Kirari! Ultra-realistic Communication Technology: Beyond 2020

Akihito Akutsu, Kenichi Minami, and Kota Hidaka

Abstract

The NTT Group is creating new value to overcome the limitations of space through research and development of Kirari! ultra-realistic communication technology. Kirari! goes beyond simply improving sound and image quality; it is aimed at achieving a sense of realism, as though the objects or people being viewed were actually there in front of the viewer. This article gives an overview of Kirari! technology, introduces some application examples, including live public viewings and other unprecedented productions, and reviews trends for beyond 2020.

Keywords: ultra-realism, media, telepresence

1. Introduction

The 5th Science and Technology Basic Plan [1] was established in a cabinet decision on January 22, 2016 and defines the new concept of Society 5.0 [2]. The fifth society, or age, in human history, Society 5.0 follows the hunting and gathering society (1.0), the agrarian society (2.0), the industrial society (3.0), and the information society (4.0) (**Fig. 1**). It is defined as follows in the integrated science and technology innovation strategy set in the cabinet decision of May 24, 2016 [3].

"A human-centered society able to deliver high-quality lifestyles, rich in energy and comfort, through advanced integration of cyberspace and physical space, achieving both economic development and solutions to social issues by providing goods and services to meet the detailed, various and latent needs of all, without disparity by region, age, gender, language, or other factors."

With advances in information and communication technology (ICT), voice and video communication with people in distant locations has been achieved and advanced, increasing in sound and image quality. It has also become possible to sense the physical space we are in, with advances in various types of
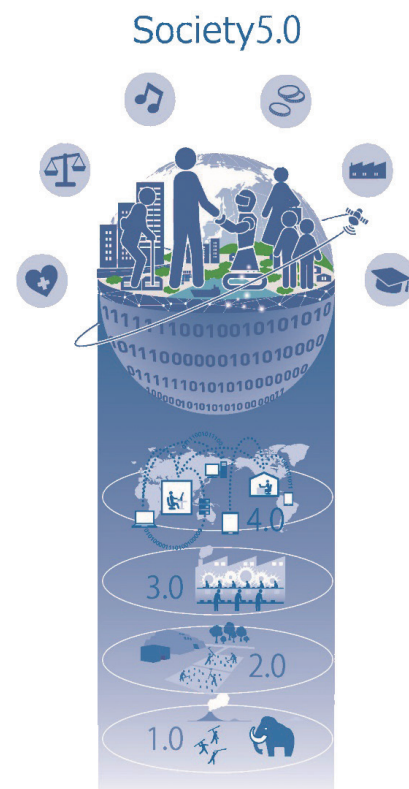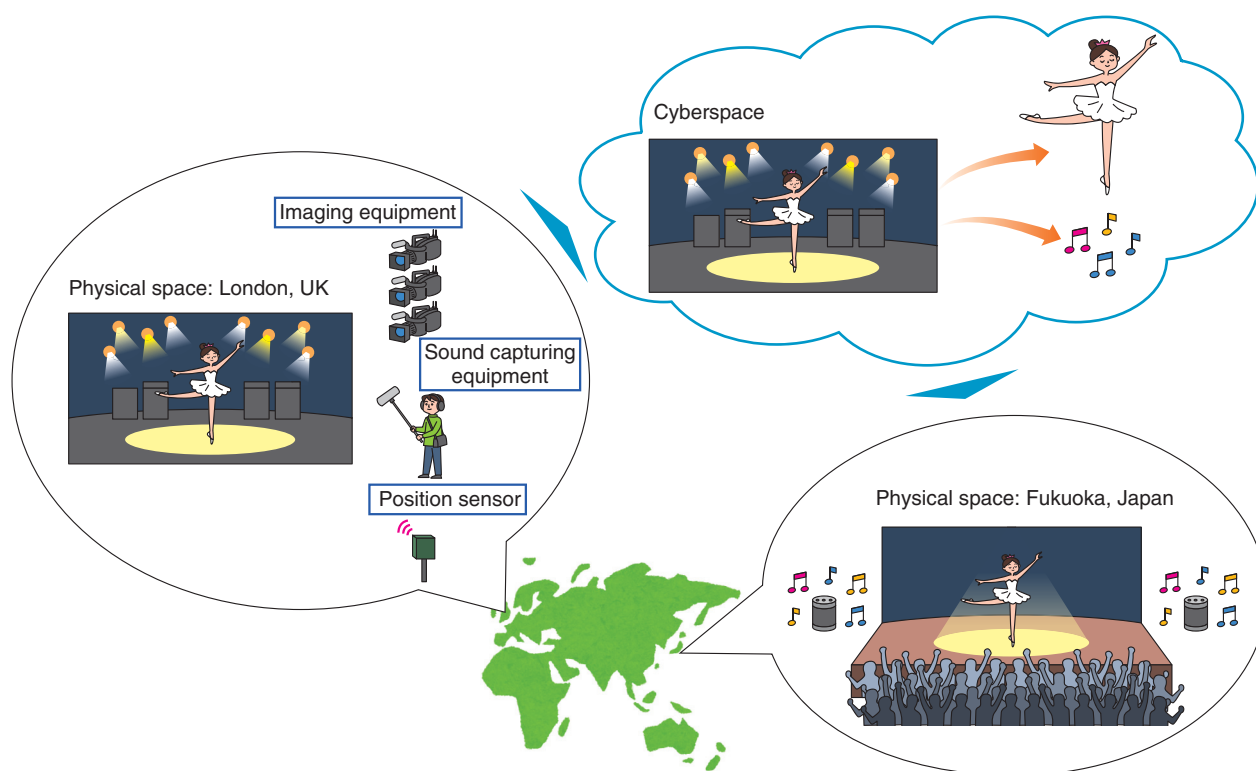

Fig. 1. Image of Society 5.0.

Fig. 2.   Kirari! overcomes boundaries of physical space and cyberspace.

sensing technologies. Media processing of such sensor data in cyberspace and the subsequent transmission and reproduction in another physical space—creating a sense that the person is right before your eyes—may well become common in Society 5.0 (**Fig. 2**). Such reproduction can be called *high realism*. The Kirari! system being developed by the NTT Group is the future of media transmission, going beyond improvements in sound and image quality, to implement ultra-realistic communication.

With Kirari!, distance can be overcome, and information and images of people and spaces can be transmitted from distant locations in real time. With Kirari!, people can experience a sporting event without having to travel to the venue, or a speaker can participate in an event from a location far from the venue. We believe Kirari! will contribute to overcoming the limitations of space in the coming Society 5.0.

Examples using Kirari! are shown in **Fig. 3**. In the examples of local live viewing of overseas sports competitions, which would not be easy to see otherwise, illusion effects have been used to create near three-dimensional (3D) viewing experiences or to provide a wider field of view than ever before possi-

ble. Astonishing performances not previously possible can also be produced. For example, by sensing the actual performer and applying media processing, another performer can be recreated and overlaid using illusion effects. By controlling the timing of the overlay, a *joint* performance featuring a performer in real time and an image of that same performer from a few minutes earlier can be implemented, which would not otherwise be possible. Through this sort of initiative, we hope to provide ultra-realistic experiences that overcome the limitations of space for various types of content.

2.   Overview of Kirari! technology

An overview of Kirari! technology is shown in **Fig. 4**. It is divided into the three aspects of the sports or performance venue, communication, and the viewing venue. Kirari! ultra-realistic communication technology is in the communication area. Information extracted using cameras, microphones, and sensors is handled by media control, media processing, and synchronization and transmitted to the viewing venue.

Example: local live viewing of overseas sports competitions
Pseudo-3D viewing experience by using illusion effects
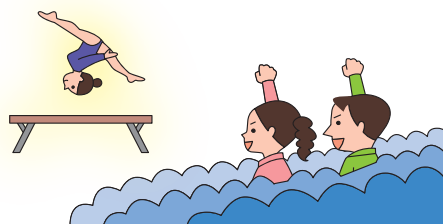
A wider image viewing field beyond a
16:9 frame

16:9 frame
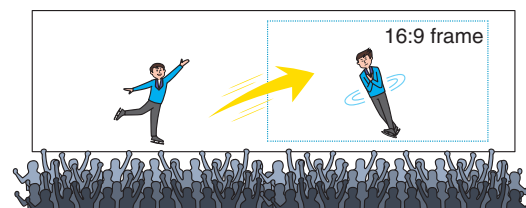
Example: experiencing a performance using unprecedented stage effects
Joint performance with one's old self

13:02:20:10
(Current)

13:01:00:00
(Past)

Fig. 3.   Application examples of Kirari!.

Media control consists of spatial information, which includes position data obtained from sensors, people in camera imagery, and associated lighting information, and time control, which is used for controlling the distribution of people in absolute time.

Media processing includes object extraction, in which people in the captured video can be extracted from the background, and audio wave-field synthesis technology [4] for highly realistic sound. For sports coverage and other performances as shown in Fig. 3, these features use illusion effects to display people in 3D and perform the processing needed to reproduce audio that seems to come from the image to the viewer's position. In the other wide-angle viewing experience in Fig. 3, images from multiple cameras are synthesized to produce an ultrawide image [5]. Then encoding is done to transmit the content efficiently.

We have created our own extension to the MPEG* Media Transport (MMT) standard, called Advanced MMT, which we use for synchronized transmission [6]. It is capable of synchronization with absolute time, so synchronized transmission to any location in the world with the same timing is possible. Advanced MMT has the role of providing a design plan for achieving ultra-realism based on video, audio, lighting, and other information. At the viewing venue, projection, speakers, lighting, and other elements are set up according to the Advanced MMT design plan.

Some of the technologies used in Kirari! are explained in the Feature Articles in this issue [7–10].

3.   Future prospects

To allow as many users as possible to experience the ultra-realism achieved by Kirari!, we have conducted proof-of-concept and live viewing trials, including sports events that seem to be happening before the viewer's eyes, unprecedented performing arts using ICT, dance performances featuring both current and past performers, and coverage of lectures

---

*   MPEG: Moving Picture Experts Group, a working group of ISO (International Organization for Standardization) and IEC (International Electrotechnical Commission) in charge of developing international standards for compression of audio and video data.
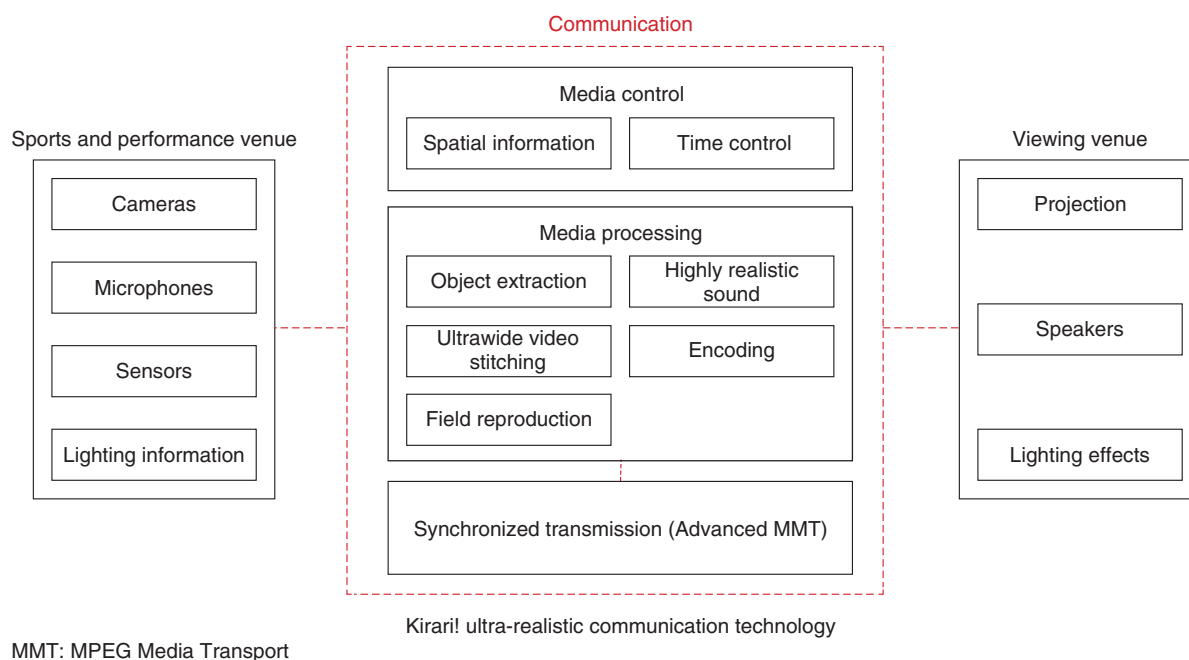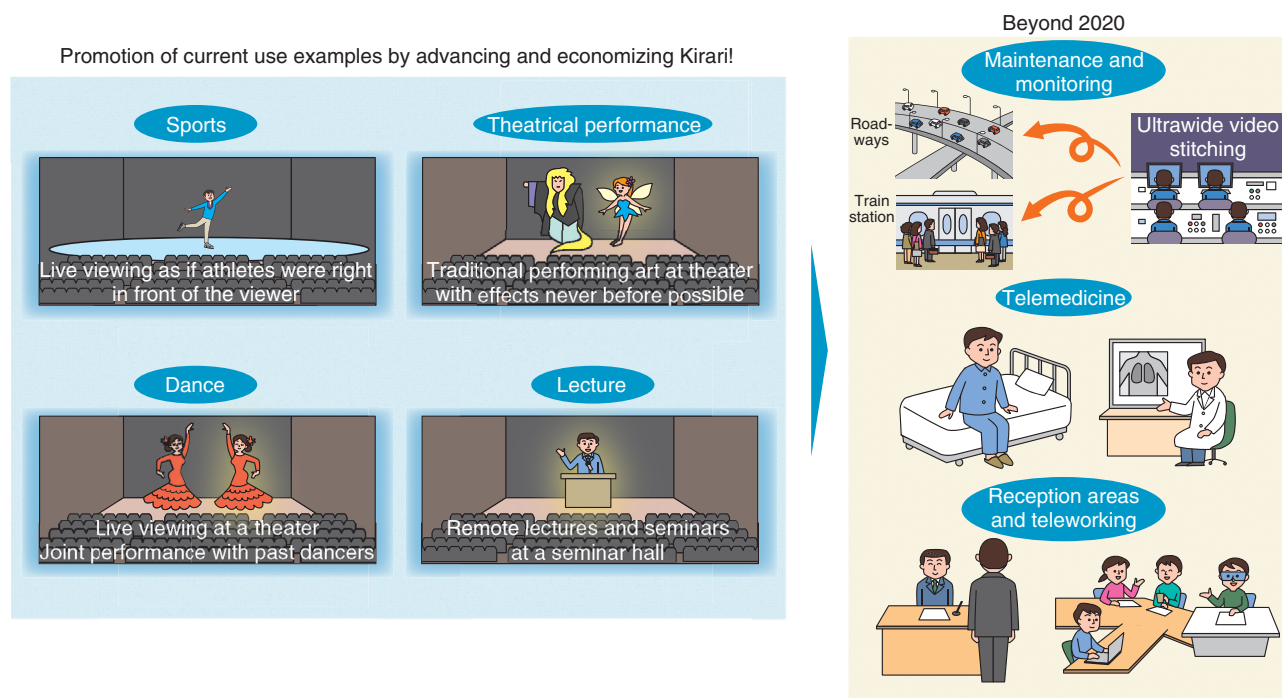
Fig. 4.   Overview of Kirari! technology.

Fig. 5.   Further development of Kirari!.

from remote locations (**Fig. 5**). One advancement to Kirari! that is needed to promote these measures is more accurate extraction of the object being viewed, even in environments with people coming and going.

Another issue is to create ultrawide compositions of high-definition video.

Measures to reduce costs are also an important issue for implementation in society. For example, we are studying ways to further compress high-definition video, and to reduce the number of media processing servers required. Also, with more advanced media control, it will become possible to automatically and accurately overlay video of performers from the past next to real performers. This could contribute to reducing the personnel costs for productions.

In addition to promoting Kirari! in society with sports and other performances, we are looking at possibilities beyond 2020. For example, in fields such as maintenance and surveillance, roadways and train stations could be monitored using ultrawide composed video, giving an experience in which items being monitored seem to be right before the viewer's eyes. Increasingly, applications are being studied in medicine for remote examinations, in business situations to replace receptionists, and to promote telecommuting. With these initiatives, we will continue to create value by overcoming the limitations of space in Society 5.0.

References

[1] Government of Japan, "The 5th Science and Technology Basic Plan," Jan. 2016.
http://www8.cao.go.jp/cstp/english/basic/5thbasicplan.pdf

[2] K. Hidaka, Y. Hasegawa, and H. Fuseda, "A Platform for Realizing a New Economy and Society: Society 5.0," Operations Research, Vol. 61, No. 9, pp. 551–555, 2016 (in Japanese).

[3] Government of Japan, "Comprehensive Strategy on Science, Technology and Innovation 2016 (Excerpt)," May 2016.
http://www8.cao.go.jp/cstp/english/doc/2016ststrategy_main.pdf

[4] K. Tsutsumi and H. Takada, "Powerful Sound Effects at Audience Seats by Wave Field Synthesis," NTT Technical Review, Vol. 15, No. 12, 2017.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201712fa5.html

[5] T. Sato, K. Namba, M. Ono, Y. Kikuchi, T. Yamaguchi, and A. Ono, "Surround Video Stitching and Synchronous Transmission Technology for Immersive Live Broadcasting of Entire Sports Venues," NTT Technical Review, Vol. 15, No. 12, 2017.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201712fa4.html

[6] Y. Tonomura, H. Imanaka, K. Tanaka, T. Morizumi, and K. Suzuki, "Standardization Activity for Immersive Live Experience (ILE)," ITU Journal, Vol. 47, No. 5, pp. 14–17, 2017 (in Japanese).

[7] H. Kakinuma, J. Nagao, H. Miyashita, Y. Tonomura, H. Nagata, and K. Hidaka, "Real-time Extraction of Objects from Any Background Using Machine Learning," NTT Technical Review, Vol. 16, No. 12, pp. 12–18, 2018.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201812fa2.html

[8] T. Isaka, M. Makiguchi, and H. Takada, "'Kirari! for Arena'—Highly Realistic Public Viewing from Multiple Directions," NTT Technical Review, Vol. 16, No. 12, pp. 19–23, 2018.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201812fa3.html

[9] M. Makiguchi and H. Takada, "360-degree Tabletop Glassless 3D Screen System," NTT Technical Review, Vol. 16, No. 12, pp. 24–28, 2018.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201812fa4.html

[10] M. Isogai, K. Okami, M. Matsumura, M. Date, A. Kameda, H. Noto, and H. Kimata, "Video Processing/Display Technology for Reconstructing the Playing Field in Sports Viewing Service Using VR/AR," NTT Technical Review, Vol. 16, No. 12, pp. 29–35, 2018.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201812fa5.html

Akihito Akutsu

Vice President, 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 research and development (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), where 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 (High Efficiency Video Coding) 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.

Kota Hidaka

Senior Research Engineer, Supervisor, Group Leader, NTT Service Evolution Laboratories.

He received an M.E. in applied physics from Kyushu University, Fukuoka, in 1998, and a Ph.D. in media and governance from Keio University, Tokyo, in 2009. He joined NTT in 1998. His research interests include speech signal processing, image processing, and immersive telepresence. He was a Senior Researcher at the Council for Science, Technology and Innovation, Cabinet Office, Government of Japan, from 2015 to 2017.

Kenichi Minami

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

He received a B.E. in electronic engineering and an M.S. in biomedical engineering from Keio University, Kanagawa in 1991 and 1993. He received an MBA from Thunderbird, Global School of Management, Arizona, USA, in 2002. He has been engaged in R&D management in the development of "Kirari!" immersive telepresence technology since 2016. During 2012–2014, he was responsible for the development of mobile application services at NTT DOCOMO. His research interests include image and audio processing, user interfaces, and telepresence technologies. He is a member of IEICE.