LIVE PRODUCTION USING THE AUDIO DEFINITION MODEL AT THE EUROVISION SONG CONTEST 2023

M. Firth

BBC Research & Development, UK

ABSTRACT

The Audio Definition Model (ADM) is a metadata standard (ITU-R BS.2076) for describing the technical properties of audio and is particularly relevant for Next Generation Audio (NGA) experiences. Whilst support for file-based ADM metadata in software is increasing and production workflows are becoming well established, live ADM production is far less developed. There are many problems to solve in order to deploy live ADM workflows as a regular method of producing content.

There are two methods of utilising ADM in live or real-time applications. These are Serial ADM (S-ADM) and ADM-OSC (ADM Open Sound Control). To investigate the implementation of these technologies and the suitability and limitations of live ADM workflows in a broadcast scenario, a trial was proposed. This took place during the Eurovision Song Contest 2023 with the aim of implementing a live end-to-end Next Generation Audio broadcast chain using ADM.

INTRODUCTION

This paper discusses workflows and processes to produce, distribute and deliver Next Generation Audio (NGA) experiences in live production scenarios. These were explored as part of an internal trial to implement and test an end-to-end NGA broadcast chain using the Audio Definition Model (ADM) [1] during the 2023 Eurovision Song Contest.

Next Generation Audio

Next Generation Audio (NGA) is an umbrella term covering a range of technologies which provide benefits over traditional channel-based audio production and broadcast. These benefits include interactivity and personalisation of experiences, such as providing multiple language options, or to address accessibility needs by supporting audio description or dialogue enhancement. It may also be for more creative use cases such as allowing the listener to select their favourite commentator for a football game. NGA experiences can adapt to different playback devices by making the best use of the device capabilities and available speakers. This functionality, combined with NGA's support for 3D spatialisation of audio, enables a more immersive experience to be offered to the listener.

There are commercial NGA technologies currently available, and support in consumer devices is increasing. However, these solutions are based on proprietary technology and are thus incompatible with each other in terms of both NGA feature sets and in the low-level encoding of audio and metadata.

Audio Definition Model

For a broadcaster, it may be undesirable to produce content in a format that is locked to a particular emission solution since this could hamper exchange of content and would constrain creative potential to the capabilities of current technology. To take advantage of new technologies in the future would require additional production effort.

The Audio Definition Model (ADM) was developed to address these issues by providing a common basis for audio metadata. It is a codec-agnostic standard (ITU-R BS. 2076 [1]) which is highly flexible, allowing support for current and future use-cases and emission codecs. In brief, ADM describes the construction of a programme and different presentations of the programme in a hierarchical structure of metadata elements. Individual audio assets of a programme are represented by "AudioObject" elements, which in turn are associated with elements to describe the format of the asset (known as its "TypeDefinition") and its properties. These TypeDefinitions include, among others, "Object" and "DirectSpeakers" types. The DirectSpeakers type is used for channel-based assets, and accompanying metadata describes the channel layout of the audio associated with the AudioObject. The Object type is used for static or dynamic objects. Accompanying metadata describes the intended position of the object in 3D space and other properties, such as its gain and diffuseness for rendering. These properties may be time-varying.

The ADM as described by ITU-R BS. 2076 [1] is primarily designed for file-based applications, however two related technologies exist to support ADM in live and real-time scenarios. These are Serial ADM and ADM-OSC.

Serial ADM (S-ADM) is standardised in ITU-R BS. 2125 [2]. It supports real-time applications of ADM by segmenting the metadata into time-delimited frames. S-ADM is transport-agnostic and capable of conveying the complete structure of ADM metadata. It can be carried using various other standards. For example, S-ADM can be encoded on to an AES3 audio channel using SMPTE standard ST 2116 [3]. This can be further encapsulated for IP transport using ST 2110-31 [4]. Alternatively, the recently published ST 2110-41 [5] standard can carry ADM over IP without prior ST 2116 encoding.

ADM-OSC is a relatively new proposal for supporting ADM in live production with a focus towards production controllers. It is based around the Open Sound Control (OSC) protocol [6] making it very simple and easy to implement. Because of this, it is already gathering a lot of interest from both software and hardware manufacturers. At the time of writing, ADM-OSC is not standardised and exists only as an informal specification [7]. It does not describe the programme structure, and it currently only defines the essential properties of an AudioObject such as gain and position. Additional feature support may be added to the specification in the future.

EUROVISION SONG CONTEST NGA TRIAL

The Eurovision Song Contest in 2023 was held in Liverpool, UK with the BBC as the host broadcaster. The nature of the production would provide a good variety of audio feeds to create immersive NGA mixes with some user-customisation options. The overall aim was to conduct an internal trial to build an end-to-end media chain for live NGA content, from production through to emission and playback on consumer devices within the BBC R&D lab. In designing such a pipeline, the trial would highlight any problem areas to address in order to implement a complete chain. As part of the trial, the intention was also to test reception and playback of live-produced NGA content on different devices and across different mediums, as well as observing adaptability to device capabilities. Interoperability testing between ADM supporting software within the chain was another key objective.

Trial Partners

BBC R&D were joined by Dolby and L-Acoustics for the trial.

Dolby provided hardware running their AC-4 encoder as the NGA emission solution. Their AC-4 codec can deliver Dolby Atmos experiences for various applications including broadcast and streaming services. They also provided equipment for DTT (Digital Terrestrial Television), HbbTV (Hybrid broadcast broadband Television) and DASH (Dynamic Adaptive Streaming over HTTP (Hypertext Transfer Protocol)) emission. For reception, they provided a consumer-grade television with AC-4 support and a smartphone with a DASH playback application also supporting AC-4.

L-Acoustics specialise in spatial audio production tools. Their L-ISA software provided production control and local monitoring during the trial. L-ISA is comprised of two pieces of software; a Controller and a Processor. The Controller allows the user to define objects in a mix and place them in 3D space, from which it generates ADM-OSC metadata which is sent to the Processor. The Processor uses this metadata to render ADM AudioObjects to a defined loudspeaker layout. The ADM-OSC metadata would also be used to drive further downstream processes for emission as part of the trial.

BBC R&D also gained valuable experience developing their own production and monitoring tools using ADM-OSC as part of the trial. The “Live ADM-OSC Production Tool” provided an alternative production controller, and local monitoring was possible through the “EAR ADM-OSC Monitor” which was based around the EBU ADM Renderer (EAR) [8]. These ran alongside the production and monitoring solutions provided by L-Acoustics allowing interoperability tests to be conducted as part of the trial.

Production Pipeline

Audio stems for individual elements of the final broadcast mix were received at the BBC R&D lab over MADI (Multichannel Audio Digital Interface) via BBC infrastructure. These stems are listed in table 1. The final video mix was also provided on an SDI (Serial Digital Interface) feed.

Audio Stem	Description	Format
Presentation*	Full broadcast mix	5.1 & Stereo
VTs	Intermissions and act introduction reels	5.1
Sound Effects	On-screen graphics, etc.	5.1
Stage	Stage mics including presenters	Stereo
External	External voting feeds	Stereo
Music	Complete music mix	Stereo
Ambience*	Soundfield venue mic pre-rendered to 5.1	5.1
Audience Spot	Audience spot mics	20x mono
Audience Mix*	Pre-mixed spot mics	5.1
Main Commentary	Commentary used for TV broadcast	Mono
Alt Commentary	Commentary used for radio broadcast	Mono
** Stems which were not used in the final NGA experience*		

Table 1 – Audio stems provided over MADI

The stems were mainly post-fade, post-FX sends from the mixing desk at the Eurovision venue in Liverpool. An additional commentary stem was also provided by BBC Radio. This would allow for the production of a personalisable experience through commentary selection. These stems, along with the final video mix, fed the NGA production pipeline shown in figure 1.

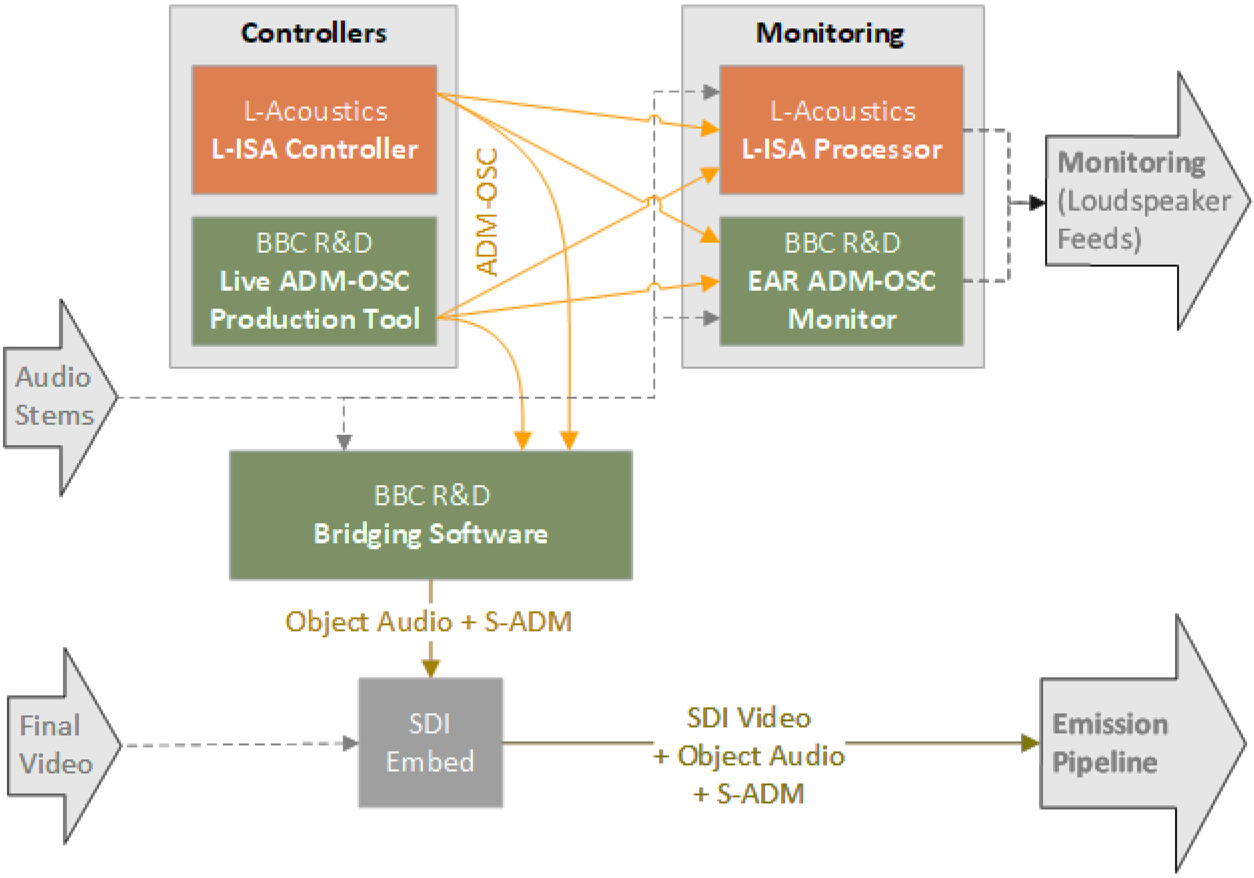

Figure 1 – Production pipeline showing software and interconnections

The production pipeline consisted of controllers, monitoring solutions, and bridging software to feed through to the emission pipeline. The controllers are essentially just metadata-generators outputting ADM-OSC. The role of the monitoring solutions is to receive audio stems and render a loudspeaker output according to the ADM-OSC metadata from controllers. Figure 1 illustrates the production software interoperability options which would be explored as part of the trial. Through the common use of ADM-OSC, both controllers should function with either of the monitoring solutions. Both controllers should also be able to drive the end-user experience by communicating with the bridging software, which ultimately feeds the emission pipeline. This would be verified during the trial.

Bridging to Emission

The Dolby AC-4 encoder ran on the Ateme Titan Live platform for the trial. The encoder requires ST 2116 [3] encoded S-ADM metadata. This had to be delivered on the 16th audio channel of an SDI feed carrying the video to the hardware running Titan Live. The remaining 15 channels could be used to carry the audio for objects within the NGA mix.

The S-ADM metadata must also conform to a specific set of rules, known as an ADM profile, to suit the encoder. Therefore, bridging software as shown in figure 1 was required in order to connect the production controllers to the emission encoder. Figure 2 provides a block diagram of the processes within this bridging software.

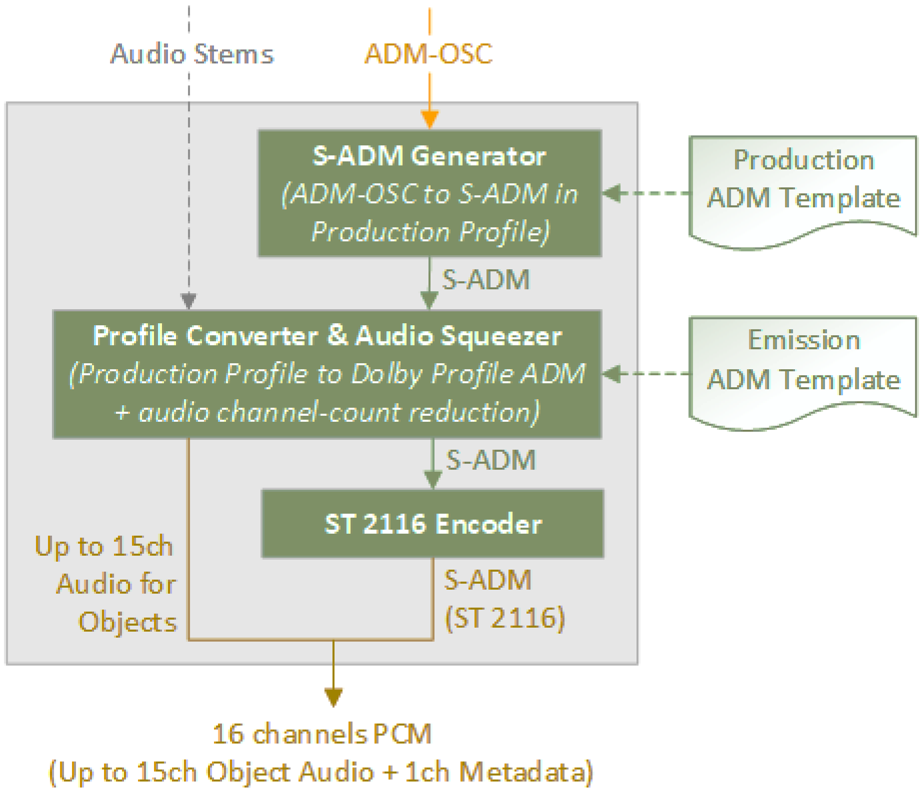

Figure 2 – Bridging software processes

The first stage within the bridging software was conversion from ADM-OSC to S-ADM. Since ADM-OSC currently only describes low-level metadata for each AudioObject and does not describe the programme structure as a whole, an ADM production template was used to provide the structure. The ADM-OSC metadata was then used to populate the low-level properties associated with each AudioObject within this template. The completed ADM template could then be used to produce S-ADM by generating time-delimited frames of metadata.

The next stage was to conform the metadata to a specific ADM profile supported by the encoder. This profile ensures that ADM metadata maps easily into AC-4. Additionally, the channel count of the production needed to be reduced. This restriction was imposed by the available channels in a single SDI connection (normally 16, but one was reserved for metadata), and by the AC-4 codec when delivering to consumer devices. Whilst AC-4 has the capability to support very high channel counts, the limitation is imposed by the decoding capabilities of target playout devices and potential bandwidth limitations of transport mediums.

Channel count reduction would require some AudioObjects to be combined. This is a two-phase process. Firstly, audio stems for combined AudioObjects must be mixed into a single audio asset, and secondly, the ADM metadata must reflect this with a single, new AudioObject. Table 2 shows how the original set of stems was reduced to one channel-based bed and two commentary objects. Ultimately, the ADM generated by this process described two presentations of the programme as set out by an ADM emission template. Both programme presentations contained the bed, but each used a different commentary object to accompany it. A 5.1.4 layout was used for the bed to provide a 3D spatial image, and creative placement of audience spot mic objects resulted in a perceptually enveloping bed. These considerations would allow the end user to switch between the two available commentaries on their device and enjoy a more immersive mix on capable hardware.

<table>
<tr><th>Audio Stem</th><th>Production AudioObject TypeDefinition</th><th colspan="2">Emission AudioObject TypeDefinition</th></tr>
<tr><td>VTs</td><td>DirectSpeakers (5.1)</td><td>►</td><td rowspan="6">DirectSpeakers (5.1.4)
(Channel-based bed using spatialisation of mono spot mics to fill height)</td></tr>
<tr><td>Sound Effects</td><td>DirectSpeakers (5.1)</td><td>►</td></tr>
<tr><td>Stage</td><td>DirectSpeakers (2.0)</td><td>►</td></tr>
<tr><td>External</td><td>DirectSpeakers (2.0)</td><td>►</td></tr>
<tr><td>Music</td><td>DirectSpeakers (2.0)</td><td>►</td></tr>
<tr><td>Audience Spot</td><td>Objects (20 of)</td><td>►</td></tr>
<tr><td>Main Commentary</td><td>Object</td><td>►</td><td>Object</td></tr>
<tr><td>Alt Commentary</td><td>Object</td><td>►</td><td>Object</td></tr>
<tr><td></td><td>= 40 Channels Required</td><td></td><td>= 12 Channels Required</td></tr>
</table>

Table 2 – Channel count reduction during profile conversion by combining multiple AudioObjects into a single AudioObject for a channel-based bed.

The final stage of the bridging software was to generate a channel of audio carrying the S-ADM metadata which can be passed over SDI. This is achieved using SMPTE standards ST 337 [9] and ST 2116 [3]. Briefly, ST 337 describes how generic data can be encoded on to a PCM audio channel conveyed over AES3 (which in turn can be embedded in SDI), and ST 2116 describes how ST 337 should be used specifically for carrying S-ADM metadata. S-ADM frames were required by the encoder every 20ms. The audio infrastructure was running at 48kHz 24-bit, meaning ST 2116 could carry 2847 bytes of payload within this time window. ST 2116 also allows for the use of GZIP compression, and so this was utilised within the bridging software, enabling an S-ADM frame to easily fit within the payload space.

Emission Pipeline

The Dolby encoder ran on the Ateme Titan Live which is a software solution for video processing and encoding. It is highly flexible and capable of running multiple output streams in parallel for different emission routes. To feed the encoder, the audio channels (including the channel carrying S-ADM metadata) were embedded on to an existing SDI feed of the final video mix, as shown in figure 1. Figure 3 illustrates the emission pipeline, showing the encoder producing output streams to support two emission routes.

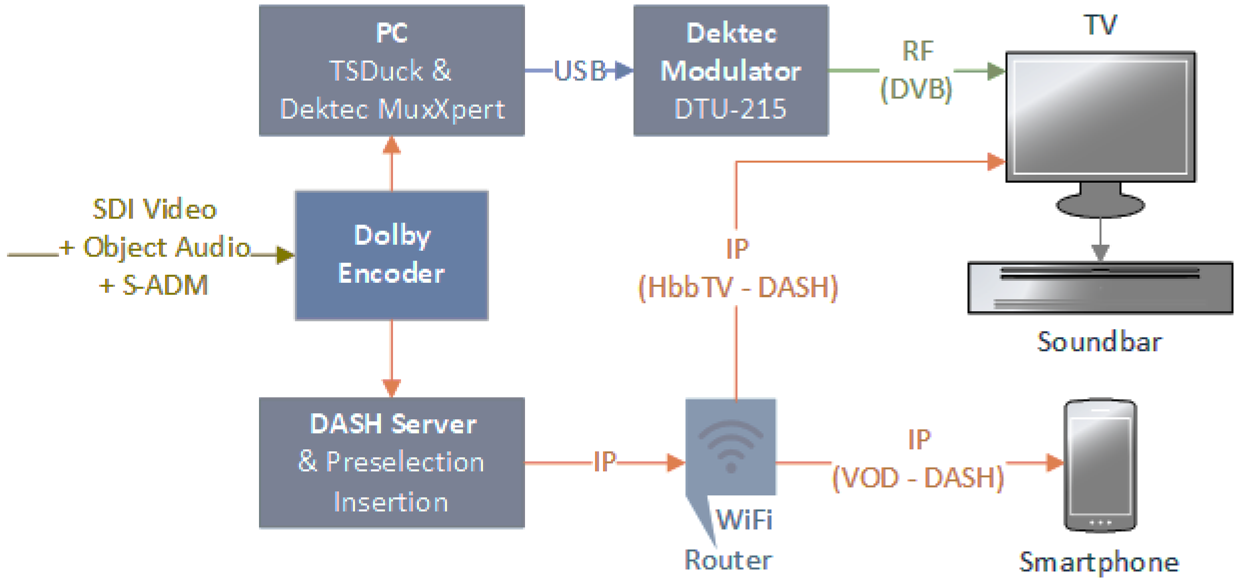

Figure 3 – Emission pipeline via encoder through to consumer devices

One stream from the encoder was for DTT (Digital Terrestrial Television) emission. This fed a small form-factor PC running "TSDuck" software to insert the mandatory tables into the transport stream to create a valid DVB-T2 stream. This was then fed to a Dektek DTU-215 modulator using the "Dektec MuxXpert" software to produce a terrestrial RF signal for over-the-air broadcast simulation.

The second stream from the encoder was fed to a DASH (Dynamic Adaptive Streaming over HTTP) server. This server's main role was to generate MPEG-DASH segments from the incoming stream. However, to provide NGA features on playback devices, some additional processing was required. This involved modifying the DASH manifest to reflect the NGA features available in the programme by use of the DASH "preselection" element. Use of this element for AC-4 is discussed in DVB Document A168 Rev.4 [10]. Dolby used a Python script to handle manifest modification for this trial.

Consumption

During the trial, NGA reception via DTT, HbbTV and a smartphone application were tested. Dolby provided a television with HbbTV and AC-4 support, and a modern Android smartphone.

Dolby's smartphone application was a streaming media player with AC-4 decoding capability which was fed from the MPEG-DASH emission route. Controls for NGA features were presented via the application's user interface.

The television was a typical consumer-grade model. This was used to receive the programme over HbbTV (using the MPEG-DASH emission route) and via DTT. The televisions in-built speakers were only capable of stereo (2.0) playback. To test adaptation to more capable devices, BBC R&D provided a Sennheiser AMBEO Max soundbar with Dolby Atmos support to connect to the television. The AMBEO Max is capable of delivering a 5.1.4 equivalent experience by using beamforming to project sound around the listening space, providing a more enveloping and immersive listening experience.

TRIAL SCHEDULE

The Eurovision Song Contest production schedule took place between the 8th and 13th of May 2023. It consisted of two semi-final shows during the week, followed by a longer final show at the weekend. There were several rehearsals for each show. Two rehearsals per show were full production run-throughs with an audience: one on the evening before the live broadcast show, and one on the afternoon of the show. The semi-finals were approximately two hours in length, and the final was approximately four hours.

The first semi-final was used by BBC R&D as an opportunity to ensure any infrastructure issues were identified and resolved, and to also test that BBC R&D software was operating as expected. Trial participants from L-Acoustics joined for the second semi-final in order to set-up their production software. Dolby also joined for the second semi-final to integrate their emission solution into the chain and provide support during the final. Therefore, the second semi-final was the first opportunity to test the complete end-to-end chain using technology from all three organisations. It was hoped that this would provide sufficient time to diagnose and troubleshoot any issues, and that a fully working end-to-end chain could be demonstrated during the final show.

FINDINGS

The trial was ultimately a success. The software from all participants worked as expected and interoperability between the production controllers, monitoring solutions, and the bridging software was proven. Both production controllers were able to operate with either monitoring solution, and both production controllers were also able to drive the bridging software and consequently the end user experience. The Dolby emission system was able to read and interpret the S-ADM generated by the bridging software and present the NGA features on the end user devices as expected.

Figure 4 – DTT Commentary Options

Using DTT delivery to a consumer-grade television, the options for alternate commentaries were available via the built-in menu system of the TV (figure 4). Likewise, for HbbTV, the same options were available via the user interface on the television (figure 5). Additionally, options for dialogue enhancement were presented. This is a feature of the AC-4 decoder whereby the level of objects signalled as “dialogue” can be boosted relative to other objects.

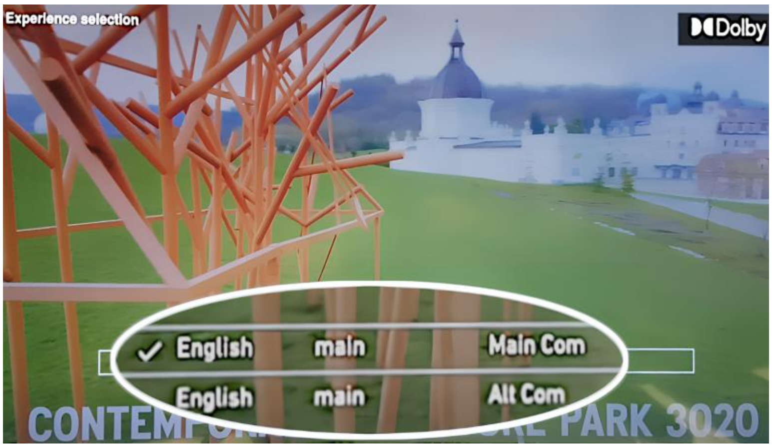

Figure 5 – HbbTV Commentary Options

Connecting the television to a soundbar with Dolby Atmos support successfully demonstrated the content adapting to a more capable device. This provided a far more immersive experience than basic stereo rendering available using the built-in speakers of the TV.

Delivery using MPEG-DASH to a smartphone application was also successful. The features available using HbbTV were also mirrored on the smartphone application (figure 6).

Figure 6 – Smartphone Commentary Options

Further Work

The trial provided an excellent insight into both workflow issues and technical considerations of implementing live NGA production using ADM-OSC [7] and S-ADM [2]. Some areas for further work were identified as a result.

Local monitoring during the trial was not completely representative of the end user experience. This is largely due to two factors. Firstly, the monitoring feed was produced using different renderers to the decoders used on consumer devices. Secondly, this rendering was driven by the original ADM-OSC metadata produced by the production tools, rather than the S-ADM metadata fed into the encoder after profile conversion by the bridging software. It is important that a producer has the ability to monitor the end-user experience, but this becomes a difficult issue to solve where there may be parallel emission routes using multiple NGA codecs as well as legacy channel-based outputs.

Further standardisation work is required for the technology encompassing the trial. ADM-OSC is currently not a standard and only includes a very limited subset of ADM functionality at the time of writing. A working group has been formed to develop the specification with the hope of reaching a standard. Work is also currently underway in the ITU to define an Emission Profile for ADM. This profile would provide a common ground for NGA emission technologies, simplifying the broadcast chain (particularly for parallel emission) by avoiding the need to conform to codec-specific ADM profiles.

This trial focused on using ADM-OSC and S-ADM in a broadcast chain using well-established infrastructure technologies such as SDI and MADI. With IP-based infrastructures becoming increasingly prevalent, it is important that the production of NGA using the SMPTE ST 2110 [4, 5] suite of standards is also investigated.

Further work is required to generalise this workflow and the software (particularly bridging software) used to form the broadcast chain. The bridging software in this trial was very bespoke to this particular use case. It used hard-coded logic to perform the ADM profile conversion from a known ADM production structure to suit the Dolby encoder. In order to generalise a live NGA production workflow, the software is required to be more versatile. Profile conversion should be customisable based on the structure of the production. There may be some automation of this step possible, which would require investigation. This task will become significantly easier when the proposed ADM Emission Profile is standardised and encoders begin to adopt this, since bespoke conversion to suit specific encoders would no longer be necessary.

CONCLUSION

This trial demonstrated live, end-to-end NGA production using software and hardware from multiple vendors in a real-world use case. This codec-agnostic approach to NGA production was only achievable using ADM and related standards and technologies. In the process, BBC R&D developed new software solutions for ADM-OSC and S-ADM, and gained new insights into what is required for full adoption of NGA production using ADM. This will direct focus for further development, particularly in generalising the workflow and software tools required to build a more flexible pipeline for other productions, and inform further standards work. The success of the trial represents a significant step forward in the real-world application of ADM software and standards.

REFERENCES

[1] ITU-R BS.2076: "Audio Definition Model"

[2] ITU-R BS.2125: "A serial representation of the Audio Definition Model"

[3] ST 2116:2019 - SMPTE Standard - "Format for Non-PCM Audio and Data in AES3 - Carriage of Metadata of Serial ADM (Audio Definition Model)"

[4] ST 2110-31 - "Professional Media Over Managed IP Networks: AES3 Transparent Transport"

[5] ST 2110-41 - "Professional Media Over Managed IP Networks - Fast Metadata Framework"

[6] Open Sound Control: http://opensoundcontrol.org/

[7] ADM-OSC: https://immersive-audio-live.github.io/ADM-OSC/

[8] EBU ADM Renderer (EAR): https://ear.readthedocs.io/en/latest/

[9] ST 337:2015 - SMPTE Standard - "Format for Non-PCM Audio and Data in an AES3 Serial Digital Audio Interface"

[10] DVB Document A168 Rev.4 - "MPEG-DASH Profile for Transport of ISO BMFF Based DVB Services over IP Based Networks" (November 2021)

ACKNOWLEDGEMENTS

The author would like to thank his colleagues at BBC R&D for their contributions to this work, and to Dolby and L-Acoustics for their participation and support in the trial. He would also like to thank the International Broadcasting Convention for permission to publish this paper.