

Telepresence Technology to Facilitate Remote Working with Human Physicality

Shigekuni Kondo, Daisuke Sato, Mitsuhiro Goto, Motohiro Takagi, and Narimune Matsumura

Abstract

Remote work allows people to free themselves from geographical and temporal constraints, and there are calls for it to be made available as an option for essential workers for whom it is not readily applicable. This article introduces our efforts to develop telepresence technology with which a person can use an artificial body or robot to sense the situation at a remote location and interact with other people and the environment in the same way as when working in the field. Thus, they can carry out their required tasks and engage in face-to-face communication without any discomfort or hindrance.

Keywords: remote working, telepresence, cybernetics

1. Introduction

In 2020, our values and common practices were changed by a new strain of coronavirus that rapidly spread throughout the world. One such change was the general acceptance of remote working, in which employees are no longer required to perform their jobs at a fixed location [1]. With the spread of remote working, it has become possible for people to carry out various tasks that would previously have been performed face-to-face without having to go to the actual location, thus releasing them from the geographical and temporal constraints associated with working in a set location. This provides people with various benefits, such as being able to freely choose where to live and perform other activities during the time they would have spent commuting to and from work. However, certain people, such as essential workers, can find it difficult to perform their jobs remotely, and even when remote working is possible, they can still experience problems such as communication failures and increased psychological burden when they have to accommodate a remote and decentralized work style.

NTT Human Informatics Laboratories is researching and developing telepresence technology to address the problems of remote working to provide the Remote World where people can have the option of working remotely when necessary in any line of work.

2. What is telepresence?

In some cases, remote work may not be an option because a person's job primarily involves on-site work that would be less efficient and productive when performed remotely or that requires face-to-face communication.

To facilitate remote working in jobs that require on-site work, it is necessary to have some way of actually performing this work on site by remote control (e.g., by machines or robots, i.e., artificial bodies). However, a problem with the current situation is that the content and timing of work operations performed at the remote site can differ from what was intended by the remote operator due to a lack of repeatability or delays in the response of the operation interface or artificial body. This can significantly affect the

efficiency and productivity of the work. It should also be possible to accurately ascertain the local situation from a remote location, but this raises the issues of the lack of information for recognizing changes in the current environment or behavior of other people on site, and the tendency for people who are present on site to feel uncomfortable and anxious about working with someone who is in a remote location.

With the telepresence technology we are developing, we aim to make it possible for a person to use an artificial body to sense the situation at a remote location and interact with other people and the environment in the same way as when working on site. Therefore, they can carry out their required tasks and engage in face-to-face communication without any discomfort or hindrance. We are currently researching and developing technologies that will be needed to make this telepresence a reality: *zero-latency media technology*, which improves the operability at remote locations by predicting motions and reactions to compensate for latency and lack of operational information, *lifelike communication technology*, which enables information presentation methods that deliver the same quality of experience as being present on site, and *embodied-knowledge-understanding technology*, which clarifies and systematizes the connections between knowledge, experience, and physical behavior to predict changes in the remote operator or in the on-site environment and ascertain information from the on-site environment that is necessary for operating from the remote location. This article introduces our efforts in developing these technologies to enable telepresence.

3. Zero-latency media technology

Zero-latency media technology is used to create a system with which human operators can exert at least 100% of their work from a remote location by supporting the remote operation of robots. The main factors that hinder people's ability to work by remote control are latency, missing information, and differences between the human and robot body structures, each of which requires countermeasures. We are currently engaged in research and development focused on countermeasures against latency, which has a particularly large impact.

In remote control work, the operator controls a robot's position, force, and speed to appropriate values on the basis of feedback information such as camera images sent from a remote location. If there are any delays in this process, the feedback needed

for remote control cannot be obtained at the required timing, reducing the accuracy with which the robot's position, force, and speed can be controlled remotely and resulting in issues such as operational errors and reduced work efficiency. For example, if an object's position shifts while trying to grasp it with a robot arm, the arm will fail to grasp it properly. If too much force is applied, the object could be destroyed, and if too little force is applied, it will not be possible to hold it well. When working on or painting an object, a shift in speed will change how the object is affected.

Zero-latency media technology addresses the latency issues of remote operations to improve work efficiency and reduce the burden on workers by implementing motion correction and cognition correction on the basis of an understanding of people and the environment (**Fig. 1**). Motion correction mitigates the effects of latency by compensating for the operational information sent by the operator before moving the robot. We are studying two approaches for implementing this compensation, predictive control and semi-autonomous control. In predictive control, we aim to reduce the latency of the entire system by predicting the operator's movements and controlling the robot on the basis of these predicted movements, making it operate in a preemptive manner. Semi-autonomous control is used to eliminate operational difficulties due to latency by estimating the operator's intentions and controlling the robot semi-autonomously on the basis of these estimations. Cognition correction mitigates the effects of latency by compensating the feedback information sent from the remote work environment before it is presented to the operator. We are studying two approaches for achieving this compensation, control-object compensation and work-object compensation. In control-object compensation, feedback information related to the robot (the controlled object) is compensated for with techniques such as augmented reality (AR), whereas in work-object compensation, compensation is applied to the feedback information about the object on which this remote work is performed.

To verify the effects of motion correction and cognition correction on mitigating latency effects, we remotely operated a simple pen-type robot and measured the time taken to accomplish the alignment task of touching points marked on a sheet of paper. The results of this experiment confirmed that the task took longer to accomplish as the delay time increased when no compensation was applied, but that compensation was effective at preventing delays from adversely affecting the task efficiency. We are

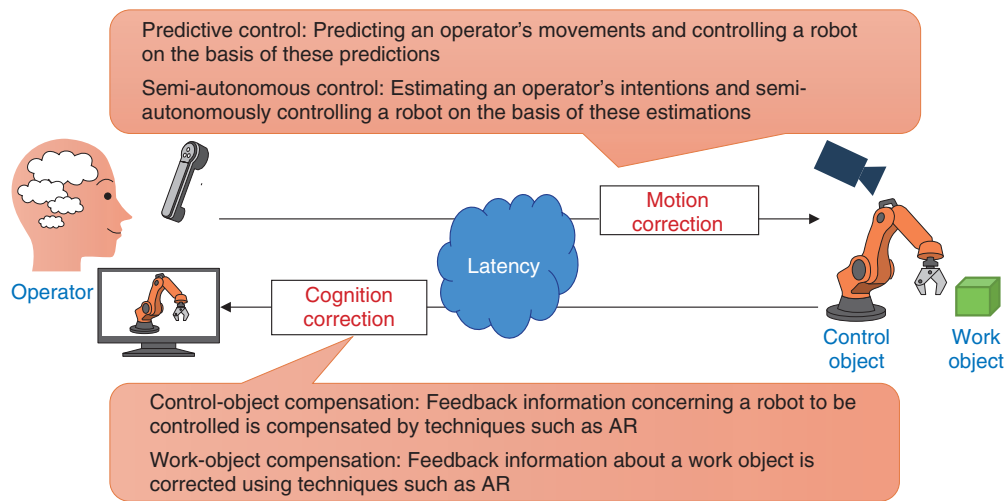


Fig. 1. Zero-latency media technology.

currently using remotely operated robot arms and humanoid robots to verify the delay-mitigation effects in more practical tasks and working to improve various correction techniques by deepening our understanding of people and the environment.

4. Lifelike communication technology

Lifelike communication technology focuses on how communication is experienced between people online. People have been presented with more opportunities to communicate online by exchanging video and audio with remote partners using tools such as web-conferencing applications. However, compared with direct face-to-face communication, this sort of online communication can tend to feel somewhat unsatisfactory. With lifelike communication, our aim is to reveal the factors that cause these dissatisfactions with online communication and provide the best possible online information that delivers the same quality of experience as face-to-face communication. In online communication, it is particularly difficult to sense the other person's presence and distance through the screen. The conventional approach to addressing these issues has been to provide high-quality and high-definition reproductions of sounds and images perceived in real space. With lifelike communication technology, instead of faithfully reproducing the audiovisual experience of communication in a real space, it is sufficient if the user ultimately feels satisfied and convinced even if the information reaching the eyes and ears is crude. Therefore,

even the users of online communication can be provided with a face-to-face experience, as if the other person was close by (Fig. 2).

In 2021, based on this technical concept, we conducted a case study to enhance online technical exhibitions [2]. Currently, visitors to online exhibitions are only able to browse the content of exhibition pages they are interested in and do not get to feel the satisfaction of receiving an explanation from a helpful staff member, as is the case with face-to-face exhibitions. We, therefore, proposed an information-presentation method that uses stereophonic sound to present visitors with an audio commentary that corresponds to the part of the page they are viewing, thus providing them with the experience of receiving explanations directly from those people while allowing them to experience the distance and position of exhibition guides and attendants in the same way as at a face-to-face exhibition. We asked several visitors with experience in participating in online exhibitions and exhibition planning to evaluate the effectiveness of the proposed method. The results indicate satisfaction with and acceptance of the exhibition. We are considering applying this method to technical exhibitions as well as online customer service and art exhibitions.

5. Embodied-knowledge-understanding technology

As part of our effort to support telepresence, we introduce our research on the understanding of

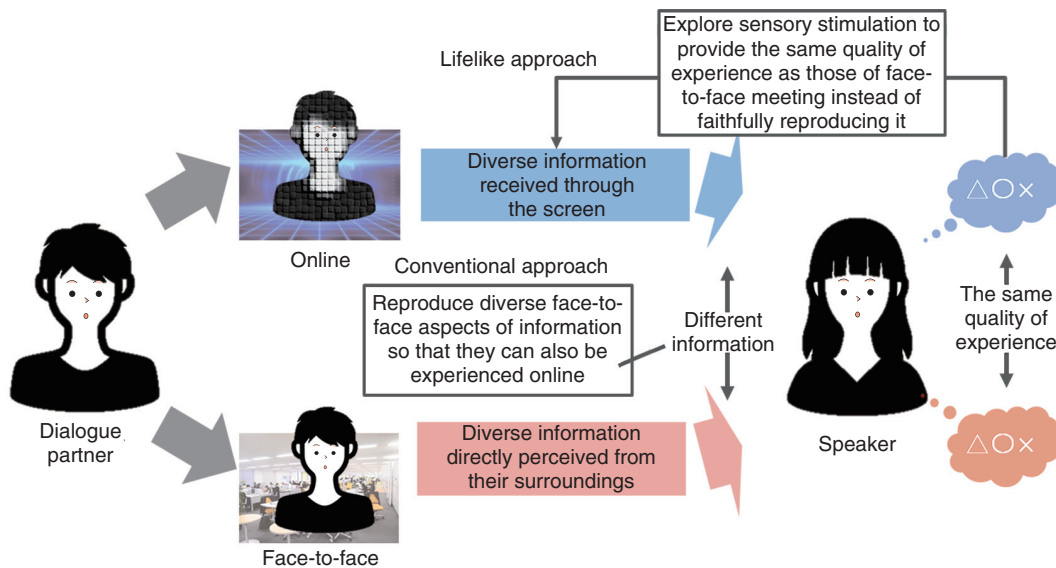


Fig. 2. Lifelike communication technology.

embodied knowledge for systematization of physical behavior. For an essential worker to perform their tasks by means of telepresence, it is necessary to reproduce the essential worker's skills on an artificial body or robot. The use of robots has been considered for tasks that are relatively simple for humans, such as restocking the shelves in convenience stores and replacing local area network cables in datacenters. However, the complex jobs of professionals, such as nurses and caregivers, involve many skills with embodied-tacit knowledge that must be acquired in the field and are therefore difficult to perform with robots. These skills, which can be considered embodied knowledge, cannot be reproduced by artificial bodies or robots because there are no clear requirements for handling them digitally, making them difficult to perform remotely. To reproduce the skills required for tasks such as nursing and caregiving that are complicated even for humans, the challenge we face is how to express these skills digitally with artificial bodies and robots.

To digitally represent the skills demonstrated in high-level professional work, it is necessary to clarify and systematize the physical behaviors that are based on the specialist expertise that these skills require. For example, when dealing with patients in a hospital ward, the situational knowledge that the patients in this ward have asthma and the occupational knowledge that the breathing of asthma patients should always be checked with a stethoscope will result in

the physical action of using a stethoscope to check their breathing. Thus, by systematically organizing the embodied knowledge that is needed in different environments and situations, and what physical behavior needs to be expressed based on this knowledge, it is possible to define how artificial bodies and robots should function in each environment and situation.

In our effort to systematize physical behavior, we are conducting research and development on how to systematically define the physical behavior information needed for the movement of artificial bodies and robots and how to acquire and express this information. Some of the physical behaviors that are targeted by physical-behavior information can be expressed as words in the form of instruction manuals, while others cannot and exist only as embodied-tacit knowledge within a particular field. There may also be discrepancies between the actions described in a manual and those that take place in the field. Our goal is to reproduce the skills of physical knowledge in artificial bodies and robots by grasping how these physical behaviors are performed in practice and combining them with expert knowledge so that they can be handled digitally. In 2021, we developed a technique for understanding the physical behavior of customers and employees in a convenience store business with finer granularity than could be achieved from manuals [3] and a dataset for recognizing the behavior of employees working in high places in the construction

of telecommunications infrastructure from a first-person perspective and evaluating whether they were adhering to safety rules defined in manuals [4]. We will investigate physical-behavior-recognition technology for grasping the actual situation in the field as well as establishing methods for linking specialized knowledge and physical behavior.

6. Conclusion

We have introduced the zero-latency media technology, lifelike communication technology, and embodied-knowledge-understanding technology that we are developing to achieve telepresence technology with which a person can use an artificial body to sense the situation at a remote location and interact with other people and the environment in the same way as when working on site. Therefore, they can carry out their required tasks and engage in face-to-face communication without any discomfort or hindrance. We will proceed with technical development

and verification in specific fields such as nursing and equipment maintenance in which it is currently difficult to work remotely and expand the concept of telepresence technology and promote its research and development to liberate individuals from physical constraints and extend their capabilities.

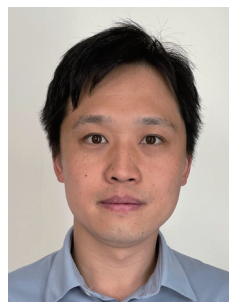
References

- [1] The Ministry of Health, Labour and Welfare, “Whitepaper on the Labour Economy 2021,” 2021 (in Japanese). <https://www.mhlw.go.jp/wp/hakusyo/roudou/20/dl/20-1.pdf>
- [2] S. Uchida, M. Goto, and H. Seshimo, “A Study of Increasing Satisfaction during Online Communication by 3D Audio,” IEICE Technical Report, Vol. 121, No. 423, MVE2021-88, pp. 255–260, 2022.
- [3] “Human behavior understanding technology for improving CX/EX—Visualize human behavior and promote better CX/EX,” Exhibits of NTT R&D Forum – Road to IOWN 2021, C27, 2021. https://www.rd.ntt/_assets/pdf/e/forum/2021/C27_e.pdf
- [4] R. Furuta, M. Takagi, Y. Sugano, and Y. Sato, “A Dataset for Work Instruction Deviation Detection from Field Work Videos,” IEICE Technical Report, Vol. 121, No. 304, PRMU2021-30, pp. 37–42, 2021.

**Shigekuni Kondo**

Senior Research Engineer, Cybernetics Laboratory, NTT Human Informatics Laboratories.

He received a B.E. in applied physics and M.S. in integrated design engineering from Keio University, Kanagawa, in 2003 and 2005 and joined NTT in 2005. His research interests include software modularity, home networks, and user interfaces. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE).

**Motohiro Takagi**

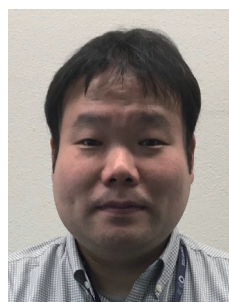
Researcher, Cybernetics Laboratory, NTT Human Informatics Laboratories.

He received a B.E., M.E., and Ph.D. from Keio University, Kanagawa, in 2009, 2011, and 2020. In 2011, he joined NTT. His research interests include human behavior understanding through machine learning, computer vision, and natural language processing.

**Daisuke Sato**

Senior Research Engineer, Cybernetics Laboratory, NTT Human Informatics Laboratories.

He received a B.E. and M.E. in mechanical engineering from Waseda University, Tokyo, in 2007 and 2009. He joined NTT in 2009 and has been engaged in the research and development of information retrieval, geographic information, and telerobotics services. He is a member of the Information Processing Society of Japan (IPSI).

**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.

**Mitsuhiro Goto**

Senior Research Engineer, Cybernetics Laboratory, NTT Human Informatics Laboratories.

He received a B.E., M.E., and Ph.D. from the University of Electro-Communications, Tokyo, in 2008, 2010, and 2022. In 2010, he joined NTT. His research interests include human-computer interaction, human-robot interaction, and non-verbal behavior of communication.
