NGA LIVE-PRODUCTION WORKFLOWS: PERSONALISED AUDIO EXPERIENCE IN BROADCASTING

T. Iwasaki, H. Kubo, T. Sugimoto, S. Oode, Y. Nakayama, and H. Okubo

NHK (Japan Broadcasting Corporation), Japan

ABSTRACT

Next generation audio (NGA) delivers personalised audio experiences on the basis of the listener's preference and loudspeaker configurations and rendering algorithms in the listening environments. The prototype of NGA production equipment was arranged to verify the NGA live-production workflow to produce the typical personalised NGA programmes. It consisted of our developed NGA console, NGA loudness meter, and MPEG-H 3DA-complied coding system supporting the Audio Definition Model (ADM) and the serial representation of ADM (S-ADM). Subjective tests were conducted to verify the reasonability of the loudness measurement method in order to develop the NGA loudness meter using several rendering conditions, including loudspeaker configurations and rendering algorithms. The results showed that differences between objective and subjective loudness of NGA programmes are comparable to those of channel-based audio programmes.

INTRODUCTION

Next generation audio (NGA) delivers personalised audio experiences on the basis of the listener's preference, reproduction setup, and listening environment. These experiences include users' emphasis on dialogue and selection of preset programmes such as programme audio with a native and multilingual broadcasting content. This scenario is enabled by rendering processes that convert a set of audio signals with associated audio metadata into a different configuration of audio signals fed to loudspeakers. Audio metadata is used to describe the positions and gains of audio objects and the programme composition. As a world standard of audio metadata, the Audio Definition Model (ADM) is specified in Recommendation ITU-R BS.2076 [1] for offline production and the serial representation of ADM (S-ADM) is specified in Recommendation ITU-R BS.2125 [2] for live production.

NGA has been adopted as broadcasting standards (e.g., in Europe and the United States) [3 – 5] and its services have already been launched. In these services, offline-produced contents are mainly provided to users. On the other hand, live contents are rarely offered, and live-production workflows are not established. Case studies of NGA live-production workflows using S-ADM were reported by the EBU [6]. The reported workflows required the handling of several protocols and guaranteed synchronization between audio signals and associated S-ADM frames.

The authors have developed the NGA console [7], NGA loudness meter [8], and coding system [9] supporting S-ADM to realize NGA live production. This paper presents the NGA live-production workflow for a typical personalised audio programme and then an experiment conducted to validate the proposed workflow. Throughout the experiment, the production system consistently signalled S-ADM, which was output synchronized with audio signals from the NGA console. Within this workflow, the NGA loudness meter adopted the proposed method for measuring the loudness of signals obtained after rendering processes. To verify the proposed measurement method, the subjective test, which included loudness matching tasks, was conducted.

LIVE-PRODUCTION WORKFLOW FOR TYPICAL PERSONALISED NGA PROGRAMME

Proposed NGA live-production workflow

A typical personalised NGA programme is one consisting of background sound and multiple dialogue objects. In producing this type of programme, the live-production workflow shown in Figure 1 is proposed. The workflow includes four steps. The first step is to select the appropriate template of ADM for programme content and to modify the template of ADM on the basis of programme composition. The second step is to load ADM with static parameters to the NGA console. The NGA console automatically assigns channels of audio objects to BUSes and faders on the basis of ADM. The third step is to mix the main programme audio (main dialogue and background sound) and to align the levels of secondary dialogue with the main dialogue. With the mixing operation, the parameters of faders are converted to ADM formats, and then the frames of S-ADM with dynamic parameters are generated. Although the number of personnel can increase for multiple dialogue objects, the automatic level adjuster for the secondary dialogue is used in this workflow to produce multiple preset programmes efficiently. The fourth step is to monitor audio signals and an S-ADM stream as broadcasting programme including multiple alternative dialogues using the S-ADM renderer.

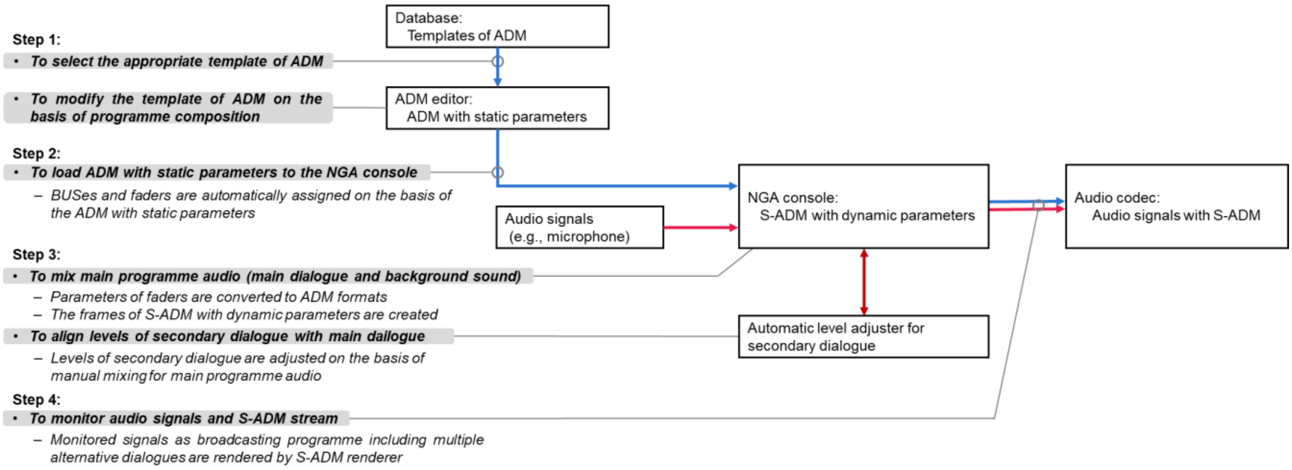

Figure 1 – Proposed NGA live-production workflow

Overview of our developed tools supporting S-ADM

To realize NGA live-production, we developed the NGA console [7], NGA loudness meter [8], and coding system [9] supporting S-ADM.

- The NGA console [7] supports the capability to generate S-ADM and audio objects whose configuration is up to 22.2 multi-channel sound and to transmit an NGA stream including up to four audio tracks for S-ADM via MADI.
- The NGA loudness meter [8] receives S-ADM stream and audio signals, renders audio signals for the selected loudspeaker configuration, and measures the loudness of the rendered audio signals as well as the channel-based audio signals. The measurement algorithm supports Recommendation ITU-R BS.1770 [10].

- The coding system [9] complies with MPEG-H 3DA baseline profile level 4. The audio stream between the encoder and the decoder is an MPEG-H 3D Audio Stream (MHAS) in the MPEG Media Transport (MMT) format over IP [11]. The decoder is equipped with the MPEG-H 3DA renderer [12]. Its interface includes functionalities to switch preset programmes and to select and enhance dialogue objects.

Experiment to verify the proposed NGA live-production workflow

To verify that the proposed workflow enabled the production of personalised audio programmes, a live-production experiment was conducted using the prototype NGA production equipment. In this workflow experiment, two personalized test-programmes were produced. The assumed production system consisted of the NGA console, the NGA loudness meter, the automatic level adjuster for secondary dialogue [13], and the coding system complying with MPEG-H 3DA baseline profile level 4. The experimental system and its diagram are shown in Figure 2 and Figure 3, respectively. Note that ADM with static parameters was created from scratch and that an audio player that output all audio materials was used instead of microphones. In this production workflow, the equipment was connected with an S-ADM-containing format specified in SMPTE ST 2116 [14]. The workflow experiment was carried out at NHK Science & Technology Research Laboratory. Two rooms, representing a broadcast station and a home, were used. The broadcast station room was equipped with loudspeakers capable of reproducing 22.2 (9+10+3), 7.1.4 (4+7+0), 5.1.4 (4+5+0), 7.1 (0+7+0), 5.1 (0+5+0), and stereo (0+2+0) audio signals as specified in Recommendation ITU-R BS.2051 [15]. In the home room, 24 loudspeakers capable of reproducing 22.2 multi-channel audio and a sound bar supporting 22.2 multi-channel audio inputs were installed.

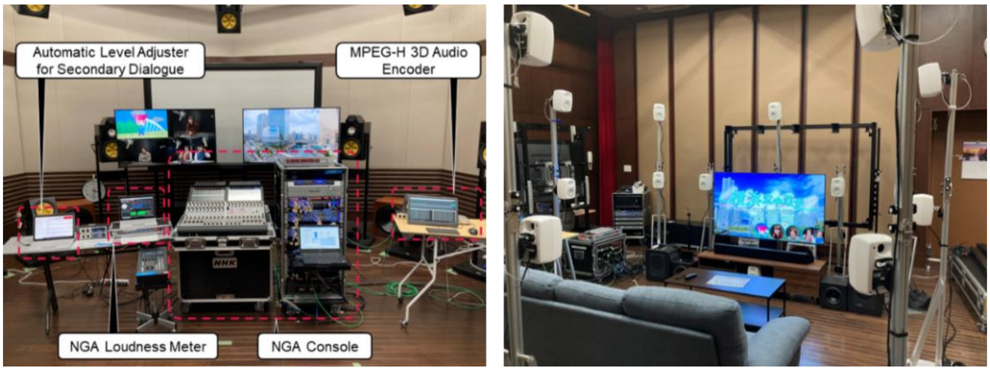

Figure 2 – NGA live-production experimental system
(left: room recreating a broadcasting station; right: room recreating a home)

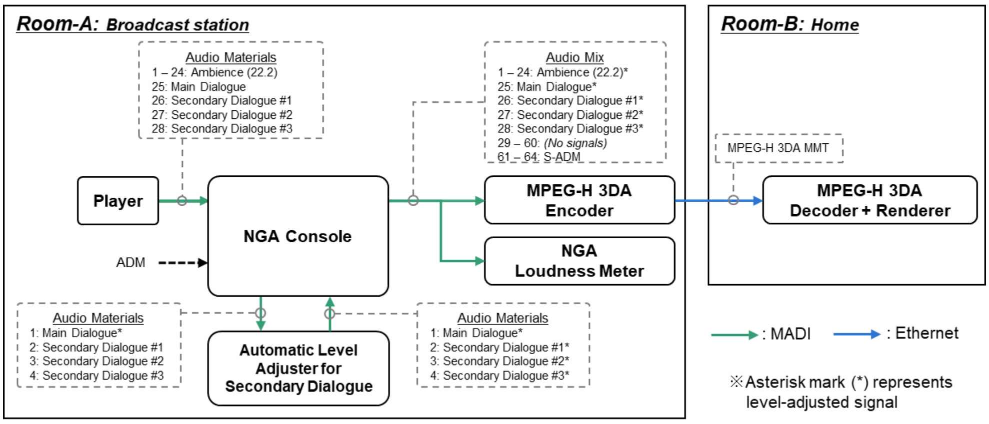

Figure 3 – Diagram of the experimental system to verify the proposed workflow

Test-Programmes to verify NGA live-production workflow

Two test-programmes were produced: “Yokohama Trip” (a travel documentary) and “Figure Skating” (a live sports event). Both programmes consisted of 22.2 multi-channel audio background sound and four mono dialogue audio objects. Users can select their preferred dialogue. Figure 4 shows samples of the programme videos. Three narrators and one pair of narrators for “Yokohama Trip” and four pairs of narrators for “Figure Skating” talked in separate booths while watching the video and listening to sounds captured from a shooting location and a venue. In the programmes, these narrators were displayed at the bottom of the video for the purpose of motivating users to switch between dialogues as well as to present switchable dialogue options to them.

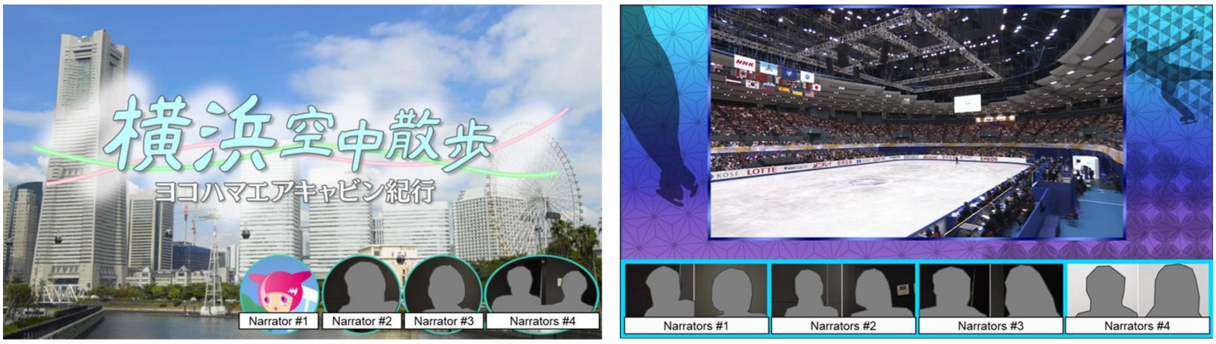

Figure 4 – Samples of programme video (left: Yokohama Trip; right: Figure Skating)

Producing NGA programme

The programme audios were produced using the NGA console and the automatic level adjuster for secondary dialogue. Table 1 shows the channel composition of audio objects in the main output for broadcasting and the combination of these objects for each preset programme described in ADM.

MADI Ch.	Format	Audio Object	Preset Programme			
			#1	#2	#3	#4
1 – 24	22.2	Background sound	✓	✓	✓	✓
25	Mono	Main commentary	✓			
26	Mono	Sub-commentary (1)		✓		
27	Mono	Sub-commentary (2)			✓	
28	Mono	Sub-commentary (3)				✓
29 – 60	*No signal*					
61 – 64	S-ADM					

Table 1 – Composition of audio objects in programme audio

The live-production experiment included the following workflow.

- ADM with static parameters was created on the basis of the programme composition (e.g., Table 1) and then loaded into the NGA console before live production.
- The main dialogue and background sounds were mixed by an audio engineer using the NGA console.

- The levels of the secondary dialogues were adjusted by the automatic level adjuster for secondary dialogue [13] on the basis of the levels of the main dialogue adjusted manually.
- The console generated S-ADM, which contained time-variant parameters for mixing operations, sequentially from ADM and output them synchronized with the audio signals in real time as a stream in accordance with the S-ADM transmission format [14].

To achieve the proposed workflow, the NGA console [16] was developed by adding the following three functions to the previous NGA console [7]. The block diagram of this console is shown in Figure 5.

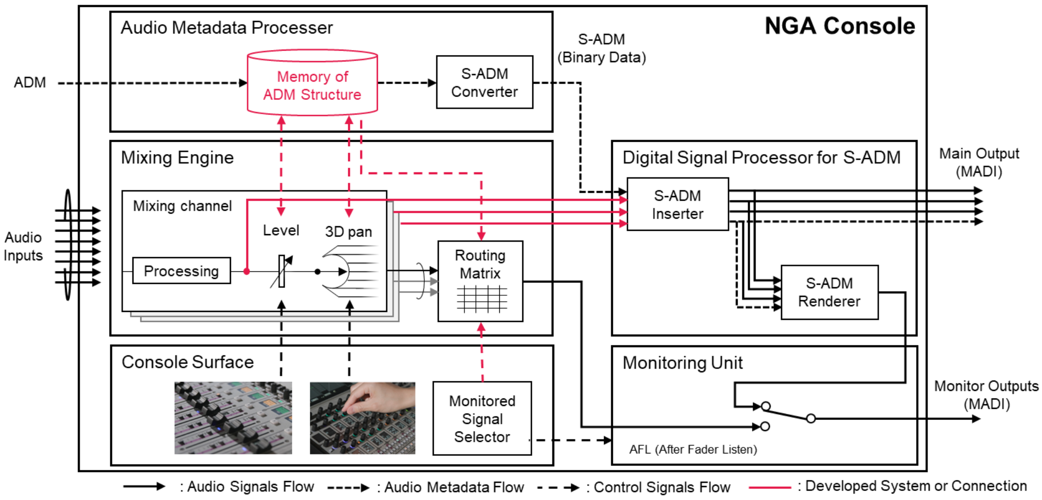

Figure 5 – Block diagram of developed NGA console

(1) Generating dynamic parameters in S-ADM updated by fader operations:

The console is equipped with a memory to hold the parsed ADM as a structure. The gains and positions of the audio objects are periodically obtained from the console surface and panning controller, and the parameters in the structures were updated. Time-varying parameters can be reflected in the programme audio by converting these structures to S-ADM.

(2) Providing low-latency auditory monitor:

The console is equipped with two monitoring processes with a built-in renderer and internal circuits for After Fader Listen (AFL). The monitoring process via the renderer is ideal in terms of checking the normality of S-ADM contained in the main output. However, the built-in renderer caused an approximately 250 ms delay in processing S-ADM and audio signals and buffering for input and output. The delays are unacceptable for mixing engineers to produce programmes. To avoid these delays, the other low latency monitoring process via AFL can be used. This system generated mixed audio through the use of the ADM structures. However, these low-latency monitoring signals are similar to, but not equal to the rendered audio signals. Therefore, the two monitoring processes may be switched depending on the purpose of the monitor: mixing of audio materials or checking the main output.

(3) Setting console parameters consistent with ADM semi-automatically:

The console analyses the loaded ADM and automatically assigned channels of audio objects described in ADM to the console BUSes. The appropriate type of bus is selected on the basis of the audio object configuration (e.g., 22.2, 5.1, mono) described in ADM. The BUSes corresponding to time-varying mixing parameters are determined by judging whether the audio objects assigned to the BUSes are dynamic or static. Although creating ADM is required in pre-production unlike channel-based audio, this function enables a console setting to be linked with ADM semi-automatically.

The levels of secondary dialogue-objects were adjusted by the automatic level adjuster for secondary dialogue. It enables the control of the levels of secondary dialogue objects on the basis of the levels of the main dialogue that is adjusted manually [13]. The purpose of this mechanism is to reflect the engineer's intention of the level balance between dialogue and background sound. This equipment measures the similarity between speech timings of the main and secondary dialogues and three different objective indicators, and then determines the adjusted value of the secondary dialogue on the basis of these values. The objective indicators are calculated as the average of the top 25% of momentary loudness in three different periods (60, 20, and 2 seconds). The similarity prevents over-adjusting of the level of secondary dialogues in cases where speech timings between the main and secondary dialogues differ significantly.

Delivering NGA programme

The coding system compliant with MPEG-H 3DA baseline profile level 4 was used to encode the audio signals containing the S-ADM from the NGA console to the 3DA stream. In the case of homes, it was used to decode the streams and then render the decoded signals on the basis of the user's customizations (e.g., selecting preset programme, adjusting levels of audio objects, etc.) with the built-in renderer. As a result, it was verified that the programme audio produced by the prototype equipment could be rendered to in response to user's controls.

MANAGING LOUDNESS OF NGA PROGRAMME

Method to measure loudness for NGA programme

The loudness measurement method specified in Recommendation ITU-R BS.1770 is widely used to align perceived loudness among programmes. However, it does not cover NGA programmes and is undergoing revision work by the ITU-R Rapporteur Group. A method for measuring the loudness for NGA was proposed, in which the conventional method was applied to audio signals after rendering processes. To verify that the proposed method was appropriate, objective and subjective tests using several rendering conditions including loudspeaker configurations and rendering algorithms were conducted [17].

Overview of objective and subjective tests [17]

Objective Test

Thirty-one materials were created by selecting 12–16 second clips from eight programmes (e.g., sports event, documentary, music, and drama). These programmes consisted of multiple objects such as multilingual dialogues, enhanced dialogues, and multi-viewpoint dialogues (e.g., one to four mono objects and 7.1.2 object) and background sound (22.2, 7.1.4, or 5.1.4). In this objective test, the loudness for 868 test items was calculated using the proposed method to confirm the differences in objective loudness among the rendered audio signals under various rendering conditions. Test items were rendered from a total of 31 test materials to six loudspeaker configurations (22.2 (9+10+3), 7.1.4 (4+7+0), 5.1.4 (4+5+0), 7.1(0+7+0), 5.1 (0+5+0), and stereo (0+2+0) as specified in Recommendation ITU-R BS.2051 [15]) using five renderers (ITU-R ADM Renderer in the polar and Cartesian coordinate systems as specified in Recommendation ITU-R BS.2127 [18], MPEG-H 3DA Renderer in the polar coordinate system as specified in ISO/IEC 23008-3 [12], and Dolby

ATMOS Renderer in the Cartesian coordinate system with and without trim functions as specified in ETSI TS 103 448 [19]).

The mean and maximum absolute differences for 31 test materials depending on loudspeaker configurations and rendering algorithms are shown in Table 2 and Table 3. The results showed that loudspeaker configurations caused differences of up to 4.1 LU (average of 0.57–1.9 LU for 31 test items) and that rendering algorithms caused differences of up to 5.9 LU (average of 0.03–2.6 LU for 31 test items).

	Mean [LU]	Max [LU]
ITU-R ADM Renderer (polar)	1.1	3.1
ITU-R ADM Renderer (Cartesian)	1.1	3.1
MPEG-H 3DA Renderer	0.57	2.1
Dolby Atmos Renderer with trim function	1.9	4.1
Dolby Atmos Renderer without trim function	0.97	3.8

Table 2 – Mean and maximum absolute difference depending on rendering algorithms

	Mean [LU]	Max [LU]
stereo (0+2+0)	2.6	5.9
5.1 (0+5+0)	1.9	3.8
7.1 (0+7+0)	1.0	2.3
5.1.4 (4+5+0)	1.0	1.9
7.1.4 (4+7+0)	0.48	1.7
22.2 (9+10+3)	0.04	0.32

Table 3 – Mean and maximum absolute difference depending on loudspeaker

Subjective Test

The subjective test was conducted to confirm correlations and differences between the objective and subjective loudness. As subjects, 21 mixing engineers participated in this test. The subjects were assigned tasks to match the loudness between the reference and test items. They adjusted the playback level of the test item using the fader until the perceived loudness of the test item matched that of the reference item. They could switch between the reference and test items as often as necessary. The reference item was “FemaleInterviewer.wav”, as used in previous loudness listening tests [20–22]. The reference item was set to –24 LKFS, and its reproduction level was 60 dB via A-weighting at the listening position. It was presented at the Front-Center (M+000) loudspeaker as a single mono audio object. These were the same terms as those in a previous ITU-R listening test [20–22]. The initial objective loudness of test items was set within the range of –24±10 LKFS. In this test, five test materials (a sports event programme, a documentary programme, two music programmes, and a drama programme) were rendered into four configurations (stereo, 5.1, 5.1.4, and 7.1.4) by using the five rendering algorithms. In total, 98 items were used (20 items for individual rendering algorithms and three items, whose initial values were different, were evaluated twice to assess the consistency of the adjustments by the subjects).

Figure 6 – Figure 10 show the results of the subjective test for ITU-R ADM Renderer (with polar and Cartesian), MPEG-H 3DA Renderer, and Dolby Atmos Renderer with and without the trim functions, respectively. The horizontal axis indicates the subjective loudness corresponding to the adjusted value obtained by the subjects. The vertical axis indicates the objective loudness corresponding to the difference between the loudness of the reference item (–24 LKFS) and the initial loudness of the test item.

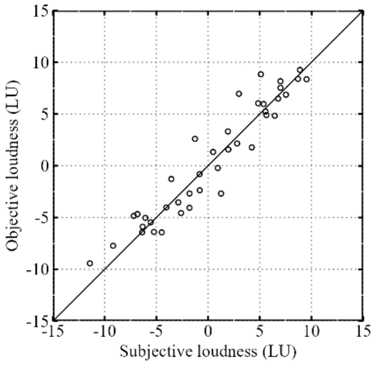

Figure 6 – Result for ITU-R ADM Renderer (polar)

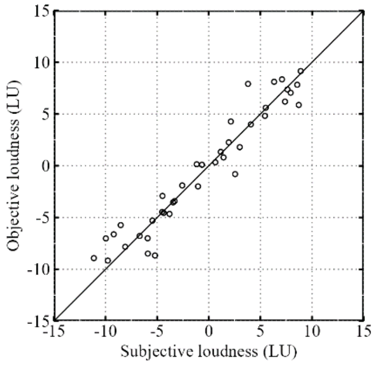

Figure 7 – Result for ITU-R ADM Renderer (Cartesian)

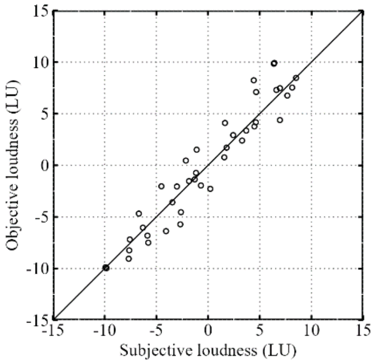

Figure 8 – Result for MPEG-H 3DA Renderer

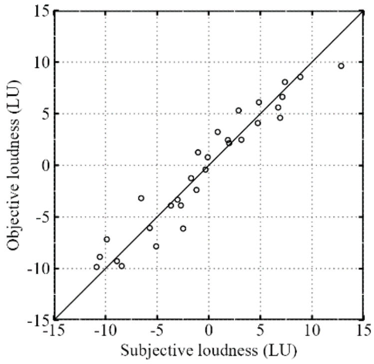

Figure 9 – Result for Dolby Atmos Renderer without trim function

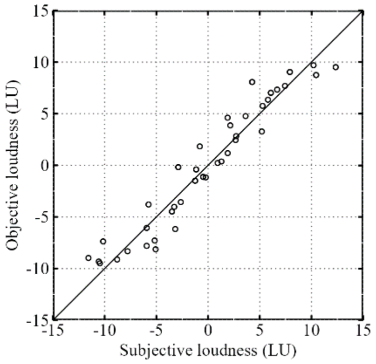

Figure 10 – Result for Dolby Atmos Renderer with trim function

Table 4 shows the statistics (CC: correlation coefficient; RMSE: root mean squared error; AAE: averaged absolute error; and MAE: maximum absolute error) for the test result depending on the rendering algorithms. The results show that CC and RMSE between the objective and subjective loudness are 0.958 and 1.7 LU, respectively, and that the absolute difference between the objective and subjective loudness is a maximum of 4.1 LU (average of 1.6–1.8 LU for five rendering algorithms).

	CC [-]	RMSE [LU]	AAE [LU]	MAE [LU]
ITU-R ADM Renderer (polar)	0.951	1.73	1.34	3.98
ITU-R ADM Renderer (Cartesian)	0.962	1.63	1.19	4.11
MPEG-H 3DA Renderer	0.957	1.73	1.32	3.82
Dolby Atmos Renderer with trim function	0.963	1.68	1.36	3.81
Dolby Atmos Renderer without trim function	0.959	1.76	1.40	3.81

Table 4 – Statistics for the result of subjective test

Consideration of managing loudness in the NGA live-production workflow

In the workflow experiment, our NGA loudness meter measured the loudness of the main output using the proposed method to measure audio signals obtained after rendering processes. This meter is equipped with a built-in renderer and can measure the loudness of audio signals under given rendering conditions: loudspeaker configuration and preset programme. The results of the subjective test showed that the correlation coefficient is 0.958. The MAEs were almost the same as those for the differences between the subjective and objective loudness in previous subjective tests for conventional channel-based audio [19–21]. This suggests that our NGA loudness meter works with almost the same differences between objective and subjective loudness as in the case of a loudness meter for channel-based audio. On the other hand, the results of the objective test showed that measured loudness differed up to 5.9 LU depending on rendering conditions, especially in the case of rendering a different configuration from the produced one. Although the use of a loudness meter equipped with rendering capability is considered to be appropriate in an NGA workflow, measuring loudness under multiple rendering conditions may be required for single programme audio. Thus, managing loudness would involve measuring loudness for each preset programme and under various rendering conditions, including at least the reference loudspeaker configuration that used in production (e.g., 22.2, a configuration of background sound), and a typical loudspeaker configuration with a small number of loudspeakers (e.g., stereo) that can cause a significant difference in loudness for a single programme simultaneously.

DISCUSSION

With the developed tools supporting S-ADM, the operation for live-production NGA test programmes that consisted of multiple dialogues with channel-based background sound was confirmed. The workflow for NGA differed from that for channel-based audio in terms of some operations: creating the ADM with static parameters in pre-production and monitoring audio signals with rendering processes. As mentioned in the proposed workflow, ADM creation would be replaced by the operation to select an appropriate template for programme content from the database and to modify the template of ADM on the basis of programme composition. In future work toward the practical use of NGA, a standard set of typical templates should be discussed and organized. The rendering processes cause unacceptable delays for mixing programme audio (e.g., 250 ms for the built-in renderer on the NGA console). To avoid delays in the rendering, two monitoring processes with different delays were used in the verified workflow. It is necessary to discuss guidelines for operating these processes in practice. In the workflow experiment, the levels of secondary dialogue objects were adjusted using the automatic level adjuster. This enabled the production of the secondary preset programmes without the additional audio consoles and mixing engineers. On the other hand, to improve the qualities of the secondary preset programmes, the needs to separately adjust the level of background sound for the main and secondary dialogues was revealed.

CONCLUSIONS

NGA live-production workflow, using production tools that handled S-ADM generated by the NGA console, was proposed. An experimental system, representing a broadcasting chain (from a broadcasting station to homes), was constructed. The programme audio that included S-ADM with dynamic parameters corresponding to the operations of the NGA console could be rendered in response to the user's controls. Although several issues (the management of the loudness, the creation of ADM, and the monitoring processes) were revealed in this experiment, they could be resolved within each step of the workflow. Thus, it was found that the proposed workflow is applicable to producing a typical NGA programme. Another advantage is that the system is connected via MADI among production equipment. This enables the implementation of NGA live-production systems while utilizing many existing SDI-based infrastructures in broadcasting stations. Future works are needed to study more complex NGA live-production workflow such as that requiring the integration of multiple ADM (e.g., ADM for studio and venue in live sports broadcasting).

REFERENCES

1. ITU-R, 2019. Audio definition model, Recommendation ITU-R BS.2076-2.
2. ITU-R, 2022. A serial representation of the Audio Definition Model, Recommendation ITU-R BS.2125-1.
3. ATSC, 2022. AC-4 System, ATSC Standard A/342 Part 2.
4. ATSC, 2022. MPEG-H System, ATSC Standard A/342 Part 3.
5. ETSI, 2017, Digital Video Broadcasting (DVB); Specification for the use of Video and Audio Coding in Broadcasting Applications based on the MPEG-2 Transport Stream, ETSI TS 101 154 V2.3.1.
6. EBU, 2021, Practical implementation of new open standards for Next Generation Audio production and interchange, EBU Tech Review.
7. Kubo, H., Nishiguchi T., Sugimoto T., and Oode S., 2021. Next-Generation Audio Console Supporting the S-ADM for Live Production of Immersive and Personalized Sound, Proceedings of the 2021 NAB Broadcast Engineering and Information Technology (BEIT) Conference.
8. Kubo H., Iwasaki T., Nishiguchi T., Oode S., and Okubo H., 2022. Development of loudness meter for object-based audio, Proceedings of Annual Convention 2022 of the Institute of Image Information and Television Engineers, 32C5 *(in Japanese).*
9. Sugimoto T., Aoki S., Oode S., Hasegawa T., Kubo H., and Okubo H., 2019, Real time audio encoding/decoding system using MPEG-H 3D Audio toward advancement of terrestrial broadcasting technology, Proceedings of the 23rd International Congress on Acoustics.

10. ITU-R, 2015. Algorithms to measure audio programme loudness and true-peak audio level, Recommendation ITU-R BS.1770-4.
11. ISO/IEC, 2014. Information technology – High efficiency coding and media delivery in heterogeneous environments – Part 1: MPEG media Transport (MMT), ISO/IEC 23008-1.
12. ISO/IEC, 2022. Information technology - High efficiency coding and media delivery in heterogeneous environments - Part 3: 3D audio, ISO/IEC 23008-3.
13. Kubo H., Nishiguchi T., and Oode S., 2021. Dialogue level auto adjustment based on multiple loudness with individual measurement periods on the production of second audio program, Proceedings of the 2021 Spring Meeting of the Acoustical Society of Japan, pp. 1355 – 1356 *(in Japanese)*.
14. SMPTE, 2020. Format for Non-PCM Audio and Data in AES3 – Carriage of Metadata of Serial ADM (Audio Definition Model), SMPTE Standard, SMPTE ST 2116:2019.
15. ITU-R, 2022. Advanced sound system for programme production, Recommendation ITU-R BS.2051-3.
16. Kubo H., Iwasaki T., Nishiguchi T., Oode S., and Nakayama Y., 2022. Development of dynamic metadata output function on the object-based audio console, Proceedings of Annual Convention 2022 of the Institute of Image Information and Television Engineers, 32C4 *(in Japanese)*.
17. Iwasaki T., Kubo H., and Oode S., 2022. Loudness of Next-Generation Audio contents depending on rendering conditions, the 154th Convention of the Audio Engineering Society.
18. ITU-R, 2019. Audio Definition Model renderer for advanced sound systems, Recommendation ITU-R BS.2127-0.
19. ETSI, 2016. AC-4 Object Audio Renderer for Consumer Use, ETSI TS 103 448 V1.1.1.
20. Soulodre G.A. and Norcross S.G., 2003. Objective Measures of Loudness, the 115th Convention of the Audio Engineering Society, paper 5896.
21. Komori T., Oode S., Ono K., Irie K., Sasaki Y., Hasegawa T., and Sawaya I., 2015. Subjective loudness of 22.2 multichannel programs, the 138th Convention of the Audio Engineering Society, paper 9219.
22. Sizle A., Sporer T., Liebetrau J., and Oode S., 2015. Progress in Standardization of 3D-Audio Loudness in ITU-R BS.1770, Proceedings of the 3rd International Conference on Spatial Audio, paper 013.