LiVRation: Remote VR live platform with interactive 3D audio-visual service

Takashi Kasuya
University of Tokyo / Takenaka Cooperation
kasuya@hongo.wide.ad.jp, kasuya.takashi@takenaka.co.jp

Manabu Tsukada
University of Tokyo
tsukada@hongo.wide.ad.jp

Yu Komohara
University of Tokyo
como@hongo.wide.ad.jp

Shigeki Takasaka
Sophisticated Engineers Department (SED)
s_takasaka@se-dept.co.jp

Takuhiro Mizuno
Alphacode Inc.
mizuno@alphacode.co.jp

Yoshitaka Nomura
Alphacode Inc.
yoshitaka.nomura@alphacode.co.jp

Yuta Ueda
CRI Middleware Inc.
ueda.yuta@cri-mw.co.jp

Hiroshi Esaki
University of Tokyo
hiroshi@wide.ad.jp

***Abstract*—Of late, various audio-visual services based on the internet are being deployed extensively. Among these, object-based audio-visual services are attracting more attention. In 2014, we had established the software defined media (SDM) consortium to investigate object-based and internet-based audio-visual services. Despite the increasing demand and popularity of live concert events, the placement of the microphone and camera limit the free-viewpoint watching of the contents of package media, such as DVDs. In this study, we design and implement an interactive 3D audio-visual service system called LiVRation, with a free-view-listen point. For subjective evaluation, 211 people were made to experience LiVRation and answer a questionnaire, subsequently. In addition, we demonstrated the system in the ”Billboard Live Hackasong 2017” hosted by Billboard Japan, and received the first prize, based on the votes of the judges as well as audience.**

I. Introduction

There is a rapid growth in internet-native audio-visual services, utilizing large-volume communication and lossless sound delivery technologies. We can easily enjoy 4K and 360-degree video with streaming services, such as YouTube and SteamVR, using VR devices. Higher quality, rich streaming media are expected to become common in the near future. On the other hand, despite the increasing demand and popularity of live and concert events, the reproduction features of package media, such as Blu-ray and DVD, are restricted by the location of the recording devices. Although we can play VR contents at home, most of the sound sources are static, and it is challenging to play dynamic high-quality streaming media, such as MPEG-4 audio lossless coding(MPEG4-ALS) [1].

Of late, the object-based approach, which represents multiple objects that exist in 3D space, has gained importance. Here, media data are decoupled from audio-visual and metadata, and transmitted to remote locations for the reproduction of 3D space. This allows flexible and possibly interactive reproduction that adapts to the receiver configuration, including head-mounted display (HMD), 3D TV, and 3D audio systems. In 2014, we had established the software defined media (SDM) consortium [1] for targeting new research areas and markets involving object-based digital media and internet-by-design audio-visual environments. The SDM is an architectural approach to media as a service, by the virtualization and abstraction of networked media infrastructure [2].

Currently, event recordings are generally recorded in package media at the positions and angles of predetermined cameras and microphones, and the recorded information is not reproducible. As the position and direction that can be viewed and listened to are highly limited, the content has significant restrictions. To solve these problems, we have developed a remote VR live system called LiVRation, which enables the reproduction of live experiences with free viewing angles, using VR and lossless delivery technologies based on SDM architecture.

This paper discusses the development and evaluation of LiVRation. The related work is described in section II. Section III describes the motivations and objectives of the paper. Section IV outlines the recording settings of live music and the system dataset. Section V presents the design, implementation, and content of LiVRation. Section VI reports its evaluation and demonstration. Section VII describes the demonstration of the system in an event, where we won the first prize. Finally, the conclusions and future challenges of this paper are presented in Section VIII.

II. Related Work

Sound recording and playback systems can be broadly divided into three categories[3]: channel-based, object-based, and scene-based. An object-based system records the sound color data of the sound source and its three-dimensional position as metadata, and renders the sound field from the loudspeaker position, in a regeneration environment. Examples include the Dolby Atmos [4] and AuroMax [5], which are being adopted in movie and home theaters. Object-based methods are being standardized in the moving picture experts group

[1] https://sdm.wide.ad.jp/

(MPEG) of the international organization for standardization (ISO) and the international electrotechnical commission (IEC)[6], [7]. Object-based methods require the location of the sound source to be recorded; however, they have the advantage of using a common microphone for recording. In addition, the sound fields can be calculated from the relative positional relationship between the audience and the reproducing sound source, enabling the audience to displace the viewer, and audio presentation following head rotation.

With the increase in IP-network speed, demonstration experiments are delivering multiple 4K60p images simultaneously. By delivering high-precision interactive video audios, the establishment of a remote collaboration system can enhance the sense of reality, such as stereotypes and emotions [8]. In addition, live delivery to digital cinema theaters using similar techniques enables theaters to be converted into public viewing venues. This will differentiate theaters from home theaters, which are also becoming more precise, and considerably enhance the value that they can provide. Business development is expected, and commercial demonstration experiments have been performed [9].

Open metadata descriptive methods, such as the resource discrimination framework (RDF) and linked open data (LOD) are currently being used. For example, in large-scale humanitarian-social tea databases [10] and knockout mouse phenotype databases [11], to facilitate the cross-sectional use of data. Vocabulary, which describes the relationship between data and the type (class) of data is called RDF vocabulary; RDF vocabulary, which describes music data, is defined as music ontology[12] and is extensively used by the broadcasting association of the United Kingdom (BBC).

III. Objectives and Requirements

LiVRation was developed to realize an interactive reproduction system for recorded music events with 3D audio-visual and free viewpoint. In addition, it dynamically handles the metadata related to singers, players, musical instruments, and contents from Twitter. We defined the requirements of LiVRation as follows:

- **3D audio-visual software-programmable environment** LiVRation is an SDM station [2], which can manage audio-visual objects as flexible systems by software-rendering. The user can himself locate the viewing position, based on which the recorded sounds are automatically synthesized in real time by software rendering.
- **Object-based audio-visual data management** A mechanism is required for decomposing, interpreting, and managing recorded media, independent of the application, into multiple viewing objects, etc. In this research, we study the design of an LOD database that responds to the requirements of the platform using 3D video and sound.
- **Social-networking service integration** Of late, audience are seeking a sense of unity with the venue and other participants, and the enjoyment of interacting with each other. They mainly use social networking services (SNSs), such as Facebook and Twitter. However, these are text-centered services, and their sense of unity is limited. Therefore, we integrate Twitter into the VR service for enhancing the sense of unity with the venue and the live experience of remote participants.
- **High Presence / Immersiveness** LiVRation can handle high-quality video, such as 4K and lossless streaming media. VR and vibration transmission (haptics) are used for more realistic live experiences. Live can be delivered in real time and viewed anywhere using the network, if a regenerative environment is established. Similarly, the sources of the recorded sounds can be delivered.
- **Interactiveness** The controller can be used to accentuate specific sounds or eliminate unnecessary ones. It can share comments in collaboration with social media.

IV. Recording and Dataset

The audio-visual data used in LiVRation was recorded at the rehearsal of Billboard Live Hackasong 2017, performed by the Japanese singer-songwriter "Kariya Seira" and her band at Billboard Live Tokyo in Roppongi, Japan, in January 2018. Billboard Live Tokyo accommodates approximately 300 people in three floors, including table seats (3F), sofa seats (4F) and casual seats (5F). Video editing and audio mixing were completed approximately 4 h after the recording, and the final presentation was made at the event. Fig. 1 shows the location of the band, and the placement of the 360-degree cameras and microphones.

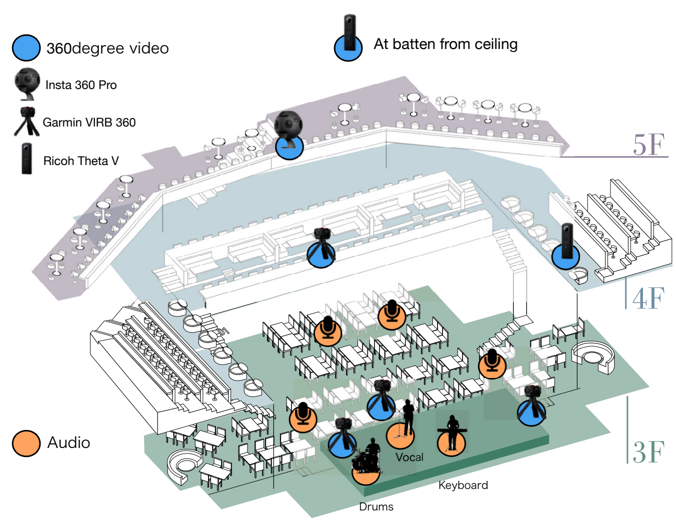

Fig. 1. Recording Environment

The band included drums, a keyboard, and bass. All the musical instruments were amplified and reconditioned by a sound reinforcement (SR) mixing console before being regenerated by a large loudspeaker. The sound pressure level was approximately 100 dBspl, depending on the location.

We installed four nondirectional microphones(DPA4090). Two microphones were placed on both sides of the stage at heights of 1 m, while the other two were connected to the ceiling at heights of 5 m. All the sounds were arranged in the SR mixing console, which digitally transfers the Head Amp bifurcations of the console to DAW. As all the recorded volume levels depend on the setting of the mixing console, a subsequent sound control task is essential. The recorded data were compiled into a vocal mix, drum mix, bass mix, and keyboard mix. In addition, we compiled the ambience mix on the left-back, right-back, left-front, and right-front of the audience seat. We used an Insta 360 Pro, four Garmin VIRB 360, and two Ricoh Theta V as the movie capture devices. All the cameras captured in 3840 x 1920@30 fps (4K). The Insta 360 Pro, which is a high-end 360-degree camera with six F2.4 fish-eye lenses, was located on the fifth floor to capture a full view of the venue. The Garmin VIRB 360, which is a 360-degree camera with two lenses, was installed in the center of the fourth floor, in front of the vocal and drums, respectively, and on the right-side of the stage. The Ricoh Theta V is also a 360-degree camera equipped with two lenses, and is light and compact; one was installed in a batten suspended on the ceiling of the venue, while the other was placed on the right hand of a fourth-floor seat.

V. LiVRation prototype

This section describes the design guidelines and implementation details of the system.

A. Design

LiVRation is designed, based on SDM architecture. Fig. 2 illustrates the design overview of LiVRation.

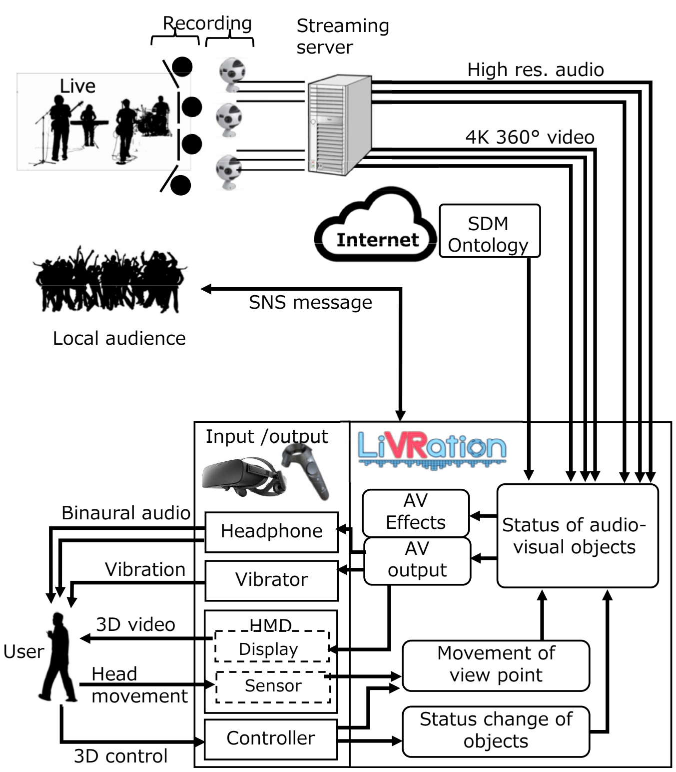

Fig. 2. Design of LiVRation

1) Overview: LiVRation reproduces a stereo-acoustic environment using HMD. MPEG4-ALS is used for delivering a streaming reproduction of high-resolution audio. The user can move in virtual space and experience acoustic space in the audible position, virtually. In addition, the user can extract/emphasize the sound of a selected musical instrument, obtain comments from Twitter, and feel the vibration from the vibration transmission devices. Fig. 4 displays a screenshot of LiVRation. A 360-degree video and audio are mapped to a spherical object displayed on the screen. An audio object is represented as a clear sphere with video effect; the length of the bar varies for each volume of the frequency component. A video image is also represented as a spherical object, with a projected video image. In addition, a Twitter message is displayed around the player(Twitter 3D barrage).

The sounds and movies described in section IV are edited and stored in the streaming server, which serves high-resolution audio and 4K 360-degree video to the LiVRation client. The client handles the audio/video streams as SDM objects representing audio and movie, and locates them in a prefabricated virtual space, as shown in Fig. 3. The user can change the view-listen point interactively in the virtual space around the stage and audience seats. LiVRation generates the view and synthesizes the binaural sound at the view-listen point. It also adds sound effects, such as sound attenuation, reverb effect, and acoustic Doppler effect.

2) SDM Ontology: As SDM applications are internet-based, the contents should be retrieved from the internet. To retrieve adequate content, we need metadata. This metadata should be published on the internet; for external use, it should include not only the video and audio of interest but also information on the location, orientation of the musical instruments, played songs and venues, and the recording processes. RDF is recommended for representing structured data on the internet. It provides a graph-based data model, whose statements are often referred to as RDF triples; they include a subject, predicate and object, implying directionality in the RDF graph. In addition, each resource/concept has a unified resource identifier (URI); hence, the RDF graph is explicitly labelled. When two identical URIs are found, their semantics are also considered identical. If they are published as open data on the internet, these types of linked data are called LOD, which accelerate the use of open data. We defined the RDF Vocabulary, "SDM ontology", for SDM applications [13]. The SDM ontology consists of five major components: ***SDMEvent***, which includes information on the entire event; ***Context***, which includes information on the program within the event; ***Target***, which includes the information to be recorded; ***Recorder***, which includes information on the recording equipment; ***MediaObject***, which is the media information to be generated.

3) Interaction: The user inputs the action using HMD controllers, and can listen to a specific sound by grabbing and pulling an audio object, such as a cappella, or a solo playing. Conversely, when an audio object is grasped and pushed to the opposite side, all the audios are activated, allowing the audio to be heard in a mixture of normal audios. In addition,

Fig. 3. SDM-object placement in virtual space

Fig. 4. Audio-object Manipulation

twisting the wrist, while grasping the audio object, can adjust the volume intensity (Fig. 4). The user can move into a 360-degree movie by selecting a video object with the pointer emitted from the controller. Of late, it has become common to interact with other audiences through SNS, such as Twitter, for a sense of unity in live broadcasting. LiVRation allows Twitter to interact with other audiences by presenting the relevant tweets in virtual space. In addition, we provide a more realistic experience by connecting a low-tone reacting oscillator through a low-pass filter from the headphone jack and attaching it to the body.

B. Implementation

Based on the design guidelines, LiVRation was implemented using the development environment and framework shown in Table. I.

TABLE I
IMPLEMENTED ENVIRONMENTS OF LIVRATION

Client PC	CPU Memory Graphics SSD	Intel Core-i7-8700K 32.0 GB GeoForce GTX 1080 Ti 480 GB
HMD		Samsung HMD Odyssey
Execution environment of client		Unity 2017.2.0f3(64-bit)
Streaming server		Wowza Streaming Engine 4.7.5
		NodeRED v0.17.5
SPARQL Endpoint		Graphdb-free-8.3.0
Time-series database		Elasticsearch v5.6.2

As depicted in Fig. 5, LiVRation clients have three functional modules that can be reflected and displayed on the screen, after receiving the contents, video and audio encoding, and rendering, as well as on receiving information from various services. Specifically, it has the following modules:

1) Rendering Modules: The streaming data from Wowza are processed and displayed by the Unity plugin. Streaming videos are mapped to the texture of the spherical object by AVPro video, which is also a Unity plugin, and reproduced as a 360-degree video. Audio is mapped to the audio object using the Universal Media player plugin, and a visual effect, which varies with the volume of the frequency component, is added. The Universal Media player is a plugin based on

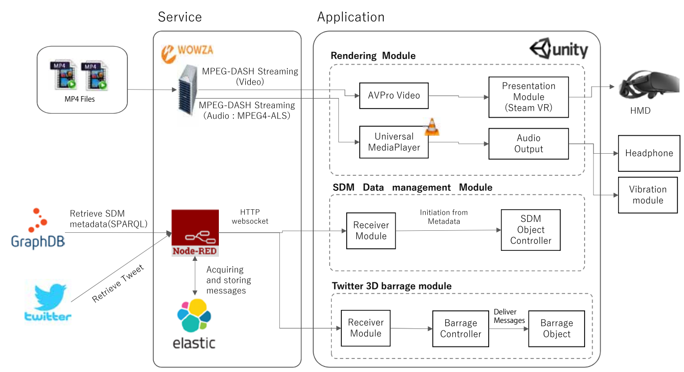

Fig. 5. Implementation of LiVRation

VLC and FFmpeg, and MPEG4-ALS can be regenerated. The visual effect is realized using the Rhythm Visualizator Pro plugin. The audio at the distance is automatically calculated by the Unity plugin, and the output audios can be reproduced by headphones or vibrating devices. The audio recorded by the 360-degree camera is muted.

Currently, there is no specific synchronization process for the playback of video and audio. The movie is played, when all the respective delivery contents are prepared in the plugin. The audio is automatically played, when the audio object is judged to be individually renewable. The replay timing can be adjusted to a certain extent by the parameters related to the buffer in the Universal Media player plugin, but due to the dispatch timing from Wowza, it is difficult to implement full synchronization with the video, by Unity. Although these advanced synchronous treatments are a challenge for the future, regeneration misalignment was not a concern in the local environment. In addition to the above functions, the HMD sensors and controllers can rotate and move through the spherical objects, and control the volume by manipulating the audio objects.

2) SDM Data Management Module: Metadata (sound sources, performers, positional information, etc.) with the SDM Ontology is acquired from GraphDB, which is a SPARQL endpoint, to the LiVRation client via the service layer. When the client is initialized by Unity and the service platform is accessed by HTTP, a prepared SPARQL query is sent to GraphDB and the results are sent to the client. Therefore, the metadata obtained is provided to the rendering module, etc. An excerpt of the ontology defined for the implemented LiVRation is shown below:

```
@prefix sdmo: <http://sdm.hongo.wide.ad.jp/sdmo/> .
@prefix sdm: <http://sdm.hongo.wide.ad.jp/resource/> .

###  Define SDMEvent
sdm:hackasong rdf:type sdmo:SDMEvent ;
  s:contentLocation sdm:billboard ;
  s:name "Billboard Live Hackasong 2017"@en ;
  s:startDate "2018/1/30 17:30" ;
  sdmo:has sdm:song1 ;
  sdmo:recording sdm:bass_mic ,
                 sdm:drum_mic ,
                 sdm:far_left_mic ,
                 sdm:far_right_mic ,
                 sdm:keyboard_mic ,
                 sdm:near_left_mic ,
                 sdm:near_right_mic ,
                 sdm:vocal_mic .

###  Define Context
sdm:song1 rdf:type sdmo:Song ;
  s:name "Colorful World" ;
  sdmo:performed_by sdm:bass ,
                    sdm:drum ,
                    sdm:keyboard ,
                    sdm:vocal ;
  sdmo:program_number "1" ;
  sdmo:recorded_by sdm:bass_mix ,
                   sdm:drum_mix ,
                   sdm:far_left_mix ,
                   sdm:far_right_mix ,
                   sdm:keyboard_mix ,
                   sdm:near_left_mix ,
                   sdm:near_right_mix ,
                   sdm:vocal_mix .

###  Define Target
sdm:vocal rdf:type sdmo:Musician ;
    s:name "Seyra Kariya"@en ;
    sdmo:attend_to sdm:hackasong ;
    sdmo:localX "0.0"^^xsd:float ;
    sdmo:localY "0.5"^^xsd:float ;
    sdmo:localZ "4.5"^^xsd:float ;
    sdmo:perform sdm:song1 .
```

SDMEvent *sdm: hackasong*, and *context sdm: song1*, which is linked by performed by (musican class) and recorded by, are linked by sound sources (AudioRecorder class). The player *sdm:vocal* is described as an annotation of each coordinate position (*sdm:local X-Z*) in which a sound source is placed. Similarly, for the sound sources, a URI for media delivery is described.

3) Twitter 3D barrage module: This is a module for displaying the comments from Twitter; it can also acquire hash tags and profile images from Twitter. The data from this module is displayed around the viewing position in succession. Similar to the SDM data management Module, the data from this module is continuously received by a websocket with the specific URI of the service platform, during application initialization. In addition, it is possible to control data visibility(display/hide) using the controller.

VI. Evaluation

We evaluated LiVRation by network analysis and subjective evaluation.

A. Network Evaluation

The contents of LiVRation include eight videos and seven audios, which were reproduced simultaneously. The reproduction time was 2 min and 14 s, and the data volume was approximately 160 M bytes and 5 M bytes, respectively.

To eliminate the influence of the internet-connection condition, we evaluated in a local network environment. Initially, all the movies were downloaded by the Unity plugin, for which the communication speed was approximately 300 Mbps. When the download was completed, only the audio streams were observed, which was approximately 15 Mbps. Audio streams can be observed continuously, until the end of the content. When using the streaming server on the internet, the movies were regenerated at approximately 30 s, with a maximum speed of approximately 100 Mbps. For household applications, cache server or middleware could be essential in local networks.

B. Subjective Evaluation

1) Method and subject: For subjective evaluation, we conducted a questionnaire survey to evaluate the viewing experience of interactive three-dimensional content using LiVRation. For instance, we investigated whether a stereoscopic effect was obtained from the 3D video/audio delivered by software, the tracking ability of the video/sound was sufficient, and the interactivity was adequate.

This evaluation was carried out in Interop Tokyo 2018 from the 13–15th of June 2018. According to the organizer, there were 143,806 participants in Interop 2018, the majority of whom were employed in information systems, network engineering, sales, and research. We asked the visitors to explain the usage of LiVRation; an experienced person was permitted to operate a device freely as a demonstration, after which data was obtained through a questionnaire.

Responses were obtained from a total of 211 people, including 181 men and 23 women, with seven unanswered questions. The age composition comprised 12 teenagers, 74 people in their 20s, 50 in the 30s, 48 in the 40s, 24 in the 50s, 2 in over 60, and there was one questionnaire without answer. 21 of them were audio-visual specialists.

2) Questionnaire items: The questionnaire items were evaluated using seven Likert scales, ranging from 1–7 (min:1, max:7), for each of the questions, Q1–Q7. The seven questions are shown in Fig6.

Questions Q1 and Q2 were regarding the fundamental three-dimensionality of the audio and video. Q3 and Q4 queried whether the combinations were perceived in correct agreement; Q3 was regarding the direction and position of the sound, when the user stopped moving, whereas Q4 was regarding the video and sound follow-up, when the user was moving. Q5 and Q6 were on the interactive viewing by the controller; Q5 queried the general ease of manipulation, whereas Q6 was regarding the visualizing and manipulation of the audio objects. Q7 queried whether the individual sounds from the audio objects were heard. Furthermore, we included a free comment field on the viewing and hearing experience.

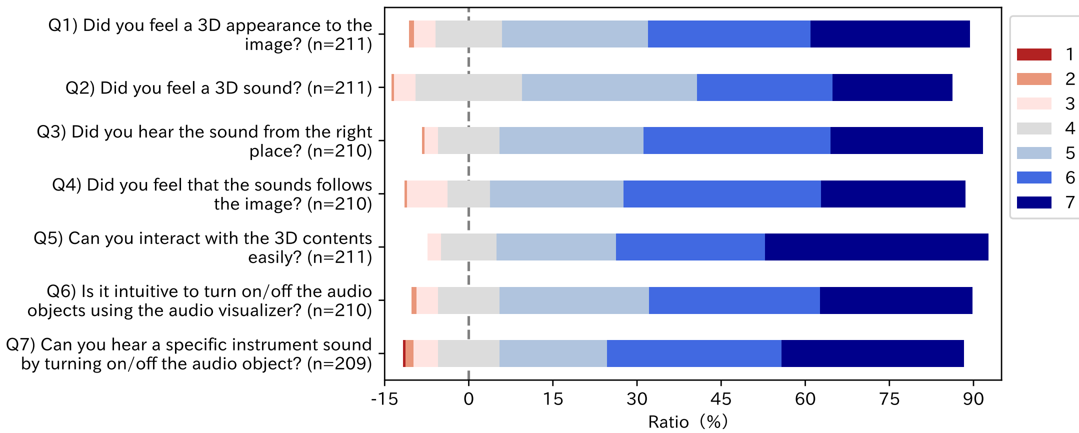

Fig. 6. Questionnaire Results with Likert scales

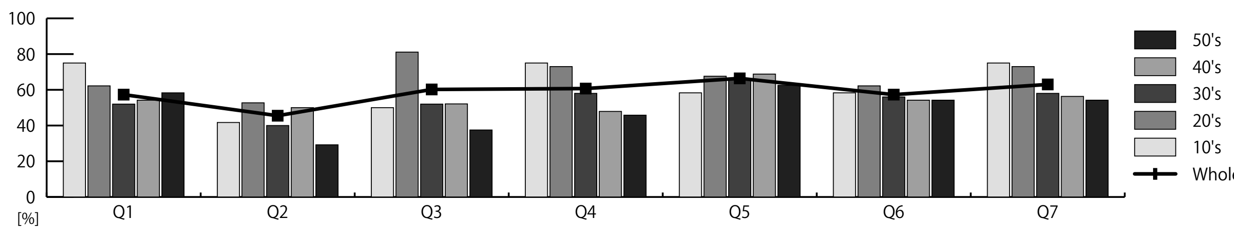

Fig. 7. Results of the generation-wise data extraction based on high scores(6 or 7)

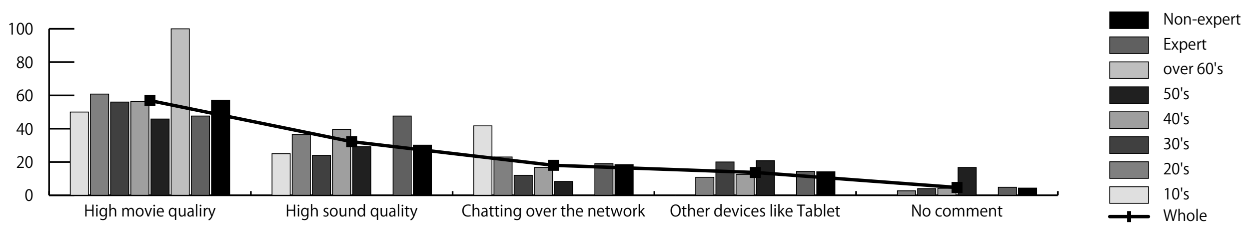

Fig. 8. Results of the requests for improvement

3) Results: Fig. 6 depicts the results. The X-axis represents the answer percentage from 17; the bar-graph position is depicted by considering value 4 at the center of the X-axis as 0%, with low scores on the left of the scale and high scores on the right.

Most items were rated 6 or 7. However, the proportion of the top 2 boxes of Q2 alone was less than half. We also analyzed the data based on gender and generation. On comparing by gender, the average point value of women was 0.34(11%) more than that of men. For generation-based analysis (Fig.7), data with high scores(6 or 7) were extracted, generation wise. Q4 and Q7 were highly rated by the 10s and 20s; Q3, in particular, was rated high by the 20s. Conversely, the ratings of Q3 and Q4, related to the voice, were relatively low, by the 40s and 50s. Data extracted based on poor scores (1 or 2) indicated a low rating for Q7 by the 50s. Younger people appeared to be more accustomed to similar VR and game productions.

In addition, we queried the users on the most required important feature in the next upgraded version (Fig.8). Majority answered, ”high movie quality”(56.9%), followed by ”high sound quality” (32.2%). However, video and audio experts considered high sound quality as important as ”high movie quality”. For teenagers and students, there was a strong tendency to emphasize ”being seen, while chatting over the network”.

There were several comments such as, ” I want to use it for watching sports” and ”I would like to try to experience, even paying a fee”. However, there were several comments on the difficulty of understanding the interface, for instance, the gesture of grabbing and pulling an audio object was not intuitive, and difficult to use. It was also indicated that the performance of the headphone and the quality of the 360-degree video hindered immersiveness.

VII. Demonstration

We demonstrated LiVRation during the "Billboard Live Hackasong 2017" held in Billboard Live Tokyo on 30th January 2018.

In the demonstrations, the audio and video views of the operators on the HMD stage were reclaimed by the projectors at the venue, and the speakers on the left and right sides of the stage. The performance was streamed on the internet, and watched by more than 3,600 people. The comments received by the auditors included, "High completeness" and "I would like to use it immediately." As a result, we received the first prize, accorded by the judges, and the audience prize determined by the number of audience votes. We believe that this award has enabled us to confirm our business potential.

VIII. Conclusion and Future work

In this paper, we have presented LiVRation, which enables the viewing of live recordings from a free perspective, using HMD, and free sound control. We have achieved the objectives and requirements defined in section III. However, problems with respect to the recording environment, regeneration environment, and HMD interface, have emerged.

In the recording environment, blown out highlights were observed, in the area where the spotlight strikes. Live venues are generally dark, and it is difficult to adjust the contrast during recording. The establishment of recording techniques in these environments is also a challenge.

Professional mixing is essential for high-quality audio content. In the prototype environment, the reproduction timing of each audio and video was configured, depending on Unity. However, in a poor network environment, the regeneration timing was markedly missed, inhibiting immersiveness. However, even in actual live environments, lag occurs at different distances. The solution may include middleware or cache for the synchronization of audio and video in local environments, with a mechanism for flexible correction at different distances using SDM ontology.

In future, we plan to improve the user interface and the immersiveness of LiVRation. In addition, we intend to examine the linkage of the SDM application with BIM (Building Information Modeling). LiVRation uses a shape model of the venue created by SketchUp[2]. Recent construction projects have generalized BIM-based designs that include information on 3D shapes and attributes. By combining BIM data with SDM ontology, we believe that an SDM application can be created easily. In addition, we would like to continue our activities with a view on the business development of LiVRation, in future.

References

[1] *Information technology - Coding of audio-visual objects - Part 3: Audio, AMENDMENT 4: Audio Lossless Coding (ALS), new audio profiles and BSAC extensions*, International Organization for Standardization (ISO) Std. ISO/IEC 14 496-3:2009, Rev. MPEG4-ALS, 2009.

[2] M. Tsukada, K. Ogawa, M. Ikeda, T. Sone, K. Niwa, S. Saito, T. Kasuya, H. Sunahara, and H. Esaki, "Software Defined Media: Virtualization of Audio-Visual Services," *IEEE International Conference on Communications (ICC2017)*, May 2017, paris, France.

[3] ITUR, "ITU-R BS. 2051-0 (02/2014) Advanced Sound System for Programme Production," *Int. Telecommun. Union, Geneva, Switzerland*, 2014.

[4] "Dolby Atmos© Specifications," Dolby Laboratories, Tech. Rep. Issue 3, 2015.

[5] "AUROMAX© Next generation Immersive Sound system," November 2015. [Online]. Available: http://www.auro-3d.com/wp-content/uploads/documents/AuroMax_White_Paper_24112015.pdf

[6] J. Herre, J. Hilpert, A. Kuntz, and J. Plogsties, "MPEG-h 3D audio — the new standard for coding of immersive spatial audio," vol. 9, no. 5, pp. 770–779, Aug. 2015.

[7] R. Bleidt, A. Borsum, H. Fuchs, and S. M. Weiss, "Object-based audio: Opportunities for improved listening experience and increased listener involvement," *SMPTE Motion Imaging Journal*, vol. 124, no. 5, pp. 1–13, Jul. 2015.

[8] S. Kim, T. Nakachi, T. Fujii, S. Emura, and Y. Haneda, "Highly realistic collaboration system by using 4K multi JPEG2000 codecs and 6-channel acoustic echo canceller," in *Proc. Int. Symp. Intelligent Signal Processing and Communications Systems*, Nov. 2012, pp. 452–457.

[9] T. FUJII, T. FUJII, S. ONO, K. SHIRAKAWA, and D. SHIRAI, "Ods live streaming technology to digital cinema theaters," *IEICE ESS Fundamentals Review*, vol. 5, no. 1, pp. 80–89, 2011.

[10] M. GOTO, "Apply linked data to large-scale humanities database," *Joho Chishiki Gakkaishi*, vol. 25, no. 4, pp. 291–298, 2015.

[11] M. E. Dickinson, A. M. Flenniken, X. Ji, L. Teboul, M. D. Wong, J. K. White, T. F. Meehan, W. J. Weninger, H. Westerberg, H. Adissu *et al.*, "High-throughput discovery of novel developmental phenotypes," *Nature*, 2016.

[12] Y. Raimond, S. A. Abdallah, M. B. Sandler, and F. Giasson, "The Music Ontology," *Proceedings of the International Conference on Music Information Retrieval*, pp. pp. 417–422, 2007.

[13] R. Atarashi, T. Sone, Y. Komohara, M. Tsukada, T. Kasuya, H. Okumura, M. Ikeda, and H. Esaki, "The software defined media ontology for music events," in *Workshop on Semantic Applications for Audio and Music (SAAM) held in conjunction with ISWC 2018*, Monterey, California, USA., Oct. 2018.

[2]https://www.sketchup.com/