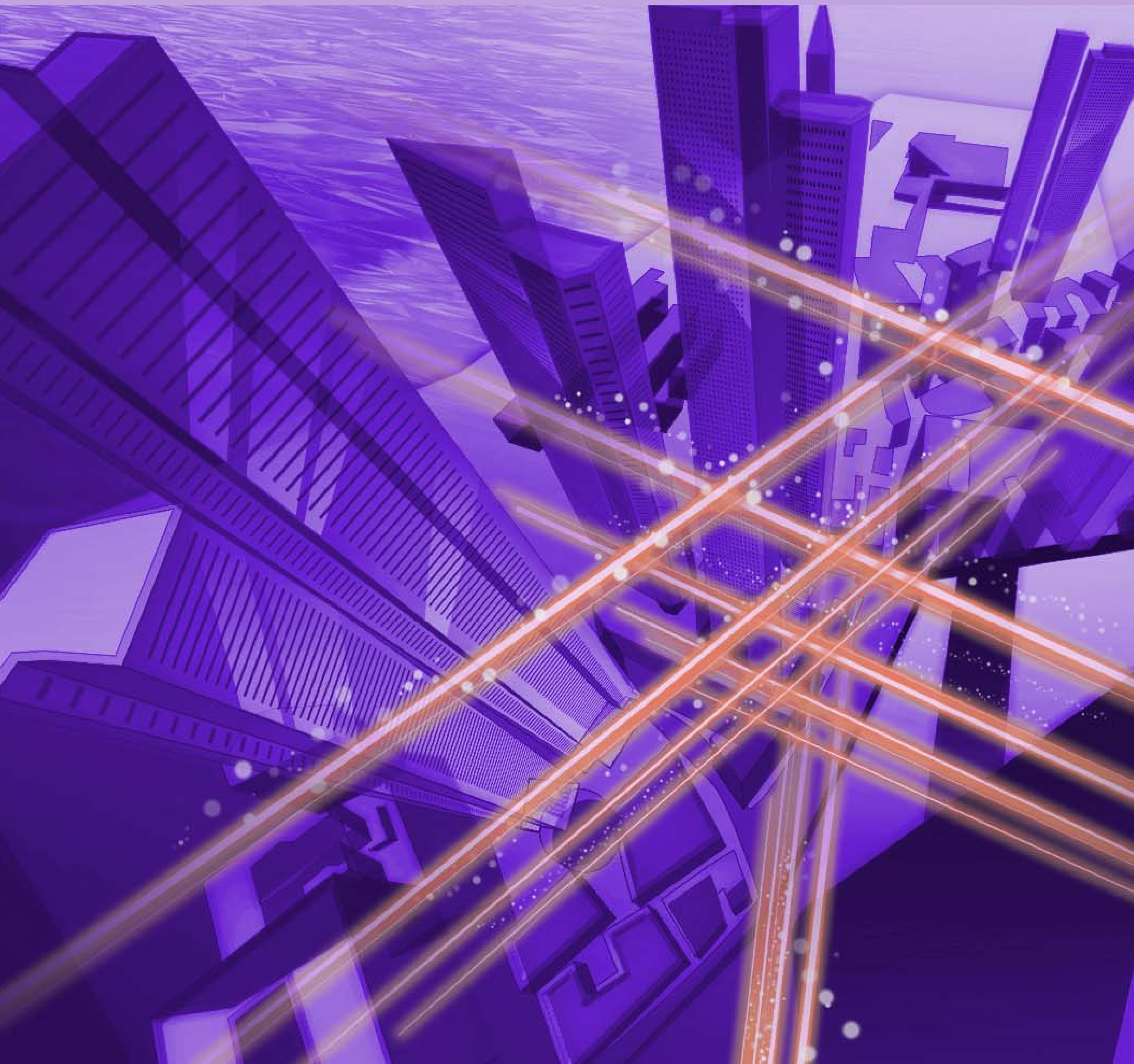
NTT Technical Review

12
2021



December 2021 Vol. 19 No. 12

NTT Technical Review

December 2021 Vol. 19 No. 12

Treasure Every Encounter to Expand Your World

Toru Takama
President & CEO, NTT Anode Energy Corporation

Overview
NTT Anode Energy is promoting the smart energy solution business, which leverages the resources and information and communication technologies possessed by the NTT Group. Two years after its establishment, the company is striving to contribute to resolving social issues, such as environmental and population problems, revitalizing industry, and building a sustainable society. We interviewed Toru Takama, president and chief executive officer of NTT Anode Energy, about the progress of business to promote and expand renewable energy and the mindset of top management.

Keywords: renewable energy, carbon neutrality, power purchase agreement

Expand sales by developing renewable energy sources and promote smart grids

—NTT's new environment and energy vision "NTT Green Innovation toward 2040" is exactly where NTT Anode Energy will play an active role.

In September 2021, NTT announced a new environment and energy vision called "NTT Green Innovation toward 2040" to simultaneously achieve the contradictory objectives of "zero environmental load" and "economic growth" by reducing environmental load due to business activities and creating breakthrough innovation. This vision states two goals: the first is an 80% reduction in greenhouse-gas emissions by the NTT Group (compared to FY2013) and carbon neutrality for the mobile (NTT DOCOMO) and datacenter sectors by FY2030, and the second is carbon neutrality for the NTT Group by FY2040. NTT Anode Energy is playing a role in achieving these goals through business development on the basis of two pillars: (i) developing renewable energy sources and expanding sales and (ii) promoting local generation for local consumption of energy, centering on "Battery Farms."

Since the establishment of the company, we have been focusing on developing *green power* as a renewable energy source. In addition to constructing and installing green power plants, we provide green power to companies that recognize the value of renewable energy. We will supply green power to DOCOMO Denki, a retail electricity service scheduled to be launched in March 2022.

The supply of green power requires storage batteries to store surplus power from green power plants and to supply power when the amount of power being generated drops. Accordingly, we are establishing Battery Farms, an environment in which green power plants can be easily connected to power-transmission and distribution networks by installing storage batteries in approximately 7300 of NTT Group's telecommunication buildings scattered across Japan.

We began providing green power to stores such as Seven & i Holdings Co., Ltd.'s convenience stores through Japan's first off-site power purchase agreements (PPAs). Under off-site PPAs, power is supplied

to the customer's facilities from power-generation facilities installed in remote locations through power-transmission and distribution networks. We also provide green power to Daiichi Sankyo Chemical Pharma Co., Ltd. and Furukawa Electric Co., Ltd. through on-site PPAs under which power is supplied from solar-power generation equipment, etc. installed next to the customer's facilities.

—What specific action will be taken within the NTT Group to achieve this vision?

The NTT Group currently uses just over 1% of the electricity consumed in Japan. I believe that our efforts to transform this 1% into green power are of great social significance. If NTT Group's electricity consumption follows its current trend, it is projected to double by FY2040 compared to FY2013. Therefore, "NTT Green Innovation toward 2040" presents a vision of achieving decarbonization and carbon neutrality of the NTT Group by FY2040 by cutting power consumption in half and converting it into renewable energy through implementing the Innovative Optical and Wireless Network (IOWN) and energy saving.

Conventional energy systems have typically been centralized systems; that is, power generated at large power plants is delivered to points of consumption via power-transmission and distribution networks. In contrast, energy systems from now on will be distributed systems. NTT Anode Energy plans to construct a distributed energy system by using NTT Group's telecommunication buildings across Japan as Battery Farms to provide locally generated renewable energy for local consumption. This system will reduce electricity loss due to long-distance power transmission and distribution and enable more efficient use of energy by, for example, controlling the power supply in accordance with supply-and-demand conditions. The construction and rapid development of Battery Farms and power plants is our highest priority.

How to seize opportunities during challenging situations

—Have you had anything to do with energy in your past work?

My position at NTT Anode Energy is my first energy-related job. I joined Nippon Telegraph and Telephone Public Corporation (now NTT) in 1981, and for 26 years starting my second year there, I have led a challenging corporate life during which I have taken on work without any briefing or handover; in other words, I have always been assigned to projects new to NTT. After completing my training period, I was assigned to a technology development department related to transmission and took charge of digital-transmission technology. At that time, telecommunication networks were being digitized, and a very accurate digital clock was needed to ensure that entire networks operated in synchronization. To meet this need, we developed the "D2 standard-clock generator," which became Japan's standard clock. (The actual device is on display at the NTT History Center of Technologies in Tokyo.) Since then, I have worked in fields such as product development, global affairs, technology development, and software development, but NTT Anode Energy is my first job in the energy field. However, most of the work I have experienced thus far has been characterized by taking new steps, such as starting new projects or joining new organizations, and the energy field is the same in the sense of being new to NTT.

Among my experiences, one that left a particularly strong impression on me was working at the World Bank in Washington, D.C. beginning in 1987 for three and a half years, where I served as a professional in charge of loan appraisal and supervision of telecommunication projects in Indonesia and Bangladesh. The World Bank has nothing to do with technological development. I arrived in Washington, D.C. on a Sunday, and since I had just arrived, I had neither found a house to live in nor other living arrangements. The next day, Monday, on greeting my boss for the first time, I was given a document of about 150 pages and told, "Tomorrow, you will attend a meeting on behalf of our section and write a report on it."

I had hoped to be able to get my life together before I started work, but that hope was dashed. Although I had studied in France before, this was my first time in an English speaking country. On top of that, my boss at the time was Indian, and I had difficulty in understanding his heavy Indian accent. The theme of the meeting that I attended was, naturally, finance, and the conversation was filled with banking terminology in English. As an engineer, I didn't know such banking jargon, so I couldn't write the expected report. As a result, I spent nine months in a private room without being given any work.

—That's a tough situation. How did you overcome that?

First, I made an effort to brush up on my English by studying at a business school and earned an MBA (master of business administration) at night school. At the World Bank, professionals have their own offices without having meetings or document exchanges, so I had no way of knowing what was going on in my own section. Therefore, I invited colleagues of my section to lunch to find out what was going on and ask if there was any work I could do. As the World Bank was a competitive workplace, they wouldn't talk about good jobs with me.

Occasionally, some people told me they had a job for me, but it was either very challenging or unpleasant. However, it was better than no work for me, so I volunteered for such a challenging job. Since I didn't

have the capability to carry it out, I invited other colleagues to lunch and asked them how to do the job. As I repeated this process over and over, I was able to receive help from different people and finally and fortunately was assigned to telecommunication projects in Bangladesh and Indonesia after nine months.

The World Bank is an organization with full of senior professionals. One of the chief economists at that time was Lawrence Summers. Later he became the U.S. Secretary of the Treasury in 1999 and the president of Harvard University in 2001. I worked with former Managing Director of Sri Lankan Telecom, and my boss with the heavy accent was the former Managing Director of Indian Development Bank. Looking back, I was fortunate to have opportunities to meet these people.

It was also during that time that I met one of the founders of Apple, Mike Markkula. Later, in 1995, I happened to be a member of the joint project between NTT and Apple. During the project, we developed and commercialized a system called Desktop Communication Conference System FM-A71, which had almost identical functionalities to today's web conferencing systems, allowing users to share images during a conference call. Web conferencing is now used for remote work as a matter of course; however, back then, the Internet was not yet widespread, and our system required an ISDN connection (integrated services digital network). Even though it did not sell well, in retrospect, I think our efforts were pioneering.

Results will always emerge over time

—Are there any moments when you feel that the work you have done thus far has made a difference in your current position?

NTT's electric-power business has a shorter history than its telecommunication business. As I mentioned earlier, I have many experiences of starting new businesses from scratch, so I think I am used to challenging situations. Even if you face a big decision or a fork in the road in regard to the job at hand, it may not be such a big issue if you look at it from a wider perspective. I believe that we should work on one thing at a time and not rush things, because results will surely come as time goes by. However, some things must be addressed immediately. For example, investment in Battery Farms and power plants that enable local generation and local consumption of energy is needed right now. I think that my past experiences in investing in overseas companies has come in handy in that regard.

I have learned from past experiences two things that are important to me in top management. One is communication, and the other is "*ichi-go ichi-e*," which means treasure every encounter, for it will never recur. I have thus far managed to overcome difficulties thanks to the help of many people and cherish the many encounters I have had. Thanks to them, I am now able to do the job I do today.

In addition to these two points, in top management you must have the ability to execute plans, make decisions, and make judgments. It is important to listen carefully to the opinions of others, including employees, and once you have made a decision, try your best to carry it out or call it off if you find that the decision is not working. For that reason, I listen to people a lot. This is because many clues can be found by listening to people. It is also obvious that there are limits to what one person can think and do alone. Therefore, it is crucial to create an environment in which people support each other. This environment cannot be created without good communication.

—Communication is also important for engineers and researchers to flourish. What is your advice to them?

I have always valued communication—both internally and externally. I invite employees to lunch to create a place for communication, and I have been holding quarterly group discussions with employees for the past 20 years. Setting a theme and discussing it creates a variety of ideas, and in a number of cases, those ideas have become reality.

Outside the company, a study group that I launched 38 years ago is still going, and a total of 800 people are registered. It is held once a month, and each session has 20 to 30 participants. I try to make the most of every moment I get to meet each of the members who attend the study sessions. This accumulation of encounters leads to work and connections that expand my world, and I am grateful that it does.

When I was head of the Advanced IP Architecture Center at NTT Communications, a technology-development organization, I set up an open lab for co-creating with venture companies by offering NTT Communications' cloud services and mentoring for free for six months. We started this project to support as many entrepreneurs as possible, and although it did not directly benefit the NTT Group, some of the organizations that used it are now successful companies. Also, when I was working at Verio, a US Internet and web hosting company, I learned about the benefits of the open-source model from their business.

I have been creating settings that lead to such open innovation in Japan and abroad as well as promoting initiatives to energize engineers. I hope that engineers will appreciate the importance of communication with various people and keep challenging themselves. I also hope that researchers in the software field will communicate with their outstanding engineering colleagues and work on many interesting projects together.

Interviewee profile

■ Career highlights

Toru Takama joined Nippon Telegraph and Telephone Public Corporation (now NTT) in 1981. After being seconded to the World Bank and Verio Inc. he had served as President of the Advanced IP Architecture Center of NTT Communications, member of the Board of NTT COMWARE, and Executive Vice President of NTT TechnoCross. He has been in his current position since June 2020.

Venture Out of the Laboratory and Test Your Research Results in the Real World

Yasushi Takatori
Senior Distinguished Researcher,
NTT Access Network Service Systems Laboratories

Overview

Future networks are expected to provide high-speed, large-capacity communications and enormous amount of computing resources that exceed the limits of the conventional network infrastructure. To meet these expectations, NTT is conducting research and development on the Innovative Optical and Wireless Network (IOWN), a future information-processing infrastructure that includes user's terminals. Yasushi Takatori, a senior distinguished researcher at NTT Access Network Service Systems Laboratories, is taking on the challenge of innovating wireless access toward the 6G (six-generation mobile communication system) era, which will be an essential component of IOWN. We asked him about his research activities and his attitude as a researcher.

Keywords: wireless access, Cradio®, extreme NaaS

Provide a natural and uninterrupted communication environment that does not force users to be aware of it

—Please tell us about your current research projects.

My current research is on using multiple wireless access systems to provide wireless access in the sixth-generation mobile communication system (6G) era, which will be an essential component of the Innovative Optical and Wireless Network (IOWN). In the field of wireless communications, the amount of communications via smartphones is increasing, and accompanying the development of the Internet of Things, a multitude of objects have become connected to the Internet. The types of wireless terminals and usage scenarios are also diversifying, as illustrated by remote control of autonomous vehicles and drones and exchange of ultra-high-definition images. Under these circumstances, the role played by wireless communications should increase dramatically in every aspect of our lives and that the amount of such communication will continue to grow. Such diversification of terminals and usage scenarios will require wireless networks to cover not only urban areas, where the population is concentrated, but also suburban areas, rural areas, and many other locations. It is also necessary to satisfy various requirements concerning wireless communication quality, which varies according to various usage scenarios.

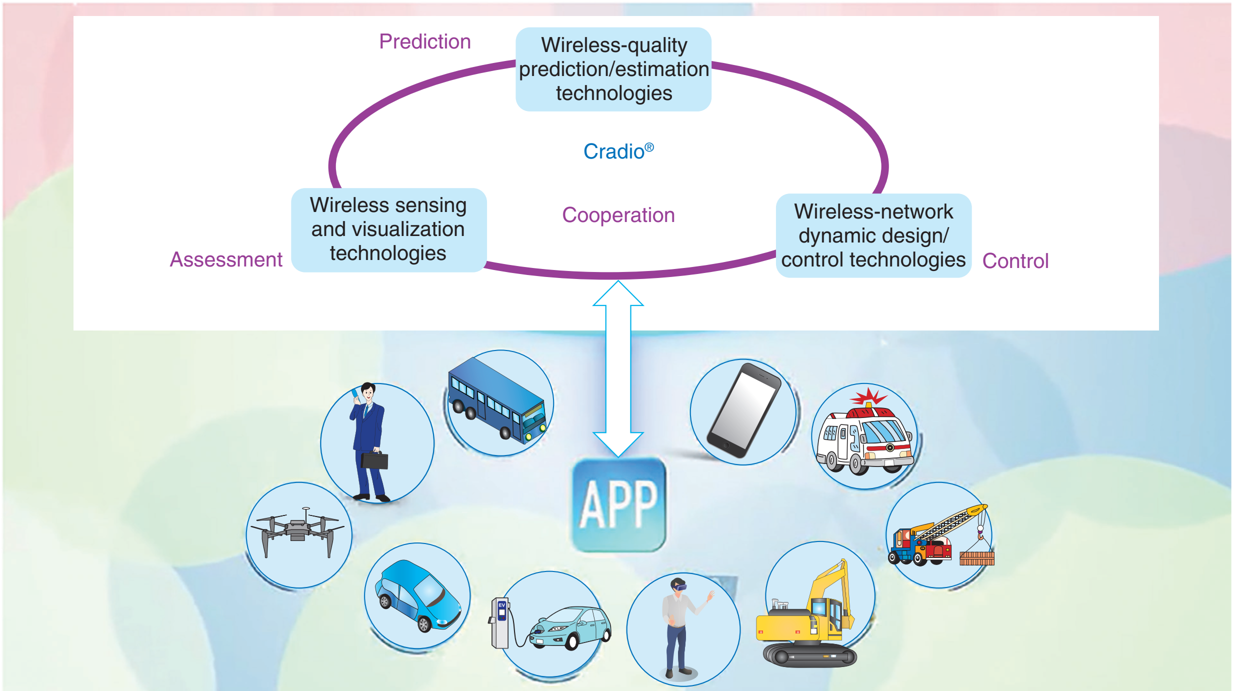


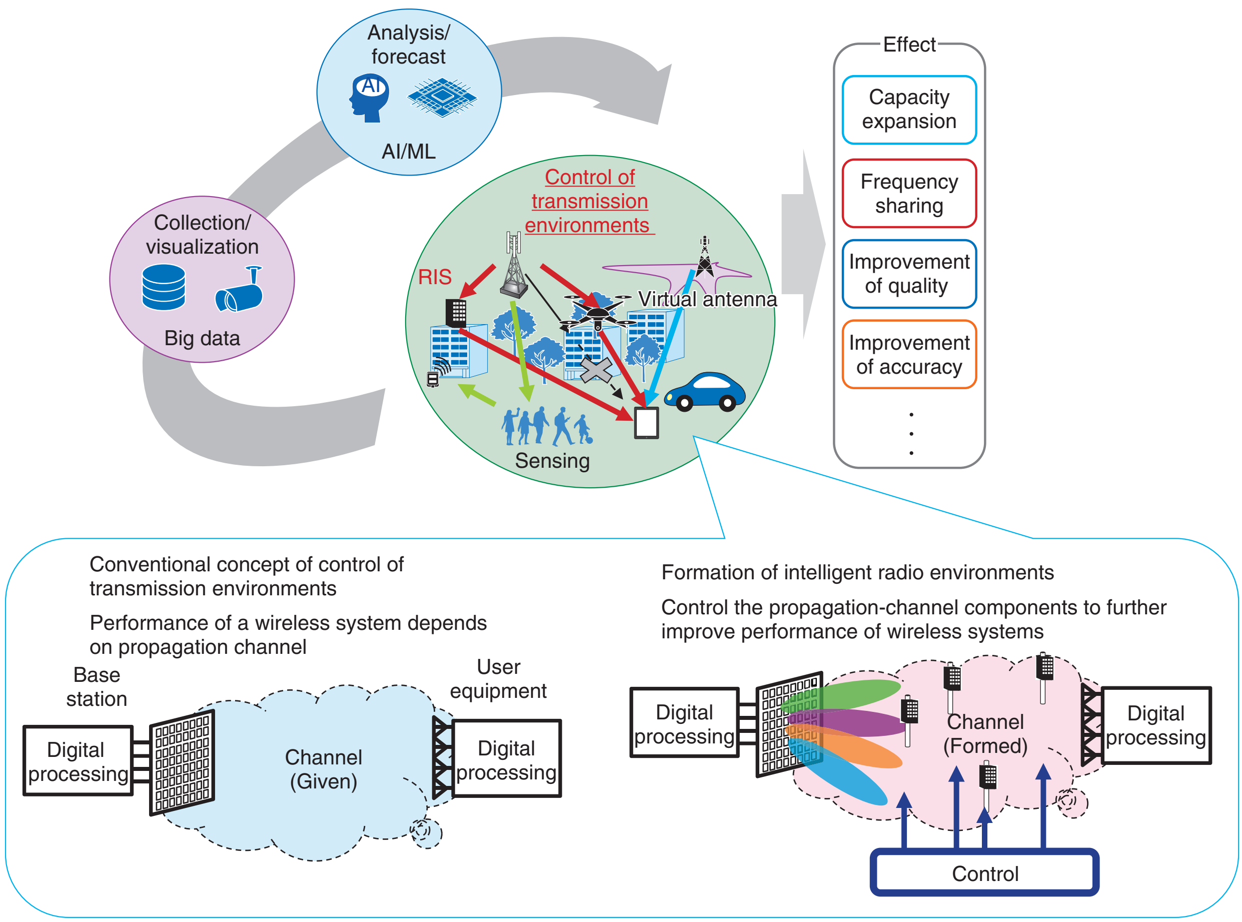
Fig. 1. Multi-radio Proactive Control Technologies (Cradio®) with various applications and social systems.

Against that backdrop, I am taking on the challenge of developing a wireless access network that offers new value with the following key phrases in mind: "achieving continuous connection," "sharing experiences," and "expanding into unexplored areas." 5G features high speed, large capacity, low latency, and simultaneous connection of many terminals. Looking toward IOWN and 6G, however, it will be necessary not only to improve the performance of each feature but also skillfully combine these features. Therefore, we are developing elemental technologies for extreme network as a service (NaaS) to deliver wireless access that satisfies individual *extreme* service requirements in a flexible manner whenever and wherever needed. One group of such elemental technologies for proactively controlling wireless access by assessing various wireless conditions is called Multi-radio Proactive Control Technologies or Cradio® (**Fig. 1**).

—You are attempting to create a communication environment in which people do not have to be aware of the wireless network. What exactly are you working on?

As one of the components of IOWN, Cradio is a group of wireless technologies that will provide users with a natural and uninterrupted communication environment without making them aware of the wireless network. It consists of the following three areas of technologies: (i) wireless sensing and visualization technologies are tasked with *assessment*, namely, collecting and visualizing wireless status and visualizing real-world status through means such as wireless sensing; (ii) wireless-quality prediction/estimation technologies are tasked with *prediction*, namely, predicting and estimating wireless communication quality that changes moment by moment; (iii) wireless-network dynamic design/control technologies are tasked with *control*, namely, designing physical locations of movable base stations, deriving optimal wireless parameters, and dynamically controlling network parameters, resources, etc. according to the environment and requirements. By developing the three areas of technologies and coordinating them in real time, we aim to satisfy user requests and control the momentarily changing wireless conditions to provide a natural communication environment that does not force the user to be aware of the wireless network.

To maximize the potential of wireless communications, we are also researching intelligent radio environment formation technology for controlling the



AI: artificial intelligence
ML: machine learning
RIS: reconfigurable intelligent surface

Fig. 2. Formation of intelligent radio environments.

environment in which radio waves propagate (**Fig. 2**), aiming at a paradigm shift from a *given propagation channel* to a *formed propagation channel* by controlling the propagation-channel components to achieve a fast and stable connection.

Up until now, we have considered that the environment is given to us and we had to build a communication system optimized for that given environment; however, we changed such perspective and started researching and developing various technologies for creating the environment ourselves. For example, when a typical reflector receives light or radio waves from a certain direction, it reflects the light or radio waves in a certain direction. However, using a recently developed reflector that can arbitrarily change the direction of reflection, we conducted an experiment to change the reflection environment.

We have used Cradio to drive autonomous farm machinery on a farm road in Iwamizawa City, Hokkaido. As the autonomous farm machinery runs across multiple wireless networks such as 5G, artificial intelligence predicts fluctuations in communication quality and automatically switches to an appropriate network before the communication quality degrades. Thus, we demonstrated stable and automatic driving of the machinery without interrupting the remote monitoring of the machinery. In coordination with NTT's Cooperative Infrastructure Platform, Cradio overcomes the differences in Internet Protocol addresses and communication schemes, thus provides connectivity independent of the particular access network. In this manner, users can use applications as usual without having to be aware of the network.

Accumulate experiments, measurements, and evaluations while confirming those results in the real world

—You were able to achieve results by changing the conventional perspective and approach, right?

Yes, we got that idea from a chat within our team. A technology can be *environment-dependent*; that is, one technology may produce good results in a certain environment but not the same results in another environment. We wondered if we had been bound by that notion and willfully set limits based on it. Therefore, we decided to overcome those limits by applying different ideas more freely, controlling the environment for ourselves, and changing our perspectives.

This shift in thinking has also been influenced by technological advances. In the past, only one type of wireless system existed in one area; however, various types now co-exist. As technology progresses, we'll likely end up with an environment in which many wireless systems overlap each other. When I pondered this situation, I thought that rather than looking at optimizing each system on a one-by-one basis, we should conduct research and development (R&D) assuming an environment in which multiple wireless systems overlap.

I'm thankful that great expectations are being placed on our R&D concerning wireless technology. This heightened expectation from society is echoed in the motivation of our researchers. People from various industries are interested in exchanging opinions on this topic. This type of stimulation can trigger new ideas and lead our R&D in the right direction. However, the more active the discussion, the more issues will arise. Although we are sequentially addressing the issues that have arisen, new issues may come up, and it will take some time to address them all. Nevertheless, we are determined to patiently repeat experiments and accumulate knowledge bit by bit. These experiments are low profile and physically demanding; however, they will provide us with unique data that no one else in the world has seen. Being able to create new technologies based on those data is rewarding.

—You conduct experiments not only in the laboratory but also in the actual field, as in the case of Iwamizawa City.

One of the interesting aspects of research activities is to actually go out into the field and conduct experiments and measurements. It is important to conduct experiments, measurements, and evaluation while confirming those results in the real world. When we consider what type of wireless environment is needed for use of better services in real life, the place for research activities is not the laboratory. Merely saying, "We have achieved various results in the laboratory." does not mean that the results are real. To be useful in the real world and meet the expectations of users, it is important to get out of the laboratory and verify our research results in the real world.

These experiments would not be possible without the collaboration of people who are facing problems, so it is very important to have an opportunity to conduct experiments in the field. In addition to conducting the experiment in Iwamizawa City, we have recently collaborated with people who are having trouble detecting damage caused by birds and other animals. The cost of the damage to crops caused by wild birds and animals is about 20 billion yen per year, of which about 80% is accounted for by wild boars and deer. Such damage also leads to secondary damage, such as loss of motivation to farm and abandonment of cultivation, which is becoming a serious social problem in Japan. As a countermeasure to such damage caused by birds and animals, we have been researching sensing technology using radio waves for detecting intrusions by animals.

The request from our collaborators was for a technology to remotely check what the sensors installed in the cages detected. When a sensor detects something, an administrator or other person is sent to check the cage; however, the person sent to check should depend on whether a wild boar, deer, or other animal is captured in the cage. Since the current system does not detect what is in the cage, the person sent to the site to check the content of the cage often finds captured birds and other animals that are not their responsibility, so another person—who is responsible for those animals—is sent to the site. Sending images wirelessly from the mountains where the cages are installed is a very difficult task because of many obstacles. However, we were able to carry out an experiment because that task matched a technology we were researching and developing.

Responding to requests from the fishing industry, we also experimented with transmitting images taken by underwater drones to the coast to obtain information on the types of fish caught in offshore nets and the condition of the nets. It is valuable experience to be able to see how well our technologies can perform in such real environments. When we asked people with actual problems to collaborate in experiments, the fact that they said, "Let's give it a try." indicates that they have high expectations for our research and technology. Being recognized in the real world in this way is rewarding for us.

Don't set limits, aim to break through them

—What do you keep in mind when you are looking for a research theme?

It is important not to set limits. When I hear someone say, "This is the most significant characteristic," I would think, "Something is wrong," and seek an approach to break the deadlock. This way of thinking came from my experience at Aalborg University in Denmark, where I studied for a year (2004 to 2005). After quoting Guglielmo Marconi's words, "It is dangerous to put limits on wireless," my supervisor asked me what limits I was thinking of overcoming in regard to future wireless technology. This question made me more determined to push the envelope, and I began to examine whether a goal is high enough or how unconventional it is when setting my research themes.

When IOWN becomes a reality, a world in which everything is interlinked will come. In this era of R&D, it is necessary to incorporate not only wireless technology but also a variety of other elements into the R&D process. This idea of incorporating R&D and perspectives from other fields has been important for a long time; however, it will become even more important in the future, and I think we should aim to adopt that idea.

If we consider only wireless technology, we tend to pursue R&D for evaluating communication quality and speed and determining how to improve it. Instead, it is important to verify those parameters in the actual environment, while keeping in mind how the user's experience and services can be improved, and create technology that works in conjunction with a real society. I also believe that we need to think of products and services that are not only easy to use from the viewpoint of the users but also easy to develop from that of the people who develop them.

Imagining the R&D world as a sphere, I think there are two approaches to R&D: one is to look at the surface of the sphere as a whole, and the other is to push toward the center of the sphere and hone one's expertise. I believe that since the sphere has only one center, the closer we get to the center by pursuing our specialties, the more our sense of commonality grows, no matter which field we pursue, and the more we can understand each other at essential points. I want to pursue R&D that pushes toward the center of the sphere.

—Do you have any words of encouragement for younger researchers?

I believe that age does not determine one's seniors and juniors in regard to research activities and hope that researchers will work hard together and improve each other's skills. When I talk to young researchers, I sometimes find it a bit of shame that although they all have very good sense and come up with the best idea very quickly, they tend to dismiss ideas that are a bit strange or didn't work out. I think it is also important to try out ideas that "might not work" with the mindset that what didn't work out this time will lead to a success down the road. The time you spend thinking about something is unique to you, and that experience in itself is an asset that you can use for your next R&D topic.

Also, don't lock your ideas away; instead, create new technologies by knocking those ideas against other ideas. I want to work with young researchers on these activities. Of course, sometimes we get stuck, and it can be difficult to make decisions about what to discard from research that has not been successful and whether to continue a research topic. In such a case, if you have the sense of "I like it" or "It's interesting" even though your judgment with logic tells you otherwise, you don't have to throw it away.

I have been with NTT for more than 26 years, and when I joined the company, there were times when our R&D on wireless technology did not bear fruit. However, when many researchers envisioned the future of devices and networks, the last-mile access network was always wireless, and I have been pursuing research believing that the time of wireless technology would come. In light of the growing interest in wireless technology, I am glad that I have been researching it for all these years.

R&D does not produce results in one or two years, so keep a long-term perspective and be persistent. I am sure that one day you will be glad you did.

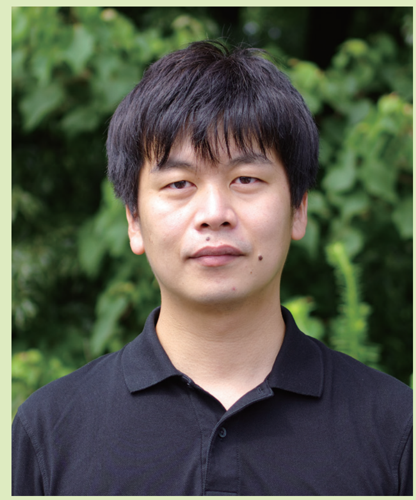
■ **Interviewee profile**
Yasushi Takatori

Senior Distinguished Researcher, Executive Manager, NTT Access Network Service Systems Laboratories.

He received a B.E. and M.E. in electrical and communication engineering from Tohoku University, Miyagi, in 1993 and 1995. He received a Ph.D. in wireless communication engineering from Aalborg University, Denmark, in 2005. He joined NTT in 1995 and has served as a secretary of Institute of Electrical and Electronics Engineers (IEEE) Japan Council Awards Committee and a vice chairman of the Wireless LAN system development project in the Association of Radio Industries and Businesses (ARIB). He was a visiting researcher at the Center for TeleInFrastructure (CTIF), Aalborg University from 2004 to 2005. He served as a co-chair of COEX Adhoc in IEEE 802.11ac from 2009 to 2010. His current research interests include future wireless access systems for IOWN. He received the Best Paper Awards from the Institute of Electronics, Information and Communication Engineers (IEICE) in 2011, 2016, and 2020. He was honored with the IEICE KIYASU Award in 2016. He received the Radio Achievement Award from ARIB in 2020 and the IEEE Standards Association's Outstanding Contribution Appreciation Award for the development of IEEE 802.11ac-2013 in 2014. He is a senior member of IEICE and a member of IEEE.

Research on Transfer Learning to Give AI the Same Versatility and Skill as Humans

Atsutoshi Kumagai
Distinguished Researcher,
NTT Computer and Data Science Laboratories and NTT Social Informatics Laboratories



Overview

Machine learning is needed to build artificial intelligence (AI), and this requires a large amount of training data. Sometimes, however, you cannot get enough high-quality training data. What's more, to prevent an AI from becoming out of date, you need to re-train it regularly to keep pace with data that is constantly changing. In this interview, Distinguished Researcher Atsutoshi Kumagai talked to us about the technology of transfer learning, which can be used to improve AI performance even when ideal data cannot be obtained.

Keywords: transfer learning, AI, anomaly detector

Transfer learning improves AI performance even without sufficient data

—What sort of technology is transfer learning?

When an artificial intelligence (AI) is given an unknown task or a task without sufficient training data, transfer learning technology helps improve performance by leveraging training data from related tasks to compensate for data shortages. We are studying the possibility of using transfer learning technology to give AI the same versatility and skill as a human. For example, we humans can do a variety of different tasks like cooking, doing simple calculations, reading and writing, talking, running, throwing a ball, and more. And when we're presented with a new task we've never seen before, we can use our past knowledge and experience to adapt to and get the hang of it in relatively few attempts.

On the other hand, recent AIs have been able to perform as well as, and in some senses exceed, humans in specific tasks for which large amounts of data can be prepared, such as image recognition. One example of this is the news that an AI that plays *go*, called AlphaGo, beat the top player in the field. However, AIs do not have the versatility or skill that humans have. This is common knowledge among researchers like myself.

That's why transfer learning technology is key when trying to build a versatile and skilled AI. The

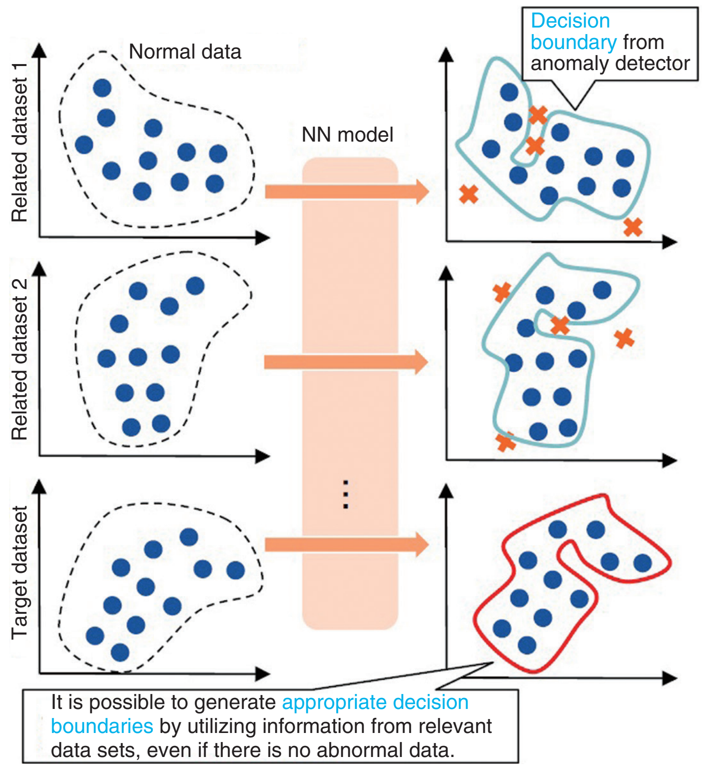


Fig. 1. High-speed creation of an anomaly detector based only on normal data.

simplest way to build a versatile AI is to list all the tasks and provide large amounts of training data for each task. However, actually listing every task is no simple task, and even if you manage to do so, gathering the large amounts of training data required for each task is not practical in terms of cost or resources. Transfer learning technology can help us build versatile and skilled AI through performance improvements using related training data to compensate for data shortages, even for unknown tasks.

—What specific research is being conducted into this technology?

The first piece of research being carried out is on technology for rapidly generating appropriate anomaly detectors in situations where only normal data is available. Anomaly detection is the task of finding abnormal data, which is data with different properties from normal data. To create a highly accurate anomaly detector, both normal and abnormal data sets are generally required as training data. However, since abnormal data is very rare in nature, and so is hard to obtain in practice, it is often not practical for teaching anomaly detectors.

The approach I'm researching uses similar data sets that contain both normal and abnormal data to generate an anomaly detector, even when only normal data is available. **Figure 1** is a diagram showing an appropriate anomaly detector being generated by inputting a target data set with no abnormal data into a trained neural network (NN) model. The NN model is trained in advance using several related data sets (1, 2...) that include normal and abnormal data to map the normal data group onto the anomaly detector.

Normally, AI requires re-training each time a new problem is presented, which is a process with a very high computational cost. However, with this approach, anomaly detectors can be generated by simply entering a new data set into an NN model, making it a good fit in cases where a real-time response is required, or in cases where computational resources are limited, but you still want to perform anomaly detection.

The second piece of research is on technology for transfer learning for tasks that change over time. Machine learning learns something called a "classifier" that classifies data. But as the data changes over time, so does the classifier. These changes are very

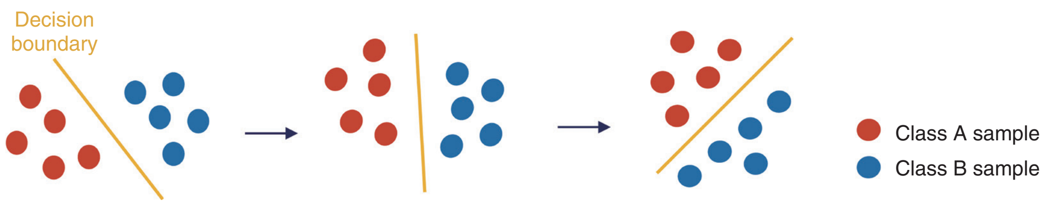


Fig. 2. Change in the decision boundary of the classifier due to a change over time.

common in real-world situations such as malware detectors, where attackers are continually developing new attack methods to slip past detectors. As a result, continuing to use malware detectors that have been trained once without updating them results in the classification accuracy continually deteriorating.

An effective method for preventing the accuracy of the classifier from degrading is to update the classifier in a timely manner using new additional data. This method is simple and very effective, but collecting additional data costs money. Such image data must be labeled, for example "this image is a cat" or "this image is a dog," but this is work that must be carried out by human beings, making it somewhat costly. In addition, there may be other factors that make it difficult to obtain labeled data, such as the need to protect personal data.

The approach I'm researching involves predicting future decision boundaries from decision boundaries learned using labeled data, without the need for any additional labeled data. **Figure 2** shows a decision boundary with labeled Class A and Class B samples that goes through three stages over time, allowing the next change to the decision boundary to be predicted. Transfer learning usually assumes there will be data available from where it will be applied, but what's interesting about this study is that it does not use additional labeled data.

—What are the current challenges you're facing?

The main challenge is how to ensure that the model is correct after training. In general machine learning, separate validation data is prepared to check the trained model. The accuracy of the learning can be assessed by how correctly it discerns between the validation data.

However, in research like this, it can sometimes be difficult to create data for validation in the first place, or it may not exist at all. In particular, there is no data that can be used for validating future forecasts, so it is difficult to determine their accuracy. This is a major challenge because safety must also be ensured when considering practical applications in the future, especially in mission-critical areas.

Aiming to build versatile and skilled AI in the future

—What fields can this research be applied to?

For example, in the field of anomaly detection, suppose there's a manufacturer that owns several factories. They build a new factory, and they want to make an AI that will automate the monitoring of equipment on site. This is equivalent to creating an anomaly detector for identifying abnormal data that differs from the normal data for the equipment. As the new factory has no operational results, it may not be possible to create a high-performance anomaly detector until enough abnormal data is collected for a rare equipment. However, if there are other factories that have been running for a long period of time, sets of normal and abnormal data from those factories can be used to create highly accurate anomaly detectors as soon as the data from the new factory is entered. In addition, to use a transmission service to centrally manage the security of your network for multiple customers, you can automatically create conditions that are tailored to your customers by simply entering the correct data.

Speaking of transfer learning technology for tasks that change over time, another example of the security issues I mentioned earlier is applying transfer learning to automatically updating anti-virus software. In addition, some companies may use a "blacklist" of sites that must not be connected to in order to protect the company from malware, and the technology we have developed can be used to create these blacklists.

E-commerce is another potential area of application. Transfer learning can be used to accurately

predict customers' hobbies and preferences as they get older, and recommend the right products. It can also be used to estimate the needs of new and infrequent users. It may be difficult to keep making predictions over a long period of time, but I think it may be possible to reduce costs by slightly increasing the time between updates. Research into increasing accuracy in situations where there are multiple data sets is applicable in many areas, so it should have a wide scope of application.

—What are the plans for future research?

At present, there are areas in which transfer learning can be used to ensure a practical level of accuracy, such as in image recognition and language processing. However, we believe that there are also blue-ocean areas where the potential of transfer learning is being overlooked and not being effectively utilized. As a medium-term goal, I would like to focus on such fields, and conduct research into broadening the scope of transfer learning. The lack of training data can be a problem in many fields, so I hope that transfer learning will be a way of solving this practical problem.

In the long term, I would like to build an AI that has the versatility and skill we discussed earlier. It will be interesting if AI become able to learn independently and do anything with just a single model, rather than humans having to teach and tune the AI for every new problem or task.

■ Interviewee profile

Atsutoshi Kumagai

Distinguished Researcher, NTT Computer and Data Science Laboratories and NTT Social Informatics Laboratories.

He joined Nippon Telegraph and Telephone Corporation (NTT) in 2012, where he researches machine learning and cybersecurity. He worked as a member of NTT Secure Platform Laboratories (2012 to June 2021) and NTT Software Innovation Center (2018 to June 2021). He has been with NTT Computer and Data Science Laboratories and NTT Social Informatics Laboratories since July 2021.

Research and Development of Digital Twin Computing for Creating a Digitalized World

Ryo Kitahara, Takao Kurahashi, Toru Nishimura, Ichibe Naito, Daisuke Tokunaga, and Koya Mori

Abstract

The Digital Twin Computing Research Project aims to make Digital Twin Computing, a key component of the Innovative Optical and Wireless Network (IOWN) initiative promoted by NTT, a reality. We established four grand challenges to achieve this objective and have been conducting research and development to meet each one. This article outlines each grand challenge and associated technologies and describes horizontal activities across these challenges.

Keywords: Digital Twin Computing, grand challenges, IOWN

1. Grand challenges for achieving Digital Twin Computing

Digital Twin Computing (DTC) is a key component of the Innovative Optical and Wireless Network (IOWN) initiative promoted by NTT [1]. It aims to make large-scale and high-accuracy predictions and trials of the future beyond the limits of conventional information and communication technologies. It also aims to achieve advanced forms of communication having new value by combining various types of high-accuracy information related to things, people, and society in the real world. The idea is to create a smart society by solving a variety of social issues affecting the world and creating innovative services.

We established "Mind-to-Mind Communications," "Another Me," "Exploring Engine for Future Society," and "Inducing Inclusive Equilibrium Solutions for the Earth and Its Social and Economic Systems" as four grand challenges for achieving DTC as announced in an NTT press release in November 2020 [2]. We have since begun specific studies on each of these grand challenges. In this article, we outline these challenges and the technologies being developed to meet them and introduce future objectives. We also describe activities designed to accelerate the meeting of these four challenges.

2. Mind-to-Mind Communications

"Mind-to-Mind Communications" is a research and development (R&D) objective to achieve a new way of communicating that will enable people to directly understand how other people perceive and feel things. This new form of communication will transcend not only language and cultural differences but also individual characteristics such as one's experiences and sensitivities.

To achieve this objective, we are undertaking the R&D centered about three key technologies: mind-to-mind modeling, mind-to-mind conversion and presentation, and state transformation and environment creation. First, mind-to-mind modeling technology uses sensor information consisting of biological signals, images, speech, etc. to infer expressions in mind-to-mind communication, that is, in ways of perceiving and feeling things. Next, mind-to-mind conversion and presentation technology converts and presents expressions in mind-to-mind communication that differ from one individual to another in a

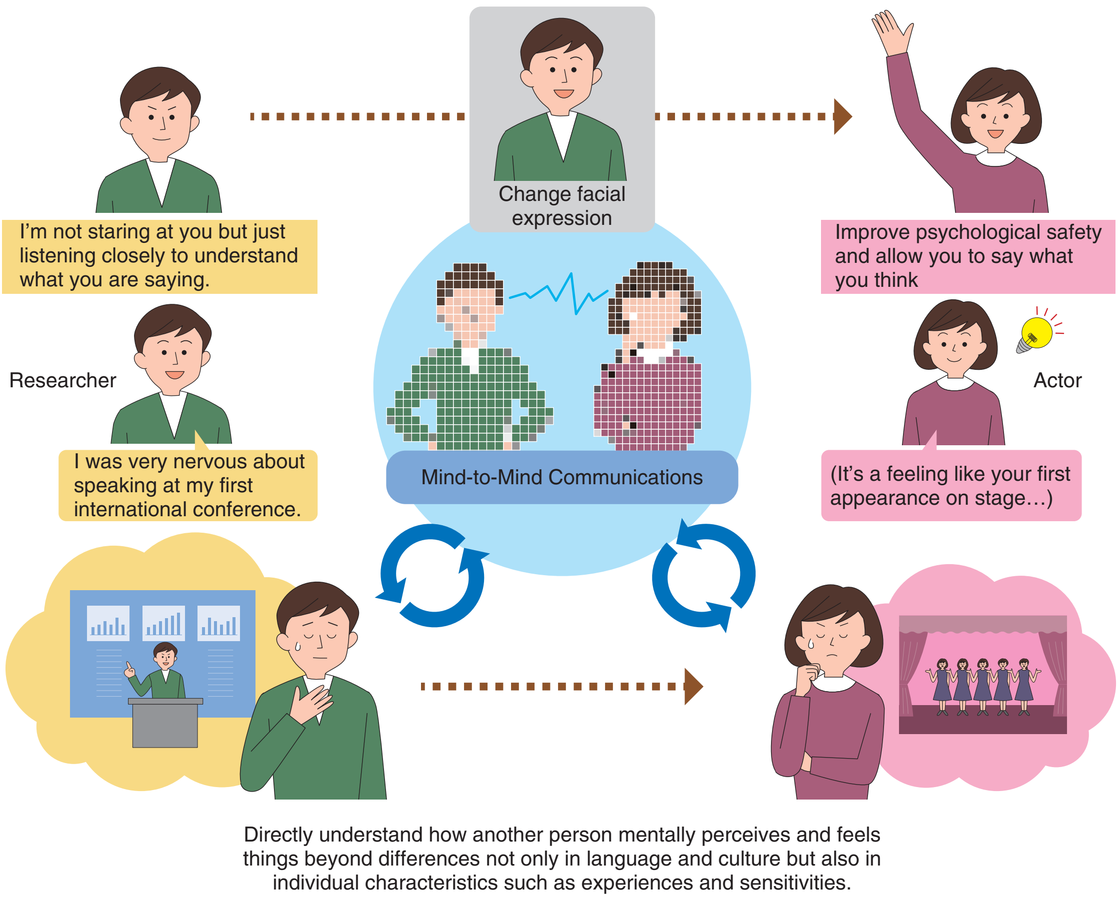


Fig. 1. Mind-to-Mind Communications.

form suitable to one's communication partner. Finally, state-transformation and environment-creation technology promotes a change in how one perceives and feels things by intervening not only in terms of the information exchanged in communication but also in regard to a person's inner world, such as receptivity, and the communication site.

Achieving this R&D objective will lead to a new stage of communication that will solve the problems of discommunication (failure to communicate one's intentions) and miscommunication (mutual misunderstanding) prevalent in past ways of communicating. This will enable a person to realistically and directly understand how another person mentally perceives and feels things beyond differences in language, culture, experiences, values, sense perceptions, and sensitivities. By improving psychological safety and promoting mutual understanding, Mind-to-Mind Communications creates an inclusive society in which people having diverse characteristics engage in activities together while exploiting those characteristics, thereby inspiring and helping each other to grow (**Fig. 1**).

3. Another Me

The loss of various opportunities in one's life due, for example, to the difficulty of balancing childcare and caregiving with work or the inability to participate in society despite an interest or desire to do so, is becoming a social problem. To dramatically increase the opportunities for a person to become more active and grow while expanding and merging the range of activities from the real world to the virtual world, we are taking up the challenge of achieving "Another Me" as another version of oneself in digital form who can participate actively as oneself in a way that transcends the limitations of the real world and who can share the results of such activities as experiences (**Fig. 2**). Our main initiatives in meeting this

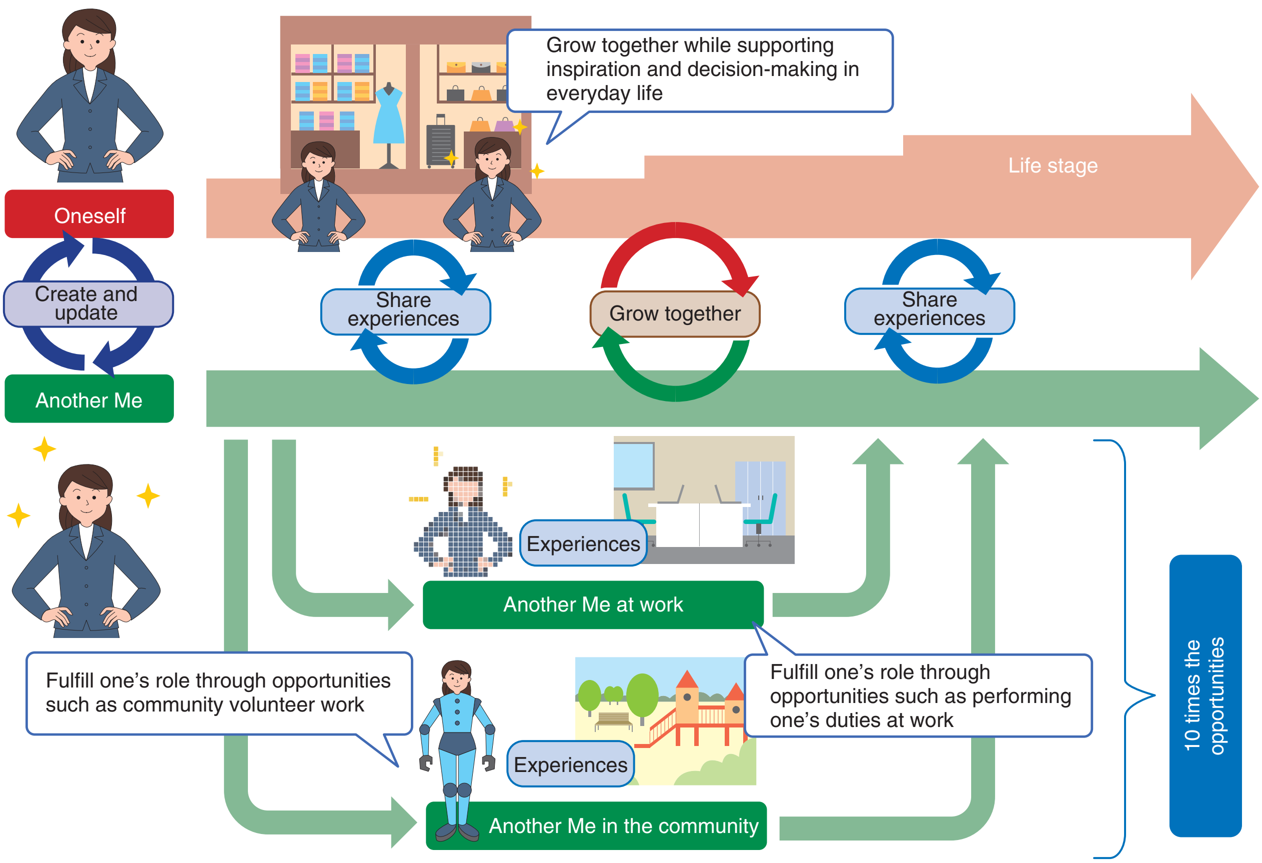


Fig. 2. Another Me.

challenge are summarized below.

(1) Transcend the limitations of the real world: To transcend the limitations of time, space, and personal ability that reduce one's opportunities in life, we seek to establish autonomous agent technology that enables Another Me to automatically perform activities in place of the user and telepresence/body-augmentation technology to accommodate remotely located users or solve ability-related issues of the physically challenged, elderly, etc.

(2) Foster the user's sense of self-agency: To make the user feel as if the activities of Another Me are those done by oneself and experience a feeling of accomplishment and growth, we seek to establish experience-transfer technology to achieve psychological teleportation and feedback. So that the actions of Another Me can be recognized by the user and others as those of the user himself/herself, we will also establish technology for creating and nurturing a human digital twin who can develop a personal image that changes from day to day and exhibit individual characteristics such as a sense of values and personality.

(3) Social acceptance research: We will investigate philosophical, ethical, and legal issues on the basis of how we view ourselves toward widespread adoption of Another Me in society.

This approach will accelerate self-realization by providing a variety of opportunities in one's life and improving one's state of mind, health, and motivation while enabling the experiences of Another Me to grow as those of one's own through many growth opportunities.

4. Exploring Engine for Future Society

Due to the increasing complexity of the social structure and an increase in unpredictable events such as natural disasters, it is becoming difficult to observe

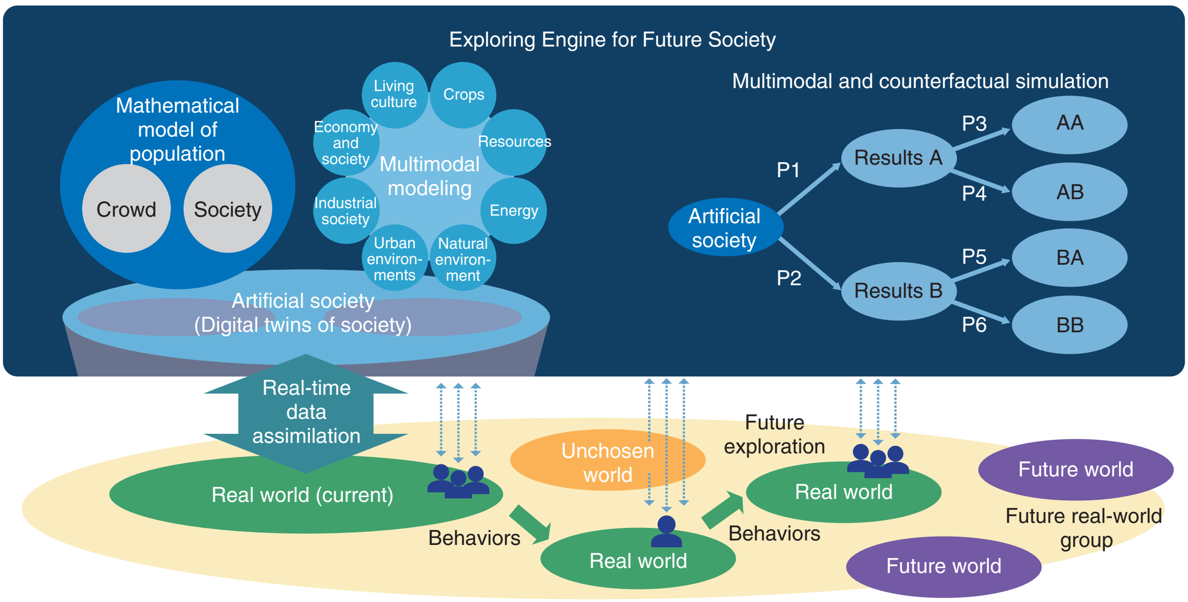


Fig. 3. Exploring Engine for Future Society.

the effects of individual behavior and cooperation on society and the natural environment and how those effects come back to affect the individual. Under these circumstances, it would be desirable to search for ways that society and lifestyles can coexist with nature. To this end, we are researching and developing an "Exploring Engine for Future Society." This engine will divide the society that people are active in into the three layers of environment, city, and people and express these layers as digital twins. It will then use these digital twins to simulate the transformation of this multilayered society.

To make this engine a reality, we are researching and developing technology for creating in real time digital twins of the real world obtainable through sensing and for modeling abstract information such as public opinion and individual consciousness. We intend to use basic technologies such as previously developed temporal-spatial data management technology. We are also researching and developing basic simulation techniques to combine the above with city and environment simulations. These technologies will make it possible to drive society's digital twins and present a variety of future scenarios and candidate measures that should be taken now by calculating backward from an ideal image of society.

Visualizing various forms that future society can take in this manner should make it possible to bolster various types of social solidarity by having not only government and corporate management but also local residents and individual citizens view the same future images. Additionally, we aim for a world in which certain types of behavior and cooperation can be proactively searched for and spontaneous actions can be encouraged so that a society desired by the individual can be achieved (**Fig. 3**).

5. Inducing Inclusive Equilibrium Solutions for the Earth and Its Social and Economic Systems

In view of climate change and social crises now occurring on a global scale, discussions have begun on reconstructing social and economic systems that can maintain an autonomous equilibrium on the Earth. This is a problem with a high degree of uncertainty, so it is important to prepare multiple options and scenarios and study and execute comprehensive approaches. With this in mind, we have taken up the challenge of presenting options with regard to the transformation of energy, the economic system, and human society. The aim is to determine changes and mutual effects (chain reactions) among the many different systems making up the Earth, such as the environment, ecosystem, energy, economic systems, and human society, and preserve an autonomous equilibrium that can maintain a balanced state in which no system will exceed a critical point and collapse.

First, to meet this challenge using as themes the

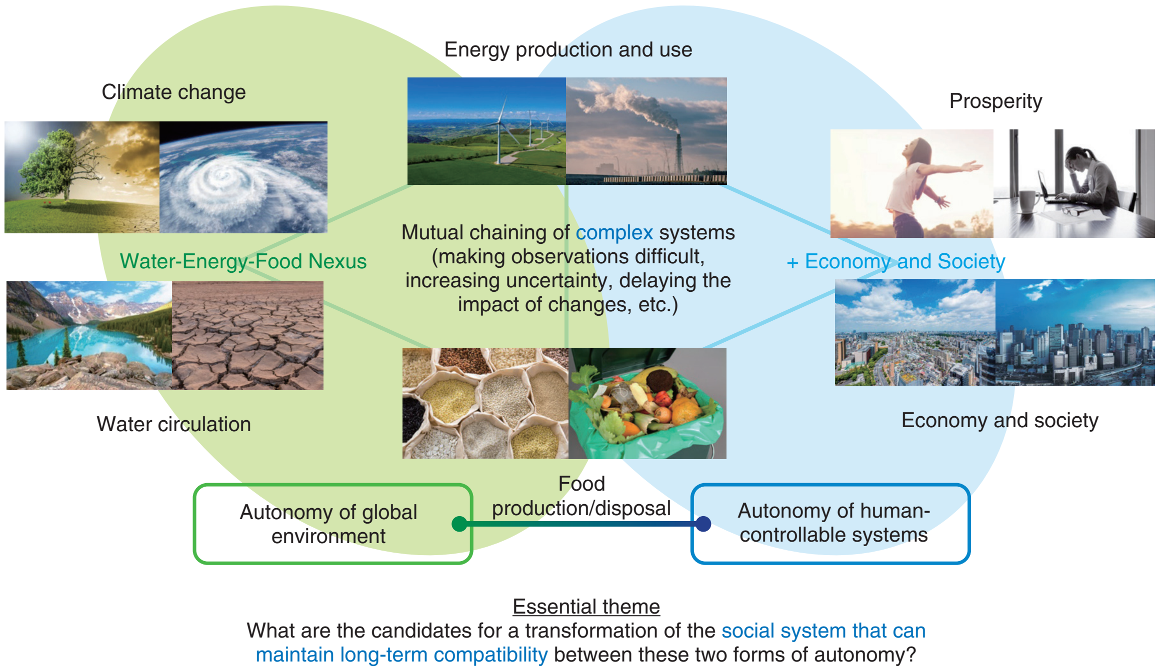


Fig. 4. Inducing Inclusive Equilibrium Solutions for the Earth and its Social and Economic Systems.

energy supply corresponding to changes in climate and water circulation along with economic systems and human society, we will undertake the R&D of (1) large-scale multimodal technology for massive coupling of model groups of many and varied phenomena and simulation of inter-model chain reactions, (2) modeling technology for designing on a global scale while taking into account inter-system coupling between the environment and economic systems, and (3) techniques for determining key indices of autonomous equilibrium and local solutions of those indices and techniques for evaluating index-value groups for each scenario.

In this way, we will be able to link many and varied large-scale models and conduct inclusive circulation simulations on a global scale of diverse and complex phenomena occurring in the real world throughout the Earth while presenting multiple options for a social transformation (**Fig. 4**).

6. Horizontal activities for the grand challenges

Technology development for meeting the above grand challenges covers a wide range of fields, so we are engaged in unified development activities across these grand challenges. We introduce three of these activities: (1) creation of a DTC reference model, (2) collaboration with partners, and (3) standardization activities.

(1) Creation of a DTC reference model

In creating a DTC reference model that compiles technical achievements across multiple projects, we have documented digital-twin concepts common to DTC while taking a comprehensive view across the four grand challenges and released the document both inside and outside NTT laboratories [3]. This document includes conceptual structures now being studied at the NTT Digital Twin Computing Research Center with respect to the definition of digital twins and configuration of human digital twins. The aim of releasing this information is to promote discussion with partner companies having a common vision and make joint explorations of even more essential DTC concepts. Going forward, we will add more research results as they become available from the NTT Digital Twin Computing Research Center to provide an even more complete reference document.

(2) Collaboration with partners

To accelerate the R&D of the above grand challenges through collaboration with partners, we will search for outside partners to accelerate DTC R&D and work on creating a collaboration scheme. Although necessary technology areas differ from one grand challenge to another, we will pursue a broad range of partnering, including tie-ups with universities, other research institutions, and startup companies targeting a wide range of fields from artificial intelligence technologies for understanding human emotions and behavior to digital-twin technologies for smart cities. The field of digital twins is relatively new, but at the same time, it is connected to a wide range of technology areas such as virtual reality (extended reality, etc.) and digital humans; therefore, many startup companies with deeply interesting technologies are appearing and growing. It is our belief that establishing collaborative partnerships with such companies will accelerate R&D and enable the early implementation of novel technologies.

(3) Standardization activities

Finally, we will engage in standardization activities to standardize and spread DTC on a global scale. As part of this effort, we will promote activities for visualizing DTC and research results at global forums. Specifically, we are participating in the Digital Twin Consortium and the IOWN Global Forum while making contributions and presentations in relation to DTC and grand challenges. Through these activities, we hope to increase NTT's presence in the digital twin field and promote standardization activities for digital-twin technologies together with global partners we can collaborate with in making DTC a reality.

References

[1] IOWN, https://www.rd.ntt/e/iown/

[2] Press release issued by NTT, "NTT Announces New R&D Projects of Digital Twin Computing," Nov. 13, 2020. https://group.ntt/en/newsrelease/2020/11/13/201113c.html

[3] NTT Digital Twin Computing Research Center, "The Digital Twin Computing Reference Model," Mar. 2021. https://www.rd.ntt/_assets/pdf/iown/reference-model_en_2_0.pdf

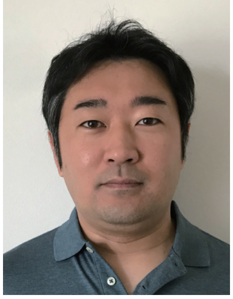
Ryo Kitahara
Senior Research Engineer, NTT Digital Twin Computing Research Center.
He received a Master's degree in information science from Tohoku University, Miyagi, in 2003 and joined NTT the same year. His research interests include image-processing technologies. Since 2020, he has been in charge of research on Digital Twin Computing at NTT Digital Twin Computing Research Center.

Takao Kurahashi
Senior Research Engineer, NTT Digital Twin Computing Research Center.
He received an M.E. in industrial engineering from Hosei University, Tokyo, in 2002 and joined NTT the same year. His research interests include human-computer interaction and robotics.

Toru Nishimura
Senior Research Engineer, NTT Digital Twin Computing Research Center.
He received a Master's degree in information engineering from Nagoya University, Aichi, in 1998 and joined NTT the same year. He has been involved in the R&D of an operation support system for information technology systems including networks. Since 2020, he has been in charge of research on Digital Twin Computing at NTT Digital Twin Computing Research Center.

Ichibe Naito
Senior Research Engineer, NTT Digital Twin Computing Research Center.
He received Bachelor's and Master's degrees in information science from Waseda University, Tokyo, in 2006. Since joining NTT in 2006, he has been engaged in R&D of distributed data stream management systems and personal information management systems. In 2010, he moved to NTT Communications, where he designed the database of operation support systems and developed an IaaS (infrastructure as a service) service. In 2013 he joined NTT Software Innovation Center and developed operation support systems for distributed object storage. His current research interests include efficient management of huge geospatial data.

Daisuke Tokunaga
Senior Research Engineer, NTT Digital Twin Computing Research Center.
He received a Master's degree in information science from Kyushu Institute of Technology, Fukuoka, in 1999 and joined NTT the same year. His research interests include large-scale application system technologies. Since 2020, he has been in charge of research on Digital Twin Computing at NTT Digital Twin Computing Research Center.



Koya Mori
Senior Research Engineer, NTT Digital Twin Computing Research Center.
Since joining NTT in 2004, he has been involved in the R&D of an Internet-of-Things application based on software technologies, such as OSGi, OpenStack, and edge computing. Since 2020, he has been in charge of research on Digital Twin Computing at NTT Digital Twin Computing Research Center.

Thought-processing Technology for Understanding, Reproducing, and Extending Human Thinking Ability

Kyosuke Nishida, Takeshi Kurashima, Noboru Miyazaki, Hiroyuki Toda, and Shuichi Nishioka

Abstract

The Human Insight Laboratory at NTT Human Informatics Laboratories aims to deepen our understanding of human thinking ability, which consists of human information understanding, human information processing, and response and behavior, and to create technologies that can reproduce these abilities on computers and expand human thinking ability. In this article, the technologies of visual machine reading comprehension, behavior modeling, speech recognition, and thought-enhancing stimulus design are introduced.

Keywords: human information understanding, human information processing, human insight

1. Introduction

The goal of Human Insight Laboratory of NTT Human Informatics Laboratories is to deepen our understanding of (1) human information understanding (from perception to cognition), (2) human information processing (for interpreting acquired information and taking the next action), and (3) reaction and action (for acting on the outside world). We create technologies to reproduce these processes on computers for expanding human thinking ability. We introduce four technologies that will lead to this goal, i.e., (i) visual-machine-reading-comprehension technology for understanding documents visually, (ii) behavior-modeling technology for considering the mechanism of human decision-making and behavior, (iii) speech-recognition technology for understanding a human's internal state, and (iv) thought-enhancing stimulus design for drawing out and expanding a human thinking ability.

2. Visual-machine-reading-comprehension technology for understanding documents visually

To further develop information retrieval and dialogue/question-answering services, the Human Insight Laboratory has been researching machine reading comprehension, by which artificial intelligence (AI) reads text written in natural language (i.e., the language used in our daily lives) and understands its meaning [1, 2].

Research on machine reading comprehension has made great progress, and according to evaluation data, AI has exceeded human reading comprehension; however, machine reading comprehension is limited in that it can only understand textual information. The PDF (portable document format) documents and presentation slides that we use on a daily basis contain not only linguistic information but also visual elements such as font size and color, figures, tables, graphs, and layout information. An integrated understanding of vision and language is therefore essential in regard to developing AI to support office work and daily life.

To achieve visual machine reading comprehension

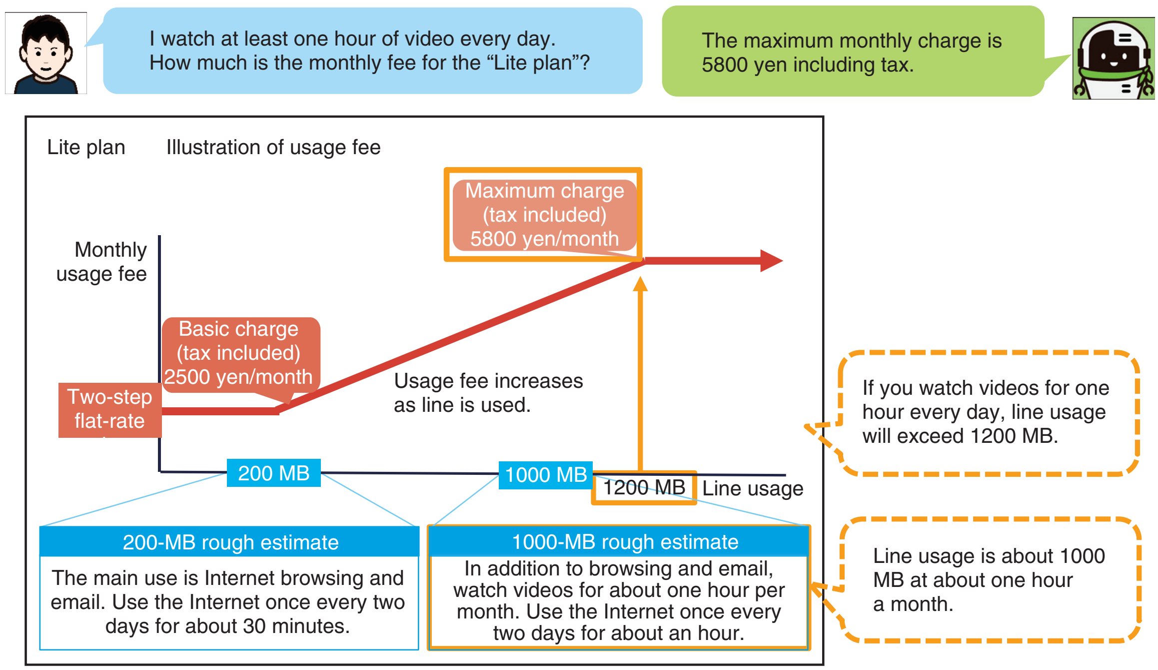


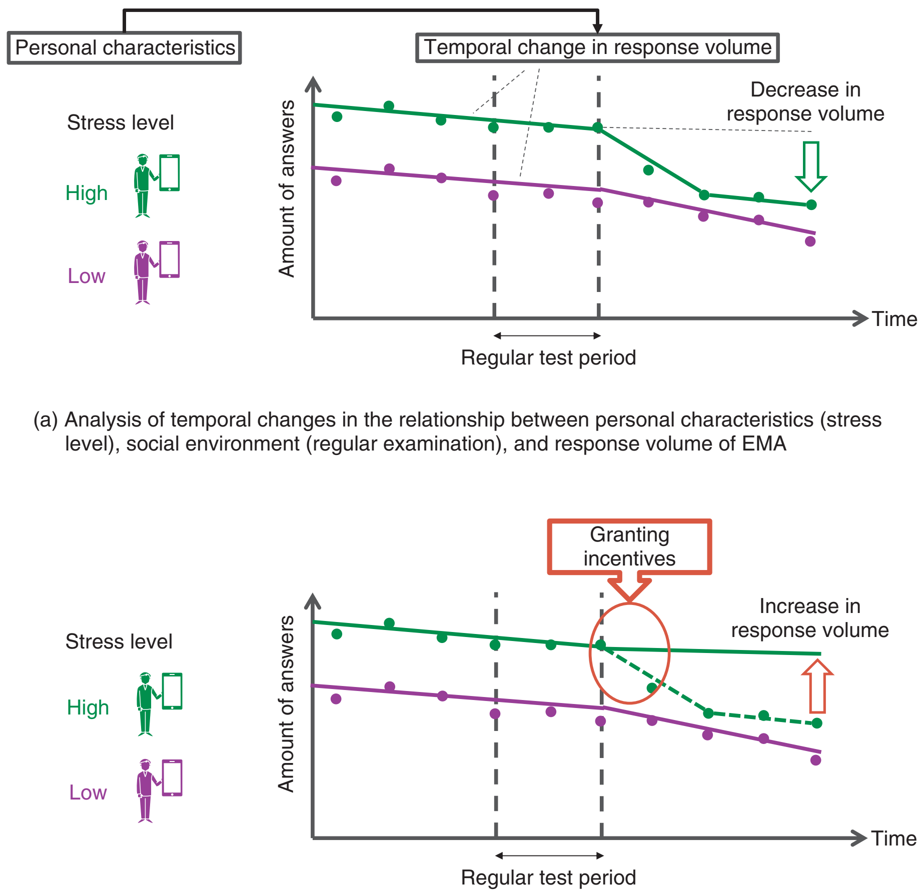
Fig. 1. Visual machine reading comprehension.

of document images (**Fig. 1**), we at the Human Insight Laboratory have constructed a dataset called VisualMRC [3] and are pushing ahead with research using it. This dataset consists of free-text question-answer data concerning document images of webpage screenshots, and regions in documents are annotated with nine classes, including title, paragraph, list, image, and caption. We conducted research using this dataset and proposed a visual-machine-reading-comprehension model that can take into account—as additional input—areas in the document (extracted using object-recognition technology) as well as position and appearance information of "tokens" (extracted using optical character recognition technology) [3]. Although this model is not yet able to match the accuracy of human question answering, we confirmed that understanding the visual information of documents improves the performance of question answering compared with that of text-only models. We will continue to work on various research projects for integrated understanding of vision and language.

3. Behavior-modeling technology for considering the mechanism of human decision-making and behavior

We are studying the mechanisms of human decision-making and behavioral judgments (including irrational judgments) by using both humanistic and sociological knowledge (such as behavioral economics) and a variety of human data that have become available with the recent spread of the Internet of Things. We are also constructing a behavioral model that reproduces human decision-making and behavior by using the obtained knowledge as a framework and aim to apply this model to enable future predictions and simulations about people. When an undesirable future is predicted, we will be able to use the simulation results to search for ways that will lead to a brighter future.

As part of our efforts to explain the mechanism of human decision-making, we have been studying the effects of factors, such as personality, values, physical and psychological states, and social environment, on a person's decision-making and behavior. Ecological momentary assessment (EMA), also known as the experience sampling method, is a survey method with which participants are sent simple questions about their situation, thoughts, emotions, and behavior via a portable device (such as a smartphone or tablet) and asked to spontaneously respond to the questions. Using the results of a 10-week EMA questionnaire administered to university students [4], we analyzed in detail the differences in the students'



(a) Analysis of temporal changes in the relationship between personal characteristics (stress level), social environment (regular examination), and response volume of EMA

(b) Give an incentive when the amount of responses decreases and encourage an increase in the amount of responses

Fig. 2. Search for strategies on the basis of results of behavioral-data analysis.

tendency to respond to the questionnaire on individual characteristics from the perspective of temporal changes [5]. For example, as shown in **Fig. 2(a)**, participants who were judged to have high stress levels on a daily basis (according to a preliminary survey taken before the start of the EMA questionnaire) tended to actively self-disclose information to their device throughout the survey period. However, our analysis revealed that immediately after a stressful event (e.g., a regular exam), the students tended to be uncooperative toward answering, and their volume of responses dropped sharply. We also found that the participants who were judged to have high integrity in the preliminary survey were cooperative and gave a large amount of responses immediately after the start of the EMA questionnaire; however, the amount of their responses tended to decrease significantly over time. By deepening our understanding of the temporal changes in such human behavior (in this case, the act of voluntarily answering a questionnaire), it will be possible to predict what will happen regarding people's behavior and outcomes, and it will be easier to make decisions on how to change the future for the better. From the viewpoint of constantly understanding a person's condition, the EMA questionnaire should secure a uniform number of responses without bias in time or participant. For example, when the amount of responses is expected to drop, as shown in **Fig. 2(b)**, it is possible to give special incentives to participants with high stress levels to encourage them to respond after the test.

The above example is just an analysis that captures one aspect of complex human behavior. In the future, we plan to make our human decision-making and behavioral models [6] more sophisticated by taking into account behavioral economic human characteristics

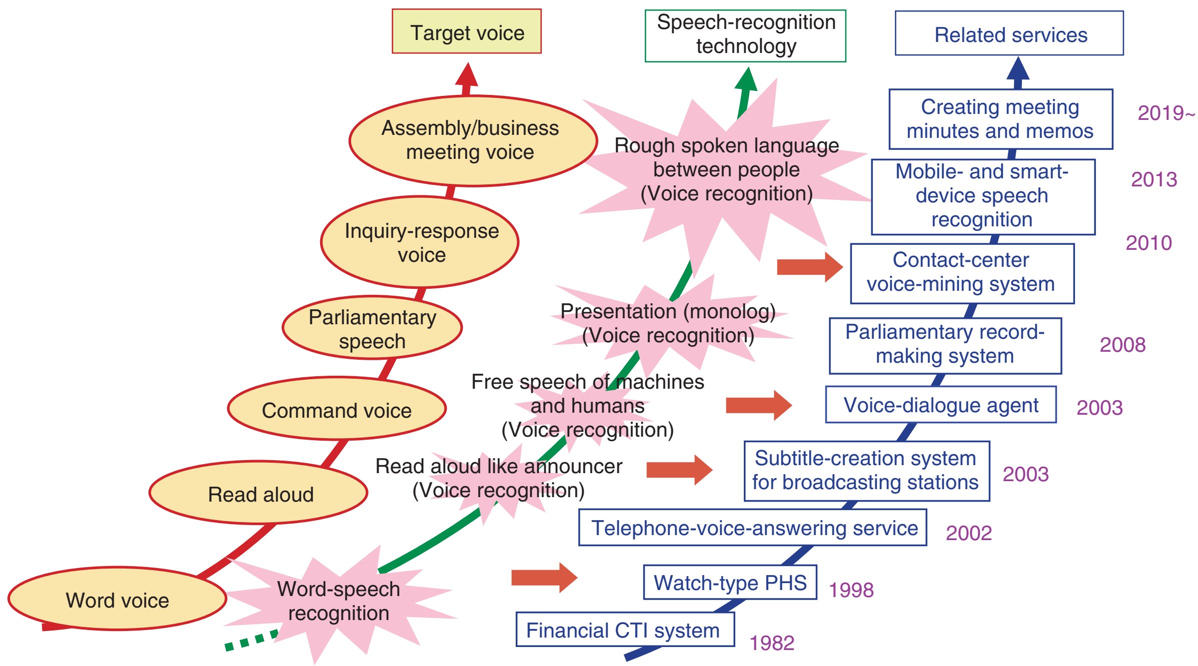


Fig. 3. Efforts concerning voice-recognition technology at NTT.

such as attitude toward uncertainty (risk-aversion tendency) and toward waiting (patience). By "sophisticated," we mean that such a model will be able to make more "human-like" decisions. We believe that because a model is "human-like," it will be possible to predict and depict a future of society, even an undesirable future.

4. Speech-recognition technology for understanding a human's internal state

With the advent of voice assistants, which allow users to operate devices by speaking into a smartphone or AI speaker, speech-recognition technology has spread rapidly around the world. The practical application of speech-recognition technology for operating computers by means of spoken dialogues using short-word exchanges began in the 1980s with the introduction of interactive voice-response systems. The introductions of car (satellite) navigation systems in the 1990s and deep-learning technology in recent years have greatly improved speech-recognition accuracy and spurred the rapid spread of voice assistants.

Natural communication between people involves longer sentences and more complex linguistic expressions, which are more difficult for speech recognition to handle, than the short sentences uttered to a voice assistant. Moreover, when the conversational partner is familiar, the tone of voice becomes more casual and utterances fragmented, which makes it more difficult to predict both sounds and linguistic expressions. Regardless of those difficulties, recognizing such conversational speech can be used for applications with significant business value, such as analyzing the content of conversations in call logs stored in large quantity at call centers and supporting conversations in real time; accordingly, various speech-recognition technologies based on deep learning are being actively studied. As shown in **Fig. 3**, speech-recognition technology has evolved through repeated improvements in the accuracy at which speech can be converted to text, expansion of its applications, and increasing complexity of the speech to be handled.

Information communicated through speech includes not only verbal information (text information) but also non-verbal information such as gender, age, emotions, intentions, and attitudes. As well as working on recognizing text information from speech with high accuracy, we are investigating technologies for recognizing and using non-verbal information and developing technologies that can extract speaker attributes (adult male, adult female, and child), emotions (joy, anger, sadness, and calmness), and questions and/or non-questions. We are also researching, developing, and practically applying technologies for estimating the degree of a customer's anger and/or satisfaction as well as the impression that the telephone operator's response gives to that customer during conversational speech consisting of two speakers (such as when a customer calls a call center).

The recognition technology for verbal and non-verbal information we are currently investigating is the first step toward reading the internal state of humans in a manner that allows us to offer more-advanced services. In everyday conversation, we use various clues, such as voice tone, line of sight, facial expression, pause length during interaction, and changes in wording, to estimate information inside a person's mind, such as their emotions and interests or friendliness and indifference, and we use that information to facilitate smooth communication. The recognition of such internal information using a machine is expected to lead to new services that target the subtleties of the human mind that cannot be expressed in words. Such services might conceivably include a voice-dialogue agent that is attuned to the emotions of the user, education that responds to the understanding and interests of each student, and detection of a patient's physical condition and mental stress from medical-interview dialogue data. It is also expected to contribute to creating a new type of communication—as targeted by the grand challenge "Mind-to-Mind Communications" set for Digital Twin Computing (DTC), which is one of the elements of IOWN (the Innovative Optical and Wireless Network) promoted by the NTT Group, that can directly understand the way people perceive and feel in a manner that transcends differences in individual characteristics such as experience and sensitivity.

5. Thought-enhancing stimulus design for drawing out and expanding a human thinking ability

Human thoughts and behaviors are not only influenced by information that is interpreted consciously, such as language and numerical values, but also by sensations and perceptions that are perceived unconsciously, such as colors, scents, and the atmosphere created by the manner of speaking. For example, smelling soap makes us want to wash [7] and changing the color of a product package makes us feel that the product is different even though it contains the same thing [8].

We are accumulating knowledge on the relationship between perceptual stimuli and human thought and behavior, and on the basis of that relationship, we are researching generation and control of perceptual stimuli with the aim of drawing out and expanding human thinking ability. In one of our previous studies, we investigated the effect of the speaker's speaking style (voice pitch, speaking speed, and inflection) on a human's psychological state and behavior [9]. In that study, we conducted a large-scale subjective-evaluation experiment on the relationship between speech that promotes products and people's consumption behavior and analyzed the relationship between speech and purchase intention by using the consumer-behavior model developed by Mehrabian and Russell [10] (**Fig. 4**) that expresses emotion as a mediator.

5.1 Experimental procedure

Via crowdsourcing, we asked 202 native speakers (male and female) of Japanese to listen to an advertisement for an electrical appliance spoken in different styles of speech. After listening, we asked them to respond to evaluation items about the emotion and purchase motivation they felt when listening to the spoken ad. We then analyzed the relationship among speech characteristics, emotions, and purchase intention on the basis of the obtained parameters (voice pitch, speaking speed, and intonation) related to the ad's speech (voice characteristics) and the responses to the evaluation items. The promotional text of the ad was based on speech uttered by a professional female speaker in a read-aloud tone, where the speech parameters were set as shown in **Table 1**. The evaluation items are listed in **Table 2**.

5.2 Results

The results of analyzing the relationship among speech characteristics, emotions, and purchase intention—by using a three-layer model based on Mehrabian and Russell's consumer-behavior model—are shown in **Fig. 5**. In line with previous studies [11], the results indicate that among the emotions, "pleasure" and "arousal" positively affect purchase intention. They also indicate that the effect of pleasure is the strongest, higher voice, faster speech speed, and higher intonation lead to higher emotion rating, and among those parameters, speech speed had the greatest

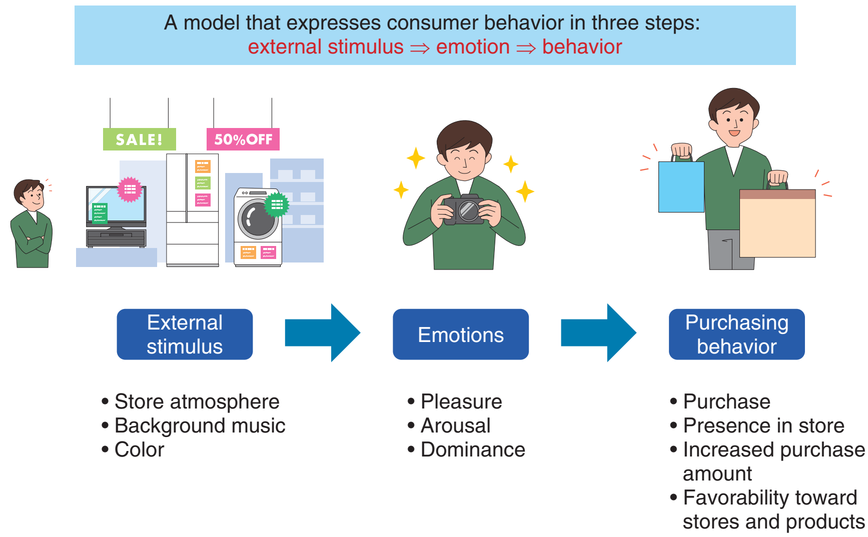


Fig. 4. Consumer-behavior model.

Table 1. Voice parameters.

Parameters of voice		Setting value
Voice pitch	Average F0 (Hz)	0.9439x, 1x, 1.059x
Speaking speed	Speaking speed (mora/s)	0.890x, 1x, 1.1225x
Magnitude of intonation	F0 variance (Hz)	0.6667x, 1x, 1.5x

Table 2. Evaluation items.

Evaluation item		Setting value
Emotion	Pleasure	Displeasure ⇔ Pleasure (7 stages)
	Arousal	Sleep ⇔ Arousal (7 stages)
	Dominance	Obedient ⇔ Dominant (7 levels)
Purchase intention		“I want to buy it very much.” ⇔ “I don’t want to buy it at all.” (7 levels)

effect on the evaluation items.

To summarize these results, we found that (i) the consumer-behavior model is valid for a speech stimulus based on read-out advertisements and (ii) the speed of speech has the strongest effect on pleasure, which has the strongest effect on purchase intention.

The above-described efforts represent the analysis of only a small part of the complex and diverse thinking processes of humans. We plan to study a wider range of factors, such as the relationship among the attributes of listeners with an auditory stimulus, and investigate non-auditory stimuli such as sight and smell. We also plan to study the effects of stimuli on human thinking phenomena—ranging from intuitive judgments to deeper stages such as effects on logical thinking and human values. Through these studies, we want to clarify what is necessary to draw out human thinking ability, enhance human potential, and contribute to the creation of a better society.

6. Concluding remarks

By refining the four technologies described in this article, we plan to understand, reproduce, and extend human thinking ability, which consists of human information understanding, human information

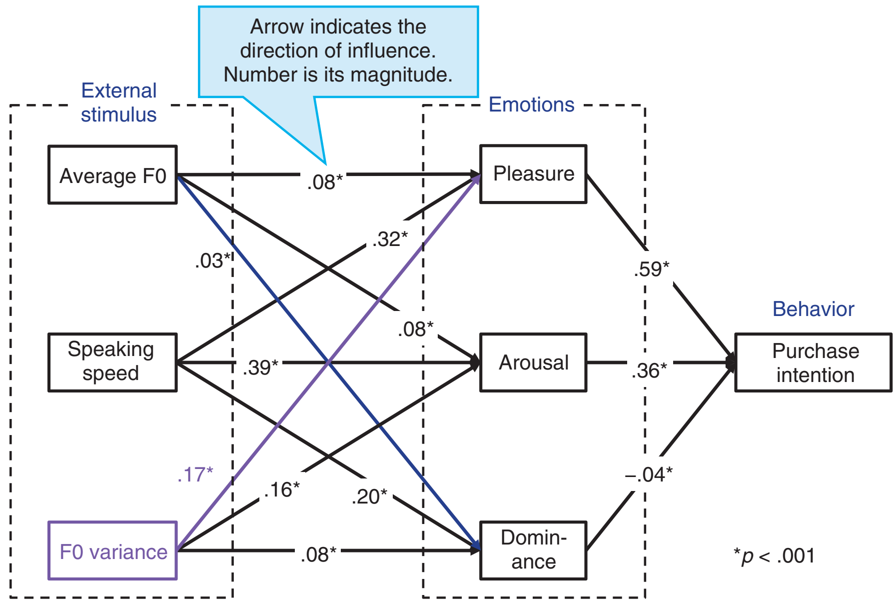


Fig. 5. Results of analysis of relationships among voice characteristics, emotions, and purchase intention.

processing, and response and behavior.

References

[1] K. Nishida, I. Saito, K. Nishida, K. Shinoda, A. Otsuka, H. Asano, and J. Tomita, "Multi-style Generative Reading Comprehension," Proc. of the 57th Annual Meeting of the Association for Computational Linguistics (ACL 2019), pp. 2273–2284, Florence, Italy, July 2019.

[2] K. Nishida, K. Nishida, M. Nagata, A. Otsuka, I. Saito, H. Asano, and J. Tomita, "Answering while Summarizing: Multi-task Learning for Multi-hop QA with Evidence Extraction," Proc. of ACL 2019, pp. 2335–2345, Florence, Italy, July 2019.

[3] R. Tanaka, K. Nishida, and S. Yoshida, "VisualMRC: Machine Reading Comprehension on Document Images," Proc. of the 35th AAAI Conference on Artificial Intelligence (AAAI 2021), pp. 13878–13888, Feb. 2021.

[4] R. Wang, F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev, and A. T. Campbell, "StudentLife: Assessing Mental Health, Academic Performance and Behavioral Trends of College Students Using Smartphones," Proc. of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2014), pp. 3–14, Seattle, USA, Sept. 2014.

[5] T. Tominaga, S. Yamamoto, T. Kurashima, and H. Toda, "Effects of Personal Characteristics on Temporal Response Patterns in Ecological Momentary Assessments," Proc. of the 18th IFIP TC13 International Conference on Human-Computer Interaction (INTERACT 2021), Bari, Italy, Aug./Sept. 2021.

[6] T. Kurashima, T. Althoff, and J. Leskovec, "Modeling Interdependent and Periodic Real-world Action Sequences," Proc. of the 2018 World Wide Web Conference, Lyon, France, pp. 803–812, Apr. 2018.

[7] K. Liljenquist, C. B. Zhong, and A. D. Galinsky, "The Smell of Virtue: Clean Scents Promote Reciprocity and Charity," Psychological Science, Vol. 21, No. 3, pp. 381–383, Mar. 2010.

[8] K. Maki, "Color Studies for Color Design," Ohmsha, 2006 (in Japanese).

[9] M. Nagano, Y. Ijima, and S. Hiroya, "Impact of Emotional State on Estimation of Purchase Intention from Advertising Speech," Proc. of INTERSPEECH 2021, Brno, Czech Republic, Aug./Sept. 2021.

[10] A. Mehrabian and J. A. Russell, "Approach to Environmental Psychology," The MIT Press, 1974.

[11] R. Donovan and J. Rossiter, "Store Atmosphere: An Environmental Psychology Approach," Journal of Retailing, Vol. 58, No. 1, pp. 34–57, 1982.

Kyosuke Nishida

Distinguished Researcher, Human Insight Laboratory, NTT Human Informatics Laboratories.

He received a B.E., M.I.S., and Ph.D. in information science and technology from Hokkaido University in 2004, 2006, and 2008 and joined NTT in 2009. He received the Yamashita SIG Research Award from the Information Processing Society of Japan (IPSJ) in 2015 and the Kambayashi Young Researcher Award from the Database Society of Japan (DBSJ) in 2017. His current interests include natural language processing and artificial intelligence. He is a member of the Association for Computing Machinery (ACM), IPSJ ,the Association for Natural Language Processing (NLP), the Institute of Electronics, Information and Communication Engineers (IEICE), and DBSJ.

Takeshi Kurashima

Distinguished Researcher, Human Insight Laboratory, NTT Human Informatics Laboratories.

He received a B.S. from Doshisha University, Kyoto, in 2004 and M.S. and Ph.D. in informatics from Kyoto University in 2006 and 2014. He joined NTT in 2006. He was a visiting scholar at Stanford University, USA from 2016 to 2017. His current research interests include data mining, machine learning, and recommender systems. He is a member of IPSJ, IEICE, DBSJ, and ACM.

Noboru Miyazaki

Senior Research Engineer, Supervisor, Human Insight Laboratory, NTT Human Informatics Laboratories.

He received a B.E. and M.E. from Tokyo Institute of Technology in 1995 and 1997. Since joining NTT in 1997, he has been engaged in research on spoken dialogue systems, speech synthesis, and speech recognition technologies. He received the IEICE Inose Award in 2001. He is a member of the Acoustical Society of Japan, IEICE, and the Japanese Society for Artificial Intelligence (JSAI).

Hiroyuki Toda

Senior Research Engineer, Supervisor, Human Insight Laboratory, NTT Human Informatics Laboratories.

He received a B.E and M.E. in materials science from Nagoya University in 1997 and 1999, and Ph.D. in computer science from University of Tsukuba in 2007. He joined NTT in 1999. His current research interests include information retrieval and data mining. He is a member of IPSJ, IEICE, JSAI, DBSJ, and ACM.

Shuichi Nishioka

Executive Reserch Engineer, Supervisor, Human Insight Laboratory, NTT Human Informatics Laboratories.

He received a B.E. and Ph.D. in engineering from Yokohama National University, Kanagawa, in 1995 and 2005. He joined NTT in 1995. His current interests include data engineering and artificial intelligence. He is a member of IPSJ and DBSJ.

Research on Symbiotic Intelligence for Achieving Symbiosis among People and between People and Machines

Tetsuya Kinebuchi, Taichi Asami, Sen Yoshida, and Ryuji Yamamoto

Abstract

With the rapidly progressing fusion of the real and cyber worlds, the manner of symbiosis among people and between people and machines is changing significantly. Information and communication technologies will make it possible to gently guide the behavior of people and groups toward a better future and shared emotions and to enable people and machines to work together through more-natural and complex interactions. An overview of the research being conducted by the Symbiotic Intelligence Research Project toward this future vision is given in this article.

Keywords: behavioral modeling, knowledge processing, emotional perception control

1. Introduction

As the real and cyber worlds rapidly converge, the manner in which people work and live is changing dramatically. Meetings and collaborative work that used to be done face-to-face in real spaces, such as classes at schools and social gatherings, are now done remotely, and social networking services (SNSs) are becoming the mainstream for chatting and gossip. Sporting events and theatrical performances, which used to require a direct trip to a stadium or theater, can now be richly experienced in cyberspace. The novel-coronavirus (COVID-19) pandemic has sped up the merging of the two (real and cyberspace) worlds.

We believe that by using the power of information and communication technologies, all people, and even people and machines, will be able to coexist more prosperously than ever before, and that this coexistence will enable people to experience even greater well-being than before. For example, we will be able to gently guide people and groups toward a better future and shared emotions or work together with machines that have a similar inner self to that of people through more natural and complex interactions. With this vision of the future in mind, we are researching and developing technologies that will contribute to (i) modeling and simulating human behavior, especially the behavior (thoughts, actions, etc.) of groups, and (ii) predicting and optimizing future social activities in which people as well as people and machines collaborate and live together. Specifically, this research and development (R&D) is based on the following three themes.

1) Group psychological and behavioral modeling: Modeling of psychological, judgmental, and social behaviors of people and their groups and simulation and optimization of those behaviors on the basis of the models.
2) Emergence and social brain: Knowledge processing that (i) accurately understands and converts the textual content into knowledge and

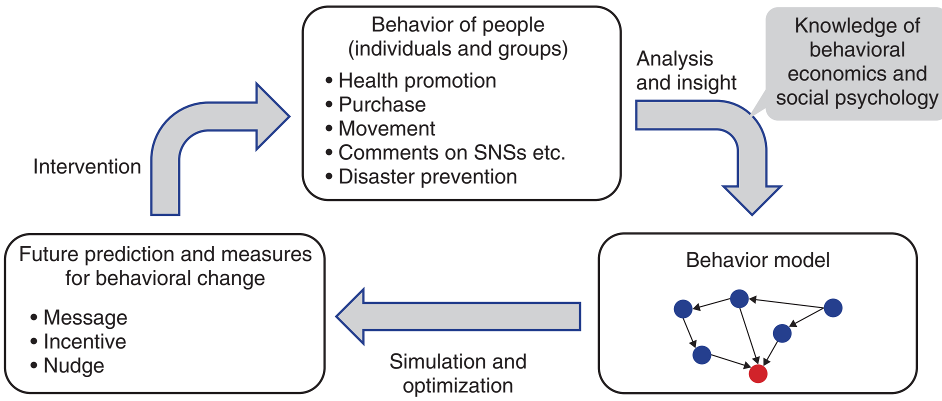


Fig. 1. Research initiatives on group psychology and behavior modeling.

(ii) provides the information required for collaboration with people in a natural exchange.

3) Emotional perception control: Explanation and sensing of an internal model of human emotional psychology (enthusiasm, sense of unity, etc.), and collaborative-work support to resolve communication discrepancies by empathizing with individual interpersonal characteristics.

Hereafter, our efforts concerning these three themes are introduced.

2. Group psychological and behavioral modeling

We are conducting R&D on the modeling of human decision-making and behavior. People make various daily decisions, such as what to eat for dinner, what to buy, whether to exercise, where to go for pleasure, and what to post on SNSs, that dictate their social activities. To predict the future of social activities and engender ever greater well-being, modeling and computer simulation of human decision-making and behavior are indispensable technologies. For example, if we could model behavior and simulate how that behavior changes with interventions such as messages and incentives, we could create apps and services that enable users to naturally make healthier, safer, or more enjoyable choices. We are analyzing human behavior, modeling and simulating that behavior, and deriving measures to change behavior on the basis of the models (**Fig. 1**).

Application of various mathematical-modeling and big-data-based optimization techniques, including deep learning, is a promising approach to model and simulate behavior. However, it is not easy to collect a large amount of data on the behavior of a wide variety of people. We believe that to elucidate the mechanisms of human decision-making and behavior, a more in-depth approach than simply applying current mathematical models is necessary.

Human behavior has been studied in the fields of behavioral economics and social psychology, and a great deal of knowledge has been accumulated. At NTT Human Informatics Laboratories, by constructing new mathematical models that incorporate this knowledge and verifying hypotheses, we aim to clarify the mechanism of human decision-making and behavior and implement efficient and highly accurate computer simulation.

3. Emergence and social brain

We are researching and developing knowledge processing that (i) accurately understands and converts the textual content into knowledge and (ii) provides the information required for collaboration with people in a natural exchange. Language modeling is a fundamental technology for accurately understanding and converting textual content into knowledge. Regarding Japanese, a language model is a representation of “the essence of the Japanese language” in a form that is easy for a computer to handle. Language modeling has been dominated by methods for creating deep-learning models through large-scale pre-training from huge amounts of text data.

We are also constructing large-scale pre-training language models for Japanese and conducting R&D to further improve their performance. Through these efforts, we aim to build a foundation for linguistic communication to enable collaboration among people as well as between people and machines while

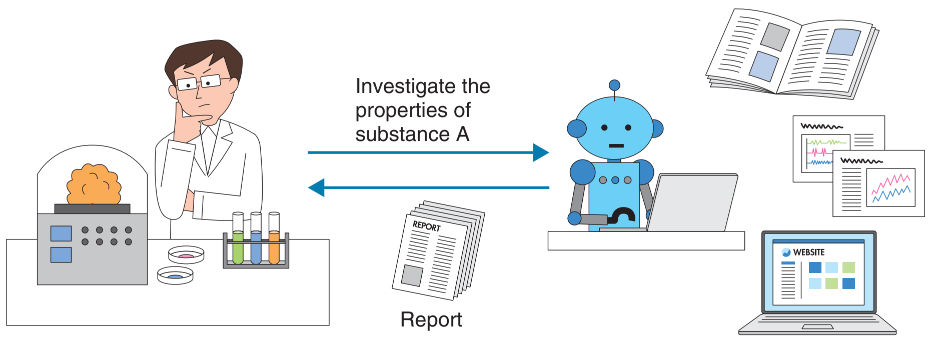


Fig. 2. Collaboration between people and machines (automatic generation of reports).

contributing to various language-processing-related services of the NTT Group.

One of the most-promising use cases for collaboration between people and machines in regard to textual content is automatic document summarization. We are developing a technology that enables highly accurate summarization using large-scale pre-training language models. This technology can, for example, generate a summary that includes keywords specified by the user and allows the length of the summary to be adjusted in accordance with the application. We believe that if we can further develop this technology to automatically generate reports concerning tasks such as examining a large amount of literature (medical care and intellectual property), it would generate great value (**Fig. 2**).

Collaboration would be more natural if people could talk to machines in the same language as they talk to other people. Technologies for interacting with users in natural language are called dialogue systems and can be broadly classified as task-oriented (which perform specific tasks) and systems that conduct so-called "chat." We are researching and developing chat-oriented dialogue systems; in particular, we are developing technology to add character to such systems so that users feel familiar with them.

4. Emotional perception control

We are researching and developing technology for improving the quality of symbiosis as a group by (i) understanding the characteristics of people from perception to emotion and (ii) intervening in communication between and among people in accordance with these characteristics. Specifically, we are developing the following two technologies.

1) Emotional-perception-control technology: Technology that expands communication (e.g., fostering a sense of unity and enthusiasm through sharing of emotions) by understanding the characteristics of human emotional expressions, estimating emotional characteristics through sensing and data analysis, and controlling emotions with perceptual stimuli tailored to emotional characteristics.
2) Collaborative-work-support technology: Technology for improving the quality of collaborative work within a group by understanding the communication characteristics of people and eliminating communication discrepancies by converting information tailored to those characteristics in a manner that induces behaviors that facilitate communication.

Regarding emotional perception control, we are developing emotion-estimation technology for estimating the emotions of individuals and groups (groups, crowds, etc.) by using biometric data, images, sounds, content, etc. and emotion-control technology for changing such emotions to a desirable state in accordance with the estimation results (**Fig. 3**).

Regarding collaborative-work support, which focuses on psychological safety as an important factor in improving the quality of collaborative work in groups, the aim is to make it easier for people with diverse characteristics to work together by enhancing their psychological safety.

For this purpose, we are developing a technology that can improve psychological safety, namely, reduce communication discrepancies, by understanding the communication characteristics of individuals and converting verbal and non-verbal information in accordance with those characteristics (**Fig. 4**).

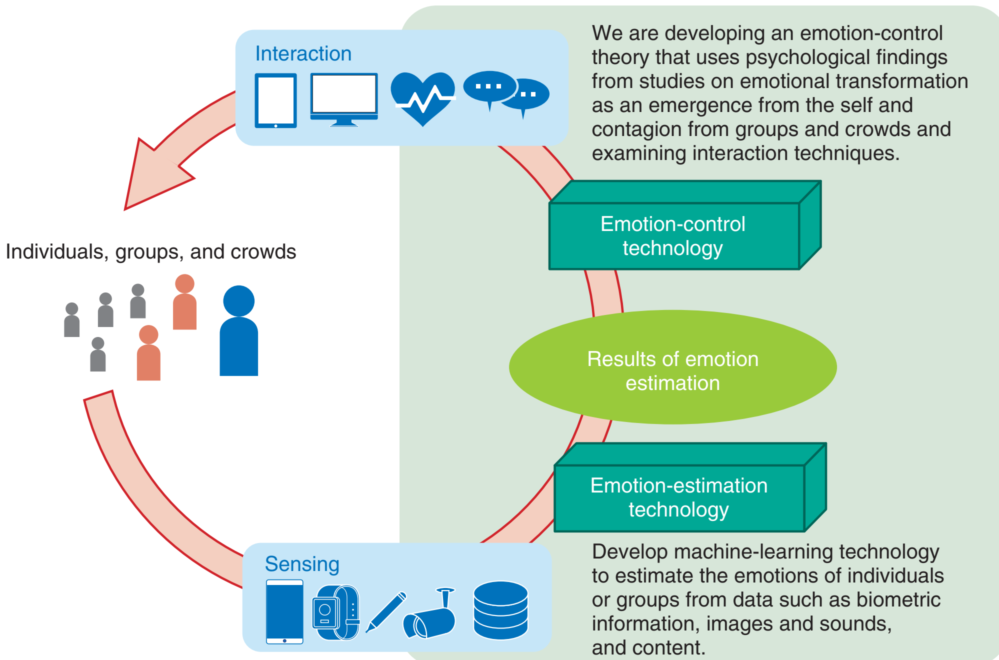


Fig. 3. Overview of technology for controlling emotional perception.

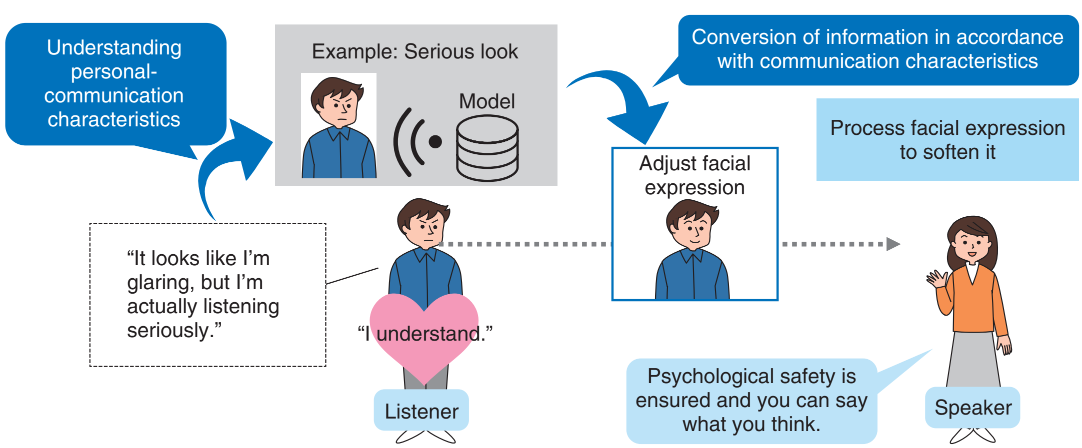
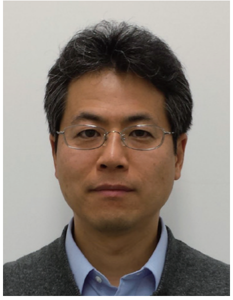


Fig. 4. Overview of technology for supporting collaborative work.



Tetsuya Kinebuchi
Senior Manager, Symbiotic Intelligence Laboratory, NTT Human Informatics Laboratories.
He received an M.S. in physics from Tohoku University, Miyagi, in 1997 and joined NTT Corporation the same year. His research interests include media processing and artificial intelligence (AI). He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) of Japan.

Sen Yoshida
Senior Research Engineer, Supervisor of Symbiotic Intelligence Laboratory, NTT Human Informatics Laboratories.
He received a B.E. and M. Info. Sci. from Tohoku University, Miyagi, in 1993 and 1995. He joined NTT Corporation in 1995. His research interests include natural language processing and AI. He is a member of IPSJ and the Japanese Society for Artificial Intelligence (JSAI).

Taichi Asami
Senior Research Engineer, Symbiotic Intelligence Laboratory, NTT Human Informatics Laboratories.
He received a B.E., M.E., and Ph.D. in engineering from Tokyo Institute of Technology in 2004, 2006, and 2019. He joined NTT Corporation in 2006. His research interests include media processing, AI, and behavioral modeling. He is a member of the Institute of Electrical and Electronics Engineers (IEEE), IEICE, the Information Processing Society of Japan (IPSJ), and the Acoustical Society of Japan (ASJ).

Ryuji Yamamoto
Senior Research Engineer, Supervisor of Symbiotic Intelligence Laboratory, NTT Human Informatics Laboratories.
He received an M.E. from Kyushu Institute of Technology, Fukuoka, in 1998. He joined NTT Corporation in 1998. His research interests include media processing and human-computer interaction.

Cybernetics Technologies for Symbiosis between People and Machines

Yushi Aono, Hitoshi Seshimo, Narimune Matsumura, Yukio Koike, and Masaaki Matsumura

Abstract

With the development of computing technology and artificial intelligence, an environment in which we can use advanced information and communication services in our daily lives is rapidly being established. By installing mechanisms that recognize and understand in a more human-oriented manner than conventional technologies, we aim to create a society in which people and machines can coexist in harmony. In this article, we introduce our research themes for achieving this aim.

Keywords: symbiosis between people and machines, motion support, embodied knowledge

1. Toward symbiosis between people and machines

As the computational efficiency of processers has increased significantly, the accuracy of image processing and recognition has improved; technology that handles cross-modal information, such as various sensors, has been refined; and automated driving of cars, which until now has only been depicted in science fiction, is gradually being achieved. The improvement in voice recognition, character recognition, and context understanding makes it possible for automatic-response engines, such as chatbots, to support counter services (contact persons), and people now carry around mobile devices (such as smartphones) with high-performance processors. These factors have led to rapid establishment of an environment in which we can use advanced information and communication technology (ICT) services in our daily lives. One of the technologies that contributes significantly to improving the accuracy of recognition and understanding is artificial intelligence (AI), which is constructed by machine learning using large amounts of real data.

Current mainstream AI technology is akin to modeling logical and objective recognition and understanding that works like the human left brain based on our experience and memory, and it has enriched our lives mainly by targeting things that can be expressed in words and those that can be symbolized such as images and sensors data.

We aim to create a society that enables people and machines to coexist in harmony (**Fig. 1**) by updating conventional AI technology with the following two steps:

(1) Elucidating the processing and action processes that work like the human right brain for things that cannot be expressed in symbols (such as sensations and actions) and for those that are subjective and cannot be uniquely expressed (such as judgments that differ from person to person)

(2) Installing the elucidated mechanisms that provide more human-oriented recognition and understanding into conventional AI technology

To achieve this aim, we introduce the following three technologies from our research themes:

- **Motion support** that supports motion by sensing human biological signals with high accuracy

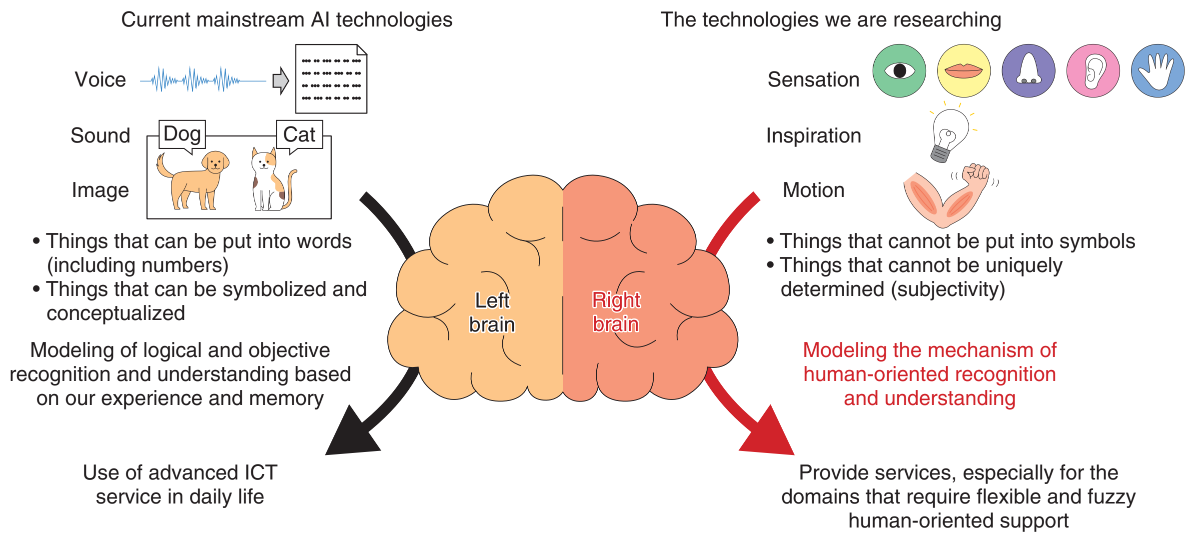


Fig. 1. Technologies we are researching.

- **Embodied knowledge** that supports the acquisition and demonstration of non-verbal and empirical skills
- **Zero latency** that controls the perceived time when a person operates a remote-controlled robot

2. Motion support

Cybernetics, proposed by Norbert Wiener in 1948, is a unified approach to communication and control concerning living organisms and machines. Paying renewed attention to this concept, we are developing technologies to support the motion of people.

Human motion is executed when motor instructions from the brain are transmitted to the muscles, which then contract on receiving the instructions. The brain then perceives and recognizes the result as a stimulus through sensory organs and repeats planning a new motion and sending its appropriate instruction. We are attempting to support this communication and control in the human body by sensing and feeding back biological signals such as electroencephalography (EEG) and electromyography (EMG).

More specifically, we aim to create a society in which people can move (live) better by reproducing (transcribing) their current motor skills temporarily and permanently on the basis of the motions of others (skilled people) and their former selves. For example, we are considering (i) training that uses electrical muscle stimulation (EMS) to reproduce the muscle motions of skilled people executing various skills and sports and (ii) rehabilitation based on records of one's motions in daily activities.

We are working on enabling the following four functions, which we consider necessary for reproducing (transcribing) human motor skills by using biological signals (see **Fig. 2**).

- Measuring biological signals and constructing kinematic models: Bio-signals (EEG, EMG, etc.) of the transcription source are sensed and a kinematic model representing the relationship between the bio-signals and the motion is constructed.
- Generating and providing sensory stimulation: Sensory stimuli (EMS, vibration, exoskeleton, etc.) that control the motion of the transferee are generated and expressed on the basis of the generated kinematic model.
- Estimating dynamics: The motion results and biological signals of the transferee (which was stimulated and moved) are sensed, and states such as posture and muscle contraction are estimated.
- Adjusting sensory stimulation parameters: The parameters of sensory stimuli are adjusted to attain better motor results and natural motor experiences on the basis of the estimation results concerning the motion state and kinematic model. New feedback based on the results is then provided to the transferee.

We had been working on bio-sensing technology

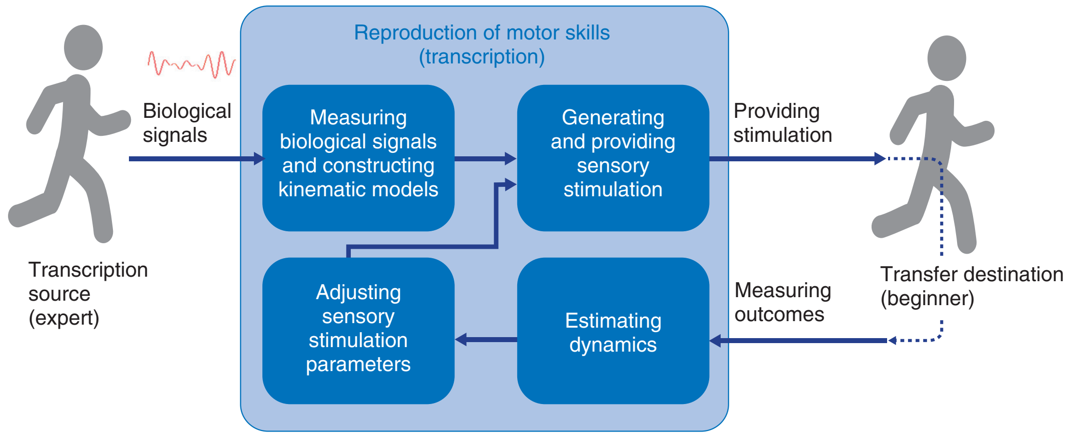


Fig. 2. Efforts to reproduce (transcribe) human-motion ability by using biological signals.

(using, for example, EMG) and feedback technology using tactile and EMS to the human body. Mechanisms of human motion, however, are complex, and various efforts are underway in the fields of physiology, brain science, and biomedical engineering to support and enhance motion, so an interdisciplinary approach is necessary. Looking to the future and using our accumulated technological strengths, we are planning to acquire new knowledge about the mechanisms of motion through creation of technology and interdisciplinary collaboration with internal and external experts. For a society in which people and machines coexist at a higher level, we will elucidate human processing and action processes related to motion and expand human capabilities.

3. Embodied knowledge

In human life and social activities, knowledge is the basis for improving and refining the quality and efficiency of people's behavior and thinking. It is therefore important for people living in the present to acquire knowledge that has been accumulated over a long period and apply it in various situations in their daily lives.

The first thing that comes to mind when people think of acquiring knowledge is studying at school or learning a subject. Such already systematized and described knowledge is called *explicit knowledge*, and everyone has experienced the widely practiced methods of acquiring and demonstrating such knowledge. However, knowledge does not mean only the explicit type. Another type is the knowledge that people learn and express that is difficult to systematize or describe, and it can only be acquired through experience rather than through mental understanding. This type is what we call *embodied knowledge*.

Regarding learning to play a musical instrument, people generally refer to instructional books that systematically describe how to handle the instrument, its functions, and how to move the fingers. Therefore, if you read all of this instructional material and understand it, will you be able to play an instrument? Of course, the answer is "no." In other words, only by moving your body (fingers, in this case) and practicing with that understanding in mind will you be able to play. This reality indicates the existence of embodied knowledge, which is the knowledge rooted in a physical activity, as opposed to explicit knowledge expressed in textbooks (**Fig. 3**).

Embodied knowledge is largely based on the senses and subjectivity of people, and it is far from being supported by ICT for acquiring and demonstrating in the manner of explicit knowledge, which is objective and easily verbalized. However, with the recent evolution of technology, it has become possible to sense and process information at high speed in a manner closer to the sensations received by the people involved. Accordingly, we are conducting research on embodied knowledge from the perspective of neo-cybernetics, which captures the mechanism of actions from the starting point of the person's cognition. Taking windsurfing as a subject, we are clarifying the mechanism of acquiring and demonstrating embodied knowledge in sports and making it possible to support this mechanism. With this in mind, we are studying an analysis method to find the latent pattern between the physical activity of the athlete, behavior

Knowledge acquired through learning

Explicit knowledge	Embodied knowledge
• Linguistic knowledge • Understanding of systematic and sequential teaching	• Non-verbal skills • Sensory and experiential (training) acquisition

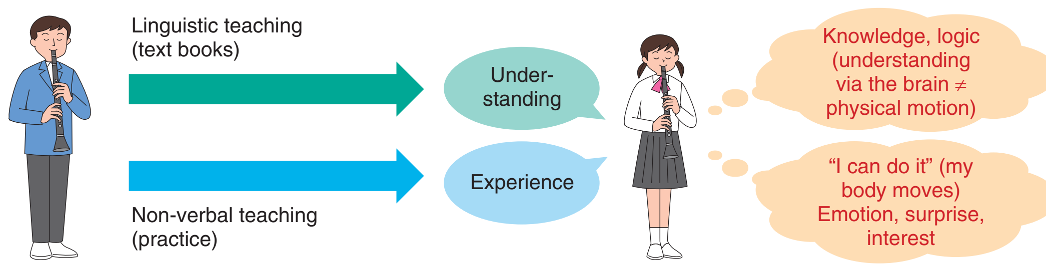


Fig. 3. What is embodied knowledge?

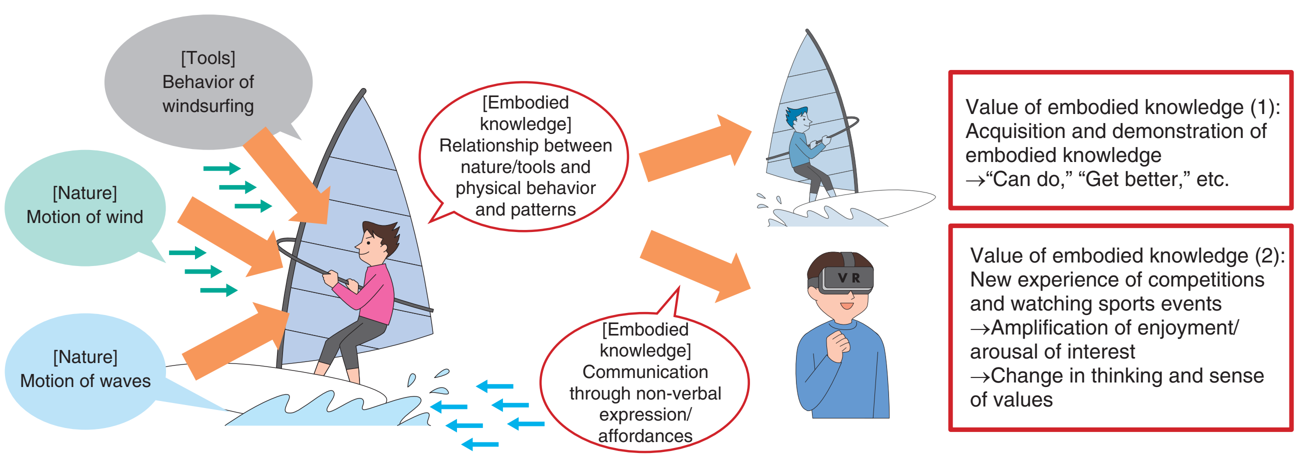


Fig. 4. Embodied knowledge concerning windsurfing.

of the natural environment and tools, and performance of the athlete. Moreover, using the data acquired from actual competitors, we are also studying how to communicate the discovered pattern to other people and how to improve the value (acquisition and demonstration of embodied knowledge or changes in interest in the sport and the natural environment) provided to the recipients of the communication (**Fig. 4**).

Although sports is typical in the study of embodied knowledge, the aforementioned skills, which are knowledge acquired and demonstrated in strong connection with physical actions, such as playing a musical instrument, drawing a picture, or making a good presentation, exist in a variety of everyday situations. Accordingly, we are aiming to create a society in which anyone can acquire and apply embodied knowledge, which has been confined within certain people, through ICT based on the perspective of the person who wants to acquire that knowledge while simultaneously studying areas other than sports.

4. Zero latency

Physical work by remotely operated robots is promising for solving the labor shortage in the service industry, such as nursing and caregiving, and for ensuring safety in dangerous work, such as at medical institutions and disaster sites, during a pandemic. However, such remote operation always involves a

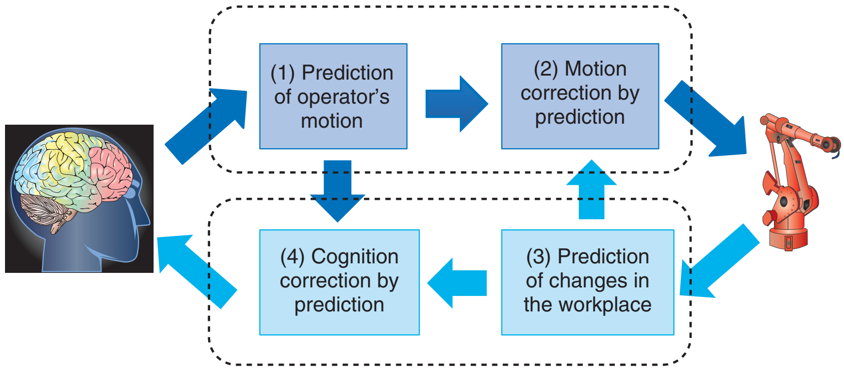


Fig. 5. Four technologies for achieving zero latency.

delay between the operator and robot, and that delay is known to have adverse effects such as work errors, reduced work efficiency, and operator fatigue. It can thus be said that while the sophistication of remotely operated robots has increased their practicality, the effects of delays are becoming impossible to ignore. The delay is caused by not only the communication but also the operation of the camera, display servo, etc., and the control of these systems. Even if a low-latency system is used and remotely controlled over a very short distance, a round-trip delay (namely, the sum of the delay between the incidence of light at the camera and screen display to the operator and the delay between the operator's operation and actuator's operation) of 100 ms or more occurs.

We therefore aim to develop a remote-operation system that can fully demonstrate the capabilities of the operator by eliminating the negative effects of delays through four technologies: prediction of the operator's motions, prediction of physical changes in the work area, correction of the robot's motions, and correction of the operator's perception (**Fig. 5**).

One example of predicting the operator's motion and correcting the robot's motion is predicting the motion of the operator several hundred milliseconds in the future and moving the robot before the operator performs the motion in accordance with the prediction. Therefore, the effect of delay is reduced (**Fig. 6**). However, it is possible that the robot's motion may be wrong (namely, unintended by the operator), so we are currently working on improving prediction accuracy by estimating the operator's intentions on the basis of the content of the work and motion of objects.

As an example of predicting physical changes in the work area and correcting the robot's motion, we

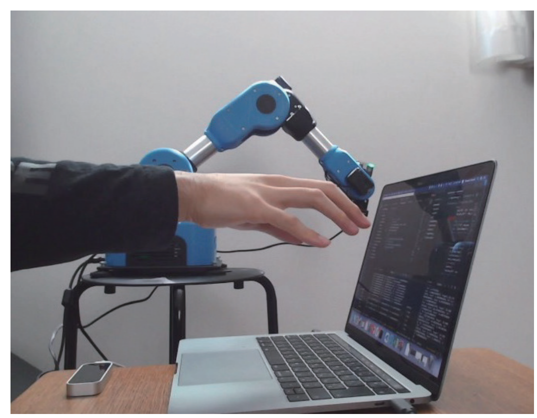


Fig. 6. Experimental system for predicting the operator's motion and correcting the robot's motion.

simulated a robot's physical structure to determine how it moves according to the operator's motions. Presenting the simulation results to the operator through augmented reality (AR) and other means allows the operator to recognize the results of the robot's motions before the robot actually moves, which reduces the effects of delays (**Fig. 7**). Since the operator will be working on the basis of AR-based predictive information, in contrast to direct work, we are currently studying how to present information to support the operator more effectively during various tasks.

5. Future developments

We are currently refining these technologies on an individual basis. In the future, we will unify these technologies to provide services especially for

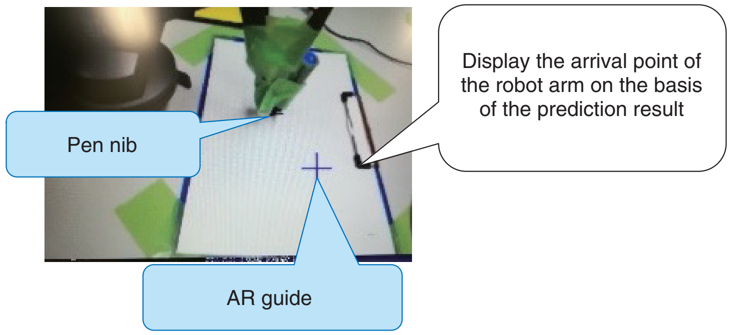


Fig. 7. Experimental system for cognitive correction using AR.

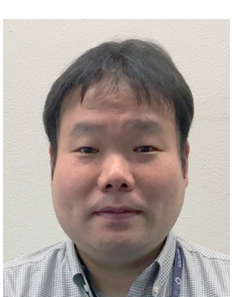
domains that require flexible and fuzzy human-oriented support.

Yushi Aono
Executive Research Engineer, Supervisor, Laboratory Manager, Cybernetics Laboratory, NTT Human Informatics Laboratories.
He received a B.E. in biotechnology and M.E. and Ph.D. in control engineering from Osaka University in 1994, 1996, and 1999. Since joining NTT in 1999, he has been engaged in research on speech synthesis, speech recognition, and spoken dialogue systems. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) and the Information Processing Society of Japan.

Yukio Koike
Senior Research Engineer, Cybernetics Laboratory, NTT Human Informatics Laboratories.
He received an M.S. in physics from Nagoya University in 2004 and joined NTT the same year. His research interests include network technology for Internet-of-Things services and biological signal processing. He is a member of IEICE.

Hitoshi Seshimo
Senior Research Engineer, Supervisor, Cybernetics Laboratory, NTT Human Informatics Laboratories.
He received a B.E. and M.E. in mechanical engineering from Waseda University, Tokyo, in 1995 and 1997 and joined NTT in 1997. His research interests include computer aided instruction, web-based learning, content distribution and navigation systems, and geographical information services.

Masaaki Matsumura
Senior Research Engineer, Cybernetics Laboratory, NTT Human Informatics Laboratories.
He received a B.E in design from Kyushu Institute of Design, Fukuoka, and M.E. in design from Kyushu University, Fukuoka, in 2005 and 2007 and joined NTT in 2007. His research interests include computer graphics, computer science, image/video coding, image processing, computer vision, data science, and machine learning.



Narimune Matsumura
Senior Research Engineer, Cybernetics Laboratory, NTT Human Informatics Laboratories.
He received a B.E. and M.E. in information science from Kyushu Institute of Technology, Fukuoka, in 2002 and 2004 and joined NTT in 2004. His research interests include human-machine interaction, robotics, and digital transformation of nursing and care.

Elemental Technologies for Supporting the Integration of Real Space and Cyberspace

Jun Shimamura, Shingo Ando, and Ryuichi Tanida

Abstract

To create a human-centered society in which everyone can lead a comfortable, vibrant, and high-quality life, we hope that a system that integrates real space and cyberspace will become a reality. In this article, a mobile mapping system (MMS) point cloud data processing technology called geoNebula™ is introduced, which analyzes data measured in real space, recognizes space and objects, and compresses the data into a compact form suitable for constructing cyberspace. This technology is being researched and developed to establish a mechanism for creating value through the sophisticated blending of real space and cyberspace.

Keywords: cyberspace, real space, MMS point cloud data processing technology

1. Building cyberspace

With the development of information and communication technologies, it is becoming possible to collect and analyze vast amounts of data from the Internet of Things. In response, the Japanese government and various companies are pursuing research and development to create a system that integrates real space and cyberspace in the manner advocated in Society 5.0. NTT has started research and development of the 4D digital platform™, which can integrate four-dimensional (4D) information (latitude, longitude, altitude, and time) as precisely as possible in real time and make future predictions from those data to implement the above system in society.

The 4D digital platform integrates sensing data about people and objects on the Advanced Geospatial Information Database with high accuracy and containing rich semantic information in a manner that enables advanced prediction, analysis, simulation, and optimization in various industrial fields (**Fig. 1**). We believe that this platform has many possible uses in areas such as alleviation of road traffic, optimal use of urban assets, and maintenance and management of social infrastructure. To develop the Advanced Geospatial Information Database, it is necessary to convert real space into data by using technology such as laser-imaging detection and ranging (LiDAR) and construct cyberspace with those data. This process is generally composed of the following three steps.

Step 1 (data collection): Real space measured using LiDAR, etc. is converted into data.

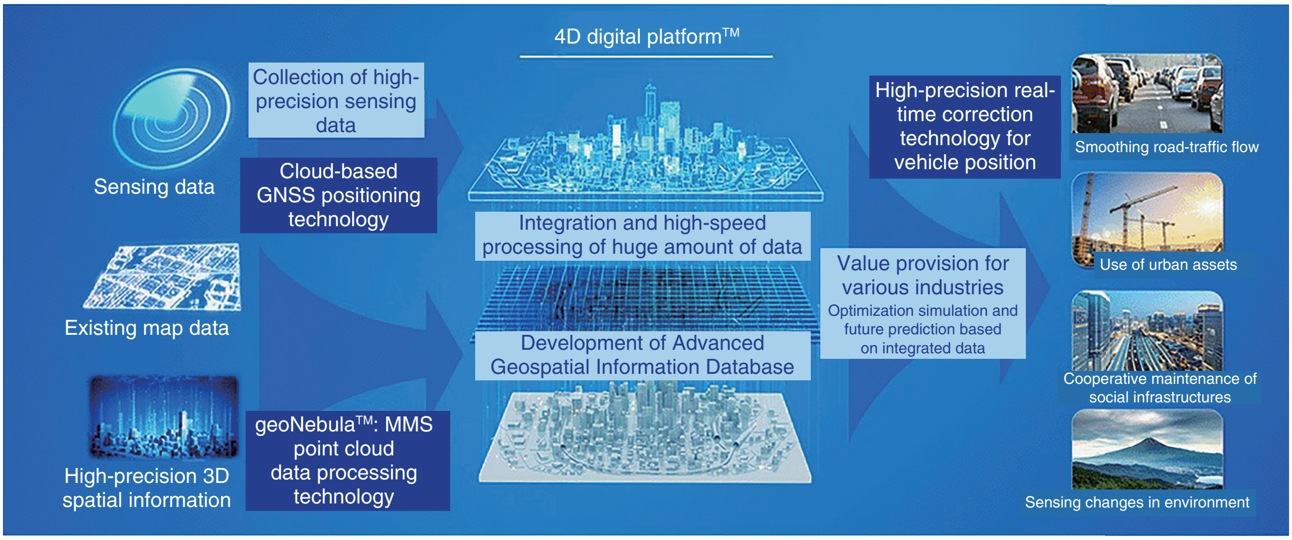
Step 2 (space and object recognition): Recognize space and objects from the measurement data and obtain their exact positions.

Step 3 (data storage): Store the finished data and original data for updating or other uses.

The latest research results concerning geoNebula™—a mobile mapping system (MMS) point cloud data processing technology we are developing to save labor and improve the efficiency of each step—are introduced hereafter.

2. Real-space structuring technology that enables labor saving and efficiency improvement of data collection and space/object recognition

With our research and development, we aim to establish the following two technologies for structuring real space from 3D point clouds measured using



GNSS: global navigation satellite system

Fig. 1. Overview of 4D digital platform™.

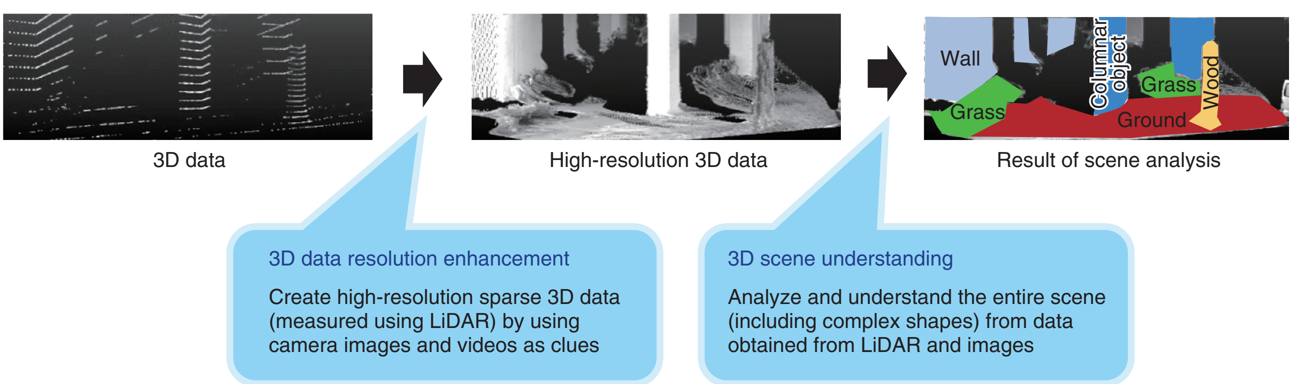


Fig. 2. Real-space structuring technology.

LiDAR, etc. (**Fig. 2**).

(1) 3D data resolution enhancement

(2) 3D scene understanding

3D data resolution enhancement is positioned as step 1 (data collection) in the three steps for constructing cyberspace listed at the end of the previous section. Due to the structure of a 3D point cloud, vacant spaces (gaps) of a certain width are created between points that are spatially adjacent to each other. This technology (3D data resolution enhancement) can acquire higher-definition 3D point clouds by generating multiple new points to fill these gaps and interpolating them [1]. 3D scene understanding is positioned as step 2 (space and object recognition). It is capable of automatically assigning semantic labels (road surface, building, etc.) to all points in a dense point cloud on a large scale and with high accuracy on a point-by-point basis [2]. These two technologies are explained in detail as follows.

2.1 3D data resolution enhancement

Less expensive LiDAR equipment tends to produce larger gaps between points. In such a case, much of the 3D-shape information of an object is lost, and the object's detailed features are obscured. The resolution of camera images has significantly increased, so it is now possible to acquire relatively dense 2D data even with inexpensive cameras. Therefore, if the information from the camera images taken separately is used appropriately, it would be possible to compensate for a sparse 3D point cloud. Although several technologies for this purpose are on the market, it is necessary to strictly align (calibrate) the LiDAR point cloud with the camera image as a prerequisite. Aligning the two sensors (LiDAR and camera) is generally difficult and requires a high level of skill. We therefore proposed a method of generating a stable and dense point cloud even with a relatively simple rough calibration. What sets our method apart from other methods is not only installation cost (low cost of equipment) but also operation cost (simple calibration) are accounted for.

2.2 3D scene understanding

Several deep neural networks (DNNs) that can be applied to non-grid data such as 3D point clouds have been proposed. 3D scene understanding is one such DNN for non-grid data, and we proposed using it to solve a task called semantic segmentation [2]. Semantic segmentation of a 3D point cloud is the process of separating all points in the point cloud into multiple segments by assigning meaningful labels to each point. By pre-training with a DNN, it is possible to estimate the label that should be assigned to each point in an unknown 3D point cloud. However, most conventional methods target point cloud data within a limited space such as indoors. Accordingly, to handle a wide area such as an entire city with a high density of point clouds, a high-spec computer is required. With that issue in mind, we devised a method for efficient semantic segmentation of large point cloud data even on a relatively inexpensive computer [2]. Combining this method with the aforementioned 3D data resolution enhancement technology makes it possible to acquire high-density point clouds inexpensively and understand an entire 3D scene in detail.

3. 4D point cloud coding technology that enables highly compressed storage of 3D point cloud data

4D point cloud coding is a technology that highly compresses and stores 3D point cloud data obtained from LiDAR and stereo cameras [3]. This technology targets 3D point cloud data concerning, for example, roads and cities collected using the Advanced Geospatial Information Database, and we are researching and developing technology for highly efficient compression of such data and their convenient utilization.

The first problem regarding accumulating and using 3D point cloud data of a wide area of a city is the size of the data. Efficiently storing a large amount of point cloud data necessitates a highly efficient compression technique. Moreover, 3D point cloud data collected moment by moment by using in-vehicle LiDAR, etc. are generated in fragments along the trajectory of a moving vehicle. Therefore, to accumulate and manage those data as a 3D point cloud database covering a wide area of a city, it is necessary to integrate the fragmented 3D point cloud data and detect temporal changes in the latest data from past data. To use such integrated “urban” data, it is necessary to implement a random-access function to retrieve 3D point cloud data at arbitrary locations. In addition to developing the compression function, we are researching and developing three main functions: data integration, temporal-change detection, and random access.

The compression function is based on the point cloud coding technology called geometry-based point cloud compression (G-PCC), which is being standardized by MPEG (Motion Picture Experts Group) and enables highly efficient compression of 3D point cloud data with a large data size. By efficiently representing the spatial location of points as an octree structure while dividing the space hierarchically, G-PCC can compress the data size up to about a tenth of the original size (**Fig. 3**).

Integration of 3D point cloud data and detection of temporal changes are explained as follows. In addition to developing the function for compressing the data of individual files, we are investigating functions for integrating the data in the compressed files and easily comparing the differences between the data. As mentioned above, 3D point cloud data compressed by G-PCC maintain the spatial location of points hierarchically in an octree structure. This hierarchical structure makes it easy to extract a rough outline of the objects in the data from the encoded data. By extracting this outline, it is possible to, for example, detect changes in 3D point cloud data at different acquisition times and combine fragmented data while taking object boundaries into account (**Figs. 4** and **5**).

Regarding the random-access function, by using the hierarchical structure of G-PCC, we are also investigating (i) a scalable function for extracting sparse summary data from dense point cloud data and (ii) a function for extracting only a part of the space. With these functions, it becomes easy to not only extract partial data but also obtain a general idea of

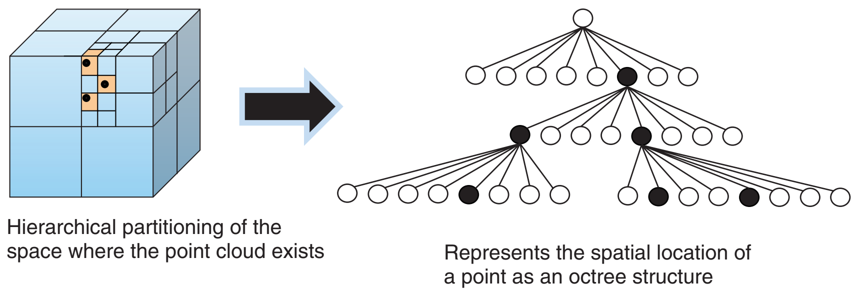


Fig. 3. Layering of point cloud data using an octree structure.

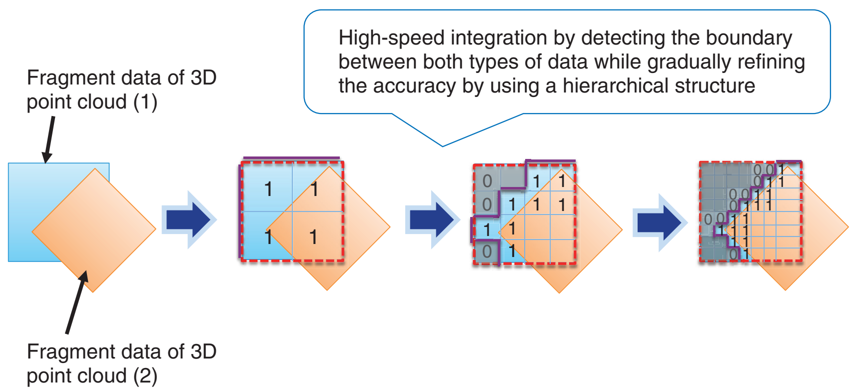


Fig. 4. Combined processing of 3D point cloud data.

the data over a wider area.

We are currently researching and developing a software engine that can handle these three functions in a unified manner. Using this engine will make it possible to store 3D point cloud data of a vast city in a compact size and easily retrieve the data for any arbitrary position and time. We plan to continue our research and development to enable faster and more efficient encoding while taking into account the fact that the volume and distribution of 3D point cloud data will further increase in the future.

4. Future developments

Using the technologies described above, we will establish technologies for constructing cyberspace that can precisely reproduce real space and embody a mechanism for value creation through the advanced integration of real space and cyberspace. Therefore, we will contribute to the creation of a human-centered society in which everyone can lead a comfortable, energetic, and high-quality life.

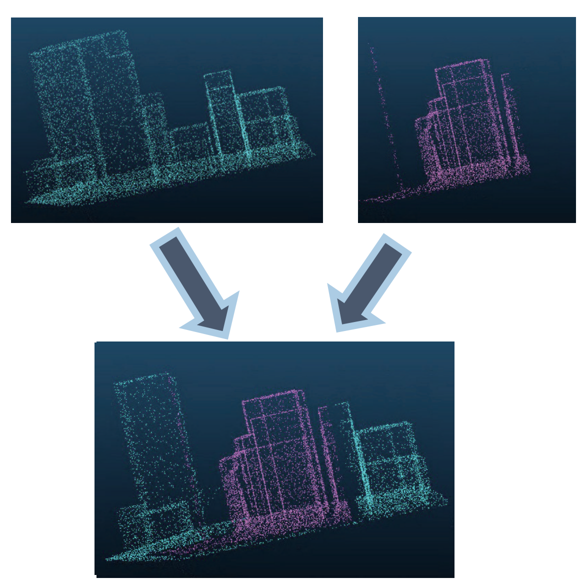


Fig. 5. Result of integrating 3D point cloud data.

References

[1] Y. Yao, R. Ishikawa, S. Ando, K. Kurata, N. Ito, J. Shimamura, and T. Oishi, "Non-learning Stereo-aided Depth Completion Under Mis-projection via Selective Stereo Matching," IEEE Access, Vol. 9, pp. 136674–136686, 2021.

[2] K. Kurata, Y. Yao, S. Ando, N. Ito, and J. Shimamura, "Aggregative Input Convolution for Large-scale Point Cloud Semantic Segmentation," The 7th IIEEJ International Conference on Image Electronics and Visual Computing (IEVC 2021), Sept. 2021.

[3] M. Watanabe, I. Igarashi, R. Tanida, and H. Kimata, "To That Place That Day: An Efficient Method for Merging Compressed Point Clouds to Create Point Cloud Data at Any Time and Place," Proc. of DICOMO 2021, pp. 357–362, June 2021 (in Japanese).

Jun Shimamura

Senior Research Engineer, Supervisor, Cyber-Physical System Group Leader, NTT Human Informatics Laboratories.

He received a B.E. in engineering science from Osaka University in 1998 and M.E. and Ph.D. from Nara Institute of Science and Technology in 2000 and 2006. He joined NTT Cyber Space Laboratories in 2000. His research interests include computer vision and mixed reality. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE).

Shingo Ando

Senior Research Engineer, Supervisor, NTT Human Informatics Laboratories.

He received a B.E. in electrical engineering from Keio University, Kanagawa, in 1998 and Ph.D. in engineering from Keio University in 2003. He joined NTT in 2003, where he has been engaged in research and practical application development in the fields of image processing, pattern recognition, and digital watermarks. He is a member of IEICE, the Institute of Image Electronics Engineers of Japan (IIEEJ), and the Institute of Image Information and Television Engineers (ITE).

Ryuichi Tanida

Senior Research Engineer, NTT Human Informatics Laboratories.

He received a B.E. and M.E. in electrical engineering from Kyoto University in 2001 and 2003. Since joining NTT in 2003, he has been engaged in research on video coding and its application technology. His research interests include image processing, video coding, and sensing data coding technology. He is a member of IEICE.

Introduction

Shingo Kinoshita

Keywords: Olympic and Paralympic Games Tokyo 2020, sports viewing, ultra-realistic communication technology, Kirari!

1. The beginning: Toward future proposals

On September 7, 2013, I was on a business trip to New York with Hiromichi Shinohara, then senior executive vice president and head of the Research and Development Planning Department, when the decision was made to bid for the Olympic and Paralympic Games Tokyo 2020. It seems like only yesterday that we were in a taxi on our way to the airport and I said, "I want to propose a future that will surprise the world at Tokyo 2020." That utterance was the beginning of the Tokyo 2020 Project at NTT's research and development (R&D) laboratories.

At past Olympic Games and other world events, NTT has proposed various visions for the future of communications technology. For instance, at the Olympic Games Tokyo 1964, we provided communication technology for the world's first live broadcast using geostationary satellites and proposed a vision of the future of television broadcasting. At the 1970 World Exposition in Osaka, we experimentally demonstrated the world's first wireless telephone system (**Fig. 1**) and proposed a vision of the future of personal communications. At the Olympic Winter Games Nagano 1998, Japan, we provided a personal handy-phone system (PHS) multimedia communication system and proposed a vision of the future of wearable communications by experimentally demonstrating the world's first wristwatch-type PHS (**Fig. 2**). At the Olympic Games London 2012 and the World Soccer Tournament in Brazil in 2014, we provided technical support for live viewing via 8K Internet Protocol transmission and envisioned the future of public viewing. For the Olympic and Paralympic Games Tokyo 2020, we wanted to propose a vision of the future that goes beyond this legacy, and that proposal became our mission.

Fig. 1. The world's first wireless telephone system tested at the 1970 World Exposition in Osaka.

Fig. 2. Wristwatch-type PHS tested at the Olympic Winter Games Nagano 1998.

2. Challenges

Immediately after returning to Japan from that trip to New York in 2013, we gathered a total of 13 researchers from various NTT laboratories and formed a study team. The ideas we had at that time were dreamy, diverse, and exciting. Now that the Tokyo 2020 Games are over, looking back at the study materials gathered by that team, I'm amazed at how many ideas have been implemented.

The 2020 Epoch-Making Project was established at NTT Service Evolution Laboratories and tasked with the mission of applying these ideas to the Tokyo 2020 Games and proposing a vision of the future. I've been the project leader since 2015. In that year, an ultra-realistic communication technology called "Kirari!", which was envisioned by Katsuhiko Kawazoe, then head of NTT Service Evolution Laboratories, and promoted its research by Akihito Akutsu, was introduced at the NTT R&D Forum. While conducting various demonstrations at events such as *kabuki* (the traditional Japanese performing art) performances at home and abroad, a global technology event in Austin, USA, synchronous performances in three cities around the world by a three-female-member techno-pop unit, the opening of the new National Stadium in Tokyo, and live viewings of the Japanese professional soccer league and U.S. professional baseball league, we continued to make proposals for the Olympic and Paralympic Games Tokyo 2020.

3. Significance of working on the Olympic and Paralympic Games Tokyo 2020

For NTT, Tokyo 2020 Games were not only a celebration of sports and world peace but also a celebration of technology and other innovations. I believe that the involvement in the Games has resulted in the following three benefits to NTT laboratories.

(1) As the largest event of its kind in the world, the Olympic and Paralympic Games attract worldwide attention.

(2) Strict demands and severe constraints are imposed at the Games, leading to improvement and demonstration of not only technical capabilities but also planning and promotion capabilities.

(3) We can take on bold challenges.

Regarding (1), the Olympic and Paralympic Games are watched by as many as four-billion people around the world, and that about six-hundred million people watched the opening ceremony of the Olympic Games Tokyo 2020 alone. Moreover, the latest technologies used at various Olympic and Paralympic Games attract attention from the public and are often used at subsequent sporting events as a legacy.

Benefit (2) has not been pointed out much so far, but it is what we felt most through actual experience. Although it is the flip side of the first reason, the more attention it gets, the stricter the requirements must be satisfied. In addition to performance, quality and safety of the technology, the world's highest levels of expression and design are required. Technology, planning, proposal, and coordination skills are also required. In particular, the range of coordination is very wide, for example, planning proposals required coordination with the Tokyo Organising Committee of the Olympic and Paralympic Games (Tokyo 2020 Organising Committee), the International Olympic Committee (IOC), and international sports federations. When providing live viewing of competition videos via technologies such as Kirari!, it is necessary to coordinate with the broadcasting-rights holders. In regard to installing filming equipment, it was essential to coordinate with Olympic Broadcasting Services (OBS), and in regard to system construction and operation, it was necessary to coordinate with those in charge of the technology, venues, and security staff in the Tokyo 2020 Organising Committee as well as with the Tokyo Metropolitan Government, local governments, and police. Each of them has specific roles and responsibilities and thus finding solutions for a breakthrough is necessary. On top of all that, when introducing such activities on the company's website or issuing news releases, we must ensure that we are within our rights as a sponsor, and we need to be creative in devising ways to express ourselves. Although our sponsor category is telecommunication services, the technologies of our research laboratories are not limited to telecommunication services in a narrow sense, such as so-called communication lines, but often cover communication services in a broader sense.

Regarding (3), being able to take on bold challenges is the most important significance for laboratories. Normally, when selecting a theme at our laboratories, we select reasonably convincing viewpoints such as technological advancement, business potential, or public acceptance. The more viewpoints you have, the more likely you are to settle on a mediocre research topic with a high probability of success and a low level of challenge. However, in the case of the Olympic and Paralympic Games, both public and companies become more positive and open-minded, and that thinking generates a better atmosphere for encouraging challenges. As a result, initiatives that would normally be unchallenged are keenly undertaken, and unexpected results and value may be created. With these three reasons in mind, we worked on various proposals for more than two years and finally obtained more than ten projects of our technology, including live viewing via Kirari!.

4. Postponement, no spectators, and never giving up until the very end: the show went on!

On March 24, 2020, just before the start of the Tokyo 2020 Olympic Torch Relay, our relief, however, was short-lived when the novel coronavirus (COVID-19)—which had been spreading since the beginning of 2020—caused the Olympic and Paralympic Games Tokyo 2020 to be postponed for one year. The postponement led to a major rethinking of the Olympic and Paralympic Games Tokyo 2020 vision, from “the most innovative Games in history, bringing positive change to the world,” to “simplification.” All the projects, including those we had proposed, became candidates for review. Fortunately, the importance of innovation through technology was understood, and all our proposed projects could be continued.

Our next concern became the presence of spectators because many of our projects were designed to be attended. On June 21, 2021, about a month before the opening of the Olympic Games, the policy was announced that the number of spectators should be within 50% of venue capacity and no more than 10,000. At that stage, we were relieved that our projects would be able to proceed without any major effects; however, about two weeks later, on July 8, the policy was changed to not allowing any spectators, except for a few venues. Throughout June, opposition to the public viewing scheduled for Yoyogi Park in Tokyo grew, and eventually most public-viewing sites across Japan decided to cancel their events. We decided to go ahead with our projects without spectators in the hope that our preparations would be complete and that we would be able to communicate with the athletes, officials, and the world through various mass media even if there were no spectators.

The Torch Relay, which started on March 25, was initially often held on public roads; however, in many cases, the relay on public roads was gradually cancelled, and only the ignition ceremony was held. NTT had also planned extended celebration events, in Osaka on April 13 and Yokohama on June 30, to which 5000 spectators were originally invited. In the end, the invitations to the general public were cancelled, and the events were streamed online only.

5. Impact of the COVID-19 pandemic

The aforementioned three benefits of working on the Olympic and Paralympic Games Tokyo 2020 were greatly changed by the COVID-19 pandemic. Therefore, I’d like to reflect on how the pandemic affected the activities of our laboratories in relation to each of the three above-stated benefits.

(1) As the largest event of its kind in the world, the Olympic and Paralympic Games attract worldwide attention

As the world’s attention grew, so did the negative attention, such as the opposition to hosting the Olympic and Paralympic Games during the pandemic. As a result, the focus of people’s attention became whether the Olympic and Paralympic Games should be held and whether there could be spectators. In contrast, people’s expectations for future proposals based on technology and sports entertainment plummeted.

(2) Strict demands and severe constraints are imposed at the Games, leading to improvement and demonstration of not only technical capabilities but also planning and promotion capabilities

The requirements and constraints that are normally stringent were made even more stringent by the pandemic. For example, when organizing events, it was necessary to take thorough measures to prevent COVID-19 infection among spectators and other people involved as well as to prepare for all possible variations of requirements and constraints. For example, if 100% audience capacity, 50% audience capacity, or no attendance is stipulated, the content of the projects had to be significantly changed. How and when to determine this change was a major issue. In addition, with public opinion changing rapidly, among the stakeholders involved in our projects, there were many differences of opinion regarding the content and timing of external announcements such as news releases, and it was difficult to adjust the content accordingly.

(3) We can take on bold challenges

Owing to the growing criticism of the Tokyo 2020 Games and the busy schedule caused by switching to the no-spectators policy just before the event, the momentum in the stakeholders, such as the Tokyo 2020 Organising Committee, to take on new challenges greatly declined. It would have been satisfying

Table 1. List of initiatives of NTT R&D.

Category	Theme	NTT R&D initiatives
1	NTT R&D technologies for viewing Tokyo 2020 Games	Sailing × Kirari!
		Badminton × Kirari!
		Marathon × ultra-low-latency communication technology
2	NTT R&D technologies that colored Tokyo 2020 Games	Torch Relay photo session × Kirari!
		Torch Relay runner support × communication control technology “Swarm”
		Torch Relay celebration stage × Kirari!
		Torch Relay regional events × speech-recognition communication technologies
3	NTT R&D technologies that supported Tokyo 2020 Games	High-efficiency Wi-Fi
		Network security
		Technologies for supporting athletes
		Technologies for supporting staff
		Technologies for supporting spectators

to see more challenging approaches such as virtual-reality viewing and remote cheering, which are based on the premise of no spectators. Despite this atmosphere, we stressed the importance of real-time, interactive remote cheering via telecommunications and proposed to implement it at the marathon and sailing events. Even though the people involved did not have any spare energy, they accepted that proposal and managed to implement it. When I saw the moment when the cheering of remote spectators reached the athletes, along with the people involved, I realized that it was worthwhile taking on the challenge.

6. Future proposals from NTT laboratories and the structure of feature articles

After the above-mentioned twists and turns, the initiatives of NTT laboratories that were ultimately implemented at the Tokyo 2020 Games fall into the following three categories (as listed in **Table 1**):

Category 1, “NTT R&D technologies for viewing Tokyo 2020 Games,” proposes a future of watching sports through “Kirari!” and other technologies, which were applied to the sailing, badminton, and marathons events.

Category 2, “NTT R&D technologies that colored Tokyo 2020 Games,” proposes a future of staging events by using Kirari! and the communication control technology called “Swarm,” which were implemented at events to celebrate the Olympic Torch Relay.

Category 3, “NTT R&D technologies that supported Tokyo 2020 Games,” introduces high-efficiency Wi-Fi and network security as well as technologies that support athletes, staff, and spectators.

These initiatives will be introduced in NTT Technical Review in the coming months.

NTT is an Olympic and Paralympic Games Tokyo 2020 Gold Partner (Telecommunication Services).

Shingo Kinoshita
Vice President, Head of NTT Human Informatics Laboratories.

He received a B.E. from Osaka University in 1991 and M.Sc. with Distinction in technology management from University College London, UK, in 2007. He joined NTT in 1991 and was a senior manager of the R&D planning section of the NTT holding company from 2012 to 2015. He is currently a visiting professor at the Art Science Department, Osaka University of Arts, and visiting executive researcher at Dentsu Lab Tokyo. He has served as a member of the Japan Science and Technology Agency (JST) JST-Mirai Program Steering Committee, member of the All Japan Confederation of Creativity (ACC) TOKYO CREATIVITY AWARDS 2021 Judging Committee, and member of the Broadband Wireless Forum Steering Committee. He has been engaged in R&D of a media-processing technology, user interface/user experience, communication protocols, information security, machine learning, service design, and technology management. Until recently, he had been in charge of NTT's Tokyo2020 initiatives, including sports-watching video technology, inclusive design for social issues, and promoting the use of ICT in *kabuki*, entertainment, and media arts such as live music.

He has been in his current position since 2021, where he manages R&D on information and communication processing of humans based on human-centered principles.

Initiatives for a New Sports-viewing Experience

Shingo Kinoshita

Keywords: Olympic and Paralympic Games Tokyo 2020, sports viewing, ultra-realistic communication technology, Kirari!

1. NTT R&D aims to create a new way to watch sports

The first Olympic Games to be broadcast were the Olympic Games Berlin 1936. Even though the receiver was a television (TV), it only had 180 scanning lines and a screen size equivalent to 19 inches. Images of the Games were shown via such screens located at 28 venues around Berlin and watched by 100 spectators. At the Olympic Games scheduled for Tokyo in 1940, it was planned to broadcast the events via screens with 441 scanning lines, frame rate of 25 frames per second, and an aspect ratio of 4:5, which would be the basis for today's TVs. However, the Olympic Games at Tokyo was cancelled due to World War II, so it became a phantom TV broadcast. In Rome in 1960, the Olympic Games were broadcast live, for the first time, to 18 European countries and, albeit one hour later, to the United States, Canada, and Japan. At the Olympic Games Tokyo 1964, NTT developed the wireless relay technology used for the first satellite relay of the Olympic Games. Since then, NTT has provided technical support for the color broadcast of all events at the Olympic Winter Games Sapporo 1972, the first 2K high-definition broadcast of the Olympic Games Seoul 1988, the 2K high-definition broadcast of all events at the Olympic Winter Games Nagano 1998, and the 8K super-high-definition public viewing at the Olympic Games London 2012.

Although video-relay technology has evolved over 80 years in the manner described above, that evolution has been the same story about devices, namely, TVs, with screens surrounded by a square frame. In other words, it has not fundamentally changed the experience of watching sports. In preparation for the Olympic and Paralympic Games Tokyo 2020, in 2015, NTT began research and development (R&D) of an ultra-realistic communication technology called "Kirari!" for creating a new sports-viewing experience. To deliver the experience of being at the venue to those who cannot attend the actual event, Kirari! is an ambitious project that removes the square frame of conventional TV and public viewing and transmits the event space itself. Since 2015, we have experimentally demonstrated Kirari! for live viewing of a variety of events in addition to sports such as the Japanese professional soccer league and U.S. professional baseball league. Example live-viewing demonstrations include *kabuki* events in Japan and abroad, a global technology event in Austin, Texas, USA, a synchronized performance by a three-female-member techno-pop unit in three cities around the world, and the opening event at the new National Stadium in Tokyo (**Fig. 1**).

Initially, Kirari! was intended to provide viewing for people who lived far away from the event venue, people who could not afford tickets, and those unable to go out due to illness. However, with the spread of the novel coronavirus (COVID-19) in 2020, people's mobility became ever more restricted, and the vision of Kirari! grew ever more important. The feeling that you could not go to an event venue even if you wanted to go or that something was missing even if you watched the event on TV or smartphone became even more pronounced in 2020 when events, such as concerts and sports, were canceled due to the COVID-19 pandemic and online distribution suddenly expanded. In the beginning, many people tried online streaming to relieve the stress of being unable to get to the

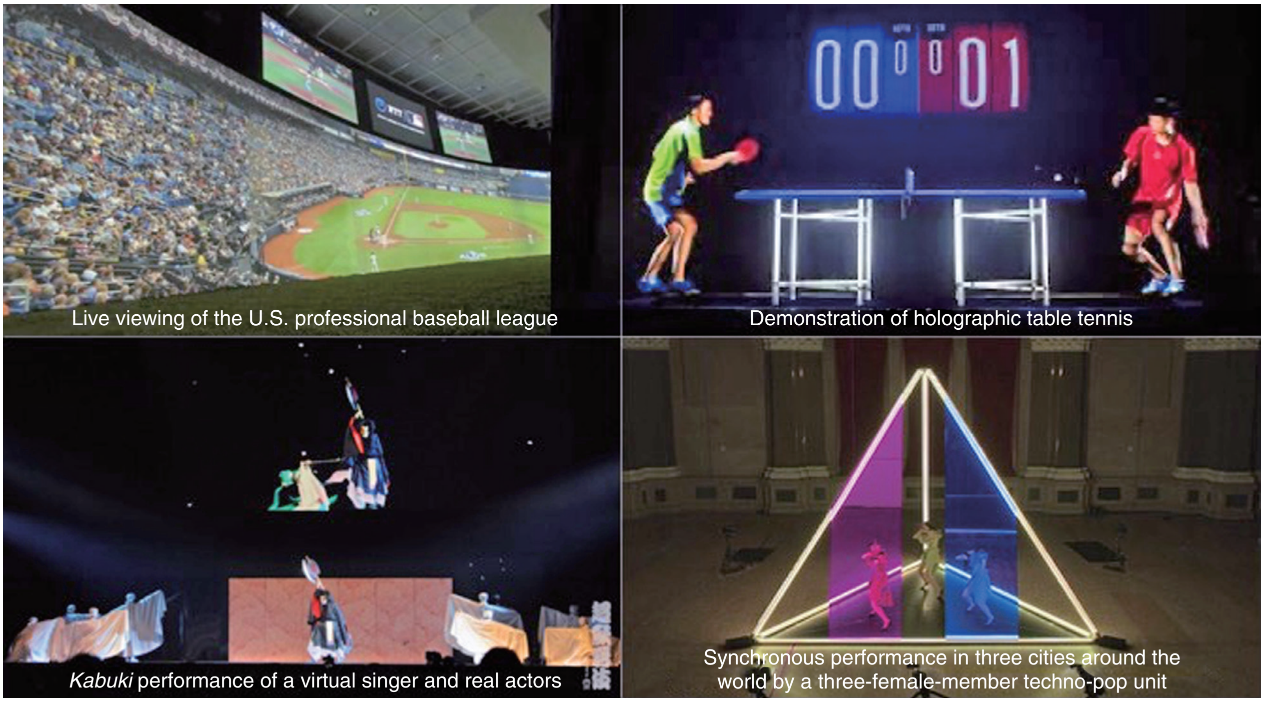


Fig. 1. Previous showcases of Kirari!.

venue, out of curiosity, or because of its ease of use. However, in accordance with the results of a survey called "Awareness Survey on Live Music Distribution" carried out by SKIYAKI Inc. in September 2020, many people who have experienced real live performances answered that real live performances are better than online-streaming ones because they convey senses of presence, unity, and specialness. For example, 77% of the survey's respondents answered, "Real live is better" because it gives "a sense of presence" (92.9%), "sense of unity" (93.6%), and "sense of specialness" (69.6%). Given those results, we considered the question, "How can we recreate those sensations?" For that reason, recognizing once again the importance of the vision of Kirari!, we decided to take on that challenge at the Olympic and Paralympic Games Tokyo 2020.

2. Sense of presence

The "sense of presence" concerning a sports event means that you feel as if you are actually there watching the event. In the actual venue, the distance between a spectator to the action is so great that you cannot see more than a speck in the distance. Although it is easier to see the action online, the feeling of "being there" is what makes the real thing so appealing. This sense of being there is based on two main factors: the feeling that "you are there" (at the venue) and the feeling that the "athletes are there" (in front of you). TVs and smartphones fit into a small square frame, so we perceive them as a separate space; in other words, we do not feel that we are sharing the event space with the athletes or feel the space that expands when we enter the venue and sit down.

To create a sense that "you are there" (at the venue), as if you were sitting in the stadium, we attempted to reproduce an expanse of space that covers the full viewing angle. To achieve this, we use the ultra-wide image-synthesis technology of Kirari!. Multiple 4K cameras are installed and multiple videos from them are seamlessly stitched together in real time to produce a single ultra-high-resolution video image. As a result, video images with a resolution of over 20K can be displayed on a display several tens of meters wide. Although it is possible to enlarge the video image from a 4K or 8K camera and project it on a huge display that covers the entire viewing angle, the image resolution will inevitably be rough, and it will be difficult to reproduce the sense of presence of the scene. It is also difficult to freely change the size of the image (in terms of height and width) in accordance

Fig. 2. Sailing × Kirari!.

with the event to be presented. NTT has conducted experiments with the U.S. professional baseball league, professional Japanese soccer league, tennis tournament in France, windsurfing world championships, and a large-scale fashion show for women in Tokyo.

For the Olympic and Paralympic Games Tokyo 2020, we used this technology at the sailing, at which the spectators were seated far from the sailing course. By using Kirari!, we transmitted the sailing course directly to an offshore display near the spectators' seats to reproduce the sensation as if the races were being held right in front of them (**Fig. 2**). The details are introduced in the article "Sailing × Ultra-realistic Communication Technology Kirari!" in this issue [1].

To create a sense of the "athletes are there" (in front of you), we attempted to create three-dimensional images of the athletes at the Tokyo 2020 Games. To achieve this, we used the "real-time extraction of objects with arbitrary backgrounds," a component technology of Kirari!. Even if there is no green screen, the athlete's image is extracted in real time and displayed in a holographic manner. Even if it is the same image, displaying it as a holograph gives the audience a stronger sense of the athlete's presence. The sense of presence was further enhanced by setting up realistic objects other than the images of the athletes, such as the badminton court or table tennis table. We conducted experiments with this technology at karate, judo, and badminton tournaments as well as at *kabuki* performances and a large-scale music event in Texas, USA.

For the Olympic Games Tokyo 2020, we used this technology at the badminton competition. Only video images of the athletes were extracted from 8K camera images of the badminton matches held at Musashino Forest Sport Plaza, and those images were transmitted to the National Museum of Emerging Science and Innovation (Miraikan), the remote-viewing venue. At this remote-viewing venue, courts, nets, and spectators' seats were set up in a manner just like that at the main venue. The holographic images of the athletes were then displayed in conjunction with those objects in a way that created a sense of the athletes' presence just like that at the main venue (**Fig. 3**). The details are introduced in the article "Badminton × Ultra-realistic Communication Technology Kirari!" in this issue [2].

3. Sense of unity

A "sense of unity" means a connection between the athletes and spectators as well as among spectators. In sports, cheering from the spectators is the most powerful force in regard to the athletes. Spectators can connect with each other through cheering, thereby creating even more excitement as they become one. For online streaming, which was expanded in 2020, efforts were made to enhance the sense of unity by, for example, displaying images of the remote audience on the stage and initiating call-and-response. In reality, however, the timing of the shouts and cheers was too disjointed, preventing a sense of unity.

The reason for that disjointedness is mainly the latency and variability of the communication. Delays of more than a few dozen milliseconds generally make music sessions difficult, and delays of more than a few hundred milliseconds make call-and-response uncomfortable. Such a delay has the following three components: the propagation delay of light is about 0.5 ms for 100 km, transmission-processing

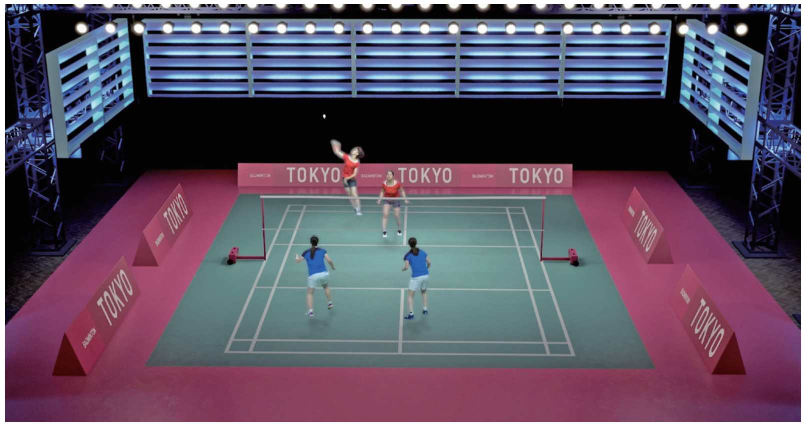


Fig. 3. Badminton × Kirari!.

Fig. 4. Marathon × ultra-low-latency communication technology.

delay is several milliseconds to several dozen milliseconds, and video-coding process takes several hundred milliseconds. With the addition of video editing and other factors, it is not uncommon for the final delay to be up to ten seconds in the case of digital terrestrial broadcasting and live streaming over the Internet.

For the Olympic Games Tokyo 2020, ultra-low-latency communication technology, which significantly reduces latency, was used at the marathon. Spectators' cheers from Tokyo were delivered without delay to the athletes in Sapporo, who were running at 5 m/s, and the challenge was to create a sense of unity that transcended distance (**Fig. 4**). The details are introduced in the article "Marathon × Ultra-low-latency Communication Technology" in this issue [3].

NTT is an Olympic and Paralympic Games Tokyo 2020 Gold Partner (Telecommunication Services).

References

[1] T. Miyatake, K. Shimizu, T. Hayasaka, M. Nakamura, M. Aoki, H. Chigira, J. Nakajima, M. Urata, K. Ogo, Y. Yoshida, and S. Kinoshita,

"Sailing × Ultra-realistic Communication Technology Kirari!", NTT Technical Review, Vol. 19, No. 12, pp. 57–68, Dec. 2021.
https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202112fa8.html

[2] M. Muto, K. Hasegawa, D. Namikawa, S. Konya, N. Hirachi, T. Nakamura, K. Suzuki, and S. Kinoshita, "Badminton × Ultra-realistic Communication Technology Kirari!", NTT Technical Review, Vol. 19, No. 12, pp. 69–77, Dec. 2021.
https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202112fa9.html

[3] S. Usui, S. Fukatsu, E. Matsumoto, M. Imoto, D. Shirai, and S. Kinoshita, "Marathon × Ultra-low-latency Communication Technology," NTT Technical Review, Vol. 19, No. 12, pp. 78–84, Dec. 2021.
https://ntt-review.jp/archive/ntttechnical.php?contents=ntr202112fa10.html

Shingo Kinoshita
Vice President, Head of NTT Human Informatics Laboratories.

He received a B.E. from Osaka University in 1991 and M.Sc. with Distinction in technology management from University College London, UK, in 2007. He joined NTT in 1991 and was a senior manager of the R&D planning section of the NTT holding company from 2012 to 2015. He is currently a visiting professor at the Art Science Department, Osaka University of Arts, and visiting executive researcher at Dentsu Lab Tokyo. He has served as a member of the Japan Science and Technology Agency (JST) JST-Mirai Program Steering Committee, member of the All Japan Confederation of Creativity (ACC) TOKYO CREATIVITY AWARDS 2021 Judging Committee, and member of the Broadband Wireless Forum Steering Committee. He has been engaged in R&D of a media-processing technology, user interface/user experience, communication protocols, information security, machine learning, service design, and technology management. Until recently, he had been in charge of NTT's Tokyo2020 initiatives, including sports-watching video technology, inclusive design for social issues, and promoting the use of ICT in *kabuki*, entertainment, and media arts such as live music.

He has been in his current position since 2021, where he manages R&D on information and communication processing of humans based on human-centered principles.

Sailing × Ultra-realistic Communication Technology Kirari!

Takashi Miyatake, Kentaro Shimizu, Tomoyuki Hayasaka, Mushin Nakamura, Masakatsu Aoki, Hiroshi Chigira, Junichi Nakajima, Masakazu Urata, Kenta Ogo, Yuki Yoshida, and Shingo Kinoshita

Abstract

NTT provided its ultra-realistic communication technology called Kirari! to the TOKYO 2020 5G PROJECT, which was implemented by the Tokyo Organising Committee of the Olympic and Paralympic Games to offer a new experience of sports viewing by using 5G (5th-generation mobile communication system). A new experience of watching sports at the sailing regatta, held at the Enoshima Yacht Harbor, is introduced in this article. In short, a sense of realism—as if the spectators were watching the races from close up—was provided by transmitting ultra-wide video images taken close to the race course to the spectators' seats set up at the harbor.

Keywords: 5G, Kirari!, live viewing

1. Overview of our initiatives at the sailing regatta

In sailing races, spectators usually have to use binoculars to watch the races (from breakwaters, etc.) because the races are held quite far from land. At the sailing regatta of the Olympic and Paralympic Games Tokyo 2020, we took up the challenge of solving this problem by using communication technology to create an innovative style of spectating that goes beyond the real world; namely, the spectators feel as if they were watching the races from a special seat on a cruise ship. Ultra-wide video images of the races (with a horizontal resolution of 12K) were transmitted live to a 55-m-wide "offshore wide-vision" display near the spectators' seats (**Fig. 1**) by using the 5th-generation mobile communication system (5G) and our ultra-realistic communication technology Kirari!.

Initiatives at the sailing regatta for the TOKYO 2020 5G PROJECT are summarized below:

- Period: July 25, 2021 to August 4, 2021
- Venue: Enoshima Yacht Harbor, Kanagawa Prefecture
- No. of events held: 10
- Total number of participants: about 4000 (only athletes and related people were allowed to watch the event)
- Organization: TOKYO 2020 5G PROJECT formed by four bodies: the Innovation Promotion Office of the Tokyo Organising Committee of the Olympic and Paralympic Games, NTT (participating as the technical supervisor), NTT DOCOMO, and Intel Corporation (**Fig. 2**).

The ultra-wide video images were also transmitted live to Tokyo Big Sight, where the main press center (MPC) was located, to convey the realism of the event to media personnel whose movements were restricted within Japan because of the novel coronavirus (COVID-19) pandemic (**Fig. 3**).

Although the sailing regatta was held without spectators, we implemented a measure called a "virtual stand" for sending support to the athletes from their far-away families and friends (**Fig. 4**).

Fig. 1. Example of live video transmission via offshore wide-vision display.

Fig. 2. Logo of TOKYO 2020 5G PROJECT.

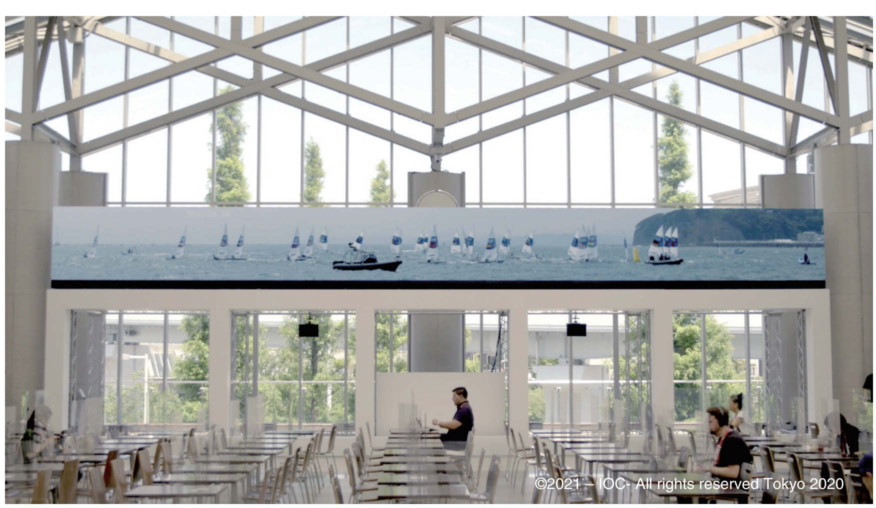


Fig. 3. Example of video transmission to MPC.

Fig. 4. Remote cheering from the virtual stand.

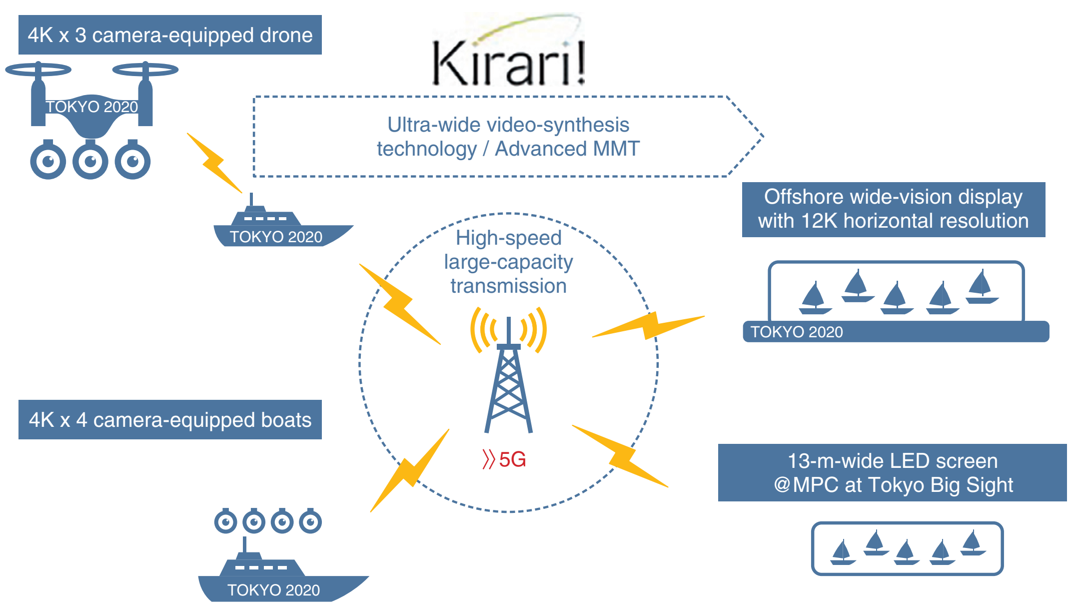


Fig. 5. Overview of the video transmission system.

2. Configuration of the video transmission system

The configuration of the video transmission system for the sailing regatta is shown in **Fig. 5**. First, the sailing action is captured from boats or a drone equipped with multiple 4K cameras. The captured video images are synthesized as ultra-wide video images with 12K resolution and transmitted in real time by using Kirari! and 5G, which enables high-speed, large-capacity transmission. The transmitted images were displayed in real time on a 55-m-wide offshore wide-vision display floating in front of the viewing area at the regatta venue and on a 13-m-wide

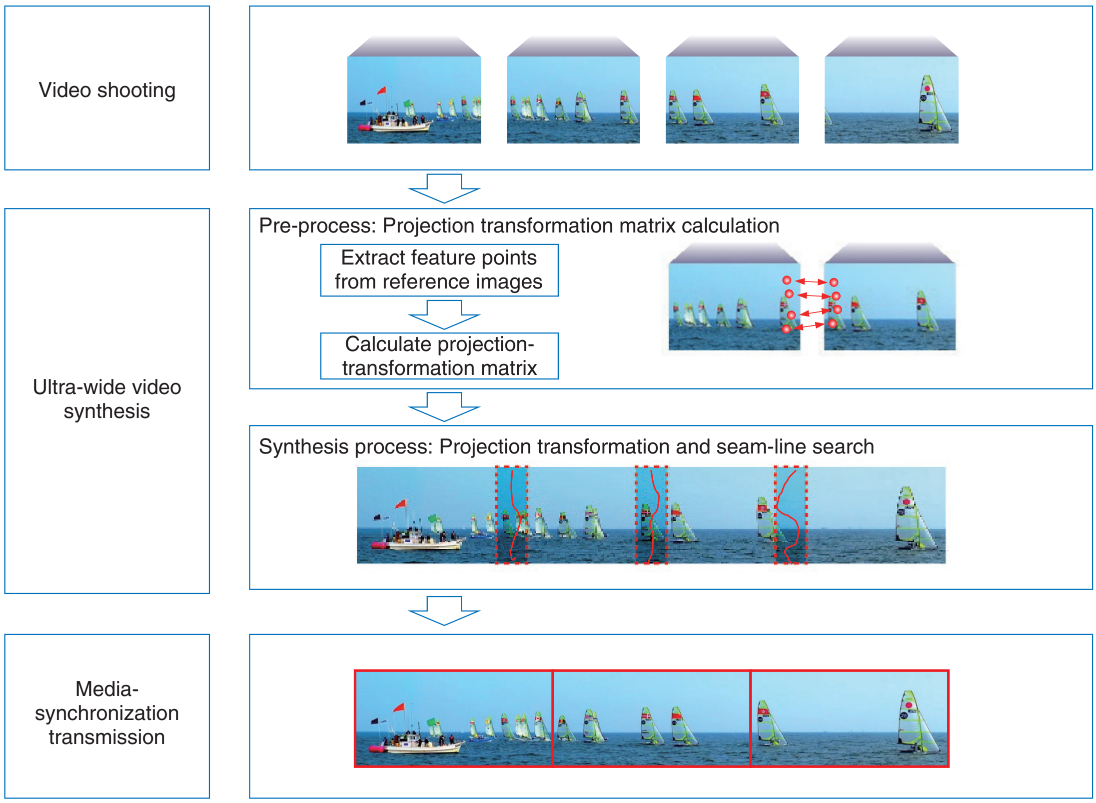


Fig. 6. Processing flow of ultra-wide-video synthesis technology.

light-emitting diode (LED) display installed at the MPC at Tokyo Big Sight.

3. Technology

3.1 Ultra-realistic communication technology Kirari!

Kirari! is a communication technology that enables spectators to experience a race or other sports events as if they were in the stadium even when they are far away from it. In this project, we implemented two technological elements of Kirari!, video transmission using ultra-wide video-synthesis technology and ultra-realistic media-synchronization technology.

3.1.1 Ultra-wide video-synthesis technology

The process flow of ultra-wide video synthesis is shown in **Fig. 6**. First, the optical axes of each of four cameras are aligned as much as possible, and the cameras are set up so that their shooting ranges overlap adjacent captured images by 20–40%. Before the actual synthesis process, as a pre-processing step to correct the disparity between each camera, a projection-transformation matrix is calculated from feature points extracted from a reference still image. For sailing, an image was taken of the competition yachts waiting just before the race as the reference image. To implement the pre-processing, we optimized the entire process from acquisition of the reference image to determining the wide-image-synthesis parameters by compressing the image data and reducing the transmission time by two-thirds, for example. Next, during the synthesis process, video frames were combined using a projection-transformation matrix calculated for each video frame to correct disparities and find appropriate seam lines in the video images (i.e., a process called "seam search"). The synthesis process is characterized by three features: synchronous distributed processing, accelerated seam search, and use of a graphics processing unit (GPU).

For synchronous distributed processing, the synthesis process is distributed to multiple servers (in accordance with the input and output conditions) so that a huge amount of video processing can be executed in real time. To synchronize the synthesis timings in each server, the same time code is given to the video frames when they are input, and this time code is propagated to the transmission devices (such as encoders) in the subsequent stages. As a result, synchronization of frames can be controlled across the entire system. The time code can be selected with either of two methods: (i) using the time code superimposed on the input video or (ii) generating the time code when the video is input.

The seam search is sped up in two ways: (i) using the movement and shape of moving bodies obtained by analyzing the video frames during the seam search or (ii) using the seam position of the past video frame to prevent the seam position from fluttering. If the seam search for the next frame is executed after the seam search for the previous video frame is completed, it becomes difficult to achieve real-time synthesis. Accordingly, a two-stage seam search is conducted on a reduced video frame as a pre-search, and the pre-search result is used in the seam search of the next frame.

For GPU-based acceleration of the seam search, almost all video processing is executed on the GPU, so processing is more intensive and faster. As a result of using the GPU, the video input from the serial digital interface (SDI) board is transferred to the GPU with minimal central-processing-unit cost, and the subsequent video processing is executed by the GPU. For example, when synthesizing 12K (4K × 3) ultra-wide video images from images captured with the four 4K cameras, it was possible to reduce the number of servers required by two-thirds (from six to two) compared with when the GPU was not used.

3.1.2 Ultra-realistic media-synchronization technology (Advanced MMT)

The video images generated using the ultra-wide video-synthesis technology are divided into multiple 4K images for transmission so that they can be handled by general display devices, encoders, and decoders. In this case, the image on the display was 12K wide and 1K high, so it was divided into three 4K images. Kirari! uses advanced MPEG Media Transport (MMT) technology to perfectly synchronize the playback timing of individual images and sounds that are transmitted in separate segments. Ultra-realistic media-synchronization technology (Advanced MMT) synchronizes and transmits multiple video and audio with low delay by using UTC (Coordinated Universal Time)-based synchronization control signals that comply with the media transmission standard MMT. This technology can (i) synchronize and display not only wide-angle video images but also multi-camera-angle video images (2K, etc.) that track a specific athlete and (ii) absorb the difference in delays due to the network (dedicated line, Internet, wireless local area networks, etc.) that transmits each camera image or audio signal and device processing. This enables an ultra-realistic viewing experience in which multiple videos and audio are completely synchronized.

3.2 Wireless communication technology

The races of the sailing regatta at the Olympic and Paralympic Games Tokyo 2020 were held on six race courses (**Fig. 7**). Two of those race courses, “Enoshima” and “Zushi,” were chosen for our initiatives. Ultra-wide video images were captured from three camera boats for the Enoshima course, and ultra-wide images were captured using a drone camera in addition to three camera boats for the Zushi course. 5G base stations were set up near both courses at the points shown in the figure.

The configuration of the network used in this project is shown in **Fig. 8**. For uplinking video from the boats to shore and downlinking from the video-editing base to the barge-mounted offshore wide-vision LED display, NTT DOCOMO’s 5G line was used as the core in conjunction with multiple wireless communications with different characteristics such as bandwidth, distance, and directionality. NTT DOCOMO’s 4G line was used for flight control of the drone, backup video, and control of the server located on the camera boats.

3.2.1 5G downlink transmission

The video transmission between the video-editing base and barge-mounted offshore wide-vision LED display was carried out using NTT DOCOMO’s 5G access premium service. A commercially available customer premises equipment (CPE) terminal was used as the receiver. During the regatta, 200-Mbit/s video transmission was continuously operated for more than eight hours a day, achieving stable video transmission.

3.2.2 5G uplink transmission

NTT DOCOMO’s 5G access premium service was mainly used for transmission of drone video. A commercially available CPE terminal was used as the receiver. Six 2K video images were transmitted at an average bandwidth of 80 Mbit/s. NTT DOCOMO’s 4G service was used for backup of video transmission at a total rate of 50 Mbit/s.

3.2.3 Transmission of video images from a drone

The system configuration of a drone used to capture video images of the races is shown in **Fig. 9**. The drone was fitted with three 4K cameras. Three streams of 4K video taken by a drone were transmitted

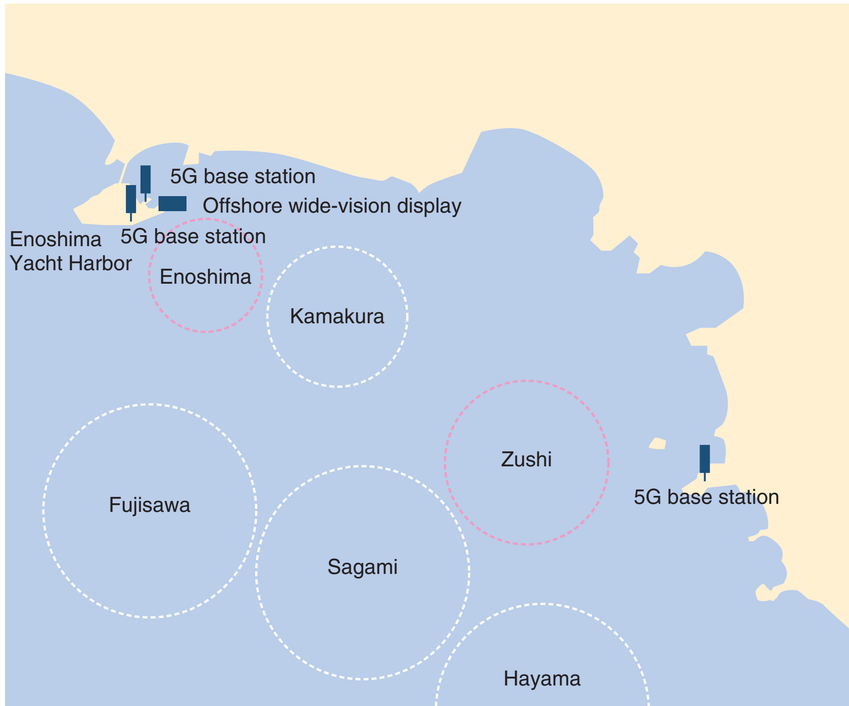


Fig. 7. Map of the race courses for the sailing regatta.

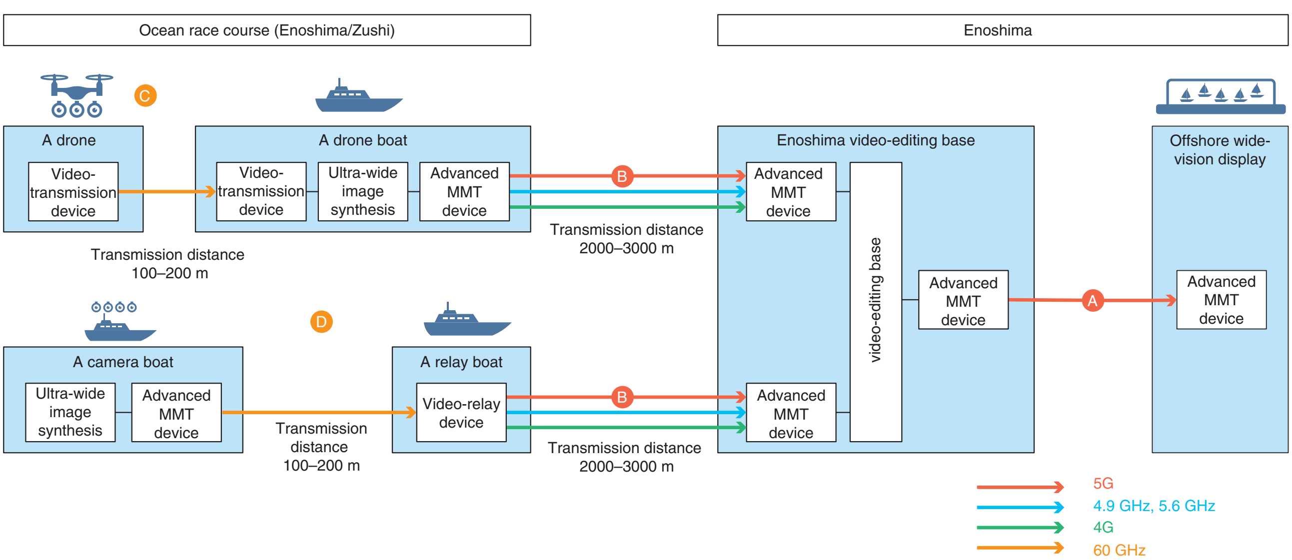


Fig. 8. Network-configuration diagram.

to a drone boat, from which the drone is launched and landed via 60-GHz wireless communication. The three video streams were time-synchronized for transmission. On board the drone boat, wide video images were synthesized and synchronously transmitted to the video-editing base at Enoshima via the 5G access premium service and Advanced MMT (Fig. 8).

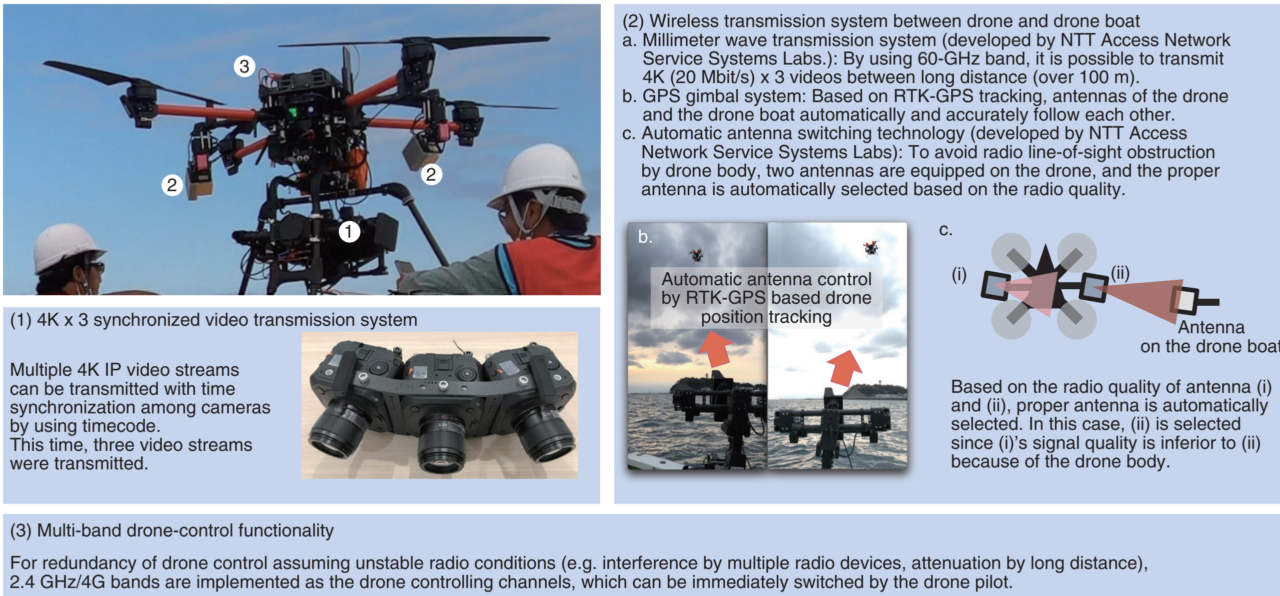


Fig. 9. Configuration of a drone system.

For the video transmission between the drone and drone boat, 60-GHz wireless-communication technology developed by NTT Access Network Service Systems Laboratories was used. For the wireless equipment, it is necessary to keep the antenna surface facing the drone all the time. To ensure more-stable wireless communication, we developed a technology that automatically controls the direction of the antenna by using real time kinematic (RTK) GPS (Global Positioning System) information so that the antenna always points in the direction of the drone. Depending on the shooting angle of the drone, the drone may create a blind spot for wireless communication. To solve this problem, we developed a technology for automatically switching the antenna in accordance with the wireless-signal strength. The above technologies enable stable video transmission between the drone and drone boat in a manner that gives the drone more freedom to shoot video. As a security measure for the drone, the flight control of the drone was duplexed using two 4G wireless signals, one at 4 GHz and the other at 2 GHz.

3.2.4 Transmission of video images from camera boats

The following requirements for transmission of video images from the camera boats were set:

1. Since the race course may be changed at short notice due to weather conditions or wind direction, the camera boats should be able to move flexibly.
2. Depending on the race course used, wireless transmission over a distance of 2 to 3 km is required.

To meet these requirements, we prepared a "relay boat"—separate from the camera boats—for long-distance wireless transmission to the land. For transmission of video between the camera boats and relay boat, 60-GHz wireless communication, which has a short transmission distance but a large data-transmission capacity, was used. To allow the relay boat to move flexibly, the boat was fitted with a wireless antenna with a relatively wide angle of directivity. In addition to 5G, 4.9-GHz and 5.6-GHz wireless communications were used to transmit video images from the relay boat to land (Fig. 8). While these wireless communications can transmit over long distances, the antenna directionality is strong, so the antenna surfaces of the transmitter and receiver must always be aligned. Two mechanisms for controlling the antenna were developed (**Table 1** and **Fig. 10**).

The type of boats was selected by applying a bit of ingenuity. Since it was necessary to film the races close up, small fishing boats were selected as the camera boats because they would not affect the navigation of the racing yachts. A large fishing boat that can suppress rocking even in open seas was selected because it has to transmit wireless signals over 2 to 3 km.

Table 1. Two systems used for controlling antennas.

Systems	Details	Radio
System 1 Remote control	We integrally molded a remote controllable surveillance camera and antenna and remotely controlled the antenna using surveillance-camera images. Specifically, we used surveillance-camera images from boats to land to remotely control the direction toward the land antenna, achieving wireless communication.	Used for radios with relatively large antenna-reception angles (30 degrees)
System 2 Auto control by GPS	We developed a technology that automatically controls the direction of the antenna by using RTK-GPS information so that the antenna always points in the direction of the boat and land.	Used for radios with relatively small antenna-reception angles (10 degrees)

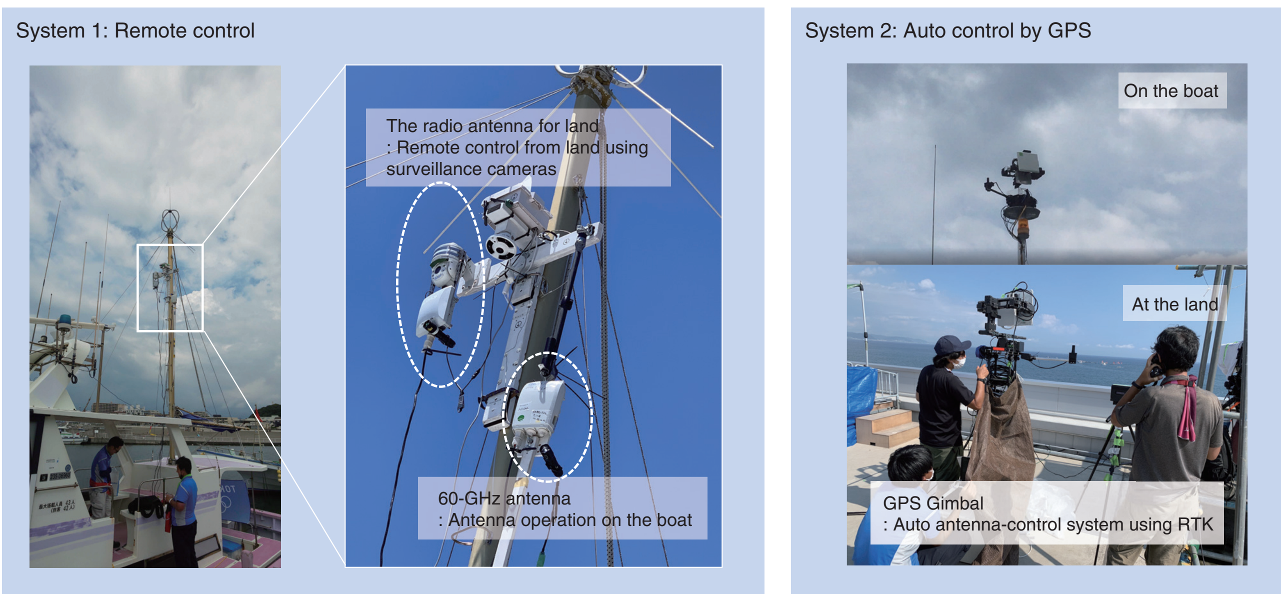


Fig. 10. Examples of using two antenna-control systems.

For the on-land antennas, it was not always possible to secure a roof top of a tall building along the coastline. Accordingly, a vehicle was fitted with a pole that could be raised high above the vehicle, and the antenna was attached to the top of the pole to ensure a clear line-of-sight to the relay boat.

The 60-GHz band enables transmission at 100 to 200 Mbit/s over a distance of 100 m (i.e., boat-to-boat distance), while the 4.9- and 5.6-GHz bands enable transmission at 50 to 150 Mbit/s over a distance of 2 to 3 km (boat-to-land distance). Communication speed depended on the condition of the ocean waves. During the approach of typhoon Nepartak, the waves were high, so it was difficult to control the antennas of the wireless transmitters, which significantly reduced communication speed.

3.3 Offshore wide-vision display

To create a sense of realism as if you were watching the races up close, highly realistic video images and a space to make use of those images are required. To meet the first requirement, actual-size images of the competition yachts, a display size that covers the entire field of view, high resolution, and a display with no noticeable dots are needed. The offshore wide-vision display was 5 m high and 55 m wide with a horizontal resolution of 12K and vertical resolution of 1K. Although the pitch size was 4.8 mm, as shown in **Fig. 11**, it was installed 40 m from the spectators' seats, so the roughness of the image was not noticeable because it exceeded the spectators' retinal resolution.

To meet the second requirement, the space in which the display is set up is also important. Since the

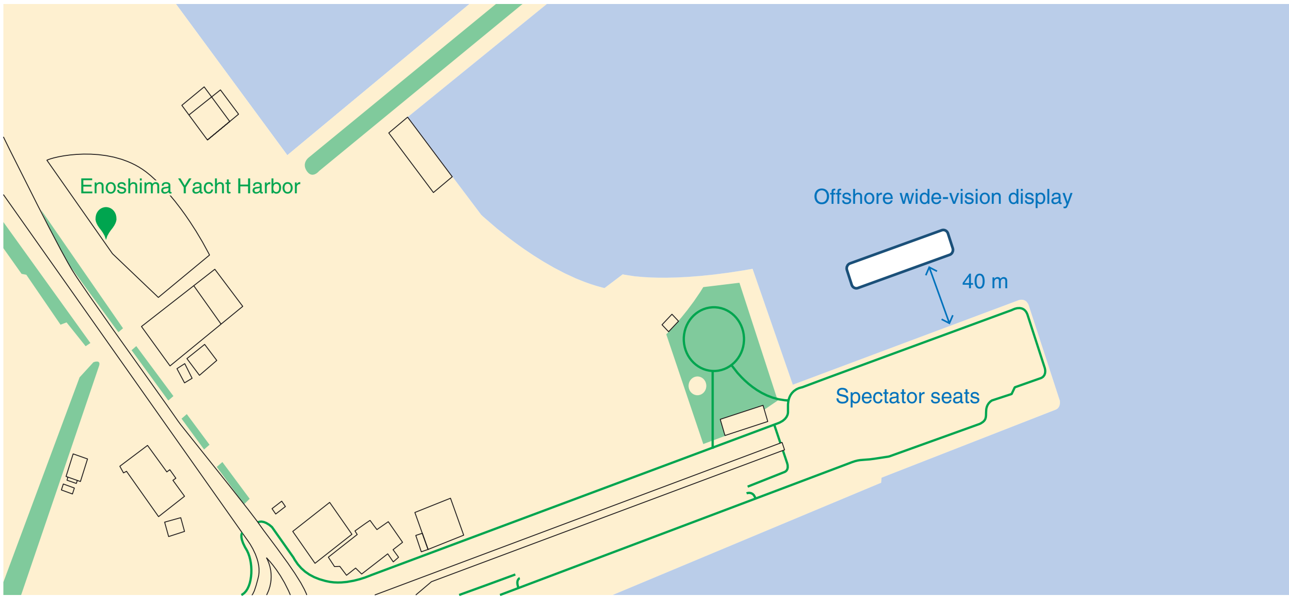


Fig. 11. Location of the offshore wide-vision display.

display is like a rectangular cut-out, it inevitably leaves an unnatural feeling in the same manner as a television. By floating the display on the ocean and merging its boundaries with the natural sky and sea, it was possible to create a natural sense of realism.

When the offshore wide-vision display was installed, we set up an organization containing not only specialized members who design and build LEDs but also those who carry out structural calculations on the basis of the undulation of the sea surface, those who navigate and moor boats, marine-procedure agents, weather forecasters, etc. and studied the following points in consideration of the weather at the actual regatta. We collected meteorological data for Enoshima in July and August over the past 10 years to understand wind speed, wind direction, wave height, etc., evaluated the strength of the designed LED display and the safety of the means of mooring the barge of the floating display, and determined the criteria for evacuating the site of the offshore wide-vision display. To meet the evacuation criteria, the display was operated under the condition that wind speed did not exceed 25 m and wave height (swell) did not exceed 1.6 m. The criteria were determined after careful considerations of the members specializing in marine weather based on not only the intensity of the LEDs and the strength of the mooring of the floating-display barge but also the weather conditions on Enoshima Island and on the waters off Miura Peninsula in case of evacuation to Odaiba in Tokyo. During typhoon Nepartak, which occurred during the regatta, weather information was constantly monitored, and possible evacuation was determined on the basis of the above-mentioned criteria. As a result of that monitoring, it was determined that no evacuation was necessary, and the display was operated without incident.

3.4 Wave-undulation correction technology

When filming from a boat on the ocean, the captured video images will flicker due to the rocking of the boat. We therefore developed a technology for compensating for this video flickering in real time. The developed technology uses optical flow to detect the direction of the video flickering and correct it in real time to minimize as much flickering as possible.

4. Results

During the regatta, the video transmission went smoothly without any major problems. Although the regatta was not attended by spectators, hundreds of people, mainly athletes and related support people, watched the video transmission every day of the regatta. We received comments from many people of those that expressed their surprise, excitement, and hopes for the future of sailing. Some comments were “I was surprised by the high sense of realism that I have never experienced before," "I really wish the general public could see this," "It made it easy to understand the race situation," and "If this style of spectating becomes common, the number of sailing fans will increase rapidly." For our aim of creating "a sense of realism, as if the race was being held in front of you," we were able to create the experience of "warping into the race space itself" by fusing the powerful, actual-size race images displayed on the offshore wide-vision display floating on the sea with the actual surface of the sea, sky, and other real spaces.

Technically, issues still had to be solved. These issues include (i) unstable wireless communication, including 5G, in certain situations and (ii) slight disturbance in the video images because the camera boats and drone could not completely absorb the undulations due to the waves and engine, respectively. However, we believe that we were successful in an environment that demands a very high level of performance.

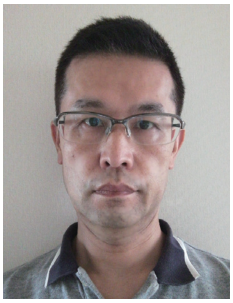
5. Concluding remarks

Using commercial 5G and our in-house ultra-realistic communication technology Kirari!, we successfully transmitted live ultra-wide video images (with 12K horizontal resolution) via a 55-m-wide offshore wide-vision display installed at the Enoshima Yacht Harbor site and a similar but smaller display at the MPC at Tokyo Big Sight in Tokyo. At both sites, the live video transmissions were highly evaluated by athletes, officials, and international media. The communication technology used in this project will enable live viewing not only within a competition venue but also to remote locations, including overseas ones. Even during the COVID-19 pandemic, this project can be ranked as a showcase for achieving a remote world that NTT is aiming for. From now onwards, we will promote further research and development on communication technologies, mainly in the fields of sports and entertainment, using 5G and the All-Photonics Network of IOWN (Innovative Optical and Wireless Network) as an infrastructure.

Acknowledgments

We thank all members of the TOKYO 2020 5G PROJECT, the Innovation Promotion Office of the Tokyo Organising Committee of the Olympic and Paralympic Games, and Intel Corporation who worked with us on the project. We also thank the Functional areas (FAs) of the Tokyo Organising Committee of the Olympic and Paralympic Games, World Sailing, and Japan Sailing Federation for their cooperation in implementing this project, and the Enoshima-Katase, Koshigoe, and Kotsubo Fishing Cooperatives for providing the necessary boats we used for filming. Finally, we particularly want to thank Tatsuya Matsui, the chief engineer in charge of EEP infrastructure, and Hikaru Takenaka, who was in charge of network construction for this sailing project.

NTT is an Olympic and Paralympic Games Tokyo 2020 Gold Partner (Telecommunication Services).



Takashi Miyatake

Senior Research Engineer, Supervisor, NTT Human Informatics Laboratories.

He received an M.E. in engineering from Kobe University in 1995. In the same year, he joined NTT and engaged in research on image processing and human interface. He also worked as a system engineer at NTT Communications and other companies. He is currently engaged in research of the ultra-realistic communication technology Kirari!.

Kentaro Shimizu

Senior Research Engineer, Supervisor, NTT Human Informatics Laboratories.

He received an M.S. in electronics and information engineering from Yokohama National University, Kanagawa, in 1999 and Ph.D. in medical research from Tokyo Women's Medical University in 2009. He joined NTT in 1999. He is a member of the Institute of Electronics, Information and Communication Engineers, the Information Processing Society of Japan (IPSJ), the Japanese Society for Artificial Intelligence, the Japan Society for Health Care Management, and the Japan Society of Computer Aided Surgery.

Tomoyuki Hayasaka

Senior Research Engineer, NTT Human Informatics Laboratories.

He received an M.Sc. in physics from Hokkaido University in 1996. He joined NTT the same year and engaged in a system engineering on web systems. He has been engaged in research on Kirari! since 2019.

Mushin Nakamura

Senior Research Engineer, NTT Human Informatics Laboratories.

He received a B.S. in mathematics and M.E. in computer science from the University of Tokyo in 2001 and 2003. He joined NTT in 2003, where has been engaged in research and development (R&D) of NGN (Next Generation Network) appliances and web-based signage systems.

Masakatsu Aoki

Senior Research Engineer, NTT Human Informatics Laboratories.

He received a B.E. and M.E. in engineering from Shinshu University, Nagano, in 1994 and 1996. He joined NTT Human Interface Laboratories in 1996 and worked on research related to computer graphics. His research interests include human-computer interaction, life logs, and visual communication.

Hiroshi Chigira

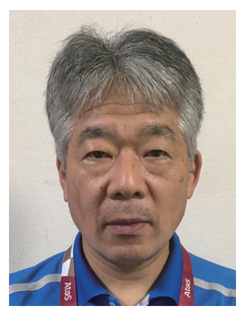
Senior Research Engineer, NTT Human Informatics Laboratories.

He received a B.E. and M.E. in engineering from Waseda University, Tokyo, in 2007 and 2009 respectively, and joined NTT in 2009. Since then he has engaged in research on human-computer interaction and developed vital sensing hardware at NTT Cyber Solution Laboratories and NTT Service Evolution Laboratories. In 2016, he joined 2020 Epoch-Making Project, where he has been engaged in several creative research projects combining NTT's cutting-edge technologies, design, and art to celebrate the Tokyo 2020 Games.

Junichi Nakajima

Senior Research Engineer, NTT Human Informatics Laboratories.

He received an M.Sc. in information science from the University of Tokyo in 1998. He joined NTT the same year, and engaged in research on video coding and media streaming systems. He also developed various devices such as videophone and set-top box at NTT EAST. He is currently engaged in research of Kirari!.



Masakazu Urata

Research Engineer, NTT Human Informatics Laboratories.

He received a B.E. and M.E. in electronic engineering from Osaka Prefecture University in 1990 and 1992. He joined NTT Information Networking Systems Laboratories in 1992 and engaged in R&D of intelligent network, e-Government, and smart card. He is currently engaged in research on a cloud computing platform and high-speed networks.

Kenta Ogo

Research Engineer, NTT Human Informatics Laboratories.

He graduated from the University of Tokyo. He joined NTT in 1995, where he was engaged in R&D on on-demand video distribution using Internet Protocol communication at NTT Human Interface Laboratories. He then launched an international content delivery network and large-scale commercial video-distribution service at NTT Communications. He is widely involved in the video distribution field such as program recommendations at NTT Plala. He then conducted research on a three-dimensional virtual society, location-linked tourism recommendations, ultra-high reliability 360-degree virtual-reality video distribution, and telepresence. He designed a 12K-wide long-distance maritime wireless video transmission system for the Tokyo 2020 Games.

Yuki Yoshida
Research Engineer, NTT Human Informatics Laboratories.

She received a B.Sc and M.Sc. from the University of Tsukuba. She joined NTT and engaged in R&D of human interface. In 2019, she joined 2020 Epoch-Making Project, where she has been engaged in service promotion to celebrate the Tokyo 2020 Games.

Shingo Kinoshita
Vice President, Head of NTT Human Informatics Laboratories.

He received a B.E. from Osaka University in 1991 and M.Sc. with Distinction in technology management from University College London, UK, in 2007. He joined NTT in 1991 and was a senior manager of the R&D planning section of the NTT holding company from 2012 to 2015. He is currently a visiting professor at the Art Science Department, Osaka University of Arts, and visiting executive researcher at Dentsu Lab Tokyo. He has served as a member of the Japan Science and Technology Agency (JST) JST-Mirai Program Steering Committee, member of the All Japan Confederation of Creativity (ACC) TOKYO CREATIVITY AWARDS 2021 Judging Committee, and member of the Broadband Wireless Forum Steering Committee. He has been engaged in R&D of a media-processing technology, user interface/user experience, communication protocols, information security, machine learning, service design, and technology management. Until recently, he had been in charge of NTT's Tokyo2020 initiatives, including sports-watching video technology, inclusive design for social issues, and promoting the use of ICT in *kabuki*, entertainment, and media arts such as live music.

He has been in his current position since 2021, where he manages R&D on information and communication processing of humans based on human-centered principles.

Badminton × Ultra-realistic Communication Technology Kirari!

Makoto Muto, Keisuke Hasegawa, Daichi Namikawa, Seiichi Konya, Nobuhiro Hirachi, Taiji Nakamura, Kenya Suzuki, and Shingo Kinoshita

Abstract

NTT provided technical support through Kirari!, an ultra-realistic communication technology, for the TOKYO 2020 Future Sports Viewing Project spearheaded by the Tokyo Organising Committee of the Olympic and Paralympic Games. This article presents results of demonstrations of a next-generation immersive technology for providing a sense of "being there," namely through the *sense of presence* and *sense of unity*, by transmitting holographic images of the badminton matches held at Musashino Forest Sport Plaza to a remote location (National Museum of Emerging Science and Innovation (Miraikan)).

Keywords: sports viewing, highly realistic experience, holographic video transmission

1. Overview of the TOKYO 2020 Future Sports Viewing Project

The beauty of watching performances live is being able to experience the sense of presence and sense of unity that can be felt when you are actually at the venue. In other words, it enables experiencing the sensation of having athletes or artists being right in front of you and sharing the same space with them when watching sports events or live concerts, rather than viewing them as images on a flat screen. Despite dramatic improvements in screen size, image quality, and resolution, online streaming and live viewing still can transmit only flat-screen images. It is therefore not possible to fully convey a sense of presence, such as the athlete's in-the-moment excitement, strength and beauty of the physical form, and sense of unity between the athletes and audience. For the TOKYO 2020 Future Sports Viewing Project spearheaded by the Tokyo Organising Committee of the Olympic and Paralympic Games, we carried out technology demonstrations aimed at delivering the experience of being at the badminton venue of the Tokyo 2020 Games to those who could not attend the actual event by using Kirari! ultra-realistic communication technology to transmit holographic images of the event (**Fig. 1**).

We initially planned demonstrations of live viewing by inviting the general public. However, to prevent the spread of the novel coronavirus, we cancelled the public demonstration and instead opened "The Future of Sports Viewing—Next-generation Immersive Technology Demonstration Program" at the National Museum of Emerging Science and Innovation (Miraikan) for the media from July 30 to 31, 2021.

2. System configuration

Figure 2 shows the overall system configuration for the technology demonstrations. We set up 8K cameras (**Fig. 3**) at the Musashino Forest Sport Plaza, the venue for the Tokyo 2020 Games badminton competitions, and transmitted video images taken during the badminton matches via a 1-Gbit/s network to the broadcast center. We then used the *real-time extraction of objects with arbitrary background*, a component technology of Kirari!, to selectively extract only the images of players and shuttlecocks from the transmitted 8K images at the broadcast center (**Fig. 4**). Using *highly realistic media synchronization*

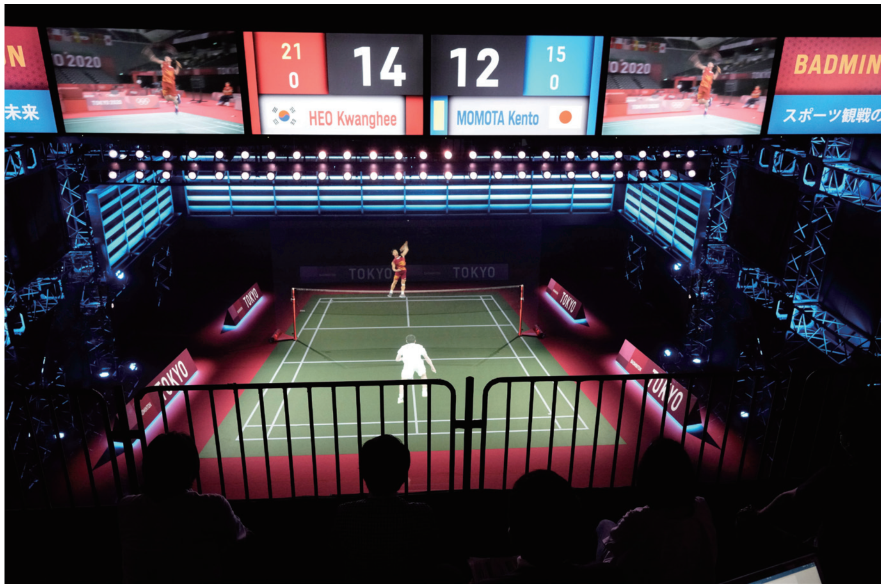


Fig. 1.   Kirari! technology demonstration experiments of the Tokyo 2020 Games badminton competition.

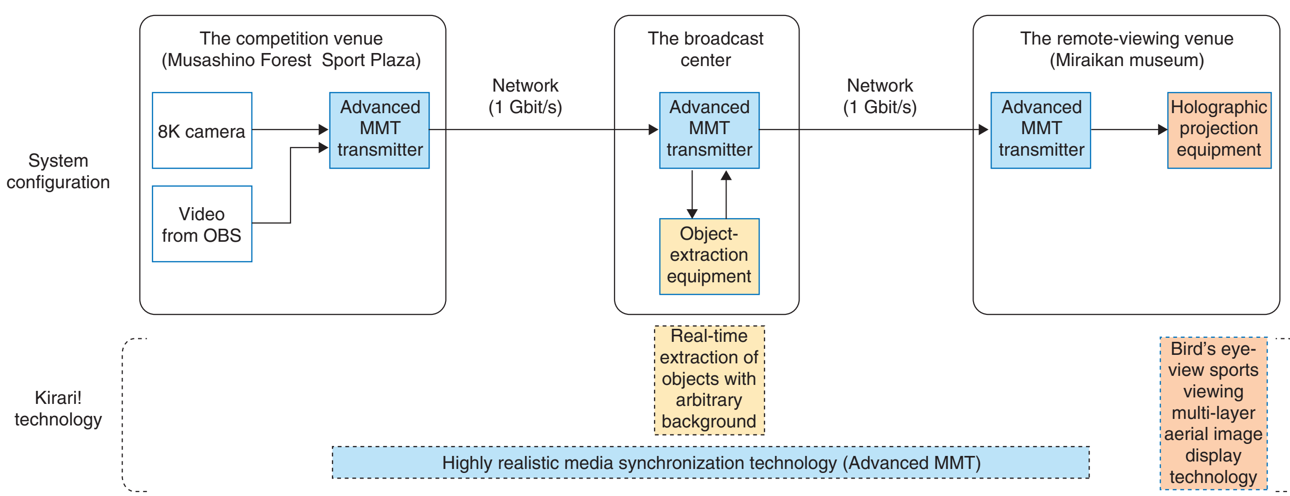


Fig. 2.   The overall system configuration.

technology (Advanced MMT (MPEG Media Transport)) [1], we then synchronously transmitted the extracted images along with multiple images that included video and audio provided by the Olympic Broadcasting Services (OBS) to the remotely located Miraikan museum.

We set up an approximately 100-seater spectator stand and a full-size court equipped with holographic projection equipment at the remote-viewing venue to recreate the actual venue (**Fig. 5**). The extracted images of players and shuttlecocks were holographically displayed by presenting players in their actual positions in front and behind the net using the *bird's eye-view sports viewing multi-layer aerial image display* technology. As a result, we were able to create a space where the badminton players appeared to be actually present in the remote-viewing venue.

Fig. 3. An 8K camera installed at the competition venue.

Fig. 4. Networking in the broadcast center.

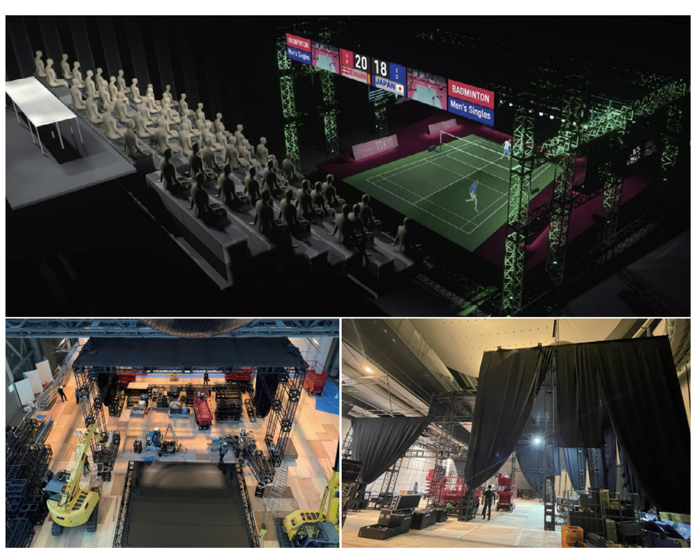


Fig. 5. The projection stage at the remote-viewing venue and state of its construction.

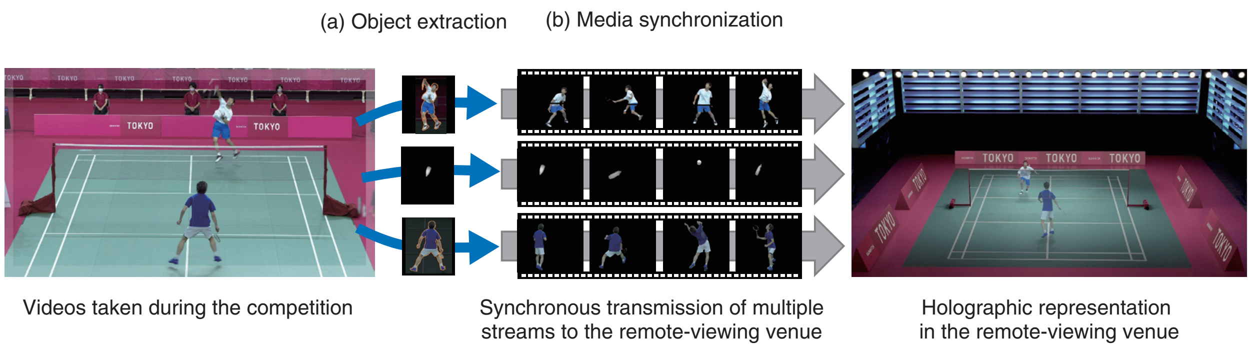


Fig. 6. Workflow of object extraction and synchronous transmission.

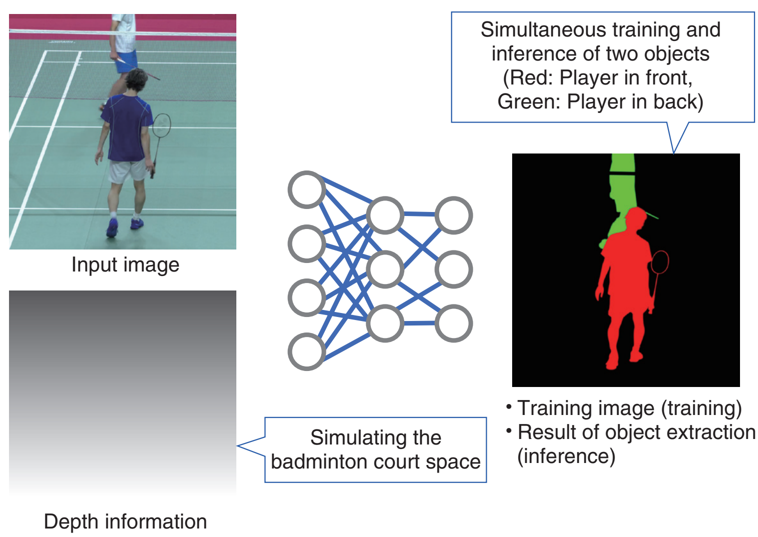


Fig. 7. Method of individual extraction of the front and back players.

3. Kirari! technologies

3.1 Real-time extraction of objects with arbitrary background

The real-time extraction of objects with arbitrary background is a technology for selectively extracting only the specific objects in images, such as players and shuttlecocks, from the videos taken during a competition [2] (**Fig. 6(a)**).

Normally, extraction of only the objects from videos requires using a green or blue background and erasing the background color by chroma keying. This technology, however, enables extracting only the objects from the raw images of the competition venue in real time without the need for any special background environment.

To apply this technology to the badminton matches, we carried out the following five improvements on our previous system [1] that is for less complicated content.

(1) Individual extraction of each player in the front and back sides of the court

Separately extracting the front and back players is not possible with the conventional technology. To achieve this, we developed a deep learning model that can simultaneously extract and infer front and back players by inputting depth information simulating the badminton court space. This enabled the extraction of individual players for a game like badminton where players are located separately in the front and back sides of the court. We therefore succeeded in stable and accurate extraction of objects (**Fig. 7**).

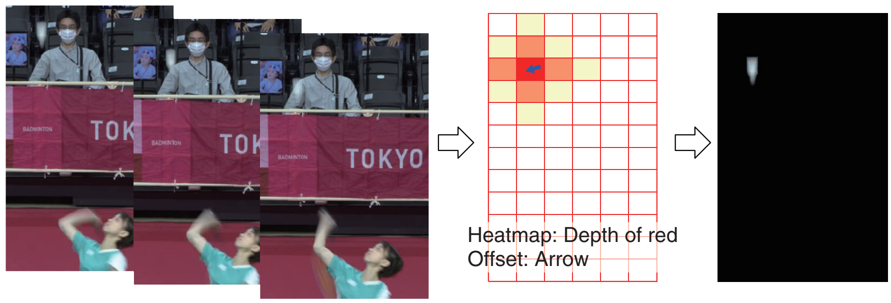


Fig. 8. Method of extracting shuttlecocks.

(2) Improvement of output-image resolution and frame rate of players' images

The conventional technology only supports resolutions of up to 4K and frame rates of up to 30 fps. The maximum resolution of player images that can be achieved using 4K cameras is only approximately 640 pixels; and the roughness of aerial images becomes evident when they are magnified and projected into life-size images. A frame rate of 30 fps does not allow tracking of rapid movements such as when the player is smashing a shuttlecock, leading to disrupted images. We therefore carried out multi-layering of the image frames to be processed and leveling of computing resources to support images from 8K, 60-fps cameras. As a result, we were able to produce smooth, high-definition images of players at 930-pixel resolution and 60-fps frame rate.

(3) Stable extraction of subtle and rapid shuttlecock movements

Since shuttlecocks are very small and move very rapidly in the video, using conventional image-recognition methods results in considerable noise and flickering and in discontinuity and incompleteness of the shuttlecock trajectory. We therefore developed extraction algorithms specific for the shuttlecock and succeeded in detecting its exact position and in accurately extracting its various contours. To detect shuttlecock position, we devised a method for learning the shuttlecock's position and movement information by inputting continuous frames in a convolutional neural network (CNN) to eliminate the effects of small objects similar to the shuttlecock appearing in the video (such as seat guide lights). The position of the shuttlecock was determined on the basis of the rough position (heatmap) and correction value (offset) obtained from the CNN (**Fig. 8**). Extraction of the shuttlecock image was carried out by generating the image using a background-subtraction method and filtering with the extracted shuttlecock position and shape and predicted shuttlecock information (position, contour, and motion blur level) in the next frame. This resulted in a precision of 90.7% and recall of 90.3%, levels that can be used for realistic sports viewing.

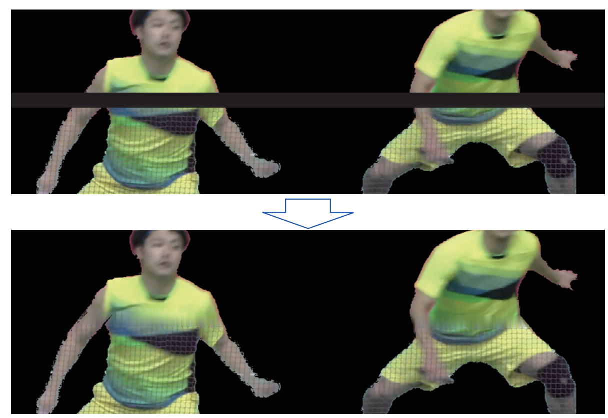


Fig. 9. Correcting the black gap zone caused by the court's net.

(4) Automatic correction of missing parts in player images

As shown in **Fig. 9**, when the player at the back of the court overlapped with the net, a black gap zone appeared on the player image. Images were corrected by inferring the missing parts on the basis of the color information above and below the gap zone.

(5) Automatic generation of player's shadow

Reproducing the players' shadows appearing in the actual venue and in the remote-viewing venue enables the creation of more natural player images. We extracted the shadows from the video based on the results of player-image extraction. We carried out this step quickly by determining the range where the

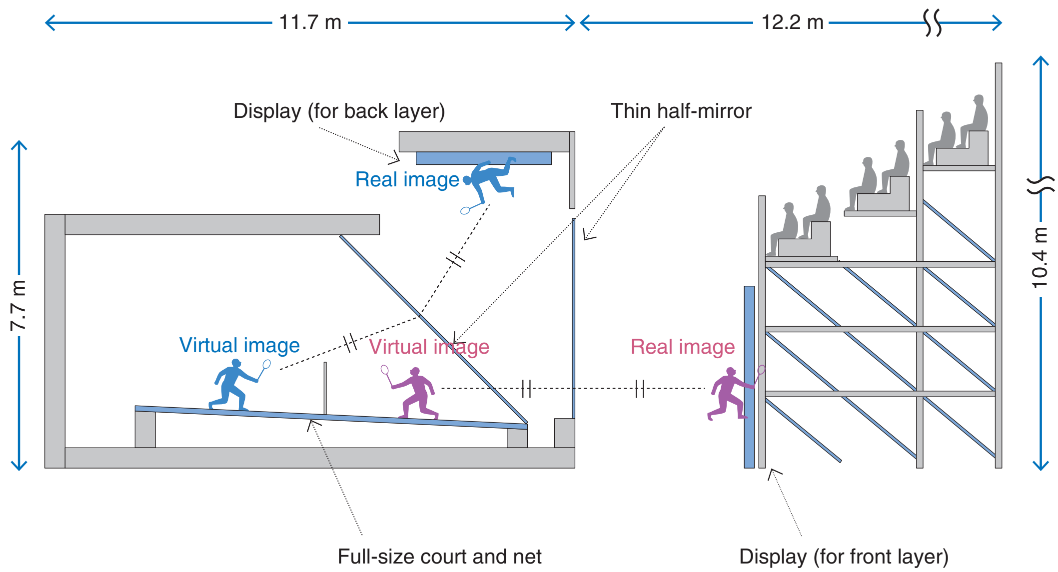


Fig. 10. Image of equipment and example of multi-layer aerial image display.

shadows might appear on the basis of the height of the player's jump and lighting conditions of the competition venue.

3.2 Highly realistic media synchronization technology

The highly realistic media synchronization technology (Advanced MMT) is NTT's proprietary technology developed by extending the media transmission standard MMT and enables the transmitting of various continuous data (streams), such as video, audio, and lighting information, while maintaining their time synchronization. For this project, we synchronously transmitted multiple streams, such as video of the competition taken at the venue, audio, images of players extracted from competition videos, extracted shuttlecock images, video obtained from OBS, etc. and displayed the necessary data on the remote-viewing venue at the appropriate timing to create a highly realistic spatial representation (**Fig. 6(b)**).

3.3 Bird's eye-view sports viewing multi-layer aerial image display technology

To deliver the experience of watching as if viewers were at the competition venue, we created a space that physically simulated the actual venue (**Fig. 10**). We set up a full-size court and net, as well as a spectator stand with the same downward viewing angle as in the actual venue. We used the bird's eye-view sports viewing multi-layer aerial image display technology for projecting the holographic representation of players at the same position within that recreated space.

With the conventional holographic projection method called Pepper's ghost, aerial images are displayed by reflecting images on the display to a half mirror tilted at a 45-degree angle. Only a single layer of an aerial image can be displayed. Single-layer display, however, cannot be used to display players in front and behind the net, as in badminton, in different locations (layers) in the projected space.

With our technology, a spectator stand at the same height as the actual competition venue was set up, and the locations of the court, net, two half-mirrors, light-emitting diode displays, and the projectors were optimized to create a realistic appearance of the players in front and behind the net. This enabled the holographically displaying of the two players at their correct positions.

The use of two half-mirrors, however, results in a discontinuous appearance of the shuttlecock, which moves back and forth between the two mirrors. Since this problem is a structural issue arising from having two physically separated layers, there is no perfect solution to it. To produce a more natural appearance, we repeatedly conducted trial-and-error experiments on the timing for switching the shuttlecock's display layers and came to the conclusion that the best timing for switching is when the player hits the shuttlecock. To determine the best timing for switching, we

Table 1. Impressions of participants.

Viewed in full scale, the sense of presence is completely different.
When the shuttle goes up high, it may be a little difficult to look where the shuttle is. However, it feels as if the player has warped to the spot.
We can enjoy the real feeling of being at the match venue.
The full scale players were so impressive that we cheered as if being at the match venue unintentionally.

combined multiple approaches, such as determining the shuttlecock position through analysis of lateral court images, determining the timing when the shuttlecock is hit through detection of hitting sounds from the audio during the match, and human operators watching the game manually switching the display layers. The combination of multiple approaches enabled the real-time switching of shuttlecock-image display positions.

4. Results of demonstrations

The technology demonstrations were conducted for four days from July 30 to August 2 during the badminton tournament finals. Press representatives participated in the viewing demonstrations for not only the recorded videos of the preliminary rounds but also for the live coverage of the men's doubles, women's singles, and men's singles events. Many participants commented that they were able to gain a highly realistic experience of the matches (**Table 1**).

The highly realistic environment with a full-size court and net and life-size holographic projections of players apparently elicited a sensory illusion that the players were actually in front of the participants. This indicates the possibility of achieving an emotional connection at a deeper level than when watching on television. The spectators were able to more easily feel the energy and vibe during the competitions or the excitement of players before the start of the matches, and even the emotions associated with winning or losing. When Kento Momota fell to his knees when he lost in the second round of his match during the group play stage, the realistic sense of his presence in the remote-viewing venue was very striking.

In terms of object-extraction performance, we succeeded in real-time processing of 8K, 60-fps videos. There are issues that need to be addressed for extraction accuracy such as when images of the two players overlap or extraction for doubles matches. Nevertheless, we succeeded in processing the images without missing parts or errors that could hinder proper viewing of the games, even for players wearing different uniforms for all the six live broadcast experiments. We also confirmed that there were no errors in the correction of the black zones caused by the net and in the addition of shadows. For the shuttlecock images, however, more improvements are necessary due to considerable discontinuity depending on the viewing position.

Table 2. Details of lag time.

Camera system	≤ 17 ms (≤ 1 frame)
MMT transmission	1600 ms (4K 59.94p BT.709, rate of encode: 40 Mbit/s, FEC: 5%)
Extraction of objects	1000 ms (including input/output buffering)
Presenting system	~ 120 ms (4 frames at 30 fps)

In terms of transmission performance, the total end-to-end delay from the competition venue to the remote-viewing venue was less than 2800 ms. **Table 2** lists the lag time details. The MMT transmission delay included network transmission delays of approximately 1 ms between the competition venue and broadcast center and approximately 0.1 ms between the broadcast center and remote-viewing venue. Delays in object-extraction processing included buffering that takes fluctuations in processing time, etc., into consideration, but the actual delay in extraction processing excluding the buffering was less than 400 ms.

Since the demonstrations involved one-way transmission of images from the competition venue to the remote-viewing venue, an approximately 3-s delay did not cause major problems. However, our future goal is to connect the two venues bi-directionally and deliver the cheers from the audience in the remote-viewing venue back to the competition venue, in which case, significant lag times would become problematic, making it necessary to further reduce processing time.

We were able to confirm that the above performance of the demonstration system were sufficient in delivering a realistic experience of the presence of the athletes in the remote-viewing venue.

5. Summary

As part of the TOKYO 2020 Future Sports Viewing Project, we applied Kirari! to badminton matches and demonstrated that it can be used to deliver a realistic experience of being at the competition venue. Although there are issues regarding accuracy of extracting player images and in the method for displaying the shuttlecock, we succeeded in demonstrating the possibility of a new sports-viewing experience, beyond what television can provide, for the Olympic Games, an event that attracts attention from all over the world.

Going forward, we will carry out improvements in extraction performance and other technical aspects, as well as deploy the technology to other games and areas, such as for music concerts. We will also study methods of minimizing lag time for bi-directional connection between competition and remote-viewing venues and conduct research and development on Kirari! and other technologies for achieving the Remote World and eventually make future proposals that will surprise the world using communication technologies

Acknowledgments

We would like to thank the members of the Innovation Promotion Office of the Tokyo Organising Committee of the Olympic and Paralympic Games for helping us resolve various issues in conducting this technology demonstration, representatives of the Miraikan museum for organizing and leading the project, and representatives of partner companies for their technical support.

NTT is an Olympic and Paralympic Games Tokyo 2020 Gold Partner (Telecommunication Services).

References

[1] Y. Tonomura, H. Imanaka, K. Tanaka, T. Morizumi, and K. Suzuki, “Standardization Activities for Immersive Live Experience (ILE),” Journal of the ITU Association of Japan, Vol. 47, No. 5, pp. 14–17, 2017 (in Japanese).

[2] H. Kakinuma, G. 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

Makoto Muto

Senior Research Engineer, NTT Human Informatics Laboratories.

He joined NTT in 2001 and engaged in research on information retrieval. He is currently researching segmentating subjectives from video for sport public viewing applications.

Keisuke Hasegawa

Engineer, NTT Human Informatics Laboratories.

He received a B.E. and M.E. in informatics from Kyoto University in 2012 and 2014. He joined NTT WEST in 2014 and engaged in the maintenance of network facilities. In 2016, he joined NTT Service Evolution Laboratories, where he has been engaged in research on media processing for Kirari!. He is a member of the Institute of Image Information and Television Engineers (ITE).

Daichi Namikawa

Senior Research Engineer, NTT Human Informatics Laboratories.

He received a B.E. and M.E. from Kagoshima University in 2007 and 2009. He joined NTT in 2009 and has been researching and developing enhanced network services and systems.

Seiichi Konya

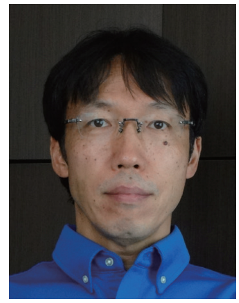
Research Engineer, NTT Human Informatics Laboratories.

He received a Master of Electronic Engineering from Hokkaido University in 1990 and joined NTT Information and Communication Laboratories. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE).

Nobuhiro Hirachi

Senior Research Engineer, NTT Human Informatics Laboratories.

He received an M.Sc. in natural science from Chiba University in 1998. He joined NTT in the same year and studied fiber optic access networks including optical access systems. He has developed an asynchronous-transfer-mode-based passive optical network system for the first of fiber-to-the-home services and received the President's commendation. He has been engaged in research of Kirari! since 2019.



Taiji Nakamura

Senior Research Engineer, 2020 Epoch-Making Project, NTT Human Informatics Laboratories.

He joined NTT DATA Communications Systems (now NTT DATA) in 1991. He has over 20 years' experience in information systems planning and development in the national public sector.

Kenya Suzuki

Senior Research Engineer, Supervisor, NTT Human Informatics Laboratories.

He received a B.E. in electro-communications from the University of Electro-Communications, Tokyo, in 1993 and M.E. in electronic engineering from the University of Tokyo in 1995. He joined NTT Human Interface Laboratories in 1995, where he has been engaged in research on human-computer interaction and multimedia applications including web-based signage system development. He is currently interested in media art projects such as Kirari!. He is a member of IEICE of Japan and the Physical Society of Japan.

Shingo Kinoshita

Vice President, Head of NTT Human Informatics Laboratories.

He received a B.E. from Osaka University in 1991 and M.Sc. with Distinction in technology management from University College London, UK, in 2007. He joined NTT in 1991 and was a senior manager of the research and development (R&D) planning section of the NTT holding company from 2012 to 2015. He is currently a visiting professor at the Art Science Department, Osaka University of Arts, and visiting executive researcher at Dentsu Lab Tokyo. He has served as a member of the Japan Science and Technology Agency (JST) JST-Mirai Program Steering Committee, member of the All Japan Confederation of Creativity (ACC) TOKYO CREATIVITY AWARDS 2021 Judging Committee, and member of the Broadband Wireless Forum Steering Committee. He has been engaged in R&D of a media-processing technology, user interface/user experience, communication protocols, information security, machine learning, service design, and technology management. Until recently, he had been in charge of NTT's Tokyo2020 initiatives, including sports-watching video technology, inclusive design for social issues, and promoting the use of ICT in *kabuki*, entertainment, and media arts such as live music.

He has been in his current position since 2021, where he manages R&D on information and communication processing of humans based on human-centered principles.

Marathon × Ultra-low-latency Communication Technology

Soichiro Usui, Shinji Fukatsu, Eiichiro Matsumoto, Maiko Imoto, Daisuke Shirai, and Shingo Kinoshita

Abstract

NTT provided ultra-low-latency communication technology to the Tokyo 2020 Real-time Remote-cheering Project, which was implemented by the Tokyo Organising Committee of the Olympic and Paralympic Games for the marathons of the Olympic and Paralympic Games Tokyo 2020 held in Sapporo on August 7 and 8, 2021. This project proposed a new means of watching sports by connecting the marathon course in Sapporo with the cheering venue in Tokyo in real time. By delivering the cheers of spectators to the athletes from the remote location in Tokyo, it was possible to create a sense of unity between the athletes and spectators, as well as a sense of realism similar to roadside cheering while maintaining safety and security. Our efforts concerning this project and the technologies that we used are introduced in this article.

Keywords: ultra-low-latency communication technology, remote cheering, IOWN

1. Overview of the project

During the novel coronavirus (COVID-19) pandemic, it has become increasingly difficult to cheer on athletes directly at competition venues. By using new communication technologies, the Tokyo 2020 Real-time Remote-cheering Project aimed to create a world in which athletes at a competition venue and spectators in a remote area can share the excitement, emotion, and sense of unity at the venue (**Fig. 1**).

As a Gold Partner of the Olympic and Paralympic Games Tokyo 2020, NTT has been contributing to the creation of the most-innovative Games in history by providing cutting-edge communication technology. In light of the situation concerning the COVID-19 pandemic, we adopted a new mission to propose a new form of sports viewing using communication services and make it a springboard for the future and decided to provide the technology for this project.

Regarding remote spectator experience, delay time is the largest challenge in transmitting cheering from remote venues to the competition venue. For marathons in particular, even a small delay can have a significant effect on the ability to transmit cheering to athletes running at 5 m/s. The total delay in watching a sporting event remotely had been several seconds each way, which includes not only the propagation delay of light but also media-processing delay such as transmission-processing delay and compression delay for video information; as a result, it has been impossible to transmit cheering to players, athletes, etc. without a delay. NTT undertook the Tokyo 2020 Real-time Remote-cheering Project with the goal of solving these problems by using its ultra-low-latency communication technology to minimize the latency of the transmission process to about 100 ms each way and ensure the transmission of cheering to the athletes.

2. Device configuration of the project

The overall device configuration of the project is shown in **Fig. 2**. We aimed to create a space in which remote spectators could feel as if they were cheering along the roadside as the runners were running by. Therefore, we installed light-emitting diode (LED) displays (approximately 50 m wide and 2 m high) at both the Sapporo marathon course (in front of Sapporo

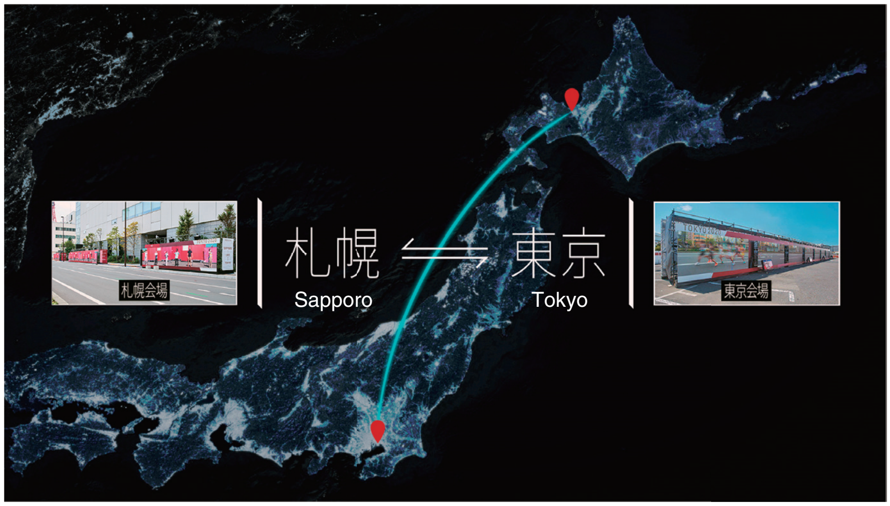


Fig. 1. Overview of the Tokyo 2020 Real-time Remote-cheering Project.

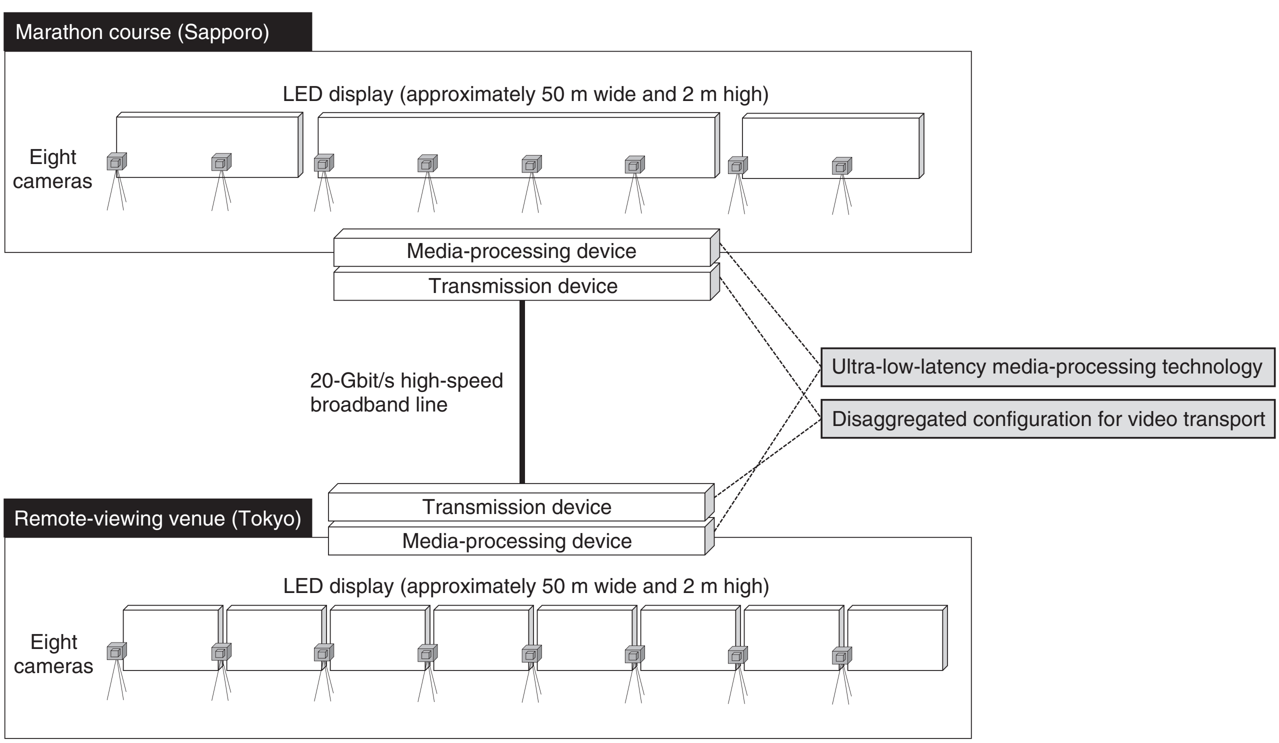


Fig. 2. Device configuration of the project.

Sosei Square) and Tokyo remote-viewing venue. The displays enabled the runners and spectators to see each other in actual size. The total width of the LED displays at both sites was approximately 50 m, but the location and spacing of the displays varied in accordance with the conditions at the installation site (**Figs. 3(a)** and **(b)**). Eight cameras were set up in front of the LED displays at both sites to capture images of spectators and runners (**Fig. 3(c)**). The 4K images from each of the two sites were transmitted to

(a) 50-m-wide LED display installed at the Tokyo venue

(b) 50-m-wide LED display installed along the Sapporo marathon course (in front of Sapporo Sosei Square)

(c) Eight cameras and microphones installed along the marathon course

Fig. 3. The setup at both venues.

Fig. 4. NTT network equipment that provided ultra-low-latency communication technology.

the other site via a 20-Gbit/s high-speed broadband line via a media-processing device using ultra-low-latency media-processing technology and transmission device using disaggregated configuration for video transport. These two technologies constitute our ultra-low-latency communication technology.

3. Ultra-low-latency communication technology

The Innovative Optical and Wireless Network (IOWN) is an innovative concept proposed by NTT as a next-generation communication platform. IOWN will enable the creation of a network and information-processing infrastructure that includes devices capable of providing high-speed, high-capacity communications and vast computing resources by using innovative technologies centered on optical technology. One of the components of IOWN, the All-Photonics Network (APN), will achieve ultra-low-power consumption, ultra-high capacity, and ultra-low latency by introducing photonics-based technology into everything from the network to devices. The ultra-low-latency communication technology used in this project uses two APN elemental technologies: disaggregated configuration for video transport and ultra-low-latency media-processing technology (**Fig. 4**).

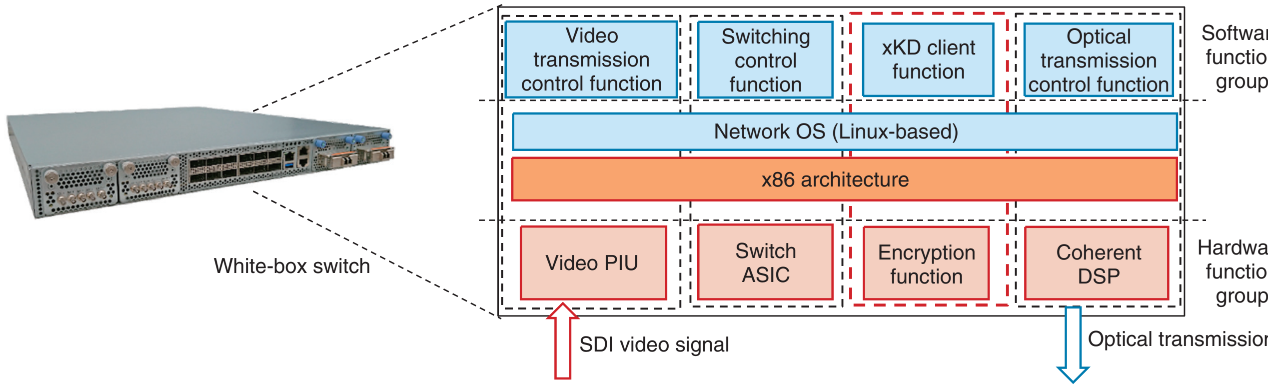


Fig. 5. Disaggregated configuration for video transport.

3.1 Disaggregated configuration for video transport

Disaggregated configuration for video transport enables flexible changes in configuration, addition of functions, and cost reduction by separating network functions that were previously provided in combination and configuring them so that each function can be controlled by standardized interfaces. We developed a plug-in unit (video PIU) in which a direct serial digital interface (SDI)-signal acquisition function is added to a white-box switch with an optical transponder for long-distance transmission. The new video PIU enables direct transmission of uncompressed video and audio in SMPTE ST 2110* format over optical long-distance transmission lines. As a result, the delay between the video input on the transmitter side and video output on the receiver side was reduced to about 1 ms, and the one-way delay (including the distance delay between Tokyo and Sapporo) was reduced to about 20 ms. We also implemented a set of functions for video routing in the network operating system, which is the control software for the white-box switch, to achieve integrated operation of network and video devices (**Fig. 5**).

3.2 Ultra-low-latency media-processing technology

Ultra-low-latency media-processing technology captures SDI video signals output from a camera on a subframe or line basis and executes video processing such as geometric transformation and composition at the video-signal level. Implementing each video process at the video-signal level instead of at the frame level makes it possible to reduce the frame latency required by conventional video processing and achieve low latency (**Fig. 6**).

In this project, when transmitting video from multiple cameras, the SDI video signals from each camera were made into subframes in line units, and the received subframes were immediately separated and displayed on multiple displays in a manner that achieved even lower latency (**Fig. 7**).

4. Results of the project

On the days of the men's and women's marathons, after a demonstration of low latency was conducted, the remote spectators cheered for the runners. Via the 50-m-wide display, the Tokyo venue provided a sense of realism to the spectators, namely, the feeling that the athletes were running right in front of them, and the spectators could show their support by clapping and waving flags. For past remote-cheering experiences, due to delays, the timing of the cheering was not synchronized with the athletes, so the spectators found it difficult to feel a sense of unity while cheering for the athletes. For the Sapporo marathon course, however, the ultra-low-latency communication technology enabled the spectators to see their cheering reaching the runners as the runners passed by. As a result, the spectators were able to feel the same sense of unity even in the Tokyo venue, which made them

* SMPTE ST 2110: A standard developed by the society of motion picture and television engineers (SMPTE) for transmitting video over Internet Protocol (IP) networks for the professional media industries.

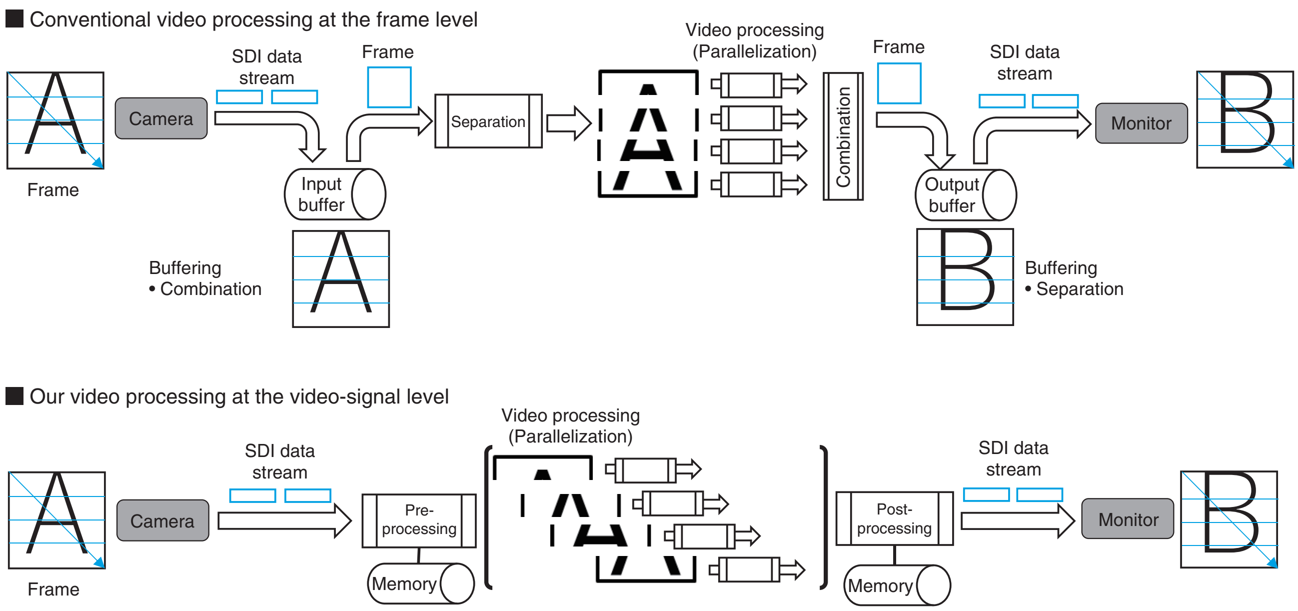


Fig. 6. Ultra-low-latency media-processing technology.

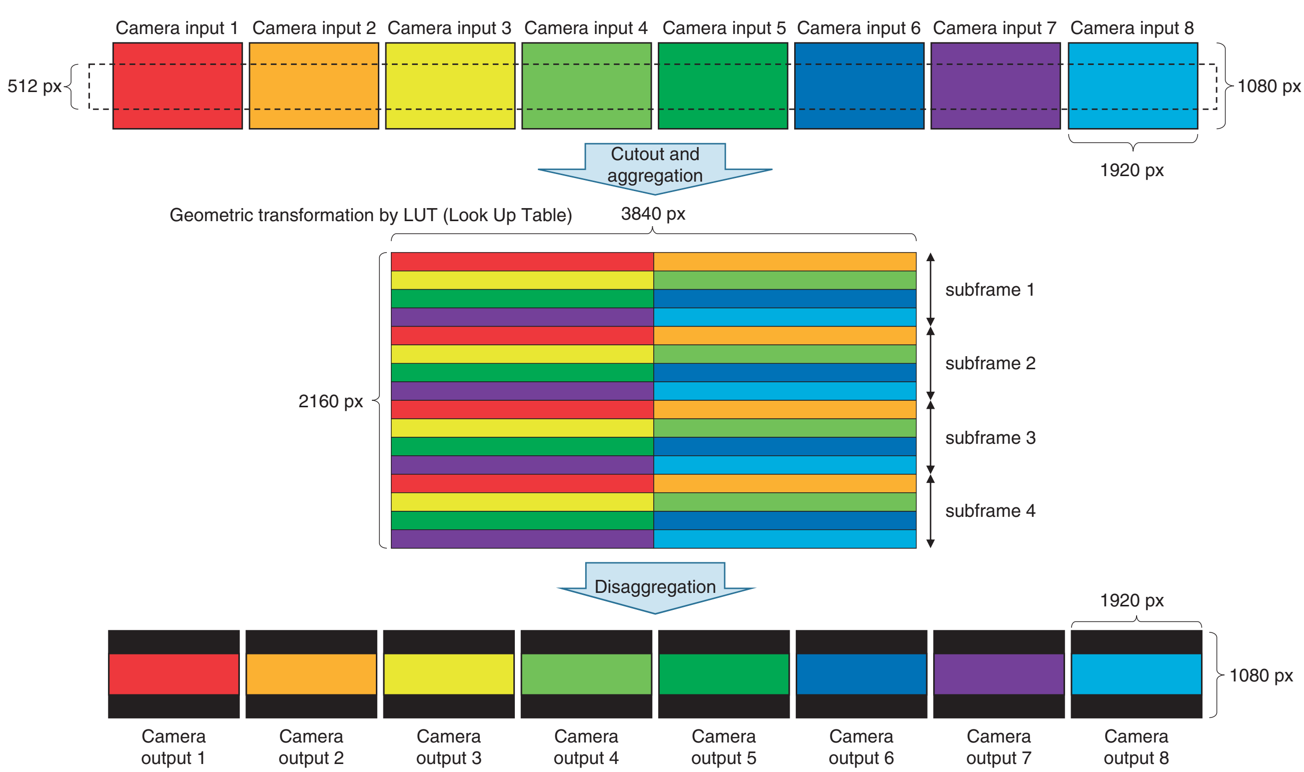


Fig. 7. Aggregation and disaggregated processing of multiple camera images.

(a) Cheering from Tokyo to Sapporo during the women's marathon

(b) Cheering from Tokyo reaching the runners during the women's marathon

(c) Cheering from Tokyo to Sapporo during the men's marathon

(d) Cheering from Tokyo reaching the runners during the men's marathon

Fig. 8. Scenes from the project during the marathons.

even more enthusiastic about cheering (**Figs. 8(a)** and **(b)**).

Through this cheering experience, the spectators at the Tokyo venue commented, "I could experience the same sense of realism and speed as cheering at the roadside," "I felt that I could cheer from a position closer to the athletes than possible at the roadside, which is usually crowded," and "Above all, it was lovely to see the cheering reach the athletes." In other words, the spectators seemed to be very satisfied with the sense of unity by having their cheering reach the marathon venue, an outcome that had been impossible with the remote cheering. The significance of this project was also confirmed by the fact that some of the marathon runners showed interest in the real-time cheering from the remote area via the display, and that result was a world's first. As expected, the delay time for the transmission process was kept to 100 ms each way. Unfortunately, typhoon Mirinae was approaching Japan, and the men's marathon was held in rainy weather on August 8th; nevertheless, after ensuring safety of the athletes and spectators, we were able to transmit cheering from Tokyo to Sapporo in spite of the rain (**Figs. 8(c)** and **(d)**).

5. Future developments

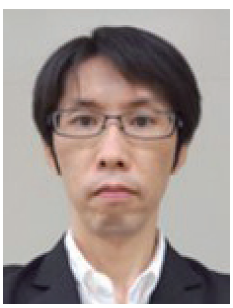
In light of the results of this project, NTT aims to create a new style of watching sports competitions that allows people who are unable to cheer at a competition venue to experience the sense of presence and unity, which could only be experienced at the venue, in a safe and secure manner. Regarding the Tokyo 2020 Real-time Remote-cheering Project, the cheering was transmitted from a special venue in Tokyo. We will accelerate our research and development so that our ultra-low-latency communication technology can be widely used and the same real-time remote cheering will be possible from homes.

NTT is an Olympic and Paralympic Games Tokyo 2020 Gold Partner (Telecommunication Services).

Soichiro Usui

Senior Manager, Strategic Business Creation Team, Research and Development Planning Department, NTT.

He received a B.S. in industrial engineering and management from Tokyo Institute of Technology in 1999. He joined NTT the same year and was involved in corporate business and network services development at NTT EAST. He was with NTT Research and Development Planning Department from 2012 to September 2021, where he was in charge of information and communication technology (ICT) business creation for the entertainment and sports domain. He has been with NTT EAST since October 2021.



Shinji Fukatsu

Senior Research Engineer, Supervisor, Cyber-World Laboratory, NTT Human Informatics Laboratories.

He received a Ph.D. in engineering from Osaka University in 2002 and joined NTT the same year. He has been engaged in research and development (R&D) of human interfaces and video streaming services as well as in planning and development of video streaming services at NTT Plala and in promoting standardization and international development of ICT at the Ministry of Internal Affairs and Communications. He is currently engaged in R&D of remote-world infrastructure technology.

Eiichiro Matsumoto

Senior Research Engineer, Cyber-World Laboratory, NTT Human Informatics Laboratories.

He received a B.E. in science engineering (information science) from Meiji University in 1999 and joined NTT the same year, from August he was engaged in the development of geographic information systems and voice/video communication systems at NTT EAST. From 2008 to 2010, he was involved in the R&D of IP television (TV)/mobile broadcasting standards and the formulation of technical standards at NTT. Since 2010, he has been engaged in the planning and development of network services and systems for video distribution systems, such as IPTV and radio-frequency TV and researching video-transmission-system technologies for adding high value to relay networks and access networks at NTT EAST. He has been at his current position since 2020.

Maiko Imoto

Research Engineer, Cyber-World Laboratory, NTT Human Informatics 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 for mobile phones and HTML5-based e-learning applications. She was involved in R&D on Kirari! immersive telepresence technology. From 2017 to 2020, she was responsible for the development of live-video distribution platforms for broadcasters and creating live viewing services for sports and entertainment at NTT Plala. She has been engaged in R&D of multimedia data interactive transmission technologies.

Daisuke Shirai

Senior Research Engineer, Supervisor, Frontier Communication Laboratory, NTT Network Innovation Laboratories.

He received a B.E. in electronic engineering, M.E. in computer science, and Ph.D. in media design from Keio University, Kanagawa, in 1999, 2001, and 2014. He pioneered the world's first 4K JPEG 2000 codec system, which enables low latency 4K60p video transmission on a Gigabit network. He has applied his expertise across multiple domains through his study of practical applications in digital audio and video broadcasting technology, image coding, information theory, networking, human-computer interaction, and software architecture. His current research topics include remote video production networks, ultra-low latency visual communication, and its security over optical transport networks.

Shingo Kinoshita

Vice President, Head of NTT Human Informatics Laboratories.

He received a B.E. from Osaka University in 1991 and M.Sc. with Distinction in technology management from University College London, UK, in 2007. He joined NTT in 1991 and was a senior manager of the R&D planning section of the NTT holding company from 2012 to 2015. He is currently a visiting professor at the Art Science Department, Osaka University of Arts, and visiting executive researcher at Dentsu Lab Tokyo. He has served as a member of the Japan Science and Technology Agency (JST) JST-Mirai Program Steering Committee, member of the All Japan Confederation of Creativity (ACC) TOKYO CREATIVITY AWARDS 2021 Judging Committee, and member of the Broadband Wireless Forum Steering Committee. He has been engaged in R&D of a media-processing technology, user interface/user experience, communication protocols, information security, machine learning, service design, and technology management. Until recently, he had been in charge of NTT's Tokyo2020 initiatives, including sports-watching video technology, inclusive design for social issues, and promoting the use of ICT in *kabuki*, entertainment, and media arts such as live music.

He has been in his current position since 2021, where he manages R&D on information and communication processing of humans based on human-centered principles.

High-accuracy Time-synchronization Technology for Low-latency, High-capacity Communications in the 5G and Beyond 5G eras

Kaoru Arai and Makoto Murakami

Abstract

High-accuracy time-synchronization technology has been attracting attention for enabling services that use high-accuracy time information, such as fifth-generation (5G) mobile communications, high-frequency financial transactions, high-definition video distribution, and smart grids. The International Telecommunication Union - Telecommunication Standardization Sector (ITU-T) Study Group (SG) 15 and the Institute of Electrical and Electronics Engineers (IEEE) mainly carry out international standardization. In this article, we introduce the time-synchronization requirements for 5G as well as next-generation communication technologies, such as Beyond 5G and quantum communications, an overview of time-synchronization technology/Precision Time Protocol (PTP), the standards of which have recently been updated by IEEE, and the latest discussion topics in ITU-T SG15.

Keywords: time synchronization, PTP, 5G/Beyond 5G

1. Background of time-synchronization technology

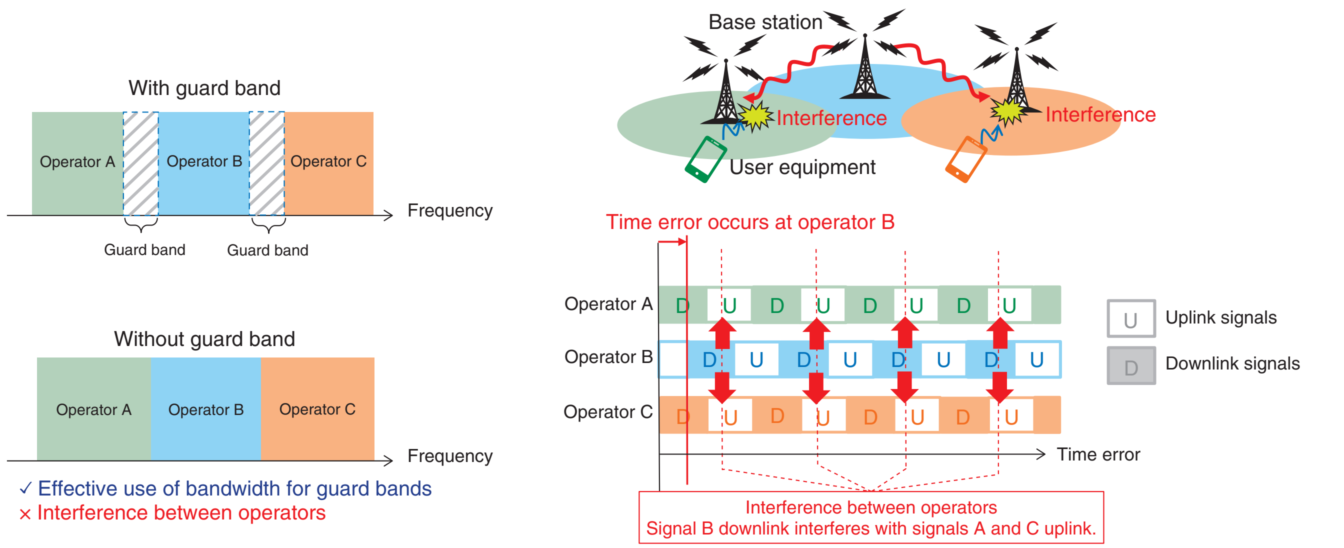
Time synchronization means that the time between systems in a communications network is the same and is consistent with Coordinated Universal Time (UTC). Time synchronization is required for fourth-generation (4G) and later mobile communications, and technology has been developed to deliver UTC-synchronized time information to mobile base stations. As information technology (IT) services have become more diversified and higher quality, the application of time information has expanded beyond the mobile field. Examples include high-accuracy time stamping in high-frequency trading in the financial sector, synchronization in distributed processing among datacenters, frame synchronization in video streaming, and time synchronization among instruments in large-scale scientific experiments such as accelerators in atomic physics.

Time-synchronization technology plays an important role in various industries and scientific fields, as described above, is discussed in many forums and standardization organizations, and is being commercialized and introduced into services. The International Telecommunication Union - Telecommunication Standardization Sector (ITU-T) has been discussing the requirements for time synchronization in communication networks and specifications for high-accuracy time-synchronization devices.

2. Use cases and requirements of time synchronization in telecommunications

We first introduce the use cases of time synchronization in telecom networks that are under consideration by ITU-T and example requirements for mobile and industrial applications.

In 4G and later mobile communications, the same frequency is time-shared and transmitted for both



(a) Frequency-slot assignment example

(b) Interference with no guard band between operators

Fig. 1. Radio-wave interference between operators due to time-synchronization errors.

uplink and downlink communications (i.e., time division duplex (TDD)) to effectively use the bandwidth of mobile base stations. For TDD, a time-synchronization accuracy of +/–1.5 µs is required. However, to maximize the use of the radio frequency band allocated to each telecom operator in 4G/5G communications, a system without inter-operator guard bands is used. Therefore, if the time-synchronization error increases, there are concerns about radio-wave interference between operators (**Fig. 1(a)**). This suggests that if the time error between certain operators increases, the uplink and downlink frame patterns may be reversed, and interference may occur with other neighboring operators (**Fig. 1(b)**). In Japan, 4G and later operate without guard bands; hence, there are concerns about the effects of interference due to time-synchronization errors [1]. To avoid this problem, it is necessary to synchronize to a common time standard (e.g. GPS (Global Positioning System)) among operators and unify the uplink and downlink TDD frame structure. NTT clarified these interference mechanisms and detailed conditions of their occurrence due to time-synchronization errors, proposed consideration of them in ITU-T Study Group (SG) 15, and reflected them in requirements for time synchronization in Recommendation G.8271 at the January 2020 meeting.

The Institute of Electrical and Electronics Engineers (IEEE) has standardized time-sensitive networking (TSN) as an industrial Ethernet standard to enable real-time monitoring and control with low latency and time synchronization and has specified standards for time synchronization in IEEE802.1AS. One proposed use case is to transmit time information over the 5G network of a telecom operator and deliver it to industrial applications assumed for TSN (**Fig. 2**). At the ITU-T SG15 meeting in April 2021, a budget of 900 ns was proposed as the time accuracy required for the 5G network specified by ITU-T. Details of the requirement will be discussed in the future.

As future use cases of time synchronization, applications to next-generation communications, such as quantum communications and quantum cryptography, are also being studied. In ITU-T, standardization of quantum key distribution (QKD) is actively being discussed. One of the methods of QKD requires high-accuracy time synchronization to detect weak photon signals from the sender at the receiver with accurate timing. Certain telecom operators are also interested in time synchronization in QKD, and requirements for synchronization technologies for these systems are expected to be discussed in the future.

3. Enhancement of time synchronization—New PTP standards

Next, we introduce trends in Precision Time Protocol

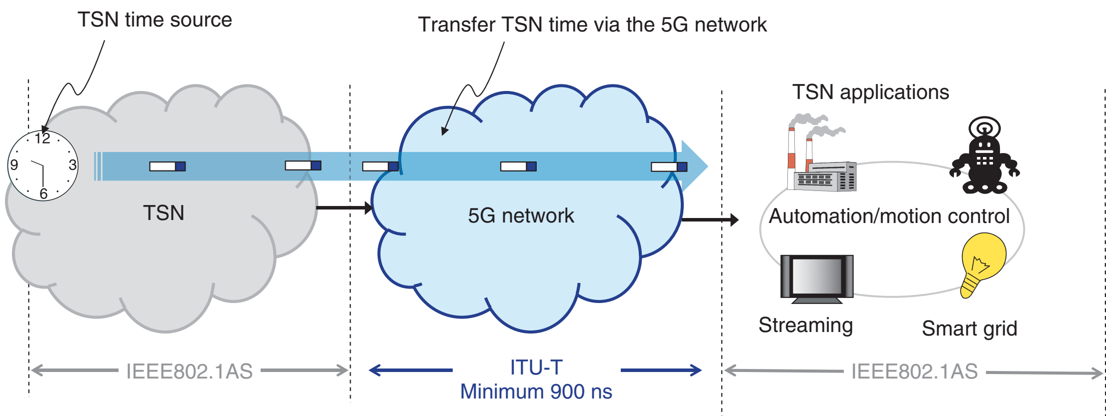
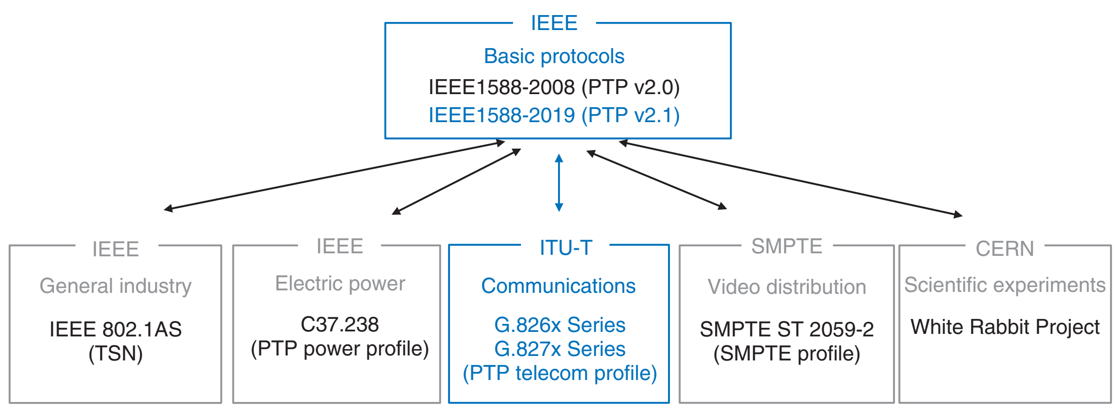


Fig. 2. Transmission of TSN time information via 5G network.



SMPTE: Society of Motion Picture and Television Engineers
CERN: European Organization for Nuclear Research

Fig. 3. Relationships among standardization bodies for PTP.

(PTP), which is a protocol to achieve high-accuracy time synchronization. PTP synchronizes time between two systems by exchanging dedicated packets that are embedded time information [2]. The basic PTP has been standardized as IEEE1588. On the basis of IEEE1588, standardization organizations and research institutes in various industrial fields maintain compatibility and extend the protocol according to requirements for various usage scenes (**Fig. 3**). The ITU-T supervises the communications field and defines PTP specifications for communications networks to meet requirements of the mobile applications discussed in the 3rd Generation Partnership Project (3GPP). IEEE supervises the industrial field and is expected to support private 5G[*1] by using the TSN standards mentioned earlier.

The current PTP was standardized as IEEE1588-2008 (Version 2.0) in 2008. Its extended specification was standardized as PTP telecom profile by ITU-T. It was then introduced to commercial communication devices and deployed to telecom networks worldwide. A new standard was also completed as IEEE1588-2019 (Version 2.1) and released in 2020 [3]. Version 2.1 is backward compatible with Version 2.0, although the standard documentation has been significantly increased from about 300 pages in

*1 Private 5G: Differing from 5G provided by telecom operators, these are 5G systems built and provided locally by companies and local governments according to the individual requirements of industries.

Table 1. Major additional functions in IEEE1588-2019.

Main perspectives	Additional functions	Major changes and additions
Flexibility	Mixed packet forwarding methods	Multicast/unicast mixable
Robustness	Multi-profile	Multiple profiles superimposable on the same network
	PTP domain redundancy	Multiple domain numbers can be specified for a PTP device
	Security	Functions added to check PTP packet for abnormalities
	Monitoring	Added slave port monitoring function
Accuracy	High-accuracy profile	Added parameters to achieve sub-nanosecond accuracy
	Pre-calibration	Added calibration function for delay asymmetry

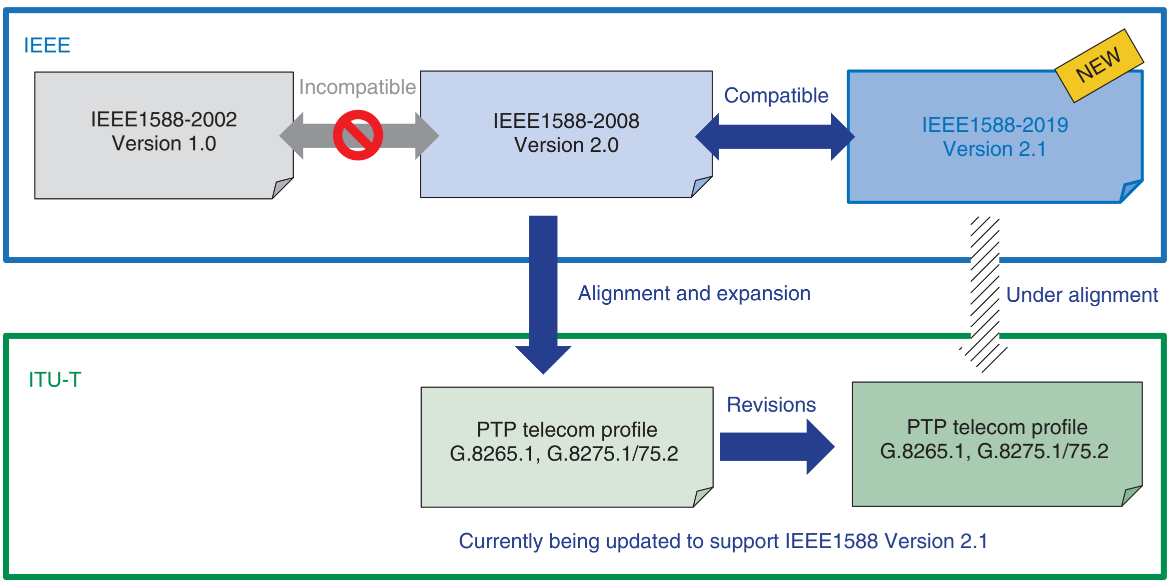


Fig. 4. Evolution of IEEE and ITU standards for PTP.

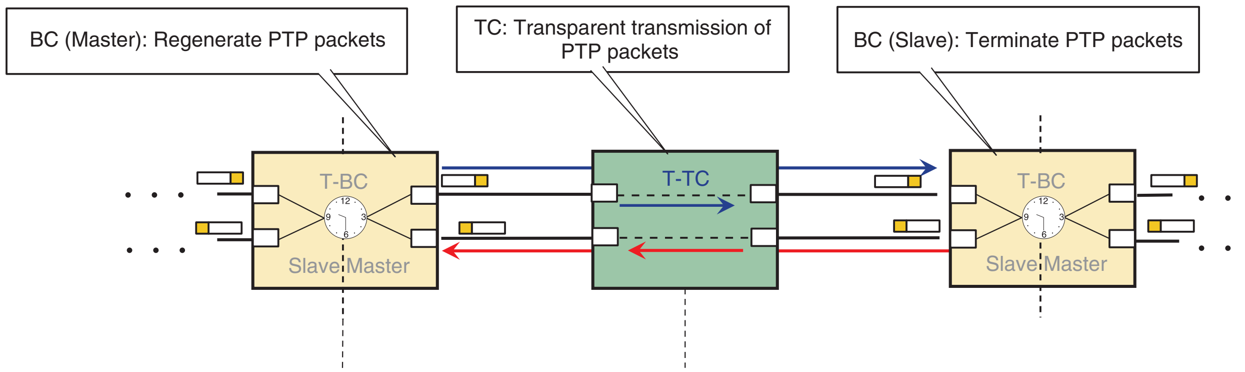
Version 2.0 to about 500 pages in Version 2.1 as many new functions have been incorporated.

In response to the increasing importance of time synchronization in applications and the emergence of high-accuracy time-synchronization requirements in various fields, IEEE1588-2019 was expanded from the previous PTP standards mainly in terms of flexibility, robustness, and accuracy (**Table 1**). Regarding flexibility, the PTP packet forwarding method allows both unicast and multicast packets to be mixed. In terms of robustness, features related to security functions, quality monitoring of PTP receiving ports, and redundancy of PTP domains[*2] were added to enhance the protocol's resistance to abnormal operations. Accuracy has been improved from the nanosecond level to sub-nanosecond (in the order of 0.1 ns) level, and a calibration function was added for optical path differences between uplink and downlink of optical fibers and time-stamp errors between egress and ingress in a PTP device, which affect synchronization accuracy. To achieve compatibility between IEEE1588-2019 and the PTP telecom profile, ITU-T is in the process of updating the recommendations (**Fig. 4**). A number of telecom profile parameters were added and modified to comply with IEEE1588-2019 at the April 2021 meeting.

4. Improving the accuracy of time synchronization—PTP devices

ITU-T has been studying to improve the accuracy

*2 PTP domain: A logical group of PTP devices belonging to a single time-reference device.



Time error per BC step

Class	Maximum time error
A	100 ns
B	70 ns
C	30 ns (newly agreed)
D	Under discussion

Time error per TC step

Class	Maximum time error
A	100 ns
B	70 ns
C	30 ns? (under discussion)

Fig. 5. Accuracy requirements for PTP devices.

of PTP devices to meet the requirements of 5G and is still discussing how to further improve the accuracy. ITU-T specifies a telecom-boundary clock (T-BC) in G.8273.2 as a device that transmits time information over a network using PTP [2]. The T-BC terminates PTP packets and relays them to the next device (**Fig. 5**). The maximum absolute time error per T-BC was specified as Class B at 70 ns, but then a new standard was newly specified as Class C at 30 ns. Class D regulations to achieve even higher accuracy have also been under discussion from 2019 onwards. The telecom-transparent clock (T-TC), which transmits PTP packets transparently without terminating them, is defined in G.8273.3 (Fig. 5). For T-TC, discussions on high-accuracy classes after Class C have also begun, and certain Class C error parameters have been agreed at the September 2020 meeting. The maximum absolute time error has not yet been specified, but there is a proposal to set it to 30 ns, which is equivalent to T-BC Class C. This is expected to be discussed in the future.

5. Improving the accuracy of time synchronization—Delay measurement

Since PTP is a protocol in which uplink and downlink latencies are assumed constant, ITU-T SG15 has been actively discussing delay of a PTP device and optical-fiber propagation delay, which are the main causes of synchronization-accuracy degradation.

Regarding delay of a PTP device, ITU-T is concerned about the effect of processing-delay variations at lower layers that cannot be eliminated by the hardware-timestamping function specified in IEEE1588 (**Fig. 6(a)**). There is an opinion that the time of delay should be specified with nanosecond accuracy for data processing at the physical layer in optical modules. It has also been pointed out that the delay variation in frame multiplexing and forward error correction (FEC) in digital signal processing used for optical signal transmission and reception is several tens of nanoseconds. Since the delay variation of a device is directly related to the time errors of PTP, a clear specification of the fixed delay and control of the variation are required. However, there is also the dissenting opinion that it is difficult to specify the fixed delay of the optical module configurations and FEC processing as a standard because of the dependency on system-vendor implementation. These issues are therefore currently under continuous study.

Regarding optical-fiber propagation delay, the delay difference between an uplink and downlink of less than 1 ns must be measured for a route of PTP packets. In optical-fiber propagation delay, the optical-path-length variation caused by the refractive index change and mechanical expansion and contraction of the optical fiber when the ambient temperature changes must be considered. The calibration function of the difference in delay specified in IEEE1588-2019

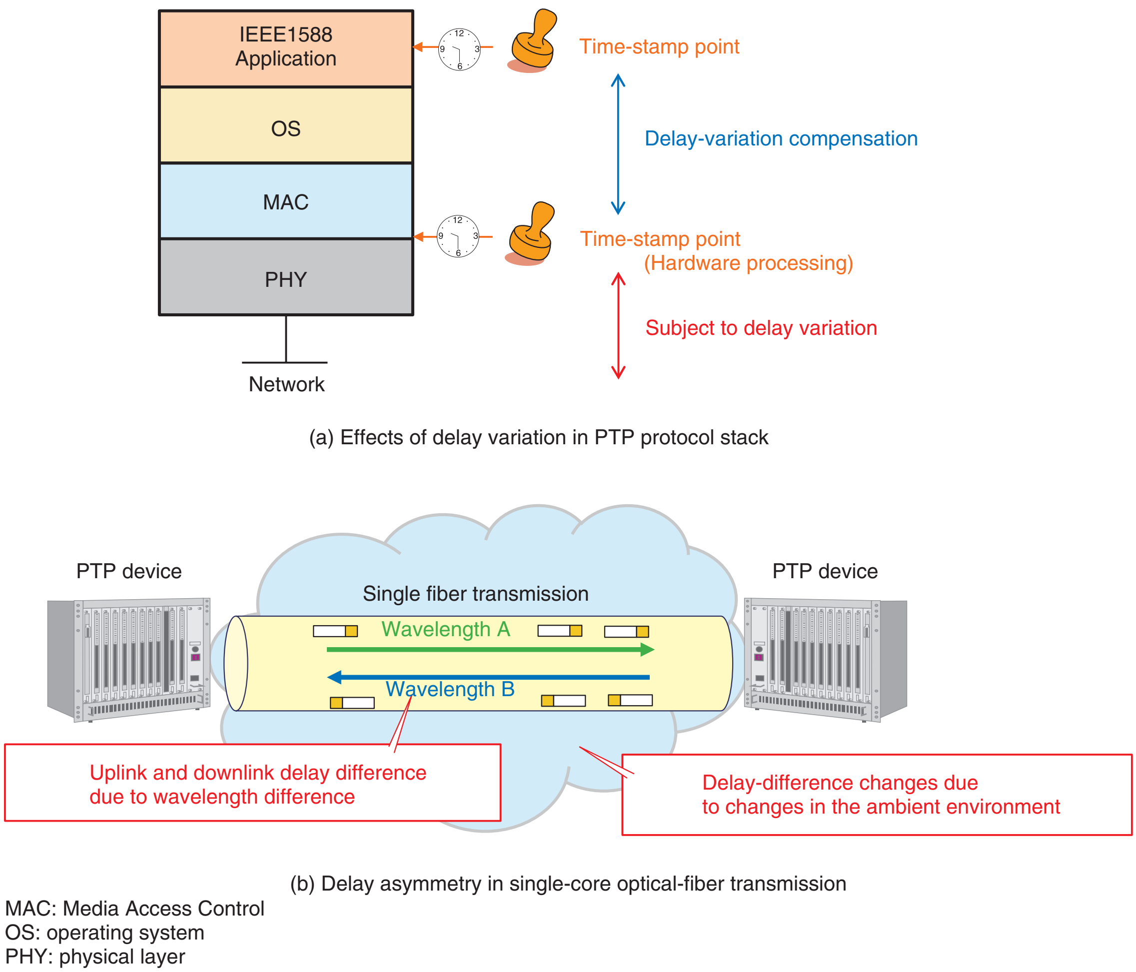


Fig. 6. Effects of device-internal and optical-fiber-transmission line delays.

is based on the assumption that PTP devices are directly connected by optical fibers and that there is no change in difference of delay during operation. However, considering complex telecom networks, the need to dynamically measure delay differences with high accuracy has been raised in ITU-T. In particular, the coherent network Primary Reference Time Clock (cnPRTC)[*3] [4], which is a concept for high-accuracy time-generation architecture that NTT has proposed and standardized jointly with other telecom operators, uses a network to connect distributed PRTC devices (time-reference devices) and compare their time information. At the April 2021 meeting, it was agreed that requirements of the time-transmission accuracy for the comparison link between PRTCs should be very high—Class A: 5 ns, Class B: 1 ns. In Class B, PTP packet transmission is presumably executed using a single optical fiber for bi-direction. Since uplink and downlink PTP packets are transmitted at different wavelengths on a fiber, there are wavelength-dispersion effects in which optical-fiber propagation delay varies depending on the wavelength (**Fig. 6(b)**). To address this, experts in various technical fields are also jointly discussing technologies to accurately measure optical-fiber propagation delay. Such delay can generally be measured using optical time-domain reflectometry (OTDR)[*4]. The use of correlation OTDR, which enables high-accuracy measurement by modulating

*3 cnPRTC: A new architecture that generates optimal time by comparing the times of distributed PRTCs with each other to improve the time accuracy of the PRTC that serves as the time reference.

*4 OTDR: A technology for measuring the transmission loss and distance of an optical fiber and detecting fracture points by injecting optical pulses into the fiber and measuring the reflected pulses.

optical pulses and correlating them at the receiver, is also under discussion. However, issues remain such as the effect of the differences between the OTDR measurement wavelength and PTP transmission wavelength and how to deal with a case making it impossible to secure the reflection path of optical pulses because of optical amplifiers deployed along the section to be measured.

6. Future synchronization technologies

As a synchronization signal reference, the use of optical clocks*5 is being considered as next-generation high-accuracy atomic clocks to replace the conventional cesium atomic clocks. Optical clocks are expected to surpass commercial cesium atomic clocks in accuracy by more than five orders of magnitude and are attracting attention as a new standard of seconds for the future. In ITU-T SG15, proposals have begun for new work items such as a new time-synchronization method using UTC-synchronized optical clocks as a time source and monitoring of operational cesium atomic clocks with high-accuracy optical clocks. NTT has proposed a next-generation synchronization network concept based on optical clocks and will continue to actively propose requirements for next-generation synchronization devices and networks.

7. Future developments

The use of 5G and Beyond 5G networks will be expected to diversify, and many IT services using highly accurate time information will emerge in the future. NTT will continue to actively contribute to the standardization of synchronization technology by searching for potential service requirements and monitoring the latest technological developments.

References

[1] M. Fujii, S. Suwa, R. Toba, and T. Toeda, "Base-station Equipment with the Aim of Introducing 3.5-GHz band TD-LTE," NTT DOCOMO Technical Journal, Vol. 18, No. 2, pp. 8–13, Oct. 2016.

[2] K. Arai and M. Murakami, "Overview of Network Synchronization Technology Standardization in ITU-T," NTT Technical Review, Vol. 14, No. 2, Feb. 2016.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201602gls.html

[3] IEEE Std 1588-2019: "IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems," 2019.

[4] K. Arai and M. Murakami, "Trends in Standardization of High-precision Time- and Frequency-synchronization Technology for Creating a 5G Mobile Network," NTT Technical Review, Vol. 17, No. 1, pp. 35–42, Jan. 2019.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201901gls.html

*5 Optical clock: The next generation of atomic clocks based on optical frequencies. Research institutes and universities worldwide have been actively working on improving the accuracy of these technologies. Optical clocks are promising candidates to become the next-generation standard to replace the cesium atomic clock, the current standard of seconds.

Kaoru Arai
Research Engineer, NTT Network Service Systems Laboratories.

He received a B.S. and M.S. in applied physics from Tokyo University of Science in 2010 and 2012. Since joining NTT Network Service Systems Laboratories in 2012, he has been researching and developing network systems such as clock supply systems and time synchronization systems. He has been participating in ITU-T SG15 activities concerning synchronization technologies since 2014. He received the Distinguished Service Award regarding standardization activities of synchronization technologies from the Telecommunication Technology Committee (TTC) of Japan in 2020. He is currently engaging in research of next-generation synchronization technologies.

Makoto Murakami
Researcher, NTT Network Service Systems Laboratories.

He received a Ph.D. in electrical engineering from the University of Tokyo in 2009. He initially engaged in the research and development (R&D) of long-haul transmission systems using optical amplifiers and coherent modulation/demodulation schemes at the emergence of those technologies. After completing development and deployment of a commercial optically amplified submarine system, he continued R&D of wavelength-division multiplexing systems to further increase the fiber-transmission capacity. From 2001 to 2003, he worked for NTT Communications, where he was involved in the construction and operation of international communication networks mainly in the Asia-Pacific region. Since 2003 he has been an active participant in ITU-T SG15 as head of the Japanese delegation and has also been involved in R&D and standardization of large-capacity optical transport networks. He is currently the chairman of the transport networks and EMC (Electro-Magnetic Compatibility) Working Group at TTC of Japan. He received the Accomplishment Award from the ITU Association of Japan in 2015 and the Distinguished Service Award from TTC in 2015.

Case Studies of Communication Problems with Video Transmission Systems Using Wireless LANs

Technical Assistance and Support Center, NTT EAST

Abstract

This article introduces recent problems that have occurred when using wireless local area networks for video transmission and solution to these problems. This is the sixty-seventh article in a series on telecommunication technologies.

Keywords: wireless LAN, channel interference, video transmission, low-emissivity glass

1. Introduction

Internet access using wireless local area networks (LANs) has become available both for indoors and outdoors. We use wireless LANs for a wide variety of purposes, which include not only accessing websites, sending and receiving emails, and social networking but also video distribution. However, when designing and constructing wireless LAN systems for video transmission such as to display security-camera images or the screen of a personal computer (PC) on a large display, it is necessary to take into account data delays and losses. This article introduces two recent problems that have occurred when using wireless LANs for video transmission and solutions of these problems.

2. Problem case with security-camera system

2.1 Equipment configuration and problem situation

A problem with a security-camera system using a wireless LAN was reported by our customer. The equipment configuration is shown in **Fig. 1**. Two wireless-LAN access points (wireless APs) were installed near the windows of a two-story office building. One of these was on the first floor and the other was on the second floor. These wireless APs were connected to four security cameras via wireless repeaters. These cameras and wireless repeaters were installed on four poles for outdoor lighting. The images from the cameras were stored in the cloud via the wireless LAN and Internet. The customer reported that video streams from three cameras, which were connected to the wireless APs on the second floor, were frequently interrupted.

The wireless LAN used the 2.4-GHz band, which is generally considered to be affected by interference. Accordingly, the maintenance engineer switched from that band to the 5-GHz band (W56), which is less affected by interference and generally used outdoors. However, interruptions were not resolved, so the Technical Assistance and Support Center (TASC) was asked to investigate the problem.

2.2 Investigation of cause

We set the frequency used for the wireless LAN to the 2.4-GHz band and measured the radio-wave environment and the received signal strength.

(1) Investigation of radio-wave environment

We used a wireless LAN tester, which was developed by TASC, to measure radio waves in the

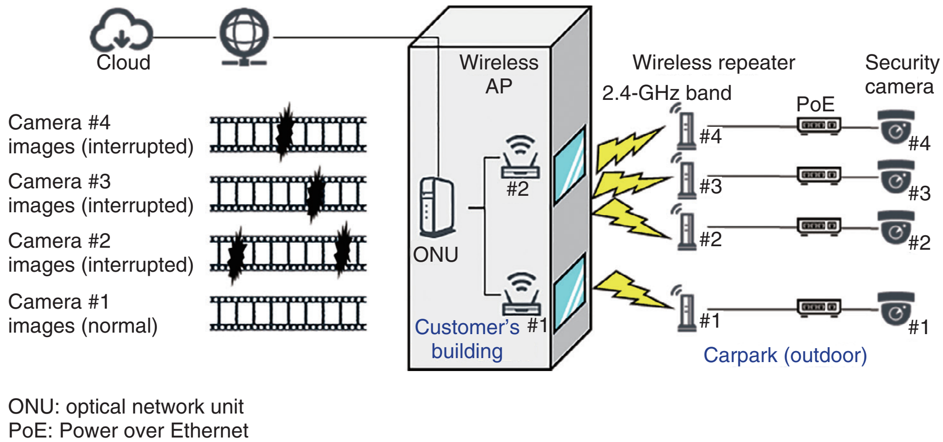


Fig. 1. Equipment configuration.

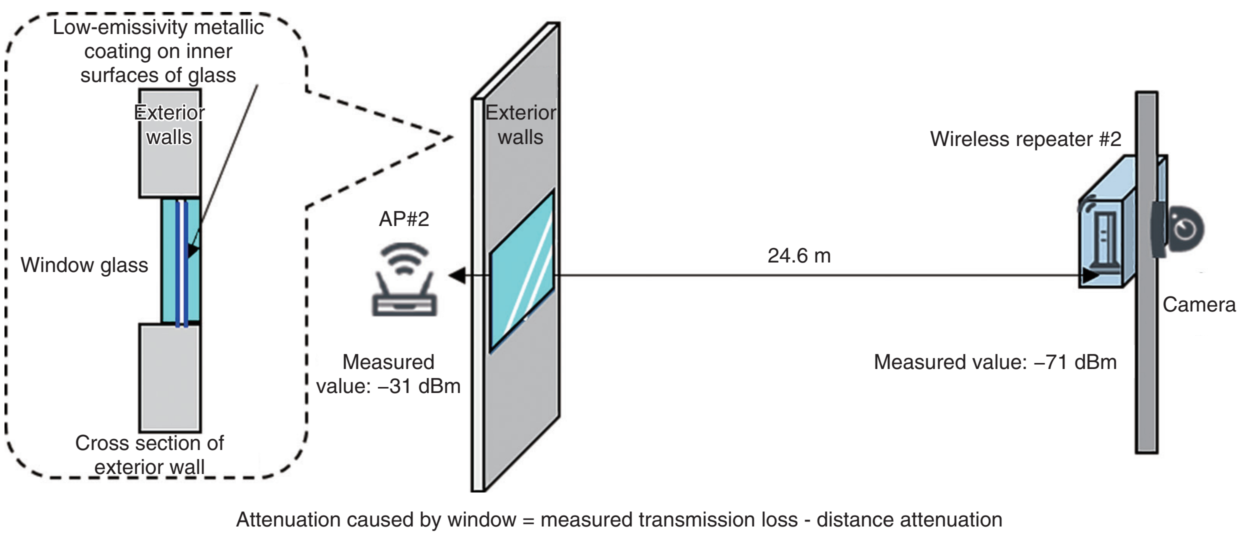


Fig. 2. Attenuation caused by low-emissivity glass.

2.4-GHz band and found no interference or other wireless-LAN waves that could cause radio-wave interference.

(2) Measurement of received signal strength

Using a measuring instrument that can measure the radio-wave strength of devices, we measured the received signal strength of the radio waves near the wireless APs and wireless repeaters. The received signal strength of the radio wave for downlink transmitted from a wireless AP near a wireless repeater was less than –60 dBm, and that for uplink transmitted from a wireless repeater near the wireless AP was very weak, ranging from –74 to –100 dBm, which was the measurement limit.

The measured radio-wave strength for downlink near the wireless APs and wireless repeaters was –31 and –71 dBm, as shown in **Fig. 2**. The results indicate that the transmission loss between a wireless AP and wireless repeater was about 40 dB. The theoretical value of distance attenuation calculated from the distance of 24.6 m between the wireless AP and wireless repeater is 28 dB (≒ 20log (24.6 m)); therefore, attenuation of 12 dB (about 1/20) was caused by the window glass in the building. The window was a type of double glazing and the inner surfaces of the glass were covered with a low-emissivity metal coating to

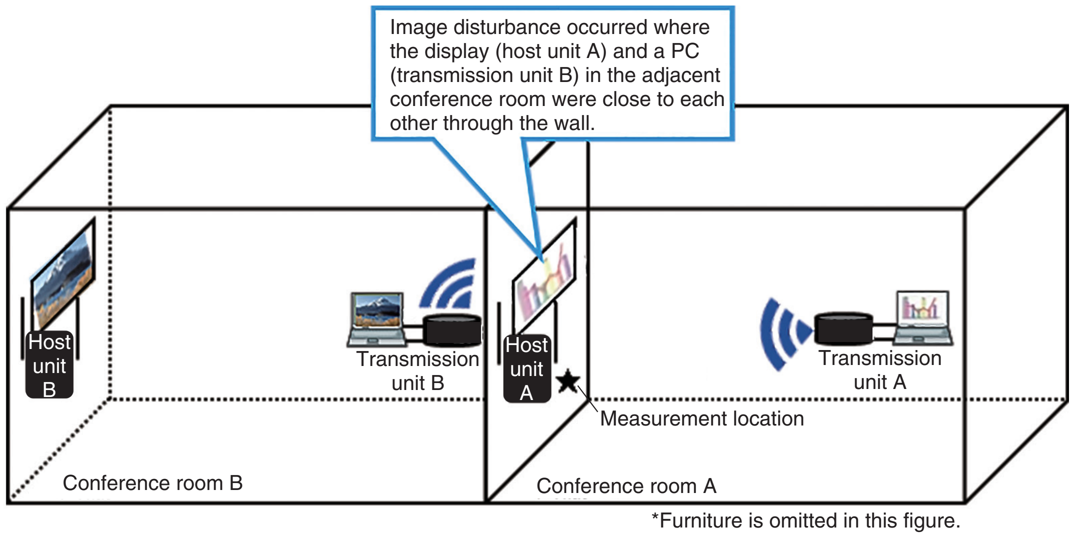


Fig. 3. Equipment configuration.

reduce radiant heat from the sun.

2.3 Cause of problem

From our investigation, it was presumed that the large attenuation caused by the low-emissivity glass interrupted the video streams from the security cameras. The radio waves for uplink from the wireless repeater were significantly weakened, so the communication between the wireless APs and wireless repeaters probably become unstable.

2.4 Solution

In this case, by moving the wireless APs away from the low-emissivity window glass, which prevents the transmission of radio waves, it is possible to increase the amount of radio-wave transmission through the wall other than the glass. When we moved the wireless APs, the received signal strength between the wireless APs and each wireless repeater increased and improved radio-wave throughput, eliminating video interruptions.

3. Problem case with video-image-transferring system

3.1 Equipment configuration and problem situation

A problem with a security-camera system using a wireless LAN was reported from our customer. The equipment configuration is shown in **Fig. 3**. A wireless display adapter (WDA), which is used to display the screen of a PC on a large display via a wireless LAN, has two units. Its host unit is connected to the display and its transmission unit is connected to the PC. At the customer's premises, many WDAs were installed in conference rooms and meeting spaces. The WDAs used the 5-GHz band of wireless LAN, and channel bonding was used to ensure the 40-Mbit/s bandwidth, which is required for video transmission.

The 5-GHz-band wireless LAN has a dynamic frequency selection (DFS) function that automatically switches the channel to be used if radar waves, etc. are received. In this equipment configuration, W52, a frequency band that is not used by radar waves, was used to avoid communication interruption due to channel switching by DFS. This band was also divided into two channel groups (36/40ch and 44/48ch) to avoid inter-channel interference. Despite these attempts, the video images displayed on the large display were disrupted in certain areas.

3.2 Investigation of cause

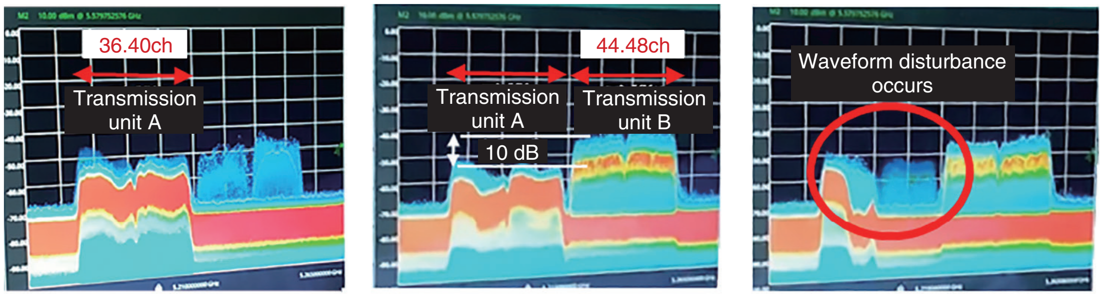
(1) Investigation of radio-wave environment

We used a spectrum analyzer to measure the radio-wave environment in the 5-GHz band, and no radio waves other than those of WDAs were measured.

(2) Investigation of transmission-waveform spectrum

We measured the transmission-waveform spectrum near host unit A when video disturbance occurred (**Fig. 4**).

The measurement results indicate that transmission



(a) Only transmission unit A is communicating

(b) Immediately after communication with transmission unit B starts

(c) 2 to 3 seconds after communication starts

Fig. 4. Measurement of transmission-waveform spectrum using spectrum analyzer.

units A and B used different channel groups and that at the measurement location shown in Fig. 3, the received signal strength from transmission unit B was about 10 dB higher than that from transmission unit A. They also indicate that two to three seconds after transmission unit B started communication, the transmission waveform from transmission unit A was disrupted (Fig. 4(c)) and, simultaneously, the video image was disrupted.

3.3 Cause of problem

It was presumed that this disturbance was caused by the radio wave from transmission unit B, which was about 10 dB higher than that from transmission unit A, reached host unit A, although conference rooms A and B were separated by a wall. The radio wave from transmission unit B reached the host unit A; therefore, the communication between transmission unit A and host unit A was affected. This disturbance is presumed due to adjacent-channel interference, which is caused by the arrival of strong radio waves on the adjacent channel.

3.4 Solution

The basic solution against problems like this case is to improve the radio-wave environment. To improve such an environment, it is necessary to ensure the distance between the wireless-LAN devices, to prevent radio-wave interference between the wireless-LAN devices by installing sheets of radio-wave shielding, to shorten the distance between the host and transmission units or not to obstruct the line of sight between the host and transmission units.

However, it was difficult to implement the above-mentioned solution in this case; therefore, we set up the same communication channel for units A and for units B and controlled the transmission timing temporally. It became possible to stabilize the communication in the wireless section and eliminate video disturbance. It should be noted that when the same channel is used for multiple wireless devices, the transmission rate will be divided in proportion to the number of devices; therefore, it is necessary to verify operation before installation of the devices to ensure that the bandwidth required for video transmission can be secured.

4. Conclusion

Recent problem cases with video transmission systems using wireless LANs and their solutions were introduced in this article. TASC develops tools that can be used in the field to analyze and identify the cause of problems in wireless-LAN and wireless-based systems.

The EMC Engineering group in TASC is continuously engaging in technical support to the field regarding problems caused by lightning, malfunction due to electromagnetic phenomena, interference of radio waves, induction from power line, and so on, to ensure reliability of telecommunication services.

External Awards

Electronics Society Award

Winners: Masahiro Nada, Toshihide Yoshimatsu, NTT Device Innovation Center; Hideaki Matsuzaki, NTT Device Technology Laboratories
Date: March 31, 2021
Organization: The Institute of Electronics, Information and Communication Engineers (IEICE) Electronics Society

For their pioneering research and development of high-speed avalanche photodiodes with a vertical-illumination structure for optical fiber communications.

FIT Encouragement Award

Winner: Yuya Omori, NTT Device Innovation Center
Date: August 27, 2021
Organization: Information Processing Society of Japan (IPSJ)

For "Complexity Reduction Based on Equivalence and Continuity of Convolutional Layers in CNN Inference."
Published as: Y. Omori, D. Kobayashi, S. Yoshida, S. Hatta, H. Uzawa, K. Nakamura, and K. Sano, "Complexity Reduction Based on Equivalence and Continuity of Convolutional Layers in CNN Inference," Proc. of 20th Forum on Information Technology (FIT2021), Vol. 1, pp. 179–180, Aug. 2021.

Best Paper Award (Research track)

Winners: Yukako Iimura and Shinobu Saito, NTT Computer and Data Science Laboratories
Date: September 8, 2021
Organization: IPSJ Special Interest Group on Software Engineering (SIGSE)

For "Practical Report on Software Development Approach for Promoting Workstyle Flexibility."
Published as: Y. Iimura and S. Saito, "Practical Report on Software Development Approach for Promoting Workstyle Flexibility," Proc. of IPSJ/SIGSE Software Engineering Symposium 2021, pp. 14–22, Sept. 2021.

Distinguished Contributions Award

Winner: Doohwan Lee, NTT Network Innovation Laboratories
Date: September 15, 2021
Organization: IEICE Communications Society

For his contributions as a peer reviewer for papers submitted to the IEICE Communications Society.

Young Scientist Presentation Award

Winner: Koji Sakai, NTT Basic Research Laboratories
Date: September 21, 2021
Organization: The Japan Society of Applied Physics (JSAP)

For "Development of 3D-cultured Neuronal Network in Graphene-based Self-folding Electrode Array."
Published as: K. Sakai, T. Teshima, T. Goto, H. Nakashima, and M. Yamaguchi, "Development of 3D-cultured Neuronal Network in Graphene-based Self-folding Electrode Array," The 68th JSAP Spring Meeting 2021, Mar. 2021.

Young Scientist Presentation Award

Winner: Ai Ikeda, NTT Basic Research Laboratories
Date: September 21, 2021
Organization: JSAP

For "Novel Superconducting $(CaCuO_2)_n/(Ca_2Fe_2O_5)_m$ Superlattices Prepared by MBE."
Published as: A. Ikeda, Y. Krockenberger, Y. Taniyasu, and H. Yamamoto, "Novel Superconducting $(CaCuO_2)_n/(Ca_2Fe_2O_5)_m$ Superlattices Prepared by MBE," The 68th JSAP Spring Meeting 2021, Mar. 2021.

Young Scientist Presentation Award

Winner: Takahiro Kashiwazaki, NTT Device Technology Laboratories
Date: September 21, 2021
Organization: JSAP

For "Terahertz-order Broadband Measurement of Squeezed Light by Optical Parametric Amplification for Ultra-fast Optical Quantum Computing."
Published as: T. Kashiwazaki, N. Takanashi, A. Inoue, T. Kazama, K. Enbutsu, R. Kasahara, T. Umeki, and A. Furusawa, "Terahertz-order Broadband Measurement of Squeezed Light by Optical Parametric Amplification for Ultra-fast Optical Quantum Computing," The 68th JSAP Spring Meeting 2021, Mar. 2021.

Young Scientist Presentation Award

Winner: Takuma Tsurugaya, NTT Device Technology Laboratories
Date: September 21, 2021
Organization: JSAP

For "Photonic Reservoir Computing Using Low-power-consumption SOA."
Published as: T. Tsurugaya, T. Hiraki, M. Nakajima, T. Aihara, N.-P. Diamantopoulos, T. Fujii, T. Segawa, and S. Matsuo, "Photonic Reservoir Computing Using Low-power-consumption SOA," The 68th JSAP Spring Meeting 2021, Mar. 2021.

Papers Published in Technical Journals and Conference Proceedings

Divide-and-conquer Verification Method for Noisy Intermediate-scale Quantum Computation

Y. Takeuchi, Y. Takahashi, T. Morimae, and S. Tani

arXiv:2109.14928, Oct. 2021.

Several noisy intermediate-scale quantum computations can be regarded as logarithmic-depth quantum circuits on a sparse quantum computing chip, where two-qubit gates can be directly applied on only some pairs of qubits. In this paper, we propose a method to efficiently verify such noisy intermediate-scale quantum computation. To this end, we first characterize small-scale quantum operations with respect to the diamond norm. Then by using these characterized quantum operations, we estimate the fidelity $\langle \psi_t | \hat{\rho}_{out} | \psi_t \rangle$ between an actual n-qubit output state $\hat{\rho}_{out}$ obtained from the noisy intermediate-scale quantum computation and the ideal output state (i.e., the target state) $|\psi_t\rangle$. Although the direct fidelity estimation method requires $O(2^n)$ copies of $\hat{\rho}_{out}$ on average, our method requires only $O(D^3\, 2^{12D})$ copies even in the worst case, where D is the denseness of $|\psi_t\rangle$. For logarithmic-depth quantum circuits on a sparse chip, D is at most $O(\log n)$, and thus $O(D^3\, 2^{12D})$ is a polynomial in n.