NEXT GENERATION TV IN BRAZIL DURING THE FOOTBALL WORLD CUP 2022

C. Cosme

Globo, Brazil

ABSTRACT

In 2016, the Sistema Brasileiro de Televisão Digital (SBTVD) Forum decided to update the current digital terrestrial television (DTT) system, using a two-step approach: extending the life span of the existing DTT system through a backward-compatible evolution (TV 2.5 Project) and starting the development of the next-generation DTT system (TV 3.0 Project). While the deployment of TV 2.5 started in 2020, the TV 3.0 standard will be finalized in 2024 and is expected to launch in 2025.

INTRODUCTION

During the FIFA World Cup 2022, SBTVD Forum members and partner companies tested and demonstrated some of the technologies of Projects TV 2.5 and TV 3.0. Every match of the tournament, from 20 November to 18 December 2022, was broadcasted using backward compatible improvements from the TV 2.5 Project and live streamed using preliminary specifications for the TV 3.0 Project. Both showcases represent a significant technological step towards a better viewing experience for digital terrestrial television in Brazil.

BRINGING THE SHOW TO BRAZIL

Globo, a Brazilian mediatech company, has acquired the rights to broadcast the FIFA World Cup 2022 Qatar in Brazil. For this purpose, an intercontinental transport system was implemented between the International Broadcast Centre (IBC) located in the city of Doha in Qatar and the Globo facilities located in Rio de Janeiro city, Brazil. The transport system operated during the World Cup period and was based on media gateways operating on a private leased network, whose purpose was to transport, among other information, the two UHD-1 feeds provided by the event's broadcast host. Once these UHD-1 feeds were received at the Globo facilities, they were used as primary signals for an audio-visual program for Globo's commercial channels, which include a free-to-air DTT system based on the ISDB-T standard and Globo's streaming service, entitled Globoplay. While in the first case, the sports events signal was distributed with HD spatial resolution (1080i), in the second case the event signal was distributed with UHD-1 spatial resolution (2160p). In both distribution services, it was also possible to deliver to the end consumer an immersive audio experience in 5.1+4H format. For this purpose, the DTT distribution simultaneously employed Dolby Atmos E-AC-3 JOC and MPEG-H Audio codecs, while the streaming service employed Dolby Atmos E-AC-3 JOC codec. Aside from this commercial operation, an entire dedicated system was set up in near closed Globo premise, henceforth called Globo Città, to enable three showcases' tracks. These showcases were open to Globo employees as well as partners and visitors from the Brazilian and international broadcast community.

SHOWCASES' TRACKS AT GLOBO CITTÀ FACILITY

The received UHD-1 feeds from Qatar had the following video format: 3840 x 2160 pixels, 10-bit depth, 50 frames per second (fps), progressive scan, colour gamut ITU-R BT.2020, and High Dynamic Range (HDR) in Hybrid Log Gamma (HLG) standard. It was necessary to perform two conversions on the received feeds, to conform them to the format commonly used in Brazil. The first conversion changed the frame rate to 59.94 fps and the second conversion changed the transfer curve to the HDR10 standard. The received UHD-1 feeds contained 16 PCM audio channels, composed of an immersive 5.1+4H mix, four mono objects (English commentator, Closed ball, PA music, and Audio follow video), and a stereo object (Fans). At the Globo premise, which received the signals directly from Qatar, all this content was collected after the above-mentioned video conversions process and was sent to the Globo Città using an uncompressed optical transport system based on 12G-SDI and 3G-SDI interfaces. The objective of employing this type of transport was to ensure that the same quality level of the UHD-1 feeds received from Qatar was also available to the systems assembled at Globo Città with the lowest possible latency. The uncompressed optical transport system was also responsible for carrying additional signals, such as mono audio objects containing the Portuguese commentator of the soccer matches and the 4K HDR production distributed on the Globoplay streaming service. All these signals fed the production system set up at Globo Città, which is illustrated in Figure 1. The production system in Figure 1 can be divided into four large functional domains but physically they were assembled in two different environments at Globo Città.

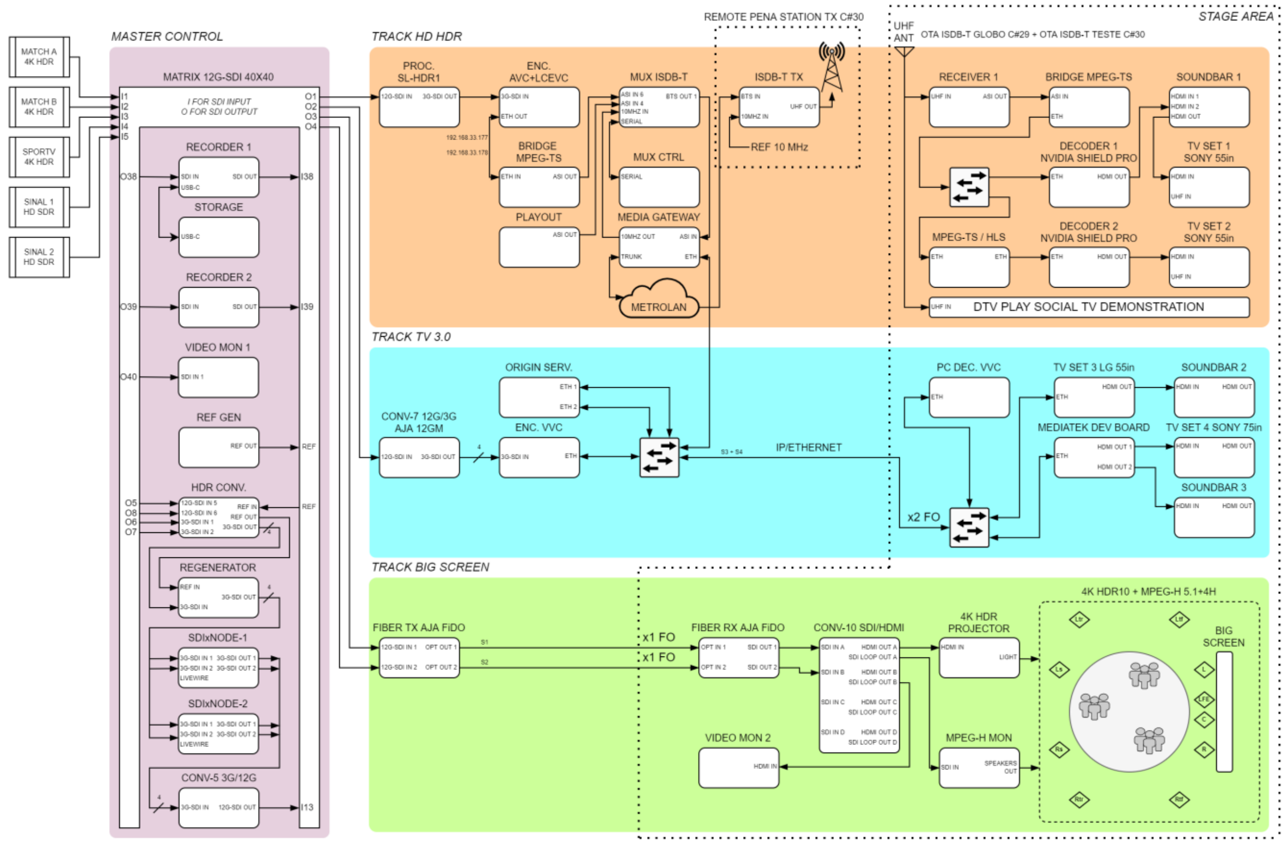

Figure 1 – Globo Città production system diagram.

The first functional domain is called Master Control (purple area in Figure 1) and was assembled in a cold room inside Globo's innovation laboratory at Globo Città. This part was responsible for receiving all signals transported by the uncompressed optical system, selecting the desired audio objects from Qatar, adding the audio object containing the Portuguese commentator, ensuring the correct lip sync with UHD-1 video content, and thus composing the baseband audio and video program for the showcases tracks. In this step was also performed the MPEG-H authoring process for the available audio content targeting to compose the desired audio scene with its respective presets and resources. Figure 2 illustrates the audio interface of the equipment responsible for the MPEG-H authoring process. In the Master Control, it was also possible to record the signals of the soccer matches for later display in the showcase tracks during the periods in which the live broadcast of a soccer match did not occur.

The MPEG-H authoring process included the creation of three audio presets, as described below:

- Preset "Padrão": composed of an immersive 5.1+4 mix, an option to select the match commentator language (English or Portuguese), and an option to enhance the sound of the kickball (Closed ball object).
- Preset "Completo": identical to the previous preset, with the addition of audio objects with fans cheering sounds (Fans object) and pre/post-match music (PA music object). In this preset besides the volume control enabled for all objects, the Fan and PA music objects were authored to allow the control of their position within the immersive 5.1+4H mix.
- Preset "Experiência Globo": identical to the previous preset. The authoring done this preset allowed the total position and volume control for all the audio objects in the audio scene.

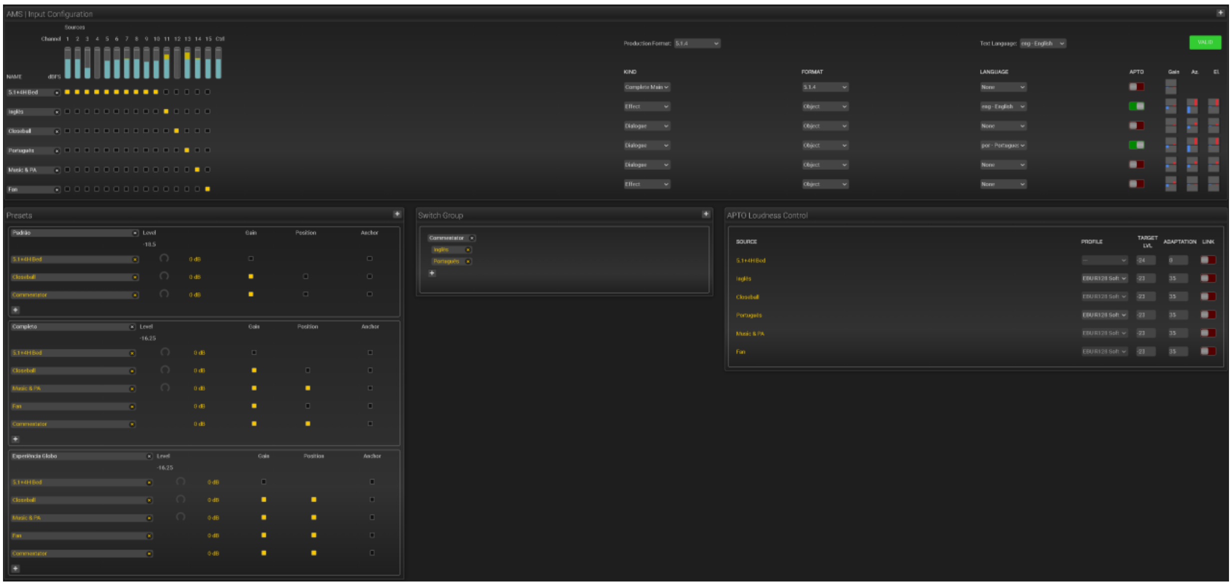

Figure 2 – MPEG-H authoring and monitoring system screen.

All elements of the Master Control functional block received and supplied video signals to a video matrix using 12G-SDI interfaces, thus allowing a simultaneous feed of the three existing showcase tracks with a UHD-1 signal as 3840 x 2160 pixels, 10-bit depth, 59.94 fps, progressive scan, colour gamut ITU-R BT.2020, and HDR10 standard. Embedded into this 12G-SDI interface was also present an audio program consisting of 16 PCM audio channels containing: an immersive 5.1+4H mix, five mono objects (English commentator, Portuguese commentator, Closed ball, Fan, and PA music), an audio track containing the MPEG-H metadata. This signal was the primary source that fed the three available showcase tracks, entitled: Track HD HDR, Track TV 3.0, and Track Big Screen.

Showcase Track HD HDR

The first showcase, named in Figure 1 as Track HD HDR (orange area) explored new technologies that can be used with the current Brazilian DTT system in a way that is backward compatible with legacy receivers while enabling new features in new receivers. This showcase includes technologies from the current Brazilian TV 2.5 specification and the video codec Low Complexity Enhancement Video Coding (MPEG-5 Part 2 – LCEVC), which is under consideration for inclusion by the SBTVD Forum. It used a live over-the-air transmission on a UHF station with regular ISDB-T equipment to showcase three significant enhancements of the existing DTT system:

- Broadcast-Broadband Integration: using DTV Play, the new IBB system available in Brazilian TV sets, to provide a multiscreen experience (Social TV application).
- Immersive and Personalized Audio: using MPEG-H Audio to provide a personalized sound experience containing a 5.1+4H immersive mix, alternative commentary options, enhanced kickball, and fans cheering sounds from arenas, among other audio objects for the user to interact with.
- Enhanced Video Quality: MPEG-5 Part 2 (LCEVC) to provide a backward compatible PQ10/SL-HDR1 video at 1920 x 1080 pixels, 59.94 fps, 10-bit video enhancement layer over the existing H.264 base layer.

In this track, the UHD-1 feed from the Master Control first was downscaled to 1920 x 1080 pixels spatial resolution (1080p) and then converted to SL-HDR1 standard. The generated SL-HDR1 metadata was inserted in the Vertical Ancillary Data Space (VANC) area of the SDI output interface carrying the 1080p signal. Next, this 1080p signal fed an experimental encoder that used the LCEVC video codec operating together with H.264 video codec. In this encoder, the input signal (1920 x 1080 pixels, 10-bit depth, SL-HDR1, 59.94 fps, and progressive scan) was converted to 1920 x 1080 pixels, 8-bit depth, SDR ITU-R BT.709, colour gamut ITU-R BT.709, 29.97 fps, an interlaced scan signal. This format was encoded by the H.264 codec and formed the base layer component of the LCEVC+H.264 encoder. The LCEVC enhancement layer encoded the video residual elements and carried the SL-HDR1 metadata to enable subsequent restoration of the original video quality present at the input of the LCEVC+H.264 encoder. The output of the experimental encoder used the MPEG Transport Stream standard as the transport protocol. The video PID carried the information from the LCEVC base layer and, thru Supplemental Enhancement Information (SEI) messages, also carried the data from the LCEVC enhancement layer. The maximum bit rate assigned to the video PID was set at 14 Mbps, a value identical to the video bit rate used in a regular ISDB-T DTT system in commercial operation in Brazil. The experimental encoder also encoded the audio content using the MPEG-H Audio codec and encapsulated it in a specific audio PID, whose bit rate was set at 540 kbps. The transport stream at the output of the experimental encoder fed the input of a regular ISDB-T multiplexer responsible to generate the ISDB-T Broadcast Transport Stream, later sent to a regular ISDB-T UHF transmitter. Another resource present in the ISDB-T multiplexer input was the Playout's signal, whose function was to transport the DSM-CC carousel containing the files of the Social TV application created. The DSM-CC carousel bit rate was set at 600 kbps.

The reception side of the Track HD HDR was set up in another physical environment at Globo Città, called Stage Area in Figure 1, allowing the attendees to watch the results obtained in this showcase. The 55-inch TV set at the bottom of Figure 3 displayed the SDR video resulting from the LCEVC base layer decoding (e.g., targeting legacy H.264 devices without LCEVC support). The 55-inch TV set at the top of Figure 3 displayed the HDR video resulting from the LCEVC base layer decoding with the enhancements from the LCEVC enhancement layer decoding plus the SL-HDR1 metadata (e.g., targeting new H.264 devices with LCEVC and SL-HDR1 support). There was also a soundbar connected to one of the TV sets allowing the showcase attendees to listen to the MPEG-H immersive audio content and interact with the personalization options available in the MPEG-H interactivity menu. To decode the LCEVC content, a NVidia Shield Pro set-top-box running an Android operating system was used, along with an ExoPlayer App player supporting LCEVC and SL-HDR1 technologies. In an adjacent location, there was yet another spot containing two commercial TV sets supporting the ISDB-T DTT system, which were used to demonstrate to the present attendees the Social TV application available in this track and to reinforce the backward compatibility characteristic of the LCEVC base layer using H.264 video codec when received by regular commercial TV sets without LCEVC support (Figure 4). The interactive application here entitled Social TV was running on the aforementioned commercial TV sets, both containing the new Brazilian IBB system so-called DTV Play.

Figure 3 – Track HD HDR showcase booth.

Figure 4 – Social TV showcase booth.

The Social TV application allowed the showcase attendees, with their mobile phones, to scan a QR Code sent by the broadcaster and thus engage in a chat room composed either by fans of team A or team B on the current broadcasted soccer match. Other existing features allowed the attendees to mirror on the TV set screen the screen of their mobile phones containing the chat room, create private chat rooms and invite their friends to participate, and receive on the chat room advertisements and/or statistics regarding the progress of the current soccer match. All these features of the Social TV application, aimed to enrich the sports content transmitted and thus increase user engagement in the consumption of the content.

The execution of DTV Play applications is already supported by Brazilian TV sets since 2021. The reproduction of immersive audio over DTT reception is also supported by some devices (e.g., TV sets and soundbars) available in the Brazilian market. DTV Play applications and immersive audio are already in use 24/7 by some Brazilian broadcasters in their regular DTT transmissions.

Showcase Track TV 3.0

The second showcase, entitle Track TV 3.0 in Figure 1 (cyan area), explored recent technologies selected for the TV 3.0 Project: H.266 Versatile Video Coding (MPEG-I VVC), Low Complexity Enhancement Video Coding (MPEG-5 Part 2 – LCEVC), MPEG-H Audio, and Dynamic Adaptive Streaming over HTTP (DASH). It was the first time these technologies were used in a complete end-to-end live production to provide 4K HDR video with immersive and personalized 5.1+4H audio. It also featured the first TV set supporting VVC, MPEG-H Audio, and DASH, as well as a complete user interface for viewers to interact with the content while experiencing the immersive sound over a soundbar.

Figure 5 – Track TV 3.0 showcase booth.

In this showcase, the UHD-1 content sent by the Master Control over the 12G-SDI interface was reformatted to four 3G-SDI interfaces and fed to an experimental encoder supporting VVC, MPEG-H Audio, and DASH technologies. This encoder encoded the UHD-1 video content using VVC codec at a video bit rate of 14 Mbps. The audio content was encoded with MPEG-H Audio codec at an audio bit rate of 540 kbps. Both compressed contents were encapsulated in a fragmented MP4 (fMP4) container using Common Media Application Format (CMAF) profile. The used transport protocol was DASH, and this encoder continuously pushed the generated manifest file and DASH-related segments to an Origin Server (on the same local IP network) using Web-based distributed Authoring and Versioning (WebDAV) protocol. From the Origin Server on the local network, the content was available for consumption.

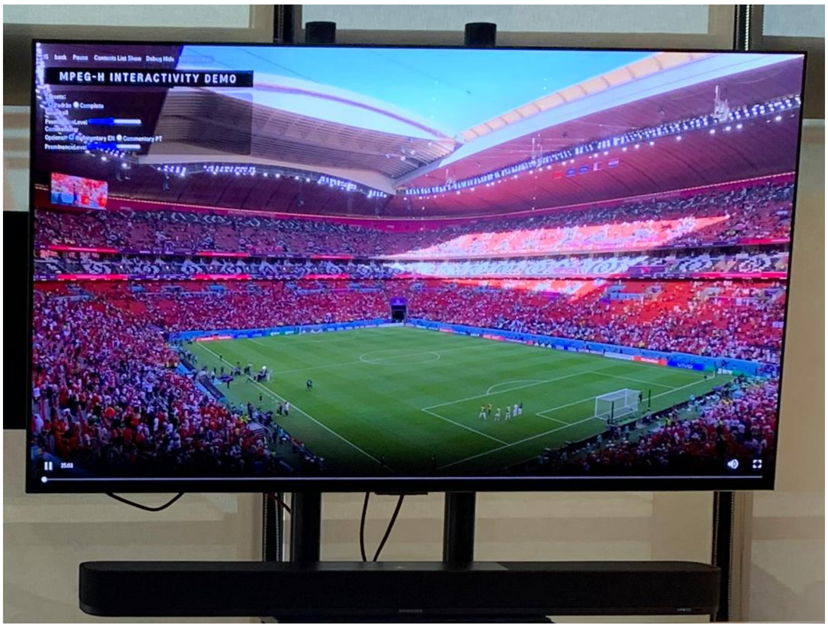

Figure 6 – TV with integrated MPEG-H user interface.

On the reception side of this showcase, it was demonstrated the real-time reproduction of the live encoded content using three setups:

- The first setup was composed of a workstation running a Windows 10 operating system and containing a modified version of the MPV player supporting only DASH and VVC decoding. The function of this setup was to demonstrate real-time VVC decoding of the live 4K HDR feed in a Windows environment.
- The second setup was composed of a MediaTek Pentonic 1000 development board, running Android operating system, and containing a version of ExoPalyer app supporting DASH and LCEVC+VVC. Seamlessly, the ExoPlayer app fetched via DASH protocol the 4K HDR content of soccer matches previously encoded with LCEVC+VVC. Here again, for comparison purposes, the maximum video bit rate was set to 14 Mbps. In this setup, 90% of the video bit rate was used for the VVC on the base layer operating at HD resolution (1080p) and 10% of the video bit rate was assigned to the LCEVC enhancement layer (2160p).
- The third setup was composed of a commercial TV set, whose hardware received the installation of an experimental firmware, thus enabling support for VVC, MPEG-H, and DASH technologies. Thus, this TV set, along with the soundbar connected to it, decode in a real-time manner the VVC 4K HDR content and the immersive MPEG-H audio being live transmitted and available on the Origin Server. It was also possible in this TV set to enable an interactive menu that allowed the showcase attendees to navigate and select the available options in the MPEG-H interactivity menu. These options were authored live in the Master Control block in Figure 1.

The showcase Track TV 3.0 explored the maturity and use of VCC, LCEVC, MPEG-H Audio, and DASH technologies (all technologies already selected for the future Brazilian DTT system) when employed in a real-time production chain for a live sports event.

Showcase Track Big Screen

The third showcase, entitled Track Big Screen in Figure 1 (green area), was composed by displaying the uncompressed UHD-1 signal from the Master Control on a 120-inch screen whose projection was performed by a 4K HDR laser beamer. This environment also had an audio layout containing ten speakers to reproduce the immersive 5.1+4H mix authored with MPEG-H on the Master Control step. It was possible in this spot to enjoy the same interactivity and personalization features provide by the authored MPEG-H Audio content. That is, besides listening to the immersive audio of the soccer matches in 5.1+4H format, the attendees could switch the narration of the soccer match between English or Portuguese, reinforce the kickball sounds, and adjust the level and position of the fans' cheering sounds in the stadium. The overall experience received incredibly positive feedback from the attendees, who felt immersed as if they were in the stands of the soccer stadiums in Qatar watching the soccer matches.

Figure 7 – Track Big Screen.

CONCLUSIONS

Major sporting events are always excellent opportunities for technological showcases involving state-of-the-art technologies. During the Football World Cup 2022, it was no different and, therefore, it was possible to explore the maturity of some technologies chosen for the future Brazilian DTT system, as well as experiment with new approaches such as in the showcase Track HD HDR. All this work illustrates the fast pace at which tense technologies are maturing while they are implemented in both hardware and software solutions.

REFERENCES

1. Brazil Administration, 2023. Implementation of second-generation digital terrestrial television broadcasting (DTTB) system in the Americas. CITEL, 41 meeting.

2. ISO/IEC JTC 1/SC 29/WG4, m62044, “Report on usage of LCEVC in Brazilian broadcasting for demonstration during Football World Cup”, January 2023.

3. Brazil, Federative Republic of. TV 2.5 and TV 3.0 Trials on the FIFA World Cup 2022. Information document submitted to ITU-R Study Group 6, February 2023.

ACKNOWLEDGEMENTS

The author of this article would like to thank immensely the partnership and joint work with the colleagues from the following companies, without whom this entire project would not have come to fruition.

- Ateme SA, Paris, France.
- Fraunhofer IIS, Erlangen, Germany.
- InterDigital, Delaware, United States.
- LG, Seoul, South Korea.
- MainConcept, Aachen, Germany.
- MediaTek, Hsinchu, Taiwan.
- Philips, Amsterdam, Netherlands.
- Samsung, Seoul, South Korea.
- Sennheiser, Wedemark, Germany.
- Spin Digital, Berlin, Germany.
- V-Nova, London, England.