

2020 Public Viewing—*Kirari!* Immersive Telepresence Technology

*Akihito Akutsu, Akira Ono, Hideaki Takada,
Yoshihide Tonomura, and Maiko Imoto*

Abstract

NTT Service Evolution Laboratories is conducting research and development on an immersive telepresence technology called *Kirari!* in order to achieve ultra-realistic services that provide an experience akin to actually being at a sporting or live venue to off-site viewers in remote locations. This article gives an overview of technologies supporting *Kirari!*, including advanced media streaming and synchronization (Advanced MMT (MPEG Media Transport)), real-time image segmentation for any background, ultra-realistic presence design (virtual loudspeakers), and super high-definition video stitching. Initiatives using these technologies are also introduced.

Keywords: ultra-realistic, immersive telepresence, media processing/synchronization technology

1. Introduction

Diverse ways of viewing sporting events beyond conventional television coverage have been introduced recently for international competitions and popular domestic sports such as baseball and soccer. These include live distribution over the Internet and public viewings at off-site venues. Further, with the spread of 4K and 8K broadcasts, people around the world are expected to share in the excitement of sporting events in the year 2020 through television and public viewings.

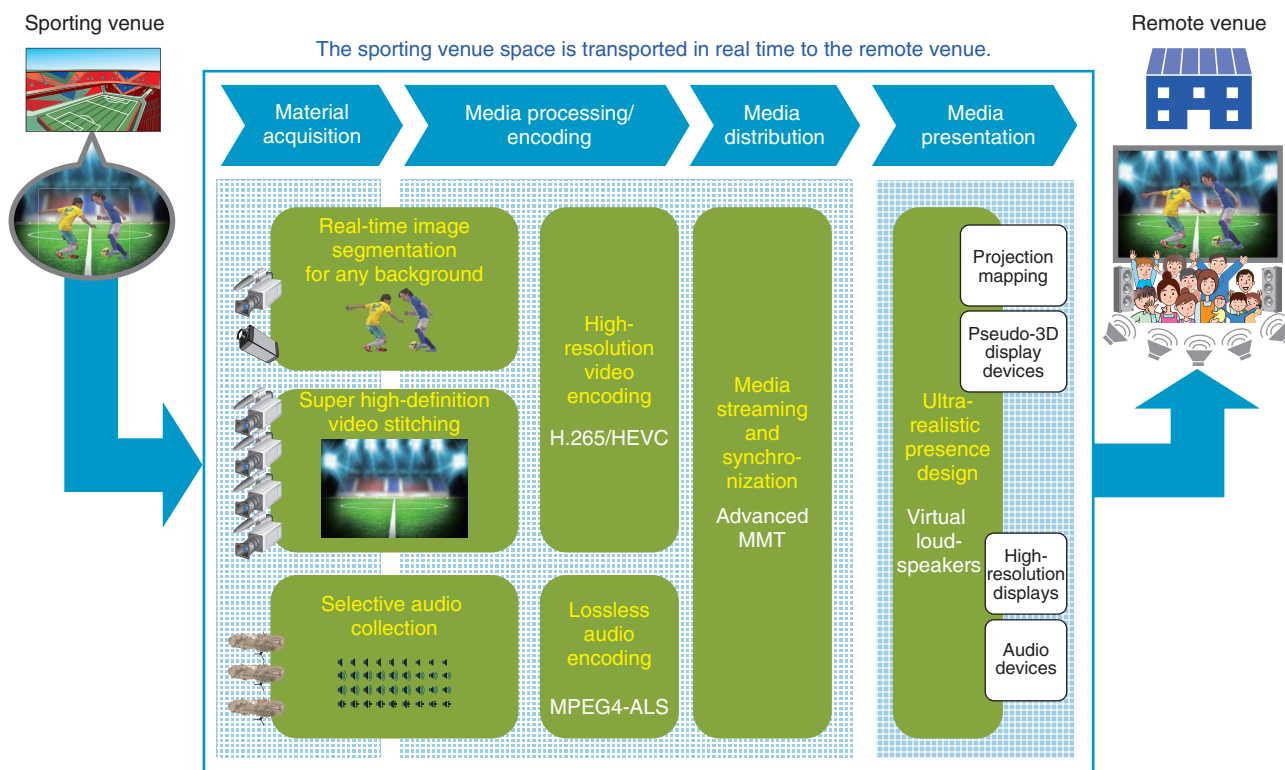
To tackle this world-wide diversification in viewing styles, NTT Service Evolution Laboratories announced our concept of *Kirari!* immersive telepresence technology in February 2015 [1]. *Kirari!* will implement high resolution and ultra-realistic presence technologies to deliver the sporting venue space in real time throughout Japan and the world. The goal of *Kirari!* is to enable viewers to experience sporting events as though they were actually at the venue from anywhere in the world.

2. Technologies supporting ultra-realistic services

Two main technical issues arose in implementing ultra-realistic real-time transportation of the video space with *Kirari!* (**Fig. 1**):

- Improving the accuracy when separating the image of a participant and the sound of the competition from the overall situation at the venue in real time
- Synchronizing the transportation of the extracted participant image and competition audio as well as stitching the overall competition background image and audio data and achieving highly realistic reproduction at the remote venue

We analyzed the technical elements needed to resolve these technical difficulties and carried out research and development (R&D) on each of them. As a result of this R&D, we were able to present real-time pseudo-three-dimensional (3D) video transmission of athletes giving a karate demonstration at the NTT R&D Forum held in February 2016, successfully achieving an ultra-realistic viewing experience (**Fig. 2**). An overview of the technical elements developed and comments on the results are given below.



ALS: Audio Lossless Coding
HEVC: High Efficiency Video Coding

Fig. 1. Kirari! technical elements.

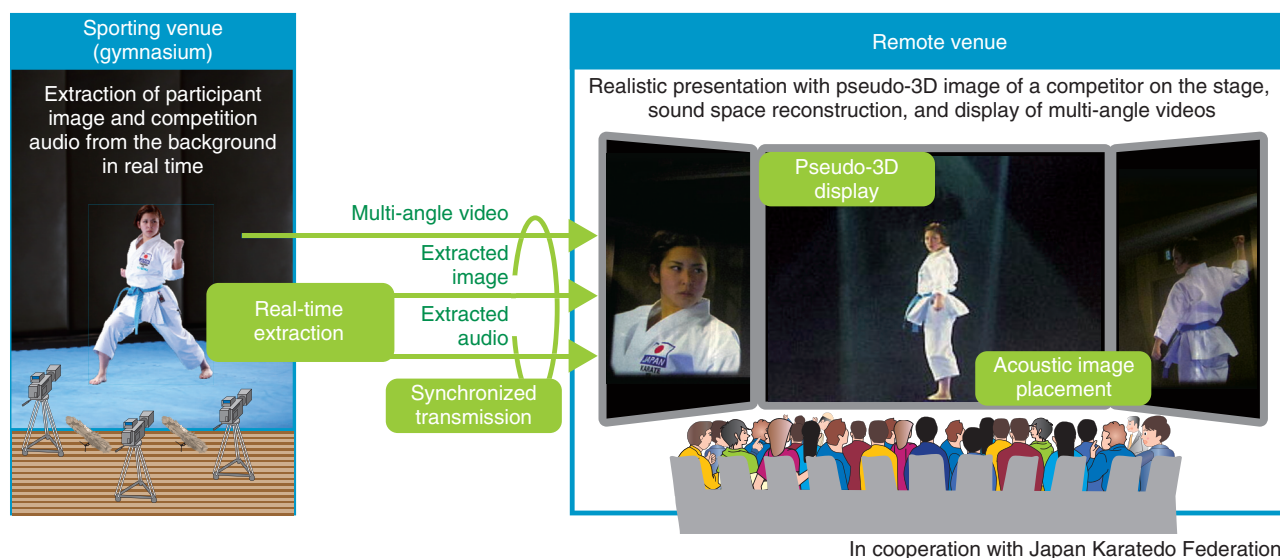


Fig. 2. Live demonstration with Kirari!

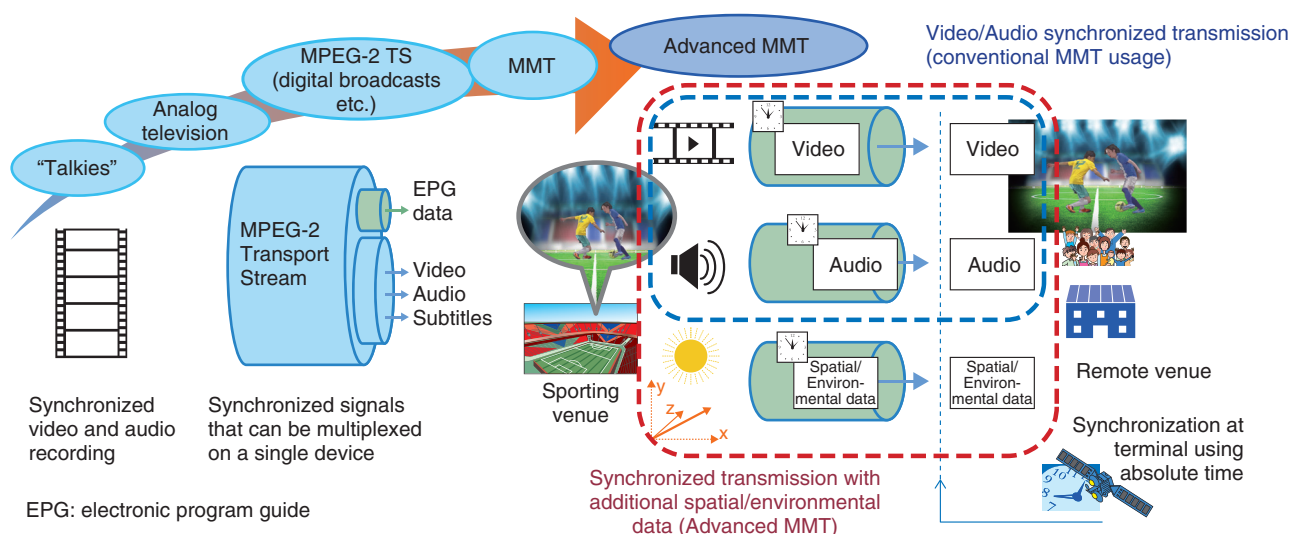


Fig. 3. Media synchronization for realistic sensation (Advanced MMT).

3. Advanced media streaming and synchronization (Advanced MMT)

We developed technology to synchronize and transport video and audio, combined with spatial information, by extending MPEG (Moving Picture Experts Group) Media Transport (MMT), which is an optimized protocol for media synchronization. We added definitions for MMT signaling to describe 3D information such as the size, position, and direction of MMT assets (Fig. 3). This technology makes it possible to correlate physical spatial parameters such as the size and position of the display device with asset data (frame pixel data) so that the space can be reconstructed with high realism at the destination at the intended size. Also, transmission of the DMX (Digital Multiplex) signals commonly used in production to control stage lighting and audio devices together with the MMT assets enables realistic presentations that accurately synchronize remote stage equipment with the media.

4. Real-time image segmentation for any background

To implement Kirari!, it was necessary to extract the desired subject from the captured video and display it in a way that resembled the real thing. We developed a technology for extracting the desired subject from any background; this technology can be used in situations such as sporting events, plays, and

lectures where it is difficult to use a green screen or other special background (Fig. 4). The first step involves extracting an approximate range of the subject using sensor data (trimap generation). The second step is to extract the subject in real time. The technology we developed to achieve this uses a framework that identifies the boundary between the subject and the background and accurately separates them using nearest neighbor search in the color space. It also uses clustering, with constraints to ensure that foreground and background colors do not overlap.

5. Ultra-realistic presence design (virtual loudspeakers)

We developed a highly realistic acoustic image placement technology that uses diffuse ultrasonic reflections to position acoustic images within the video image, where speakers cannot be placed. It can position virtual sound sources realistically using fewer speakers and producing a larger listening area than with earlier surround-sound speaker systems (Fig. 5). Degradation of ultrasonic characteristics is handled by using electrodynamic speakers to complement low frequencies. This technology enables virtual sound sources to be generated at any location on large, full-size images displayed on a large screen, so that realistic effects can be generated using a simple configuration such as having voices or competition sounds originate from their source on the screen.

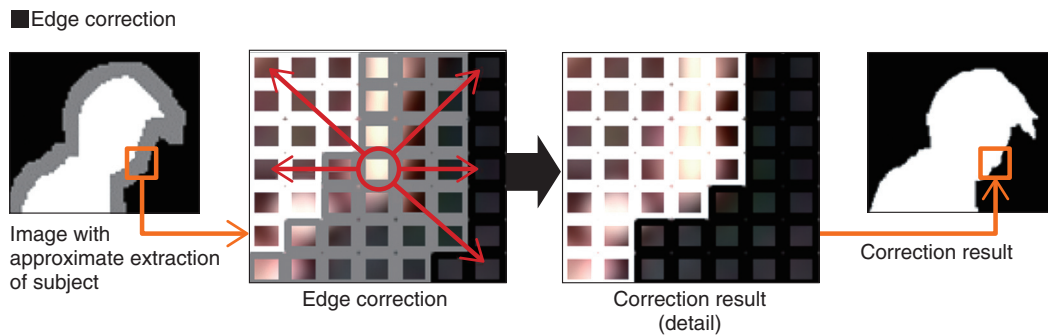


Fig. 4. Real-time image segmentation for any background.

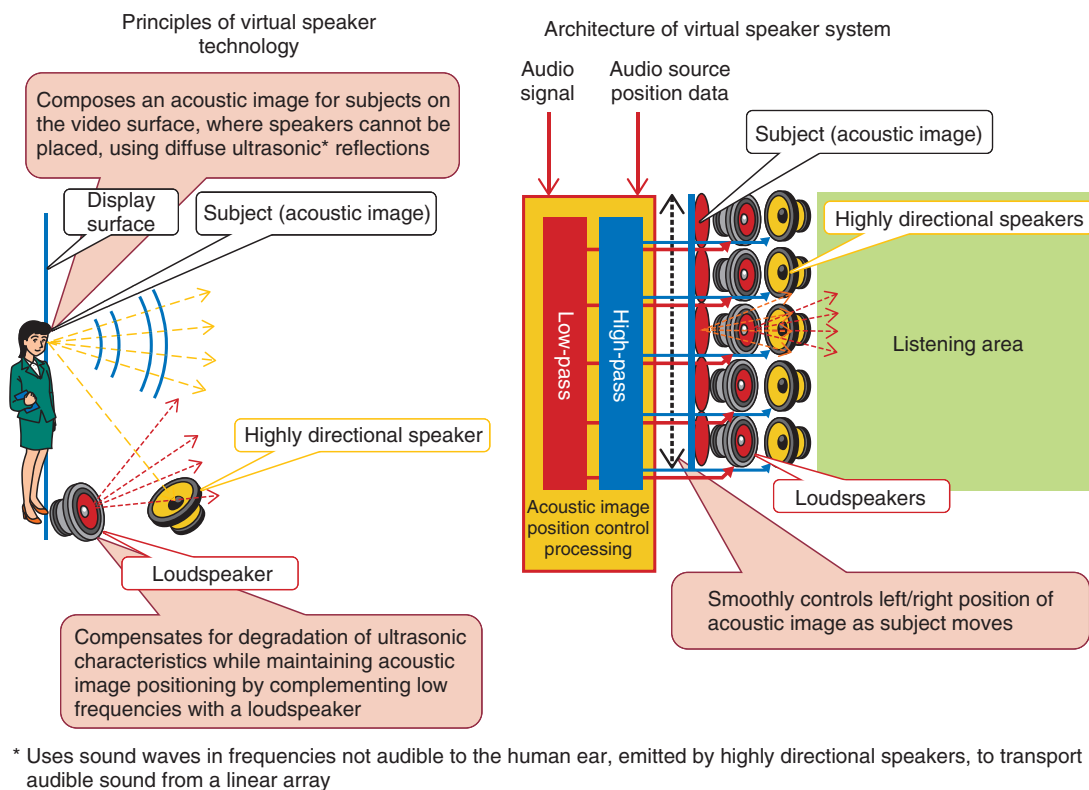


Fig. 5. Ultra-realistic presence design (virtual loudspeaker).

6. Super high-definition video stitching

As a first step toward achieving a highly realistic service with ultra-wide video not possible with ordinary 16:9 television video, we established algorithms and a system architecture to capture video using five 4K cameras arranged horizontally, and to compose them in real time (Fig. 6). To process the large

amount of 4K image data at high speed in real time, the data are partitioned and a mechanism is used that enables successive frames to be processed without waiting for positioning from the previous frame, which is necessary to suppress flicker. These innovations made real-time (4K60p) processing possible.

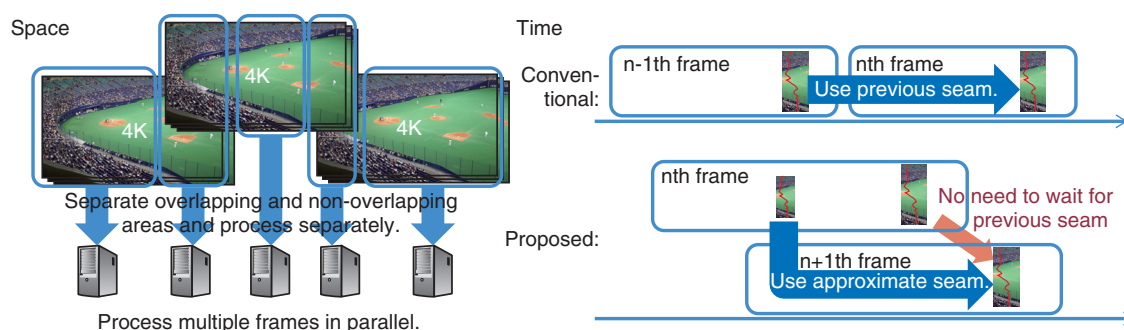


Fig. 6. Ultra-wide image composition.

7. Future development

We plan to conduct real-time transportation trials on these technologies, focusing mainly on individual sports. Beyond sporting events, we also intend to expand our horizons to other types of event that have been difficult for many people to appreciate remotely. This will include public viewings of traditional performances, regional festivals, intangible cultural assets, music concerts, and live coverage of lectures.

In further R&D, we plan to expand the scope of Kirari! to support competition with multiple athletes (competitions with large overlapping subjects) so that in the future, large numbers of competitors, or multiple subjects moving over a wide area, can be pre-

sented. This will be applied to sports such as judo and soccer that are currently difficult to present in this way.

We are conducting R&D on elemental technologies needed to provide ultra-realistic services with the goal of making viewing experiences close to actually being at a sporting venue available anywhere in the world by the year 2020.

Reference

- [1] A. Akutsu, K. Hidaka, M. Inoue, N. Ito, T. Yamaguchi, S. Fujimura, and A. Nakadaira, "Delivering Technologies for Services that Deliver the Excitement of Games Worldwide," NTT Technical Review, Vol. 15, No. 7, 2015.
<https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201507fa2.html>



Akihito Akutsu

Executive Research Engineer, Natural Communication Project, NTT Service Evolution Laboratories.

He received an M.E. in engineering from Chiba University in 1990 and a Ph.D. in natural science and technology from Kanazawa University in 2001. Since joining NTT in 1990, he has been engaged in R&D of video indexing technology based on image/video processing, and man-machine interface architecture design. From 2003 to 2006, he was with NTT EAST, where he was involved in managing a joint venture between NTT EAST and Japanese broadcasters. In 2008, he was appointed Director of NTT Cyber Solutions Laboratories (now NTT Service Evolution Laboratories). He worked on an R&D project focused on broadband and broadcast services. In October 2013, he was appointed Executive Producer of 4K/8K HEVC at NTT Media Intelligence Laboratories. He received the Young Engineer Award and Best Paper Award from the Institute of Electronics, Information and Communication Engineers (IEICE) in 1993 and 2000, respectively. He is a member of IEICE.



Yoshihide Tonomura

Senior Research Engineer, NTT Service Evolution Laboratories.

He received an M.S. in electronics engineering from Nagaoka University of Technology, Niigata, in 2004 and a Ph.D. from Tokyo Metropolitan University in 2010. Since joining NTT, he has been engaged in R&D of multimedia data transmission technologies. From 2011 to 2012, he was a visiting scientist at MIT Media Lab, Massachusetts, USA. He received the SUEMATSU-Yasuharu Award of IEICE in 2016 and the 26th Telecommunications Advancement Foundation TELECOM System Technology Award in 2011.



Akira Ono

Senior Research Engineer, Supervisor, Natural Communication Project, NTT Service Evolution Laboratories.

He received an M.E. in computer engineering from Waseda University, Tokyo, in 1992. He joined NTT in 1992 and began researching and developing video communication systems. From 1999 to 2010, he was with NTT Communications, where he was engaged in network engineering and creating consumer services. He moved to NTT Cyber Solutions Laboratories (now NTT Service Evolution Laboratories) in 2010. He is working on the Kirari! immersive telepresence technology.



Maiko Imoto

Researcher, Natural Communication Project, NTT Service Evolution Laboratories.

She received an M.S. in computer science from Ochanomizu University, Tokyo, in 2011. She joined NTT the same year and began studying geographical information search and HTML5-based web applications. She has been working on the immersive telepresence technology called Kirari! since 2014.



Hideaki Takada

Senior Research Engineer, Supervisor, Natural Communication Project, NTT Service Evolution Laboratories.

He received a B.E. in information and communication technology from Tokai University, Shizuoka, in 1995, an M.E. in information systems from the University of Electro-Communications, Tokyo, in 1997, and a D.S. in global information and telecommunication studies from Waseda University, Tokyo, in 2007. He joined NTT Integrated Information & Energy Systems Laboratories in 1997. He has been researching 3D visual perception and 3D display technology. He received the Best Paper Award from the 3-D image conference in 2001 and 2005, the Young Researcher's Award from the Virtual Reality Society of Japan in 2002, the Achievement Award from IEICE in 2003, a Commendation by the Minister of Education, Culture, Sports, Science and Technology of Japan in 2006, and Technical Committee Prize Paper Awards from the Institute of Electrical and Electronics Engineers (IEEE) in 2014 and 2016. He is a member of IEEE and the Institute of Image Information and Television Engineers.