Multichannel Audio Transmission over IP Network by MPEG-4 ALS and Audio Rate Oriented Adaptive Bit-rate Video Codec

Yutaka Kamamoto, Noboru Harada, Takehiro Moriya, Sunyong Kim, Takahiro Yamaguchi, Masanori Ogawara, and Tatsuya Fujii

Abstract

This article describes an experiment of lossless audio transmission over an Internet protocol (IP) network and introduces a prototype codec that combines lossless audio coding and variable bit rate video coding. In the experiment, 16-channel acoustic signals compressed by lossless audio coding (MPEG-4 Audio lossless coding (ALS) standardized by the Moving Picture Experts Group) were transmitted from a live venue to a café via the IP network. At the café, received sound data were decoded losslessly and then appropriately remixed to adjust to the environment at that location. The combination of high-definition video and audio data enables fans to enjoy a live musical performance in places other than the live venue. This experiment motivated us to develop a new prototype codec that guarantees high audio quality. The developed codec can control the bit rates of both audio and video signals jointly, and it achieves high audio and video quality.

1. Introduction

Network quality has improved recently, and the storage size has also been rapidly expanding. The Next Generation Network (NGN) can provide high quality, secure, and reliable services [1]. This higher bit rate, almost error-free and low-delay communication network makes it possible to transmit high-definition content almost in real time [2]. In this environment, the use of lossless codecs (coders/decoders) is widespread these days [3]–[7]. Audio lossless coding can perfectly reconstruct the signal from the bit stream. Users are able to choose not only the efficiency of lossy coding, at the sacrifice of quality, but also the reliability of acoustic signals by lossless coding because the network bit rate and disk space are also increasing. However, since the size of bit streams encoded by a lossless coder depends on the characteristics of the input signals, the lossless coding results in a variable bit rate that cannot be controlled. This is in contrast to lossy coding such as MPEG[*1]-2/4 advanced audio coding (AAC) that is applied for portable music players and broadcasts [3], [4], [8]–[10].

One of the most promising applications of broadband networks including NGN is high-quality video and audio transmission. Digital cinema is one potential application of this, as well as content distribution from popular theaters and music halls. We sometimes refer to such a content delivery system as *other digital stuff* or *online digital sources* (ODS) [11], [12]. In Japan, several musical artists such as the Takarazuka

*1 The moving picture experts group (MPEG) is a working group of international organization for standardization (ISO) and the international electrotechnical commission (IEC) that develops standards for coded representation of digital audio and video and related data.

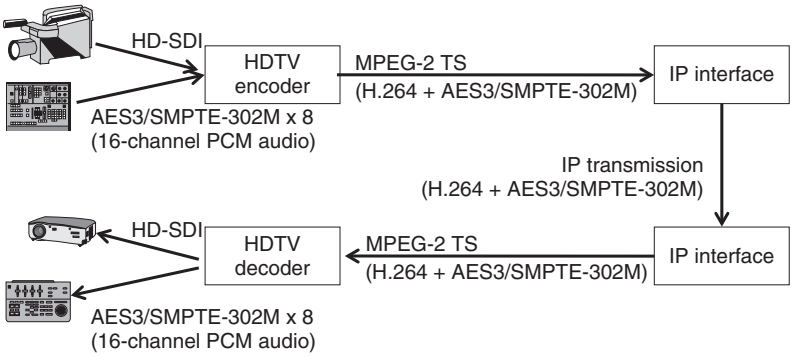

Fig. 1. System diagram of existing settings.

Revue Company [13], X-Japan, and L'Arc~en~Ciel [14] have provided live performances in real time to fans located in other places such as movie theaters. Video signals are normally compressed by ITU-T H.264/MPEG-4 AVC[*2], and audio signals by AAC to save the bit rate. Even though these transmissions are of music content, a lossy codec has been used for sound data. In addition, multichannel audio signals from the live stage are sometimes mixed down to two channels, and then the processed stereo data are transmitted to the other venue. Lossy compression and down-mixing are reasonable when the speaker settings are defined and the acoustic characteristics are the same in each place of delivery, but obviously such assumptions are not realistic. Consequently, there is room to improve the audio quality of ODS. One feasible idea is to transmit the sound of the musical instrument as-is (i.e., without down-mixing or lossy coding), and to down-mix at each local site.

To achieve a way to provide high-quality music, we carried out an experiment of lossless transmission of sound data. MPEG-4 Audio lossless coding (ALS) [15]–[17] was used to losslessly encode the 16-channel musical instrument data because it is an international standard technology and supports multichannel audio signals. The compressed audio data were transmitted from a live music venue (sender) to a café via an Internet protocol (IP) network. Although the bit rate of lossless audio became larger than that of the lossy one, we did not have to worry about degradation of sound waveforms. At the café, the transmitted sound data were decoded without any loss and appropriately remixed to adjust to the environment of the site. Lossless audio coding allowed the listeners in the offsite venue to enjoy the concert with customized audio data.

The experimental result from this trial motivated us to develop a new prototype codec that can control the bit rate of video and audio. Audio quality is guaranteed, and higher video quality is achieved by making use of extra bits saved by lossless audio compression.

The remainder of this article is as follows. We report on the lossless audio transmission experiment in section 2 and introduce the prototype codec in section 3. Finally, we conclude the article in section 4.

2. Investigation of lossless compression efficiency in demonstrative trial

2.1 Configuration

As described in the previous section, we carried out an experiment of lossless transmission of sound data to provide high-quality music. We used the high-definition television (HDTV) encoder/decoder (HV9100 series by NTT Electronics Corp. (NEL)) and the IP interface (NA5000 by NEL) as shown in **Figs. 1** and **2**. The HDTV encoder output an MPEG-2 transport stream (TS), which included eight pairs of AES3/SMPTE-302M[*3] (i.e., 16-channel pulse-code modulation (PCM) audio signals) bitstreams that were input from a digital mixer and an H.264 bitstream that consisted of the encoded high-definition serial digital interface (HD-SDI) data from a high-vision camera.

*2 ITU-T H.264/MPEG-4 AVC was prepared jointly by international telecommunication union telecommunication standardization sector (ITU-T) and by MPEG.

*3 AES3 is a standard that specifies the transport of digital audio signals between professional devices (published by the Audio Engineering Society); SMPTE-302M is a standard that specifies the transporting of AES3 data in an MPEG-2 transport stream for television applications (published by the Society of Motion Picture and Television Engineers).

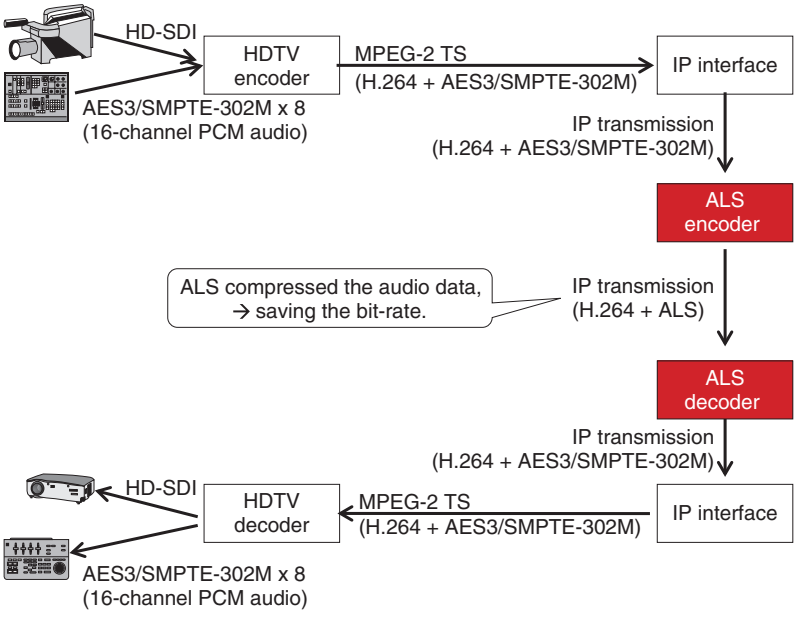

Fig. 2. System diagram of the experiment.

The MPEG-2 TS bitstream was converted to IP packets by the IP interface. Audio and video data were transmitted over the IP network. After transmission, an IP interface and HDTV decoder carried out the reverse operations. Finally, we obtained high-vision data and 16-channel PCM audio signals.

Although this existing setting (Fig. 1) can provide high-quality sound, it wastes network resources because there are redundant audio signals. From an ecological and economical viewpoint, it would be preferable to reduce the cost of the network without any degradation of audio quality.

For this experiment, we added the ALS encoder/decoder to save the bit rate, as shown in Fig. 2. Before IP transmission, the ALS encoder losslessly compressed eight pairs of the AES3/SMPTE-302M bitstreams. After the ALS decoder had received the encoded IP packets, it reconstructed the original IP packets and sent them to the IP interface. Then, the IP interface and HDTV decoder operated as if no processing had taken place.

The total bit rate was set to 50 Mbit/s (**Fig. 3**). The audio bit rate was about 18 Mbit/s because 16-channel signals were digitized at 48 kHz and 20 bits (four more bits were required for AES3/SMPTE-302M). Therefore, 32 Mbit/s was needed for the target bit rate of H.264 (video), which was not varied during the experiment. Then, the audio data were losslessly encoded by ALS, and the bit rate of the ALS bitstream was reduced to less than 18 Mbit/s, depending on the input signals. The transmitted ALS bitstreams were perfectly decoded to PCM audio signals.

2.2 Experimental results

This experiment was conducted on March 19, 2009 from 7:00 to 10:00 PM. There were about 80 listeners at the live venue and around 100 at the café. The sender and receiver sites are shown in **Figs. 4** and **5**, respectively.

The bit rate of the ALS bitstreams on average for each minute is shown in **Fig. 6**. The audio signals are losslessly compressed to 21% (3.78 Mbit/s)–60% (10.87 Mbit/s) of the original data size. The bit rate increased as the performance progressed because the number of musicians (i.e., musical instruments) also increased. In summary, the ALS was able to save about 11.2 Mbit/s on average, so the total bit rate became around 40 Mbit/s.

Sound data were appropriately remixed for adjustment to the environment of the receiving site. In this experiment, it was easy for the mixing engineers (also known as public address or sound reinforcement engineers) at the café to carry out their remixing tasks because the original sound data came from the live venue.

We received positive feedback from the invited

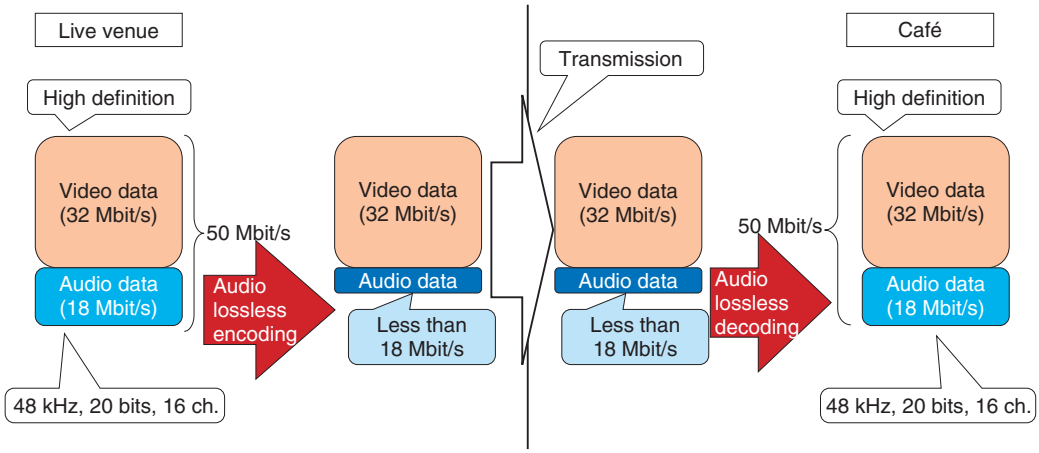

Fig. 3. Conceptual diagram of the experiment.

Fig. 4. Live venue in Shimokitazawa, Tokyo (sender site).

Fig. 5. Café in Aoyama, Tokyo (receiver site).

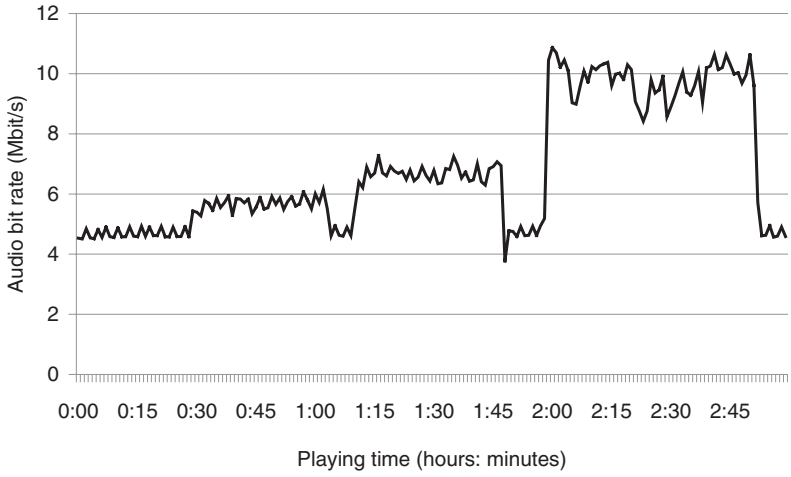

Fig. 6. Bitrate of compressed audio.

guests, and we believe that assigning a higher bit rate for multichannel audio signals was key in the success of the experiment. Our experiment suggests that the combination of high-definition video and audio data, especially ODS with lossless audio coding such as the ALS, will enable fans to enjoy live musical performances from offsite locations. We think this trial is a good example of a content delivery service with high-quality audio. As far as we know, this is the first experiment on real-time lossless audio transmission (especially ALS) with HD video via an IP network.

3. Prototype of audio rate-oriented adaptive bit-rate video codec

3.1 Concept of the developed codec

The results of the experiment described in the previous section indicate that we can reduce the bit rate of audio signals by approximately half by using ALS. For real-time streaming, we should nevertheless retain the worst-case bit rate (i.e., PCM rate) for audio. We developed a prototype of an audio bit-rate-oriented adaptive bit-rate video codec to determine whether we could make efficient use of the reduced bit rate, as shown in **Fig. 7**. The bit rate saved by using ALS can then be contributed to H.264.

Let V be the bit rate of video, A be that of audio, and C be that of losslessly compressed audio (i.e., $C \leqq A$). As shown in **Fig. 8**, the conventional codec needs $V + A$ bit/s to transmit the data. By contrast, the developed codec can provide the additional bits for the video codec. Therefore, the video codec can use $V + (A - C)$ bit/s, and the audio codec needs only C bit/s. The video quality is therefore enhanced. The total bit rate for the TS is the same, $V + A$ bit/s, and the audio quality is also the same because we use lossless coding.

3.2 Preliminary experiment with the developed codec

The prototype encoder supports.

- HD-SDI signal input → H.264 bitstream output
- Three AES3/SMPTE-302M inputs (i.e., up to 6 channels) → MPEG-4 ALS bitstream output.

The decoder supports.

- H.264 bitstream input → HD-SDI signal output
- MPEG-4 ALS bitstream input → three AES3/SMPTE-302M outputs.

In the preliminary experiment with the developed codec, the TS rate was set to 25 Mbit/s because we assumed 30 Mbit/s for the NGN operation for IP transmission (which is reasonable for users) and because forward error correction usually requires

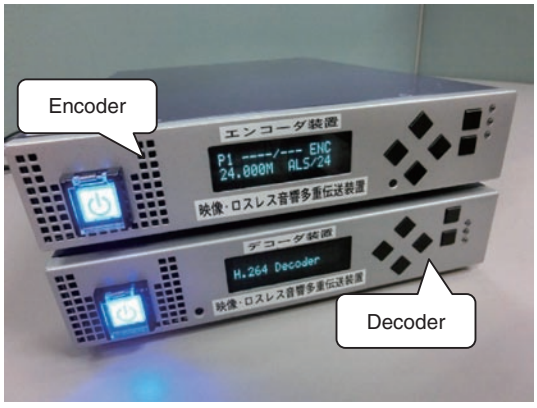

Fig. 7. The fabricated encoder and decoder.

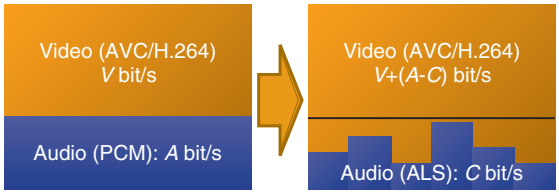

Fig. 8. Conceptual diagram of prototype codec.

10% of the bit rate. Content from an opera performed and provided by a famous Japanese opera company was used as input signals.

A diagram of the settings for the preliminary experiment is shown in **Fig. 9**. A Blu-ray player output an HDMI (high-definition multimedia interface) signal, and the converter divided it into HD-SDI and AES3/SMPTE-302M signals. The prototype encoder produced the H.264 bitstream from the obtained HD-SDI signal, of which the target bit rate depends on the bit rate of audio data compressed by the ALS, and produced the ALS bitstream from the AES3/SMPTE-302M signals. The decoder then output the HD-SDI signal and reconstructed the AES3/SMPTE-302M signals losslessly.

We captured TS bitstreams and analyzed them in order to evaluate the performance of the fabricated codec. Without lossless audio coding, the bit rate of audio requires 6.9 Mbit/s (48 kHz x 24 bits x 6 channels), which means that the remaining bit rate is only 18 Mbit/s for video. In contrast, as shown in **Fig. 10**, the bit rate of video can use more than 18 Mbit/s (statistically 18.40 (mean) ± 0.55 (standard deviation) Mbit/s) because the bit rate of audio is

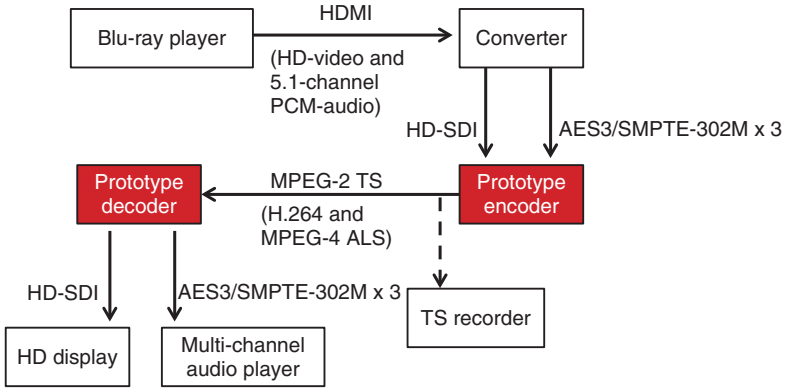

Fig. 9. Configuration for preliminary experiment.

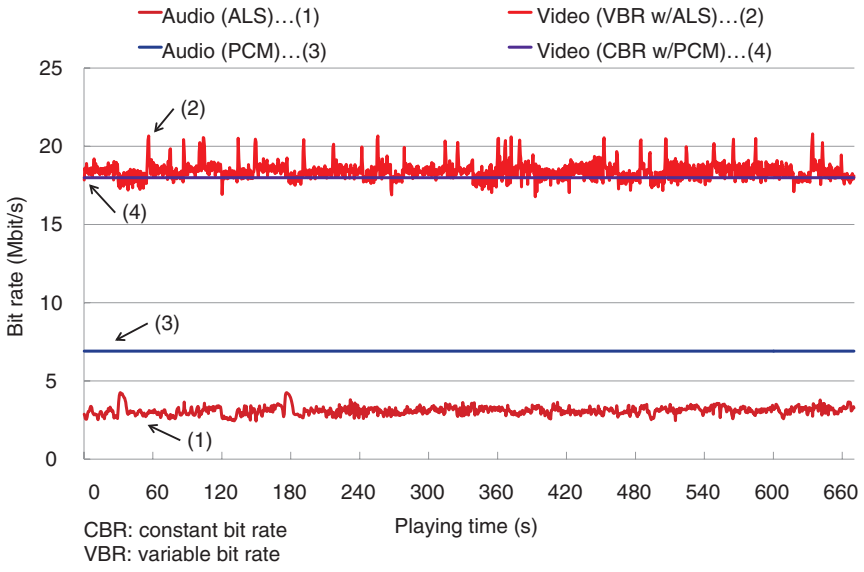

Fig. 10. Bit rates in preliminary experiment.

compressed to around 3 Mbit/s (statistically 3.10 (mean) ± 0.26 (standard deviation) Mbit/s). In summary, video quality is improved by utilizing lossless audio compression.

This codec is still a prototype, so we decided to use low-risk, low-return settings for the bit rate and adaptive bit-rate control. We can achieve better quality by fine-tuning these settings.

4. Conclusion

We conducted an experiment of lossless audio transmission via an IP network. Multichannel audio signals compressed by MPEG-4 ALS and high-definition video signals were transmitted from a live venue to a café to provide high-quality music for an off-site audience. This experimental result motivated us to develop a prototype codec that controls the bit rate between video and audio. Audio quality is guaranteed, and higher video quality is achieved by making use of extra bits saved by the lossless audio compression, while the video encoder is able to use the remaining bits subject to the constant total bit rates. After the prototype codec is refined, we will be able to transmit higher quality content such as operas, musicals, and concerts. Japanese broadcasting

standards support MPEG-4 ALS for downloading 22.2-channel high-definition audio [18], and the developed audio/video codec was used to transmit the high-quality multichannel audio signal of a Takarazuka Revue performance [19]. Thus, the described technologies are expected to be widely used in the near future.

Acknowledgements

We thank all of the engineers who assisted with the development and experiments.

References

[1] S. Esaki, A. Kurokawa, and K. Matsumoto, "Overview of the Next Generation Network," NTT Technical Review, Vol. 5, No. 6, 2007. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr200706sf1.html

[2] K. Kawazoe, R. Kakinuma, Y. Haneda, D. Minoura, S. Minamoto, and H. Ishimoto, "Platform Application Technology Using the Next Generation Network," NTT Technical Review, Vol. 5, No. 6, 2007. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr200706sf3.html

[3] A. Spanias, T. Painter, and V. Atti, "Audio Signal Processing and Coding," John Wiley & Sons, Inc., 2007.

[4] S. Salomon and G. Motta, "Handbook of Data Compression," Springer, 2010.

[5] M. Hans and R. W. Schafer, "Lossless Compression of Digital Audio," IEEE Signal Processing Magazine, Vol. 18, No. 4, pp. 21–32, 2001.

[6] K. Konstantinides, "An Introduction to Super Audio CD and DVD-Audio," IEEE Signal Processing Magazine, Vol. 20, No. 4, pp. 71–82, 2003.

[7] B. H. Kuzuki, N. Fuchigami, and J. R. Stuart, "DVD-Audio Specifications," IEEE Signal Processing Magazine, Vol. 20, No. 4, pp. 72–90, 2003.

[8] ISO/IEC 13818-7, "Information technology—Generic coding of moving pictures and associated audio information—Part 7: Advanced Audio Coding (AAC)," Dec. 1997.

[9] ISO/IEC 14496-3, "Information technology—Coding of audio-visual objects—Part 3: Audio," Dec. 1999.

[10] ARIB STD-B32, "Video Coding, Audio Coding and Multiplexing Specifications for Digital Broadcasting," May 2001.

[11] S. Kim, M. Ogawara, T. Fujii, Y. Kamamoto, N. Harada, and T. Moriya, "Requirements for Developing Ultra-Realistic Live Streaming Systems," Proc. of the 2009 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS 2009), pp. 175–178, Kanazawa, Japan.

[12] H. Takahashi, D. Shirai, T. Murooka, and T. Fujii, "Multipoint Streaming Technology for 4K Super-high-definition Motion Pictures," NTT Technical Review, Vol. 5, No. 5, 2007. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr200705le1.html

[13] Press release on live video transmission of Takarazuka Review performance (in Japanese). http://www.ntt.co.jp/news/news07/0711/071114a.html

[14] Information on live video transmissions of L'Arc~en~ciel performance (in Japanese). http://www.larc-en-ciel.com/m/news/larc/other/l_ot_ar-08.html

[15] Information technology—Coding of audio-visual objects—Part 3 Audio, 3rd Ed. Amendment 2: Audio Lossless Coding (ALS), new audio profiles and BSAC extensions, ISO/IEC Std. 14496-3:2005/AMD.2:2006, March 2006.

[16] T. Liebchen, T. Moriya, N. Harada, Y. Kamamoto, and Y. Reznik, "The MPEG-4 Audio Lossless Coding (ALS) standard—Technology and applications," Proc. of the 119th Audio Engineering Society Convention, Paper #6589, New York, NY, USA, 2005.

[17] Y. Kamamoto, N. Harada, T. Moriya, S. Kim, M. Ogawara, and T. Fujii, "Experiment of Sixteen-Channel Audio Transmission Over IP Network by MPEG-4 ALS and Audio Rate-Oriented Adaptive Bit-Rate Video Codec," Proc. of the 129th Audio Engineering Society Convention, Paper #8302, San Francisco, CA, USA, 2010.

[18] ARIB STD-B45, "Content Download System for Digital Broadcasting," Apr. 2010.

[19] H. Yamane, A. Yamashita, K. Kamatani, M. Morisaki, T. Mitsunari, and A. Omoto, "High-presence Audio Live Distribution Trial," NTT Technical Review, Vol. 9, No. 10, 2011. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201110fa4.html

Yutaka Kamamoto

Research Scientist, Moriya Research Laboratory, NTT Communication Science Laboratories.

He received the B.S. degree in applied physics and physico-informatics from Keio University, Kanagawa, in 2003 and the M.S. and Ph.D. degrees in information physics and computing from the University of Tokyo in 2005 and 2012, respectively. Since joining NTT Communication Science Laboratories in 2005, he has been studying signal processing and information theory, particularly lossless coding of time-domain signals. He additionally joined NTT Network Innovation Laboratories, where he developed the audio-visual codec for ODS from 2009 to 2011. He has contributed to the standardization of coding schemes for MPEG-4 Audio lossless coding (ALS) and ITU-T Recommendation G.711.0 Lossless compression of G.711 pulse code modulation. He received the Telecom System Student Award from the Telecommunications Advancement Foundation (TAF) in 2006, the IPSJ Best Paper Award from the Information Processing Society of Japan (IPSJ) in 2006, the Telecom System Encouragement Award from TAF in 2007, and the Awaya Prize Young Researcher's Award from the Acoustical Society of Japan (ASJ) in 2011. He is a member of IPSJ, ASJ, the Institute of Electronics, Information and Communication Engineers (IEICE), and IEEE.

Noboru Harada

Senior Research Scientist, Moriya Research Laboratory, NTT Communication Science Laboratories.

He received the B.S. and M.S. degrees from the Department of Computer Science and Systems Engineering of Kyushu Institute of Technology, in 1995 and 1997, respectively. He joined NTT in 1997. His main research area has been lossless audio coding, high-efficiency coding of speech and audio, and their applications. He additionally joined NTT Network Innovation Laboratories, where he developed the audio-visual codec for ODS, from 2009 to 2011. He is an editor of ISO/IEC 23000-6:2009 Professional Archival Application Format, ISO/IEC 14496-5:2001/Amd.10:2007 reference software MPEG-4 ALS, and ITU-T G.711.0. He is a member of IEICE, ASJ, the Audio Engineering Society (AES), and IEEE.

Takehiro Moriya

NTT Fellow, Moriya Research Laboratory, NTT Communication Science Laboratories.

He received the B.S., M.S., and Ph.D. degrees in mathematical engineering and instrumentation physics from the University of Tokyo in 1978, 1980, and 1989, respectively. Since joining the Nippon Telegraph and Telephone Public Corporation (now NTT) in 1980, he has been engaged in research on medium to low bitrate speech and audio coding. In 1989, he worked at AT&T Bell Laboratories, NJ, USA, as a visiting researcher. Since 1990, he has contributed to the standardization of coding schemes for the Japanese Public Digital Cellular system, ITU-T G.729 and G.711.0, and ISO/IEC MPEG MPEG-4 General Audio and MPEG-4 ALS. He is a Fellow member of IEEE and a member of IPSJ, IEICE, AES, and ASJ.

Sunyong Kim

Research Engineer, Media Innovation Laboratory, NTT Network Innovation Laboratories.

She received the B.E. and M.E. degrees in information science from the University of Tokyo in 2002 and 2004, respectively. She joined NTT in 2004 and has been researching conversation scene analysis on highly realistic remote collaboration systems. She is a member of IPSJ and IEICE.

Takahiro Yamaguchi

Senior Research Engineer, Media Innovation Laboratory, NTT Network Innovation Laboratories.

He received the B.E., M.E., and Ph.D. degrees in electronic engineering from the University of Electro-Communications, Tokyo, in 1991, 1993, and 1998, respectively. He joined NTT in 1998 and has been researching super-high-definition image distribution systems. He is a member of IEICE and the Institute of Image Information and Television Engineers (ITE).

Masanori Ogawara

Senior Research Engineer, Supervisor, Media Innovation Laboratory, NTT Network Innovation Laboratories.

He received the B.E. and M.E. degrees in electrical engineering from Keio University, Kanagawa, in 1992 and 1994, respectively. He joined NTT in 1994. His current research interests include reliable IP transmission technologies and network supported content creation collaboration systems. He is the director of a collaborative working platform development project. He received a paper award from IEICE in 1999. He is a member of IEICE.

Tatsuya Fujii

Senior Research Engineer, Supervisor, Group Leader of Media Processing Systems Research Group, Media Innovation Laboratory, NTT Network Innovation Laboratories.

He received the B.S., M.S., and Ph.D. degrees in electrical engineering from the University of Tokyo in 1986, 1988, and 1991, respectively. He joined NTT in 1991. He has been researching parallel image processing and super-high-definition image communication networks. In 1996, he was a visiting researcher at Washington University in St. Louis, MO, USA. He is a member of IEICE, ITE, and IEEE.