Multichannel Video Program Distributor
PID	Packet Identifier
PMT	Program Map Table
QAM	Quadrature Amplitude Modulation
RIST	Reliable Internet Streaming Transport
SAP	Stream Access Point
SCTE	Society of Cable Telecommunications Engineers
SRT	Secure Reliable Transport
VOD	Video On Demand
UDP	User Datagram Protocol

Bibliography & References

[1] ANSI/SCTE 35, Digital Program Insertion Cueing Message for Cable.

[2] ANSI/SCTE 104, Automation System to Compression System Communications Applications Program Interface (API).

[3] ANSI/SCTE 214-1, MPEG DASH for IP-Based Cable Services Part 1L MPD Constraints and Extensions.

[4] ANSI/SCTE 223, Adaptive Transport Stream

[5] Phillips, A. and M. Davis, "Matching of Language Tags", BCP 47, RFC 4647, September 2006.

[6] Phillips, A., Ed., and M. Davis, Ed., "Tags for Identifying Languages", BCP 47, RFC 5646, September 2009.

[7] ETSI TS 103 285 V1.3.1, Digital Video Broadcasting (DVB); MPEG-DASH Profile for Transport of ISOBMFF Based DVB Services over IP Based Networks.

[8] [HLS I-D] IETF RFC 8216bis, HTTP Live Streaming

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

[10] ISO/IEC 23009-1: Information technology – Dynamic adaptive streaming over HTTP (DASH) – Part 1: Media presentation description and segment formats

[11] ISO 639-2, Codes for the Representation of Names of Languages - Part 2: Alpha-3 Code

[12] ISO 639-3, Codes for the Representation of Names of Languages - Part 3: Alpha-3 Code for Comprehensive Coverage of Languages

[13] [RIST] VSF TR-06-1, Reliable Internet Stream Transport (RIST) Protocol Specification (Simple Profile), October 2018

[14] [SRT] M.A. Sharabayko, J. Dube, J.S. Kim, J.W. Kim, The SRT Protocol, March 2021, available at https://datatracker.ietf.org/doc/html/draft-sharabayko-srt-00

[15] Jan Van Der Meer, Fundamental and Evolution of MPEG-2 Systems: Paving the MPEG Road, Whiley, 2014.

[16] NorDig Unified Requirements for Integrated Receiver Decoders for use in cable, satellite, terrestrial, and managed ITV based networks, 27 October 2018, www.nordig.org

[i] The supplemental property for media language default for audio should be set at the MPD or Period Level
[ii] The supplemental property for media language default for commentary should be set at the MPD or Period Level

