Ultra HD Forum Guidelines

Blue Book – Ultra HD Production and Post Production

Ultra HD Forum

8377 Fremont Blvd., Suite 117

Fremont, CA 94538

UNITED STATES

VERSION 3.3.0

Sep 10, 2024

1. Foreword

The Ultra HD Forum Guidelines provides a holistic view of modern media systems, their mechanisms and workflows, and how those are impacted by the latest generation of improvements – the “Ultra HD” technologies, those that take media beyond the limits established at the start of this millennia, as characterized in large part by the video resolutions and the dynamic range offered for media in of “high definition” (i.e., ITU-R Rec. BT.709). The Forum considers Ultra HD to not only be any UHD media (i.e., 4K resolution, or higher), but also HD-resolution media with enhancements such as High Dynamic Range, Wide Color Gamut, etc. Ultra HD is a constellation of technologies that can provide significant improvements in media quality and audience experience. In addition, the Forum collaborates in promoting the understanding of the various deployments and delivery methods for Ultra HD media that continuously evolve around the world.

This work represents over eight years of collaborative effort by the membership of the Ultra HD Forum. The Guideline books would not have been possible without the leadership of Jim DeFilippis, who represents Fraunhofer, and chair of our Guidelines Work Group with invaluable support from the co-chair, Pete Sellar of Xperi as well as technical assistance from Ian Nock of Fairmile West Consulting, chair of the Interop Working Group.

Our gratitude to all the companies listed in the Acknowledgments that have participated in this effort over the years and specifically to Nabajeet Barman (Brightcove), Elena Burdiel Pérez (Fraunhofer), Andrew Cotton (BBC), Jean Louis Diascorn (Harmonic), Richard Doherty (Dolby), Felix Nemirovsky (Dolby),Chris Johns (Sky UK), Katy Noland (BBC), Bill Redmann (InterDigital), Yuriy Reznik (Brightcove), Chris Seeger (Comcast/NBCUniversal), Adrian Murtaza (Fraunhofer) and Alessandro Travaglini (Fraunhofer).

This document, *Ultra HD Production and Post Production* (Blue Book), is one of a series of books, referred to as the Rainbow Books, that compose the Ultra HD Forum Guidelines. If any of these terms sound unfamiliar, follow the link below to the Black Book. If a particular standard is of interest, links such as the one above are available to take you to the White Book, where references are collected.

The Rainbow Books are, in their entirety:

White Book Guidelines Index and References

Red Book Introduction to Ultra HD

Orange Book Foundational Technologies for Ultra HD

Yellow Book Beyond Foundational Technologies

Green Book Ultra HD Distribution

Blue Book **Ultra HD Production and Post Production**

Indigo Book Ultra HD Technology Implementations

Violet Book Real World Ultra HD

Black Book Terms and Acronyms

Updates in this new version of the Ultra HD Forum Guidelines are described on the following page.

I hope you will enjoy reading today.

If you want to know more about Ultra HD, and join our discussions on how it can be deployed, I invite you to join the Ultra HD Forum. You can start by visiting our website: www.ultrahdforum.org.

Dr. Yasser Syed, President, Ultra HD Forum
Sept 2024

1.1 Changes from version 3.2 to 3.3

What's new in the September 2024 version of the UHDF Guidelines Blue Book, *Ultra HD Production and Post Production* (v3.1), edited by Jim DeFilippis and Chris Seeger.

The *Ultra HD Production and Post Production* is the fifth of the series of Rainbow Books on the Guidelines for Ultra HD. The scope and purpose of this book is to describe the use of Ultra HD technology for production and post production of UHD content.

While most of the information in this edition is material from the previous version of the Blue Book of the Guidelines (v3.2), the following information has been updated:

Section 7.1.3 - Update to Table 1, Example Display Luminance Values for HLG on Different Displays

Section 7.5.1 - Revision to Single Master Workflow for UHD HDR-SDR Productions

Section 8.1.1 - Revision to Conversion between PQ10 to HLG10

Section 8.1.2 - Revision to Conversion between HLG10 to PQ10

Section 8.3 - Revision to Conversion HDR to SDR: Down-Mapping and update to Fig 5, Two Conversion Workflows Through HDR to SDR Down-mapping

Section 9.2 - Revision to Conversions from SDR/BT.709 to PQ10 and HLG10

Section 10.5 - Revision to Channel-based, Backward Compatible, Immersive Audio Post Production

Section 10.6 - Revision to Still Image Content Creation with HDR and SDR and update to Table 9, CICP Image Characteristics

Section 11.2 - New information, Viewing HDR-SDR Images in Close Proximity

Section 12.2.2 - Update to NGA Live Production use and trials

Section 12.3.3.3 - Update to MPEG-H Audio Workflows

References have been updated as well.

Each section includes references to more detailed information contained in the companion Rainbow Books. We've added a reference section to make it easier to navigate to both external and internal references. The references have been updated as well.

We hope this new format will be helpful in understanding UHD technologies as well as planning for new or expanded Ultra HD services.

Jim DeFilippis and Pete Sellar,

Guidelines Working Group Co-Chairs, Ultra HD Forum, September 2024

2. Acknowledgements

We would like to provide the acknowledgement to all the member companies, past and present, of the Ultra HD Forum who have contributed in some small or large part to the body of knowledge that has been contributed to the Guidelines Color Books, including the specific subject of this book.

ARRIS	ATEME	ATT DIRECTV
British Broadcasting Corporation	BBright	Beamr
Brightcove Inc.	Broadcom	B<>COM
Comcast / NBCUniversal LLC	Comunicare Digitale	Content Armor
CTOIC	Dolby	DTG
Endeavor Streaming	Eurofins Digital Testing	Fairmile West
Fraunhofer IIS	Harmonic	Huawei Technologies
InterDigital	LG Electronics	Mediakind
MovieLabs	NAB	Nagra, Kudelski Group
NGCodec	Sky UK	Sony Corporation
Xperi	Technicolor SA	Verimatrix Inc.
V-Silicon		

3. Notice

The Ultra HD Forum Guidelines are intended to serve the public interest by providing recommendations and procedures that promote uniformity of product, interchangeability and ultimately the long-term reliability of audio/video service transmission. This document shall not in any way preclude any member or nonmember of the Ultra HD Forum from manufacturing or selling products not conforming to such documents, nor shall the existence of such guidelines preclude their voluntary use by those other than Ultra HD Forum members, whether used domestically or internationally.

The Ultra HD Forum assumes no obligations or liability whatsoever to any party who may adopt the guidelines. Such an adopting party assumes all risks associated with adoption of these guidelines and accepts full responsibility for any damage and/or claims arising from the adoption of such guidelines.

Attention is called to the possibility that implementation of the recommendations and procedures described in these guidelines may require the use of subject matter covered by patent rights. By publication of these guidelines, no position is taken with respect to the existence or validity of any patent rights in connection therewith. Ultra HD Forum shall not be responsible for identifying patents for which a license may be required or for conducting inquiries into the legal validity or scope of those patents that are brought to its attention.

Patent holders who believe that they hold patents which are essential to the implementation of the recommendations and procedures described in these guidelines have been requested to provide information about those patents and any related licensing terms and conditions.

4. Contents

5. List of Figures

List Of Images

6. List of Tables

List Of Tables

7. Ultra HD Production and Post Production

The Ultra HD Forum is concerned with establishing viable workflows both for Real-time Program Services and On Demand content that was originally offered live. Real-time Program Services (aka Linear TV) make frequent use of Pre-recorded material, such as edited inserts, interstitials, etc., which involve production and post-production.

Live content has specific requirements and operating practices that are unlike Digital Cinema, Blu-ray™ disc mastering, or other Pre-recorded content practices. Ultra HD workflows and technologies that are designed for these other delivery methods may not apply to Live content production.

Production practices for Foundation Ultra HD audio are similar to those used in current HD content creation. Audio follows multi-channel workflows established for multi-channel 5.1 surround delivery using (as appropriate) AC-3 [29], DTS-HD [111], E-AC-3+JOC (an instance of Dolby Atmos®[1]) [35], HE-AAC [27], or AAC-LC emission [27]. Although 2.0 stereo sound is possible, Foundation Ultra HD content is considered premium content, and it is therefore recommended that productions provide at least 5.1 channels.

Foundation Ultra HD, production practices for closed captions and subtitles are also similar to those of HD content creation. Closed captions and subtitles follow workflows established for CTA- 608[18]/CTA-708[19], ETSI 300 743[20], ETSI 300 472[21], SCTE-27[22], or IMSC1 [23] formats.

The remainder of this section will focus on mechanisms for producing the video components of the content.

As content is produced, it is useful to know in advance for which service mode(s) the content is intended. Equally, service providers planning to deliver Foundation Ultra HD content need to have an understanding of the formats in which the content will be supplied.

The following is a diagram providing an overview of the content production and distribution workflow for Real-time Program Services and potentially capturing Live content for later distribution via an On-Demand Service.

[1] Dolby, Dolby Atmos, Dolby Digital and Dolby Vision are trademarks of Dolby Laboratories.

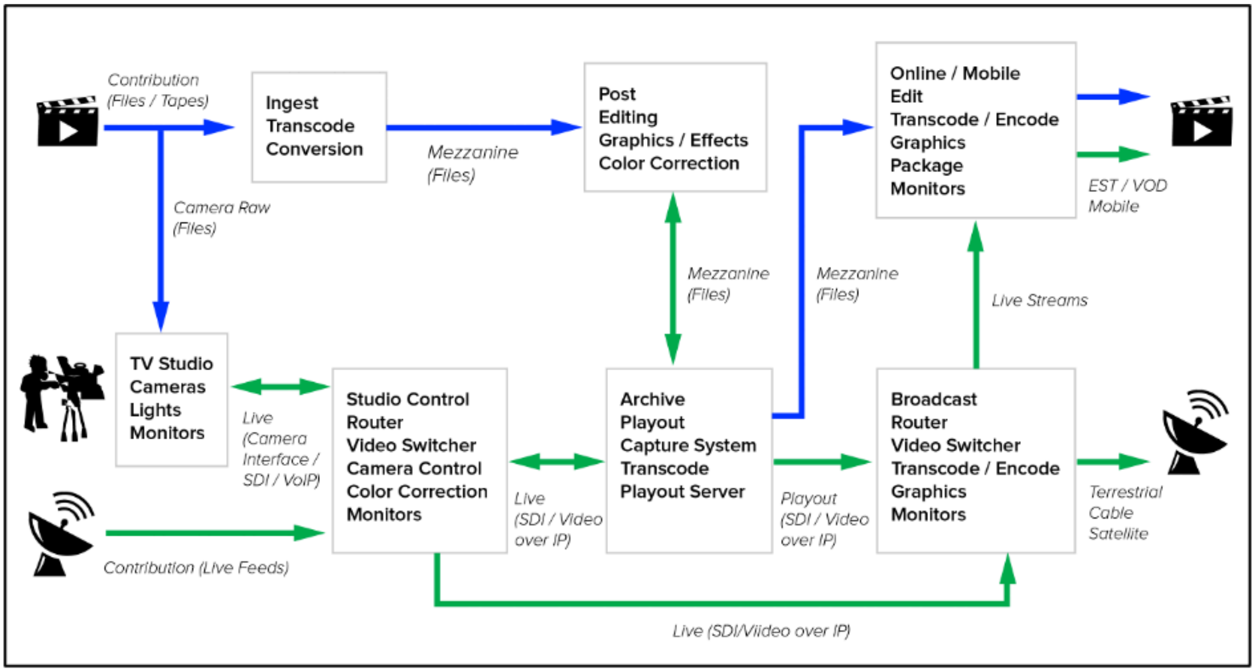


Figure 1. Content Production and Distribution Workflow

7.1. HDR and WCG Technologies

There are many terms in use in the field of HDR television. This section explains the existing terminology and how terms are used in these Guidelines.

Note that currently some UHD displays are capable of accepting BT.2020 [3] content, but as of this publication, no direct view display is available that is capable of rendering the full gamut of colors in the BT.2020 system colorimetry. It is assumed that in these cases, the device employs "best effort" gamut mapping tailored to its particular display characteristics, and thus these devices are considered BT.2020 compatible.

7.1.1. Foundation Ultra HD HDR Technologies

The following HDR systems are recommended for Foundation Ultra HD due to their conformance with the criteria listed in Section 4 above:

- HLG10: HDR systems or content employing Hybrid Log-Gamma (HLG), the wide color gamut specified in BT.2100 [5] and 10-bit depth

- PQ10: HDR systems or content employing Perceptual Quantization (PQ), the wide color gamut specified in BT.2100, and 10-bit depth
- HDR10: HDR systems or content employing PQ10 and further including or capable of providing SMPTE ST 2086 [10], MaxFALL, and MaxCLL static metadata

It should be noted that Real-time Program Services are typically comprised of both live and pre-recorded content, and it is not recommended that service providers alternate between SDR and HDR signal formats or mix different HDR formats. See Section 8.1 for details.

7.1.2. Perceptual Quantization (PQ) and PQ10 HDR

One HDR transfer function set defined for use in television is the "Perceptual Quantization" (PQ) developed by Dolby. PQ is defined as a reference display referred transfer function, or EOTF. The transfer function is designed to minimize visibility of banding on a display between the brightness ranges of 0 to 10,000 cd/m^2. SMPTE standardized the PQ EOTF in ST 2084 [9] and the ITU standardized the EOTF, OETF, and OOTF transfer functions in ITU-R BT.2100 [5]. ST 2084 specifies the shape of the curve over a range of 0 to 1 but does not specify SDI code values for the 0 and 1 values. Table 9 in ITU-R BT.2100 describes the code mapping for both narrow range and full range code values[2].

The Ultra HD Forum has defined the term PQ10 to refer to an HDR system employing the PQ transfer functions and wide gamut colorimetry specified in BT.2100 and 10-bit depth. PQ10 content or systems may or may not be metadata-capable. For example, HDR10 is a Foundation Ultra HD PQ10-based format that has the capability of including certain static metadata (see Section on HDR 10).

The PQ signal is "display-referred", meaning that the pixel-encoded values represent specific values of luminance for displayed pixels. The intent is that only the luminance values near the minimum or maximum luminance capability of a display are necessarily adjusted to utilize the available dynamic range of the display. Some implementations may apply a "knee" at the compensation points in order to provide a smoother transition from the coded values to the display capabilities; e.g., to avoid "clipping".

[2] CTA 861-I [31] offers additional information about code ranges.

When default display settings are engaged, PQ enables pixel values in the mid-range, including skin tones, to be rendered on a display at the same (absolute) luminance level that was determined at production.

For example, if a scene was graded on a 1000 cd/m^2 grading monitor and then displayed on a 4000 cd/m^2 display, the skin tones can be rendered at the same luminance values on the 4000-nit display as on the 1000-nit monitor per the grader's intent, while the speculars and darker tones can be smoothly extended to take full advantage of the 4000-nit display. See also Section 7.5..

The Ultra HD Forum InterOp testing has shown that consumer displays vary significantly, especially when different viewing "modes" are selected (e.g., "Sports", "Vivid", or "Cinema/Movie" picture modes). Research has also shown that ambient light in the viewing room can have an impact on humans' perception of a displayed image. Displays used in non-reference viewing environments may employ adjustments to the PQ curve in order to provide some compensation for the difference between the actual viewing environment and the reference viewing environment.

PQ content requires a down-mapping step in order to provide acceptable SDR quality. See Section 8.3 for a deeper discussion of backward compatibility, including the pros, cons and open questions that apply to various possible methods.

7.1.3. Hybrid Log-Gamma (HLG) and HLG 10

Another HDR transfer function defined for use in television is "Hybrid Log-Gamma" (HLG), developed by the BBC and NHK. This is defined as a camera capture transfer curve, or OETF. This curve was designed to provide HDR while maintaining a degree of backward compatibility with SDR/BT.2020 displays. The HLG curve was first specified in ARIB STD-B67 [105][3], and later in ITU-R BT.2100 [5]. The two OETF specifications are identical. BT.2100 additionally specifies the display EOTF and the OOTF, which are not defined in the ARIB standard.

[3] The original v1.0 ARIB specification referred to the breakpoint between the logarithmic and gamma portions of the HLG curve as the "reference white level", which some manufacturers incorrectly assumed to be the signal level for "HDR Reference White". Later versions of the ARIB specification have removed the term "reference white level" to avoid any such confusion. The "HDR Reference White" level for both HLG and PQ production is defined in ITU-R BT. 2408 [8].

The Ultra HD Forum has defined the term HLG10 to refer to an HDR system employing HLG, wide gamut colorimetry specified in BT.2100 and 10-bit depth with black at code 64, and nominal peak at code 940[4].

HLG is a "scene-referenced" HDR technology that uses pixel-encoded values that are intended to represent picture luminance levels relative to each other in a given scene. The intent is that all the displayed pixel luminance values may be adjusted in a defined manner to compensate for specific display capabilities (e.g., peak luminance) and viewing environments in such a way that the values retain their perceptual appearance relative to one another.

As the eye's sensitivity to light intensity is approximately logarithmic, a power (or "gamma") law is applied to the HLG relative scene-light pixel values, to scale them to span the luminance range of the display; thereby approximating the same relative subjective brightness values that were determined at production. Since no one luminance range is "fixed", the (absolute) displayed luminance of mid-range values, including skin tones and reference white, will increase or decrease to scale along with all other values in order to better preserve the appearance of the relative brightness values of the pixels in the scene.

For example, if a scene was graded on a 1000 cd/m^2 grading monitor and then displayed on a 2000 cd/m^2 display, the skin tones and other scene elements will be brighter on the 2000 cd/m^2 display than on the 1000 cd/m^2 monitor. However, given that all pixel values are adjusted through the "gamma law", the overall image remains perceptually similar to the original "look" selected by the grader. See also Section 6.1.10[5].

As mentioned above, Ultra HD Forum InterOp testing has shown that consumer displays vary significantly, especially when different viewing "modes" are selected (e.g., "Sports", "Vivid"" or "Cinema/Movie" modes). Research has also shown that ambient light in the viewing room can have an impact on humans' perception of a displayed image. Section 6.2 of ITU-R report BT.2390 [6] describes how the display gamma may be adjusted to provide some compensation.

[4] CTA 861-I [31] offers additional information about code ranges.

[5] It is well known that changes in brightness of a display may affect the perception of image color (Hunt Effect) [100] as well as image contrast (Stevens) [101]. Research on the Color Appearance Model (CAM), where brightness exceeds 700 cd/m^2, agrees with Stevens observations; however as of 2020 the BBC has found no evidence of significant changes to the perception of color or contrast in their research using HLG with displays of different peak brightness (Katz)[102], additional research is on-going on these effects with respect to HDR imagery.

HLG is, by design, a relative system, meaning that the displayed luminance of reference white, skin tones, gray levels, and all other signal levels will vary depending on the peak luminance of the display. Displayed luminance values will also differ from SDR displays. A variable gamma accounts for viewer adaptation. Table 1 below illustrates the relative nature of HLG, for HDR Reference White (75% HLG) and a mid-gray level of 38% HLG.

Table 1. Example Display Luminance Values for HLG on Different Displays

Display peak luminance (cd/m^2)	400	600	1000	2000	4000
Reference gamma (see ITU-R BT.2100 [5] Note 5f)	1.03	1.11	1.20	1.33	1.45
Displayed luminance for HDR Reference White, 75% HLG (cd/m^2)	101	138	203	343	581
Displayed luminance for mid-gray at 38% HLG (cd/m^2)	17.3	20.76	26.2	35.51	48.36

Content produced using HLG can be displayed on SDR/WCG(BT.2020) devices with a degree of compatibility that may be judged acceptable for programs and services according to Report ITU-R BT.2390 [6], and subjective tests performed by the EBU, RAI, IRT and Orange Labs [88]. Backward-compatible HLG is only intended to support SDR/BT.2020 [3] (WCG) displays and not intended for displays which only support SDR/BT.709 [2]. See Section 10.4.2 of the Violet Book for a deeper discussion of backward compatibility, including the pros, cons and open questions that apply to various possible methods.

7.1.4. ITU Recommendation on HDR, ITU-R BT.2100

In July 2016 ITU-R Study Group 6 approved the publication of a Recommendation on HDR for use in production and international program exchange and is known as ITU-R BT.2100 [5]. This recommendation includes the following specifications:

- Spatial resolutions: 1080p, 2160p, 4320p
- Frame rates: 24/1.001, 24, 25, 30/1.001, 30, 50, 60/1.001, 60, 100, 120/1.001, 120
- System Colorimetry: Same as BT.2020 [3]
- Reference Viewing Environment for critical viewing: 5 cd/m^2 luminance of surrounds and less than or equal to 5 cd/m^2 for luminance of the periphery (the remaining area outside of the surrounds which affect eye adaption and cause reflections on the display)
- Reference non-linear transfer functions EOTF, OOTF, OETF: PQ and HLG
- Signal Format: $Y'C'_BC'_R$, IC_TC_P, and RGB.
- Color sub-sampling: same alignment as specified in BT.2020
- Bit depth value ranges:
    - 10 bit, Narrow (64-940) and Full (0-1023) ranges
    - 12 bit, Narrow (256-3760) and Full (0-4095)
    - Floating point signal representation: Linear RGB, 16-bit floating point

BT.2100 is the primary reference document on HDR for use in production and international program exchange. Note that a full signal specification will need to include the following attributes: spatial resolution, frame rate, transfer function (PQ or HLG), color signal format, integer (10 or 12 bits, narrow or full range) or floating point.

Not all of the parameter values listed above have been deployed as yet, but are included as informative details.

7.2. Static HDR Metadata-SMPTE ST 2086, MaxFALL, MaxCLL

SMPTE has specified a set of static metadata in the SMPTE ST 2086 [10] Mastering Display Color Volume Metadata (MDCV) Supporting High Luminance and Wide Color Gamut Images standard. Parameters included indicate the characteristics of the mastering display monitor. The mastering display metadata indicates that the creative intent was established on a monitor having the described characteristics. If provided, the implication is that the artistic intent of the content is within the subset of the overall container per the metadata values. The mastering display characteristics include the display primaries and white point as x,y chromaticity coordinates, and the maximum and minimum display luminance. For example, the metadata may indicate that the system colorimetry of the mastering display uses the DCI-P3 color gamut in a BT.2020 [3] container, and the luminance range is a subset of the 0 to 10,000 cd/m^2 representable in PQ.

The Blu-ray Disc Association and DECE groups have defined carriage of two additional metadata items:

- MaxFALL – Maximum Frame Average Light Level; this is the largest average pixel light value of any video frame in the program
- MaxCLL – Maximum Content Light Level: this is the largest individual pixel light value of any video frame in the program

Static metadata may be used by displays to control color volume transforms of the received program to better reproduce the creative intent as shown on the mastering display, given the capabilities of the display device. However, for MaxFALL and MaxCLL static metadata specifically, there are limitations for use with live broadcasts since it is difficult to determine a program's maximum pixel light values in a program that is on-going and has no determined endpoint. According to the UHD Alliance, today's mastering practices may generate outlier values unintentionally, causing the content's associated MaxCLL value to be higher than expected. As a response to that, some content providers use statistical analysis to calculate a MaxCLL value that is more representative of the statistically significant brightest pixels contained in the image sequence. The Ultra HD Forum further suggests that such statistical methodology may also apply to MaxFALL.

It is worth noting that code levels below minimum mastering display luminance and code levels above maximum mastering display luminance were likely clipped on a professional reference monitor and therefore any detail in pictures utilizing code levels outside this range were likely not seen in the content production process.

7.2.1. HDR10

HDR10 is a PQ10-based HDR format. The term "HDR10" is in widespread use and has been formally and consistently defined by several groups, including DECE and BDA, as:

- Transfer function: Rec. ITU-R BT.2100 [5] (PQ)
- System Colorimetry: BT.2100
- Bit depth: 10 bits
- Metadata: SMPTE ST 2086 [10], MaxFALL, MaxCLL

Several delivery formats (e.g. Ultra HD Blu-ray™ and CTA's HDR10 profile) have specified delivery using the above parameters with the metadata mandatory but are still considered to be using HDR10 for the purposes of this document. While ATSC A/341 does not use the term "HDR10," when the PQ transfer function is used, the static metadata video parameters described therein are consistent with HDR10 as defined herein.

Note that some displays ignore some or all static metadata (i.e., ST 2086 (MDCV), MaxFALL, and MaxCLL); however, HDR10 distribution systems must deliver the static metadata, when present.

7.2.2. HDR10 Metadata Generation

HDR10 includes the static metadata described in Section 6.1.5. MaxFALL and MaxCLL metadata could be generated by the color grading software or other video analysis software. In Live content production, MaxFALL or MaxCLL metadata is not generated during the production process. By definition, it is not possible to generate MaxFALL or MaxCLL for a Live program because these cannot be known until the entire program is produced (i.e., after the program is over). A special value of '0' (meaning, "unknown") is allowed for MaxFALL and MaxCLL. It may be possible to set limits on the output and thus pre-determine MaxFALL and MaxCLL even for Live content production. For example, if MaxFALL and MaxCLL metadata values are provided, a video processor could be inserted in order to reduce brightness of any video frames that would exceed the indicated values (similar to the way audio processors are often inserted to enforce audio loudness and peak levels).

If it is desired to use HDR10 in Live content production, but the production facility does not support carriage of the metadata, then default values for ST 2086 [10], MaxFALL and MaxCLL should be determined and entered directly into the encoder via a UI or a file. SMPTE 2086 metadata could be set to values that represent the monitors used for grading during production of the content.

7.2.3. HDR10 Metadata Carriage

SMPTE ST 2086 metadata is carried with Mastering Display Color Volume (MDCV) metadata.

In HEVC and AVC bitstreams, MDCV metadata is carried via SEI message, (see DASH-IF Guidelines for Interoperability Points, ISO/IEC 23008-2 [16], section D.2.28 (Syntax) and D.3.28 (Semantics). Complementary metadata that often can accompany MDCV metadata are MaxFALL/MaxCLL or "Content Light Level Information" (CLLI) which is carried in a static SEI message (sections D.2.35 (Syntax) and D.3.35 (Semantics), [26]). It is expected that the decoder/display will employ “best effort” to tone map the content optimally using this metadata. Where the metadata doesn’t exist or is dropped, the end-to-end chain will operate in a manner indistinguishable from PQ10 content (i.e. PQ without metadata).

Both MDCV metadata exists to optimize both tone and color mapping on target displays that differ from the display that created the content. CLLI metadata exists so that a target display can optimize the tone mapping based on characteristics of the content itself.

Both the MDCV and CLLI (MaxCLL/MaxFall) SEI messages can be represented by their corresponding ANC messages as specified in SMPTE ST 2108-1 [48] and can accompany each video frame to which they apply. An SMPTE ST 2108-1 ANC message is a simple, bit-accurate encapsulation of the corresponding SEI message, to ease handling by an encoder or decoder; as a result, SMPTE ST 2108-1 is also able to support certain dynamic metadata technologies (see Section 12). This same metadata can be embedded in a SMPTE ST 2110 IP data stream (SMPTE ST 2110-40 [47]). Additional dynamic metadata systems may also travel down this pipeline (see Section 6.1.12)

Both MDCV and CLLI can be embedded in QuickTime, MP4 and MXF. In QuickTime and MP4 they are embedded in separate MDCV and CLLI atoms(QuickTime) or boxes(MP4). In MXF Image Essence Picture Descriptors are used to embed both MDCV and CLLI static metadata (UL values are defined in SMPTE ST 2067-21:2016 [38]).

As of 2024, W3C has added the ability to signal MDCV and CLLI in still image files as described in PNG 3rd Edition.[6]

Note that in the metadata pipeline there may be multiple points where the metadata could get lost so the values are often not added until the very end of the production process. In transmission, there may be multiple encoders in the end-to-end broadcast chain. In the event that HDR10 is used and the metadata does not reach the decoder, default values are often used.

7.3. Signaling Transfer Function, System Colorimetry and Matrix Coefficients

The system colorimetry, transfer function (SDR or HDR), and matrix coefficients must be known to downstream equipment ingesting or rendering content in order to preserve display intent. This is true for file transfers in file-based workflows and in linear content streams in linear workflows.

[6] See https://github.com/w3c/PNG-spec/blob/main/Third_Edition_Explainer.md

In file-based workflows, mezzanine file formats such as IMF/MXF (stored in picture essence descriptors),QuickTime, MP4 (stored in NCLC atoms) are often used.

As of 2024, W3C has added the ability to embed CICP in still image files as described in PNG 3rd Edition. This ability is supported in a number of content creation packages.

In baseband SDI based video, CICP values are transmitted via VPID (Video Payload Identifier) within HD-SD, SMPTE ST 292 [99]. In ST 2110 IP-based video, a Media Type Parameter embeds the CICP values. In linear compressed transmission workflows, these values are typically signaled in the VUI of an H.265/HEVC or H.264/MPEG-4 AVC bitstream. Details on SEI and VUI messaging are available in the HEVC specification [26], in particular, Appendix D (SEI) and Appendix E (VUI).

The tables below (Table 2, Table 3 and Table 4) summarize HEVC Main10 Profile bitstream SDR, PQ and HLG indicators. (In HEVC and AVC specifications, the bitstream elements are bolded and italicized to distinguish them from temporary variables and labels). As shown in Table 4 there are two methods of signaling the HLG transfer function.

Table 2. File based Signaling for SDR/BT.709

	System Identifier	BT. 709 YCC	BT. 709 RGB	Full-Range 709 RGB	BT. 601 525	BT. 601 625
Color properties	Color primaries	BT.709	BT.709	BT.709	BT.601	BT.601
	Transfer Characteristics	BT.709	BT.709	BT.709	BT.709	BT.709
	Signal Format	Y′CbCr	R′G′B′	R’G’B’	Y′CbCr	Y′CbCr
Other	Full/narrow range	Narrow	Narrow	Full	Narrow	Narrow
	4:2:0 chroma sample location alignment	Interstitial	N/A	N/A	Interstitial	Interstitial
CICP parameters Rec. ITU-T H.273 ISO/IEC 23091-2 [113] (QuickTime/ HEVC/AVC)[1]	ColorPrimaries	1	1	1	6	5
	TransferCharacteristics	1	1	1	6	6
	MatrixCoefficients	1	0	0	6	5
	VideoFullRangeFlag	0	0	1	0	0
SMPTE MXF parameters SMPTE ST 2067-21[38][2]	Color primaries	06.0E.2B.34.04.01.01.06.04.01.01.01.03.03.00.00			06.0E.2B.34.04.01.01.06.04.01.01.01.03.01.00.00	06.0E.2B.34.04.01.01.06.04.01.01.01.03.02.00.00
	Transfer Characteristic	06.0E.2B.34.04.01.01.01.04.01.01.01.01.02.00.00				
	Coding Equations	06.0E.2B.34.04.01.01.01.04.01.01.01.02.02.00.00	N/R	N/R	06.0E.2B.34.04.01.01.01.04.01.01.01.02.01.00.00	
	Full/narrow level range	Inferred (indicated in black reference level, white reference level, Color range)				
	4:2:0 chroma sample location alignment	Inferred (ChromaLocType=0)	N/A	N/A	Inferred (ChromaLocType=0)	Inferred (ChromaLocType=0)

Table 2 Notes:

1. QuickTime/MP4/HEVC/AVC: ITU-T Series H.Supplement 19, Usage of Video Signal Type Code Points. https://www.itu.int/ITU-T/recommendations/rec.aspx?rec=13895

2. SMPTE UL's available: https://registry.smpte-ra.org/view/published/labels_view.html

Table 3. File-Based Signaling for SDR/BT.2020

Color properties	**Color Primaries**	**BT.2020**	**BT.2020**
	Transfer Characteristics	**BT.2020**	**BT.2020**
	Signal Format	**Y′CbCr**	**R′G′B′**
Other	**Full/narrow range**	**Narrow**	**Narrow**
	4:2:0 chroma sample location alignment	**Co-sited**	**N/A**
CICP parameters Rec. ITU-T H.273 | ISO/IEC 12091-2 [113] (QuickTime/ HEVC/AVC)[1]	**Color Primaries**	**9**	**9**
	TransferCharacteristics	**14**	**14**
	MatrixCoefficients	**9**	**0**
	VideoFullRangeFlag	**0**	**0**
SMPTE MXF parameters SMPTE ST 2067-21[38][2]	**Color Primaries**	**06.0E.2B.34.04.01.01.0D.04.01.01.01.03.04.00.0**	
	Transfer Characteristic	**06.0E.2B.34.04.01.01.0E.04.01.01.01.01.09.00.00**	
	Coding Equations	**06.0E.2B.34.04.01.01.0D.04.01.01.01.02.06.00.00**	**N/R**
	Full/narrow level range	**Inferred (indicated in black reference level, white reference level, color range)**	
	4:2:0 chroma sample location alignment	**Inferred (ChromaLocType = 2)**	**N/A**

Table 3 Notes:

1. QuickTime/MP4/HEVC/AVC: ITU-T Series H.Supplement 19, Usage of Video Signal Type Code Points. https://www.itu.int/ITU-T/recommendations/rec.aspx?rec=13895

2. SMPTE UL’s available: https://registry.smpte-ra.org/view/published/labels_view.html

Table 4. File-Based Signaling for HDR/BT.2020

	System Identifier	BT2100 PQ YCC	BT. 2100 HLG YCC	BT.2100 PQ IC_TC_P	BT.2100 PQ RGB	BT.2100 HLG RGB
Color properties	Color primaries	BT.2020 / BT.2100	BT.2020 / BT.2100	BT.2100	BT.2020 / BT.2100	BT.2020 / BT.2100
	Transfer Characteristics	BT.2100 PQ	BT.2100 HLG	BT.2100 PQ	BT.2100 PQ	BT.2100 HLG
	Signal Format	Y′CbCr	Y′CbCr	IC_TC_P	R′G′B′	R′G′B′
Other	Full/narrow range	Narrow	Narrow	Narrow	Narrow	Narrow
	4:2:0 chroma sample location alignment	Co-sited	Co-sited	Co-sited	N/A	N/A
CICP parameters Rec. ITU-T H.273 ISO/IEC 23091-2 [113] (QuickTime/HEVC/AVC) [1]	Color Primaries	9	9	9	9	9
	Transfer Characteristics	16	18	16	16	18
	MatrixCoefficients	9	9	14	0	0
	Video Full-Range Flag	0	0	0	0	0
SMPTE MXF parameters	Color Primaries	06.0E.2B.34.04.01.01.0D.04.01.01.01.03.04.00.00				
	Transfer Characteristic	06.0E.2B.34.04.01.01.0D.04.01.01.01.01.0A.00.00	06.0E.2B.34.04.01.01.0D.04.01.01.01.01.0B.00.00	06.0E.2B.34.04.01.01.0D.04.01.01.01.01.0A.00.00	06.0E.2B.34.04.01.01.0D.04.01.01.01.01.0A.00.00	06.0E.2B.34.04.01.01.0D.04.01.01

SMPTE ST 2067-21[2]						.01.01.0B.0 0.00
	Coding Equations	06.0E.2B.34.04.01.01.0D.04.01.0 1.01.02.06.00.00		06.0E.2B.34.04. 01.01.0D.04.01.0 1.01.02.07.00.00	N/R	N/R
	Full/narrow level range	Inferred (indicated in black reference level, white reference level, Color range)				
	4:2:0 chroma sample location alignment	Inferred (ChromaLocType = 2)	Inferred (ChromaLocType = 2)	unknown	N/A	N/A

Table 4 Notes:

1. QuickTime/MP4/HEVC/AVC: ITU-T Series H.Supplement 19, Usage of Video Signal Type Code Points. https://www.itu.int/ITU-T/recommendations/rec.aspx?rec=13895
2. SMPTE UL's available: https://registry.smpte-ra.org/view/published/labels_view.html

In one method, the SDR transfer function indicator is signaled in the VUI and the HLG transfer function indicator is transmitted using an alternative transfer characteristics SEI message embedded in the bitstream. In this way, an "HLG aware" STB or decoder/display would recognize that the bitstream refers to content coded with HLG (since it is indicated by the preferred_transfer_characteristics syntax element of the SEI). If an "HLG aware" STB is connected to a TV that does not support HLG, the STB would transmit the SDR indicator over HDMI to the TV. If it is connected to a TV that supports HLG, the STB would copy the transfer function value in the SEI (to indicate HLG) and transmit this over HDMI to the TV.

In the other method, the HLG transfer function indicator is directly signaled in the VUI in the same way PQ or SDR would be signaled.

In theory it is possible to achieve a lossless conversion between the two methods of signaling HLG by flipping the VUI transfer function characteristics indicator value and inserting or removing the alternative transfer characteristic SEI.

Using the first method (i.e., including the SDR transfer function indicator in the VUI and the HLG transfer function indicator in the SEI) enables some level of backward compatibility with SDR/WCG displays. Service providers may also deem results to be acceptable on SDR/WCG displays using the second method (i.e., including the HLG transfer function indicator in the VUI). Service providers may wish to test both methods.

QuickTime, MP4, MXF and PNG 3rd Edition file wrappers use MPEG CICP (Coding Independent Code Points) Index values for signaling. MXF Image Essence Descriptors uses labels corresponding to these same index values. ICC tags also add the ability to embed CICP Index values as well.

As described in Section 11.1, service providers should convert or remap all content into a single, consistent system colorimetry and transfer function. Setting the initial values in the encoder should be adequate assuming the encoder is dedicated to a single format.

As of 2023, while there are methods for signaling system colorimetry, transfer function, matrix coefficients and HDR-related metadata through the end-to-end supply chains, however due to legacy devices with limited support for these capabilities it is extremely difficult to successfully ensure that the signaling and metadata survive through the entire broadcast linear production pipeline. Because gaps exist with legacy devices using SDI or HDMI interfaces and software that doesn't consistently support file-based signaling in the mezzanine file wrapper, verification of proper signaling and accurate video passthrough is recommended. The Ultra HD Forum observes that standards bodies have attempted to address these issues and provided documentation, much of which is referenced herein, but some of which is still under development.

7.4. IC_TC_P Color Representation

The expanding range of display technologies, from various color primaries to increasing dynamic range, is creating a marketplace where color management is becoming increasingly important if artistic intent is to be maintained. In many applications, traditional color transformations may not be possible due to limitations in bandwidth, speed, or processing power. In these cases, image processing such as blending, resizing, and color volume transform must be performed on the incoming signal. With growing color volumes and the increasing need for color processing, distortions already known to be caused by non-uniformity of standard dynamic range (SDR) non-constant-luminance (NCL) $Y'C'_BC'_R$ (hue linearity and constant luminance) will become more prevalent and objectionable.

The IC_TC_P color representation is a more perceptually uniform color format based on the human visual system. The improved decorrelation of saturation, hue, and intensity make IC_TC_P ideal for the entire imaging chain from scene to screen. IC_TC_P follows the same operations as NCL $Y'C'_BC'_R$, making it a possible drop-in replacement. These color processing improvements are achieved by utilizing aspects of the human visual system and by optimizing for lines of constant hue, uniformity of just-noticeable-difference (JND) ellipses, and constant luminance. The perceptually uniform design of IC_TC_P allows for complex tasks such as color volume transform to be easily performed on HDR and WCG imagery with minimal error.

IC_TC_P is included in BT.2100 [5] and is being deployed by OTT service providers as well as implemented by numerous consumer TV manufacturers. Another derivative of IC_TC_P, ITP, is described in ITU-R BT.2124 [107] and is described in Section 8.4 for use as an objective color metric that can be useful in testing the entire broadcast pipeline.

7.5. Peak Brightness Considerations

Luminance values can be coded as “absolute values” on a scale of 0-10,000 cd/m^2 (e.g., PQ format) or as “relative values” (e.g., RAW, gamma, S-log or HLG). The optimal coding method depends on the application and other considerations.

Cameras can be capable of capturing a higher contrast range than reference monitors and consumer displays are capable of rendering. When content is graded, the grader makes judgments according to what he or she sees within the luminance range of the reference monitor, which may have a peak brightness of 1,000 or 2,000 cd/m^2, for example. In theory, if the consumer display characteristics match that of the reference monitor, the consumer will see what the grader intended. In practice of course, consumer HDR displays have varying peak brightness performance, black level and color gamut, so it is up to the display (or source device) to map the color volume of the transmitted content to best match the capabilities of the particular display.

As a PQ signal may carry pixel values as high as 10,000 cd/m^2, it is helpful in the display to indicate the actual maximum pixel value that are expected and are creatively pertinent as a basis for color volume transforms. HDR10 systems are capable of providing static metadata to assist mapping the transmitted content for the consumer display, enabling a given display to optimize the color volume mapping based on its capability. The use of HDR10 metadata may help to optimize display mapping, but as the metadata may not be practical in Live broadcast content, the use of PQ10 may be preferred for Live content. Although a system objective is to preserve creative intent with limited perceptual alteration all the way to the consumer display, some perceptual alteration will occur because of different display capabilities both in peak brightness and minimum black levels.

HLG does not require tone-mapping metadata, and instead has a specified relationship between overall system gamma (implemented as part of the display EOTF) and nominal peak display luminance. An overall system gamma of 1.2 is specified for HLG when displayed on a 1,000 nit monitor. BT.2390 [6] states that the artistic intent of HLG content can be preserved when displaying that content on screens of different peak luminance (even when that display is brighter than the mastering display), through changing the system gamma. BT.2100 [5] provides a formula to determine overall system gamma based on the desired display peak luminance in the reference viewing environment. Section 6.2 of BT.2390 [6] further specifies how the system gamma should be adapted for darker or brighter viewing environments. To preserve creative intent, these formulae are recommended.

7.5.1.Single Master Workflow for UHD HDR-SDR Productions

Report ITU-R BT.2408 [8] Guidance for operational practices in HDR television production suggests a reference level for “HDR Reference White” be set to 75% of the HLG signal level (equivalent to 203 cd/m^2 with an HLG display normalized at 1,000 cd/m^2). 203cd/m^2 is 58% of the PQ signal level (PQ represents absolute light levels and thus reference white will be 203 cd/m^2 on a reference display).

In a single-master HDR/SDR live production, where the core production is HDR and there are dual program outputs of HDR and SDR, mid-gray is often be used to set camera exposure while reference white serves as a good anchor point around which the rest of the image can be subjectively shaded to preserve the desired detail in the SDR down-mapped images.

A 203 cd/m^2 reference white was chosen to provide sufficient headroom for “highlights” while allowing comfortable viewing in HDR/WCG and after conversion to SDR/WCG. This reference white level is within the range of the typical peak luminance level of flat panel consumer SDR televisions.

Some experts are considering a more flexible range for HDR reference white to maintain an acceptable SDR average picture level (APL) when down mapped. When selectively down mapping the HDR anchor point into SDR, the APL will change.

Down mapping HDR reference white to SDR at a lower signal level can allow the preservation of additional highlights in HDR but also a lower SDR APL. Conversely, down mapping HDR reference white into SDR at a higher signal level will cause more clipping with loss of detail and a higher SDR APL. HDR reference white serves as a visible anchor point for conversions to-and-from SDR and ultimately establishes the ability to “roundtrip” through the entire workflow in a “single-master workflow” (SDR->HDR->SDR).

HLG can support nearly an additional stop of dynamic range and a larger color volume by utilizing the commonly referred "super-white" signal range above 100% (between 10-bit code values 941 and 1019). EBU R 103 specifies using up to 105% of super-whites in HLG that can preserve an additional 391cd/m^2 on a normalized HLG display with a peak white of 1,000cd/m^2.

It is important to consider that some reference displays can reproduce HLG super-whites signals above 100%. . Since many HLG displays do not reproduce signals in the super-white region of HLG10, the details above the HLG 100% signal level will not be presented on those monitors.

Note that future displays may become available with higher peak brightness capability compared with those available today. Content that is expected to be of interest to consumers for many years to come may benefit from retaining an archive copy coded in "absolute values" of light (e.g., PQ) or original camera capture format (e.g., RAW, log) so that future grades of highlights and speculars with a higher luminance range can be produced and delivered to viewers using a format that supports these ranges.

7.5.2. Studio Video over IP

In 2017, the Society of Motion Picture and Television Engineers (SMPTE) published the first of its SMPTE ST 2110 suite of standards for "Professional Media Over Managed IP Networks", which exploits IP networking equipment for transport of video, audio, and metadata within television facilities, in addition to or in lieu of traditional serial digital interfaces (SDI). The SMPTE ST 2110 suite is built upon the Real time Transport Protocol (RTP) as described in RFC 3550 [42], with the suite providing sufficient constraints such that best-effort IP connections can succeed where historically, bandwidth and synchronization was provided by SDI connections, which offered assured point-to-point latencies.

The SMPTE ST 2110 suite currently comprises five documents: SMPTE ST 2110-10 [43] specifies the network interface requirements, overall system timing model, and session description protocols. SMPTE ST 2110-20 [44] specifies the format for uncompressed video essence, while SMPTE ST 2110-21 [45] identifies timing characteristics for these video streams and requirements for compliant senders and receivers. SMPTE ST 2110-30 [46] specifies the format for uncompressed digital audio. Lastly, SMPTE ST 2110-40 [47], published in 2018, specifies carriage of SMPTE ST 291-1 Ancillary Data (i.e., ANC messages). Each of the video, audio, and metadata essences described in SMPTE ST 2110-20, -30, and -40 are individual streams that are synchronized in accordance with SMPTE ST 2110-10.

8. Conversions Between Transfer Functions

A receiver may not be able to switch seamlessly between HDR transfer functions; therefore, it is recommended that only one transfer function be used for a given Real-time Program Service in Foundation Ultra HD. This section offers guidelines to service providers to convert HDR content from PQ10 to HLG10 or vice versa in order to maintain a single transfer function in the service. Equations and best-practices for these conversions can be found in ITU-R BT.2390 [6].

8.1. Conversion between HDR formats

8.1.1. PQ10 to HLG10

It is possible some providers may receive pre-recorded programs that are produced in PQ10 (a Foundation UHD format) but may require conversion to the HLG10 format for their specific distribution pipeline. Since PQ10 (which uses the PQ transfer function) can contain peaks as high as 10,000 cd/m2, BT.2408 [8] section 6.4 suggests that PQ10 content be pre-normalized to a nominal peak luminance of 1000 cd/m2 prior to conversion. Once PQ10 is normalized to 1000 cd/m2, it is possible to perform the required conversion to HLG10 a 3D LUT. Note, due to the conversion, there can be a small loss of color volume in HLG10 if signals are clipped to the HLG nominal signal range (x=1.0) as noted in Table 5 below, taken from BT.2408 Section 6.5.

Table 5 illustrates that there are color transforms can be designed to preserve a larger color volume in HLG10, by exploiting the 10-bit code values above nominal video range (code values 941-1019, often referred to as “super-whites”), thus minimizing the color volume differences between PQ and HLG with a 1,000 cd/m2 peak white. Because the narrow range 10-bit code value for PQ at a white level of 1000 cd/m2 is approximately 723 (for Y’of Y’Cb’Cr’), nominal video range usage is all that is necessary for PQ.

Note: some reference displays may not process HLG10 code values beyond narrow/nominal video range (e.g. 10-bit code values 941-1019). Results of conversions which include super whites are therefore not assured to reproduce the image highlights and speculars in the super-white region.[7]

[7] Note, this issue may also be evident in some consumer displays.

Table 5. BT.2408 Signal ranges and achievable color volume for PQ and HLG on a 1000 cd/m2 nominal peak luminance display

Color	BT. 2100 PQ Luminance[1] (cd/m^2)	BT. 2100 HLG Luminance (cd/m^2)	
Max value of color	x = 0.75	x = 1.0	Max non-linear signal, E', to match PQ luminance
{x,x,x} // Peak white	1,000.0	1,000.0	R'=G'=B'= 1.000
{x,0,0} // Peak red	262.7	201.1	R' = 1.041
{0,x,0} // Peak green	678.0	627.3	G' = 1.012
{0,0,x} // Peak blue	59.3	33.7	B' = 1.086

Table 5 Notes:

1. PQ nominal peak white at 1,000 cd/m^2
2. In practice some displays might not achieve a luminance higher than their nominal peak value.

Note that delivery of a Ultra HD HDR program, encoded with HEVC, it is possible to encode the PQ-related metadata (SMPTE ST 2086 [10], CLLI) in a SEI message even though HLG is signaled as the transfer function. At this time, that practice is not defined in any current standard, and thus it should be avoided or employed with caution due to unexpected behavior by the consumer displays.

8.1.2. HLG10 to PQ10

It is possible that Live programs may be produced in HLG10 (as defined in the Orange Book [O02] Foundation Ultra HD format); however, some service providers may favor PQ10 delivery making it necessary to convert HLG10 encoded content into PQ10. This conversion can be accomplished with a static or programmable 3D LUT prior to encoding. Note that this conversion involves a translation from a scene-referred relative signal (HLG10) to a display-referred absolute signal (PQ10). As such, a target peak luminance needs to be established as part of the LUT conversion function. ITU-R Report BT.2390 [6] recommends using a nominal peak luminance of 1,000 cd/m^2, so that value may be used; i.e., the nominal peak value of the HLG signal (100% HLG) with a system gamma of 1.2 would map to the PQ value of 1,000 cd/m^2. At this level, HLG10 will preserve overshoot levels above nominal video range (up to 1810 cd/m^2 white at 109%) where the HLG10 overall system gamma equals 1.2. However the EBU recommends a peak of 105% (preserving up to 1391 cd/m^2 white) (EBU r103, "Video signal tolerance in Digital Television Systems [106].).

8.2. Conversion of SDR to HDR: Up-Mapping

The Figure 3 and Figure 4 below represent the results of up-mapping display-light tests of a camera color chip chart (X-Rite Color Checker® 709-2014, See Figure 2) between BT.709 [2] SDR and BT.2100 [5] HLG using the Sony® HDRC-4000 HDR Production Converter Unit. Blue circles represent the original color after being passed through the SDR signal chain, i.e. the BT.709 OETF followed by BT.1886 [75] EOTF, with a 2x linear scaling applied to adapt the displayed luminance from the usual 100 cd/m^2 SDR nominal peak white to the HDR Reference White level of 203 cd/m^2. Green triangles represent the converted colors after applying the HLG EOTF to the converted signal to calculate display light. Figure 3 is a plot of scene-light conversion from the XRite ColorChecker ® test pattern and Figure 4 is a plot of display-light conversion. The chip chart provides a limited set of target colors and luminance levels. Other test patterns, such as the Sarnoff "Yellow Brick Road" (see Figure 6) with larger color and light ranges can provide additional color and luminance information on color accuracy. The 2D plots show the importance of using the appropriate converter for a particular application.

When attempting to match HLG, PQ or SDR images as they would be seen through a display (EOTF), a display-light conversion is the appropriate choice. Scene-light conversion is the appropriate choice if your goal is to match the look between different source image OETFs. The differences between scene-light and display-light conversion are greatest in an HLG system because of the way gamma is applied. Care should be taken to select the appropriate conversion method which matches your goals.

Figure 2. X-Rite Color Checker Chip Chart

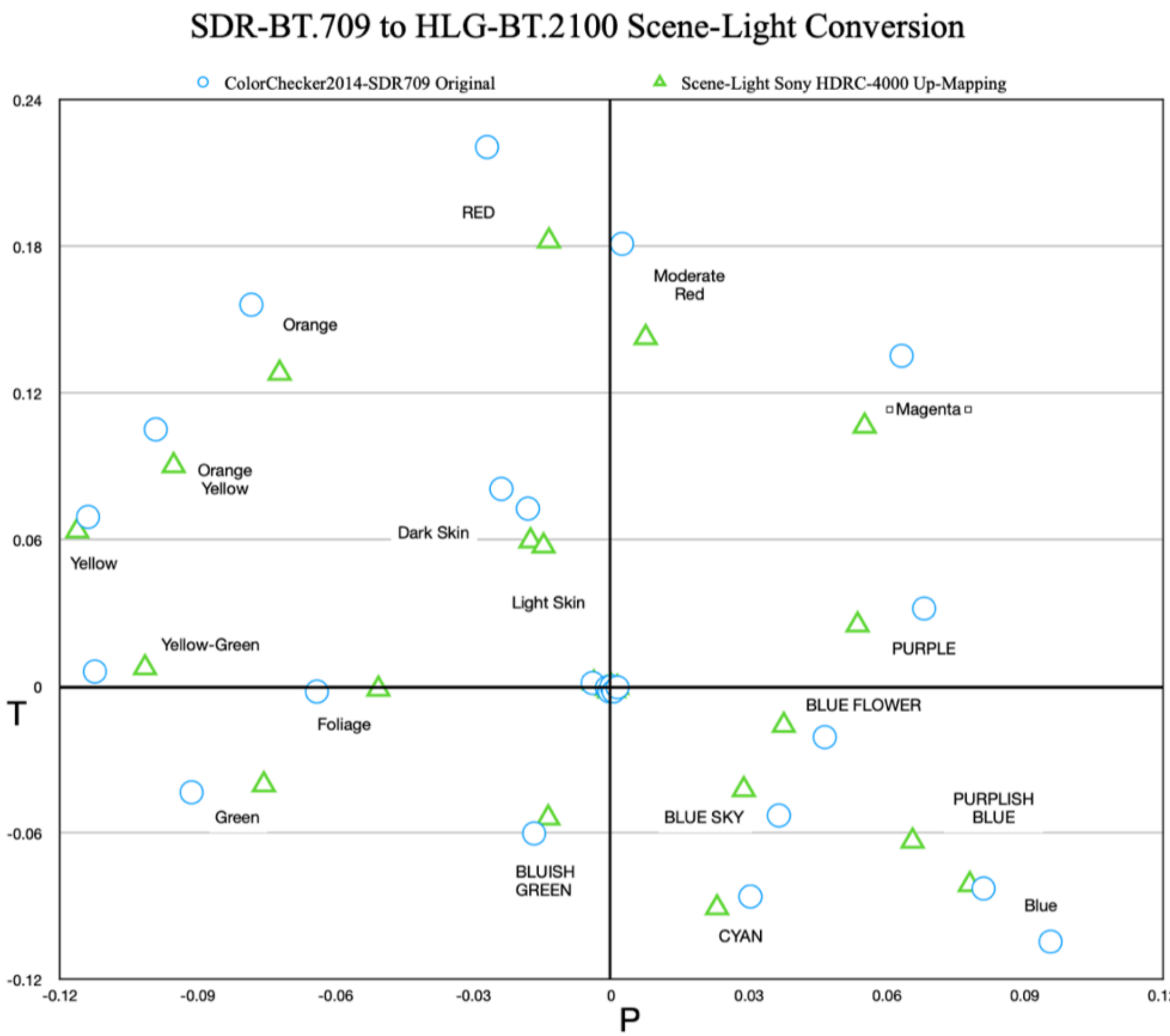


Figure 3. Display-light analysis of scene-light conversions from SDR to HLG, with errors

Note: Errors caused by selecting an inappropriate conversion type for the purpose. Errors of a similar size occur when using a display-light conversion to match cameras.

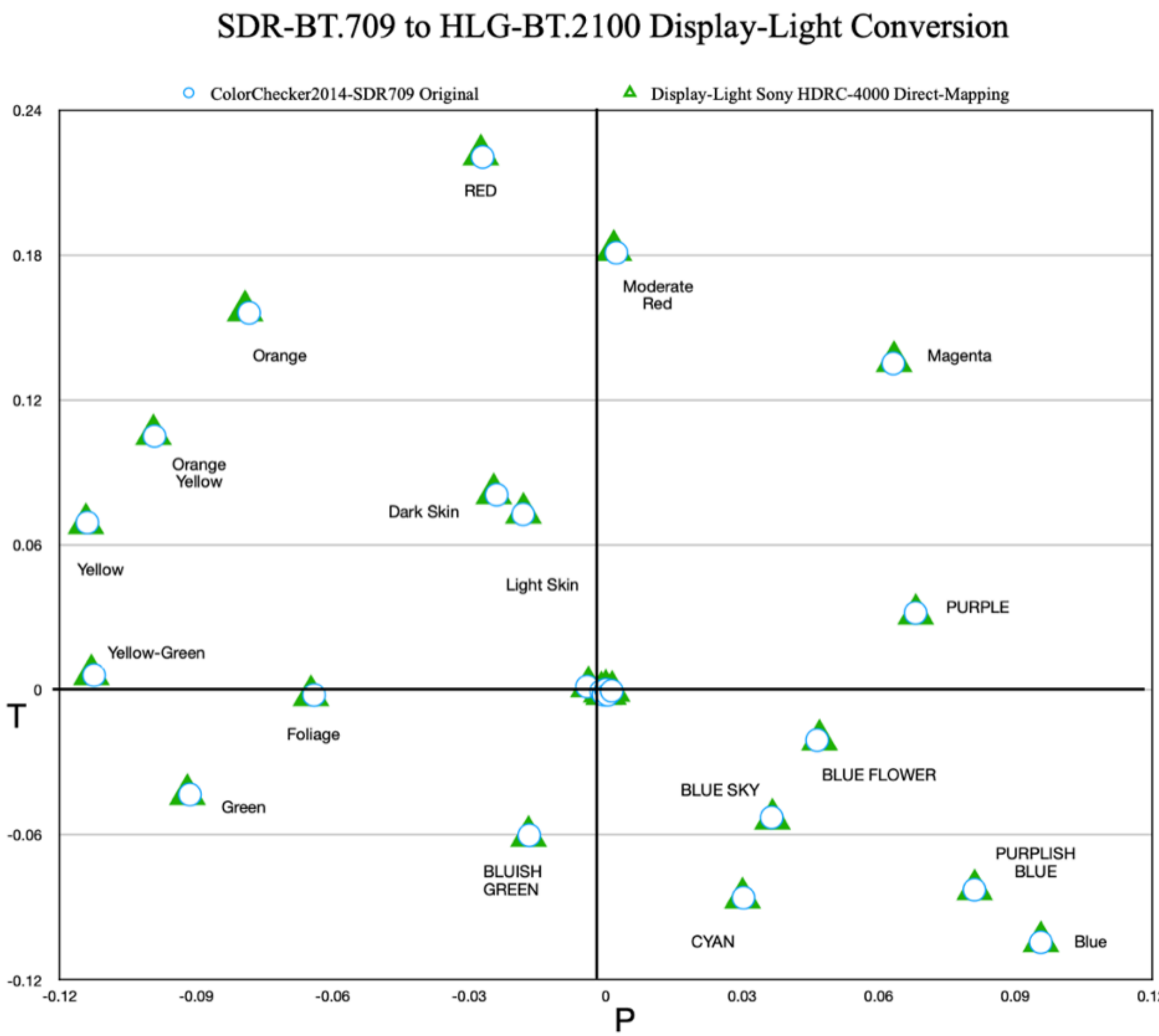


Figure 4. Display-light analysis of display-light conversions from SDR to HLG, good match

Note: showing that a good match can be achieved when an appropriate conversion is used.

Table 6. ∆E-ITP Values for the Up-mapping from SDR to HLG

ITP Difference Original vs LUT Converters	
Colorchecker CSM Suite HLG 709 colors in 2020 Container	Sony HDRC4000 Display Referred SDR->HLG ΔE-ITP
dark skin	0.6107
light skin	0.4276
blue sky	0.5727
foliage	0.4997
blue flower	0.4987
bluish green	0.3500
orange	0.9263
purplish blue	0.4047
moderate red	0.8893
purple	0.3227
yellow green	0.5079
orange yellow	0.5393
blue	0.4771
green	0.5339
red	0.7252
yellow	0.3631
magenta	0.7047
cyan	0.3159
white	0.2364
Gray 2	0.2208
Gray 3	0.2589
Gray 4	0.2547
Gray 5	0.2193
Gray 6	0.1750

8.3. Conversion HDR to SDR: Down-Mapping

Two common approaches have emerged from the HDR to SDR format conversion (see Figure 5 below for the conversion flow using each method):

- Approach A (Linear Downmapper): Attempts to maintain the subjective appearance of the HDR content when converted to SDR, and viewed on a reference display set to a peak white of 203 cd/m^2 with a gamma of 2.4 conforming to ITU BT.1886 [4]. This approach is described in Indigo Section 10.3 [I01] and has been adopted by NBCUniversal, Fox Sports, and others. This approach and others use approximately the same mid-tone mapping during up and down mapping. As an example, this tone mapping is similar to Sony's Camera setting for STD5[8] that has existed for quite a while (although they use scene light mapping of color) . The mapping uses a .5x linear scaling (see Figure 5) when converting from HDR to SDR. It mimics the same linear scaling in BT.1886 such that the HDR shadow's to mid-gray align with the original HDR luminance levels if the SDR display is set at a typical consumer SDR TV's luminance level (See paragraph below Fig 5*)*.

- Approach B (Gamma Adapted Downmapper): Attempts to maintain the subjective appearance of the HDR content when converted to SDR, and viewed in a reference BT.2035 environment (100 cd/m^2 with a display gamma of 2.4). Examples of this type of converter are described in ITU-R Report BT.2446 [128] and the approach has been adopted by Sky for its live Premier League coverage and the BBC in their licensed conversion LUTs.

- Some other approaches apply subjective looks in the LUT versus applying painting controls in the camera itself. Some examples could include:
  - Stretching of highlights during SDR to HDR conversion. To make this work with a roundtrip, an inverse function needs to be applied during the HDR-to-SDR conversion
  - Applying an additional "look" within the transform. Examples could be a manufacturer's such as Sony "Mild", or "Live" settings which add a more saturated look in the downmapper to compensate for the less saturated look of the HLG OOTF specifically.

[8] See Sony S-Cinetone Whitepaper, https://pro.sony/s3/2020/03/24095333/S-Cinetone-whitepaper_v2.pdf

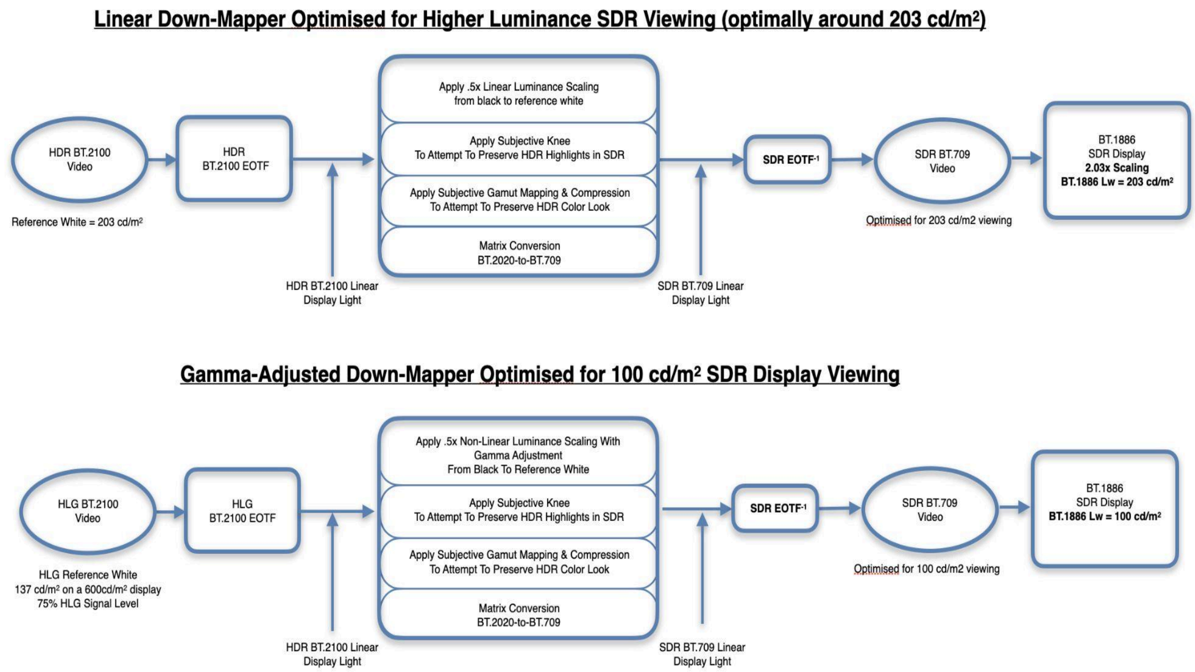


Figure 5: Two Conversion Workflows Through HDR to SDR Down-mapping

Note: Subjective conversions may use different perceptual spaces, gamut and luma compression or re-order processing

The 203 cd/m^2 shading approach provides the ability to place SDR and HDR displays or multi-viewer images in close proximity so that the content appears well matched between the HDR and SDR displays (minus the highlights that are possible in HDR). Because the adjustment of the SDR display uses the EOTF of SDR displays(BT.1886 [4] with L_w adjustment), it may improve the match between HDR and SDR when viewed on consumer displays in typical consumer viewing conditions. A recent survey of TV peak luminance levels[9] measured an average SDR peak luminance of 243 cd/m^2 for LCD screens and 218 cd/m^2 for OLED screens when in their cinema modes, within an overall range of 125-416 cd/m^2. However, if 203 cd/m^2 SDR content is viewed on a 100 cd/m^2 display, mid-tones will appear subjectively slightly darker. Conversely, when content is graded with a 100 cd/m^2 SDR display, content viewed on a consumer TV may show a better match to the HDR version especially in the lower to mid-tone region.

[9] https://nabanet.com/naba-nbcu-movielabs-tv-luminance-industry-survey/

Discussion is under-way with industry bodies to address the differences in SDR levels arising from the two approaches, which should be formalized for international program exchange. Techniques are being investigated that should ensure that signals derived using either approach appear the same on consumer TVs.

The human visual system response is very complicated and thus the evaluation of a color mapping will be subjective when deciding how to convert from a larger color volume to a smaller one. Some suggest that it may be preferable to allow a hue shift in order to preserve saturation. The specific design of 3D LUTs make different choices regarding gamut/luma compression and when to allow hue shifts to preserve saturation. It is important to test the implementation of every conversion for objective accuracy as well as subjective perceptual choices.

Figure 7, Figure 8 and Figure 9 use 533 sample points from the Sarnoff® Yellow Brick Road HLG Test Pattern of the "Color Space & Monitor (CSM) Test Pattern Suite" (max luminance constrained to 2300 cd/m^2) (Figure 6) to compare three different HLG-BT.2100 [5] to SDR-BT.709 [2] conversion methods. Each plot displays the results in ITP space (see Section 7.4) with plots of the color components (T/P). The T/P plots demonstrate different subjective choices in color mapping, by objectively visualizing hue shifts and saturation against reference plots. It demonstrates the importance of assessing the results of down-mapping using objective measurements in addition to subjective observations on reference displays to fully evaluate how well the down-mapping methods are able to preserve the original intent.

Figure 7, Figure 8 and Figure 9 show reference paths towards the primaries of BT.2020 [3] (in blue), BT.709 (in green), and the conversion (in red). In addition, an outline of the extent of the color spaces in T/P space are shown (dotted lines).

Figure 7: Method 1 is an example of a conversion from HLG to SDR that doesn't fully saturate some colors based on specific color mapping choices. Figure 8: Method 2 is an example of a conversion from HLG to SDR, where BT.2020 colors are immediately converted (hue shifted) so that they follow BT.709 primaries' paths. In Figure 9: Method 3 is a conversion that follows BT.2020 primaries while "in-gamut" for BT.709 and then performs a slight hue shift to align with the BT.709 primaries to achieve optimal saturation and preserve original artistic intent as close as possible within the confines of SDR.

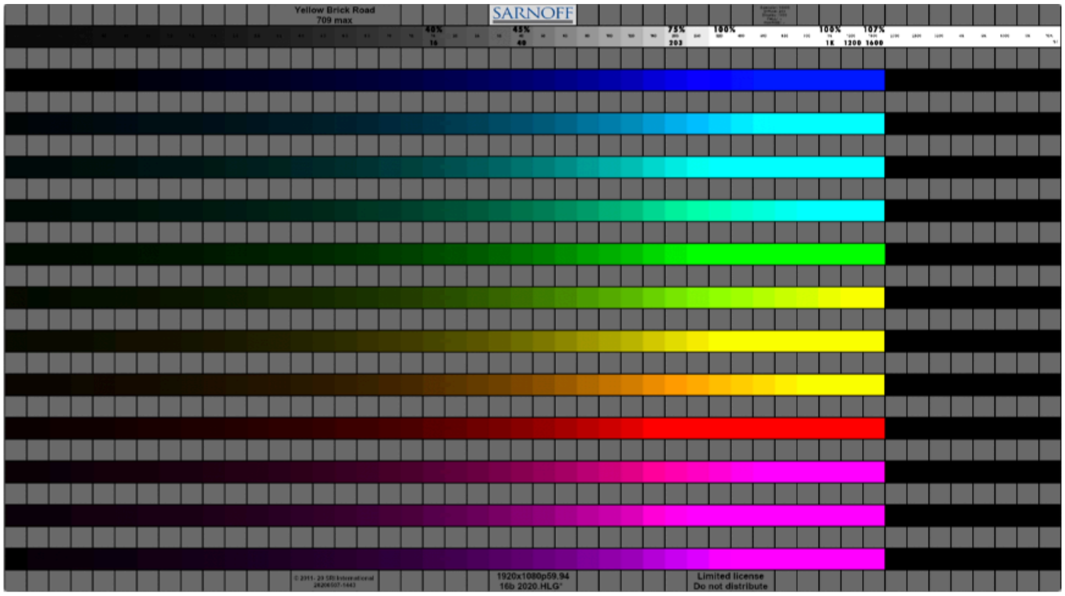


Figure 6. Sarnoff "Yellow Brick Road"© Test Chart (SDR)
(courtesy of SRI International)

Notes: Maximum luminance constrained to 2300 cd/m^2 (1810 cd/m^2 peak for HLG at 109%)

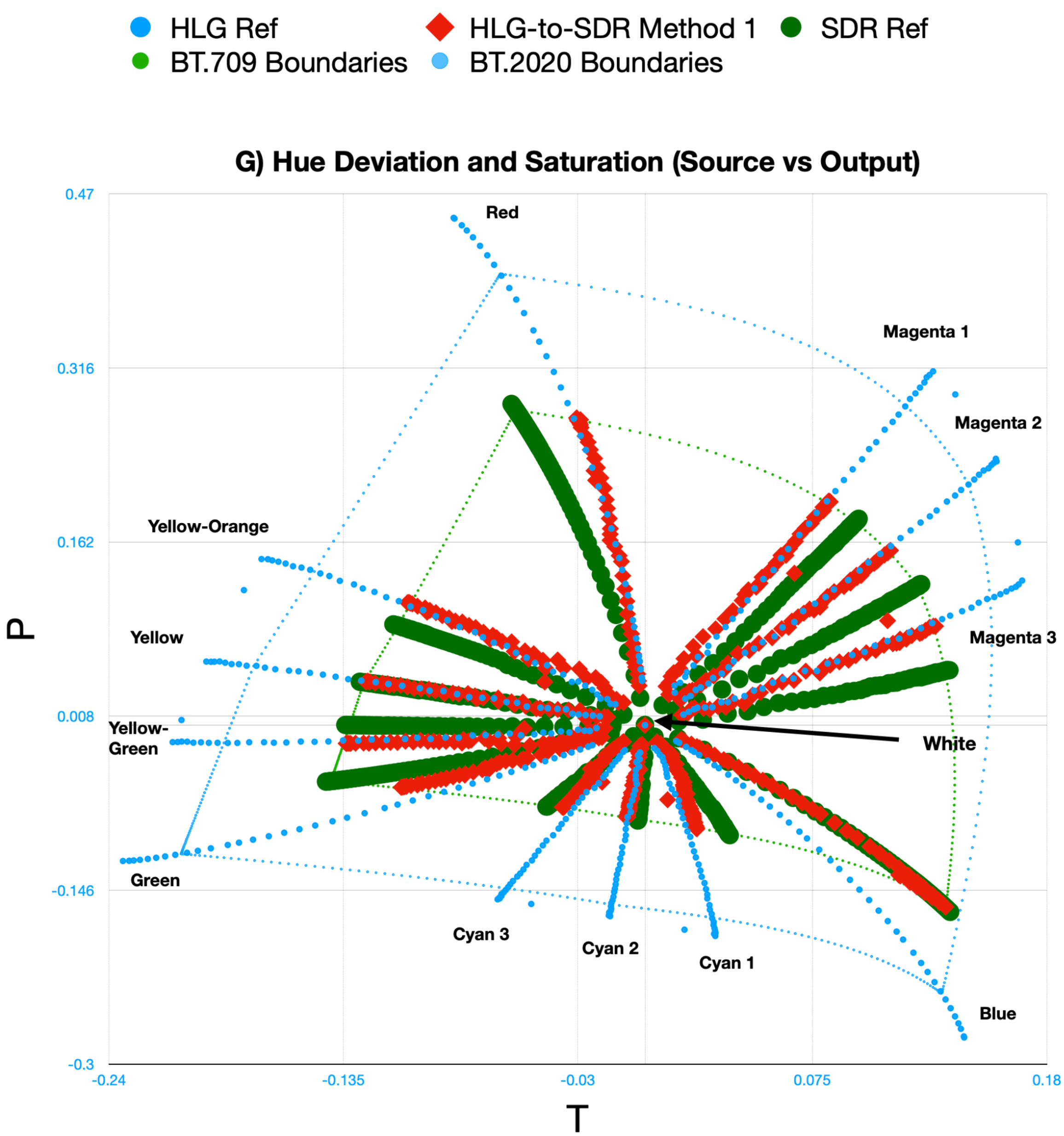


Figure 7. METHOD 1 - HLG to SDR Conversion

Note: Follows BT.709 primaries for blue which saturates BT.709 blue completely, Green is slightly hue shifted between BT.2020 and BT.709 and doesn't fully saturate; Red follows BT.2020 primaries and dead-ends on the edge of BT.2020

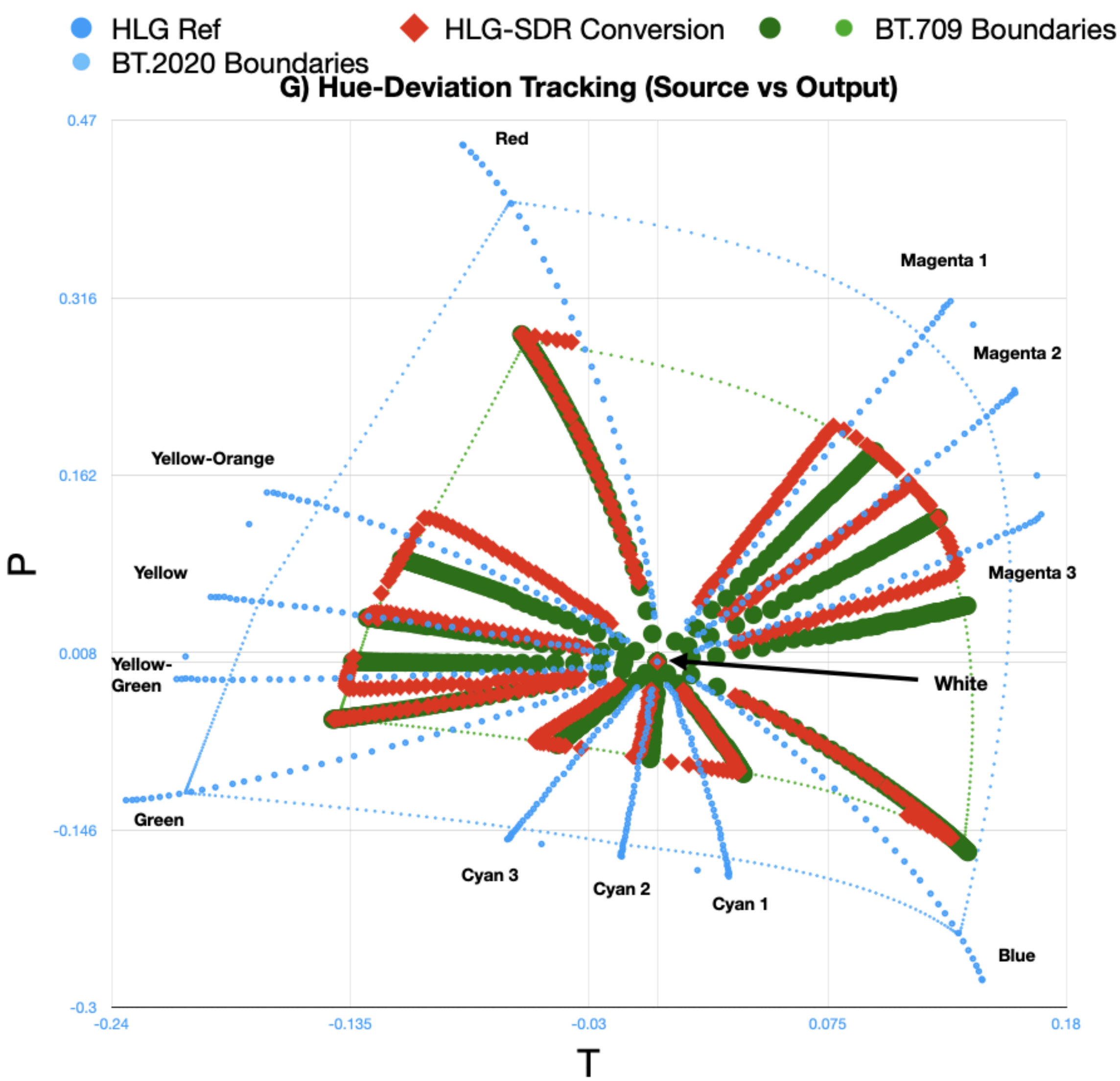


Figure 8. METHOD 2 - HLG to SDR Conversion

Note: Immediately follows BT.709 primaries (hue shifts from BT.2020) and saturates BT.709 colors completely

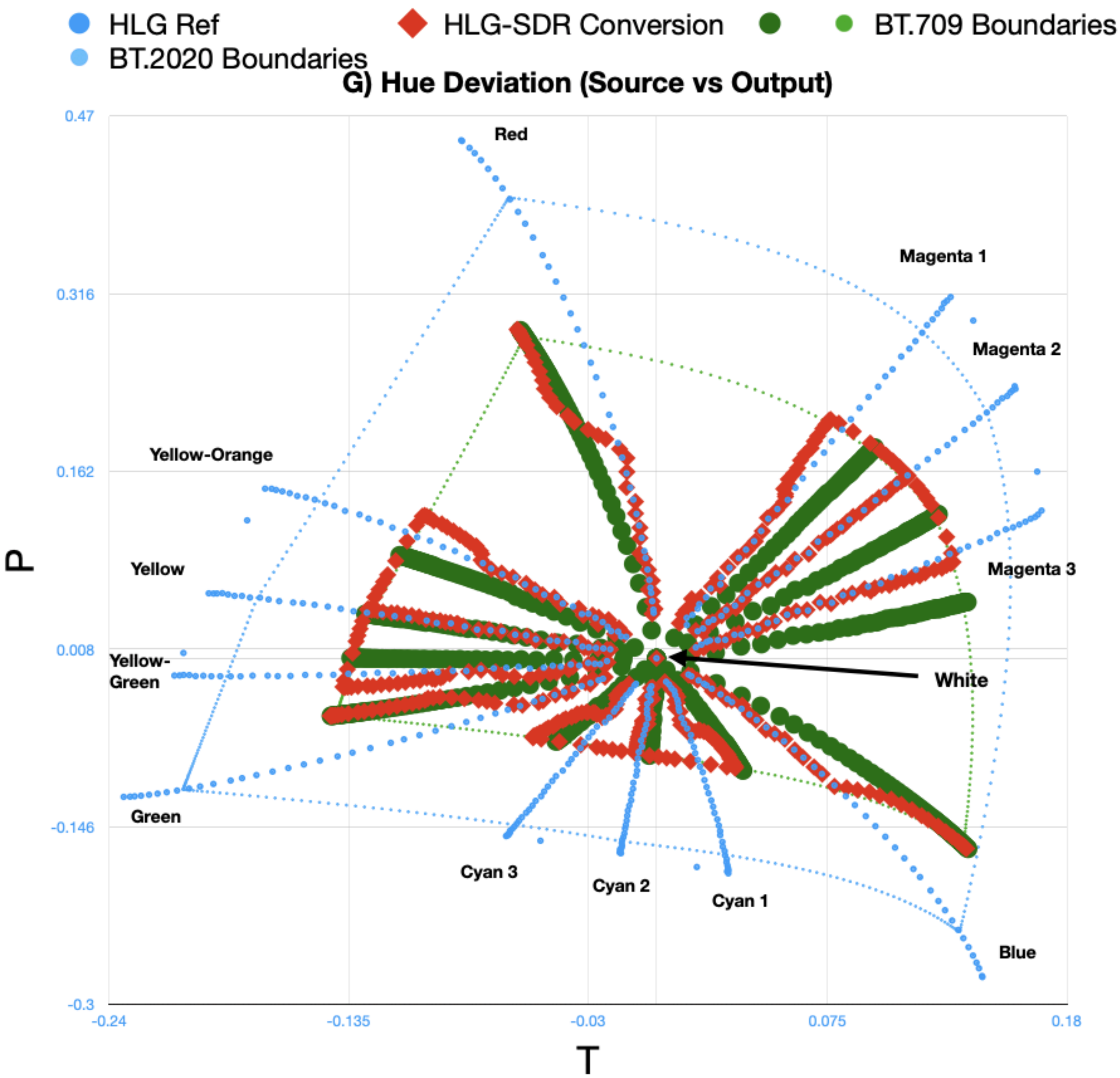


Figure 9. METHOD 3 - HLG to SDR Conversion

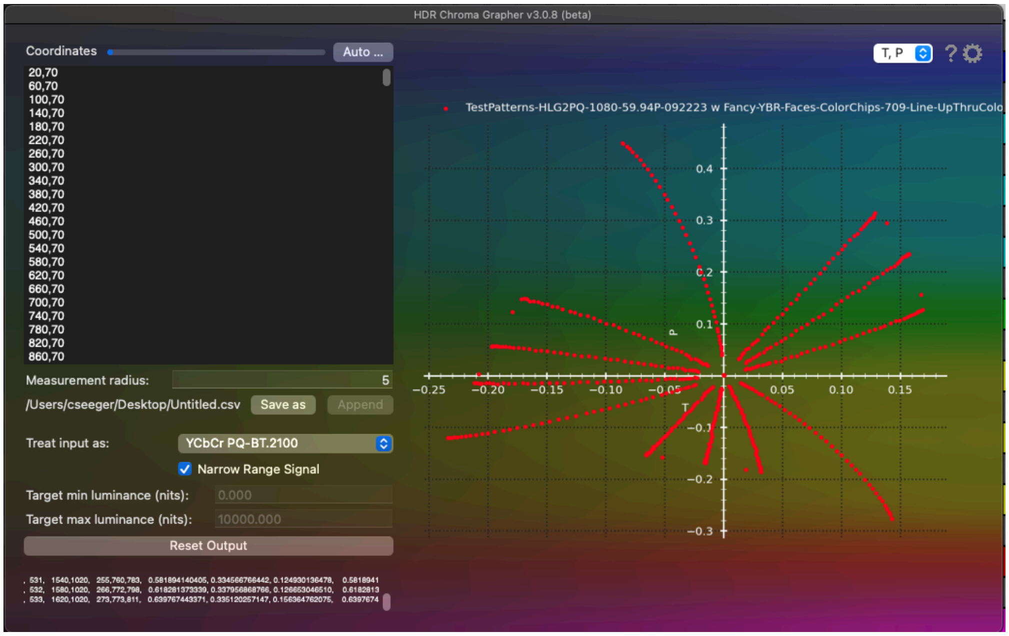
Note: Preserves original BT.2020 colors in shadows; full saturation of BT.709 in higher luminance colors

8.4. Objective Color Metrics

Traditionally video shaders/graders relied on waveform monitors and vector scopes along with subjective viewing on reference displays to establish the appropriate conversion parameters between one image format and another. Most often such conversions are provided through a "3D look-up table" (LUT). The ITU has established a new recommendation BT.2124 [107], "Objective metric for the assessment of the potential visibility of color differences in television", which provides an objective measurement of the color differences between the original image color and the converted image color. BT.2124 [107] uses a ΔE_{ITP} methodology to measure color accuracy. A ΔE_{ITP} value of 1 indicates a Just-Noticeable-Difference (JND) between the original and the converted image color, assuming the most sensitive state of adaptation. The measure is based on IC_TC_P, specified in ITU-R Recommendation BT.2100 [5] and illustrated in Figure 10. ∆E-ITP uses a derivative of IC_TC_P color representation called ITP, which uses PQ as the transfer function and BT.2020 [3] as the color space. ITP provides a large container for normalizing both the light and color.

In order make ITP measurements a video player known as "Vooya" it had a plug-in architecture that could be leveraged to build a file-based tool for normalizing conventional video into ITP. The plug-in:

- Uses a csv text file with pixel locations of the areas to be measured
- Has a definable filter radius to gather an average value based on the center pixel identified in the csv file list
- Accepts SDR-BT.709 [2], HLG or PQ YCbCr 10-bit code video sources
- For SDR:
    - Can use gamma 2.2 or 2.4 sources
    - Can be assigned a peak luminance value for the normalization (since SDR is scene referred) into ITP.
- For HLG can be assigned a peak luminance value for normalization into ITP.
- For PQ, has not parameterization settings since it is an absolute transfer function
- Vooya has new functionality including:
    - Absolute chromaticity measurements using u'v' from CIE1976
    - Radius Filter to average the values of multiple pixels around a center point.
    - Maximum error results based on the radius filter measurement (especially useful in lossy compressed workflows that are common in broadcasting).
- Below is the interface for the Vooya plug-in. For further detail, please read the eBook.



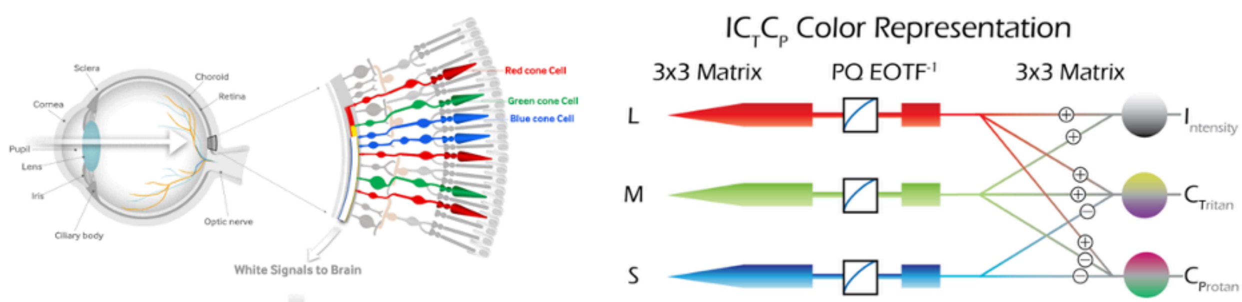


Figure 10. IC_TC_P and the Relationship to the HVS [10]

Using ΔE_{ITP}, we can compare a source vs processed signals to verify no unintended alterations of color have occurred. Alteration could occur because of poor choices in mathematical precision, using the wrong matrix between YC_BC_R and RGB (and a number of other possible errors or quality choices). For instance, lower bit-rate codecs will have additional color accuracy errors in many cases.

[10] Image credit: Dolby White Paper-What is ICtCp, (figure 1). https://professional.dolby.com/siteassets/pdfs/ictcp_dolbywhitepaper_v071.pdf

ΔE_{ITP} uses a very simple scale where any number above one ("1") represents a noticeable difference to the human visual system at its most sensitive state of adaptation. This makes it much easier to quantify perceptible issues from comparative plots.

The formula for calculating ΔE_{ITP} is given by equation 8.4(a). It is possible to compute further metrics derived from ITP to examine other aspects of the color difference. ΔTP_{ITP} (equation 8.4(b)) is the color difference assuming equal intensity. ΔI_{ITP} (equation 8.4(c)) is the intensity difference, assuming equal chroma. ΔH_{ITP} (equation 8.4(d)), known as the hue difference, is an approximation to the distance between two colors on the T/P plane if it is assumed they have the same chroma magnitude. The approximation is most accurate when the hues are very similar.

These equations have been used for the measurements[11] that are the basis of the plots in shown in Figure 11 through Figure 15:

a) ΔE_{ITP} Color Difference (Source vs Output).

$$\Delta E_{\mathrm{ITP}} = 720\sqrt{(S_i - 0_i)^2 + (S_t - O_t)^2 + (S_p - O_p)^2} \qquad \text{eq. 8.4(a)}$$

ΔTP_{ITP} Color Difference - Chroma ONLY (Source vs Output)

$$\Delta TP_{\mathrm{ITP}} = 720\sqrt{(S_t - O_t)^2 + (S_p - O_p)^2} \qquad \text{eq. 8.4(b)}$$

c) ΔI_{ITP} Color Difference - Intensity ONLY (Source vs Output)

$$\Delta I_{ITP} = 720\sqrt{(S_i - 0_i)^2} \qquad \text{eq. 8.4(c)}$$

d) ΔH_{ITP} Hue Difference (Constant Intensity and Chroma)

[11] A Spreadsheets for plotting and measurement are available via this link: https://github.com/digitaltvguy/NBCU-UHD-HDR-SDR-Resources-Table-of-Contents

$$\Delta H_{ITP} = 720\sqrt{(S_t - O_t)^2 + (S_p - O_p)^2 - \{\sqrt{S_t^{\,2} + S_p^{\,2}} - \sqrt{O_t^{\,2} + O_p^{\,2}}\}^2}$$

eq.8.4(d)

Below are the examples with results of video passthrough of a broadcast legalizer in a simple frame sync that was being used to realign audio. Different quadrants of a 2160P signal were apparently processed incorrectly which caused a change in color accuracy. The plots below show ΔE_{ITP}, ΔTP_{ITP}, ΔI_{ITP}, I_{ITP}, ΔH_{ITP} for tracking overall color, Chroma, Intensity, Constant Intensity/Chroma changes and finally identifies hue-shifts using T/P color components from ITP.

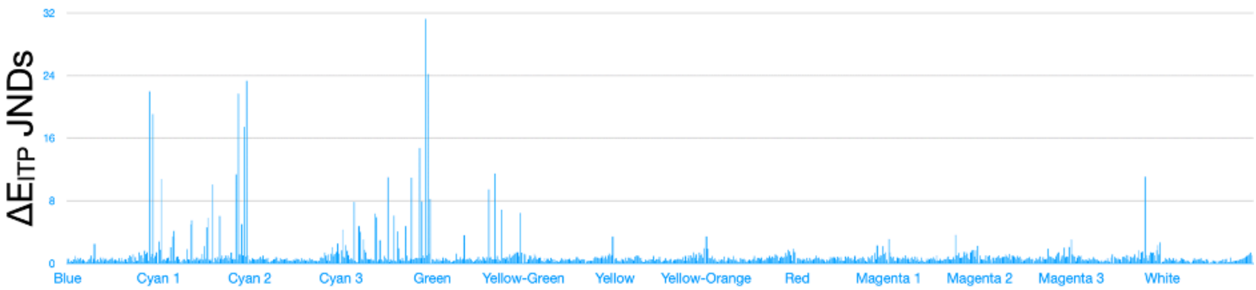


Figure 11. ΔE_{ITP} Color Difference (Source vs Output)

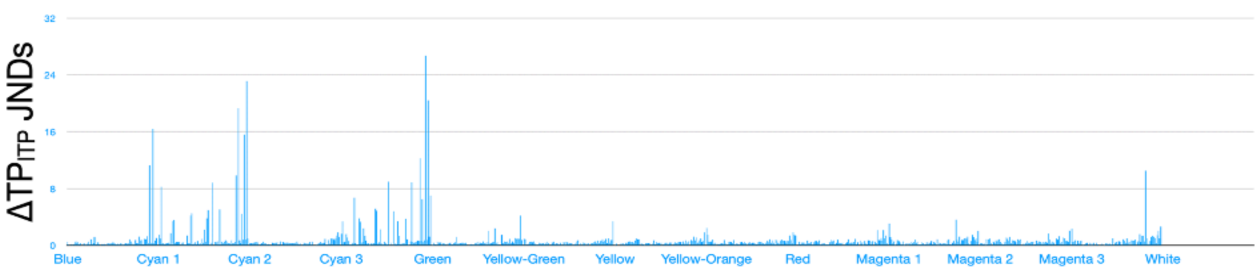


Figure 12. ΔTP_{ITP} Color Difference - Chroma ONLY (Source vs Output)

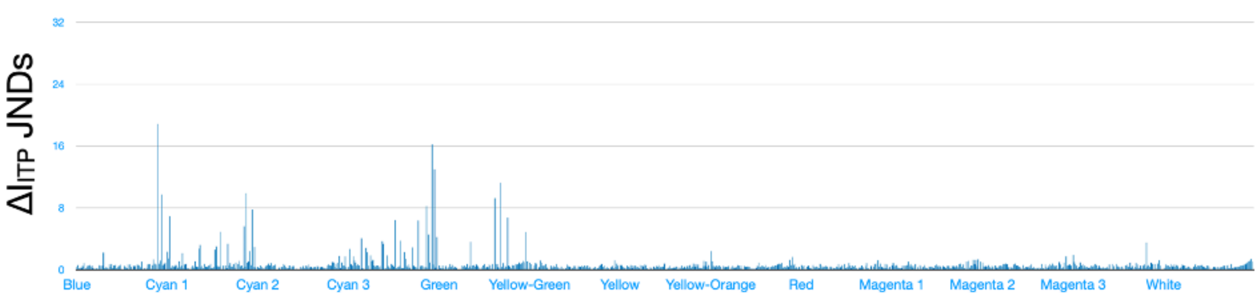


Figure 13. ΔI_{ITP} Color Difference - Intensity ONLY (Source vs Output)

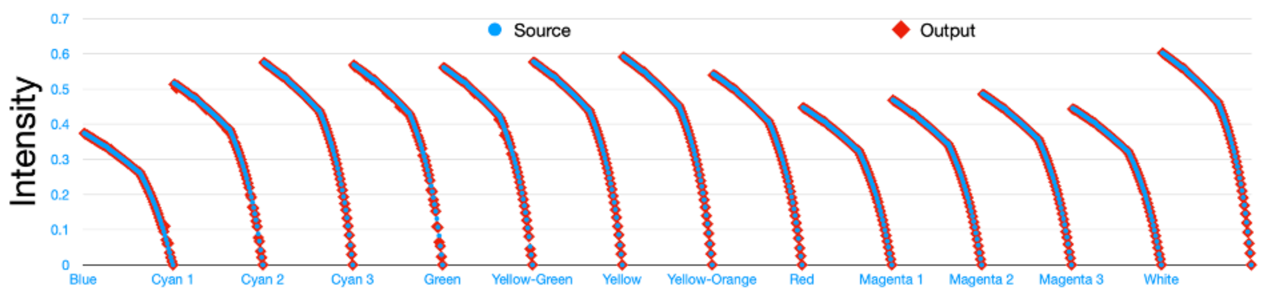


Figure 14. Intensity Tracking (Source vs Output)

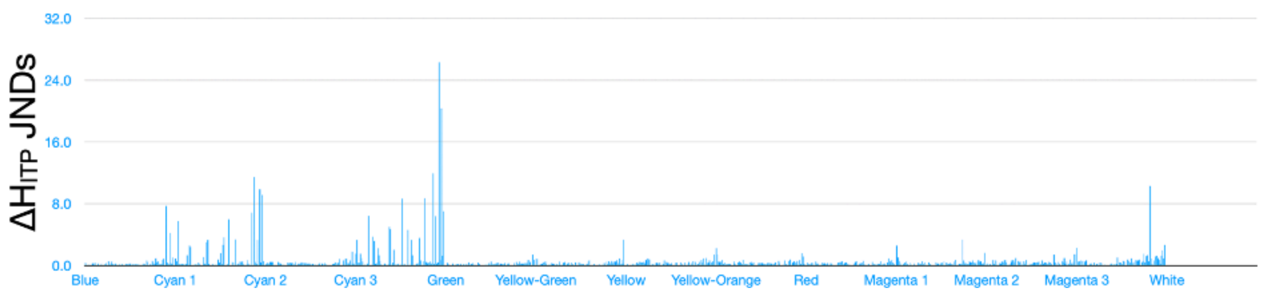
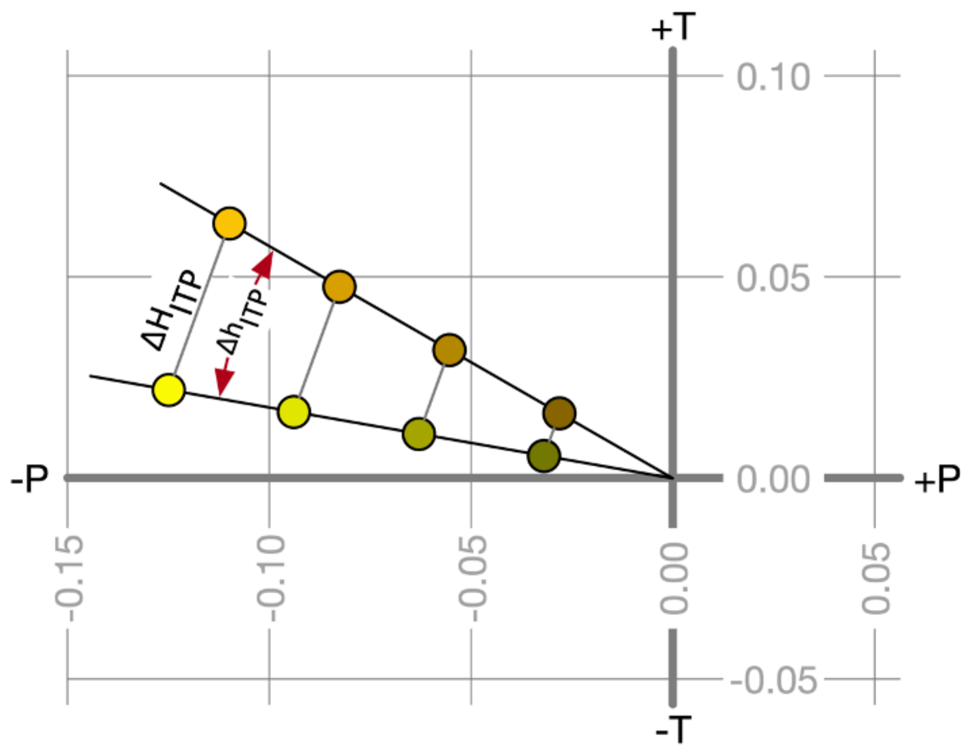


Figure 15. ΔH_{ITP} Hue Difference (Constant Intensity and Chroma)

For measuring the visibility of hue shifts, we use plots of the chroma components T/P within ITP (see Figure 16 and Figure 17). Care should be taken, however, when assessing images displayed at different brightness, e.g an HDR original image and its down-mapped SDR version or an HLG image on different brightness displays, as a change in luminance will affect the T/P components. The effect can be seen in the source points shown in Figure 17 from the Sarnoff "Yellow Brick Road" Chart, which contains rows of colors with the same hue and saturation across a luminance staircase. In these cases, u' v' of CIE Colorimetery-Pt 5: 1976 [116] can be used to determine absolute chromaticity (see Figure 18). T/P plots can provide a useful indication of hue shifts, and also allow us to understand and evaluate color conversion strategies.



The samples in each color difference pair have the same chroma, for being the same distance from the origin. The hue angle difference $\Delta \mathbf{h}_{ITP}$, in this illustration from yellow to yellow-orange, is constant for these pairs, but the hue difference $\Delta \mathbf{H}_{ITP}$ increases proportionally to a pair's chroma.

Figure 16. Hue Difference (ΔH_{ITP}) and Hue Angle (Δh_{ITP})

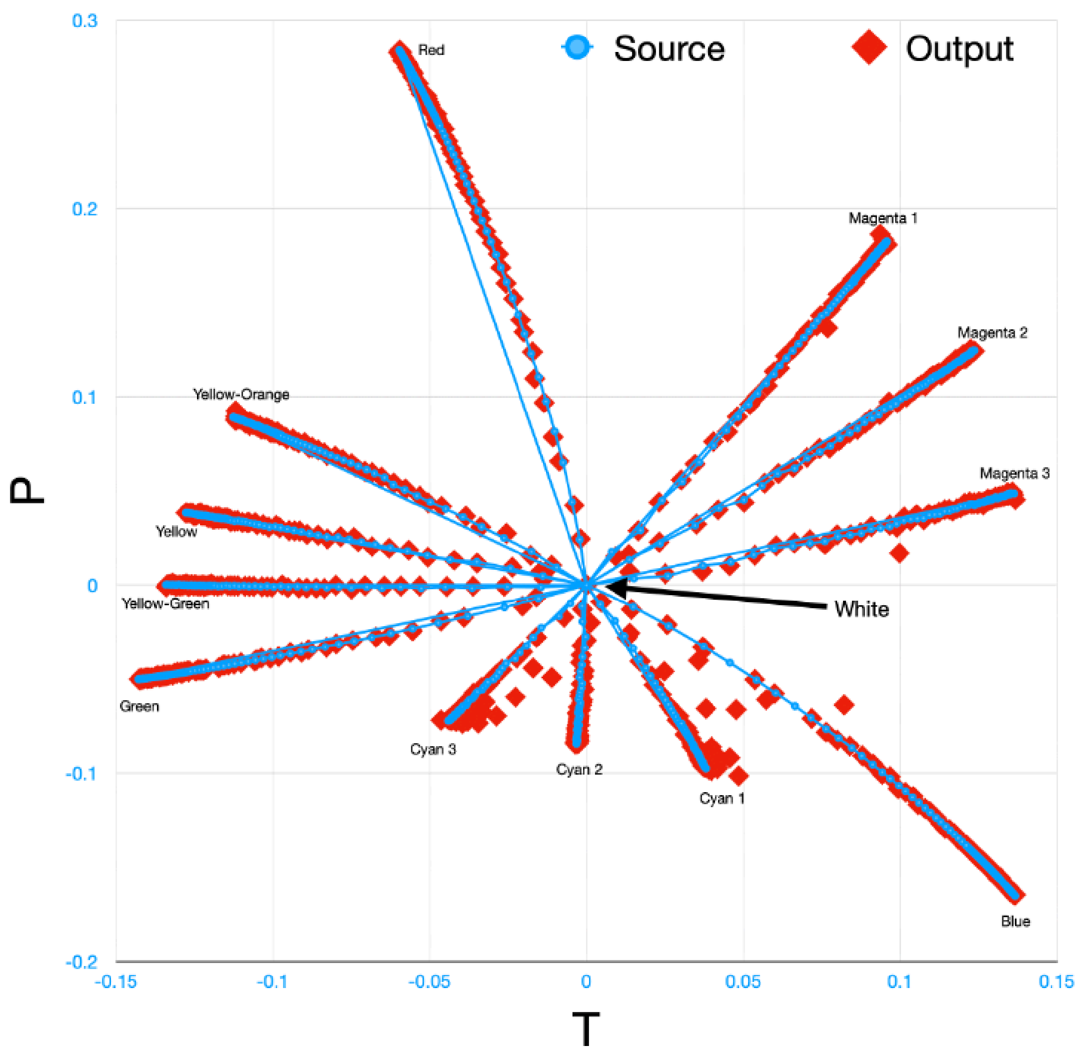


Figure 17. Example T/P Plot with Major Hue Shifts.

Note: PQ pass-through using a pseudo BT.2020 color space transform on the Sarnoff "Yellow Brick Road" Chart

Figure 17 shows how a two-dimensional plot of the T and P components can be used to assess color accuracy.

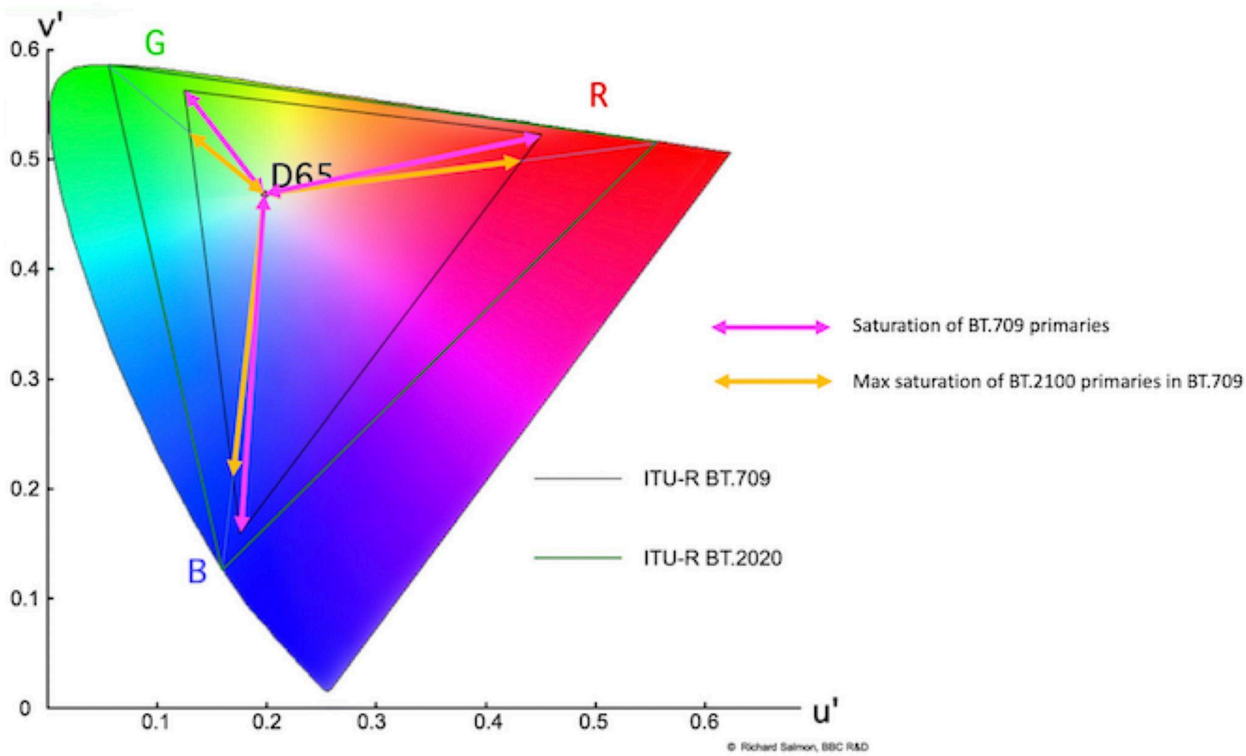


Figure 18. CIE 1976 u’ v’ color space with Gamut Boundaries, BT 2020 and BT 709

Note: color primary paths that may introduce gamut clipping during color conversion from BT.2100 to BT.709

Figure 18 shows the BT.709 [2] and BT.2020 [3] color gamuts in CIE 1976 u’v’ color space, with arrows to indicate the maximum saturation of both sets of primaries within the BT.709 gamut. The figure illustrates one of the great challenges of BT.2020 to BT.709 color space conversion. In order to bring colors that are close to the BT.2020 primaries into the BT.709 gamut, some gamut mapping and compression is clearly needed. The maximum saturation possible in either gamut is different for different hues, with peaks at the three primaries. Further, lines of constant hue (not shown in the figure) are generally curved in u’v’ space.

One option for a converter is to desaturate towards the white point. However, since the two sets of primaries have different hues, this results in the BT.2020 primaries suffering a significant loss of saturation. There will also be a slight hue shift, since the de-saturation does not follow lines of constant hue. An alternative would be to desaturate along lines of constant hue. However, these lines are not parameterised, so such a converter is hard to implement precisely, resulting in a need for significant color tuning. To minimize de-saturation of the BT.2020 primaries, another alternative would be to shift the hue of highly saturated colors towards the BT.709 primaries, introducing a larger hue shift but only for those colors that cannot be represented within the BT.709 boundary.

The difficulty is compounded when luminance is also considered. The size and shape of the gamut boundary changes with luminance, so a small luminance change is a further possibility for bringing low or high luminance colors into the BT.709 volume.

The ITP color representation addresses some of the hue-linearity problems associated with u'v'. The T/P plots in this section are one way to examine hue shifts introduced by a converter.

8.5 Tools and Resources for Objective Color Metric Measurement

Tools and techniques have been developed by Dolby, NBCUniversal, Sarnoff Labs with some optimizations provided by Interdigital. Their goal is to make objective color metric measurement easier for the broadcast engineering professional. The mathematics were provided by Dolby for inclusion into this general-use toolset (methodology as defined in ITU-R BT.2124 [107]). NBCUniversal funded and oversaw the project and built reference plotting spreadsheets for visualization. Sarnoff Labs (SRI) and NBCUniversal developed specialized test patterns.

Below are Github links the links which resources, eBooks, Tools and spreadsheets for plotting:

Reference training documents can be found here:
https://github.com/digitaltvguy/UHD-HDR-SDR-Objective-Color-Metrics-For-Broadcasting

https://github.com/digitaltvguy/NBCUniversal-UHD-HDR-SDR-Single-Master-Production-Workflow-Recommendation-LUTs

Tools can be found here (Vooya and HDR Chroma Grapher):

https://www.offminor.de/downloads.html

9. Real-Time Program Service Assembly

Real-time Program Services consist of a linear, pre-scheduled stream of content that is assembled in real-time for distribution to consumers such as a broadcast television channel, a cable network, etc. Real-time Program Services consist of Live and/or Pre-recorded content and may also include graphic overlays, such as station logos, emergency text crawls, etc.

A Real-time Program Service may be assembled at one or more locations. At Production, graphics and slow-motion replays may be added to Live content. At the Broadcast Center (see Green Book-Section 7 [G01] , Figure 1 and Table 1) and service provider, interstitials, logos, news or emergency alert crawls, etc. may be combined to produce the final product delivered to the consumer. It is also possible that assembly can occur at the consumer device, such as for targeted advertisements and rendering closed captions.

9.1. Maintaining Dynamic Range and System Colorimetry Parameters

Different dynamic range and system colorimetry parameters should not be mixed in a Real-time Program Service. For example, service operators should not shift between HLG10, PQ10 or SDR/BT.709 [2]. Decoders require time to adjust to different encoded data settings – as much as 2 seconds or more has been observed by Ultra HD Forum members – causing a poor consumer experience. OTT providers offering ABR streams must also ensure that the adaptive bitrate streams are all of the same transfer curve and system colorimetry[12]. Similar to the linear progression of a program through time, the progression of rendering successive levels of ABR streams requires quick adjustments on the part of decoders. If a Real-time Program Service must contain a switch point between dynamic range and system colorimetry, it is recommended that such switches be performed overnight or in a maintenance window and black frames be inserted at switch points to allow decoders and viewers to adjust to the new content characteristics.

It is possible to direct map SDR/BT.709 content into HLG10 or PQ10, to up-map SDR/BT.709 content to HLG10 or PQ10 and vice versa, and to convert content from PQ10 to HLG10 or vice versa (see ITU-R report BT.2408 [8]). The following subsections offer guidelines for normalizing content in the headend for this purpose. See also Section 11.3 in the Violet Book [V02] for conversion possibilities in consumer equipment for backward compatibility.

[12] See Guidelines for Implementation: DASH-IF Interoperability Points [16] Section 6.2.5, Adaption Set Constraints.

9.2. Conversions from SDR/BT.709 to PQ10 and HLG10

In multiple areas of the production chain, there may exist a requirement to utilize sources that consist of different formats (SDR/BT.709 [2], PQ10 [O03], HLG10 [O02])" when constructing a Real-time Program Service (see Figure 19). Mixing of this kind may take many forms as it does today in SDR only environments (using different colorimetry like BT.709 and BT.2020 [3]). To make it possible to combine such elements, SDR/BT.709 sources must be converted into the PQ10/HLG10 formats with respect to both an industry compliant transfer function (e.g., HLG or PQ per BT.2100 [5]) and system colorimetry (i.e., BT.709 primaries "mapped" into the BT.2020 [3] container). Such conversions can utilize:

- Direct Mapping: SDR/BT.709 content is decoded and repackaged into PQ10 or HLG10 containers, but while changing the system colorimetry, remapping does not change the color gamut or the dynamic range of the content; the content is simply mapped across to the equivalent color and brightness values.
    - When remapping SDR to HDR for PQ and HLG, the level of reference white should be considered. Diffuse white in SDR content is typically about 642.5mV (92% of SDR Peak white level. 700mV) leaving very little headroom for specular highlights. In the case of HLG, the BBC and NHK recommend the reference level for HDR graphics white (aka "reference white") be set to 75% (equivalent to 203 cd/m^2 on a 1,000 cd/m^2 reference display). This was chosen as it leaves sufficient headroom (25% of the signal range) for "specular highlights" and allows comfortable viewing when HLG content is shown on HDR/WCG and SDR/WCG displays.
- Up-Mapping: SDR/BT.709 is decoded and then enhanced/modified to emulate PQ10/HLG10 and repackaged as above. Blind (not supervised) up-mapping can lead to undesirable results depending on the technology so care should be used when converting SDR to HDR. Conversion algorithms or equipment should be carefully evaluated prior to their use. ITU-R BT.2446 [128] describes techniques for SDR to HDR conversion.

Each method has particular advantages and disadvantages, and one or the other may be preferable under different circumstances. If the service provider intends to simulcast the Foundation Ultra HD service in SDR/BT.709 for backward compatibility, then remapping may be preferable, because content that was originally SDR/BT.709 will remain exactly as intended in the legacy version of the service. Conversely, the service provider may prefer that all the segments of the Real-time Program Service look as uniform as possible to ensure a consistent consumer experience, and thus up-mapping may be appropriate. For example, up-mapping may be preferred for Live production mixing SDR/BT.709 and PQ10/HLG10 cameras, high-quality SDR content, and legacy content like advertisements.

Under some circumstances, (e.g. viewing in a darkened room) HDR displays and content can cause discomfort if consideration is not given to the viewer's vision adapting to the average light level on the screen at any given moment. For instance, if a feature program has a low average light level such as a night scene and the picture is abruptly cut to a high average luminance scene in an interstitial, the viewer may experience discomfort similar to that experienced with changing light conditions in nature. When advertisements are inserted into content, consideration should be given with respect to transitions from dark to light.

The described SDR to HDR conversions typically will be performed by dedicated devices using a combination of 1D and 3D LUTs or other appropriate algorithms or technology. Devices of this type may be used for both SDR/BT.709 to PQ10/HLG10 or vice versa as the production and the capability of the equipment in use requires. Typically, the conversion LUTs should be matching pairs that provide a subjective round-trip when SDR is up-mapped to HDR and then down-mapped back to SDR so as to retain the desired artistic intent.

A real-time dedicated conversion device is essential for some use cases, which may be encountered throughout the production chain, such as:

- Mix of SDR/BT.709 and PQ10/HLG10 live sources
- Broadcasting live events (typically sports) in PQ10/HLG10 may require a relatively high number of cameras and it is probable that all these cameras will not be HDR-capable. In that situation, SDR/BT.709 cameras can be utilized if the conversion process is implemented either at the output of the SDR/BT.709 camera or at the input of the mixer.
- SDR/BT.709 interstitials in a PQ10/HLG10 program
- With SDR/BT.709 interstitials, the interstitial content will likely come from a playout server. In this case the conversion process has to be implemented either at ingest, at the output of the playout server, or at the input of the mixer.
- Use of SDR/BT.709 content
- Extensive libraries of SDR/BT.709 content may be used, for example a live sports match production that includes archive footage from previous matches; such material needs to be converted to PQ10/HLG10 (using a near real-time file-to-file converter) before entering the video pipeline

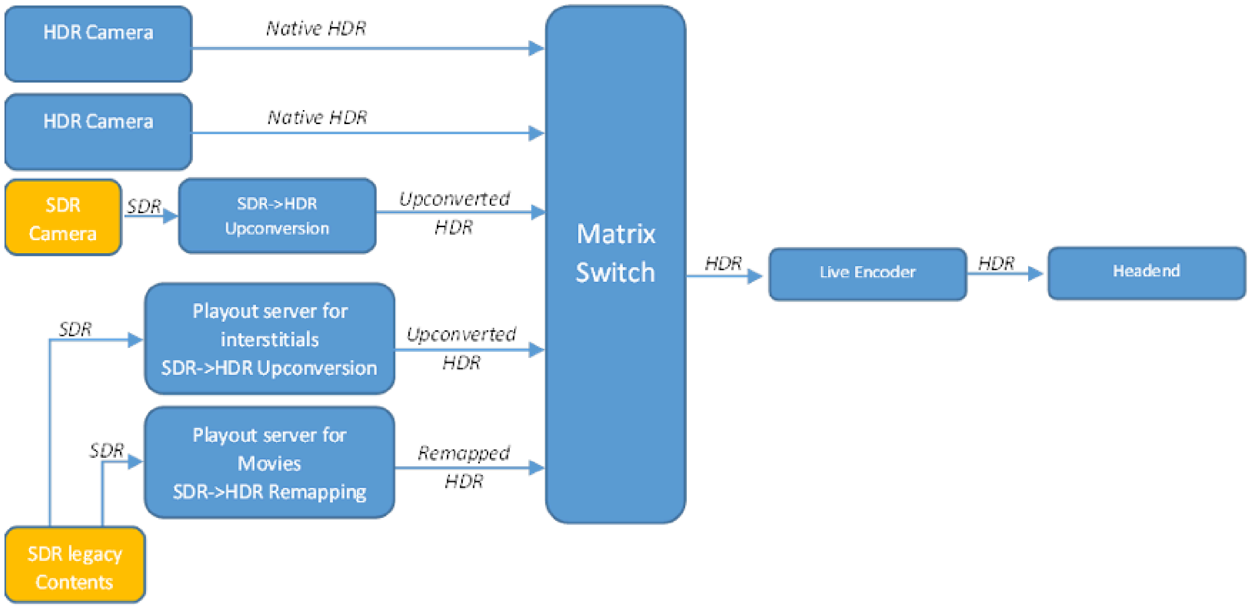


Figure 19. Sample Live Workflow with Mixed Format Source Content

Converting content to a common set of HDR and WCG technologies may occur at different points in the workflow from Production to the Consumer premises.

At the Production stage, multiple versions of the content may be produced. In the case of Pre-recorded content, producers have the opportunity to create multiple versions of the content applying creative judgment to each version. This is recommended for all Pre-recorded content and especially for interstitial material that may be inserted into a variety of Real-time Program Services, which may be operating with different HDR/WCG configurations. Live content may also be output in multiple formats; however, time constraints may prevent highly detailed artistic input. Automated conversion technologies can be “tuned” by the content creator to make the conversion the best it can be for the given program.

At the Broadcast Center or service provider stage, content may be converted to the provider’s chosen HDR/WCG configuration using automated tools. This may not include program-specific creative input; however, professional equipment may produce results that the service provider deems acceptable.

At the consumer premises, color volume transform may be possible for the purpose of backward compatibility (with limitations, see Sections 10 [V03] and 10.4 [V04] of the Violet Book). Also, display devices may “map” content internally to best suit the display characteristics. Both of those processes operate on a content stream with one, constant dynamic range and system colorimetry. Real-time Program Service providers should not expect consumer equipment to render seamless transitions between segments of content that have different transfer function or system colorimetry.

9.3. Conversion from PQ10/HLG10 to SDR BT.709

This operation is needed for PQ10 and HLG10. This method may be employed at the headend prior to final distribution to provide a backward compatible feed to legacy 2160p/SDR/BT.709 [2] TVs[13] and legacy HD networks. (See also Section 11 of the Violet Book [V05] on Format Interoperability including STB conversions.)

It is possible to do this conversion using a 3D LUT mechanism as described in Section 9.2. Another method is an invertible down-mapping process described in ETSI TS 103 433 [33], in which HDR/WCG content is down-mapped in real time to SDR/BT.709 at or prior to the emission encoder. Since commercially deployed in 2016, this down-mapping process is considered to be a Foundation Ultra HD technology.

9.4. Avoiding Image Retention on Professional and Consumer Displays

9.4.1. Background

As per EBU recommendation R 129 “Advice to Broadcasters on Avoiding Image Retention on TV Production Displays” [89], to minimize the risk of static image retention or premature ageing of displays, broadcasters and other content providers should take note of the following.

Broadcasters encounter ‘static’ images in a number of situations, including on-screen channel identifications, interactive application flags, banner displays, screens displayed when radio services are being received, program guides as well as longer-term text inserts such as sports scores.

[13] HLG offers a degree of backward compatibility with SDR/WCG displays that support BT.2020 [3] color spaces. HLG however, does not offer true backwards compatibility for BT.709 [2] displays without additional processing as described in this section. Both DVB and ARIB require support of BT.2020 system colorimetry in legacy 2160p/SDR TVs, so it may be reasonable to expect that many of these units are BT.2020-compatible, and thus able to render HLG content.

This type of content is likely to remain an editorial feature of many broadcasts. It should also be noted that if image retention issues occur in production or broadcaster environments, it can also occur on viewers' television screens.

9.4.2. Definition of Static Images

For the purpose of this document an image is deemed to be static if any part of the screen is occupied by any part of the image for more than a total of six hours in any 12-hour period on more than one occasion in a seven-day period.

If an image is not static the risk of a retained image being formed from it is low. To assist in ensuring that images are not static, certain specific practices might be considered, including:

- Moving the position of images on the screen from time to time in order that the definition of 'static' is not met.
- Instigating a time-out of static images where appropriate.

9.4.3. Recommendations

The luminance or fully saturated chrominance (in the case of high color static images) value of any static image should be restricted to a value equal to the average picture level of the screen in order to minimize the risk of forming a retained image.

Two alternative methods of achieving this are:

- To use a technique known as 'Linear–key mixing' that overlays the static image as a partly transparent image over the picture content. The 'added image volume' level that sets the apparent transparency should not be set any higher than a level necessary to make the added image acceptably visible.
- To limit the signal level of the static image to no more than:
    - 40% of peak white for standard dynamic range (SDR) static images
    - 47% of reference white1 (i.e. 35% of peak white signal level) for Hybrid Log-Gamma (HLG) high dynamic range (HDR) static images
    - 37% of peak white for ST.2084 Perceptual Quantizer (PQ) high dynamic range (HDR) static images

Further, it is recommended that the use of saturated color static images be avoided wherever possible and particularly where one is laid over the other.

10. Production for Pre-Recorded HDR Content[14]

This section focuses on the creation and distribution of pre-recorded content intended for inclusion in a Real-time Program Service. This includes all content filmed, captured digitally, or rendered in CGI and delivered via a file format to a service provider. Real-time Program Services may be delivered via MVPD (satellite, Cable, or IPTV), OTT and DTT. See Section 10.5 for Channel-based Immersive Audio production of Live content that is recorded for subsequent re-showing.

The following diagram depicts the interfaces and primary functions within the pre-recorded content creation process.

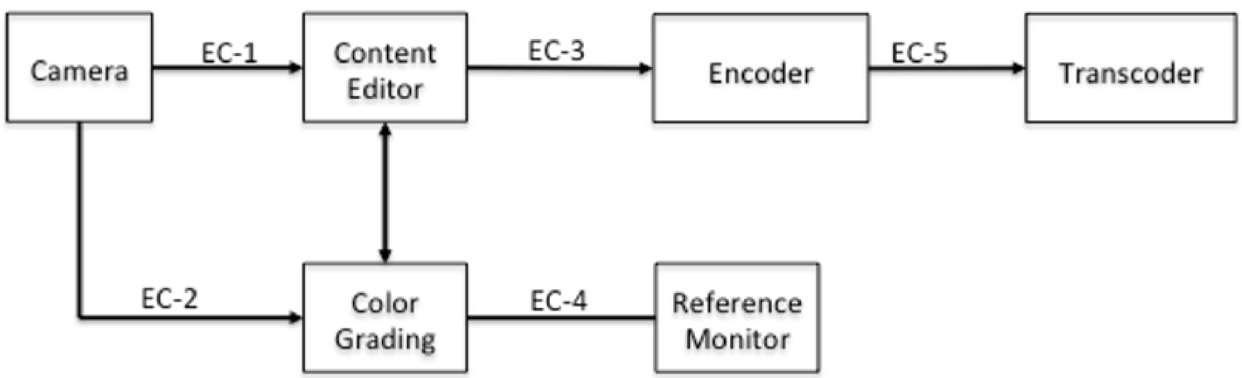


Figure 20. Pre-recorded Content Production Workflow and Interfaces

The scope of this guideline includes definition of six functions and five interfaces. The functional descriptions are described within sub-sections below, and the interface descriptions are described at a high level in the following Table 7.

[14] This section is for the production of non-real time content, not intended for pre-recorded content used in Single-Master live productions. See Section 11 Production for Live Content.

Table 7. Pre-recorded Content Interface Descriptions

Reference Point	Content Creation Functions	Reference Point Description
EC-1	Camera – Content Editor	Raw or log camera footage is ingested into the editing program. This interface may optionally contain metadata from the camera used by the editing program to provide guidance to timeline creation parameters.
EC-2	Camera – Color Grading	Raw or log camera footage is ingested into the color grading solution. This interface may optionally provide camera metadata describing the dynamic range resolution or other camera specific information from the captured footage.
EC-3	Content Editor – Encoding	Edited and color graded content is exported to an encoder for master and/or mezzanine level creation. This includes audio, video, essence, timing and CG content. Often the encoding function is included as part of the Content Editor.
EC-4	Color Grading – Professional Reference Monitor	This interface varies based on the content grading environment. It can include SDI, HDMI 2.0, Thunderbolt, DisplayPort and DVI.
EC-5	Encoder – Transcoder	This interface includes all aspects of a finished video delivered in file format for ingest to a transcoder. The transcoder uses this information to create distribution formats. If the transcoder is receiving uncompressed assets, it may also be an encoder.

10.1. Camera Requirements

10.1.1. High Dynamic Range and Wide Color Gamut Considerations

Material should be acquired at the highest resolution, dynamic range and color gamut of the camera in an operational data format best optimized for the image acquisition application in use. Although the primary consideration of this document is Foundation Ultra HD, capturing in the native camera format allows for future content grades being used for distribution of additional Ultra HD technologies.

There are numerous production examples that require the use of the camera specific raw or log data format from the camera sensing device. This is due to the requirement to acquire the highest level of picture quality to sustain the levels of post-production image processing usually encountered in high-end movie or a ‘made for TV’ drama or documentary with use of color correction, visual effects and the expected high levels of overall image quality.

There are equally numerous applications for episodic and live capture that employ camera-specific logarithmic (log) curves designed to optimize the mapping of the dynamic range capability of the camera sensors to the 10-bit digital interface infrastructure of production and TV studio operations. Such log 10-bit signals are often of the 4:2:2 color-sampling form and are either generated internally in the camera head or locally created in the camera control unit. These images are usually recorded utilizing high-quality mastering codecs for further post-production of the captured scenes or, in the case of live transmissions, minimally processed for real-time transmission to viewers.

Camera native log curves are typically designed by camera manufacturers to match the performance characteristics of the imaging sensor, such as sensitivity, noise, native dynamic range, color spectral characteristics, and response to extreme illumination artifacts. It should be noted that data describing the gamut and transfer function characteristics encoded at the camera must be passed down the production chain, particularly in cases where not all cameras used are operated with the same parameters. In post-produced content, such information is typically carried in file metadata, but such metadata is not embedded in the video signal.

10.1.2. Imaging Devices: Resolution, Dynamic Range and Spectral Characteristics

In the creation of Foundation Ultra HD content signals, ideally, the image sensing devices should have a sensor resolution and dynamic range equal to or greater than a pixel count commensurate to the signal format.

In the area of spectral characteristics, the more advanced sensing devices will exhibit characteristics approaching the system colorimetry of BT.2020 [3], while more typical devices will produce color performance approximating the DCI-P3 gamut or just beyond the gamut of BT.709 [2].

Additional considerations for 2160p capture include:

- Not all lenses are suitable for capturing 2160p resolution and the MTF of the lens must be sufficient to allow 2160p capture on the sensor.
- Nyquist theory applies to camera sensors and care may be needed in selecting cameras that achieve target 2160p spatial resolution performance.
- When transmitting film content, the 16:9 aspect ratio of Foundation Ultra HD does not correspond to the wider aspect ratio of some movies. The two alternative methods of re-formatting (full width or full height) represent the same compromises that exist in HD transmission.
- Content originating from 35mm film will translate to Foundation Ultra HD differently than digital sources, e.g. film grain, achievable resolution.

10.2. Reference Monitor Considerations[15]

Use of 2160p, HDR and WCG imply the need for reference monitors that will allow production staff to accurately adjust devices or modify content. Previously in SD and HD TV, there were accepted working practices using professional monitors and standard test signals, such as color bars, PLUGE, sweep, etc. Digital Cinema employs slightly different techniques, using calibrated monitors and LUTs to replicate the viewing characteristics of different consumer viewing conditions.

[15] Note that this section is not meant for pre-recorded content used for Single-Master UHD HDR/SDR production. See Section 11.2 Viewing of HDR and SDR Images in Close Proximity

The recommendation for Foundation Ultra HD is to agree on practical standardized methods for use of Reference Monitors for Real-time Program Service production.

For Foundation Ultra HD, a reference monitor can ideally render at least the following: resolutions up to 3840x2160, frame rates up to 60p, BT.2020 [3] system colorimetry (ideally at least the P3 gamut), and HDR (i.e., greater than or equal to the contrast ratio that could be derived from 13 f-stops of dynamic range). It should be noted that as with HD, consumer display technology is likely to progress beyond the capabilities of current generation professional displays. As such, instruments such as waveform monitors and histogram displays are essential tools to ensure optimum Foundation Ultra HD delivery.

10.3. On-Set/ Near-Set Monitoring

Viewing conditions for HDR monitoring:

- While it is difficult to establish parameters for viewing conditions for on-set monitoring, it is advisable to follow the recommendations for setup of an on-set monitor as described in BT.814 [14] or the latest version of an equivalent standard. BT.2100 [5] contains some specifications on reference viewing conditions (e.g. 5 nit surround, >1,000 nit peak white, D65 White Point)).

Dynamic range of on-set monitor:

- It is recommended to have display devices for on-set monitoring capable of at least 600 cd/m^2 of peak brightness. Some RGB OLED mastering monitors are capable of 1,000 cd/m^2 of peak brightness. Note that future HDR content delivered to the consumer may be intended for substantially greater than 1,000 cd/m^2 peak display brightness.

10.4. Color Grading

10.4.1. HDR Grading Workflow

Professional color grading should take place in a controlled environment on a professional monitor whose capability is known, stable and can be used to define the parameters of the ST 2086 [10] Mastering Display Color Volume Metadata.

Current industry workflows such as ACES [50] are recommended to be used where the operator grades behind a rendering and HDR viewing transform, viewing the content much as a consumer would. The work would result in a graded master that, when rendered with the same transformation, will result in a deliverable distribution master. (See also Section 9.1 of the Indigo Book [I02] on ACES.)

10.4.2. Grading Room Configuration (for non-live pre-production)

When there is a need to prepare both HDR and SDR video productions, which share the same physical environment and require mixing segments of different dynamic range characteristics in the same master, it is important to ensure the use of equivalent illumination levels encountered in a conventional grading environment. This is because it is important to review both the SDR and HDR rendering of the images to guarantee a level of consistency between them. The black and dark tones for both the HDR and SDR video pictures are a particular concern as the ability to discriminate the information they contain is highly affected by viewing conditions.

Secondary monitors should be turned off so as to not impact room brightness during grading or the client viewing monitor must be positioned so as to not impact the bias lighting observed by the colorist.

10.4.3. Grading Monitor Configuration

A professional mastering monitor should be specified and configured according to the required deliverable: System Colorimetry, Transfer Matrix, EOTF, White Point and RGB level range (Full or Narrow).

- BT.2100 [5] defines an HDR reference display as one that is capable of peak luminance at or above 1,000cd/m^2 which also conforms to the HLG or PQ specifications and supports the display of BT.2020 [3] color primaries (ideally including the extents of the P3 color space)
- BT.2035 [145] defines an SDR reference viewing environment which uses an SDR display that supports the BT.1886 EOTF which defines a peak luminance adjustment (using L_w) between 100-250 cd/m^2.
- BT.2129 [149] defines program production monitoring environments.

10.5. Channel-based, Backward Compatible, Immersive Audio Post Production

This section describes the production of content with channel-based immersive Audio containing height-based sound elements, that is backward compatible with some existing 5.1 surround decoders. This example is for post-produced content (file-based workflow) which can then be used in a real-time program assembly system and distributed in a linear television distribution system. Commercially deployed immersive audio systems that support channel-based, backward compatible audio production include E-AC-3+JOC[35] [16]. AC-4 and MPEG-H can be used as a mezzanine codec that can be recompressed while assuring sample alignment between all audio channels. E-AC-3+JOC(DDPJOC) [35] is designed for the last stage encoding which a consumer device will decode.

A high-level diagram of a post-production mixing environment for this format is shown below:

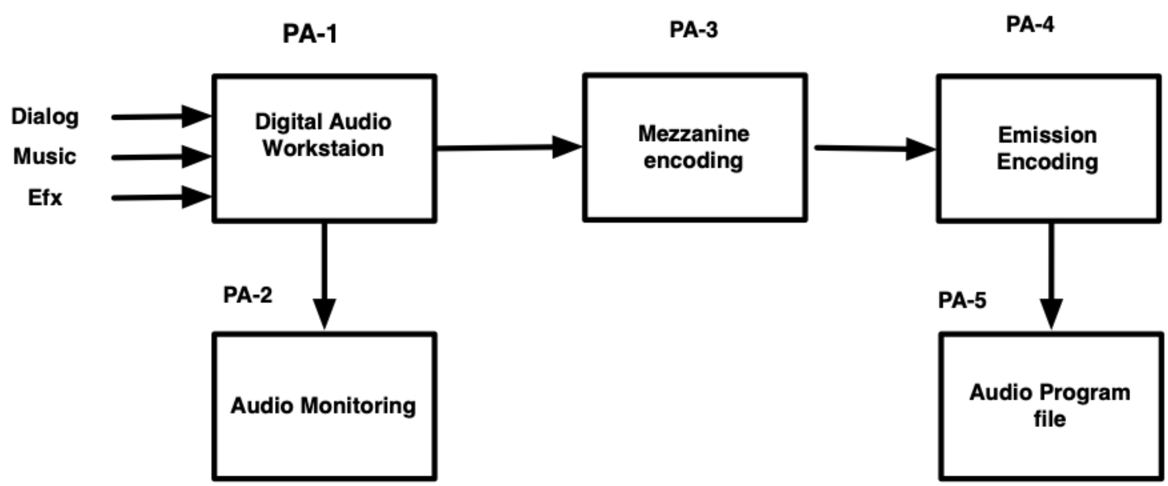


Figure 21. Channel-based Immersive Audio Post-Production

Table 8. Channel Based immersive Audio Workflow Elements

[16] Something about MAT2 from Felix.

Reference Point	Content Creation Functions	Reference Point Description
PA-1	Digital Audio Workstation	A mixer uses a digital audio workstation to create groups of audio tracks consisting of dialog, music and effects. Each group consists of channel-based audio beds (e.g. 5.1, 7.1.4).
PA-2	Audio Monitoring	The Immersive Audio mix is rendered to multiple different speaker layouts including 7.1.4 and 5.1 using hardware film-mastering style tools or commercially available software tools
PA-3	Mezzanine Encoding	Immersive Audio and 5.1 channel-based renders are encapsulated into a mezzanine file. Use of AC-4, DTS UHD or MPEG-H
PA-4	Emission Encoding	Immersive Audio is encoded into formats such as E-AC-3+JOC[17] for final emission delivery
PA-5	Audio Program File	SMPTE ST 337 [36] embedded in a delivery file.

10.5.1. Immersive Audio Monitoring Environment

Immersive Audio monitoring should take place in a controlled environment consisting of a treated listening room and a calibrated speaker environment. The recommended speaker configuration for immersive mixing and monitoring is 7.1.4 as per Recommendation ITU-R BS.2051[18][129] (System G minus the Left and Right screen loudspeakers) calibrated to a reference sound pressure level appropriate for the size of the mix room. Typical levels range

[17] See Violet Book, Section 11.5 [V07] regarding backward compatibility of E-AC3+JoC.
[18] System G minus the Left Screen and Right Screen loudspeakers would be used for 7.1.4, while System D could be chosen for 5.1.4.

from 76 dB SPL C-weighted, slow response for a room less than 1,500 cubic feet or smaller (i.e., many outside broadcast trucks) to 85 dB SPL, C-weighted, slow response for rooms 20,000 cubic feet or larger (i.e., many film stages).

10.5.2. Immersive Audio Mastering

The channel audio data is recorded during the mastering session to create a ‘print master’ which includes an immersive audio mix and any other rendered deliverables (such as a 5.1 surround mix). Program loudness is measured and corrected using ITU-R BS.1770-4 [37] methodology.

10.6. Still Image Content Creation with HDR and SDR

Introduction

In order to produce native HDR and SDR graphics it is important for a content creator to be able to identify specific image characteristics so that they can be displayed, converted or archived properly. “Coding Independent Code Points” (CICP) defined in ITU-T H.273 (2021) [114] are essential for correct identification of these image characteristics in a PNG or TIFF file which are commonly used in production and distribution.

Video wrappers and streaming codecs (QuickTime, MXF, mp4, AVC, HEVC) have had the ability to use CICP, however this ability has been lacking in uncompressed image file formats such as PNG and TIFF. The High Efficiency Image File Format (HEIF) has the ability to carry the CICP parameters and thus can populate the VUI and SEI messages within HEVC.

Below is Table 9 for the specific image characteristics identified in CICP:

Table 9. CICP Image Characteristics

CICP	Examples	Purpose
Color Primaries	BT.709 [2], BT.2020 [3]	Identify image color space
Transfer Function	SDR, PQ, HLG	Identify light mapping formula
Matrix Coefficents	0=Identity(RGB) 1=BT.709 18=BT.2020	Identify equation for conversions between YC_BC_R and RGB
Full Range Flag	1=Full 0=Narrow	Identify data range (identifies signal range quantization)

There are two methods that are in use for signaling CICP in still image files. PNG 3rd Edition establishes cICP, mDCv, and cLLi chunks. ICC profiles v4.4 includes a new CICP tag. All of this functionality is new and may require testing for use in various workflows.

Current Status:

- CICP support is read-only in many locations of MacOS and 3rd party apps. PNG chunks will not be written on export but will be translated when creating a movie from a still image file.
- MacOS does not read the full-range flag in PNG files
- Authoring of CICP for the still image files can be done by applying ICC Profiles within Photoshop.
- Final Cut Pro and Compressor can read the PNG CICP chunk and TIFF ICC profiles with CICP.
- 3rd party applications that use the MacOS API for the CICP chunk can read the CICP signaling (Affinity Photo, Pixelmator/Photomator)

In May, an effort that Apple started was ratified by ISO (ISO/TS 22028-5 [148]). It specifies the use of CICP and many other parameters for proper identification of “extended colour encodings for digital image storage,manipulation and interchanges”.

10.6.1. PNG 3rd Edition (CICP Chunk)

PNG 3rd Edition which is about to be released by W3C has the ability to signal CICP image characteristics explicitly.

MacOS has implemented PNG 3rd Edition with the exception of the full-range-flag. Also, a number of applications are now supporting CICP (Affinity Photo and Pixelmator/Photomator).

Here are two helpful links to describe this new functionality in PNG 3rd Edition:

W3C PNG 3rd Edition (Github Repository): https://www.w3.org/TR/png-3/

PNG-3rd Edition Explainer:
https://github.com/w3c/PNG-spec/blob/main/Third_Edition_Explainer.md#labelling-hdr-content

10.6.2. TIFF (ICC Profiles CICP Tag)

ICC Specification ICC.1:2022 / Profile Version 4.4.0.0 was released in 2022. Adobe Photoshop and MacOS recognize the tags if properly attached to a TIFF file.

ICC Profile v.4.4.0.0 (ICC.1-2022-05)

11. Production for Live Content

The Ultra HD Forum's objectives include understanding the implications of creating and delivering Foundation Ultra HD content all the way through the production chain. It is important that the whole system is understood as technical decisions to deploy a particular Ultra HD production approach upstream may have implications for downstream delivery equipment. The reverse is also true that downstream delivery constraints may affect production techniques.

Live content examples include sports, news, award shows, etc. Pre-recorded content is captured and produced in advance (see Section 10). Examples include soap operas, sitcoms, drama series, etc. Real-time Program Services may include Live programs and Pre-recorded programs, and Pre-recorded content – such as advertising – may be inserted within Live programs. Real-time Program Services may be delivered via MVPD (Satellite, Cable, or IPTV), OTT and DTT and are delivered on a schedule determined by the service provider.

Unlike Cinema, Ultra HD Blu-ray™ discs or On Demand, Real-time Program Services involve performing complex manipulation of images and audio in real-time at multiple points in the delivery chain, including graphics, captions, virtual studios, Live sound mixing, logos, chroma-keying, DVE moves, transitions between sources, encoding, decoding, transcoding, ad insertion, voice over production, monitoring, conditioning, rendering, closed captioning, standards conversion, etc.

11.1. Live Production in Trucks or Studio Galleries

At the live event venue or studio, the action is captured and prepared in near real-time. This section addresses the processes that may be required to assemble the live show prior to compression for distribution to a "headend" or other central facility. Such video processes may include:

- Synchronous live baseband inputs from cameras to production switcher
- Racking of cameras (for chroma/luma/black balance)
- Transitions between cameras (mix, wipe, DVE, etc.)
- Keying (chroma key, linear key, alpha channel, difference key, etc.)
- Overlay of graphics, text, etc.
- Slow motion and freeze frame
- Editing (for action replay/highlights)
- Use of virtual sets

Simultaneous production of both 2160p/HDR and HD/SDR output may be employed, and assets may be shared, e.g., HD graphics or 2160p graphics for both outputs. It is recommended to maintain a single HDR transfer function and system colorimetry. See Section 8.2 and Section 8.3 for details on converting assets from SDR to HDR or vice versa.

Performing the above functions on Foundation Ultra HD content may be different than for HD/SDR content. Foundation Ultra HD content may involve higher spatial resolution, frame rates up to 60p, and HDR/WCG.

- Graphics created using SDR-BT.709 [2], sRGB [150], AdobeRGB [151] and other typical graphic file formats may need re-mapping to an HDR transfer function depending on the desired creative intent.
- HLG has the characteristic of providing a degree of backward compatibility with legacy SDR devices, both professional and consumer. However, care must be taken to ensure the signal being displayed is encoded with the appropriate system colorimetry preset in the display. For instance, for HLG10, a displayed picture on a legacy display that uses BT.1886 [4] / BT.709 [2] system colorimetry will be incorrect. See Section 8 of the Yellow Book [Y01] for details and caveats.

11.2 Viewing HDR-SDR Images in Close Proximity[19]

Some Single-Master UHD HDR-SDR productions may take advantage of the reference monitor contrast adjustment range where there is a need to unify the reference white luminance levels during monitoring/shading/grading where HDR and SDR displays or images are in close proximity.

Approach A

In this case, an HDR (HLG or PQ) display will use the BT.2100 [5] reference setting (1,000nits) and the SDR display contrast will be adjusted to 203nits.

Approach B

In this case, an SDR display will use the BT.2035 [145] reference setting (100nits) and an HLG display will use a peak white luminance of between 300-600nits and the appropriate overall system gamma.

[19] This topic is being documented in some standards bodies and subject to update at a later date. It is currently described in BT.2408-7 [8] Annex 9 and 10.

11.3. HDR Production with encodings other than PQ and HLG

For episodic television, as with features, colorists typically work in an underlying grading space adopted by the project. Usually, this grading space comprises a log encoding, often with a color gamut native to a principle camera. The content master is created and (ideally) preserved in this project grading space. Conversion to a distribution encoding, e.g., PQ, is performed by an output viewing transform and viewed on a specific mastering display during grading, whereby the grade targets a particular peak luminance.

Similarly, live productions establish a single project grading space, often choosing one that follows their camera's native transfer function. Shading and vision mixing occurs within that selected space.

When contributions having different transfer functions and/or color gamuts are to be combined, there is a concern about cumulative quantization effects. One way to address this is to defer conversion away from the native system colorimetry as long as possible. A more general solution is to transform the various contributions to a common linear encoding as a shading space. In either case, the produced result is converted to the required delivery format, suffering the quantization of conversion only once.

In either case, a PQ master can be obtained with a corresponding output viewing transformation. Likewise, to obtain an HLG master, an appropriate output viewing transformation is used. Then the PQ master targets 1000 cd/m^2 and the HLG master is shown on a 1000 cd/m^2 nominal peak luminance display, they are typically expected to visually match each other. ITU-R report BT.2390 [76], Sec 7.3, specifies 1000 cd/m^2 as the reference condition to ensure matching between PQ and HLG mastered content. (Note: see Section 8.1. for more information on conversion between PQ and HLG content)

Other workflows may grade in a distribution space, e.g., PQ or HLG, which provides one distribution master directly. Other distribution masters are obtained by performing a conversion to the target space, however such conversions are sometimes difficult to match well. Further, some distribution spaces impose limitations that compromise whether a content archive is future proof.

11.4. Channel-based, Backward Compatible, Immersive Audio Production

For a live event, immersive audio can be created using existing mixing consoles and microphones used during the event. Using additional console busses, height-based ambience and effects can be added to a traditional 5.1 or 7.1 channel mix which can then be encoded by an ETSI TS 103 420 [35] compliant E-AC-3 encoder. A local confidence decoder can be employed to check typical down mixes, including the backwards compatible 5.1 channel render described in ETSI TS 103 420. During normal mixing, this confidence decoder can serve as a useful continuity check and display (i.e. to make sure the mix is still "on-air"), though due to normal latencies it will likely be found to be impractical to be kept in the monitor path full time. A high-level diagram of a live mixing environment was used at a major televised event[20] using E-AC-3+JOC [35]. is shown in Figure 22 below.

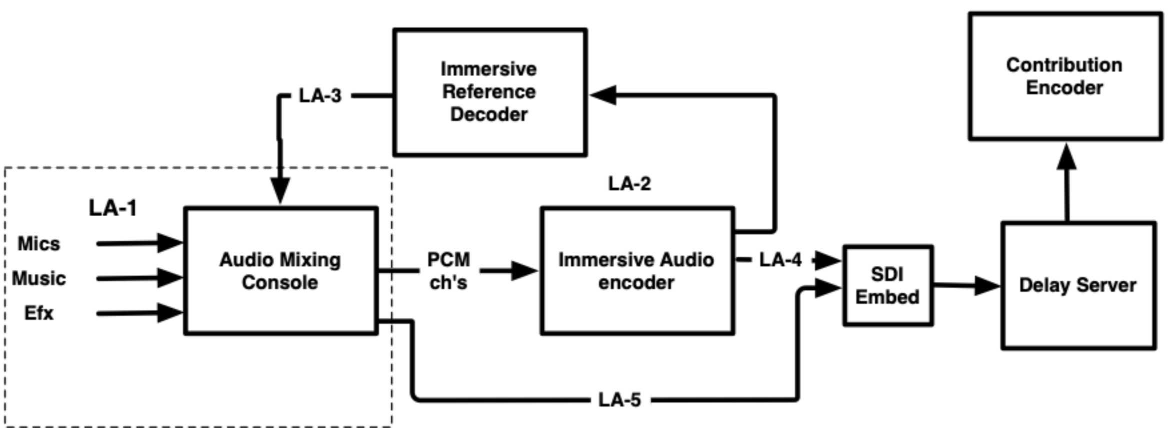


Figure 22. Channel-based Immersive Audio Live Production

[20] See Section 13.3. NBCUniversal Olympics and 2018 World Cup [V08]

Table 10. Description of Live Channel Based Audio Workflow Elements

Reference Point	Content Creation Functions	Reference Point Description
LA-1	**Audio Capture and mixing**	**Microphones are placed throughout the live venue. Microphone feeds are brought into the audio mixing console in the same fashion as 5.1 production. Sound elements are mixed into a final immersive program.**
LA-2	**Immersive Audio Authoring**	**Audio groups are created in the mixing console representing the immersive mix as channel-based (e.g. 5.1.2, 5.1.4, 7.1.4) audio[21].**
LA-3	**Immersive Audio Monitoring**	**The Next-Generation Audio Processor renders the audio to 5.1.2, 5.1.4 or 7.1.4 speaker feeds and sends these feeds back to the audio mixing console for monitoring in the live mix environment.**
LA-4	**Immersive Audio Encoding**	**Atmos immersive program, encoded as E-AC-3+JOC [35] is delivered as a 5.1 channel bitstream + parametric side-data (steering data) to the contribution encoder[22] transported over either a MADI or SDI link.**
LA-5	**Legacy Audio Delivery**	**Stereo or 5.1 complete mixes may be created at the audio mixing console and delivered via traditional means.**

[21] The downstream Atmos Channel-Based Immersive emissions encoder, using E-AC-3 + JOC will render a legacy 5.1 audio program. It is recommended to verify the rendered 5.1 audio program using a suitable E-AC-3 decoder in the monitor chain.

[22] See Section 11.5 of the Violet Book [V07] regarding backward compatibility.

In this channel-based immersive audio example, a Dolby Atmos enabled E-AC-3+JOC encoder generates a compressed bitstream containing 5.1 channels of audio that is backwards compatible with legacy E-AC-3 decoders. In parallel, the encoder generates an additional set of parameters (as specified in ETSI TS 103 420.) that are carried in the bitstream, along with the 5.1 audio, for use by a Dolby Atmos E-AC-3+JOC decoder. The full Atmos decode process reconstructs the original channel-based immersive audio source from the 5.1 backwards compatible representation and the additional parameters. A typical channel-based immersive program can be encoded, for example, at 384-640 kbps total, thus fitting into existing emission scenarios.

For discussion of Next Generation Audio for live production see Section 12.2.2.

12. Additional Ultra HD Technologies beyond Foundation Ultra HD

Methods and challenges for adding non-Foundation Ultra HD technologies into real-time program assembly vary from technology to technology. Further, an enhancement Ultra HD technology might be continually incorporated into the program stream (i.e., part of the “house format”) or could be incorporated intermittently. Ultra HD Forum will continue to study these cases and intends to provide further information in future versions of these Guidelines.

12.1. Dynamic HDR Metadata

For PQ HDR content[23], as described in Section 7.2.5 of the Orange Book [O01], HDR10 provides the static metadata elements in a PQ10-based HDR format as specified by SMPTE ST 2086 [10], MaxFALL, and MaxCLL. That section identifies a number of limitations with these particular HDR metadata values, notably the difficulty with setting these values in a live environment, real-world experience suggesting that these values have been set to artificial numbers to force certain looks on consumer displays, and the inability to correctly set these values given limitations of mastering displays.

In addition to these limitations, the values of MaxFALL and MaxCLL are also very limited in that they are only currently specified to provide single values for the entirety of the program. The dynamic range of both narrative and live content can vary dramatically from scene to scene. As a result the static, program-wide metadata values, as strictly defined, are of limited use for a great deal of content that does not have a static, unchanging dynamic range. Interoperability tests show that receivers can recognize changes in the static metadata within the duration of a program; however, it is yet unknown how frequently or quickly such changes can be recognized. For example, it is not expected that static metadata would change on a frame-by-frame basis.

[23] HLG10 does not specify any display metadata as it is based on normalized scene-light, rather than the absolute luminance of the signal seen on the mastering display, as described in Section 7.2.2 of the Orange Book [O02]. As such, the headroom (measured in f-stops) for HLG highlights above HDR Reference White, is approximately constant regardless of the display’s nominal peak luminance. Moreover the HLG10 display EOTF, which is fully specified by the ITU-R BT.2100 [5], includes a variable display gamma to provide adjustment for a specific display’s peak brightness capabilities, along with eye adaptation; thereby allowing HLG to function in brighter viewing environments. Thus static or dynamic metadata is not required for HDR productions using HLG10.

Finally, there is no standardized way of utilizing these values in the final consumer display, so displays differ significantly in reproduction of the image. In practice some displays may ignore the values altogether. This is not consistent with the goal of displaying the image as close to the creative intent as possible on the target display.

A number of Dynamic Metadata methodologies have been developed to address the limitations of PQ10 and HDR10. Dynamic Metadata refers to metadata that describes the image at a much finer temporal granularity, scene-by-scene or even frame-by-frame and produces significantly more information about the mastering and creative intent of the scenes. In addition, most of these methodologies provide detailed information about tone mapping in the consumer display with the goal of consistent images across different manufacturers' displays. The methodologies also are designed to preserve creative intent, with the final displayed image being as close to the mastered image as the consumer display has the ability to reproduce.

Some of these methodologies go further by capturing the metadata during the color grading session and passing that metadata to consumer displays to better reproduce the creative intent. Most metadata schemes also provide for automatic metadata creation, which is useful in workflows for live content.

In general, several of these dynamic metadata schemes are additive, in that they provide additional information about the carried PQ10 image, and the HDR10 static metadata remains intact alongside the dynamic metadata. In some cases, this can provide a simple backwards compatibility to an HDR10-only display - the dynamic metadata is simply ignored.

Finally, many of these methodologies have considered how the signal can be backwards compatible with SDR displays and have built-in methods for conversion. See Dolby Vision™ described in Section 8.1 [I03] and SL-HDR2 described in Section 8.4 [I04] of the Indigo Book.

SL-HDR1 is another HDR dynamic metadata technology, which serves a different purpose than Dolby Vision or SL-HDR2. SL-HDR1 is intended to enable the service provider to emit an HDR/2020 [3] service in an SDR/709 format that can be "reconstructed" to HDR/2020 by the receiver. HDR/2020 receivers that can interpret the SL-HDR1 metadata can present the HDR/2020 format to the viewer. The SDR/709 content can be displayed by receivers that cannot display HDR/2020. In this way SL-HDR1 provides a measure of backward compatibility for both HLG and PQ-based HDR content. It should be noted that SL-HDR1 requires 10-bit encoding, and so may not help address legacy SDR/709 receivers that are only capable of 8-bit decoding. See Section 8.3 of the Indigo Book [I05].

12.2. Ultra HD-Next Generation Audio (NGA)

Next Generation Audio (NGA) offers new tools for content creators and new experiences for consumers. Immersive audio offers a sense of being completely surrounded by the aural experience: above, below, and all around the listener. NGA can not only make these experiences possible in a home theater environment with optimally placed speakers, but also via headphones, soundbars, or even sub-optimally placed speakers to an extent. NGA can also allow the consumer to personalize the experience, such as increasing the dialog level relative to the music and effects or choosing a particular dialog track such as alternate language or commentary. See Section 7 of the Indigo Book [I06] for details about Dolby AC-4, DTS-UHD and MPEG-H NGA systems.

12.2.1. NGA Post Production

Maybe the most important production aspect is that NGA systems bring new paradigms in authoring and mixing the audio content requiring additional training for sound engineers. For delivery of NGA immersive and personalized experiences, authoring of metadata plays an essential role in production. Using plugins for DAW (Digital Audio Workstations) in post-production or metadata authoring and monitoring units for live production, the metadata can be authored together with the audio data delivered to the final audio encoders. In post-production the metadata can be stored as an XML file according to the ITU-R Audio Definition Model (ADM) [72]. It is anticipated that ADM profiles with tailored feature sets will be necessary at each stage of the NGA broadcast and content production chain to ensure full interoperability and quality of service.

Post production and delivery of NGA assets with ADM metadata in Broadcast Wave (BWAV) files is now widely supported. ITU-R Recommendation BS.2088 [73] specifies the mapping of ADM XML data to 64 bit BWAV files via RIFF chunks. Recent work in SMPTE has focused on defining MXF and IMF mappings for ADM metadata to support NGA and other advanced audio applications. SMPTE ST 2131, currently in development, will define mapping of BWAV RIFF chunks, including RIFF chunks containing ADM metadata, to MXF files to enable interchange of ADM between MXF and BWAV files. An IMF plug-in for ADM metadata (SMPTE ST 2067-204) is also in development.

A serial representation of ADM (S-ADM) has been defined by the ITU in ITU-R BS.2125 [123] to support use of ADM in linear workflows such as live or real-time production for broadcasting and streaming applications. To enable interchange between file based ADM and live/linear applications SMPTE has specified a mapping of S-ADM to MXF (SMPTE ST 2127-1 [146] and

SMPTE ST 2127-10 [147]) and is in the process of developing an IMF plug-in for S-ADM metadata (SMPTE ST 2067-203). Together with the ADM MXF and IMF file mappings these specifications seek to fully enable all NGA functionality in post production and distribution and allow seamless interchange of audio essence and metadata between live and file workflows with no loss of function.

12.2.2. NGA Live Production

A key consideration of implementing NGA for live production is the new requirements for dynamic metadata authoring and mixing the audio content requires additional training for sound engineers. For delivery of NGA immersive and personalized experiences, authoring and carriage of dynamic metadata plays an essential role in production.

Live productions using 3G or 12G SDI interfaces with a minimum of 16 PCM audio tracks have been used for delivery of immersive audio programs, with 4 height channels and several commentary tracks as objects. More channels may improve the user experience and allow for more interactivity. Dynamic audio objects can be added to the program mix, when the production tools have features such as dynamic spatial panning.

Additional mechanisms for efficient and secured metadata delivered over an SDI interface are provided for specific NGA systems, see Sections 12.2.3.2 and 12.2.3.3.

As noted, a serial version of ADM (S-ADM) has been defined by the ITU to support NGA and other advanced audio applications in real-time production using serial interfaces. A number of additional standards have been defined or are being developed to define transport of S-ADM in serial interfaces. For traditional SDI based infrastructures the use of SMPTE ST 337 [36] to transport non-PCM data over AES3 has been leveraged to transport S-ADM. SMPTE ST 2116 [124] defines transport of S-ADM over AES3 using a unique ST 337 data type. GZIP encoding has been included in the specification to allow many NGA applications to require only a single AES3 channel for S-ADM data transport. SMPTE ST 2109 [127] specifies the transport of more generic audio metadata transport over AES3 allowing metadata beyond S-ADM that may be required in some NGA applications (e.g. codec specific metadata). The example workflows described in the section below show the use of these standards to enable NGA applications in SDI based infrastructures.

The use of ST 337 for metadata not only allows S-ADM data to be carried in AES3 and SDI interfaces, but also in other interfaces that can transparently convey AES3 or SDI signals. This includes standards for conveying AES3 and SDI signals over IP interfaces. The SMPTE standard that defines transport of AES3 in SMPTE 2110 networks is of particular importance as it allows S-ADM data to be transported in SMPTE 2110 networks in a manner compatible with legacy SDI networks. For NGA applications that utilize a hybrid of SDI and SMPTE 2110 this is the preferred solution for transporting S-ADM as minimal conversion between interfaces is required. AES3 has been used to transport compressed audio for many years and the issues/solutions for dealing with non-PCM are well known so this approach is well tested.

The use of AES3 to transport metadata is not without limitations however, including the relatively limited data bandwidth of a single AES3 channel. While GZIP reduces S-ADM data requirements enough to support many current NGA applications, more advanced applications can easily exceed the data capacity of an AES3 channel leading to additional complexity in the use of multiple AES3 channels to convey the S-ADM data. For all IP infrastructures that do not require legacy SDI compatibility a more optimal transport solution utilizing SMPTE ST 2110-41 (currently in development). SMPTE ST 2110-41 supports transport of generic data over SMPTE 2110 networks without the legacy limitations of SDI ancillary data (SMPTE ST 2110-40 [47]). SMPTE ST 2127-2, also in development, describes transport of audio metadata, including S-ADM, via SMPTE ST 2110-41. The audio metadata is transported in frames consistent with the audio metadata described in SMPTE ST 2127-1 [146] allowing for seamless interchange with S-ADM audio metadata in MXF files. S-ADM also has mechanisms to reduce data rate through different frame modes, which may provide an alternative to using multiple AES3 channels.

From the many trials conducted in France, Germany and U.S. and the commercial broadcasts in South Korea and Brazil using NGA, it has been found that integration of authoring and monitoring units in common SDI infrastructures (including remote facilities) is straightforward and sound engineers become familiar with mixing immersive content and authoring metadata relatively quickly. For example, Korean broadcaster SBS produced the 2018 Football World Cup using MPEG-H with immersive and interactive audio services (Korean commentary, English commentary, and stadium atmosphere alone).

Additional trials were conducted in 2018 using object-based audio and interactive audio services (including separate language tracks in some cases): by NBC during Winter Olympics using E-AC-3 plus JOC [35], by France Television during Roland Garros Tennis Open using both MPEG-H [70] and AC-4 [65] , as well as by the EBU during the European Athletics

Championships using both MPEG-H and AC-4. Since 2019, trial broadcasts of MPEG-H included: “Rock and Rio” (Brazil 2019), Eurovision Song Contest (Israel 2019), the French Tennis Open (France 2019) and the Football World Cup (Brazil 2022).

For the Tokyo 2020 Olympic broadcast, NBC/Comcast (US), Astro (Malaysia), beIN Sports (MENA) broadcast audio using Dolby Atmos over AC-4. NHK (Japan) delivered audio in 22.2 channels with AAC, and MBC (Korea) broadcast the Olympics using MPEG-H audio.

12.2.3. NGA Production tools and workflows

12.2.3.1. Introduction

To achieve a healthy and growing ecosystem with NGA content flowing from production to the end consumer, a few essential needs must be addressed in the content creation and interchange parts of the workflow. Top level needs are:

- Independence from the delivery codec solutions - ensuring open solutions from multiple vendors secures scaling of NGA content interchange and archival.
- Workflow flexibility - enabling several alternatives when transitioning from today’s workflow towards NGA enabled workflows.
- Interoperability - between traditional SDI based systems and new IP transport and cloud-based infrastructures.
- Graceful coexistence - with existing 2.0 and 5.1 content.
- Extensibility - can grow as use cases of NGA emission systems evolve.

To address these needs, several pieces must necessarily come together. A fundamental part of the solution is for the industry to collaborate around open standards that define the metadata along with the definition of how these are carried in different container formats. This is discussed further in the section on relevant standards.

Furthermore, several open-source projects are ensuring the availability of necessary NGA components to the different solution providers in the workflow, which combined with trials and interoperability testing events can facilitate both the broad adoption of NGA and interoperability between these different vendor realizations of the technology.

This section describes several “real world” use cases of NGA tools and workflows implemented and available for “real world” UHD productions.

The following information is arranged in three main sections:

- Dolby Workflows
- MPEG-H Workflows (Real-time Workflows & File-based Workflows)
- Workflows using S-ADM in Production

12.2.3.2 Dolby Technology Workflows

Case Study 1: Static metadata over SDI infrastructure

Polish public broadcaster Telewizja Polska (TVP) decided to deliver Dolby Atmos via Dolby AC-4 [56] next-generation audio for a temporary UHD channel for all Euro 2020 matches. NGA has been used in all the live transmissions as well as rebroadcasts of recorded material.

In this workflow, the existing 16-channel SDI embedded audio infrastructure was used in combination with static metadata to enable NGA content to be carried prior to final transmission encoding.

After conducting internal NGA tests using AC-4, including a successful DD+ Atmos transmissions on TVP1 HD in 2018, the Polish national broadcaster, TVP, decided to make NGA a part of their UHD transmissions on the TVP 4K channel. Their goal is to deliver special events with NGA experiences by using their standard contribution and 16-channel SDI embedded audio in-house infrastructures.

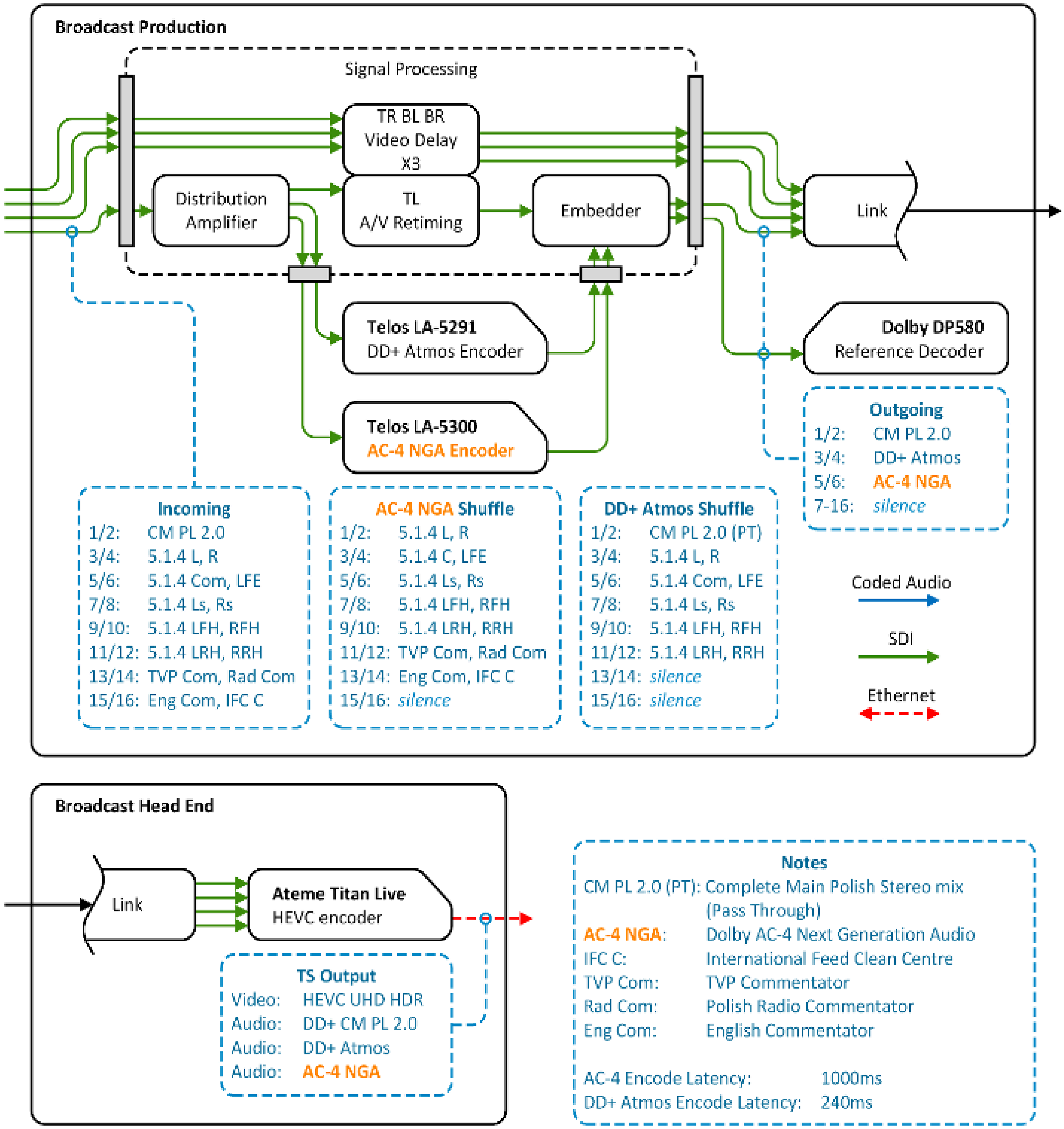


Figure 23. TVP 4k, Live Euro transmission with Atmos

Within the 16-channel SDI Infrastructure, TVP is able to carry a 2.0 Complete Main for legacy purposes as well as a 5.1.4 clean bed and three additional commentaries. The emission codec used takes care of downmixing to 5.1 upon playback if needed.

As all UHD TVs in Poland support DD+ and a large proportion also support AC-4, TVP decided to broadcast two Atmos streams simultaneously: DD+ Atmos with a TVP commentator and AC-4 Atmos with the possibility of personalizing the experience for TV viewers with TVs that support AC-4.

As highlighted in the Figure 23 workflow diagram, only two audio encoders are used. In this case a Telos LA5291 is used for encoding an immersive soundtrack using E-AC-3-JOC [35], and a Telos LA5300 for AC-4 NGA encoding. Both encoders are equipped with an audio-shuffler on the input to handle audio routing if needed. Since this is a workflow without additional dynamic metadata flow, different encoding modes on both encoders can be triggered by GPI. The E-AC-3-JOC encoder will encode a 5.1.4 immersive Complete Main, while the NGA encoder will encode a 5.1.4 immersive bed with three additional commentary substreams.

The resulting experiences are:

- Stadium Ambience Only
- Stadium Ambience + Polish Commentary
- Stadium Ambience + Radio Commentary
- Stadium Ambience + English Commentary

All experiences enable the “Dialogue Enhancement” feature to allow end-users to adjust the dialogue level to their needs.

Once the audio is encoded and embedded into the SDI, the signal will be sent to the distribution hub. At the distribution hub, an ATEME Titan Live encoder is responsible for the video encoding. As the audio is already encoded, the video encoder only needs to pass-through the already encoded audio. The emission encoder feeds DVB T2 free to air transmissions with 49 transmitters and coverage of more than 90% of the Poland population.

According to Polish regulation, NGA capable AC-4 is mandatory for DVB-T2 UHD HEVC IDTV receivers. The receivers enable personalization of the sound reception including soundtrack selection, dialogue enhancement and mixing main with additional audio broadcast as audio objects. It seems that these features are very attractive for the viewers and may be implemented in the main HD services.

Poland’s national plan foresees a switch from DVB-T to DVB-T2 in 2022.

Case Study 2: Dynamic metadata live event

French national broadcaster FranceTV is preparing to launch a UHD service over DVB T2. A key aspect of the service is Next Generation Audio and the use of metadata in open standard distribution formats to author and control the NGA experiences.

In this workflow, the metadata is authored at the original production site and is carried through the workflow to configure the emission encoder at the platform operator's site. Metadata is also used to automatically reconfigure the emission audio encoder when playout switches to pre-recorded advertisements in stereo. The emission encoder feeds DVB T2 (free to air) transmissions.

Figure 24 illustrates the workflow used in this case study. From the Roland Garros tennis arena, both 2.0 and 5.1 mixes are created along with English commentary and the Umpire's audio. In total 14 audio channels are then contributed to FranceTV as multiple audio streams all using ST 302 24 bit PCM audio.

At FranceTV the audio is brought into a ST 2110 network, where the 5.1 is up mixed to 5.1.4 using a Linear Acoustic UPMAX ISC, the French commentary is added when available, and the NGA metadata is authored as S-ADM [124] using PMD studio. All this comes together into an ATEME Titan Live Encoder which contributes 16 audio channels to Cognac-Jay in a transport stream as multiple audio streams all using ST 302 with 24-bit PCM audio and where channel 16 is used for the S-ADM metadata.

The two or three NGA experiences that are created are:

- Stadium Ambience Only
- Stadium Ambience + English Commentary
- Stadium Ambience + French Commentary (when available)

At Cognac-Jay the transport stream feeds the ATEME Titan emission encoder, although during commercial breaks the advert-streams, which also include S-ADM, are spliced in prior to the encoder. The emission encoder feeds both the DVB-T and DVB-S transmissions.

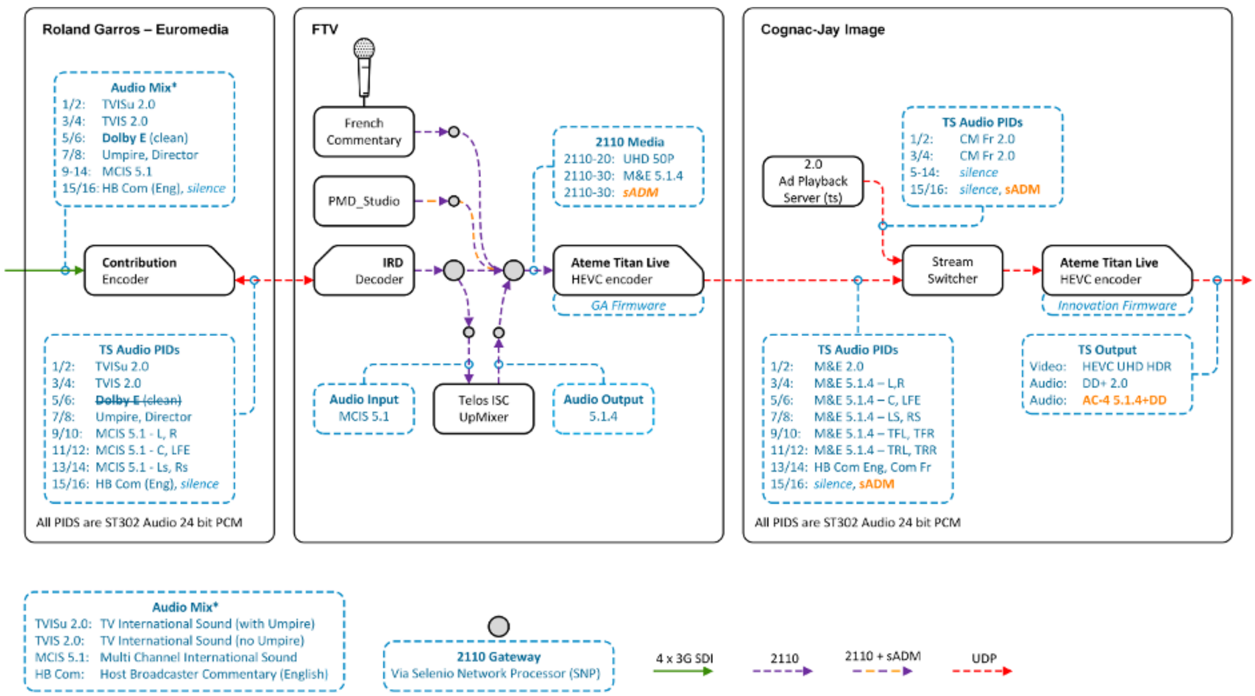


Figure 24. Dynamic Metadata Live Event example workflow

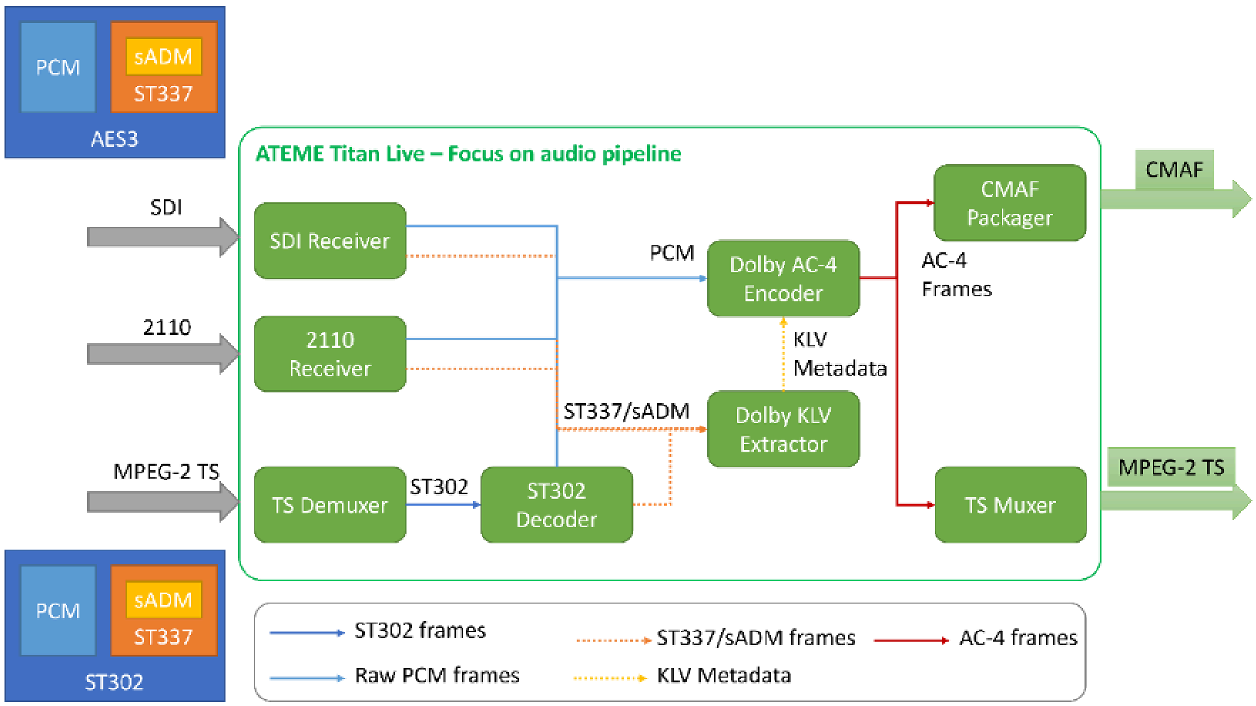


Figure 25. Encoding pipeline using an ATEME Titan Live Encoder

In Figure 25 the audio encoding pipeline is illustrated. For this use case, the following inputs are used: MPEG 2 TS, SDI or SMPTE ST 2110. Independent of the input type, the first step is to extract the raw PCM and the S-ADM ST 337 [36] frames. Those two types of data are then routed differently. Whilst the PCM frames are sent directly to the AC-4 encoder, the S-ADM frames need an additional processing step to extract the KLV metadata. The encoded AC-4 frames are forwarded to a CMAF packager and/or a TS muxer. These units will dynamically generate the signaling either using MPEG-DASH manifest declaration or MPEG-2 TS descriptors.

Case Study 3: Dynamic Metadata in a SMPTE ST 2110 environment

In this workflow, dynamic metadata is used inside a SMPTE ST 2110 [43-47] environment to control NGA experiences for live sports.

At Swiss broadcaster Schweizer Radio und Fernsehen (SRF), the discussions concerning personalized sound experiences and the ongoing aim of optimizing the use of bandwidth found their solutions in the technologies associated with Object-Based Audio. Additionally, SRF stepped into the future of IP-based workflows in broadcast by developing Europe's first all-IP UHD OB Van (EBU Technology & Innovation Award 2019) and building a new Playout and Switching center completely based on SMPTE ST 2110. One of the lessons learned from this innovation was the need for flexible metadata-handling in an IP environment.

Here we will highlight these two main issues and how they got married in a proof of concept in collaboration with Dolby Laboratories Inc.

The choices we made in our proof-of-concept were those available to us and with good support from the vendors and manufacturers concerned. Other solutions are evidently possible.

For our purposes, one of the biggest advantages of OBA is its ability through metadata to separate audio beds and other audio objects such as commentaries in different languages, audio description and so on. The idea of transmitting a blockbuster movie, concert or live sports event with one multichannel audio bed and a variety of three or four different languages plus audio description (AD) seems obvious. Whereas the thought of using static metadata for fixed rendering ratios for these programme elements was not an attractive option for us, the idea of delivering them with dynamic metadata came to mind.

The solution was found by taking the mixing parameters of an audio console (or any other audio processor) and using this data to control an authoring tool generating the desired metadata. Using a standard metadata protocol (in our case, S-ADM) allows the use and routing of these metadata in any standardized broadcast environment.

Open standards and protocols ensure the most interoperability between different products and as technology advances accelerate, they are becoming even more important for us as a broadcaster to embrace.

Test Scenario:

Our test was designed to prove how dynamic metadata could be used for a multilingual sports transmission. Typically, in such a programme we have an international ambience sound object (IS) and several commentary objects. Each object is transported once only, with accompanying metadata, but at the receiver side we have a different representation for each commentary. The appropriate mix for the IS and the individual commentary is controlled through the metadata.

This means that the IS object needs three different gains in the metadata, one for each representation, whilst each Commentary also needs a gain.

Table 11. Representations Available

Representation 1:	IS + Commentary 1
Representation 2:	IS + Commentary 2
Representation 3:	IS + Commentary 3

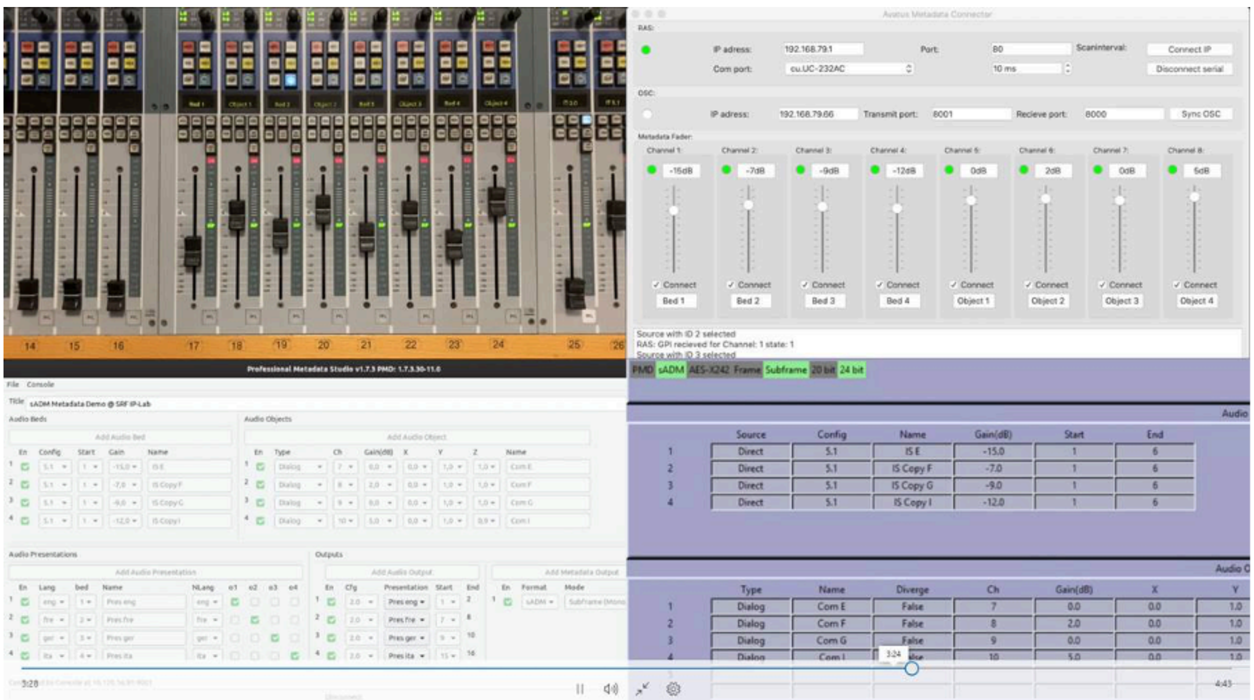


Figure 26. Setup Screen Captures

Notes:

Top Left: Faders on the Desk controlling metadata only
Top Right: Protocol converter between desk and PMD Studio
Bottom left: PMD Studio with different gains for the same IS input
Bottom right: AM Viewer on the receiver side

Transport of dynamic metadata in a SMPTE ST 2110 environment:

S-ADM is an appropriate choice for such object-based audio metadata. It is clear to us that even in a modern facility there is always the need for backward compatibility as legacy SDI infrastructure and links are often encountered. A full IP–based workflow is of course the goal.

To simplify operational use, it is important to be able to generate metadata directly from a mixing desk. As there are nowadays so many controllers and web interfaces in a control room, having yet another UI for the metadata control is not an option to contemplate.

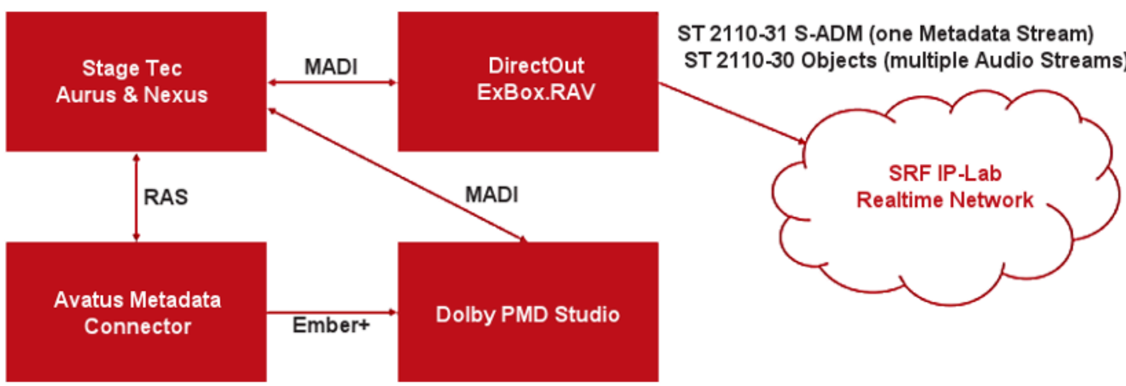


Figure 27. S-ADM setup in the audio control room

As is shown in Figure 26, things rapidly get complicated, even in a very simple setup. The first challenge was to get the fader positions from our console to the metadata authoring tool (Dolby PMD Studio, in our case). There are different protocols possible, but no common standard.

Even with the chosen Ember+ protocol, parameters vary substantially between different products and vendors, as was the case here. As a result, no direct connection was available, and we had to develop our own protocol converter to interface our console with the PMD Studio.

The generated S-ADM is carried from the PMD Studio as an AES-subframe via a transparent MADI link to an AES67 Gateway (DirectOut ExBox.RAV) and is transported as a ST 2110-31 [126] stream through our IP test environment.

Some research is still needed in this field to standardize parameters so as to reduce complexity. There is now ongoing work to develop ADM-OSC, where ADM-compatible messages are carried via OSC[24] . As OSC is widely used in audio gear and with a defined S-ADM profile, the vocabulary of each device is compatible. In local setups, a pure OSC workflow could be a viable solution as many renderers for multichannel PA setups already use OSC.

Signal path overview:

Referring to Figure 28 the delivery from origin to reproduction works as expected, even if there is conversion to and from an embedded SDI link through gateways. The open-source Dolby AM-Viewer decoded the metadata stream without any problems.

The downsides of this method may be summed up as scalability and bandwidth consumption. In a modern IP infrastructure this is not an issue though, but we think it is not a smart solution to transport dynamic metadata of a few kbit/s in a ST 2110-31 stream of approximately 3 Mbit/s. Also, as S-ADM data rates grow, the limitations of a constant bandwidth tunnel can incur even more inefficiencies. A solution with a SMPTE ST 2110-41 (under development) metadata stream will mitigate this problem.

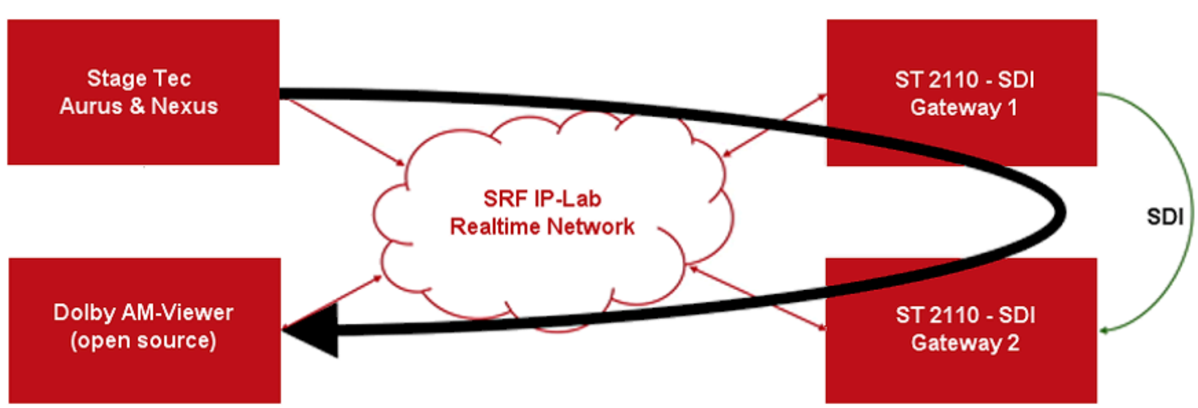


Figure 28. Signal Path overview

[24] Open Sound Control, http://opensoundcontrol.org/ .

12.2.3.3 MPEG-H Audio Workflows

MPEG-H Audio for real-time workflow:

The MPEG-H Audio system [57] is designed to work with today's streaming and broadcast equipment using SDI-based workflows as well as with future IP-based infrastructure. A typical MPEG-H Audio real-time workflow for live productions based on SDI infrastructure appears as in Figure 29. It consists of a chain containing an MPEG-H Authoring Unit and a Video Broadcast Encoder. In real-time scenarios, the authoring of MPEG-H Audio scenes and the metadata are handled by the AMAU as described in Section 25.3.1*{need to check this is incorrect}.*

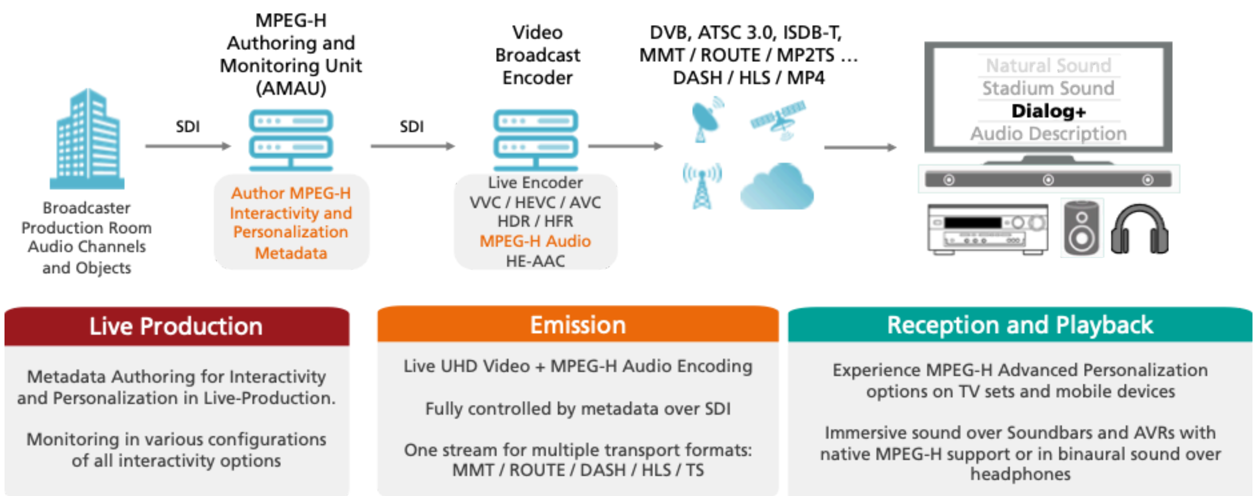


Figure 29. Typical MPEG-H Workflow for Live Productions

MPEG-H Audio for real-time Contribution using SDI-based workflow:

To ensure the integrity of metadata in an SDI environment in any production step, the metadata is delivered over SDI using the so-called "Control Track"[25] . The Control Track is a "time-code like" audio signal and can be treated as a regular audio channel. This ensures the synchronization of metadata with its corresponding audio and video signal. The Control Track is designed to survive A/D and D/A conversions, level changes, sample rate conversions or frame-wise editing. The Control Track does not require audio equipment to be put into data mode or non-audio mode in order to pass through as it was introduced for efficient and robust usage of MPEG H Audio in SDI-based workflows from the beginning.

The metadata for the audio signal is collected into packets synchronized with the video signal and is modulated with analog channel modem techniques into a Control Track signal that fits in the audio channel bandwidth. This signal is unaffected by typical filtering, resampling, or scaling operations in the audio sections of broadcast equipment.

A detailed tutorial on how to handle MPEG-H Audio in live production can be found online in the MPEG H Audio Live Production Tutorial series[26] .

MPEG-H Audio authoring and encoding is also supported in SMPTE ST2110 compliant workflows. According to SMPTE ST 2110 [43-47] , in IP-based workflows the media essences such as video, audio and metadata are transmitted over separate RTP connections and PTP (IEEE 1588, SMPTE ST 2059) is used to synchronize the different essence streams. The transmission of the PCM audio essence (SMPTE ST 2110-30 [46]) is already established andf the transport of metadata including serialized ADM (S-ADM) metadata (SMPTE ST 2110-41[27]) is already standardized.

The serial representation of the Audio Definition Model, according to ITU-R BS.2125 [123] defines a segmentation of the original ADM for use in linear workflows such as live production for broadcasting and streaming applications. Similar to the MPEG-H Control Track, one S-ADM frame contains a set of metadata describing at least the audio frame over the time period associated with that frame. S-ADM has the same structure, attributes and elements as those of ADM, as well as additional attributes to specify the frame format. The S ADM frames are non-overlapping and contiguous with a specified duration and start time.

---

[25] Reference: R. Bleidt et al. "Development of the MPEG-H TV Audio System for ATSC 3.0," in IEEE Transactions on Broadcasting, vol. 63, no. 1, March 2017, https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7874294

[26] MPEG-H Audio: Live and Broadcast Tutorial series, https://vimeo.com/showcase/7538846

[27] SMPTE ST 2110-41, Fast Metadata Framework

For illustration purposes, a simplified ST 2110 Audio and Metadata over IP workflow for MPEG H Audio is depicted in Figure 30.

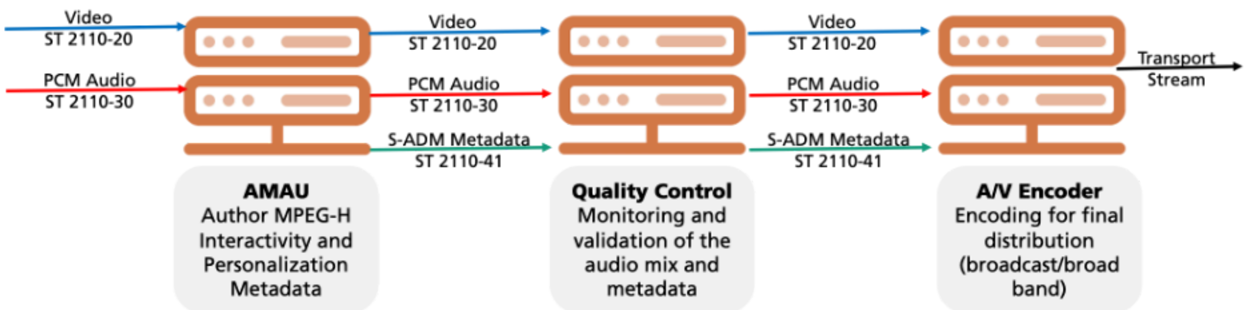


Figure 30. MPEG-H Audio production workflow based on SMPTE ST 2110 (simplified)

For storage and playout of S-ADM based content, the standardization of the transport of PCM audio essence and S-ADM metadata inside the Material Exchange Format (MXF, SMPTE ST 377-1 [36]) is underway at the time of writing. The MXF Format is optimized for content interchange or archiving by creators and/or distributors and provides a complete framework for the transport of NGA. An example application of MXF in practice is shown in Figure 31.

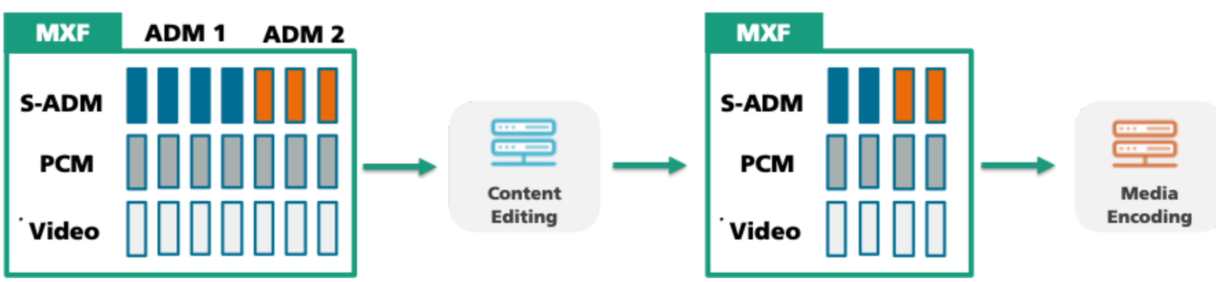


Figure 31. Example of MXF based workflow using ADM metadata

MPEG-H Audio for file-based workflows:

In file-based workflows, MPEG-H content can be created via means of file-based processing using either stand-alone software or integrated in a variety of tools such as DAW, server and platforms. In those tools several processes are executed including measuring the loudness levels of the audio components, setting of the user interactivity options and authoring of the

NGA MPEG-H scene. After completion of the authoring, the metadata and the audio data are exported together as an "MPEG-H Master" file. An example of such implementation is the MPEG-H Authoring Suite, a software suite designed for authoring MPEG-H Audio scenes[28] .

Using the tools of the MPEG-H Authoring Suite or other post-production tools, an "MPEG-H Master" can be exported as a bundle of metadata and audio content. MPEG-H Audio metadata contains all control information for user interactivity and also all necessary information that the playback device needs for reproduction and rendering to ensure the expected experience on any platform.

MPEG-H Master content can be created from multiple source formats including object-based productions, ADM and legacy channel-based material, either by authoring new projects or by ingesting existing material. The ingest step can enhance stereo content with Dialog+ as described above or convert an existing BWF/ADM file (e.g., created with Dolby Atmos Production Suite) to MPEG-H ADM Profile and enhance the content with advanced personalization and interactivity options.

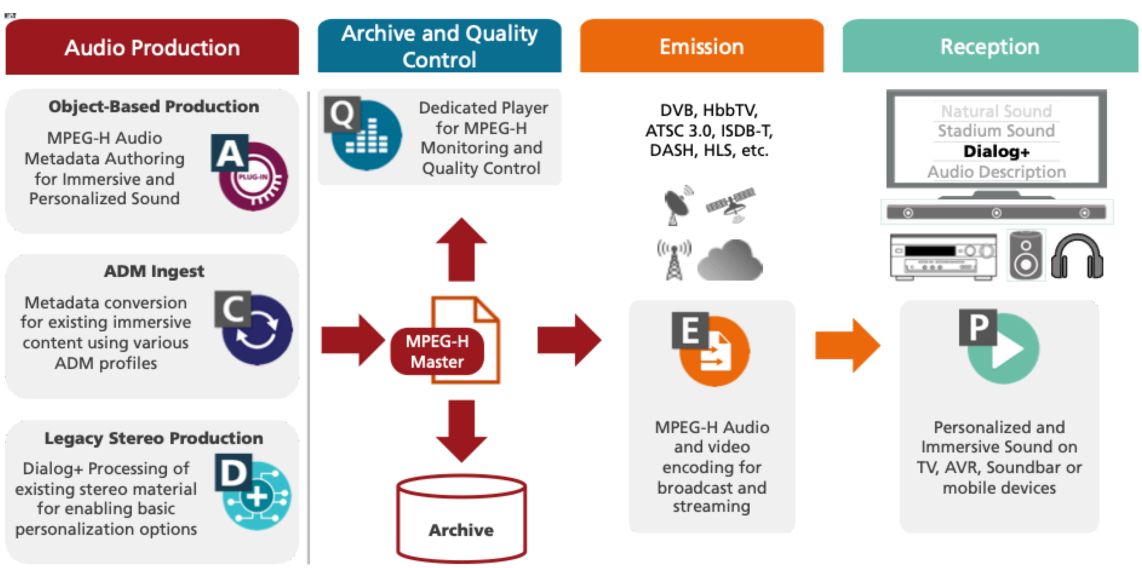


Figure 32. MPEG-H Audio Post-Production workflow (simplified)

The MPEG-H Master can be exported in the following formats:

[28] The MPEG-H Authoring Suite, https://mpegh.com/mas/

- **MPEG-H BWF/ADM:** An MPEG-H BWF/ADM file (short for Broadcast Wave Format with embedded Audio Definition Model metadata) is a multichannel wave-file which contains all the audio and metadata of the MPEG-H Audio scene. The exported BWF/ADM file is compliant to the MPEG-H ADM Profile[29] .

- M**PEG-H Production Format (MPF)**: An MPF file is a multichannel wave-file which contains all the audio and metadata of the MPEG-¬H Audio scene. The metadata is stored in the “Control Track", which is one of the audio tracks in the multichannel wave-file and contains a “time-code like” signal that is robust against sample rate conversions or level changes. The Control Track is used for live productions using SDI-based workflows (see [MPEG-H Audio for real-time Contribution using SDI-based workflows).](#)

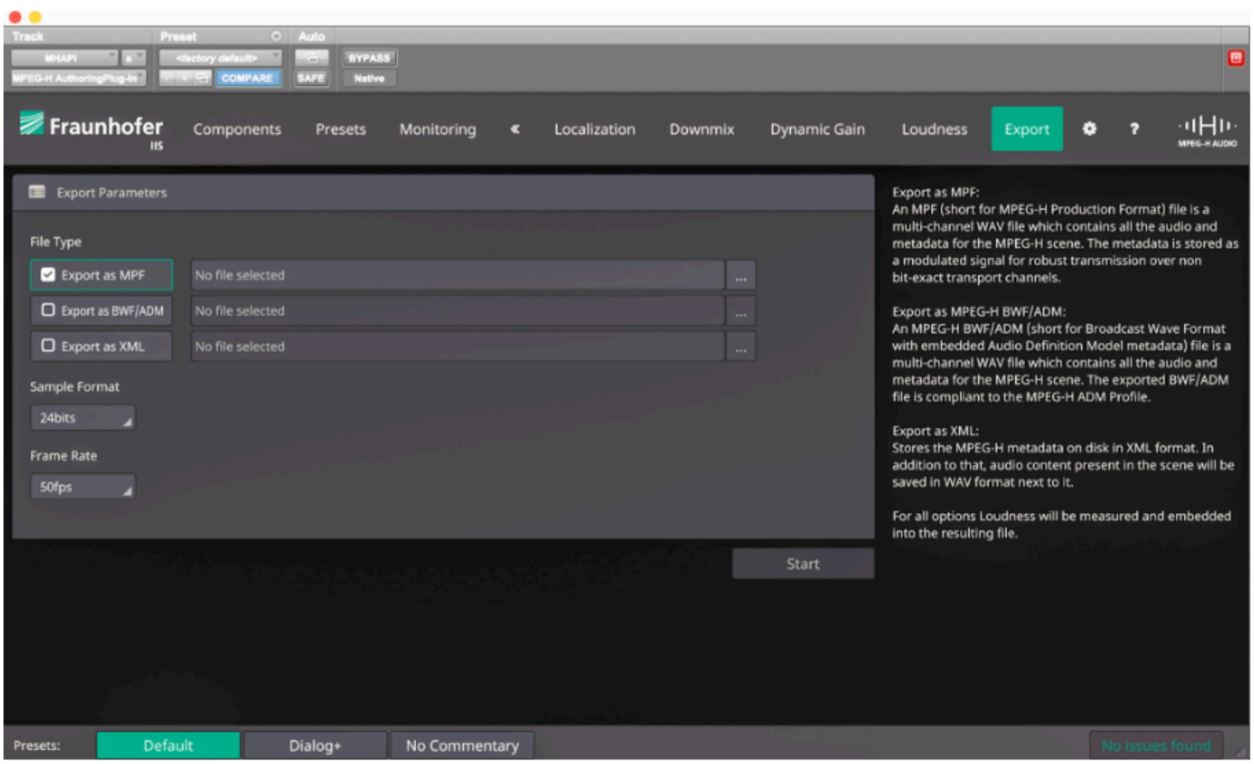


Figure 33. Export window of the MPEG-H Authoring Plug-in

[29] The MPEG-H ADM Profile, https://www.iis.fraunhofer.de/en/ff/amm/dl/whitepapers/adm-profile.html

The MPEG-H Masters can be used for both live and offline encoding or stored for archive purposes. Interoperability is made possible thanks to transcoding tools which allow conversion from and to MPF (MPEG-H master file) and ADM formats.

As shown in Figure 33 the MPF format is suitable for existing SDI-based live workflows and it may be used for all other purposes such as post-production, content exchange and archiving. On the other hand, the ADM format is ideal for content storage, content exchange using MXF based workflows, and with its serialized option for IP-based workflows according to SMPTE ST 2110 [43-47] suite of standards.

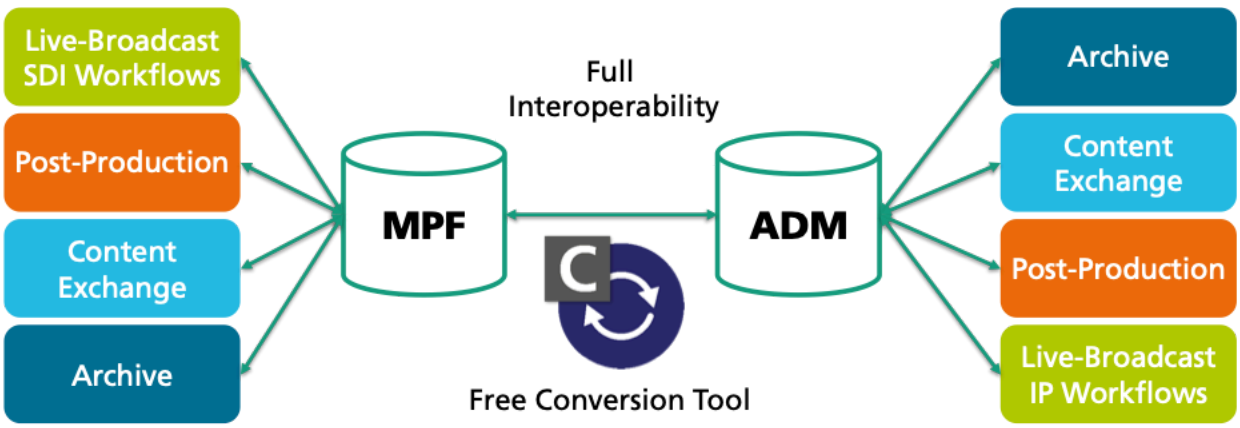


Figure 34. Full Interoperability between MPEG-H Masters and ADM-based workflows

MPEG-H Audio Production Tools:

MPEG-H operations, both real-time or file-based, are supported by a comprehensive variety of hardware and software tools which allow flexible and efficient authoring and monitoring of NGA content. This section provides short descriptions of the most common tools in use at the time of this writing.

Tools for MPEG-H Audio real-time Workflows:

In a real-time production environment, as well as during post-production, a 3D Audio scene needs to be created and monitored. At the same time metadata needs to be generated to describe the audio scene and to enable the features provided by MPEG-H. For the purpose of authoring and monitoring MPEG-H Audio content,as well as exporting the generated metadata, a device class called Audio Monitoring and Authoring Unit (AMAU) has been developed.

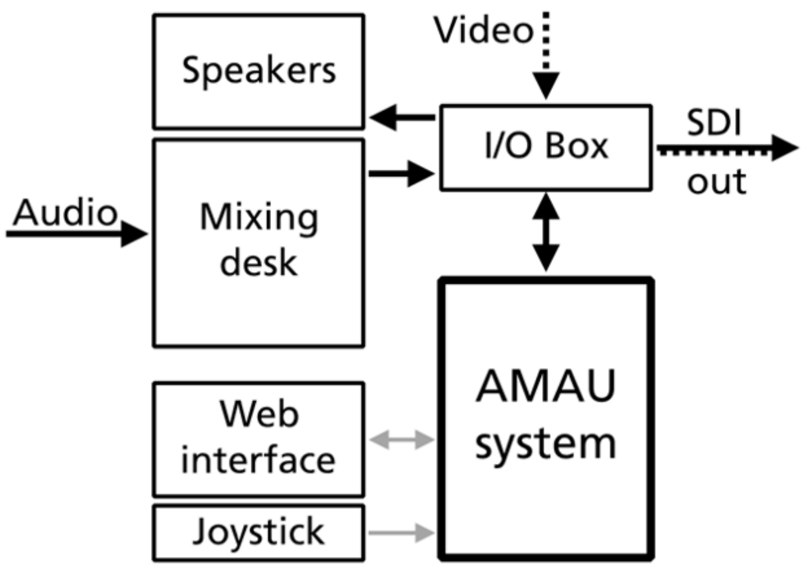


Figure 35. Signal flow of an MPEG-H AMAU System

AMAU systems can be integrated into the existing SDI, MADI or future IP based signal flows which are used for broadcast. In content production, audio signals from the mixing console are fed into the AMAU using I/O converters. Within the AMAU, metadata creation and rendering take place. The AMAU can be controlled using a web interface or hardware controller; typically, the output of the AMAU is an SDI signal including 15 channels of audio and the Control Track on channel 16 (see Figure 35). In an IP based production facility, the audio and metadata would be transmitted in a container over IP. A video signal is passed through for visual reference and for synchronization porpuses. After the audio and video signal leaves the AMAU it can be passed on to the master switch.

An AMAU device can also be used after the master switch for monitoring purposes of the MPEG H scene and its related metadata, right before the emission encoder in order to check the quality of the signal. At this point, usability of interactivity features or loudness parameters can be checked.

There are different solutions available in the market used for authoring and monitoring of MPEG-H Audio metadata in real-time workflows. Some examples are :

- Authoring and Monitoring System by Linear Acoustic (AMS)[30]
- AIXpressor by Jünger Audio (MMA)[31]
- Spatial Audio Designer Processor - Processor by New Audio Technology (NAT SAD-P)[32]
- iAM-12G-SDI by Wohler[33]
- MIXaiR™ by Salsa Sound[34]

The use of AMAU devices allows productions that support all MPEG- H features without changing the entire workflow, as most of existing equipment still can be used and does not need to be replaced.

For the purpose of monitoring MPEG-H Audio content, many different speaker layouts from stereo to 7.1+4H can be connected to an AMAU system. Additionally, all interactivity options and the audio quality can be monitored during production using an emulation of end-user receivers and with different reproduction configurations.

AMAU systems measure the loudness and true peak values of all channels, objects, output buses and formats, as well as every created Preset in real-time. With the resulting data, correction values are added to the metadata stream compliant with the applicable loudness regulation. The measurement of all generated DRC profiles and real-time loudness correction are also included. Additionally, AMAU production tools support the user with visualizations of all crucial measurement values.

A video tutorial series is available to guide you step-by-step through the live authoring, monitoring, and encoding (see footnote 19).

MPEG-H Encoders and Decoders:

Several manufacturers support MPEG-H Audio encoding/decoding in their product families including ATEME, KaiMedia, Sumavision, Mainconcept, DS Broadcast, Spin Digital, Korg, NEC, and Sapec.

Tools for MPEG-H Audio File-based Operations:

---

[30] https://www.telosalliance.com/on-air-tv-audio-processing/monitoring/linear-acoustic-ams
[31] https://www.telosalliance.com/signal-distribution-format-conversion/audio-interfaces/junger-audio-aixpressor
[32] https://www.newaudiotechnology.com/products/spatial-audio-designer-processor/
[33] https://www.wohler.com/product/iam-12g-sdi/
[34] https://www.salsasound.com/products/mixair

Typically, DAWs are used in audio post-production operations. Most common DAWs support the integration of AAX, VST or AU based plugins, which extend the capabilities of the host. Plugins such as the MPEG-H Authoring Plugin – part of the freely available MPEG-H Authoring Suite – or the Spatial Audio Designer (SAD) by New Audio Technology can be integrated into the existing post-production workflows for MPEG-H audio content creation.

When the production workflow allows, file-based operations provide several benefits. Compared to real time panning, post-production operations allow for more advanced object movements because automation can be edited and retrieved. The number of objects that move at the same time can be higher for the same reasons. Another benefit is that they can overcome the DAWs bus width limitation, e.g., enabling a high number of objects to loudspeaker layouts not natively supported in the DAW, such as 5.1+2H, 5.1+4H or 7.1+4H.

Furthermore, to monitor the different presets or switch groups that have been defined, a full MPEG H renderer is included in the authoring tools. Finally, when the mix is finished and all metadata entries are authored, the session can be exported to an MPEG-H Master. During this stage, the loudness of all components and presets is measured offline and embedded into the corresponding metadata fields.

MPEG-H software tools also allow conforming to legacy channel-based audio content. For example, in case a stereo or 5.1 audio mix should be prepared for MPEG-H broadcast, the mix does not need to be modified as only metadata needs to be generated. For this, stand-alone tools such as the MPEG-H Authoring Tool (MHAT), part of the MPEG-H Authoring Suite[35], are available. It is also possible to monitor the created scene and presets and export the final MPEG-H Master.

To control created MPEG-H Masters with or without an accompanying picture, the MPEG-H Production Format Player, part of the MPEG-H Authoring Suite, can be used to ensure that the quality of the mix and the authoring is matching the expectations.

MPEG-H Authoring Suite for file-based Operations:

The MPEG-H Authoring Suite (MAS) is a comprehensive yet powerful set of tools to create MPEG-H Audio content. It covers all steps of the production, delivery, and playback chain. The easy-to-use tools allow authoring, conversion, creating, monitoring, analyzing, encoding, and playback of AV- or audio-only content.. These tools support the MPEG-H ADM Profile, as well

[35] The MPEG-H Authoring Suite, https://mpegh.com/mas/

as binaural monitoring for immersive audio reproduction over headphones. The MPEG H Authoring Suite can be freely downloaded. An exhaustive tutorial on how to use the MPEG-H Authoring Suite is also provided[36]. The main components of the MPEG-H Authoring Suite are:

- MPEG-H Authoring Plug-in (MHAPi)
- MPEG-H Authoring Tool (MHAT)
- MPEG-H Conversion Tool (MCO)
- MPEG-H Production Format Player (MPF Player)
- MPEG-H Info Tool (MHIT)
- MPEG-H Encoding and Muxing Tool (MHEX)
- MPEG-H VVPlayer (MHVP)

While dialog enhancement (DE) can be directly enabled for new MPEG-H Audio productions using object-based audio, it is also important to enable DE for existing stereo content (e.g., archive material). This would offer a consistent user experience and the option to improve intelligibility for hearing impaired audiences. To achieve this, different solutions for Dialog separation can be used with the MPEG-H Audio system. Interoperability is ensured through the BWF/ADM open format.

MPEG-H Audio is also supported in widely-used professional audio software solutions, such asa Minnetonka AudioTools Server, DaVinci Resolve by Blackmagic Design, Nuendo by Steinberg and Pro Tools Ultimate by Avid.

MPEG-H Utility Tools:

MPEG-H Dialog+ demonstrator:

Dialog+ is based on artificial intelligence and reduces the volume of music, sound effects or background noise to put spoken dialogue in the foreground. Developed by Fraunhofer IIS, the technology is easy to integrate into existing production processes. It does not replace the original audio mix, but instead offers an alternative that viewers can switch to if they so choose. Complementing MPEG-H Audio object-based productions, the main application of Dialog+ is increasing the speech intelligibility of existing material transmitted in both legacy and NGA services at the same time, allowing for cost-effective end-to-end workflows.

[36] Video Tutorial MPEG-H Audio Authoring Suite, https://www.vimeo.com/user/41127038/folder/1871919MPEG-H

The Dialog+ technology was tested in a public trial in October 2020 together with German public broadcaster WDR where the audience was able to try out and evaluate their listening experience[37].

MPEG-H Info Tool:

The MPEG-H Info Tool (MHIT)[38], part of the MPEG-H Authoring Suite, provides an overview of all information associated with content masters that are supported by the MPEG-H Audio system. These range from file-related information (e.g., date of creation, file size, etc.) to a representation of the metadata which is carried in the file (e.g.: labels, number of presets, etc.). The Scene Summary provides a list of information with several levels of detail. The unfolded tree view shows all the metadata of an MPEG-H Scene that is relevant for experts who need a deeper insight into the details. In addition to providing information, the MHIT is also equipped with automated tests for ADM-based content.

The MPEG-H Info Tool can help with this task by providing automated tests of ADM-based content. The free tool runs a conformance framework equipped with exhaustive sets of checks derived from ITU-R BS.2076-2 [72], ITU-R BS.2088-1 [73], and ITU-R BS.2125-0 [123] as well as from all ADM profiles related to the MPEG-H Audio system. After analysis, the tool displays a summary of the conformance checks' findings and indicates whether the provided content is supported by the MPEG-H Audio system.

At the time of this writing, additional MPEG-H info tools include BBright UHD-Play and DekTec StreamXpert[39] .

EAR Producton Suite:

The EAR Production Suite (EPS) is a joint open-source development of BBC R&D and IRT consisting in a set of VST plugins for digital audio workstations (DAWs) that enable sound engineers to produce immersive and personalizable content using the Audio Definition Model (ADM) format and to monitor it using the ITU ADM Renderer. The EAR Production Suite enables the import and export of ADM files, compliant to the EBU ADM Production profile. The VST

[37] Experiencing Dialog+ on linear television, https://www.audioblog.iis.fraunhofer.com/dialogplus-on-linear-television
[38] MPEG-H Info Tool, https://www.iis.fraunhofer.de/en/ff/amm/dl/software/mpegh-info-tool.html
[39] BBright UHD Player, https://www.audioblog.iis.fraunhofer.com/nab-2019-mpegh-bbright
DekTec StreamXpert, https://dektec.com/products/applications/streamxpert/

plugins are currently optimized for the Reaper DAW, which features an extension interface that is used to import and export ADM files within a BW64 container.

For more information go to: https://ear-production-suite.ebu.io/

MPEG-H Use Cases:

The MPEG-H Audio system has been successfully integrated into various HD and UHD production workflows. In the following, several case studies, where the MPEG-H Audio system has been used during live broadcast or streaming of major events, are described. This gives a brief overview of the operational requirements and processes required during an NGA production and broadcast.

Globo broadcast using MPEG-H Audio during Rock in Rio:

Rock in Rio, one of the largest music festivals in the world, with over a million people attending, was broadcast live over the ISDB-Tb terrestrial TV service, HLS streaming and 5G using MPEG H Audio in 2019[40] . It was the first live end-to-end broadcast using MPEG-H Audio in Brazil.

Globo, the largest media group in Brazil, successfully used MPEG-H Audio for a live broadcast over the enhanced ISDB-Tb system during one of the world's biggest music festivals, Rock in Rio. Using the MPEG-H Audio format, the musical performances on the two main stages, "Mundo" and "Sunset," were delivered over the air with immersive and personalized sound in the Rio de Janeiro area. In addition, an enhanced video quality was provided using Advanced HDR by Technicolor. A brief overview of the Rock in Rio setup is given in Figure 36

[40] A world's first: Globo successfully used MPEG-H Audio for ISDB Tb terrestrial and 5G broadcast during Rock in Rio,https://www.audioblog.iis.fraunhofer.com/globo-rockinrio-mpegh-isdbtb-5g

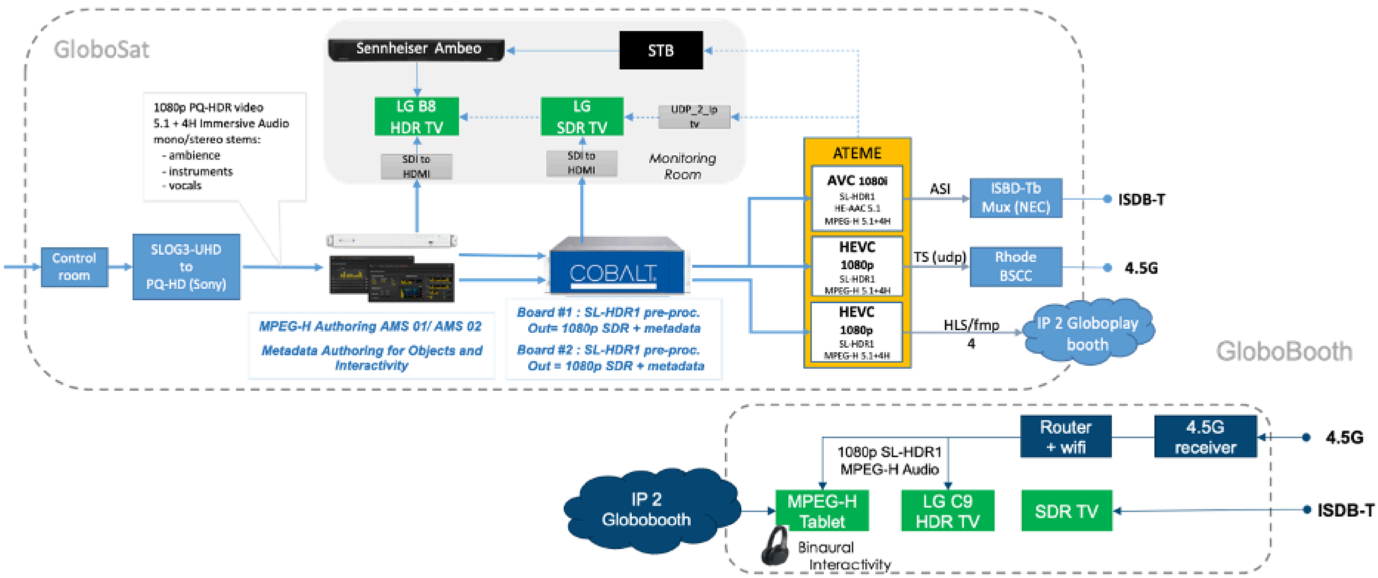


Figure 36. Rock in Rio 2019 setup using MPEG-H Audio

With MPEG-H Audio, the viewers were able to experience MPEG H immersive sound of the concerts as if they were present at the venue and moreover had the opportunity to interact with the audio content and to benefit from its advanced personalization options using tablets and binaural rendering over headphones at the Globo and Globo play booths at the festival venue. In the offices of Globosat, Fraunhofer demonstrated the ease of use of MPEG-H, bringing immersive sound into any living room with the MPEG-H-enabled Sennheiser Ambeo soundbar, which offers studio-grade sound quality.

The immersive mix using 5.1 surround plus four height channels was produced by Globosat sound engineers. Additionally, Globosat made available several mono and stereo stems for ambience, instruments or vocals. Fraunhofer sound engineers used these stems to enrich the live mix with audio objects, enabling personalization of the user experience by enhancing the level of the ambience or moving the vocals into the height loudspeaker layer. Besides immersive audio, personalization is one of the key features of MPEG-H Audio to dramatically improve the attractiveness of a service.

The same MPEG-H Audio production was simultaneously delivered over multiple distribution platforms. Besides the ISDB-Tb terrestrial broadcast, an HLS streaming service using MPEG H Audio was offered, with commercially available LG TV sets playing back the MPEG-H Audio stream. And with the support of Rohde & Schwarz, MPEG-H Audio was embedded in an experimental broadcasting UHF channel for the first 5G broadcast transmission field test in Brazil. ATEME's TITAN Live encoder created all three services in parallel. According to the SBTVD specification, the ISDB Tb service also contained an HE-AAC audio stream, which was used for legacy receiving devices, while the MPEG H audio stream could be consumed with MPEG-H enabled receivers and soundbars.

EBU test with MPEG-H Audio during the Eurovision Song Contest:

A good example of a case study where the MPEG-H Audio capabilities have been tested in a complex and very challenging real-time workflow is the production of the Eurovision Song Contest (ESC) in 2018 and 2019[41] . The Eurovision Song Contest is an international song competition produced by the European Broadcasting Union (EBU) and one of the most viewed non-sporting events worldwide.

Eurovision Song Contest 2018:

Held in Altice Arena Lisbon, Portugal and watched live by 186 million viewers, the production was supported by the national Portuguese broadcaster Rádio e Televisão de Portugal (RTP). The EBU and Fraunhofer IIS collaborated to conduct a field test for immersive and interactive live audio production, based on the MPEG-H Audio system. The goal of the field test was to evaluate future production and reproduction scenarios. The authoring and monitoring of the immersive mixes were done live for demonstration purposes as well as executed in post-production after the event. The immersive mix was not broadcasted in this case but distributed live locally during the event.

The legacy audio production format for the international feed at ESC was a 5.1 surround sound mix and an additional 2.0 stereo mix. Both were created by the host broadcaster's OB van on location. The signals used were a mix of microphone feeds and pre-produced material, such as special sound effects (SFX) and trailers for the participating competitors. In total, 232 microphones were employed by the broadcaster to capture sound from the main stage, hosts, interview partners and arena audience. Based on these signals, international feeds were mixed. 26 ambience microphones were put into different zones throughout the audience, both in front of the stage and on the rear floor area, aiming to the upper tiers and downwards from the PA rigs.

[41] Immersive and Personalized Audio at the Eurovision Song Contest, https://tech.ebu.ch/news/2019/05/immersive-and-personalized-audio-at-the-eurovision-song-contest

Positioned this way, a versatile mix of audience reactions can be created, surrounding the listener by blending more diffuse and more direct ambience signals and giving the mixer the possibility to select the most appropriate sounds. Microphone signals representing the diffuse upper part of the arena, preferably with minimum direct PA sound, were missing in the conventional production workflow to create an immersive experience. For that, a Hamasaki Square microphone was additionally placed underneath the roof in the center of the arena about 25 meters above the floor. Minimizing direct sound from PA and audience by its polar patterns and direction, it provided the upper layer ambience signal by adding just four additional microphones. A 3D audio mixing room was set up, receiving all OB van's signals including sub-mixes and additional commentator feeds from a second service provider, as shown in Figure 37.

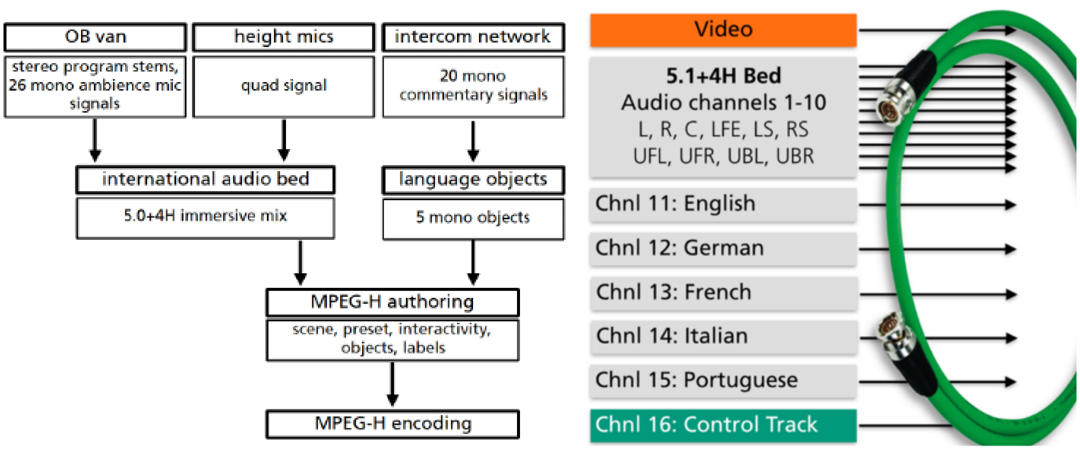


Figure 37. Schematic audio production and authoring workflow of the ESC MPEG-H field test

For creating an immersive, three-dimensional sound impression, the Hamasaki square signals were panned 100% to the upper layer of the loudspeakers, which created the intended "dome" effect. Microphone signals from the upper part of the arena that are panned to the middle layer for regular 5.1 audio production, have been used differently for the immersive mix. Here, they were panned slightly upwards to bring them to their natural position in the sound field, about 15 degrees upwards from the arena floor.

Once the immersive ambience was created, remaining stems like music and hosts were mixed in a legacy way, panning them in the mid-front line of the immersive loudspeaker configuration.

Additionally, a selection of the 20 available commentary feeds were added to the audio scene, whereupon personalization options were configured and monitored.

The resulting 5.1+4H mix could be monitored using an AMAU or DAW with MPEG-H enabled plugins to monitor downmix from 5.1+4H to 5.1 and stereo as well as the binauralized playback. The field test showed that the already existing infrastructure for sound capturing and processing only needs a few additional changes to be able to produce immersive and interactive sound. In the described case, four additional microphones and loudspeakers plus an MPEG-H enabled authoring tool were required. The resulting immersive and interactive audio mix received positive feedback compared to the legacy stereo or 5.1 production, as stated by involved producers and audio engineers during on-site demonstrations.

Eurovision Song Contest 2019:

For the 2019 Eurovision Song Contest (ESC) in Tel Aviv the EBU and its technology partners did a parallel immersive live production of the contest, which was mixed and transmitted using the MPEG-H Audio NGA system. Compared with the 2018 edition, in 2019 the focus was on demonstrating a complete workflow including the transmission part. EBU members on site in Tel Aviv were able to experience the MPEG-H Audio enhanced show as a DASH livestream and as a DVB-compliant live broadcast stream. The latter was also delivered live to broadcasters abroad via the Eurovision FINE network. The "produce once, deliver anywhere" feature is one of the great advantages of the MPEG-H Audio system.

During the ESC 2019 broadcast trial Fraunhofer's sound engineers used the System T S500m broadcast console provided by Solid State Logic to produce the 5.1+4H immersive mix of the show and handle the additional host and commentary signals in four different languages as audio objects.

Figure 39. Audio Monitoring Room for Immersive Sound Production during ESC 2019

The MPEG-H immersive and interactive mix was done in parallel with the main broadcast program production. The MPEG-H production room (see Figure 39) was located next to the main OB Van, receiving most of the microphone feeds together with OB Van stereo mixes. Similar to the event in Lisbon, Fraunhofer IIS sound engineers have placed additional Hamasaki Square microphones on the ceiling of the arena for better ambience capturing (see Figure 40).

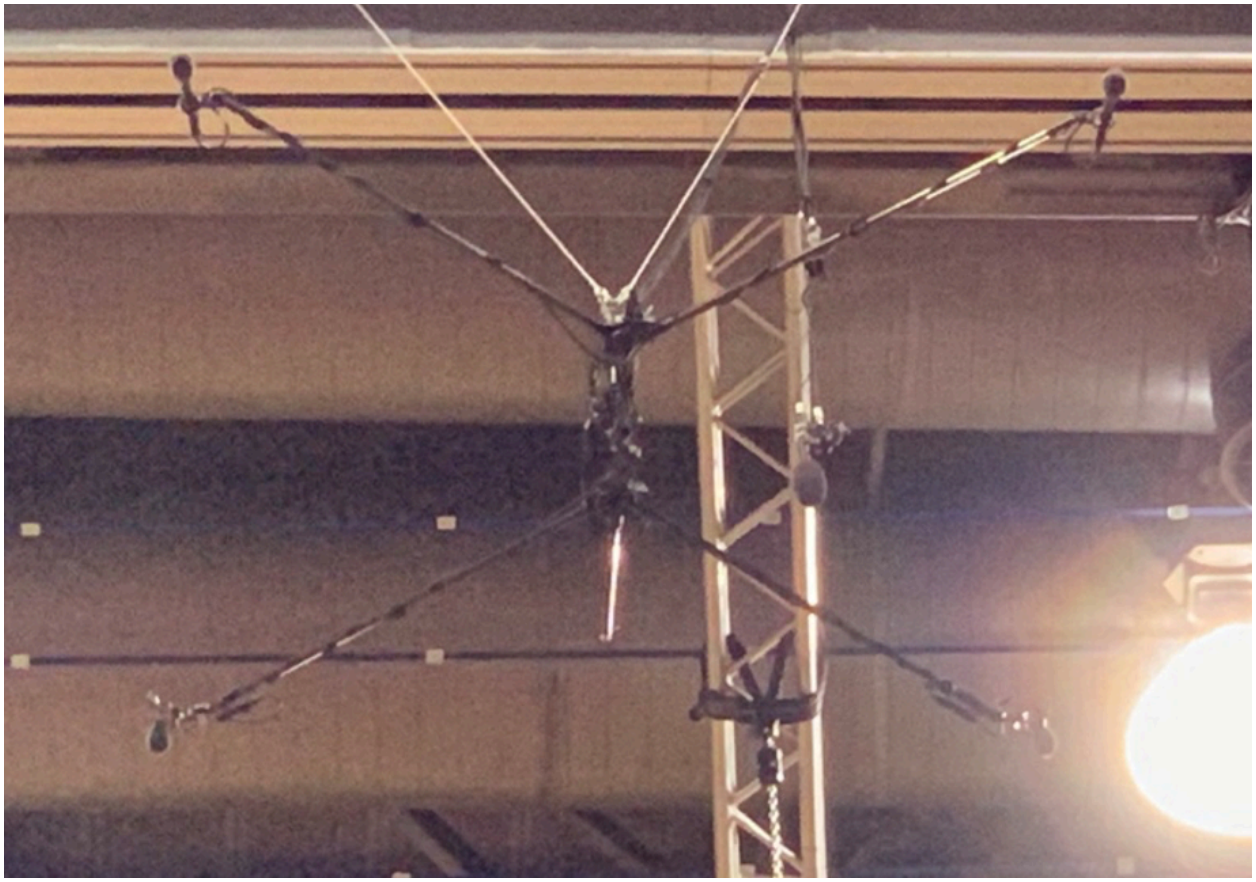


Figure 40. Additional Hamasaki Square and Ambience microphones on the arena's ceiling

The Hamasaki Square was mounted in the middle of the arena at 26-meter height, above the coverage of the PA system. This delivered a decorrelated ambience with as little music spill as possible and the final 5.1+4H immersive mix was created by pan and delay of the ambience microphone signals and up-mix of the vocals and effects via an Illusonic IAP16. The mix was monitored live using a Genelec 5.1+4H speaker setup.

The MPEG-H metadata was authored live using both Telos Alliance Linear Acoustic AMS and Jünger Audio MMA. One AMAU was used for the main production while the second one was used as fallback and for ad-insertion. The metadata was authored for offering the user multiple options to choose between different versions of the content (presets): a default preset, a dialog enhancement preset and a venue preset reproducing the experience of being present in the arena. For each preset, various interactivity features were enabled, such as level and position interactivity.

During production, the sound engineers were able to create new presets "on the fly", change the interactivity ranges for any available audio object, disable interactivity for some presets or even completely switch to a different configuration in a seamless way. ATEME's TITAN Live UHD broadcast encoder was used to create a DVB-compliant broadcast transport stream and, at the same time, DASH and HLS streams for delivery over the internet.

Figure 41. MPEG-H Audio Immersive and Interactive demonstration during the ESC 2019

The MPEG-H programs were demonstrated in Tel Aviv, Geneva, and Madrid, using the Sennheiser Ambeo soundbar (see Figure 41) able to decode and playback the MPEG-H streams for creating an immersive experience without the need to place loudspeakers all around the room. Additionally, the viewers were able to switch between different versions of the content, select their preferred language and level of dialog, as well as use the advanced menus to adjust the balance between the ambience of the arena and the dialog to their personal preference. Moreover, during the demonstration visitors had the chance to move the audio objects “away from the screen” to another part of the room.

MPEG-H Audio broadcast over DVB-T2 during the French Tennis Open:

In 2018 and 2019, France Télévisions set up a dedicated Ultra HD event channel during the French Tennis Open in Paris using MPEG-H Audio for producing a sound immersive and personalized user experience[42] . The Ultra HD channel was offered to French viewers over DVB-T2 jointly with TDF and at the same time over DVB-S2 with the support of Fransat. The 4K resolution and high dynamic range (HDR) program was associated with MPEG-H Audio immersive and interactive features. Moreover, an HLS streaming service was setup and using an LG TV set supporting MPEG-H Audio, the streaming services was showcased in Paris.

During the tournament, France Télévisions provided a 5.1+4H immersive sound bed for the court atmosphere together with an additional object for the commentary in French, allowing the viewers to personalize the content.

As described in the articles listed in footnote 31, on site the production truck of France Télévisions (La Fabrique – Toulouse) produced in parallel the international HD signal in addition to the Ultra HD video signal. The sound from the Philippe Chatrier court was captured by means of ultra-directional microphones which cover the field, the players and the referee, associated with several microphones placed in the stands, including an ORTF-3D mic from Schoeps which serves as main sound capture for the ambiance.

To generate the metadata necessary for the adaptation and interactivity of the Next Generation Audio MPEG-H signal, the MPEG-H Authoring and Monitoring unit from Jünger (MMA) was integrated in the OB van, as shown in Figure 42.

[42] Successful Terrestrial And Satellite Reception Of MPEG-H Audio During The Roland Garros French Open, http://idfrancetv.fr/successful-terrestrial-and-satellite-reception-of-mpeg-h-audio-during-the-roland-garros-french-open/

Roland Garros 2019 – An Ultra HD event channel with MPEG-H Audio: http://idfrancetv.fr/roland-garros-2019-an-ultra-hd-event-channel-with-mpeg-h-audio/

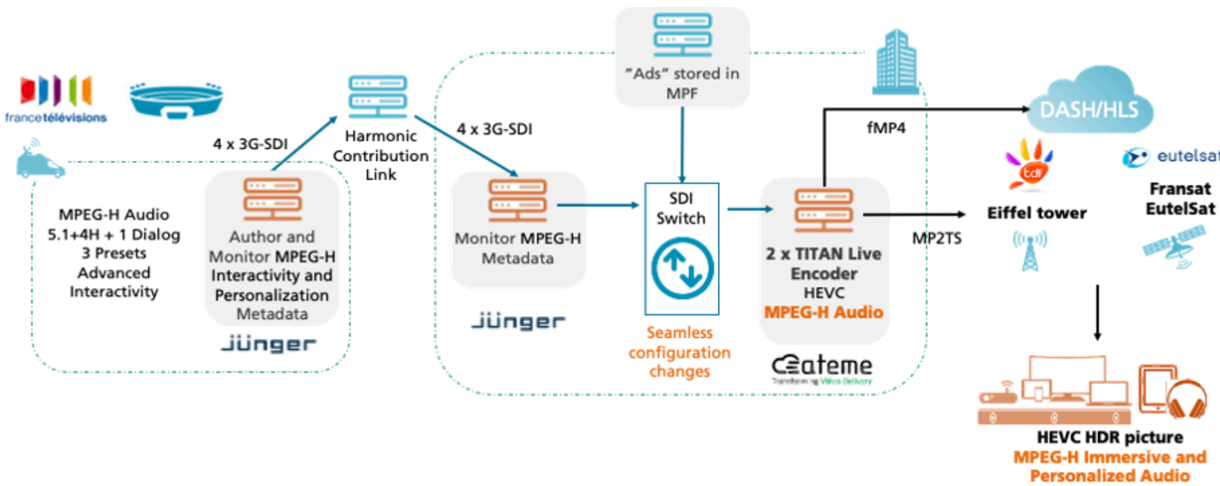


Figure 42. Setup of the UHD FTV Channel using MPEG-H Audio during French Tennis Open

The Ultra HD signal and the associated audio were then transmitted over IP by a single-mode optical fiber from the venue to France Télévisions headquarter. As shown in Figure 43, the MPEG-H Audio system allowed for seamless ad-insertion during the tournament. Pre-recorded clips were played back during the break of the live feed and every transition was seamless for the viewers.

It is the final control room of France Télévisions, the CDE (Center of Diffusion and Exchange) which recovered the signal in order to realize the switch of live and replays, trailers, advertisements, insertion of logos, scrolling texts, etc. then adapted the characteristics of this chain to the different distribution networks.

As shown in Figure 43 the MPEG-H Audio feed was demonstrated at the Philippe Chatrier court, where the various personalization features could be experienced:

- Switch between different mixes: default mix, dialog enhanced version as well as a venue version without a commentator;
- Enhance the dialog level or even change the position of the dialog in the 3D space.

Figure 43. MPEG-H Audio during French Tennis Open 2019: Demonstrations at the Philippe Chatrier court

All interactivity options were controlled by the sound engineer in the OB-van using Jünger's MMA. In the control room at the FTV broadcast center (see Figure 44), several ATEME Titan encoders have been used for encoding the various feeds. Additionally, a soundbar have been installed for reproducing the MPEG-H immersive and interactive sound.

The MPEG-H Audio stream was included in the regular live Ultra HD broadcast of the Philippe Chatrier court.

Figure 44. MPEG-H Audio during French Tennis Open 2019: Control Room

Multiple MPEG-H Audio programs at the European Athletics Championships:

In 2018 the EBU, together with EBU members BBC, France Télévisions, IRT, RAI and ZDF, as well as several technology provider partners, conducted a complex broadcast trial in Europe at the European Athletics Championships in Berlin[43]. The trial demonstrated the world's first live production and distribution of Ultra High Definition (UHD) content, with High Frame Rates (HFR), High Dynamic Range (HDR), and Next Generation Audio (NGA). The Fraunhofer IIS team, together with technology partners, supported the onsite sound engineers of EBU member stations in order to familiarize them with the features and the new possibilities of MPEG-H Audio.

The various feeds were multiplexed for a live transmission via the Eurovision satellite network to RAI's experimental test bed in the Aosta Valley, Italy, and via the Eurovision fiber infrastructure to the European Championships Broadcast Operations Centre (BOC) based at BBC Glasgow.

MPEG-H Audio was used for several different feeds and the Fraunhofer IIS engineering team worked closely with technology partners ATEME, Jünger Audio, MediaKind, Kai Media, b<>com

[43] SMPTE Periodical – European Athletics Championship: Lessons from a Live, HDR, HFR, UHD, and Next-Generation Audio Sports Event, https://ieeexplore.ieee.org/document/8673671
MPEG-H Audio Trial During European Athletics Championships 2018, https://www.audioblog.iis.fraunhofer.com/mpegh-trial-eac-ebu

and Qualcomm to demonstrate the complete set of features and capabilities of the MPEG-H Audio system (see Figure 45).

Figure 45. European Athletics Championships 2018 in Berlin: Control Room

Various devices from different manufacturers supported MPEG-H Audio work in the real-time chain and were fully interoperable, including authoring and monitoring units and post production plug-ins, contribution links as well as emission and playback devices, and demonstrating the benefit of open standards.

As illustrated in Figure 46, three different MPEG-H Audio services were running in parallel, with content produced and delivered live, using available equipment from different manufacturers operated in existing SDI-based broadcast facilities.

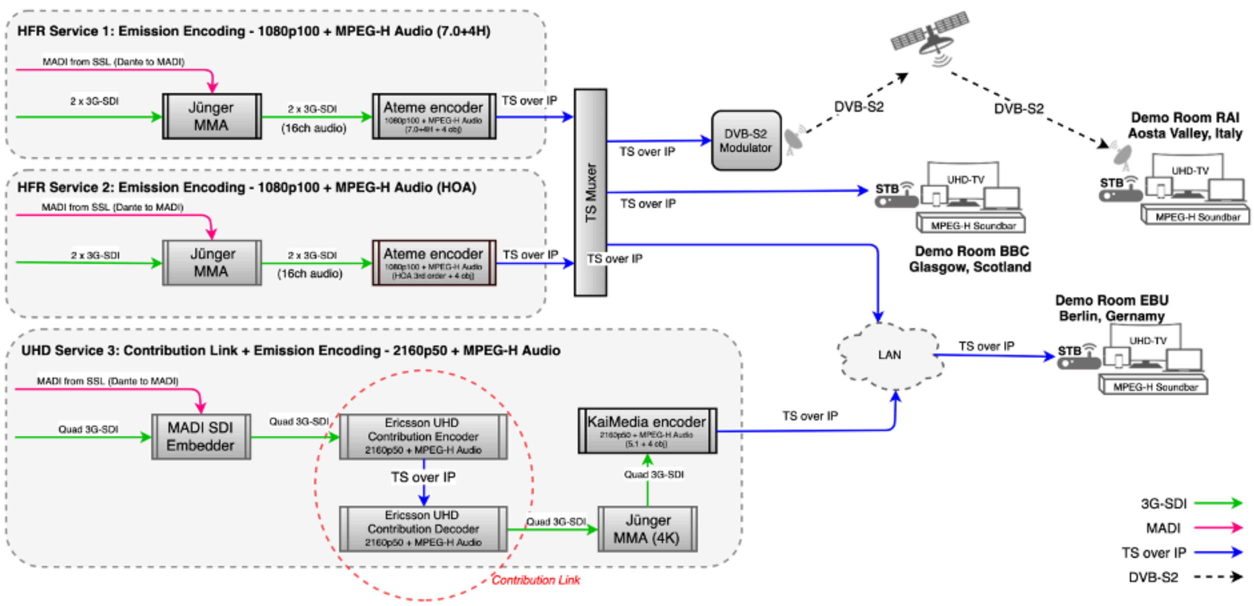


Figure 46. MPEG-H Audio Transmission setup during European Athletics Championships 2018

Service 1: "MPEG-H Audio with 1080p100 HFR and HLG video":

The main MPEG-H Audio service was designed to show the benefit of immersive and interactive sound together with High Frame Rate (HFR) and High Dynamic Range (HDR) video for sport events. This offered the viewer a truly immersive experience. The NGA mix (7.0+4H) was created by BBC, France Télévisions, RAI and ZDF sound engineers using SSL's new console with support for 3D Immersive Audio.

In addition to the immersive sound bed, four interactive mono object signals were used for two commentaries and two audio descriptions in English and French. The live metadata authoring of the MPEG-H immersive audio content was made possible through the use of the Multichannel Monitoring and Authoring (MMA) system provided by Jünger Audio. The personalization and interactivity options were enabled through the MPEG-H Audio metadata, which treats individual elements of an audio mix as objects that can be adjusted by the viewer afterwards.

For the live broadcast, the TITAN encoder was provided by Ateme, which supports MPEG-H Audio and enabled the use of immersive and personalization features of NGA in the program production and distribution.

Service 2: “MPEG-H Audio (HOA) with 1080p100 HFR and HLG video”:

Together with Qualcomm, a second MPEG-H service was provided, using scene-based production tools offered by b<>com. The MPEG-H Audio native support for Higher Order Ambisonics (HOA) allows the use of a scene-based bed instead of a channel-based one, so for this service the immersive mix was created using a scene-based (3rd order Ambisonics) bed together with the same additional four objects.

Similar to the main service, a second Jünger MMA system was used for metadata authoring, together with an additional Ateme encoder for the emission.

Service 3: “MPEG-H Audio with 2160p50 HLG video, including a contribution chain”:

The third service enabled the usage of a contribution link based on MPEG-H Audio before the UHD emission encoder. The purpose of a contribution link is to provide compression which is relatively transparent and provides robust and secure transport of the metadata, allowing downstream manipulation of the signals and of the metadata. The contribution link consisted of the AVP 2000 Contribution Encoder and MediaKind Content Processing (MKCP) Decoder, both devices manufactured by MediaKind. The contribution encoder provides certain authoring capabilities, allowing the broadcaster to manipulate the metadata to be fed into the bitstream. This is done through the GUI of the encoder.

For the emission part, the Kai Media emission encoder was used due to the Korean company encoder manufacturer having integrated MPEG-H Audio for use in the deployment of ATSC 3.0 UHD in Korea. Further, this encoder enabled DVB transmissions by incorporating MPEG2-TS for the transport layer.

A third MMA system was provided by Jünger Audio and used between the contribution decoder and the emission encoder for monitoring of the audio scene. The service also provided a different audio setup compared to the other two MPEG-H programs, using a 5.1 downmix together with the two commentaries and audio descriptions provided in English and French.

Metadata Authoring and Playback:

For all three services, the MPEG-H metadata that was authored enabled users to personalize their audio experience. First of all, they had the option to choose between different versions of the content (presets): a default preset, a dialog enhancement preset and two presets for the audio description. In addition, various interactivity features such as level and position interactivity, were enabled for each preset. During production, the sound engineers had the opportunity to change or adjust live all interactivity features using the MMA systems (see Figure 47, left). They were able to create new presets on the fly, change the interactivity ranges for any available audio object, disable interactivity for some presets, or even completely switch to a different configuration in a seamless way.

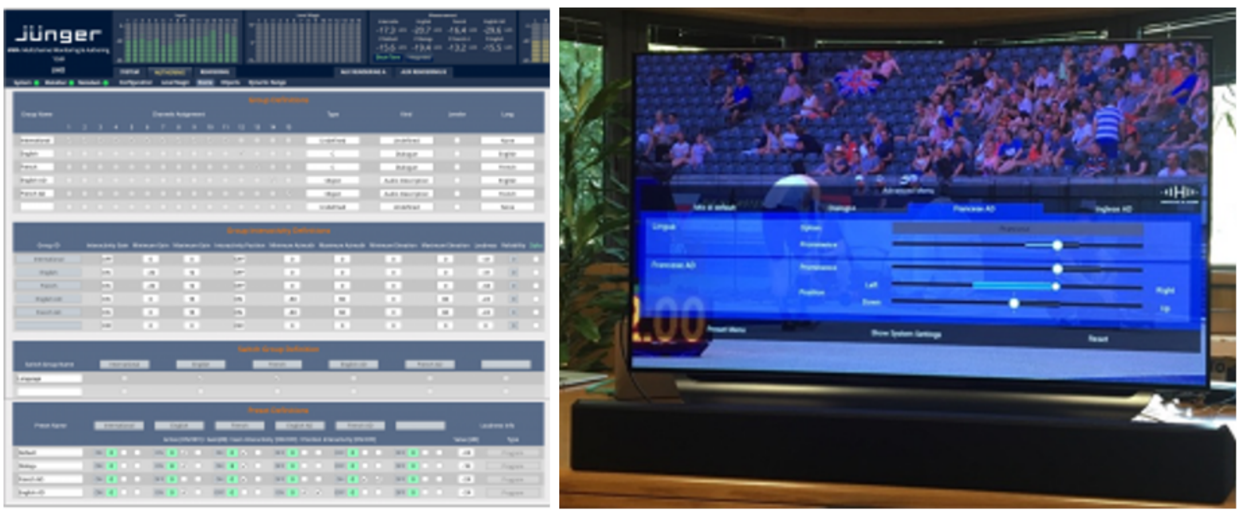


Figure 47. European Athletics Championships 2018 in Berlin: Control Room

The MPEG-H programs were demonstrated in Berlin, Glasgow and the Aosta Valley, using an immersive soundbar that was able to decode and playback the MPEG-H streams and thus generating an immersive experience without the need for loudspeakers all around the room (see Figure 47, right). Additionally, with MPEG-H Audio the viewers could switch between different versions of the content, select their preferred language and level of dialogue and use the advanced menus to adjust the balance between the stadium sound and commentary to their personal taste. Moreover, during the demonstration the users were also able to move the audio description “away from the screen” to the left/right or above, resulting in a higher separation from the main dialog localized in the center channel.

12.2.3.4 Workflows Using S-ADM in Production

The BBC, L-Acoustics, and Dolby trialed the use of S-ADM in live production at the Eurovision Song Contest 2023[44].

The test chain

The live production of audio was trialed using raw audio feeds from Eurovision in Liverpool, UK, all the way through to domestic output devices in laboratories at BBC R&D. The trial was performed in the BBC's immersive listening room[45] in Salford, UK, and getting audio feeds from Liverpool was possible.

The diagram in [Figure 48](#) shows audio being sent from the venue's mixing desk in Liverpool via London to Media City in Salford. Everything else in the diagram was set up in the BBC's Media City lab. Dashed boxes show whether it was BBC, L-Acoustic, or Dolby technology in use.

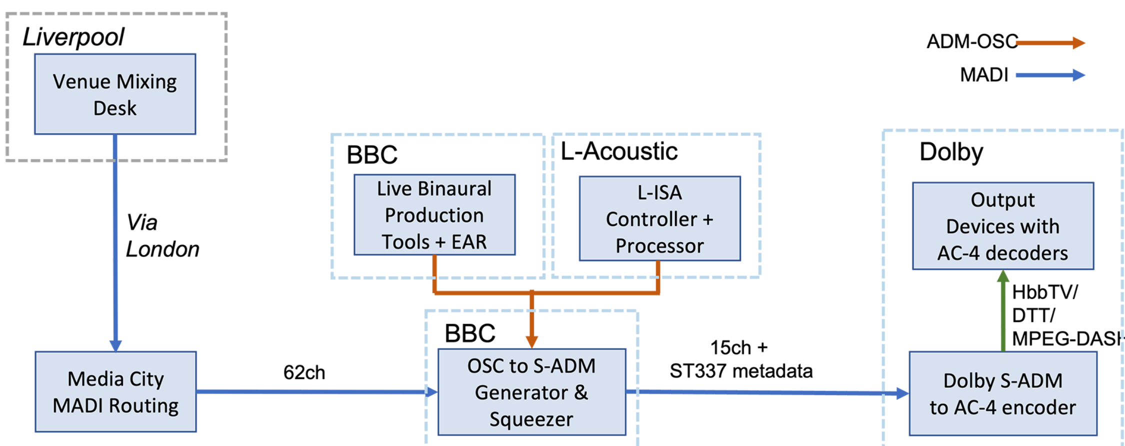


Figure 48: The Test Setup

[44] https://www.bbc.co.uk/rd/blog/2023-06-eurovision-next-generation-audio
[45] https://www.bbc.co.uk/rd/blog/2015-09-audio-research-update-number-6

The audio was received in the listening room over MADI (Multichannel Audio Digital Interface) and consisted of 62 channels of a variety of different signals from the mixing desk at the venue. These included the music, presenters on stage, pre-recorded effects, TV and radio commentary feeds and an array of audience microphone feeds to give atmosphere. S-ADM metadata was generated to identify the purpose of each feed (or channel) and how to render it in the final user experience. This included gain information to describe how loud each feed should be in the mix, and position information to describe where each feed should be perceptually placed in the mix. This S-ADM metadata is carried as data in an additional audio channel to an encoder.

The NGA emission encoder delivers the content to consumer devices. The 62 channels of audio with S-ADM metadata are too much for an NGA emission encoder, which are more limited in the number of channels they can support. Therefore, the channels had to be 'squeezed' down to no more than 15 audio channels, as well as S-ADM metadata containing the correct information for the emission encoder. A squeezer is a processor that converts an input set of audio channels and associated S-ADM metadata into a smaller number of output audio channels and simplified S-ADM metadata. A squeezer may be largely automated, but may require some configuration and/or operator input.

A squeezer was specifically designed for the event, so not generalised for all types of conversions to NGA emission encoders. This particular squeezer performed a simple panning of all the object-based channels to a 5.1.4 channel layout. All the channel-based beds were mixed into the same 5.1.4 channel layout. The object-based commentary feeds were kept separate to allow them to be interacted with.

The squeezed down audio and S-ADM metadata were then sent to the Dolby encoding and playout system, where it was AC-4 encoded before being decoded by the consumer devices in the lab.

Production using ADM-OSC

The production from the 62 channels of audio was handled by L-Acoustic's L-ISA Controller and Processor software. The L-ISA Controller allows a sound producer to set position and gain parameters for each of the input audio channels (which were treated as audio objects) with a straightforward user interface (see Figure 49). The Controller outputs real-time positional metadata using the Open Sound Control (OSC) protocol. The ADM-OSC flavor of OSC was used, which contains ADM-compatible messages for easy interfacing with ADM tools.

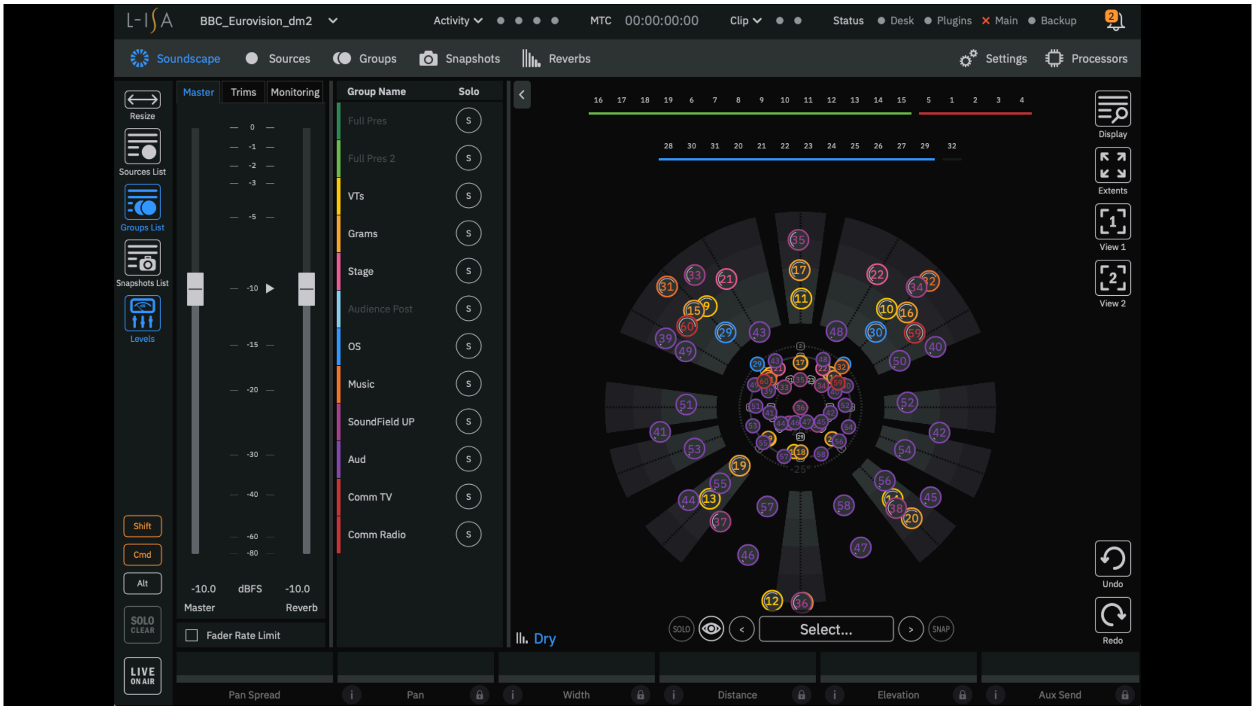


Figure 49: L-ISA Controller user interface

The L-ISA Processor receives the input audio and the ADM-OSC metadata, and can render the audio to suit a specified loudspeaker layout. This allows the producer to monitor the mix in fully immersive audio, and position the 3D location of the audio objects in real-time.

As well as the L-ISA setup, the BBC's [Live Binaural Production Tool][46] was also used (see [Figure 50]). This was originally built for binaural productions of the BBC Proms and adapted to output ADM-OSC for this trial. Like L-ISA, it also allows audio objects to be positioned in 3D space.

[46] https://www.bbc.co.uk/rd/blog/2019-07-proms-binaural

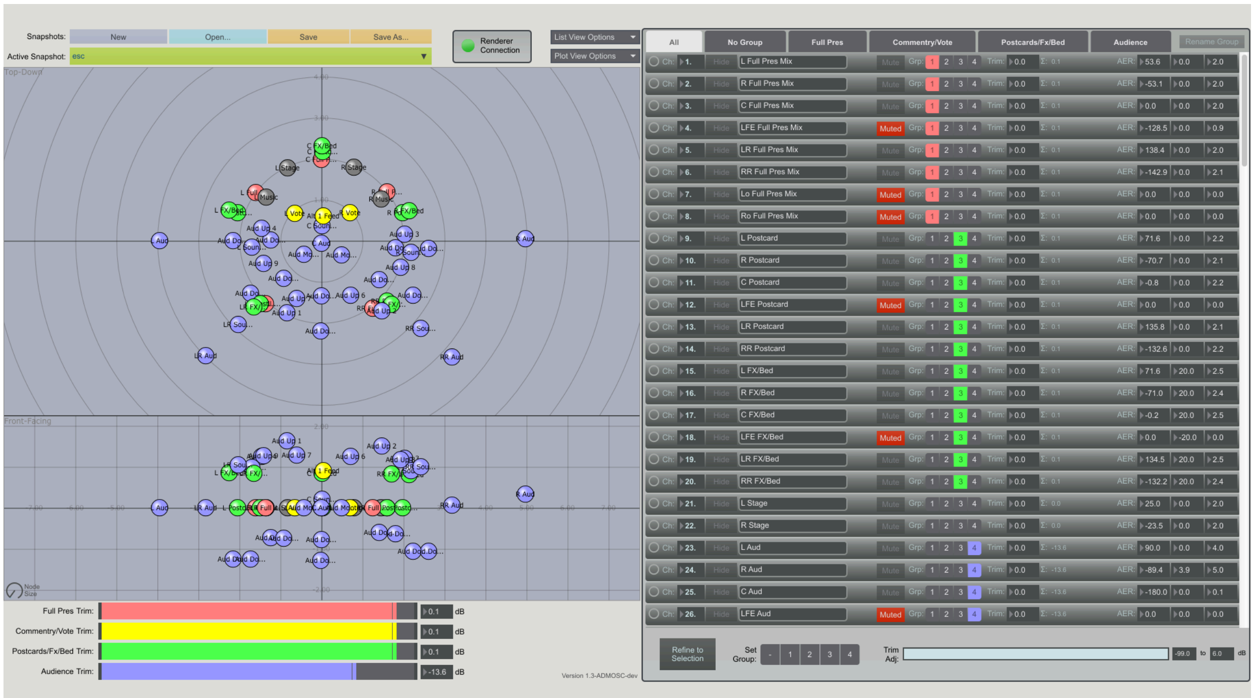


Figure 50: BBC's Live Binaural Production Tool

Either tool can connect to the S-ADM generator and squeezer, showing that any software that generates ADM-OSC can be used in the production set-up. It is also possible to render the audio to an array of speakers via the EAR[47] (EBU ADM Renderer), using some additional software that was developed to accept ADM-OSC in to EAR.

Interfacing with the Dolby AC-4 emission system

It was necessary to send the S-ADM metadata and 15 channels of audio to the Dolby AC-4 encoder, so a method of carrying that signal was required. The SMPTE has been working on several different methods of carrying audio, video and metadata. For this trial, the SMPTE ST 337 [36] and SMPTE ST 2116 [124] standards were used, which allow S-ADM metadata to be carried as data in an audio channel. In the trial setup, the first 15 channels of the output carried audio, with the 16th carrying the compressed S-ADM metadata.

[47] https://www.bbc.co.uk/rd/blog/2019-08-libear-ebu-audio-renderer-next-generation-open-source

An Ateme Titan Live encoder converted the ST337 input into a Dolby AC-4 bitstream, which was then delivered in a DVB-T transport stream for over-the-air TV reception. It was also sent via a DASH server (Dynamic Adaptive Streaming over HTTP) for internet-based delivery to both a smartphone and to a television supporting HbbTV (Hybrid broadcast broadband TV) in the lab.

The presented programme

Among the 62 audio channels fed into the lab were the BBC One (television) and BBC Radio 2 commentary feeds. This enabled the trial to provide the option to choose between the commentaries using the user interface on the TV or a smartphone (see Figure 51). The rest of the audio (including the music, pre-recorded videos and audience noise) was mixed by the squeezer to a 5.1.4 channel layout, providing an immersive experience on capable devices. Since the television only had a basic pair of built-in speakers, it rendered the programme in stereo. However, by hooking up an Atmos-enabled soundbar to the television, it was possible to enjoy a more enveloping experience, demonstrating how the content can adapt to different devices.

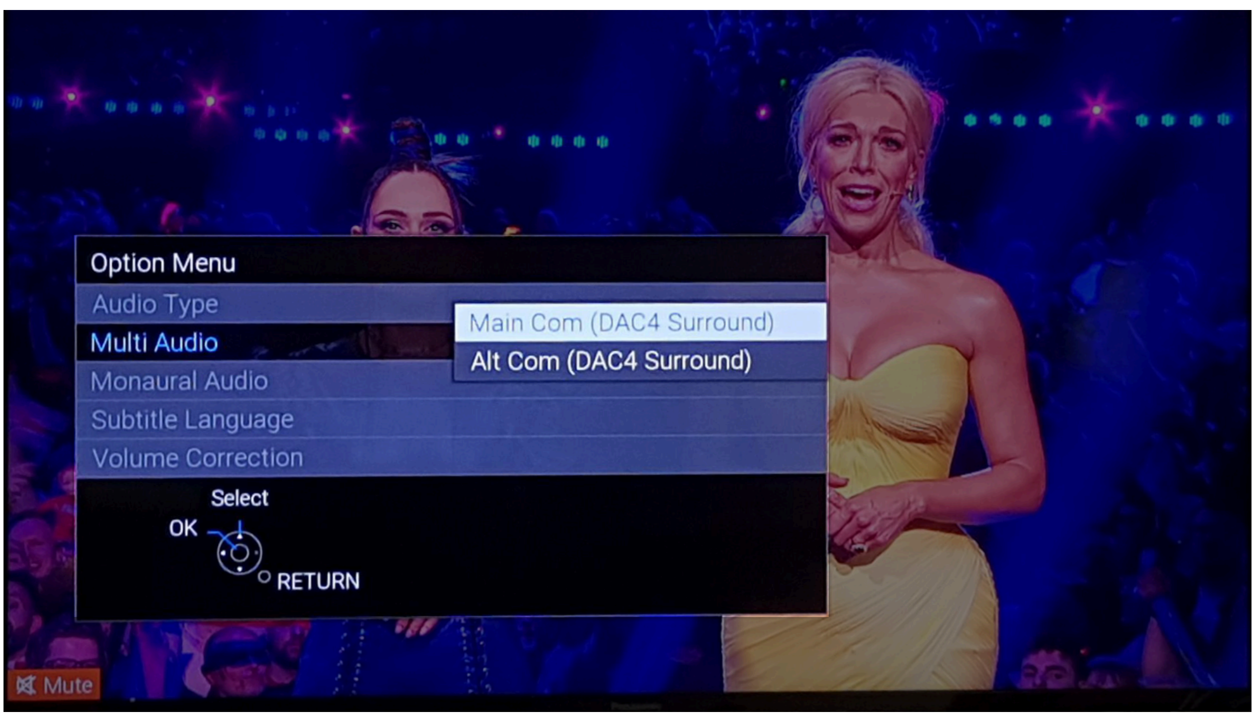


Figure 51: User interface showing commentary options

Conclusions

These tests show that S-ADM can be used in live productions, enabling the use of NGA for broadcasting and providing listeners with an enhanced experience. One of the main aims was to test interoperability between different systems, and this was successful.

As a lot of audio was captured from the contest, the tests can be repeated with other systems, in particular the MPEG-H NGA codec, as well as S-ADM over IP using SMPTE ST 2110-41 (another SMPTE standard for delivering metadata). Further development of the tools that generate the S-ADM metadata and perform the squeezing processes is planned.

13. References

[2] Recommendation ITU-R BT.709-6:2015, “Parameter values for the HDTV standards for production and international programme exchange”, July 2015, https://www.itu.int/rec/R-REC-BT.709-6-201506-I/en

[3] Recommendation ITU-R BT.2020:2015, “Parameter values for ultra-high definition television systems for production and international programme exchange”, Oct 2015, https://www.itu.int/rec/R-REC-BT.2020-2-201510-I/en

[4] Recommendation ITU-R BT.1886:2011, “Reference electro-optical transfer function for flat panel displays used in HDTV studio production”, March 2011, https://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.1886-0-201103-I!!PDF-E.pdf

[5] Recommendation ITU-R BT.2100, “Image parameter values for high dynamic range television for use in production and international programme exchange”, July 2018, http://www.itu.int/rec/R-REC-BT.2100

[6] Report ITU-R BT.2390-2, “High dynamic range television for production and international programme exchange”, https://www.itu.int/pub/R-REP-BT.2390-2016 (companion report to ITU-R Recommendation BT.2100)

[8] Recommendation ITU-R BT.2408-7:2023. “Guidance for operational practices in HDR television production”, Sept 2023., https://www.itu.int/pub/R-REP-BT.2408-7-2023

[9] SMPTE ST 2084:2014, “High Dynamic Range Electro-Optical Transfer Function of Mastering Reference Displays”, Aug 2014, https://my.smpte.org/s/product-details?id=a1BVR0000008kaj2AA

[10] SMPTE ST 2086:2018, “Mastering Display Color Volume Metadata Supporting High Luminance and Wide Color Gamut Images”, April 2018, https://my.smpte.org/s/product-details?id=a1BVR0000008kal2AA

[14] Recommendation ITU-R BT.814-4:2018, “Specifications of PLUGE test signals and alignment procedures for setting of brightness and contrast of displays”, July 2018, https://www.itu.int/rec/R-REC-BT.814-4-201807-I/en

[16] DASH-IF Interoperability Points: Guidelines for Implementation, version 4.3, November 2018,
https://dashif.org/docs/DASH-IF-IOP-v4.3.pdf

[18] CTA-608-E S-2019, “Line 21 Data Services”, Dec 2019,
https://shop.cta.tech/products/line-21-data-services#:~:text=ANSI%2FCTA%2D608%2DE%20S%2D2019%20is%20a,of%20the%20NTSC%20video%20signal.

[19] ANSI/CTA-708-E S-2023, “Digital Television (DTV) Closed Captioning”, Aug 2013,
https://shop.cta.tech/products/digital-television-dtv-closed-captioning

[20] ETSI 300 743, “Digital Video Broadcasting (DVB); Subtitling systems”, v1.6.1, July 2018,
https://www.etsi.org/deliver/etsi_en/300700_300799/300743/01.06.01_20/en_300743v010601a.pdf

[21] ETSI 300 472, ”Digital Video Broadcasting (DVB); Specification for conveying ITU-R System B Teletext in DVB bitstreams”, v1.3.1, May 2003,
http://www.etsi.org/deliver/etsi_en/300400_300499/300472/01.03.01_60/en_300472v010301p.pdf

[22] ANSI/SCTE-27 22016 (R2021), “Subtitling Methods for Broadcast Cable”, 2021,
https://wagtail-prod-storage.s3.amazonaws.com/documents/ANSI_SCTE_27_2016_R2021.pdf

[23] W3C Recommendation: “TTML Text and Image Profiles for Internet Media Subtitles and Captions (IMSC1)”, v1.1,Nov 2018 updated April 2020,
https://www.w3.org/TR/ttml-imsc1.1/

[26] ISO/IEC: Doc. ISO/IEC 23008-2:2020 “Information technology -- High efficiency coding and media delivery in heterogeneous environments -- Part 2: High efficiency video coding”, Aug 2020,
https://www.iso.org/standard/75484.html

[27] ISO/IEC 14496-3:2019, “Informational Technology, Coding of Audio-Visual objects, Part 3: Audio”, Dec 2019,
https://www.iso.org/standard/76383.html

[29] ETSI 102 366 v1.4.1 (2017-09), "Digital Audio Compression (AC-3, Enhanced AC-3) Standard", Sept 2009, https://www.etsi.org/deliver/etsi_ts/102300_102399/102366/01.04.01_60/ts_102366v010401p.pdf

[33] ETSI TS 103 433-1 v1.2.1 (2017-08) "High-Performance Single Layer Directly Standard Dynamic Range (SDR) Compatible High Dynamic Range (HDR) System for use in Consumer Electronics devices (SL-HDR1)", http://www.etsi.org/deliver/etsi_ts/103400_103499/10343301/01.02.01_60/ts_10343301v010201p.pdf

[35] ETSI TS 103 420 v1.2.1 (2018-10), "Object-based audio coding for Enhanced AC-3 (E-AC-3)". https://www.etsi.org/deliver/etsi_ts/103400_103499/103420/01.02.01_60/ts_103420v010201p.pdf

[36] SMPTE ST 337:2015, "Format for Non-PCM Audio and Data in AES 3 Serial Digital Audio Interface" https://my.smpte.org/s/product-details?id=a1BVR0000008kc22AA

[37] Recommendation ITU-R BS.1770-5, "Algorithms to measure audio programme loudness and true-peak audio level", Nov 2023. https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1770-5-202311-I!!PDF-E.pdf

[38] SMPTE ST 2067-21:2023, "Interoperable Master Format — Application #2E," 2016 https://pub.smpte.org/doc/st2067-21/20221124-pub/st2067-21-2023.pdf

[42] Internet Engineering Task Force (IETF) RFC 3550, "RTP: A Transport Protocol for Real-Time Applications", https://www.ietf.org/rfc/rfc3550.txt

[43] SMPTE ST 2110-10:2022, "Professional Media over IP Networks: System Timing and Definitions" https://pub.smpte.org/doc/st2110-10/20220328-pub/

[44] SMPTE ST 2110-20:2022, “Professional Media over IP Networks: Uncompressed Active Video”
https://pub.smpte.org/doc/st2110-20/20221214-pub/

[45] SMPTE ST 2110-21:2022, “Professional Media over IP Networks: Traffic Shaping and Delivery Timing for Video”
https://pub.smpte.org/doc/st2110-21/20221214-pub/

[46] SMPTE ST 2110-30:2017, “Professional Media over IP Networks: PCM Digital Audio”
https://ieeexplore.ieee.org/document/8167392

[47] SMPTE ST 2110-40:2018, “Professional Media over IP Networks: SMPTE ST 291-1 Ancillary Data”
https://ieeexplore.ieee.org/document/8353279

[48] SMPTE ST 2110-30:2017, “Professional Media over IP Networks: PCM Digital Audio”
https://pub.smpte.org/doc/st2110-30/

[50] SMPTE ST 2065-1:2020, “Academy Color Encoding Specification (ACES)”
https://pub.smpte.org/doc/st2065-1/20200909-pub/

[56] ATSC: A/342-2:2024-0, “AC-4 System”, April 3, 2024,
https://www.atsc.org/wp-content/uploads/2024/04/A342-2-2024-04-AC4-System.pdf

[57] ATSC: A/342-3:2024-4, “MPEG-H System”, April 3, 2024
https://www.atsc.org/wp-content/uploads/2024/04/A342-3-2024-04-MPEG-System.pdf

[65] ETSI TS 103 190-2 (2015-09), “Digital Audio Compression (AC-4) Standard Part2: Immersive and personalized audio”, September 25, 2015,
http://www.etsi.org/deliver/etsi_ts/103100_103199/10319002/01.01.01_60/ts_10319002v010101p.pdf

[70] ISO/IEC: 23008-3, "Information technology – High efficiency coding and media delivery in heterogeneous environments – Part 3: 3D audio", Aug 2022,
https://www.iso.org/standard/83525.html

[72] ITU-R BS.2076-2, “Audio Definition Model”, Oct 2019,
https://www.itu.int/rec/R-REC-BS.2076-2-201910-I/en

[73] ITU-R BS.2088-1, “Long-form file format for the international exchange of audio programme materials with metadata”, October 2019, https://www.itu.int/rec/R-REC-BS.2088/en

[76] Report ITU-R BT.2390-3, “High dynamic range television for production and international programme exchange”, March 2023, https://www.itu.int/pub/R-REP-BT.2390-11-2023

[88] EBU Tech Report 038, March 2017, “Subjective evaluation of HLG for HDR and SDR distribution” https://tech.ebu.ch/publications/tr038

[89] EBU Recommendation 129, November 2018, “Advice to Broadcasters on Avoiding Image Retention on TV Production Displays”, https://tech.ebu.ch/docs/r/r129.pdf

[99] SMPTE ST 292-1:2018 1.5 Gb/s Signal/Data Serial Interface https://pub.smpte.org/doc/st292-1/20180412-pub/

[100] R. W. G. Hunt, "Light and Dark Adaptation and the Perception of Color*," J. Opt. Soc. Am. **42**, 190-199 (1952), https://www.osapublishing.org/josa/abstract.cfm?uri=josa-42-3-190

[101] J.C. Stevens and S.S. Stevens, “Brightness Function: Effects of Adaptation,” J. Opt. Soc. Am. 53, 375-385 (1963), https://doi.org/10.1364/JOSA.53.000375

[102] J. Kautz, H. Kim, T. Weyrich, “Modeling Perception under Extended Luminance Levels”, ACM TOG, Vol 28 Issue 3, August 2009, https://dl.acm.org/citation.cfm?id=1531333

[105] ARIB STD-B67 Version 2.0, “Parameter Values for the Hybrid Log-Gamma (HLG) High Dynamic Range Television (HDR-TV) System for Programme Production”, January 2018, https://www.arib.or.jp/english/std_tr/broadcasting/std-b67.html

[106] EBU R 103, “Video signal tolerance in Digital Television Systems”, Version 3.0, June 2020, https://tech.ebu.ch/docs/r/r103.pdf

[107] ITU-R BT.2124, “Objective metric for the assessment of the potential visibility of colour differences in television”, January 2019, https://www.itu.int/rec/R-REC-BT.2124-0-201901-I/en

[111] ETSI TS 102 114 v1.6.1 (2019-08), “DTS Coherent Acoustics; Core and Extensions with Additional Profiles”, August 2, 2019, https://www.etsi.org/deliver/etsi_ts/102100_102199/102114/01.06.01_60/ts_102114v010601p.pdf

[113] DEPRECATED: See Reference [107]

[114] Recommendation ITU-T H.273: 2024 | ISO/IEC 23091-2:2019, “Coding-independent code points for video signal type identification”, Aug 2024, https://www.itu.int/rec/T-REC-H.273-202407-P/en

[116] ISO/CIE 11664-5:2024, “Colorimetery-Part 5: CIE 1976 L*u*v* colour space and u’,v’ uniform chromaticity scale diagram”,June 2024, https://www.iso.org/standard/86224.html

[123] Recommendation ITU-R BS.2125 “A serial representation of the Audio Definition Model”, April 2020, https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.2125-1-202205-I!!PDF-E.pdf

[124] SMPTE ST 2116:2019, “Format for Non-PCM Audio and Data in AES3 — Carriage of Metadata of Serial ADM (Audio Definition Model)”, Jan 2020, https://pub.smpte.org/doc/st2116/20191018-pub/

[126] SMPTE ST 2110-31:2022, ‚SMPTE Standard – Professional Media Over Managed IP Networks: AES3 Transparent Transport, Nov 2022, https://pub.smpte.org/doc/st2110-31/20220624-pub/

[127] SMPTE ST 2109:2019, SMPTE Standard – Format for Non-PCM Audio and Data in AES3 – Audio Metadata, August 2019, https://pub.smpte.org/doc/st2109/20190709-pub/

[128] Report ITU-R BT.2446-1, “Methods for conversion of high dynamic range content to standard dynamic range content and vice-versa”, March 2021, https://www.itu.int/dms_pub/itu-r/opb/rep/R-REP-BT.2446-1-2021-PDF-E.pdf

[129] Recommendation ITU-R BS.2051-3, “Advanced sound system for programme production”, June 2022, https://www.itu.int/rec/R-REC-BS.2051-3-202205-I/en

[145] RecommendationST 2094-20:2016, SMPTE Standard-Dynamic Metadata for Color Volume Transform-Application #2, July 2016 https://pub.smpte.org/doc/st2094-20/20160706-pub/l

[146] SMPTE ST 2094-30:2016, SMPTE Standard-Dynamic Metadata for Color Volume Transform-Application #3, July 2016, https://pub.smpte.org/doc/st2094-30/20160706-pub/

[147] SMPTE ST 2127-10:2022, SMPTE Standard - Mapping Metadata-Guided Audio (MGA) signals with S-ADM Metadata into the MXF Constrained Generic Container, July 2022, https://pub.smpte.org/doc/st2127-10/20220309-pub/

[148] ISO/TS 22028-5:2023, Photography and Graphic Technology-Extended Colour Encodings for Digital Image Storage, Manipulation and Interchange, Part 5: High Dynamic Range and Wide Colour Gamut encoding for Still Images (HDR/WCG), https://www.iso.org/standard/81863.html

[149] ITU-R BT.2129:2009, "User requirements for a Flat Panel Display (FPD) as a Master monitor in an HDTV programme production environment". May 2009, https://www.itu.int/dms_pub/itu-r/opb/rep/R-REP-BT.2129-2008-PDF-E.pdf

[150] IEC 61966-2-1:1999, Multimedia systems and equipment - Colour measurement and Management - Part 2-1: Colour management - Default RGB colour space - sRGB, 2014, https://webstore.iec.ch/en/publication/6169

[151] IEC 61966-2-5:2007, Multimedia systems and equipment - Colour measurement and Management - Part 2-5: Colour management - Default RGB colour space -opRGB, (AdobeRGB), 2007, https://webstore.iec.ch/en/publication/6175

[O] **Orange Book** – Foundational Technologies for Ultra HD

[O01] Section 7.2.5, HDR10

[O02] Section 7.2.2, Hybrid Log-Gamma (HLG) and HLG10

[O03] Section 7.2.1 Perceptual Quantization (PQ) and PQ10

[Y] **Yellow Book** – Beyond Foundational Technologies

[Y01] Section 8, Considerations in Using Technology Beyond Foundation

[G] **Green Book** – Ultra HD Distribution

[G01] Section 7, Introduction

[I] **Indigo Book** – Ultra HD Technology Implementations

[I01] Section 10.3, NBCU Single-Mastr HDR-SDR Workflow Recommendations

[I02] Section 9.1, Dolby AC-4

[I03] Section 8.1, Dolby Vision

[I04] Section 8.4, SL-HDR2

[I05] Section 8.3, SL-HDR1

[I06] Section 9, Monographs on NGA

[V] **Violet Book** – Real World Ultra HD

[V01] Section 10.4.2 Backward Compatibility

[V02] Section 11.3 Down Conversion at the STB

[V03] Section 10, Decoding and Rendering

[V04] Section 10.4, Considerations fro Ultra HD Technologies beyond Foundation Ultra HD

[V05] Section 11, Format Interoperability

[V06] Section 13, Real World Foundation Ultra HD Deployments

[V07] Section 11.5, Interoperability of Atmos Immersive Audio

[V08] Section 13.3. NBCUniversal Olympics and 2018 World Cup

(End of Blue Book)