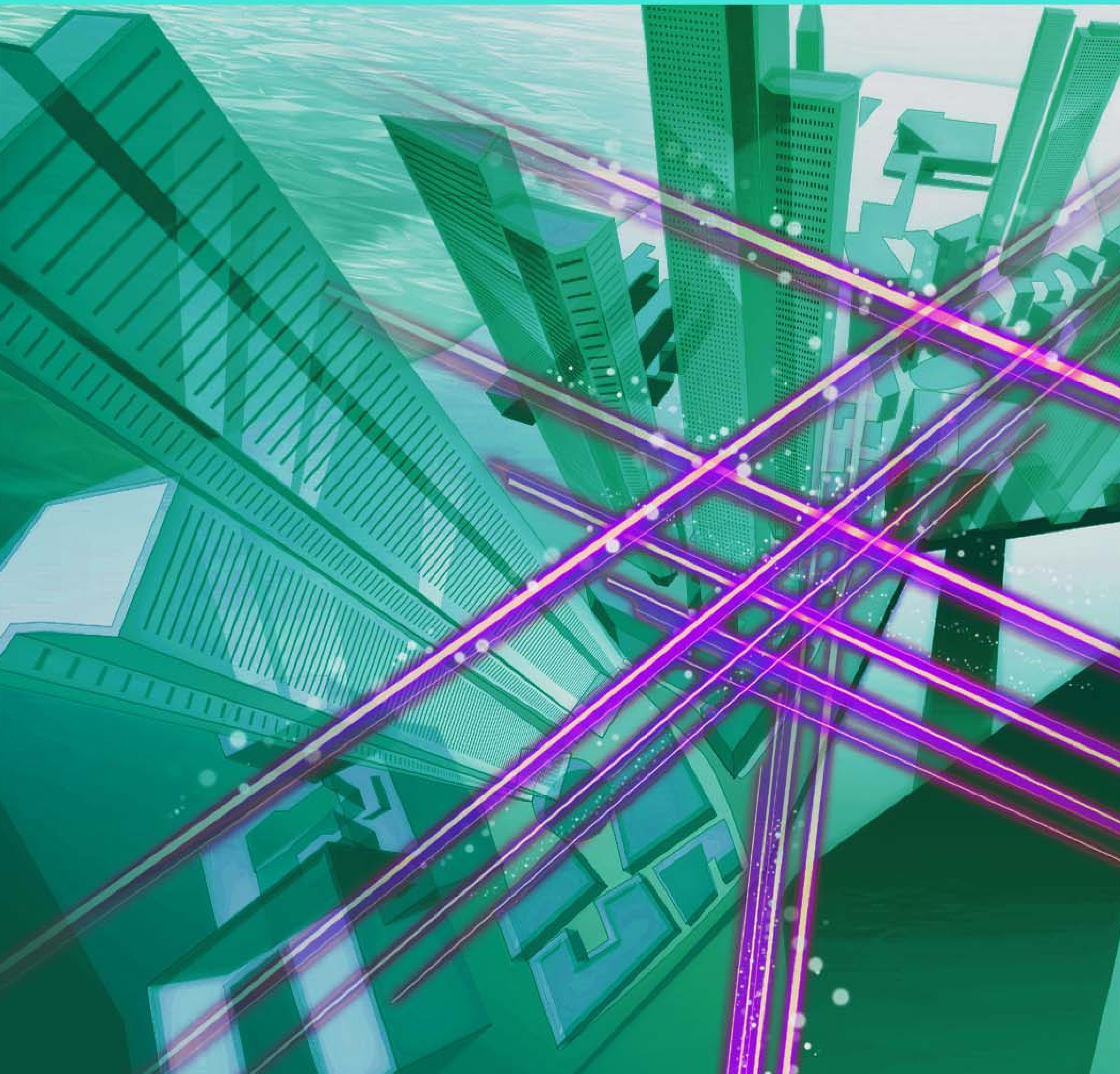
NTT Technical Review

4
2024



April 2024 Vol. 22 No. 4

NTT Technical Review

April 2024 Vol. 22 No. 4

The Role of Top Management Is to Generate a Perturbation without People Being Aware of It

Akira Okada

Senior Vice President of R&D, Head of NTT Science and Core Technology Laboratory Group

Abstract

NTT's laboratories are engaged in a wide range of research and development from basic research to applications, and divided into four laboratory groups. As one of those laboratory groups, NTT Science and Core Technology Laboratory Group carries out research and development with three missions in mind: "Conduct research and development on cutting-edge technologies to expand NTT's business domains," "Create new principles and concepts that will revolutionize society," and "Research and develop technologies that are friendly to the global environment and people." We asked Akira Okada, senior vice president of R&D, head of NTT Science and Core Technology Laboratory Group, about the technology strategy of the laboratory group and his mindset as a top manager.

Keywords: basic research, human resources development, R&D management

Cutting-edge research and development with an eye on society in 2050

—Please tell us about NTT Science and Core Technology Laboratory Group.

NTT Science and Core Technology Laboratory Group is engaged in cutting-edge research with the vision of "Creating a society toward 2050 by connecting everything, nurturing the Earth, and building a world where people can live vibrantly."

The laboratory group is composed of four laboratories. To create innovative network systems, NTT Network Innovation Laboratories researches and develops technologies for diverse frequency bands and transmission media to pioneer a new communication paradigm. NTT Device Technology Laboratories conducts research and development (R&D) of devices and materials that create new value through the convergence of photonics and electronics by using its clean rooms to realize next-generation communications, information-processing infrastructure, and technologies that are sustainable and enrich people's lives. NTT Communication Science Laboratories researches media, information processing, and human science to construct fundamental theories that address the essence of humans and information, develop media and information processing technologies that break down communication barriers, and bring about changes in society. NTT Basic

Research Laboratories researches materials science, physical science, and quantum science to create new principles and concepts that overcome the limitations of current technologies and explores technology seeds that will become future core technologies.

By expanding their knowledge of science and technology and thinking outside the box, researchers at these laboratories are striving to create the world's first and best technologies or technologies that amaze everyone.

—NTT Science and Core Technology Laboratory Group is engaged in a wide range of basic research. Would you introduce some of the research topics that are currently attracting attention?

In the information-processing field, we are focusing on photonics-electronics convergence technology, which is the backbone of the All-Photonics Network (APN). The APN is one of the three elements of the Innovative Optical and Wireless Network (IOWN) and uses optical signal processing in an end-to-end manner in contrast to conventional networks in which signals are transmitted by repeating conversion between electrical and optical signals. In fact, the results of our laboratory group's research on photonics-electronics convergence technology were the starting point for envisioning the IOWN concept, goals of which include a hundred-fold increase in power efficiency by introducing this technology. The development of an optical-electrical-optical converter (optical transistor) reported in Nature Photonics in 2019 is a stepping stone in this research field.

Artificial photosynthesis is also attracting attention as a sustainable technology. As its name suggests, artificial photosynthesis is a technology that artificially recreates the natural photosynthesis mechanism that uses sunlight to generate energy from carbon dioxide (CO_2). We believe that artificial photosynthesis can contribute to the creation of a carbon-recycling society. In 2023, we developed an artificial photosynthesis device achieving the world longest continuous carbon fixation for 350 hours. The cumulative amount of carbon fixed by the CO_2 conversion reached 420 g/m^2, which surpasses the annual amount of carbon per unit area fixed by a tree (Japanese cedar). We are currently striving to improve the performance of our artificial photosynthesis device by increasing both the efficiency of the reaction and the longevity of its electrodes. We are also attempting to establish this device as one of the technologies for reducing CO_2 emissions using solar energy through outdoor tests.

Setting a new research problem by identifying a problem and boldly tackling a challenging problem

—I think that because basic research is so pioneering, it is sometimes difficult to convey its importance, isn't it?

Basic research serves as the foundation for the development of new technologies, products, and systems through the discovery and elucidation of mechanisms of theories, phenomena, natural laws, and the like in a manner that approaches the essence and existence of things. Therefore, since we cannot directly "see" basic research in our daily lives, it is difficult to convey its importance to people. Everyone can understand that research that wins a Nobel Prize in physics is amazing research; however, it is often difficult for the general public to understand how the results of such research are implemented in and useful to society. To widely convey the importance of basic research to the general public, I believe that researchers need to have the ability to convey their research to non-experts in easy-to-understand and simple language.

With that importance in mind, we are conducting educational activities to communicate our technology in an easy-to-understand manner. NTT is unique in being one of the few Japanese corporate research institutes that conduct basic research. Since this uniqueness is one of NTT's strengths, NTT laboratories are engaged in activities that firmly emphasize the cutting-edge nature of our technologies by bringing our researchers to the forefront.

Specifically, we use metaphors to communicate our research to the general public in a way that is easy to understand even for those without specialized knowledge, and we are increasing exposure of our research via various media, including news releases, to make it easier for our message to reach as many people as possible. We are also presenting our research results and progress at conferences and symposiums while actively participating in lectures at general venues.

It has been discussed that Japan's research capabilities are declining. Technology in the basic-research field cannot be developed after two or three years of research. Although we are conducting research with an eye on 2050, we also set milestones for 2030, 2035, and 2040 indicating—as concretely as possible—the technology we are aiming for and the value it will create. Considering the nature of basic research, we try to create an environment in which researchers can tackle challenging themes from a long-term perspective and proactively propose what they want to research.

I constantly advise researchers to avoid using the word "difficult" when they are pursuing their research. I say that because the moment we say the word "difficult," we sometimes stop thinking. From a different perspective, what is "difficult" is something has not been done before or something no one has been able to achieve, and that situation provides an opportunity for us to take on a challenge. We can overcome a challenging problem by taking approaches from a variety of angles and mobilizing our knowledge and out-of-the-box thinking and means. I believe that adopting this stance can create a new paradigm and lead to the improvement of our research capabilities.

—You are encouraging the creation of research seeds and drawing out the potential of researchers. What do you consider important in R&D management?

Solving problems is not the only way to develop research capability. I believe that, setting a new

research problem by identifying a problem and valuing the attitude of boldly tackling a challenging problem will lead to developing true research capability. Taking this approach may require patience, but it is a significant undertaking of NTT Science and Core Technology Laboratory Group. I would like us to continue taking this approach.

Applied research and its practical implementation, which are the next steps following basic research, are also important processes, so from a management standpoint, I want to ensure that basic and applied research are running as two wheels on one axel. We should not only engage in basic research because we are responsible for it but also do our part to ensure that we do not become ignorant of the real world and be isolated from society. That is, we should play our roles while taking a bird's-eye view and learning about applied research and the operations necessary for its practical implementation.

From a management standpoint, we must also consider the value we create and its return. At risk of being too frank, I have to emphasize how important it is that we understand and communicate the value that our research provides. How the value of research is determined depends on the party determining it; however, it is important for researchers themselves to first identify the value of their research and believe in it. Since research does not tend to produce results in a planned or smooth manner, it is important for researchers to determine to some extent the pinnacle they are aiming for and pursue it.

Take initiative and be passionate while respecting others

—In light of your past experience, tell us what you value most.

I joined NTT laboratories in 1993 after completing my doctoral course. When I was a student, I studied creating devices by exploiting the properties of materials. After joining NTT labs, I was involved in researching new optical functional devices useful for optical communications and fabricating such devices in the clean room at the Musashino R&D Center, which gave me valuable experience. In 1997, to broaden my horizons, I spent a year as a visiting scholar at Stanford University, where I studied optical communications and transmission. That experience made me realize the importance of taking a bird's eye view of things.

At the time, a major movement toward optical communications began emerging, such as the establishment of the Optical Internetworking Forum, the global industry forum regarding optical networking. Since Stanford University is located in Silicon Valley, I was able to observe this movement firsthand and felt my horizons expanding. After returning to Japan, I worked on a research project to propose a network system from a device perspective, and I had the opportunity to learn the importance of investigating devices from the network system. Through these activities, I acquired a wide range of research skills and knowledge as well as human connections.

After that, I suddenly found myself in charge of human-resources development at NTT Photonics Laboratories, which was a completely unknown field to me. As I often say over drinks, that time was the only time in my NTT life that I lost weight (although it was only for my first month). I was concerned about everything because I did not know what to do. Although I briefly returned to a research job as a group leader, until I assumed my current position, I also had the opportunity to interact with many people in NTT operating companies when I was with the general affairs department, where I endeavored to connect NTT laboratories and operating companies in terms of human resources. I have also learned various things at each milestone, including the globalization of NTT, such as the acquisition of Dimension Data (now NTT Ltd.), a global system integration company.

From these experiences, I realized the importance of being a little broader than the research I am currently involved in and take an interest into a variety of areas. I also think that it may be better not to think too much when asked to change. I have come to my current position through a process that I never expected to experience as a researcher. Maybe that is because others can judge me better than I can judge myself. I think we generally tend to avoid or resist change. In the case of researchers, they sometimes need to abandon or change their research themes for various reasons. This situation may be tough for some researchers; even so, we should see change as an opportunity to enrich ourselves as human beings, and by moving up the steps presented by that opportunity, we will become stronger people. Everything depends on our mindset, and there is no doubt that change makes people grow.

—Finally, what do you think is the role of top management? And would you give a message to researchers inside and outside the company?

I believe that the role of top management is to generate a perturbation without people being aware of it. As the world is changing, top management needs to make changes (i.e., generate perturbations) for their staff. I believe that at times of change, by communicating with staff in a way that they do not notice the change (that is, by not making them aware of the perturbation), they can move forward without being aware of the change. Once people make progress, they will understand the change as they progress, realize that the change is interesting and fun, progress in their understanding, and develop a mindset of taking initiative in and having passion for their research.

A Nobel Prize laureate in physics Arthur Leonard Schawlow once said, “The most successful scientists are not the most talented. But they are the ones who are impelled by curiosity.” I quote these words every time I give a lecture to research staff at my laboratory group. Even if one’s knowledge and abilities are not sufficient, there is always room to improve, and taking initiative and having passion rooted in curiosity are the keys to success. However, you should not recklessly rush forward with only your own thoughts; listening to and accepting advice from others and respecting others are also key factors in improving your research capability.

The words of Goro Yoshida, the first director of the Electrical Communication Laboratory, “Do research by drawing from the fountain of knowledge and provide specific benefits to society through commercial development,” are the basic philosophy of NTT laboratories today. Keeping this philosophy in mind, we at NTT Science and Core Technology Laboratory Group strive to invent concepts that differ from past ones and constantly create something new that originates from us while collaborating with parties inside and outside NTT for creating things that cannot be done by us alone. We also want to become an organization that carves its own path to the future.

To our researchers. NTT’s laboratories are places where you can take on various challenges without fear of failure. The themes of cutting-edge and basic research may not be understood or communicated well at first, but top management and your seniors are patient and thorough in their efforts to unravel and understand them. Therefore, please take initiative in and be passionate about your research.

To our research partners. Our research is like a rough stone—although it may be unique and of great value, it may not seem attractive. When this rough stone is washed downstream and collides with other stones, its surface becomes smooth, shiny, and beautiful, sometimes shines like a diamond. If that happens, the stone’s value will be widely recognized by people and it will contribute to society. This process requires time-consuming polishing and processing. By combining our rough stones (i.e., cutting-edge technologies) with your technologies, we want to contribute to creating a sustainable, safe, and secure society that embraces diversity, namely, the goal of the IOWN concept.

Interviewee profile

■ Career highlights

Akira Okada joined NTT in 1993. He became a senior researcher at NTT Photonics Laboratories in 2003, senior manager in charge of human-resources management and development at the General Affairs Department in 2009, senior manager in charge of research promotion at the Planning Department, NTT Science and Core Technology Laboratory Group in 2015, vice president, head of NTT Device Innovation Center in 2016, and vice president, head of NTT Device Technology Laboratories in 2017. He has served as head of NTT Science and Core Technology Laboratory Group since July 2022 and was appointed senior vice president of R&D in June 2023.

Toward a More Accurate and Easy-to-use System for Rapidly Evolving Machine Translation

Masaaki Nagata
Senior Distinguished Researcher, NTT Communication Science Laboratories

Abstract

Chat Generative Pre-trained Transformer (ChatGPT), a generative artificial intelligence (AI) chatbot developed by OpenAI, is rapidly gaining attention worldwide. ChatGPT generates sentences as a response to input sentences or words and outputs the sentences in the language of the input sentences unless otherwise specified. However, it can be used for machine translation; that is, it outputs a translated sentence when the output language is indicated. Masaaki Nagata, a senior distinguished researcher at NTT Communication Science Laboratories, has been researching natural-language processing and its application, machine translation, for more than 20 years. We asked him about the trends and characteristics of translation using large language models (LLMs) and machine translation using a Japanese-English bilingual patent corpus, which is about to be commercialized. We also elicited his thoughts on the research process and ideas as being a result of encounters.

Keywords: machine translation, bilingual corpus, large language model

Machine translation has evolved from statistical machine translation to neural machine translation and translation using LLMs

—Could you tell us about the research you are currently involved in?

I'm researching—in the field of machine translation—a Japanese-English bilingual corpus and word alignment in that corpus.

In the previous interview in this journal (June 2021 issue), I talked about the period of transition from statistical machine translation to neural machine translation. With neural machine translation, which has been rapidly gaining in popularity over the past few years, it is important to collect bilingual data. Therefore, we created bilingual patent data based on our experience of creating a Japanese-English bilingual corpus called JParaCrawl containing over 20 million sentence pairs through web crawling. Patents are public documents, and all documents of patent applications in Japan, USA, and other countries are publicly available. Since we had the expertise to create large-scale bilingual data, we were able to use it to create a Japanese-English bilingual patent corpus with over 300 million sentence pairs (**Fig. 1**).

One possible application of this corpus is creating and checking related documents when applying for

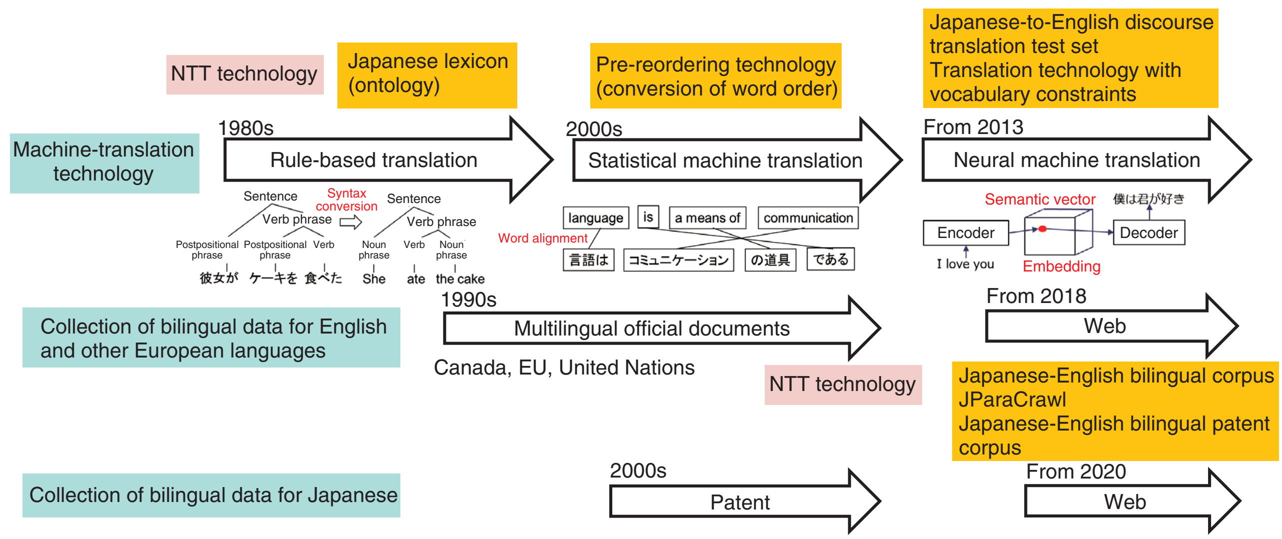


Fig. 1. History of machine translation and research and development at NTT.

patents overseas. For translation of patent-related documents such as patent applications prepared when applying for a patent overseas, it is necessary to maintain the relationship between a modification and the modified subject and to select appropriate technical terminology to secure the necessary rights. Logical precision and rigor are required, even if it means sacrificing some level of fluency, which is important in general translation. To evaluate if our corpus satisfies these requirements, we ranked the translation results of three types of artificial intelligence (AI)-based translation engines: (i) a patent-specific translation engine built using our Japanese-English bilingual patent corpus, (ii) a general-purpose translation engine, and (iii) a patent-specific translation engine by using Japanese and English patent applications of Mitsubishi Heavy Industries (including claims) as evaluation data. The quality of translation was ranked by employees working on patent applications in the intellectual-property departments of both Mitsubishi Heavy Industries and NTT. The average rankings of the three engines were respectively 1.5, 2.0, and 1.9, which indicates that the highest ranking was achieved by the patent-specific translation engine built using our bilingual patent corpus. An automatic evaluation based on similarity to the correct translation prepared in advance by a professional translator showed a significant improvement in translation by the patent-specific translation engine built using our bilingual patent corpus (scoring 57.5 out of 100 points) compared with 38.6 points for the general-purpose translation engine and 44.0 points for the other patent-specific translation engine. Mirai Translate, an NTT Group company, plans to provide a patent-specific translation engine built using our bilingual patent corpus as a service.

Although the accuracy of machine translation has increased considerably, it tends to decrease as sentences become longer, as in the case of patent claims. Our recent research theme is to clarify whether it is possible to improve the accuracy of translation of long sentences by using the world's most accurate word-alignment technology that we have created or to create a system that automatically identifies mistranslations or potential mistranslations. To improve translation accuracy in fields other than patents, we are considering the use of large language models (LLMs), which have been gaining popularity over the past year.

—How do you use an LLM for translation?

To put it simply, machine translation using an LLM is like asking a multilingual speaker who has learned a lot of things in English to paraphrase sentences in one language into another language. To understand the difference between translation using an LLM and conventional neural machine translation, it is necessary to look at the history from neural machine translation, to transformers, to the appearance of LLMs. In neural machine translation, the meaning of a sentence is represented by a real-number vector with about

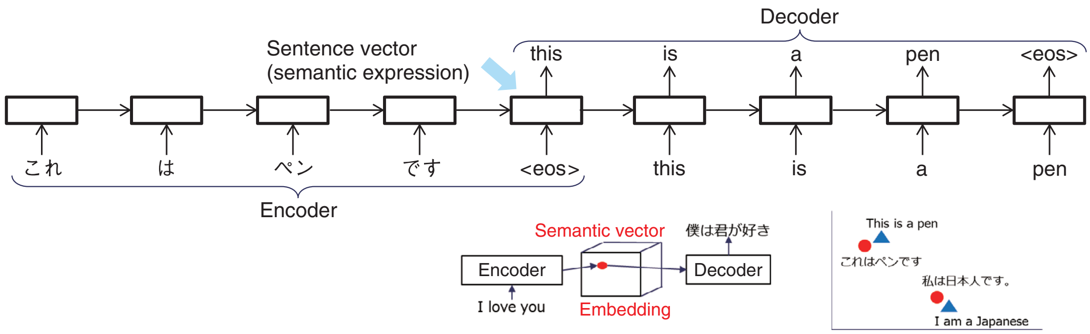


Fig. 2. Neural machine translation (RNN encoder-decoder model).

1000 dimensions then translated by coupling an encoder, which converts an input sentence in one language into a numeric vector, with a decoder, which converts this numeric vector into an output sentence in another language. In regard to the early version of neural machine translation, an encoder and decoder are configured as a recurrent neural network (RNN) into which words are input one by one, the internal state of the RNN is updated, and words are output one by one according to the internal state (**Fig. 2**). The internal state of the RNN when the special symbol <eos> (representing the end of the input sentence) is input contains information about all the words in the input sentence, so it is treated as a semantic representation of the input sentence (semantic vector). Training this RNN encoder-decoder model with a large amount of bilingual data will, for example, place the semantic vectors of the sentence pair "this is a pen" in English and its translation in Japanese ("kore wa pen desu") in close proximity. In other words, the similarity of sentences corresponds to the distance between semantic vectors in vector space. Neural machine translation translates by sharing one vector space between two languages in this manner.

Encoders and decoders using an RNN are afflicted with two problems. First, they represent input sentences as fixed-length vectors, so long sentences are difficult to represent; second, they input words one by one and output words one by one, so parallelization is difficult. To solve these problems, a neural network called a transformer was devised (**Fig. 3**), which uses an attention mechanism to selectively collect information from multiple targets. A transformer is composed of an encoder, which selects information from all the words in the input sentence and reconstructs each word vector of the input sentence, and a decoder, which selects information from each word vector of the reconstructed input sentence as well as all the word vectors in the previous output and determines the next word to output. A transformer is more accurate than an RNN encoder-decoder model, and the attention mechanism allows for parallelization, which makes it possible to train translation models on large amounts of bilingual data and to scale the translation models. Our Japanese-English bilingual patent corpus also uses a transformer to create translation models.

An LLM called Generative Pre-trained Transformer (GPT)-3, on which OpenAI's ChatGPT is based, has been attracting attention. As its name suggests, GPT-3 is a transformer model but consists of only a decoder, and the model is scaled and trained with a large amount of text data. The decoder-only transformer model—called a language-generation model or simply language model—is a neural network that uses an attention mechanism to select information from all the previously output word vectors and determines the next word to output. For GPT-3, a model with 96 layers and 175 billion parameters is trained with text composed of 0.3 tera tokens. It is called an LLM because it is approximately 1000 times larger than previous translation models and language models. ChatGPT, an interactive chatbot based on GPT-3, can be used comfortably in Japanese, and for some reason, it can also translate Japanese to other languages. However, most of the training data used for GPT-3 is in English, and Japanese accounts for only about 0.2% of the total data. Although it is not yet clear why ChatGPT can be used comfortably in languages other than English, it is probably due to the fact that all the languages included in the GPT-3 training data share one semantic vector space at a fine-grained level such as words (**Fig. 4**).

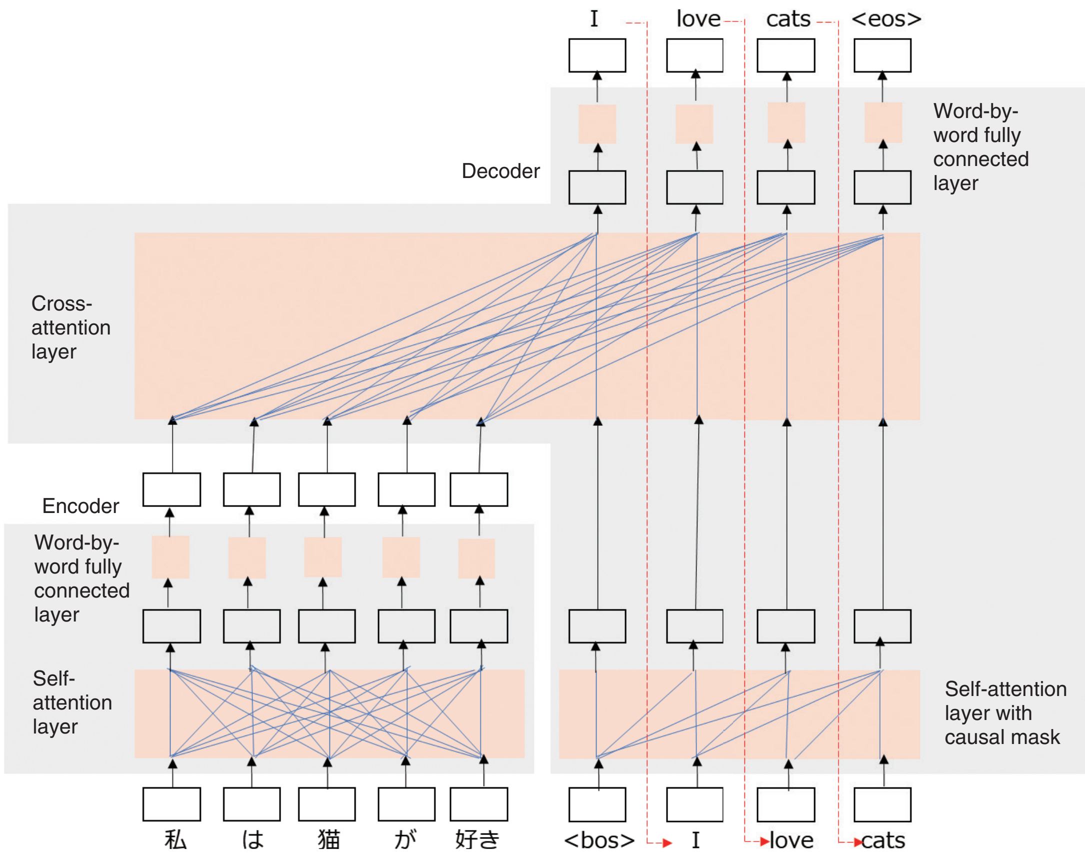


Fig. 3. Neural machine translation (transformer).

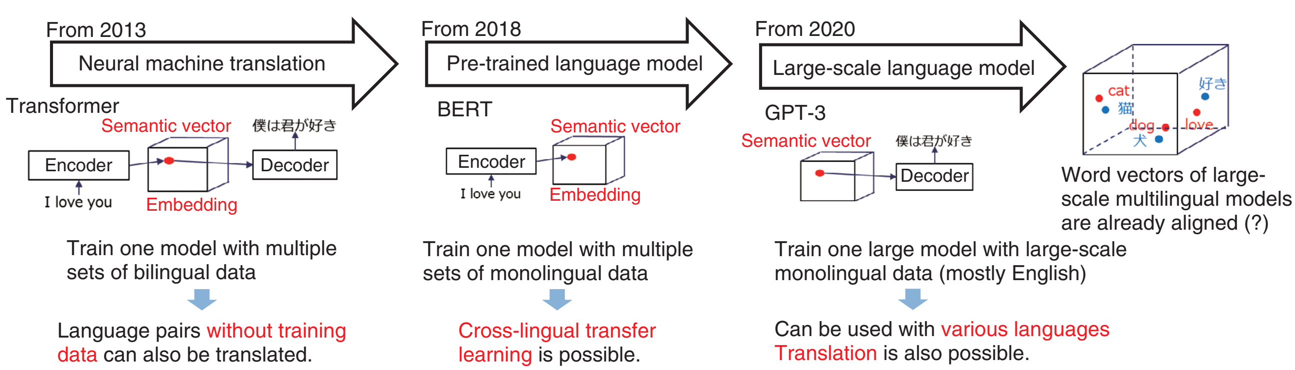


Fig. 4. Multilinguality of language models.

LLM translation can be likened to, for example, a bilingual person reflexively speaking English after hearing something in Japanese.

—ChatGPT is evolving rapidly, but what does the future hold for machine translation?

At the present stage, translation accuracy of machine translation using LLMs is equivalent to that of conventional neural machine translation; however, compared with neural machine translation, LLM-based machine translation uses 1000 times more parameters, so the processing is naturally slower. Due to its large scale, it is also difficult to prepare and operate a system for LLM-based machine translation; thus, I predict that conventional neural machine translation will not disappear. However, it is also true that translations using ChatGPT are popular with users. I think this popularity is due to the fact that ChatGPT can be controlled using natural language and can do things such as consider longer contexts, handle specified style and terminology, and convert input/output formats.

Under these circumstances, research topics concerning machine translation include evaluation of the quality of machine translation using LLMs, detection and correction of errors, confirmation of factuality, and detection of biases such as gender bias. The current trend of creating huge, supreme intelligence is expected to reach its limit at some point. I therefore believe that our future challenge will be to (i) create instruction-tuning data that reflects the usage of machine-translation systems by professional translators and (ii) develop a smaller LLM "translation assistant" with around 10 billion parameters that enables users to instruct machine translation and its peripheral functions in natural language.

The research process and ideas are a result of encounters. Increase the chances of encounters through trial and error

—What do you keep in mind as a researcher?

In my experience, I feel that much of the research I have conducted while aiming at achieving a result presumed in advance has not been successful. Some things I tried led to new results. For example, for the aforementioned LLM-based translation, fine-tuning the LLM with ordinary bilingual data did not work; however, through trial and error, I found a pattern that worked. Such a good process and good ideas are a result of encounters.

Therefore, to have many chances for such encounters, I value meeting with people, including those involved in joint research, and trial and error. In joint research, we try to avoid overlapping work as much as possible and repeat trial and error without preconceptions; however, each researcher may do what they want, and sometimes that leads to good results. I also think that many encounters occur during periods of technological innovation. During periods of innovation, little established practice exists, so it is inevitable that people do a lot of trial and error. With the advent of LLMs, we are currently in the midst of a period of innovation, so I want to focus on trial and error from the perspective of translation.

—What is your message to younger researchers?

My original research theme was natural-language processing, but for the past 20 years, I have been working on themes related to machine translation, which is one type of natural-language processing. I enjoy learning foreign languages, so this field is one that I am interested in. Basic research involves a long-term endeavor, so unless you can maintain your interest in it, you will not be able to work on it for a long time. If you research in the same field for a long time, the external factors surrounding your research will change from time to time, so you will have to study to keep up with those changes; however, by keeping your purpose and goal in mind, you can move forward without losing sight of yourself.

In the previous interview, I said something like "do something different from others and be as controversial as possible in a good way." I'm not saying one should be different from others. As researchers, we must always be aware of—even within the same field—tackling a new research problem, gaining new knowledge, and creating new technologies, which mean doing something different from others. To tackle a new research problem, you have to understand the current situation, study cutting-edge technology, then go through the trial-and-error process. Since this trial-and-error process cannot be done alone in many cases, it is necessary to try various directions in collaboration with different people, which requires discussions. I think a solution to a problem can be found by repeating this process of trial and error.

■ **Interviewee profile**

Masaaki Nagata received a B.E., M.E., and Ph.D. in information science from Kyoto University in 1985, 1987, and 1999. He joined NTT in 1987. His research interests include morphological analysis, named entity recognition, parsing, and machine translation. He is a member of the Institute of Electronics, Information and Communication Engineers, the Information Processing Society of Japan, Japanese Society for Artificial Intelligence, the Association for Natural Language Processing, and the Association for Computational Linguistics.

Achieving Complete Wireless Communication through Estimation and Prediction Technology for Radio Wave Propagation Characteristics

Motoharu Sasaki
Distinguished Researcher, NTT Access Network Service Systems Laboratories

Abstract

Today's network is continuously evolving through a variety of wireless communication systems. Wireless communication is now available just about anywhere—it has become an integral part of everyday life and brought us many benefits. At the same time, wireless communication is predicted to become increasingly complex in the future, so there is a need for further evolution of the wireless network to deal with this complexity. NTT Distinguished Researcher Motoharu Sasaki says, "It is important that the network be easy for people to use without having to think about it." In this article, we talk with him about his activities as a rising researcher including his research on "achieving complete wireless communication through estimation and prediction technology for radio wave propagation characteristics."

Keywords: wireless communication, radio wave propagation, incentive control

Achieving "complete" wireless communication overcoming the problems of a complex wireless network

—Dr. Sasaki, what kind of research is "achieving complete wireless communication through estimation and prediction technology for radio wave propagation characteristics?"

In "achieving complete wireless communication through estimation and prediction technologies for radio wave propagation characteristics," I am mainly carrying out two types of research. The first is technology for instantaneously estimating radio wave propagation characteristics and predicting future characteristics using various types of sensing data. The second is incentive control technology that uses the above-mentioned technology in such a way that users and system providers of wireless communication systems unintentionally act in a cooperative manner throughout the entire system. With these two types of technologies, our aim is to achieve "complete" wireless communication that links and integrates a variety of wireless communication systems by advanced means.

To give some background to this research, public

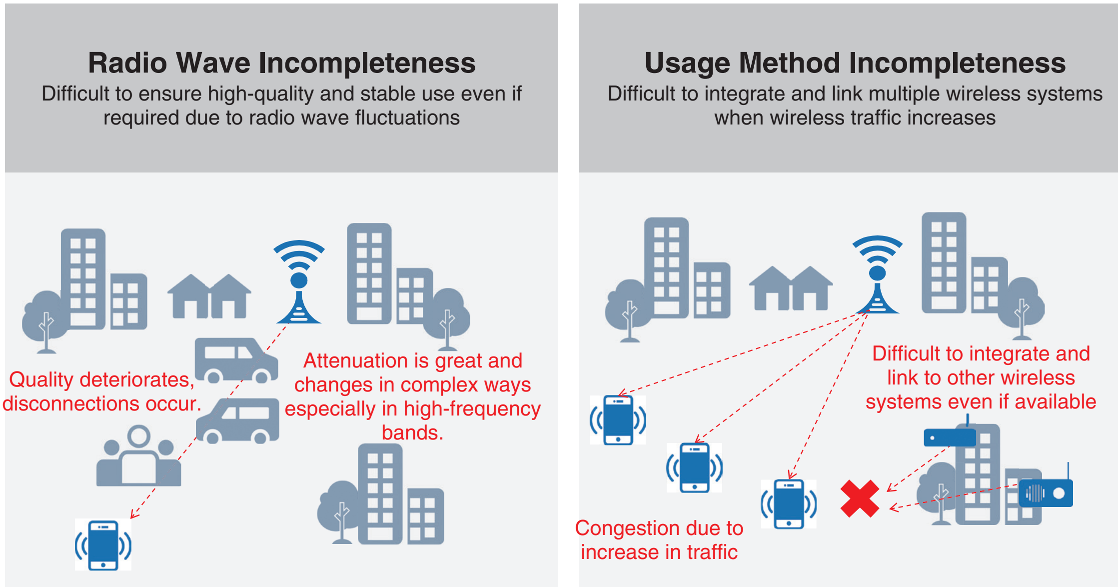


Fig. 1. Problems in a complex wireless network.

cellular is now shifting from the fourth-generation mobile communication system (4G) and Long Term Evolution (LTE) to the fifth-generation mobile communication system (5G). Similarly, a wide variety of wireless communication systems are being implemented over a broad range of frequency bands including private cellular and wireless local area networks (LANs) as well as wireless communication systems oriented to the Internet of Things (IoT) such as low power wide area (LPWA). Going forward, we can expect an even further expansion of diverse applications such as IoT and machine-to-machine (M2M) communications, so we believe that wireless communication systems and the way they are used will become all the more complex. Under these conditions, there is a need to process user communication requests that are constantly changing, and to meet this need, it is imperative that such a complex wireless network be properly used. However, the skillful use of a complex wireless network is not easy—there are a variety of problems that must be dealt with such as the inability to ensure high-quality and stable use due to fluctuating radio wave conditions (radio wave incompleteness) and the difficulty of easily using another available wireless system when traffic temporarily increases (usage method incompleteness) (**Fig. 1**). To overcome these problems in such a complex wireless network, I am researching means of achieving complete wireless communication through estimation and prediction technologies for radio wave propagation characteristics.

—What are you specifically researching?

In my research and development (R&D) activities after entering NTT, I worked to grasp and model radio wave propagation characteristics, the foundation for constructing a wireless communication system. This work often involved problems related to wireless communication at NTT operating companies, and the consultation I was asked to give frequently came down to radio wave characteristics with questions like "Radio waves are invisible, so how can I understand how they propagate?" Additionally, I often heard the comment "Even if radio waves are, in fact, propagating correctly, communication quality may still deteriorate due to usage conditions of other wireless communication users." I took these actual comments from operating companies as an opportunity to begin my current line of research.

We can grasp the behavior of invisible radio waves by instantaneously estimating a variety of radio wave conditions and predicting their future state. Specific approaches in this regard use not only static information such as building, computer aided design (CAD), and building information modeling (BIM) data but also various types of sensing data including dynamic

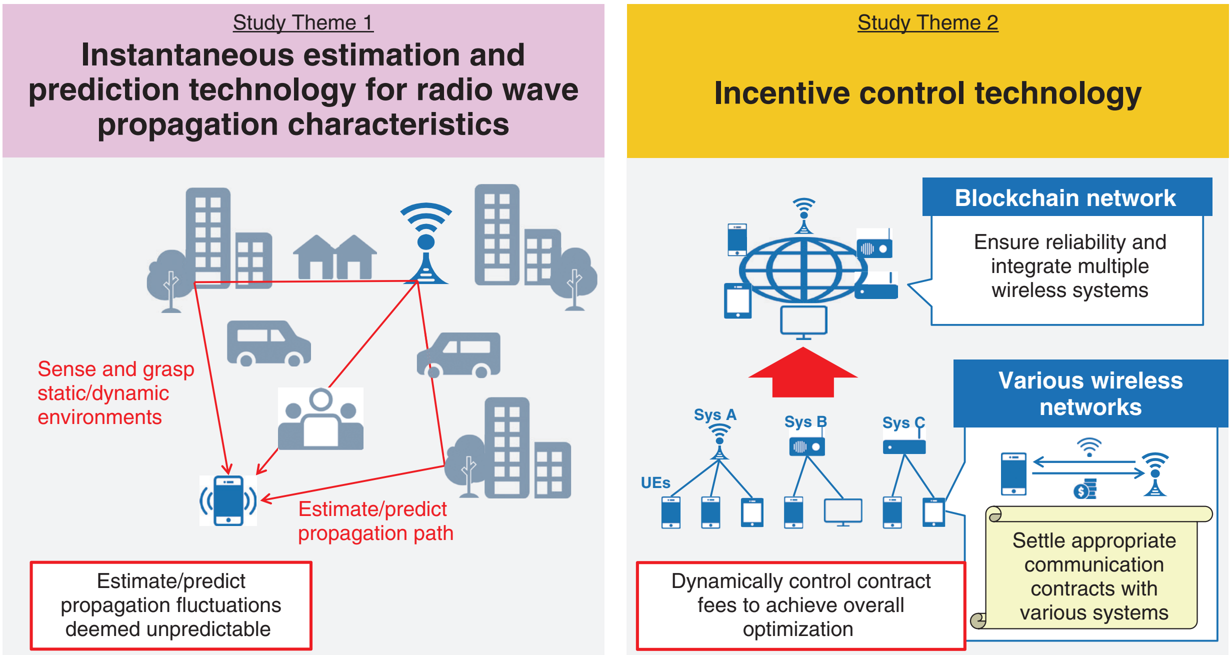


Fig. 2. Research overview.

data such as images, light detection and ranging (LiDAR), and positioning data. In addition, parameters of various types of fluctuation distributions based on these data can be used to make predictions including the probabilistic behavior of radio waves. In this way, it becomes possible to instantaneously estimate changes in radio wave propagation and predict future characteristics, which has not been possible up to now. With this technology, I believe it will be possible to grasp the radio wave propagation environment surrounding each terminal and base station and optimize wireless communication.

I am also undertaking incentive control technology that skillfully sets incentives for operators and users of wireless communication systems so that they unintentionally cooperate with each other throughout the entire system. Here, we use a mechanism for wireless network sharing using blockchain technology so that users and system providers naturally cooperate with each other and a variety of wireless communication systems operate as if they are a single integrated system. My aim here is to achieve a transition from a "world in which radio resources are isolated and users compete for them" to a "mutually cooperative world" (**Fig. 2**).

—What points have been particularly difficult in your research?

In technology for estimating and predicting radio wave propagation characteristics, and particularly in a commercially used network like public cellular, the difficulty is in estimating and predicting those characteristics amid unknowable information. For example, past studies on radio wave propagation characteristics usually assumed that both base-station positions and terminal positions were known, so it was easy to infer the degree to which radio waves would attenuate based on distance. In my research, however, there are also cases in which base-station positions and detailed parameters on the network side must be handled as unknowns. For this reason, my main pursuit is to replace such unknowns with information that can be estimated and predicted through data analysis of various types of peripheral information and parameters that can be grasped from terminals.

Next, in incentive control technology, I am researching and developing technologies using blockchain, but jumping into this new area combining wireless communications and blockchain is difficult. Up to now, there has been no precedent in incorporating the blockchain concept in the wireless communication field. There is therefore a need for

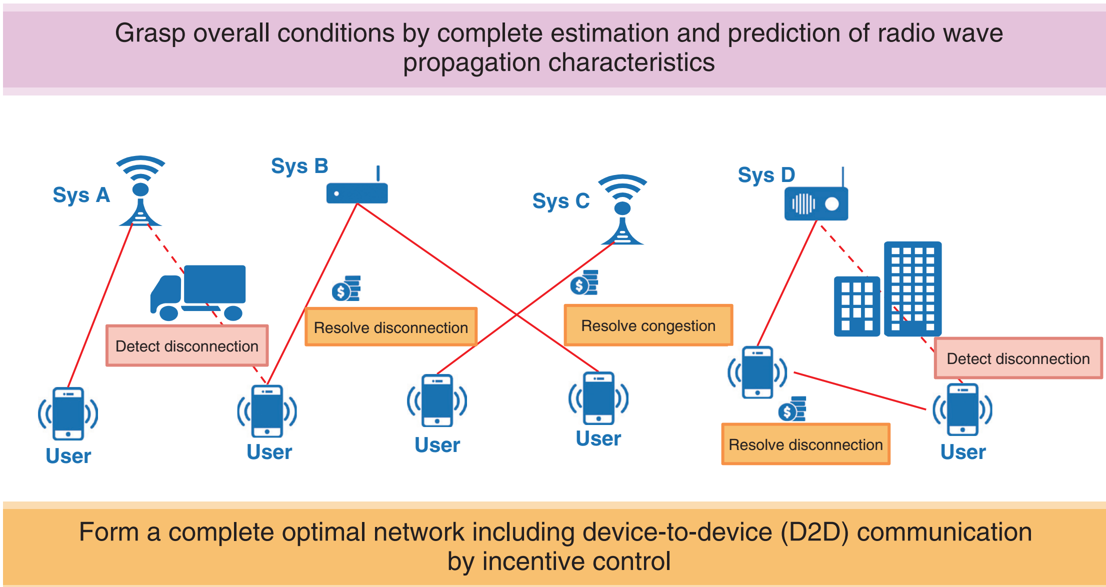


Fig. 3. Worldview targeted by research themes.

knowledge in both wireless communication and blockchain to turn this knowledge into services at NTT operating companies. However, a single organization in an operating company that attempts to deal with this need on its own has a large hurdle to overcome. I am also aware that conformity with the current legal system must be considered when talking about a mechanism for sharing wireless LAN access points (APs) using blockchain.

—Please tell us your research achievements to date.

With respect to estimation and prediction technology for radio wave propagation characteristics, I have established technology for predicting wireless communication quality such as received power about five seconds into the future. I have also used this technology for proof of concept and other demonstrations with an operating company with the aim of stabilizing wireless communication in mobile robots. At present, I am studying means of extending this capability to technology for predicting wireless communication quality such as transmission speed from received power. Regarding incentive control technology, I have performed a verification experiment of technology for sharing wireless LAN APs held by various individuals. This experiment was announced in an NTT press release issued in April 2023. Although there have been previous studies using blockchain mechanisms in the field of communications, there have been almost none on this type of blockchain application. Against this background, this research has been received with amazement and praise from many quarters—there are high expectations for the future implementation of this technology.

Looking to the future, I plan to expand my estimation and prediction technology for propagation characteristics, verify its effects through verification experiments conducted with NTT operating companies and others, and promote its implementation as technology for the Innovative Optical and Wireless Network (IOWN) era. Additionally, once incentive control technology is established in FY2024, I would like to apply it to actual services (**Fig. 3**).

Obsessed with research that produces results and accelerates the coming of IOWN

—Please explain the relationship between your research and NTT's IOWN vision.

The use of wireless communication is one important element of delivering ultra-high-speed and ultra-low-latency communication services up to user terminals as declared by the IOWN vision. In the present wireless communication system, each network operator constructs a system and provides services by selecting whatever is necessary for each of the features it can provide such as throughput, communication distance, power consumption, and stable quality. In the IOWN world, the aim is to enable a user to connect naturally to different wireless systems without having to be aware of those differences and to obtain the wireless communication performance needed. This concept is called “extreme Network as a Service (NaaS),” and in my research, I am working on creating those basic technologies.

At NTT Access Network Service Systems Laboratories, R&D of a group of control technologies called Cradio® is progressing to achieve extreme NaaS. These technologies will combine multiple wireless communication systems that can be used in a natural way. All of these research efforts will help accelerate the meeting of objectives and will make a contribution as technologies that become the IOWN platform.

—What is your impression of NTT laboratories?

NTT Access Network Service Systems Laboratories performs widespread R&D supporting access circuits from NTT central offices to customers. In addition to world-leading academic and standardization achievements, it makes many practical contributions in a direct manner such as by solving service-related problems thanks to its strong links with NTT operating companies. In this way, I understand why NTT Access Network Service Systems Laboratories is known for “world-leading R&D of the most advanced field technologies.” In the first year of my master’s course, I came to NTT laboratories on an internship program and spent four weeks in doing desktop studies by simulations and experiment evaluations using prototype equipment. Carrying out these studies in an area different from what I was familiar with at school was difficult, but I learned what it felt like to make steady progress in R&D with a view to practical applications in contrast to what I was doing at school. It was at this time that I resolved to enter NTT. A big difference between NTT laboratories and other companies is that it provides an environment in which you can be involved in infrastructures that anyone can use in contrast to creating individual products like a manufacturer. I am interested in work that can have a major impact on society. Since entering NTT, I have been able to participate in various types of R&D that can actually be applied to a wide range of systems, so I think my feelings during my internship were not wrong at all.

In addition to NTT Access Network Service Systems Laboratories, there are other laboratories at NTT involved in a variety of R&D areas. These laboratories gather together talented people having diverse areas of specialties, which I think is undoubtedly the greatest strength of NTT laboratories. In fact, there are times when I make progress in blockchain-related research based on advice from individuals at other laboratories, for which I am extremely grateful.

—Dr. Sasaki, could you leave us with a message for researchers, students, and business partners?

In my daily research life, I place great importance on producing results. R&D is often an endeavor covering a long period of time, and even if you are somewhat behind in your daily research, that should not have a great impact on your overall schedule. However, if such delays accumulate, it may be too late to recover from the overall delay once you notice what’s happening. For this reason, I think it is vitally important to undertake your work while thinking “how can I quickly move forward in my research and produce results in a timely manner?” Additionally, while it is certainly important to proceed logically in your daily R&D efforts, I think that, in the end, working with enthusiasm and discipline will lead to results that make a difference.

Today’s wireless communication field is evolving rapidly from a technical perspective, and it seems that catching up is difficult even if adopting just a single wireless communication system and its standards. On top of that, a variety of wireless communication systems are constantly appearing and becoming increasingly complex. In this way, while wireless communication has come to be used throughout daily life providing much convenience, I think its full potential has yet to be tapped.

Current networks are constructed in a way that people must think about how they can be used and adapt accordingly. On the other hand, I think that making it so that "people can use the network with ease without having to think about it" is essential to drawing out the full potential of wireless communication. This is what I keep in mind in my daily research activities. The ultimate goal of my research is a "world in which all wireless communication systems behave like a single network that can be used without having to think about it." To achieve such a world, there is still a need for multifaceted R&D and trial-and-error experiments up to practical implementation. My wish is to carve out this new world while becoming colleagues with many people and collaborating with individuals from a wide range of backgrounds.

■ **Interviewee profile**

Motoharu Sasaki received his M.E. degree from Kyushu University in 2009 and entered NTT in the same year. He is a member of NTT Access Network Service Systems Laboratories. He received his Ph.D. degree from Kyushu University in 2015 and has been a Distinguished Researcher at NTT Access Network Service Systems Laboratories since 2022. He is engaged in the research and development of instantaneous estimation technology and future characteristics prediction technology for radio wave propagation characteristics. He received the Best Paper Award from the Institute of Electronics, Information and Communication Engineers (IEICE) in 2013, the Young Engineer Award from the Institute of Electrical and Electronics Engineers (IEEE) Antennas and Propagation Society Japan in 2016, and the Encouragement Award from the ITU Association of Japan in 2019.

NTT Human Informatics Laboratories: Researching and Developing Technologies That Nurture True Humanity

Kota Hidaka

Abstract

Based on the human-centric principle, NTT Human Informatics Laboratories is engaged in research and development related to new forms of co-existence between the real world and cyberworld. In the Feature Articles in this issue, we introduce NTT Human Informatics Laboratories' latest endeavors.

Keywords: human-centric, humanity, human functions

1. Mission of NTT Human Informatics Laboratories

At NTT Human Informatics Laboratories, our mission is to "research and develop technologies that nurture true humanity" while our vision is to "enable the information-and-communication processing of diverse human functions based on the human-centric principle." Of the characteristics that humans possess, we focus on six: sense perception, sensitivities, thoughts, behavior, body, and the environment. We are developing elemental technologies that convert these characteristics into data and use them in information-and-communication processing. To actualize the Innovative Optical and Wireless Network (IOWN) concept, we are using the above human-centric functions in the integrated research on Digital Twin Computing (DTC) and the remote world.

When we note the changes surrounding us, the following comes to the forefront: the emergence of generative artificial intelligence (AI), greater miniaturization and accuracy of brain-computer interface devices, disillusionment with the metaverse and Web3, and post-capitalism. In consideration of these trends, NTT Human Informatics Laboratories has extracted the following actions to focus on: (1) accelerating research with general-purpose AI that uses the brain as a black box, (2) initiating research with general-purpose AI that uses the brain as a white box, (3) pursuing the metaverse's essential and universal value, and (4) accelerating research directly tied to the humanities. On the basis of these areas, we have set large language models (LLMs) and neurotech/cybernetics as priority elemental technologies to accelerate our applied research. We have established "Project Metaverse" and "Project Humanity" as priority use cases that present the application of integrated research (**Fig. 1**).

2. Priority elemental technologies and use cases

We describe NTT Human Informatics Laboratories' directions of four priority elemental technologies and use cases. The first priority elemental technology/use case, LLMs, shows the reality that is possible with general-purpose AI, as exemplified by the arrival of OpenAI's Generative Pre-trained Transformer 3 (GPT-3) and its use in a variety of fields. We are researching and developing a proprietary LLM that will analyze the mechanisms of the brain by leveraging NTT Human Informatics Laboratories' decades of research in natural language processing.

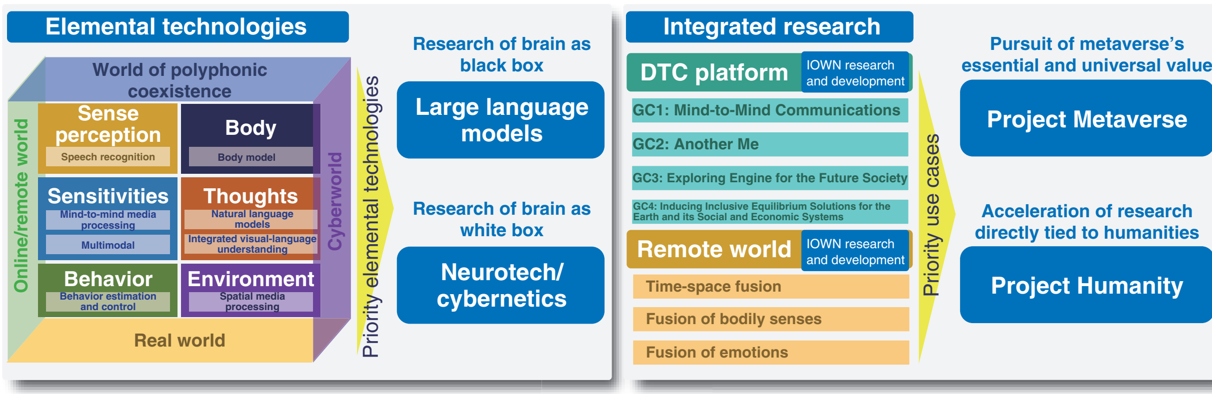


Fig. 1. Priority elemental technologies and use cases.

By quickly implementing general-purpose AI in society, we will be at the forefront of the new AI era, as well as take on the challenge of innovating the world with general-purpose AI.

For the second priority elemental technology/use case, neurotech/cybernetics, we will use the LLM we have developed, in addition to leveraging knowledge of the workings of the body that we have cultivated to date, to acquire tacit knowledge and develop intuitive human-machine interfaces.

For Project Metaverse, our third priority elemental technology/use case, with our LLM as the base, we will create a new form of "tele-" that transcends distance and time to provide encounters not considered possible before, work and leisure experiences that do not currently exist, and space to know oneself as a human being.

The fourth priority elemental technology/use case, Project Humanity, uses neurotech/cybernetics as the base to promote diversity and inclusion. It also provides ways to support people around us, such as family, friends, co-workers, and people with illnesses and disabilities as well as those supporting them by creating a world that can acquire human functions while respecting the wishes of the user (**Fig. 2**).

In this issue's Feature Articles, we report on NTT Human Informatics Laboratories' latest research and development (R&D) efforts on the above priority elemental technologies/use cases and activities to apply them. In this article, we introduce NTT's proprietary LLM.

3. NTT's proprietary LLM "tsuzumi"

The arrival of LLMs, such as GPT, has increased the feasibility of general-purpose AI. In November 2023, NTT Human Informatics Laboratories released "tsuzumi," NTT's proprietary, highly efficient LLM. This LLM addresses the issues faced by GPT: language-model size, reliability of information, model extensibility, applicability to non-language modalities, and power consumption associated with large-scale training. The following features differentiate tsuzumi from other LLMs: (1) compact language models (reduced cost), (2) superiority in Japanese language processing, (3) improved customizability, and (4) multimodal support (LLM with physical sensory capabilities).

Regarding tsuzumi's compact language models (reduced cost) (Feature 1), tsuzumi is available in two versions: ultralight with 0.6 billion parameters and light with 7 billion parameters. These versions can run high-speed inference with just one central processing unit (CPU) and one inexpensive graphics processing unit (GPU), respectively. In terms of the cost of using the GPU cloud for training, tsuzumi's ultralight version costs 1/300 and the light version costs 1/25 that of GPT-3 (175 billion parameters). For inference processing, the ultralight version costs 1/70 and the light version costs 1/29 that of GPT-3 (based on NTT's estimations).

The tsuzumi LLM's superiority in Japanese language processing (Feature 2) is due to NTT Human Informatics Laboratories' accumulated expertise from decades of research in natural language processing. The result is tsuzumi's high performance even with a small parameter size. In benchmark testing of LLMs using Rakuda, tsuzumi outperformed GPT-3.5 and the top Japanese LLMs (**Fig. 3**).

For improved customizability (Feature 3), tsuzumi

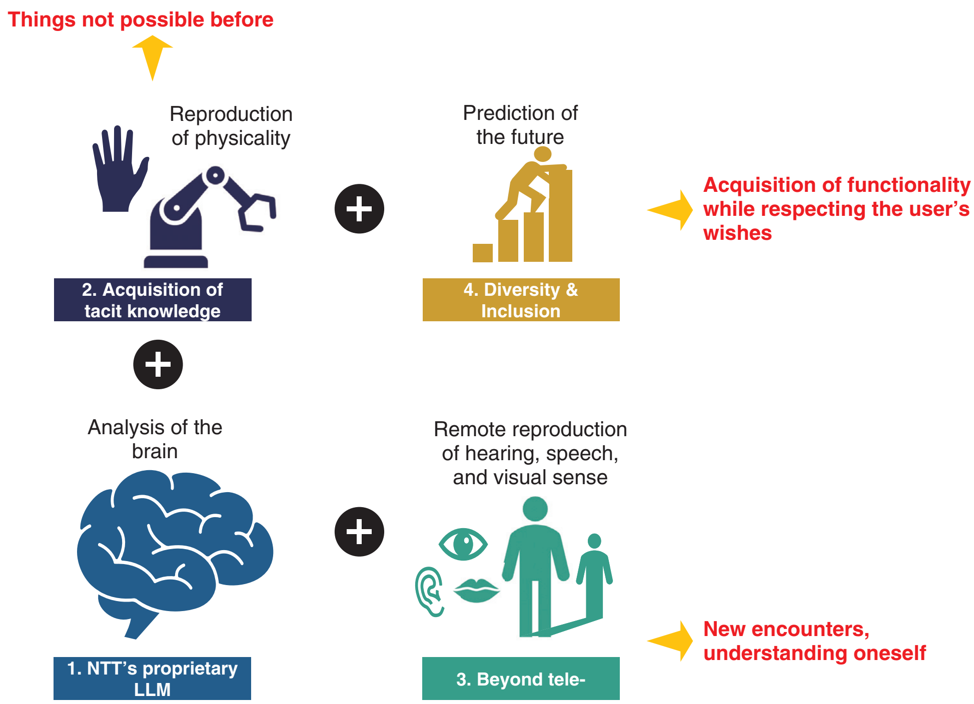
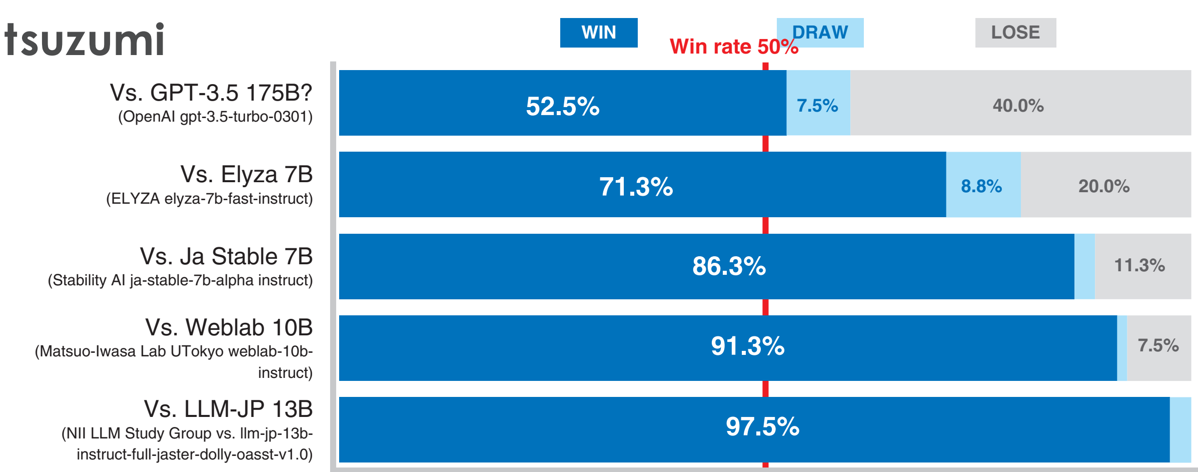


Fig. 2. Acquisition of human functions.



* Rakuda benchmark: https://yuzuai.jp/benchmark Ran on 22 Oct 2023.
Scored by evaluating each LLM using two-model comparisons by GPT-4 (40 questions x 2 in order of presentation).
Except for llm-jp, model output uploaded on the site were used in the evaluation.
llm-jp is excluded because input repetition and termination tokens are done as post-processing due to settings described in the model card of Hugging Face.
Evaluation scores were obtained by comparing all models listed on the leaderboard dated 27 Sep 2023 and tsuzumi-7b using two-model comparisons by GPT-4 and ranking the results using Bradley-Terry strengths.

Fig. 3. Superiority of tsuzumi in Japanese language processing.

offers three tuning methods to flexibly respond to different requirements such as accuracy and cost: prompt engineering, full fine-tuning, and adapter tuning (**Fig. 4**). These tuning methods enable industry-specific customization at low cost.

The tsuzumi LLM's multimodal support (Feature 4) enables acquisition of a wider range of knowledge by not only language but also other modes of input

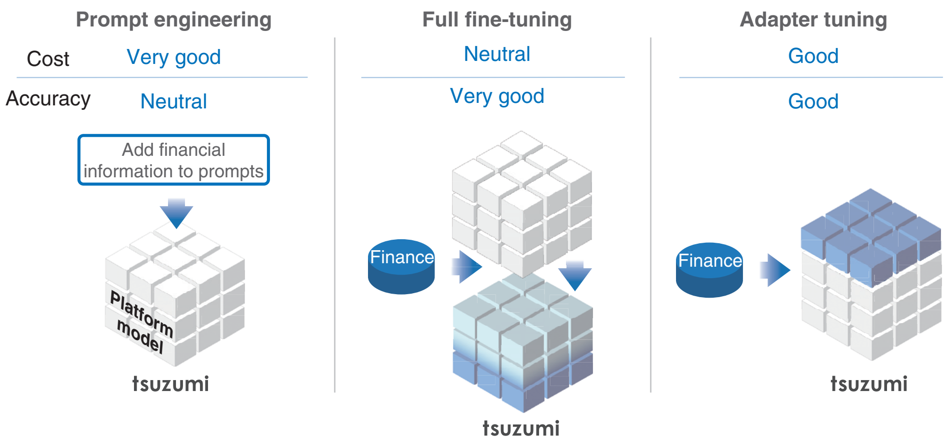


Fig. 4. Three different tuning methods.

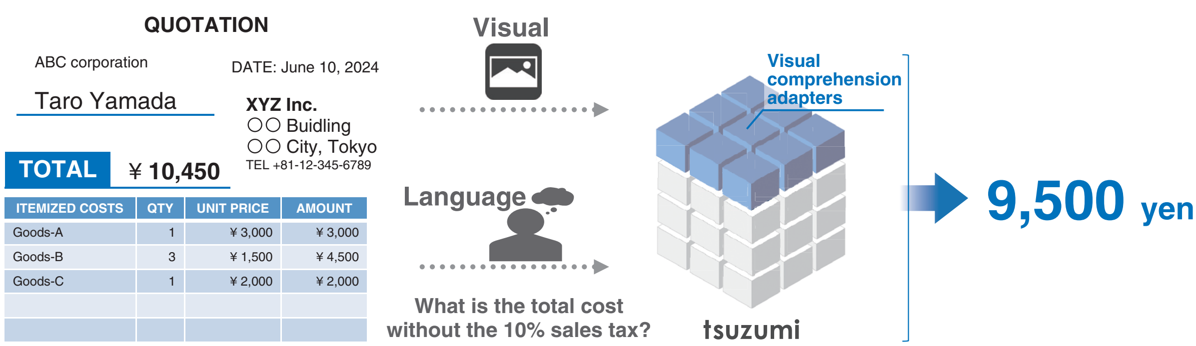


Fig. 5. Multimodal support.

and output, such as images, video, sensor data, nuances in speech, and facial expressions (**Fig. 5**). We are creating use cases centered on tsuzumi by linking it with NTT Human Informatics Laboratories' voice-, image-, and video-processing technologies, which have a history of more than 40 years, and using the knowledge of sensor and actuator technologies and cognitive psychology cultivated through our research on robotics.

4. Synergy between tsuzumi and IOWN

IOWN's high-capacity, low-latency network provides the foundation for connecting the geographically dispersed resources necessary for an LLM. It also contributes to power saving in LLM operations. In the future, social issues will be solved not by using one massive LLM that knows everything but by linking together small LLMs with specialized knowledge and personas, including tsuzumi.

5. Message to everyone

It has been three years since the outbreak of the COVID-19 pandemic, transforming our lives. It made us realize once again that we cannot take for granted the idea that we can meet anyone, anytime. We now understand that when there are restrictions on non-essential/non-urgent activities, feeling happy is not always considered essential or urgent. Our surrounding environment is changing at a dizzying pace. When the Reiwa era (Japanese era based on the reigns of emperors) began in 2019, we could not have imagined what things would be like five years later. NTT's company name includes the word "tele-," meaning "at a distance," and NTT's business has been focused on communication that transcends distance. Amid the changes over the past few years, we have continued to contribute to spreading and advancing the remote world by using information and communication technology. Our lives are still not yet settled, but it is time to examine from a human-centric perspective what has come about in the past few years and what we should leave behind for future generations. When we cannot see our loved ones, are there ways of communicating that transcend the barriers of time and distance? How much should machines imitate humans? Are there areas where we can allow machines to exceed human capabilities? How should technology provide support for people to experience convenience and happiness? Through collaboration with external parties, we will continue to ask ourselves these questions. By repeatedly testing our theories, we are focusing our resources on technologies that we desire to leave behind for future generations, even generations 100 years from now. The best from NTT Human Informatics Laboratories is still yet to come.

Kota Hidaka

Vice President, Head of NTT Human Informatics Laboratories.

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

Toward Enabling Communication Connecting Mind and Mind, Body and Body, and Mind and Body

Hiromu Miyashita, Naoki Hagiyama, and Yushi Aono

Abstract

To advance research on more intuitive interfaces that connect people to people and people to machines and to achieve interfaces that are user-friendly for everyone regardless of age or disability, we focus on researching neurotechnology using brainwaves and cybernetics using techniques such as electromyostimulation. In this article, we introduce our latest research.

Keywords: neurotech, cybernetics, human interface

1. Introduction

NTT Human Informatics Laboratories has long been actively researching ways to facilitate smooth and rich communication, and concepts such as interfaces using bio-signals, enabling computer operations based on mental imagery, have long existed. There has recently been significant progress in miniaturizing and improving the precision of devices that acquire vital information such as brainwaves and electromyography and that generate movement by directly electrically stimulating muscles. Therefore, the results of our research are becoming possible to introduce into society. On the basis of the evolution of such devices, we at NTT Human Informatics Laboratories are researching neurotechnology and cybernetics.

In this article, we highlight recent research achievements, including:

- "Sensibility-analysis technology based on similarity of brain representations" for visualizing states of mind contained in brain activities and using them to enhance communication
- "Facial-expression-recognition tendency-modification approach" for facilitating smooth communication by controlling facial expressions through subtle electrical stimuli
- "Motor-skill-transfer technology" for introducing the ability to move oneself or one's avatar as desired using vital information such as brainwaves and electromyography

2. Sensibility-analysis technology based on similarity of brain representations

We aim to enable mind-to-mind communication beyond language, behavior, and cultural norms. Our research focuses on developing brain-representation visualization technology, which interprets and expresses various types of brain information (i.e., brain representations) such as emotions and cognitive states extracted from brain activities. We have proposed several interfaces, including a system that takes inputs from brain activities and converts them into aura effects associated with avatars in the metaverse. These interfaces enable a direct and visual understanding of brain representations. As part of these efforts, we developed a sensibility-analysis technology based on similarity of brain representations that reveals personal sensibilities and differences in perception between individuals.

This technology sequentially presents images to the user wearing an electroencephalogram (EEG) headset. It then conducts similarity analysis of brain activities on the basis of the EEG data recorded when the user views each image. By using the analysis

Fig. 1. Example of a sensibility map created from brain activity when viewing images of fashion styles.

results, a sensibility map is created, where elements are arranged linearly with similar items placed close together and dissimilar ones farther apart (see **Fig. 1**). The sensibility map visually illustrates users' subjective interpretations of styles and features of the image content. For example, it conveys ideas such as "Images located near each other have a similar impression" or "Images separated by only one image evoke a unique feeling."

On the sensibility map, elements with particularly high similarity are drawn to be included in a single group, and their overlap and inclusion relationships are represented as a Venn diagram. Furthermore, labels representing each group are generated from the captions of included images. These labels linguistically describe the user's perspective when viewing these images.

As an application to creative activities and capabilities (creativity), we implemented a function that takes multiple images on the sensibility map as input and combines their semantic features to generate a new image. With the advancement of generative artificial intelligence (AI), non-professionals can now achieve high-quality expressions. However, it is challenging for the average user to vividly imagine what they want to express then transcribe it into an instruction text (prompt). By understanding the user's sensibility and generating images through EEG analysis and visualization, it will become possible to directly incorporate the user's sensibility into generative AI, improving output.

We conducted multiple evaluation experiments on several categories of images such as paintings and label designs. Regarding the generation of landscape paintings, this technology using brainwave analysis outperformed conventional prompt-based inputs. Users said that they felt they could express what they wanted to express, showing significantly higher evaluation scores in terms of expression accuracy, sense of identity, and preference for generation results. This superiority can be attributed to the difficulty in linguistically describing characteristics and impressions of landscapes painted in artistic styles such as realism and impressionism.

When comparing the results of similarity analysis of brain activities with the data from subjective evaluations (adjectives of impressions received from images), we observed considerable variation among individuals in the items that showed correlations. However, certain items showed consistent correlations across individuals, including perceptual expressions (e.g., "bright" and "hard") and subjective evaluations (e.g., "beautiful").

The results suggest the potential to extract not only aesthetic values but also individual-specific impressions from brain representations. By improving methods that use brainwave analysis for generation, our approach enables even those who are not proficient

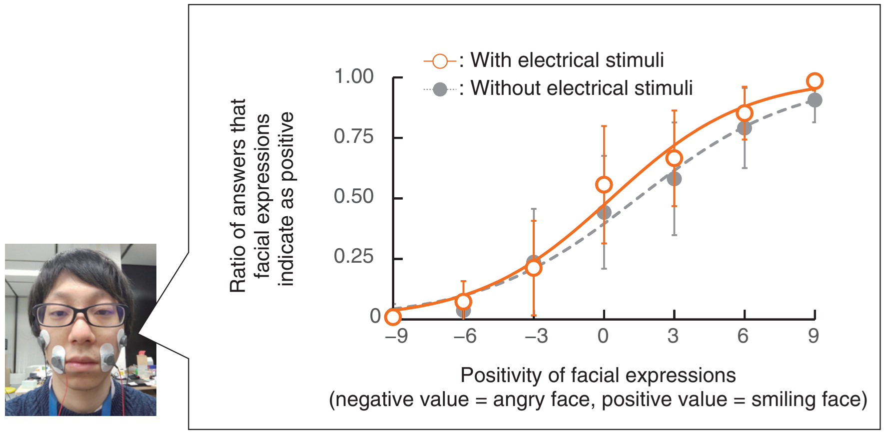


Fig. 2. Application of electrical stimuli to the cheek's facial muscles and resulting change in facial-expression recognition.

in using AI for expression to achieve their desired creative expressions.

Interfaces incorporating brain information, including the sensibility-analysis technology based on similarity of brain representations introduced in this article, are anticipated to enhance people's creativity and expressive desires, contributing to an information infrastructure that will foster richer communication. We will continue our research and development efforts to create a society where individuals with diverse characteristics share sensibilities, fostering mutual growth and creativity with AI.

3. Facial-expression-recognition tendency-modification approach

We aim to support smooth communication by modifying the recognition tendencies of facial expressions through providing subtle stimuli. As there are individual differences in how people interpret emotions from others' facial expressions, these differences can sometimes lead to communication discrepancies. Therefore, we explore methods of transforming facial-expression-recognition tendencies and promote desirable perceptions.

Our approach focuses on the facial feedback hypothesis, which suggests that the movement of facial muscles creating expressions can affect a person's emotions through feedback to the brain [1]. Previous studies related to this hypothesis suggested that moving facial muscles affects the elicitation and recognition of corresponding emotions [2]. For example, moving the zygomatic major muscle in the cheek can induce positive emotions such as happiness, making it easier to perceive others' expressions as positive. Building upon this phenomenon, we hypothesized that applying subtle stimuli to facial muscles could induce facial feedback effects and modify facial-expression-recognition tendencies.

We conducted experiments to investigate the modification of facial-expression-recognition tendencies when providing subthreshold electrical stimuli to facial muscles. Participants were presented with facial expressions displaying various intensities of anger or laughter. They judged whether the facial expression was positive or negative, and the results were compared between conditions with and without electrical stimuli. The results suggested that when electrical stimuli were applied to the facial muscles in the cheeks, participants were more likely to judge facial expressions as positive compared with the condition without electrical stimuli (see **Fig. 2**).

We aim to further understand the relationship between the type and location of stimuli and the modification of facial-expression recognition. We will also attempt to explore and identify more effective stimuli.

4. Motor-skill-transfer technology

For many years, we have been advancing research in motor-skill-transfer technology, using devices such as electromyography sensors to measure and record the movements of human muscles and transfer movements to oneself and others using myoelectrical stimulation. Through experiments, we demonstrated the effectiveness of sensing and transcribing skilled practitioners' movements to beginners, for example,

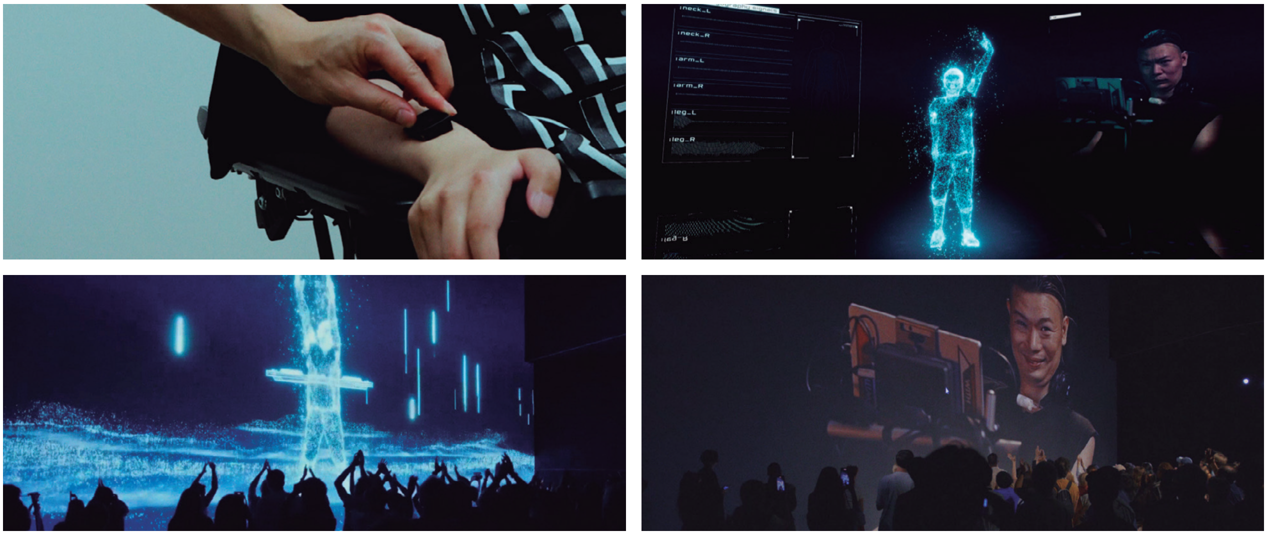


Fig. 3. Live performance by a DJ/artist with ALS.

allowing individuals to efficiently perform correct movements, particularly in activities that are challenging to master, such as musical instrument playing or sports. This technology has potential applications in situations where age or disabilities hinder individuals from moving as intended. By transcribing the movements of healthy individuals, including their past selves, those facing challenges in movement due to aging or physical impairments may regain motor skills. Our research has expanded beyond muscle movement transcription to include movement control based on brainwaves and avatar control within virtual reality environments such as the metaverse.

In the field of movement control based on brainwaves, we developed the Neuro-Motor-Simulator, a model that captures the intricate mechanism of muscles moving and actions being performed on the basis of brain commands. Our aim is to create a technology that accurately reproduces movements from brainwave signals. This technology holds the potential to evolve into an artificial spine for individuals with spinal cord injuries, allowing them to regain motor capabilities.

We are also exploring the control of avatars in metaverse spaces by sensing brainwaves and electromyography. Collaborating with a DJ and artist who has amyotrophic lateral sclerosis (ALS), a condition in which muscle movement gradually becomes difficult, we conducted experiments to control avatars in metaverse space by sensing electromyography from minimally responsive muscles. This research involved attempting live performances within the metaverse (see **Fig. 3**).

5. Future initiatives

At NTT Human Informatics Laboratories, we will accelerate the advancement of various areas of neurotechnology and cybernetics research beyond the technologies introduced in this article. Our goal is to create new communication technologies that directly connect minds and minds, bodies and bodies, and minds and bodies. Through these technologies, we aim to build a world where people can understand each other regardless of gender, age, culture, interests, etc. We also strive to create a world where anyone, regardless of age or disability, can move their body as envisioned.

References

[1] F. Strack, L. L. Martin, and S. Stepper, "Inhibiting and Facilitating Conditions of the Human Smile: A Nonobtrusive Test of the Facial Feedback Hypothesis," J. Pers. Soc. Psychol., Vol. 54, No. 5, pp. 768–777, May 1988. https://doi.org/10.1037/0022-3514.54.5.768

[2] W. Sato, T. Fujimura, T. Kochiyama, and N. Suzuki, "Relationships among Facial Mimicry, Emotional Experience, and Emotion Recognition," PLOS ONE, Vol. 8, No. 3, p. 57889, Mar. 2013. https://doi.org/10.1371/journal.pone.0057889

Hiromu Miyashita

Senior Research Engineer, NTT Digital Twin Computing Research Center.

He received an M.A. in computer science from Keio University, Kanagawa, in 2010. He joined NTT Cyber Solutions Laboratories (currently, NTT Human Informatics Laboratories) in 2010 and studied human-computer interaction, computer vision, and their applications.

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.

Naoki Hagiyama

Researcher, Symbiotic Intelligence Research Project, NTT Human Informatics Laboratories.

He received a B.E. and M.E. from Hiroshima University in 2017 and 2019 and joined NTT in 2019. His research interests include emotion recognition, biological signal processing, and computer-supported cooperative work.

Project Metaverse: Creating a Well-being Society through Real and Cyber Fusion

Naoto Abe, Hiroshi Chigira, Kenta Ogo, Takayoshi Mochizuki, Yoko Ishii, Shin-ichiro Eitoku, Lidwina Andarini, Katsuhiro Suzuki, Narichika Nomoto, and Atsushi Fukayama

Abstract

NTT Human Informatics Laboratories is researching and developing the "metaverse of the IOWN era" in which the real and cyber worlds converge on the Innovative Optical and Wireless Network (IOWN), NTT's next-generation communication infrastructure featuring ultra-high capacity, ultra-low latency, and ultra-low power consumption. This article introduces the latest research and development initiatives at NTT Human Informatics Laboratories, which are divided into space and humans (avatars), the main components of this metaverse.

Keywords: metaverse, Another Me, XR

1. Background

The word "metaverse" that appears in the title of this article originally appeared in the science-fiction novel *Snow Crash* published in 1992 as the name of a fictitious communication service in virtual space. It went on to become a commonly used noun as a variety of virtual-space services made their appearance with advances in technology. Today, it is a word that generally refers to virtual-space services on the Internet in which people can interact with each other via their alter egos called "avatars."

Metaverse-like virtual-space services have been quite diverse, beginning with chat services offering interaction via two-dimensional (2D) virtual space such as Fujitsu Habitat constructed in 1990 and extending to services such as Second Life (Linden Lab) that include sales of land and items and close-to-reality economic activities through currency in virtual space. More recently, they have come to include multiplayer online games such as Fortnite (Epic Games) and Animal Crossing: New Horizons (Nintendo) as well as business-oriented Horizon Workrooms (Meta Platforms) and Mesh for Microsoft Teams (Microsoft).

The metaverse concept and metaverse-like services have existed for some time, but there is still no uniform definition. To date, leaders of big tech companies and experts have offered a number of definitions and opinions, but after the American venture capitalist Matthew Ball compiled and presented the core attributes of the metaverse and features that it should provide in 2020, a single set of guidelines began to take shape [1, 2] (**Table 1**).

It has been pointed out that "several decades will be needed to build an infrastructure along with an investment of several tens of billions of dollars" before an extensive metaverse equipped with these attributes and features can be realized. Nevertheless, implementation of the metaverse in society is progressing as reflected in (1) the evolution and reduced cost of technologies related to virtual reality (VR),

Table 1. Seven features that the metaverse should have [1, 2].

Feature	Overview
1. Be persistent	• It never resets, pauses, or ends.
2. Be synchronous	• Prescheduled and self-contained events will happen just as in real life. • The metaverse will be a living experience that exists consistently for everyone in real time.
3. Be without any limit in the number of concurrent users	• Everyone can be part of the metaverse and participate simultaneously in a specific event, place, or activity.
4. Be a fully functioning economy	• Individuals and businesses will be able to create, own, invest, sell, and be rewarded for work that produces value recognized by others.
5. Be wide-ranging	Be an experience that spans • both the digital and physical worlds and private and public networks, and • open and closed networks.
6. Be interoperable	• Avatars, items, etc. can be freely carried between different platforms.
7. Have contributions from a wide range of enterprises and individuals	• Content and experiences will be created and operated by a wide range of contributors such as individuals, volunteer groups, and commercial enterprises.

augmented reality (AR), and cross reality (XR), the core technologies of a metaverse, and (2) growing expectations of virtual space even closer to reality due to the enhanced and enriching means of communication and diverse experiences that became possible during the recent COVID-19 pandemic.

2. Project Metaverse

At NTT Human Informatics Laboratories, we aim for a society in which everyone is in a state of well-being. Our mission is to "research and develop technologies that nurture true humanity" and our vision is to "enable information-and-communication processing of diverse human functions based on a human-centric principle." One of our initiatives toward making this vision a reality is Project Metaverse. As touched upon above, the main objective of most metaverse-like services had been to provide experiences different than those of the real world. It has therefore been difficult to directly apply experiences on a metaverse to the real world. While a variety of metaverses have come to provide services, the number of metaverses that can be used by a single user is limited due to physical constraints, which makes it difficult to enjoy a variety of experiences.

Against the above background, we would like to provide users with a wide array of experiences equivalent to those of the real world on multiple metaverses by achieving (1) ultra-real virtual space and cyber/real intersecting space and (2) avatars with identity and autonomy. Therefore, it should be possible to overcome not only space-time and physical constraints but also cyber/real barriers while increasing the opportunities of encountering the values of others. These achievements should lead to a state of well-being not just on an individual level but also for society as a whole.

In this article, we introduce specific Project Metaverse activities and related technologies from the following two perspectives:

- "Spatial representation and spatial fusion" to seamlessly connect the cyber and real worlds
- "Another Me" to enable diverse experiences in the cyber world

3. Spatial representation and spatial fusion

3.1 Spatial representation: constructing virtual space sensed with all five senses

Efforts at reproducing actual buildings and structures, historic sites, and landscapes in virtual space that can be experienced as a real metaverse have thus far been centered on tourism mainly with regional development and revitalization in mind. With these endeavors, a variety of methods have been considered from the research level to commercial services. However, it can be said that disparity between the real and cyber worlds with regard to a sense of presence is a common problem hindering a state of enjoyment and immersion in such a metaverse. In particular, in an experience such as enjoying scenery and nature while moving through a broad area, it would be costly to reproduce such a visually detailed, extensive space. It would also be difficult to reproduce the experiences and bodily sensations that would occur when actually visiting that location solely on the basis of visual information. For these reasons, it has

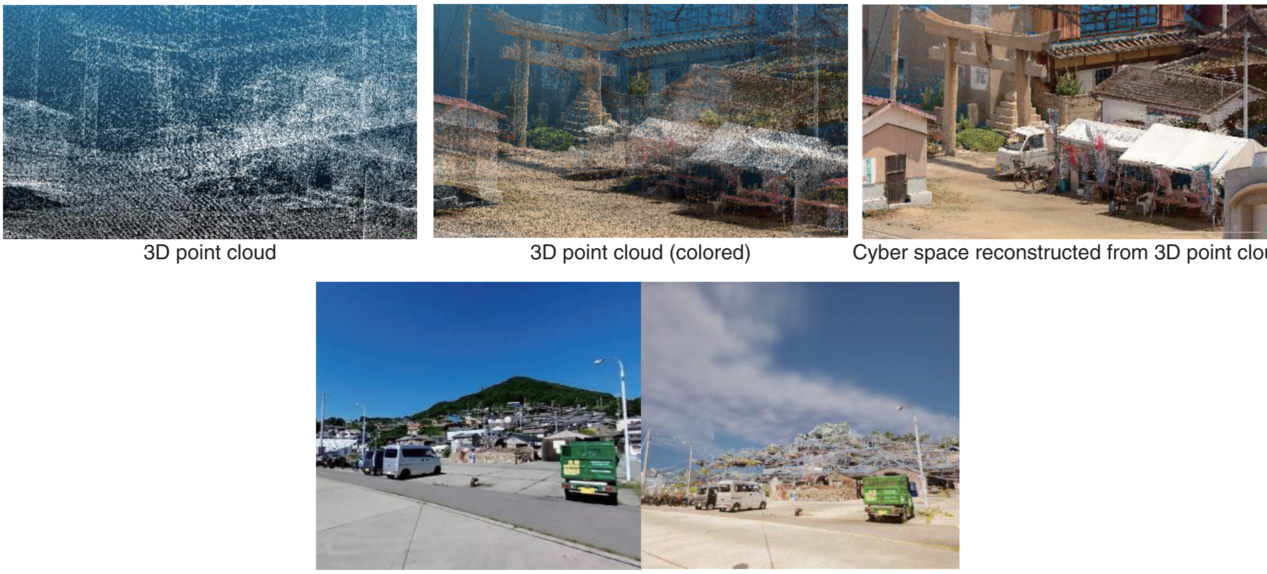


Fig. 1. Example of spatial reproduction using point cloud data (Ogijima Island).

thus far been difficult to provide a high sense of presence.

In light of the above, we undertook the construction of a virtual space that can be sensed with all five senses targeting Ogijima Island, a small island in the center of Japan's Seto Inland Sea [3]. In this project, we constructed a visually detailed, extensive virtual space while keeping costs down as much as possible by adopting (1) point cloud data obtained from a sensor technology called light detection and ranging (LiDAR) that can extensively and easily measure and grasp actual space and 3D point cloud media processing technology for analyzing and integrating that data, both of which are also used in self-driving systems, and (2) photogrammetry technology for creating realistic three-dimensional computer graphics (3DCG) from multiple photographs taken from a variety of angles (**Fig. 1**). In combination with the above, we used stereophonic acoustics technology and a newly developed walking-sensation presentation technique based on sounds and vibrations (**Fig. 2**) to faithfully reproduce and integrate sensory information "at that location" to stimulate other senses such as hearing and touch in addition to sight. Thus, we made it possible to provide users with the sensation of "actually being there" via the five senses. We consider that making experiences in virtual space equivalent to those in the real world as described above will make it possible to seamlessly use those experiences in the real world.

3.2 Spatial fusion: constructing intersecting space spanning two worlds

An avatar that acts as one's alter ego in a metaverse generally exists within a closed virtual space, and the range of its communication is likewise limited to virtual space. A user in the real world is therefore unable to confirm the state of such an avatar, which may give rise to a phenomenon called "echo chamber" or "filter bubble," which may generate division or discord between the cyber and real worlds.

We have therefore been researching technologies that will enable users (avatars) active in the cyber world and those in the real world to overcome the barriers between the two worlds and communicate with each other. One of these is High-precision VPS, which draws a positional correspondence between information in the cyber world (content and avatars) and the real world. Another is Glasses-free XR AISE-KI system, which enables communication with avatars and virtual characters in the cyber world without having to wear special equipment such as VR goggles.

The former technology uses the same technology that has recently come to be used in AR services, but by combining virtual space finely and extensively reproduced by point cloud data, as described above, with a smartphone or AR glasses equipped with a

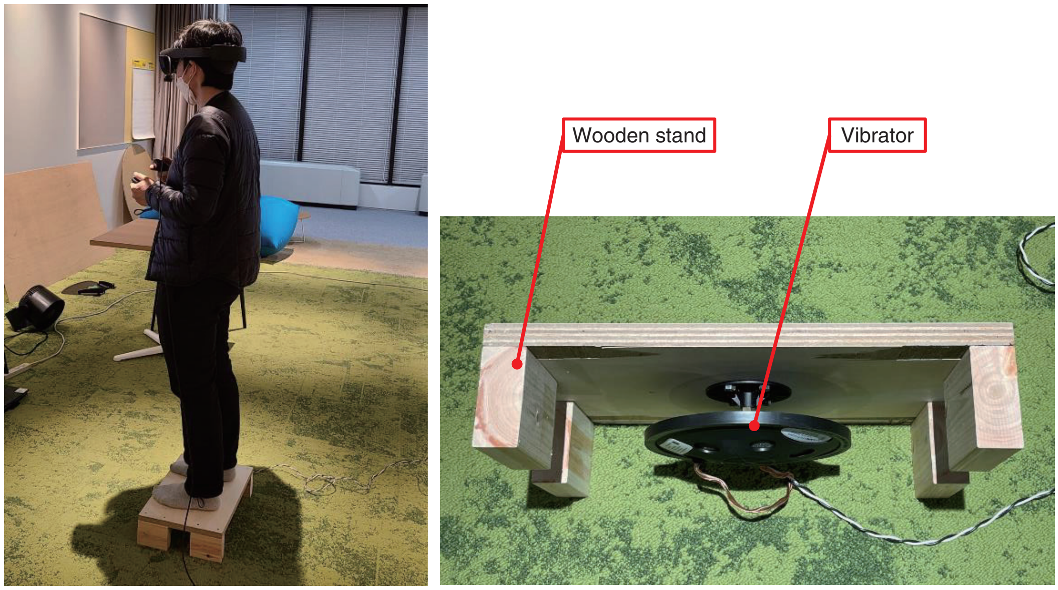


Fig. 2. Presentation of walking sensation using sounds and vibrations (experiencing scene).

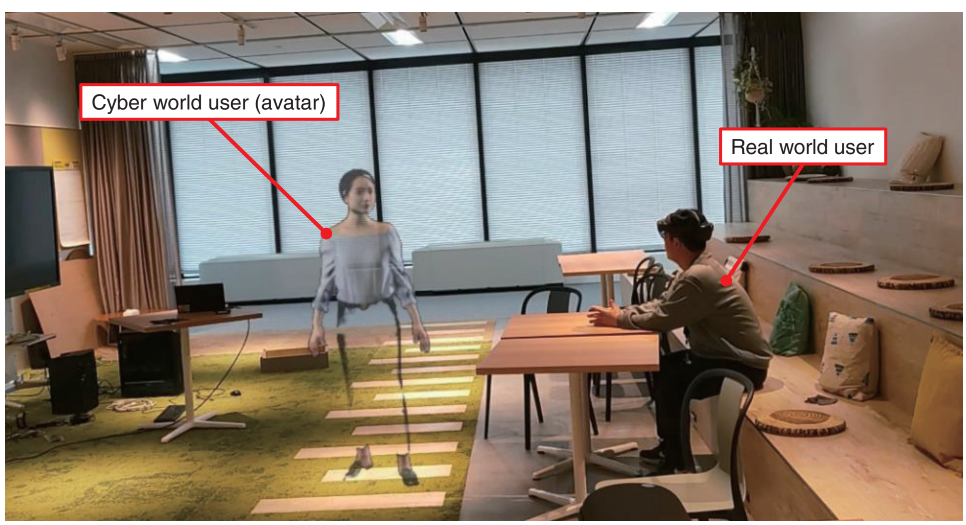


Fig. 3. Communication using High-precision VPS (image).

camera, it becomes possible to position users in the cyber and real worlds correctly in mutual space and enable them to communicate freely in the same space (**Fig. 3**). The latter technology will achieve intersecting space, enabling users in the real world to sit next to or share a table with an avatar, etc. in the cyber world through a multilayer 3D arrangement design of a half mirror, display, and a group of physical objects such as a desk or sofa. This technology will enable natural bidirectional communication between multiple users existing in both worlds without having to wear VR goggles or other equipment.

A Glasses-free XR AISEKI system was introduced at the Cho-Kabuki Powered by NTT event in April 2023 in the form of a shared-seating magic lantern tea house where conversation with virtual diva Hatsune Miku could be enjoyed while sitting alongside her in the same space. Visitors and experiencers evaluated

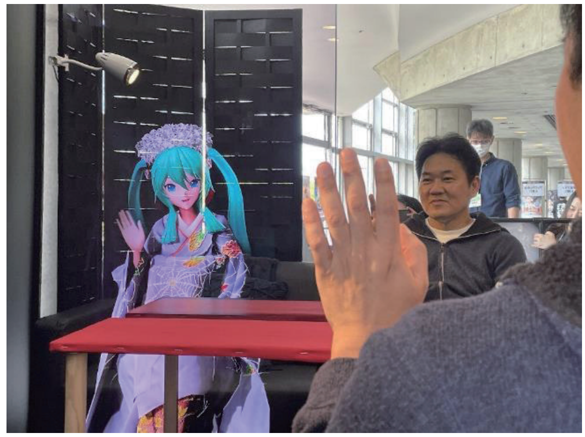


Fig. 4. Glasses-free XR AISEKI system (Experience sitting next to a hologram. Just like a sci-fi movie).

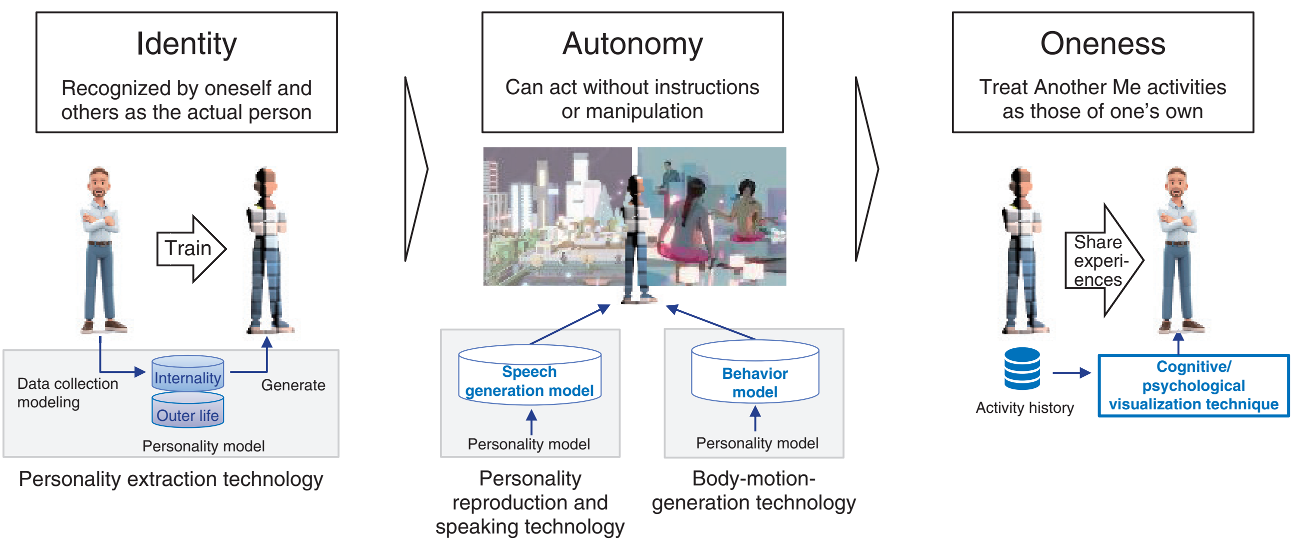


Fig. 5. Another Me (identity, autonomy, oneness).

the system highly giving comments such as "I could see Hatsune Miku in 3D!" and "Being so close to her gave me a thrill!" [4] (**Fig. 4**). We plan to achieve a more compact and frameless system that will make it even easier to use with a more immersive experience.

4. Another Me

4.1 Identity, autonomy, and oneness of Another Me

We aim to achieve a society in which everyone can overcome constraints such as time and space and even handicaps and expand the opportunities in their lives. Another Me simultaneously possesses "identity" that reproduces not only a person's appearance but also behavior and internality, "autonomy" whereby Another Me acts on its own in the metaverse beyond the constraints of the real world, and "oneness" in which the results of such activities are shared with the real person as an actual experience (**Fig. 5**). To achieve a level at which one can share their identity with their avatar to the point of feeling that the avatar is "me," we have undertaken research even from a philosophical point of view and have found that this sense of identity includes a desire to connect with other people and society. We can envision a variety of use cases in this regard such as consulting with an avatar in place of the real person using that person's specialized knowledge or carrying out work in collaboration with that person. On the basis of such use cases, we have been researching and developing elemental technologies for Another Me and working to implement Another Me in society using those technologies. In this article, we focus on three of these technologies—body-motion-generation technology, personality extraction technology, and personality reproduction and dialogue technology—and introduce our activities in applying them.

4.2 CONN for expressing natural body motion

We have been linking the facial expressions and full-body motion of the digital human "CONN" (**Fig. 6**), a virtual personality co-created by NTT Communications Corporation, NTT QONOQ, INC., and Toei Company, Ltd., with the body-motion-generation technology of Another Me and conducting experiments.

Body-motion-generation technology enables the generation of natural and human-like facial expressions and full-body motion while speaking from a small amount of data [5]. Generating a model from body motion, and the speech and its content uttered during such motion makes it possible to automatically generate body motion from uttered speech only. The model is trained to generate body motions in accordance with the meaning of the utterances made, smooth out body motions that reflect semantic coherence, and generate not only body motions that appear regularly but also those that appear in special contexts.

In the digital human CONN, human-like, natural body motion has been achieved by collecting and modeling the body motions of people acting as models and the speech uttered during those motions and generating motions tailored to the utterances made by CONN.

Fig. 6. Digital human CONN.

4.3 Personality extraction and reproduction by meta-communication service MetaMe™

With the aim of enabling anyone to have Another Me as one's alter ego, NTT Human Informatics Laboratories is testing and implementing personality extraction technology and personality reproduction and dialogue technology in MetaMe™*, a meta-communication service developed by NTT DOCOMO [6].

Personality extraction technology trains a vector for each individual that embeds information from that person's behavior log concerning personality and values that affect behavior [7] (**Fig. 7**). This technology enables behavior in tune with that person's values to be reproduced in Another Me. Therefore, similarity between values could be compared to find people with similar viewpoints or create a group with a diversity of viewpoints.

Personality reproduction in dialogue technology reproduces dialogue of a specific individual from a small amount of training data. This technology achieves this by combining persona-chat technology that reproduces dialogue according to the user's profile and adapter technology that learns personal features (tone, phrasing, etc.) from a small amount of dialogue data (**Fig. 8**).

Using these technologies, we developed a prototype of a non-player character (NPC) as an alter ego. An alter ego NPC communicates beforehand with another user on behalf of oneself so that communication between those two users becomes activated when they actually come to meet on MetaMe™, which can lead to the creation of new encounters or other opportunities. We are currently conducting a proof-of-concept experiment with NTT DOCOMO to see whether new social connections can be established between users.

* "MetaMe" is a trademark of NTT DOCOMO, INC.

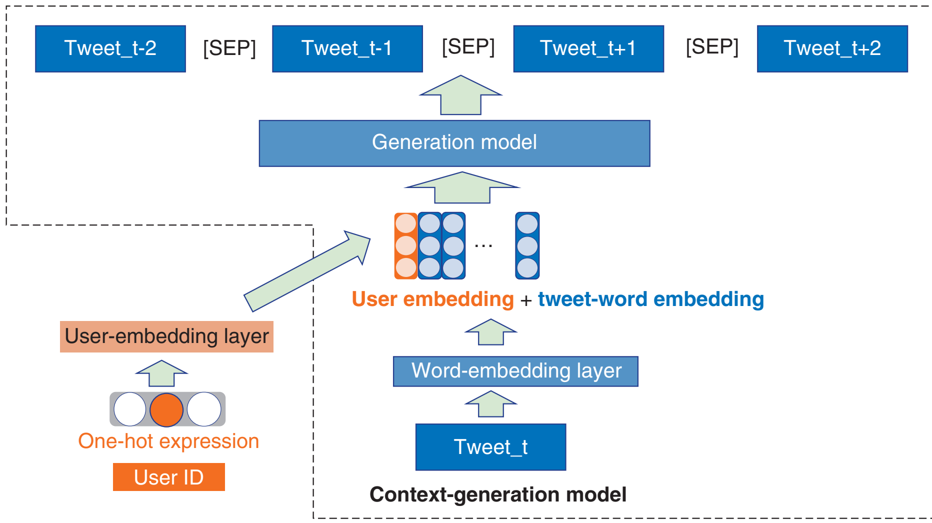


Fig. 7. Personality extraction technology.

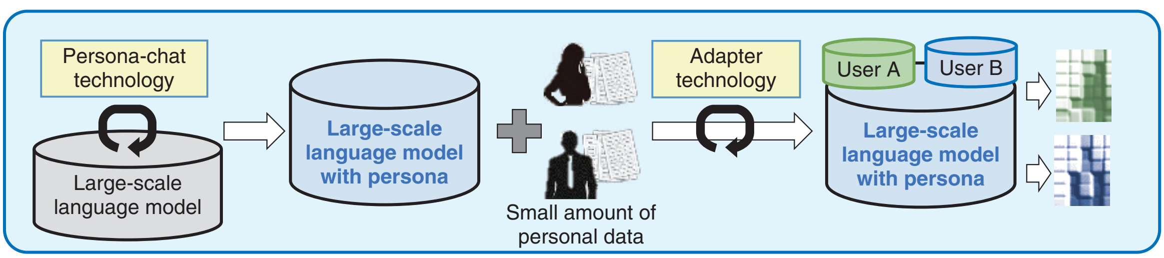


Fig. 8. Personality reproduction in dialogue technology.

5. Conclusion

As described in this article, most metaverse-like services provided thus far have been confined to virtual spaces with the result that the experiences provided have been different from those of the real world. However, as touched upon in Matthew Ball's "Seven Features that the Metaverse Should Have," user experiences in the metaverse of the future will span both the real and cyber worlds by seamlessly linking and merging these two worlds.

In our Project Metaverse initiative introduced in this article, we aim to merge real-world and cyber-world experiences, enhance the "connection" between people and people and between people and society, and create an enriching and prosperous society that embraces diversity. Please look forward to future progress in these efforts.

References

[1] M. Ball, "The Metaverse: What It Is, Where to Find It, Who Will Build It," 2020.

[2] The Japan Research Institute, Limited, "Metaverse Overview and Trends—Toward Use in Business Scenarios" (in Japanese). https://www.jri.co.jp/MediaLibrary/file/column/opinion/pdf/13531.pdf

[3] Press release issued by NTT on Nov. 15, 2022 (in Japanese). https://group.ntt/jp/newsrelease/2022/11/15/221115b.html

[4] Press release issued by NTT on Apr. 29, 2023 (in Japanese). https://group.ntt/jp/newsrelease/2023/04/29/230429a.html

[5] C. Ahuja, D. W. Lee, R. Ishii, and L. P. Morency, "No Gestures Left Behind: Learning Relationships between Spoken Language and Freeform Gestures," EMNLP: Findings, pp. 1884–1895, Nov. 2020. https://doi.org/10.18653/v1/2020.findings-emnlp.170

[6] Press release issued by NTT on Jan. 17, 2024 (in Japanese). https://group.ntt/jp/newsrelease/2024/01/17/240117a.html

[7] A. Otsuka, K. Hama, N. Nomoto, R. Ishii, A. Fukayama, and T. Nakamura, "Learning User Embeddings with Generating Context of Posted Social Network Service Texts," Social Computing and Social Media, pp. 106–115, 2023. https://doi.org/10.1007/978-3-031-35915-6_9

Naoto Abe

Senior Research Engineer, NTT Human Informatics Laboratories.

He received a Ph.D. in engineering from Hokkaido University in 2006 and joined NTT the same year. He is engaged in research and development regarding the metaverse. His research interests include information systems and applications for communications, accessible maps, and behavior change.

Hiroshi Chigira

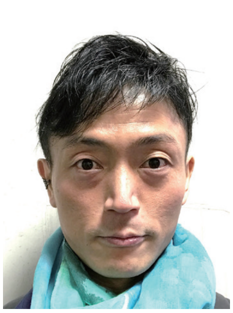
Senior Research Engineer, NTT Human Informatics Laboratories.

He received a B.E. and M.E. in engineering from Waseda University, Tokyo, in 2007 and 2009 and joined NTT in 2009. In 2016, he joined the 2020 Epoch-Making Project, where he has been engaged in several creative research projects combining technologies, design, and art to enable creative live sports viewing and artistic performance production for the Tokyo 2020 Olympics.

Kenta Ogo

Research Engineer, NTT Human Informatics Laboratories.

He received a B.E. from the University of Tokyo. He is engaged in research and development of IP streaming, video recommendation service, 360-degree VR telepresence with CAVE, 3D Social Metaverse, and machine learned AI chatbot service. He invented the “XR AISEKI” system that enables XR hologram experience with no VR headset required. The super natural fusion of real and cyber worlds is his research interest.



Takayoshi Mochizuki

Senior Research Engineer, Group Leader, NTT Human Informatics Laboratories.

He received an M.E. in human science from Waseda University, Tokyo in 2001 and joined NTT the same year. He is engaged in research and development aimed at creating a co-creation space (hyper-realistic cyberspace) where people can play an active role through free movement between real space and cyberspace.

Yoko Ishii

Senior Research Engineer, NTT Digital Twin Computing Research Center.

She received an M.S. in information systems from the University of Electro-Communications, Tokyo, in 2004 and joined NTT the same year. She received a Ph.D. in engineering from the University of Electro-Communications in 2009. Her research interests are human computer interaction and augmented reality.

Shin-ichiro Eitoku

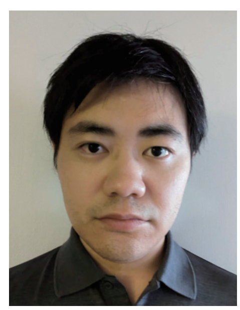
Senior Research Engineer, NTT Digital Twin Computing Research Center.

He received an M.E. and Ph.D. in information science and technology from the University of Tokyo in 2006 and 2013 and joined NTT in 2006. His research interests include information systems and multimedia systems for communications.

Lidwina Andarini

Research Engineer, NTT Digital Twin Computing Research Center.

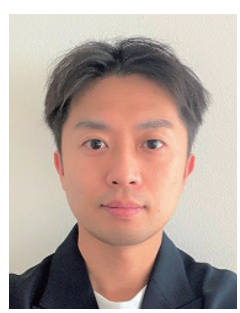
She received an M.E. from Nara Institute of Science and Technology. Her research interests include system platform development and human computer interaction technology. She is also involved in global business partnerships efforts within NTT labs.



Katsuhiro Suzuki

Senior Research Engineer, NTT Digital Twin Computing Research Center.

He received an M.S in engineering from Keio University, Kanagawa, in 2017 and joined NTT the same year. His research interests include virtual reality and human computer interaction technology.



Narichika Nomoto

Research Engineer, NTT Digital Twin Computing Research Center.

He received an M.S in media science from Keio University, Kanagawa, in 2007 and joined NTT the same year. His research interests include natural language processing.

Atsushi Fukayama

Senior Research Engineer, Group Leader, Another Me Research Group, NTT Digital Twin Computing Research Center.

He received an M.S. in informatics from Kyoto University in 1999 and joined NTT the same year. After working on media recognition technology, research and development of human computer interaction technology, and practical application development of network services, he began leading the Another Me Research Group at the NTT Digital Twin Computing Research Center in 2021.

Project Humanity: Providing Intimate Support to Respect the Humanity of Individuals

Mariko Nakamura, Ryosuke Aoki, Asuka Ono, Iwaki Toshima, Kengo Okitsu, Atsushi Ando, and Takeshi Mori

Abstract

At NTT Human Informatics Laboratories, we are engaged in Project Humanity aiming at solving problems on the basis of the human-centric principle of respecting the humanity that each person values and in a way that does not burden the user. In this article, we introduce five case studies in our effort to achieve Project Humanity.

Keywords: communication, human understanding, motor reconstruction, neurodiversity, spinal cord injury

1. Enrichment of communication for people living with ALS

As amyotrophic lateral sclerosis (ALS) progresses, cognition remains normal while muscle strength throughout the body gradually loses function. People with ALS can live their lives to the full term by using a ventilator, but they lose their voice due to the tracheostomy surgery required to put on the ventilator. Thus, the choice to continue living comes with the trade-off of losing speech. Due to the double loss of communication through both spoken language and physical expression, many people fear disconnection from society and lose hope of living. Globally, more than 90% of ALS patients refuse to wear a ventilator.

We have been developing cross-lingual text-to-speech technology, a type of text-to-speech technology that reproduces a person's voice from a recorded voice. This technology makes it possible to communicate in multiple languages while maintaining the person's unique tone of voice. In 2022, a DJ artist who is unable to speak due to ALS, was able to share dialogue and a musical performance in English using his tone of voice. In 2023, he took on the challenge of communicating nonverbally by slightly moving his body.

For communicating his bodily expressions through an avatar in the metaverse, NTT Human Informatics Laboratories used motor-skill-transfer technology based on the artist's biosignal information. On stage at the Ars Electronica Festival in Linz, a festival of art, advanced technology, and culture, the artist with ALS performed via an avatar remotely manipulated from Tokyo by just slightly moving his body. He stirred up the crowd by reproducing his bodily expressions that excited the audience before the progression of ALS.

Surface electromyography (sEMG) can be measured with sEMG sensors when muscles are moved even slightly. When the artist with ALS moved the parts of his body that could be moved slightly, even a few millimeters, sEMG measurement could be taken (**Fig. 1**).

When a person intentionally moves a part of the body, muscles in other parts of the body may react. We observed the muscles of other parts of the body

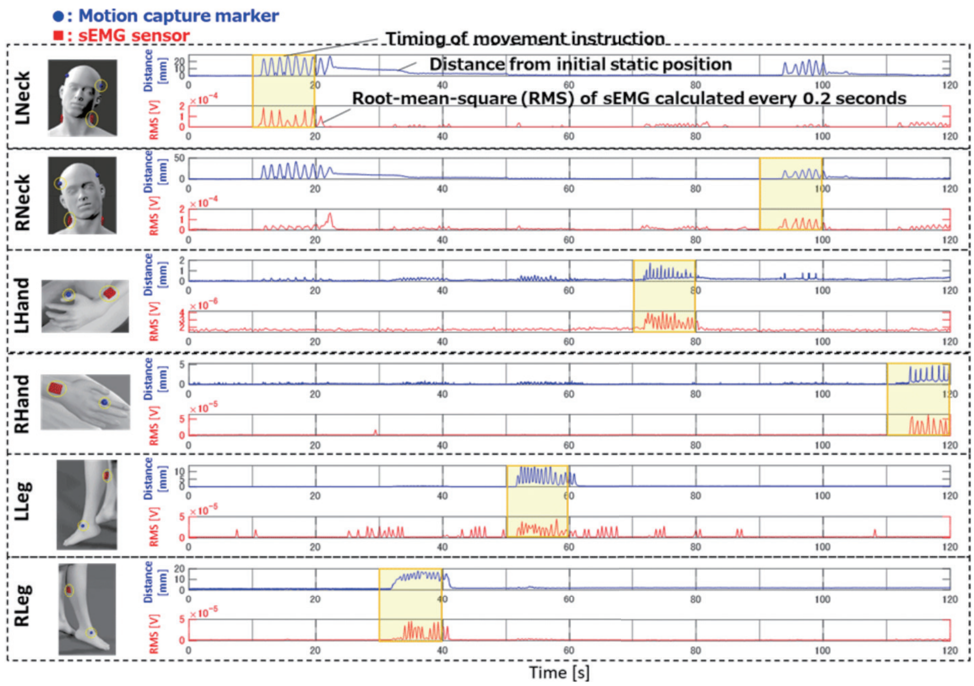


Fig. 1. Time-series changes in the marker position at each body part as the distance from the initial static position (blue line) and the root mean square (RMS) calculated with the sEMG sensor (red line). The yellow highlights indicate the body parts instructed to move.

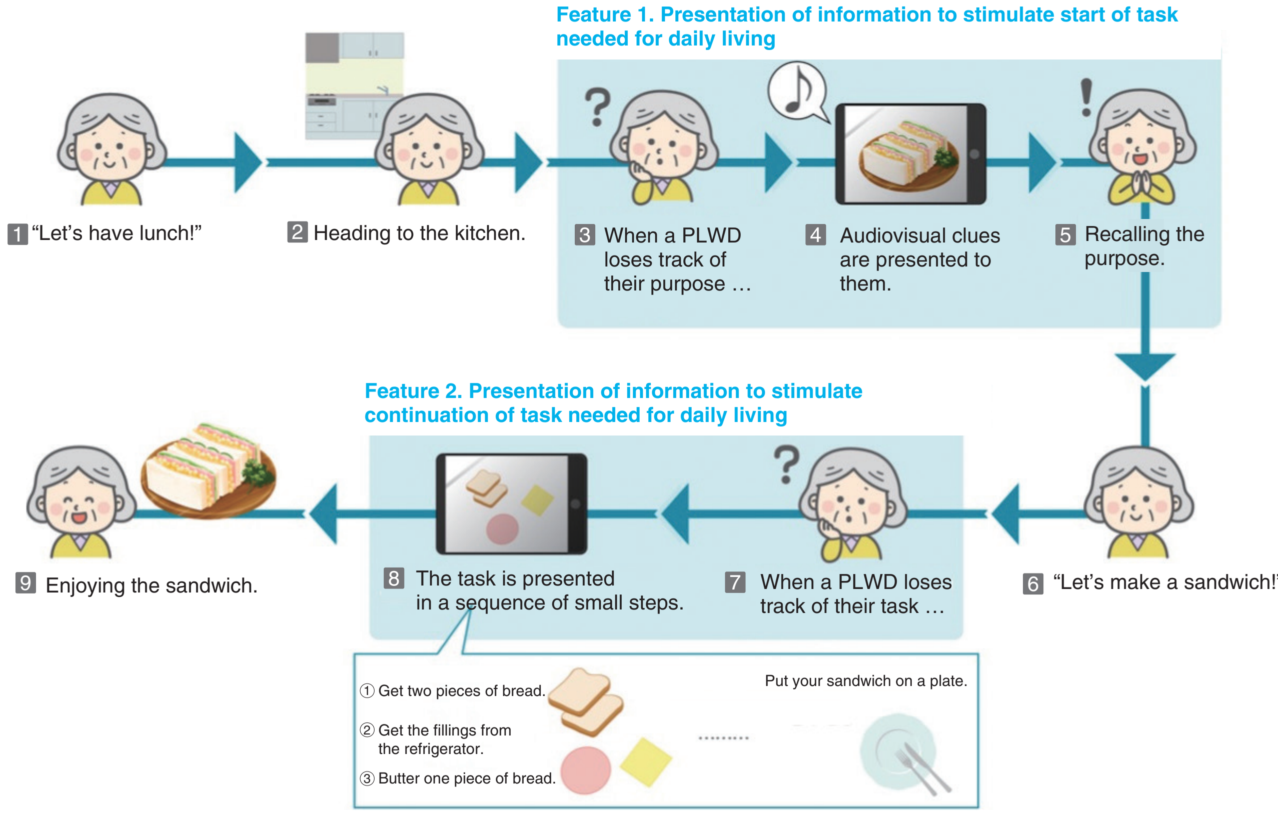
reacting during the interval (shown in yellow) in which the artist with ALS moved the indicated body part. Therefore, to obtain the sEMG of the muscles of the body part the person intends to move, it is necessary to calibrate the sEMG during the state of muscle enervation, such as at rest, as the reference values, and set a threshold for each body part to determine muscle contraction. The resting state of muscles immediately after the artist performed was set as the reference values, and the thresholds were set in accordance with his exertions during a performance.

Muscle fatigue is a point that should be considered when converting continuous sEMG into commands to operate an avatar in the metaverse. Muscle strength and endurance in individuals with severe physical disabilities deteriorate. It is thus necessary to enable intentional operating commands for the metaverse while conserving energy in the moving body parts. We thus adopted the policy of avoiding avatar commands that come from long, continuous muscle contractions or the degree of strength of muscle contractions. An avatar command is issued on the basis of the determination of muscle contraction of each body part. An avatar movement in response to an operation command is maintained for a certain period. If the same bodily operation command is repeated during this time, the duration of the reflected avatar movement is extended. Such a policy may seem ineffective because it does not reflect the reality of body movements. However, the artist with ALS who experienced this technology felt the avatar was moving as he had intended. He felt the actual sense of operating the avatar through his own body movements.

2. Development of information-prompt technology that supports the independence of people with dementia in their daily living

The decline in cognitive functions, such as memory and language, makes it challenging for individuals with dementia[*1] to perform tasks that were once routine, necessitating assistance in daily living. Even

*1 Dementia: "A syndrome due to disease of the brain, usually of a chronic or progressive nature, in which there is disturbance of multiple higher cortical functions, including memory, thinking, orientation, comprehension, calculation, learning capacity, language, and judgement. Consciousness is not clouded. The impairments of cognitive function are commonly accompanied, and occasionally preceded, by deterioration in emotional control, social behavior, or motivation." (From the International Classification of Diseases Version 10)



PLWD: person living with dementia

Fig. 2. Features of information presentation technology to promote independence in people with dementia.

though the number of elderly people with dementia is increasing, there is a shortage of care professionals, and the difficulty of providing daily living support that meets the needs of each person has become a social issue. In terms of future societal trends, as a result of the enactment of the Basic Act on Dementia*2 in June 2023 and the expected revision of the long-term care insurance system*3 in FY2024, it is expected that the deployment of technology to address dementia will accelerate. While there are many technologies designed to watch over people with dementia, there are few technologies designed to assist them in their daily living. Developing such technologies in the field of dementia is thus highly important.

At NTT Human Informatics Laboratories, in collaboration with the University of Western Sydney and Deakin University in Australia, we have been conducting research and development (R&D) of information-prompt technology that encourages individuals with dementia to perform activities of daily living (ADLs) independently. Specifically, we presented notifications and procedures related to tasks of ADLs to users with dementia through tablet devices and investigated how to effectively prompt the tasks using information such as images and sounds (**Fig. 2**).

Our results indicate the effectiveness of auditory prompts and the importance of dividing tasks appropriately on the basis of their complexity and familiarity [1]. The results also revealed user opinions highlighting the need to customize task division according

*2 Basic Act on Dementia: Abbreviated name of "The Basic Act on Dementia to Promote an Inclusive Society." "A law that establishes the basic principles related to dementia-related measures. It clarifies the responsibilities of the national government, local governments, etc., and sets forth the formulation of plans in relation to dementia measures, as well as the items that form the foundation of dementia measures." (translation by the author) (From "Basic Act on Dementia to Promote an Inclusive Society," Ministry of Health, Labour and Welfare of Japan (MHLW))

*3 Long-term care insurance system: "A system in Japan in which members of society as a whole support one another in caring for the elderly." (translation by the author) (From "Overview of the Long-term Care Insurance System," MHLW) to each individual user's capabilities and preferences, and suggested that conventional support systems that provide uniform assistance to users are insufficient. We have also been studying the design of a notification system to facilitate attention management for completing tasks independently and studying an interface design for users to customize the system in accordance with their own lifestyle. Our results will be presented in future papers.

On the basis of the knowledge gained from the above collaborative research and other efforts, we have begun an investigation to expand this knowledge into an approach that takes into account "purpose in life" and "inclusive society," which lie beyond "independence," as our effort going forward. We are revising the concept of design with the aim of developing technologies that support communication for maintaining social connections with others in one's community and promoting psychological transformation to positively embrace dementia. Our efforts are currently at the stage of organizing issues, taking into account the perspectives of people with dementia by conducting surveys through questionnaires and interviews [2]. Collaboration with local high school students in the framework of an educational program is also underway to brainstorm ideas for addressing dementia-related challenges. By engaging in R&D while collaborating with individuals with dementia and high school students, we seek to create technology that is rooted in humanity living with dementia and in humanity shaping the future dementia society. In the future, we plan to apply this technology to care robots*4 and information and communication technology that can comprehensively support living and life experiences of individuals with dementia from cognition to emotion.

3. Toward neurodiversity

Neurodiversity means perceiving differences between individuals in their brains and nervous system and various characteristics that arise from them as diversity, mutually respecting the differences, and making use of the diversity in society. For society, including communities such as workplaces, it means mutual respect and understanding and providing an environment in which it is comfortable for anyone to live. We expect that in communities where mutual respect and understanding is promoted, there is greater psychological safety and more active flow of information between members. As a result of such changes in the environment, greater creativity will be possible in companies, with diversity and psychological safety as the engine.

At NTT Human Informatics Laboratories, in collaboration with NTT Claruty Corporation, a special-purpose subsidiary of NTT, we discussed and agreed that elimination of miscommunication is most important in increasing psychological safety and allowing demonstration of creativity in a more diverse workplace. We believe that the largest cause of miscommunication lies in ambiguous expressions used in conversations during work. The workplace environment would greatly improve if the risk level of ambiguous expressions could be determined and pointed out during meetings. We are thus developing a prototype tool that notifies users ambiguous expressions. We have made improvements to the tool by using findings from its trial use. For example, a user requested that the tool should work well with other tools (web conferencing tools and various support tools). We thus learned the features needed for such a tool to satisfy a highly diverse workplace.

In developing this prototype, the following design concepts were implemented on the basis of discussions with NTT Claruty.

- Any person in any position should be able to look at the same screen and mutually confirm the notification of an ambiguous expression.
- During its use, a feedback mechanism should be provided for the notification of the ambiguous expression. The criteria for notification should be easily modifiable.
- In determining the risk level of an ambiguous expression, more consideration should be given to members whose communication errors could easily lead to serious issues such as leaving their jobs, for example, members on the autistic spectrum (members with autism spectrum disorder).
- Interference with current support software should be minimized.
- The tool should have a highly flexible configuration for conversation interventions that meet the characteristics of each individual.

The screen of this prototype for ambiguous expression detection is shown in **Fig. 3**. All speech is transcribed using speech-recognition software. When an ambiguous expression is discovered, users are notified through a pop-up window in which the word is highlighted by changing its color, font, etc., in line

*4 Care robot: "Care equipment in which robot technology is applied to support patients in being independent and reduce the burden on care partners." (translation by the author) (From "Development and Dissemination of Care Robots," MHLW)

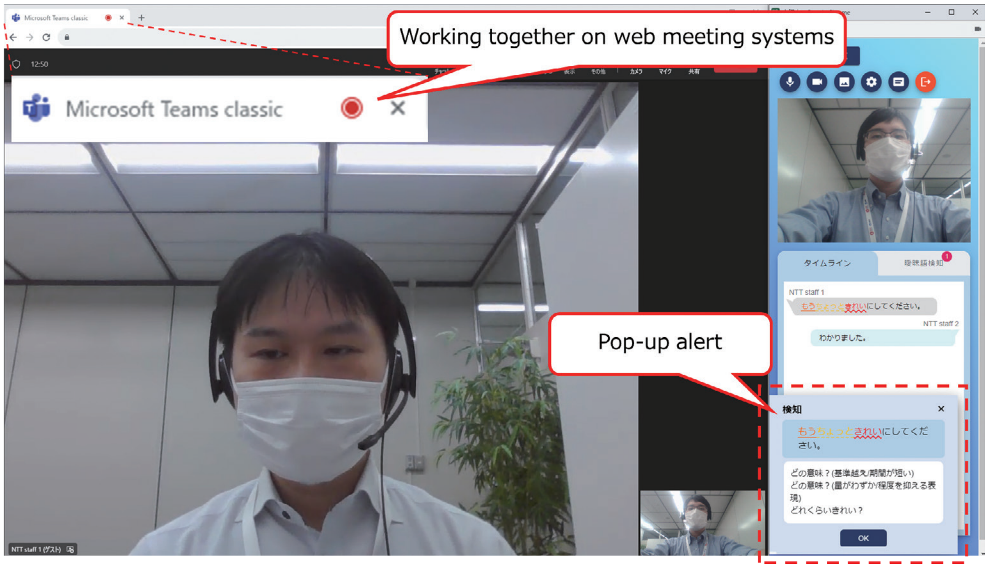


Fig. 3. Prototype of a communication support tool based on ambiguity detection.

with the level of assessed risk of miscommunication. (The risk level is a numeric value that expresses the predicted severity if miscommunication due to the ambiguous expression occurs. Its initial value is determined on the basis of the opinions of specialists in the field of employment support for people with disabilities and adjusted on the basis of feedback by trial users of this tool.) The tool is used to promote more detailed discussion of ambiguous expressions. According to a survey after the trial use, users reported that this tool increased their sense of self-efficacy. They gave comments such as "This tool made it easier to form clear agreements" and "It made it easier to understand instructions" (**Fig. 4**).

Going forward, we plan to use an improved version of the tool and confirm its performance in terms of user experience and effectiveness. We will study adding functions that enable understanding of the internal state of the user by continuing discussions with stakeholders. Such functions will enable understanding of the user's level of understanding of conversations, their familiarity with work, and their degree of concentration.

4. Movement support for people with spinal-cord injuries

The number of people with quadriplegia due to spinal-cord injury and who have difficulties with daily activities is increasing annually. In Japan, it is reported that about 6000 people per year acquire such injuries [3]. Movement support for people with spinal cord injuries has thus become an important issue. As an effort worldwide to address this issue, invasive brain-computer interfaces (BCI) are being developed mainly by research medical institutes to restore the patient's ability to move. To restore movements with invasive BCI, electrodes are surgically implanted in the brain and muscles (or nerves), and electrical stimulation is applied to the muscles on the basis of the results of analysis of brain activity information.

At NTT Human Informatics Laboratories, while respecting individuals' diverse lifestyles, we believe that movements on one's initiative using one's body is one of the components of well-being in daily living and in society. We are engaged in the development of invasive BCI technology to restore arm movements, which are essential for daily activities such as eating. Specifically, we are studying technology that outputs electrical muscle stimulation from brain activity to restore muscle-coordination movements. Our goal is to restore not only relatively simple movements such as wrist flexion, which has been achieved in previous research, but also daily-life movements that require more complex muscle coordination such as drinking water from a cup.

To implement this technology, we are developing artificial intelligence (AI) technology that extracts from brain activity what kind of muscle activity is intended on the basis of the pattern of muscle synergy,

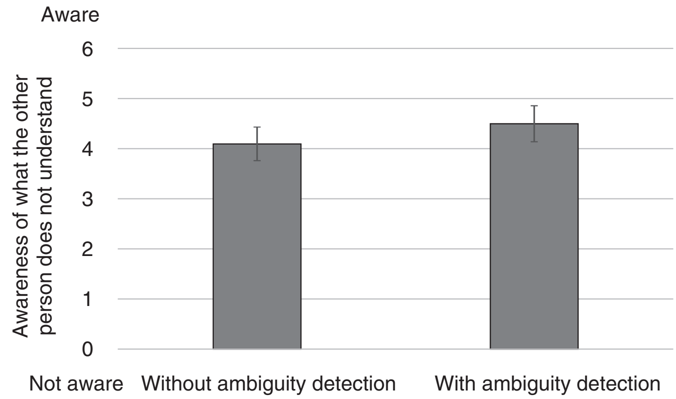


Fig. 4. Increased frequency of awareness.

which is a mechanism of muscle coordination that the spinal cord possesses. Muscle synergy refers to the pattern of coordination of multiple muscle activities that appears during a body movement. It is considered to be the result of the brain's control of nerves in the spinal cord that are supplied to multiple muscles. On the basis of this muscle synergy pattern, we constructed a model that converts brain activity to muscle activity on the basis of biological data obtained from people with healthy arms. By incorporating the estimated muscle synergy in the BCI system for individuals with spinal-cord injury, we expect it can compensate for the muscle coordination impaired due to the injury and restore movements that require complex muscle coordination.

Conventionally, the estimation of muscle synergy has been conducted solely on the basis of muscle-activity data, resulting in estimations of how muscles have coordinated. We believe that muscle synergy that reflects nerve connections in brain-spinal cord-muscles are easier to control from the brain. We have thus developed a muscle-synergy-estimation method that takes into account brain-activity data acquired at the same time as muscle-activity data. Specifically, our method uses a deep learning model that features layers to simulate muscle synergy and estimate muscle synergy in the process of training muscle activity output from brain activity input. Comparing the conventional method with our method by using electrocorticography of the monkey motor cortex, the region of the brain involved in muscle activity, and muscle-activity data, we confirmed that our method can estimate muscle synergy with greater accuracy.

Going forward, we will investigate the method of electrical stimulation of multiple muscles to achieve actual muscle activity from the estimated muscle activity. We also plan to develop and test a BCI system in combination with the muscle-synergy model and the method of electrical muscle stimulation to confirm the effectiveness of the system in restoring muscle-coordination movements.

5. Depression-symptom-detection technology

Support for mental health has become a major social topic. Depression (major depressive disorder), while familiar, is known to severely impair daily living. It has been reported that about 6 out of every 100 Japanese have experienced depression, and survey results indicate that Japanese society loses about 2 trillion yen annually due to depression in its members. The number of mental disorder cases has been on the rise as telework spreads and more people live alone. Reports state that the number of patients with depression more than doubled between 2013 and 2020. Because the more severe depression becomes, the harder it is to treat; thus, early detection and treatment are crucial. However, people around a person with depression and the afflicted person often do not recognize depression symptoms, and the case worsens [4, 5].

NTT Human Informatics Laboratories is developing AI technology that can detect symptoms of depression in daily life and promote early treatment. As a part of this effort, we have been working to create technology that can easily detect depressed mood, a major symptom of depression, by applying media-processing technologies cultivated at NTT laboratories over a long time. The technology detects depression from the responses of the user's voice and their

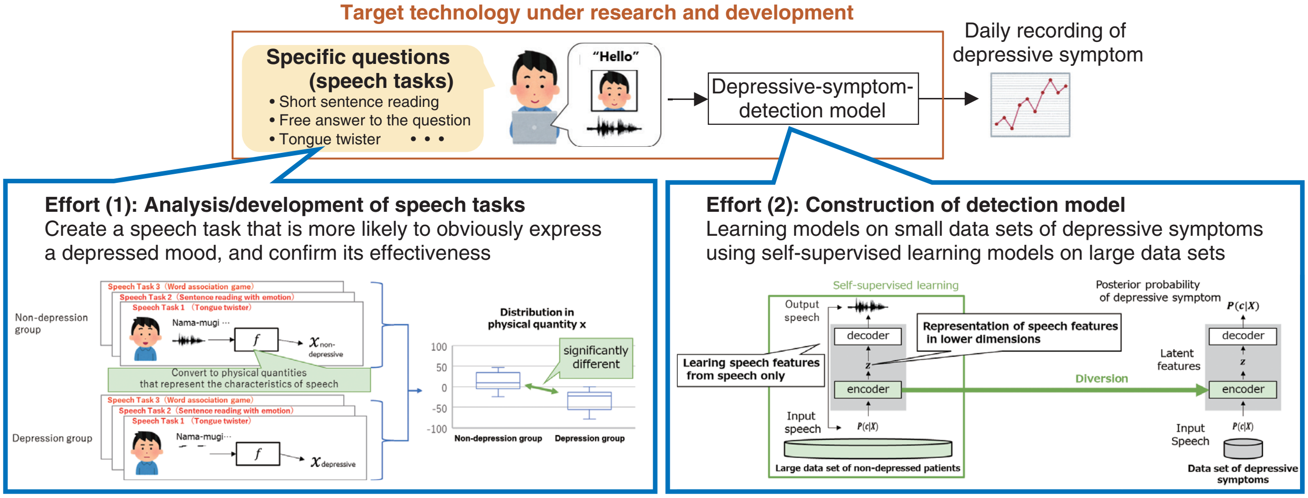


Fig. 5. Efforts to create depressed-mood-detection technology.

expressions in response to specific questions (**Fig. 5**).

We collected and analyzed voice and video data of people with depressed mood. To date, we have collected voice data from more than 100 people, including those diagnosed by physicians as having a depressed mood. Comparison and analysis of the voice features of people with depressed mood and those without revealed the expressions of features such as the following in those with depressed mood: (1) lack of emotional expression (little difference between normal voice and voice when expressing happiness or anger), (2) lowered cognitive function (delay in answering questions, fewer words spoken), and (3) decline in articulatory function (difficulty speaking quickly) [6]. On the basis of these findings, we created multiple questions (speech tasks) suitable for detecting depressed mood and confirmed their effectiveness (Effort (1)). We are also constructing an AI model that can detect depressed mood (Effort (2)). This model is being developed on the basis of self-supervised learning*5 to detect depressed mood with high accuracy from a small dataset of voice data belonging to people with depression symptoms. The AI model is trained in advance on the features of general speaking voices by using a large dataset of the voices of people without depression symptoms. It facilitates learning by the AI model of voice features of people with depressive symptoms.

We plan to develop services that prevent mental disorders and contribute to early recovery by using the above depression-symptom-detection technology to stimulate early mental care while monitoring daily living conditions. For the mental health management of employees in a company, for example, we plan to develop an AI agent service that can detect employees having depression symptoms from their daily communication and connect them to early treatment. Therefore, this technology can contribute to promoting health management in companies by implementing early prevention of mental disorders, reducing the need to take leave or retire due to depression.

References

[1] M. J. Radnan, R. Brookman, S. Parker, L. Hoon, H. Kawata, A. Ono, and C. Harris, "User Interface Preferences of Assistive Technology Device for People Living with Dementia at Home," MARCS Conference Series: Ageing, Westmead, NSW, Australia, Aug. 2022.

[2] A. Ono, A. Yamanaka, S. Seko, and H. Matsukawa, "Communication Problems for People Living with Dementia," Society for Cocreationology 7th Annual Conference, pp. 66–69, 2023 (in Japanese).

[3] N. Miyakoshi, K. Suda, D. Kudo, H. Sakai, Y. Nakagawa, Y. Mikami, S. Suzuki, T. Tokioka, A. Tokuhiro, H. Takei, S. Katoh, and Y. Shimada, "A Nationwide Survey on the Incidence and Characteristics of Traumatic Spinal Cord Injury in Japan in 2018," Spinal Cord, Vol. 59, pp. 626–634, 2021. https://doi.org/10.1038/s41393-020-00533-0

[4] OECD, "Tackling the Mental Health Impact of the COVID-19 Crisis: An Integrated, Whole-of-society Response," May 2021. https://oecd.org/coronavirus/policy-responses/tackling-the-mental-health-impact-of-the-covid-19-crisis-an-integrated-whole-of-society-response-0ccafa0b/

[5] M. Sado, "What Are Society's Losses As a Result of Depression?—Cost-of-illness Study of Depression," Psychiatria et Neurologia

*5 Self-supervised learning: A machine learning method that uses a large amount of unlabeled data to acquire in advance features from the data for the model. After a model is generated through self-supervised learning, it is possible to build an accurate inference model by using a re-training method such as fine-tuning, using a small amount of labelled data.

Japonica, Vol. 116, No. 2, pp. 107–115, 2014 (in Japanese).
[6] Y. Murata and A. Ando, "Speech Analysis of Depression Patients for Selection of Speech Tasks to Detect Depression Symptoms," Acoustical Society of Japan 2023 Spring Research Presentation Conference, 3-4P-4, 2023 (in Japanese).

Mariko Nakamura
Senior Research Engineer, NTT Human Informatics Laboratories.
After joining NTT, she researched environmental impact assessment of telecommunication services. She then become interested in human-centered design in communication while studying risk-communication regarding chemical substances, disasters, and food. She is currently promoting projects in collaboration with different companies in various industries trying to solve communication issues with new technologies from NTT R&D.

Ryosuke Aoki
Senior Research Engineer, NTT Human Informatics Laboratories.
He received an M.S. and Ph.D. in information sciences from Tohoku University, Miyagi, in 2007 and 2014 and joined NTT in 2007. His current research interests include human-computer interaction, interaction design, research through design, medical engineering, human augmentation, and wearable/ubiquitous computing.

Asuka Ono
Researcher, NTT Human Informatics Laboratories.
She received a B.S. and M.S. in design from Kyushu University, Fukuoka, in 2016 and 2018. After joining NTT in 2018, she focused on hearing assistive technology with tactile speech enhancement from 2019 to 2020. From 2021 to 2022, she worked on prompting technology to support ADL for individuals with dementia. She is currently investigating communication challenges faced by people with dementia. She envisions contributing to the empowerment of individuals with diverse abilities through the development of inclusive technological solutions rooted in cognitive science. In addition to her research, she actively engages in science programs for high school students.

Iwaki Toshima
Senior Research Engineer, NTT Digital Twin Computing Research Center.
He received an M.E. from the University of Tokyo in 2002 and joined NTT the same year. He received a Ph.D. in engineering from Tokyo Institute of Technology in 2008. His research interests include human-computer interaction, robot audition, and digital twin computing.

Kengo Okitsu
Senior Research Engineer, Cybernetics Laboratory, NTT Human Informatics Laboratories.
He received a B.E. and M.E. from Tokyo Institute of Technology in 2008 and 2010. His research interests include brain machine interfaces.

Atsushi Ando
Senior Research Engineer, Human Activity Understanding Research Group, Human Insight Laboratory, NTT Human Informatics Laboratories.
He received a B.E., M.E., and D.E. from Nagoya University, Aichi, in 2011, 2013, and 2021. Since joining NTT in 2013, he has been engaged in research on speech recognition and non-/paralinguistic information processing. He received the Awaya Prize Young Researcher Award from the Acoustical Society of Japan (ASJ) in 2019. He is a member of ASJ, the Institute of Electronics, Information and Communication Engineers (IEICE), the Information Processing Society of Japan (IPSJ), and the International Speech Communication Association (ISCA).

Takeshi Mori
Senior Research Engineer, Human Activity Understanding Research Group, Human Insight Laboratory, NTT Human Informatics Laboratories.
He received a B.E. and M.E. from Tokyo Institute of Technology in 1994 and 1996 and D.E. from University of Tsukuba in 2007. Since joining NTT in 1996, he has been engaged in research on speech and audio processing algorithms. He is a member of ASJ and Institute of Electrical and Electronics Engineers (IEEE).

High-definition AI Inference Technology for Detecting a Wide Range of Objects with One Camera at One Time

Hiroyuki Uzawa, Shuhei Yoshida, Yuukou Iinuma, Saki Hatta, Daisuke Kobayashi, Yuya Omori, Yusuke Horishita, Tomoki Onoe, Ken Nakamura, and Koji Yamazaki

Abstract

Object detection in high-definition video is required in video artificial intelligence (AI) applications for edge/terminals to detect a wide range of objects with one camera at one time. Although various AI inference schemes for object detection (e.g., You Only Look Once (YOLO)) have been proposed, they have a limitation regarding the input image size, thus need to shrink the input high-definition image into that limited size. This collapses small objects, making them undetectable. This article introduces high-definition AI inference technology we previously proposed for solving this problem, with which multiple object detectors cooperate to detect small and large objects in high-definition video.

Keywords: AI, object detection, high-definition video

1. Introduction

Object detection for identifying and locating objects in images plays an important role in video artificial intelligence (AI) applications for edge/terminals. One application is automated beyond-visual-line-of-sight (BVLOS) drone flight. BVLOS flight means flying beyond the operator's visual range and enables a drone to cover far greater distances. For BVLOS flight, a drone has to detect objects in images input from a mounted camera in real time for safe flight, especially to avoid flying directly over passersby or cars under the flight path. Another application is camera surveillance, and the object detection for determining suspicious people from a crowd of people has to be done in a terminal to comply with personal-information-protection requirements, such as the General Data Protection Regulation (GDPR). Another application is road-traffic monitoring.

Object detection in high-definition video enables detecting a wide range of objects with a single camera at one time. This makes it possible to detect passersby and cars under the flight path from higher altitude for automated BVLOS drone flight and detect suspicious people from a greater distance for camera surveillance. In a high-definition image such as full HD and 4K, since objects near and far from the mounted camera can coexist in the same image due to a wide angle of view, both large and small objects can be included in the image. Therefore, object detection in high-definition video has to be able to detect not only large but also small objects with high accuracy.

Various AI inference schemes for object detection have been proposed, such as You Only Look Once

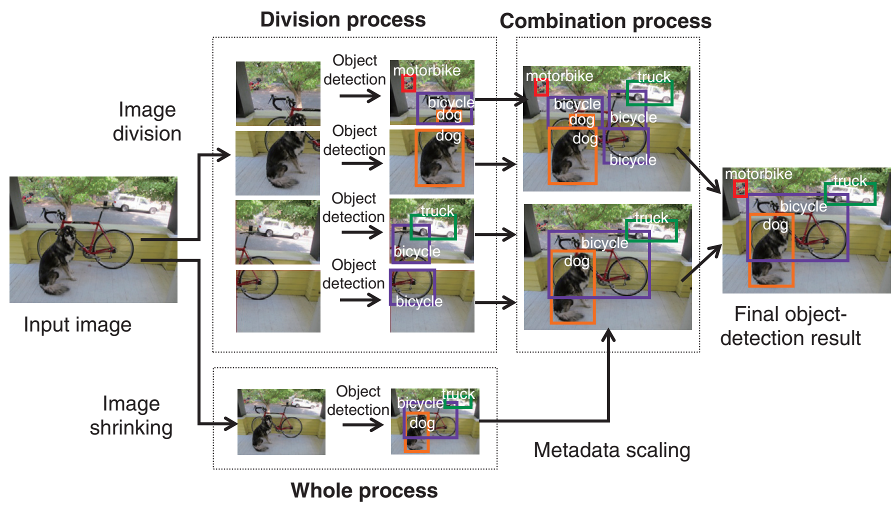


Fig. 1. Our high-definition AI inference technology for object detection.

(YOLO) [1] and Single Shot MultiBox Detector (SSD) [2]. Such an inference scheme has a convolutional neural network (CNN) that predicts object bounding boxes and class probabilities in a single evaluation. The input-layer size of a trainable CNN is limited to suppress the computational complexity and training-process difficulty. For example, the largest input-layer size in the standard YOLO model YOLOv3 is 608 × 608 pixels. Thus, even if a full HD image is input, the input image is shrunk to the limited input-layer size and inference is executed with the shrunk input image. Large objects in the input image can be detected, but small objects are collapsed and difficult to detect. Dividing the input image into a limited size can also be considered to prevent shrinking the input image [3–8], but this means large objects that straddle the divided images cannot be detected because the characteristic parts for identifying objects are also divided. In other words, such schemes are unsuitable for object detection in high-definition video.

We introduce high-definition AI inference technology we previously proposed [9, 10] for solving this problem, with which multiple object detectors cooperate to detect small and large objects in high-definition images. This technology is suitable for hardware implementation because all object detectors can be executed in parallel for real-time operation. In addition, any AI inference scheme for object detection can be applied, and re-training for applying our technology is not necessary.

The remainder of the article is organized as follows. In Section 2, we give details of our technology and present the results of an evaluation of the technology in Section 3. We conclude this article in Section 4.

2. Our high-definition AI inference technology

2.1 Overview of our technology

Our high-definition AI inference technology [9, 10] enables both small and large objects to be detected in high-definition images through cooperation among multiple object detectors (**Fig. 1**). An input image is divided into images of limited size, and objects are detected on the basis of an inference scheme for every divided image. In parallel, object detection is also done in the whole image shrunk to a limited size to detect large objects that straddle the divided images. By combining these object-detection results, a final object-detection result is obtained. In other words, the undetected objects in the shrunk whole image are interpolated from the object-detection results obtained from the divided images. Therefore, both small and large objects can be detected in high-definition images.

We explain the mechanism by which high-definition object detection can be achieved with our technology by using YOLO as an example scheme. YOLO has a CNN for detecting objects included in an input image. The CNN divides the input image into $S \times S$ grids and predicts bounding boxes and class probabilities for each grid. The resolution for detecting objects depends on only the grid size. Our technology reduces the grid size. **Figure 2** shows an

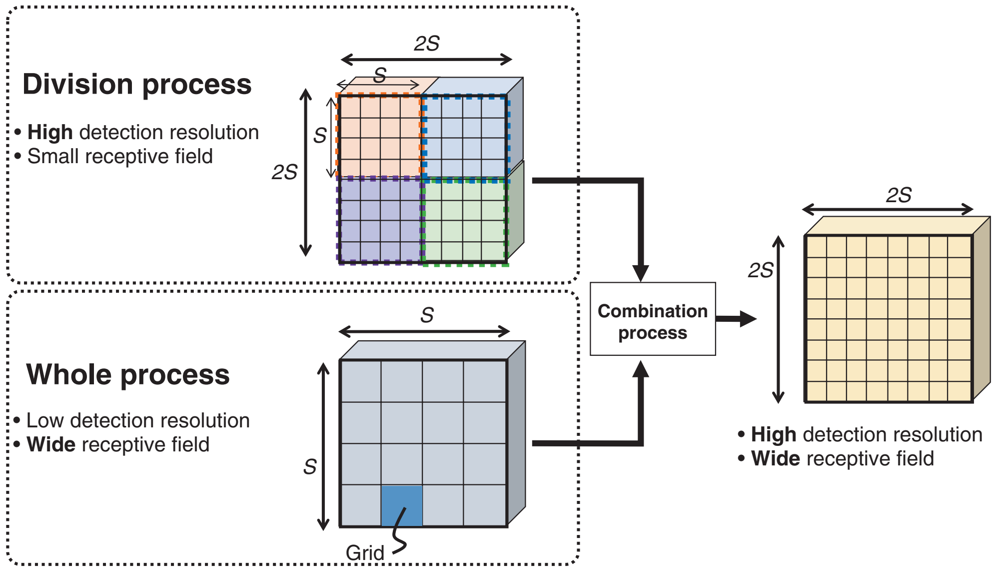


Fig. 2. High-definition object-detection mechanism.

example in which the input image is divided into four images. In the division process, since each divided image has $S \times S$ grids, the four total divided images can be regarded as $2S \times 2S$ grids. This means that the grid size is reduced to half. However, the division process causes the receptive field to narrow because it does not execute the convolutional operations for the whole image, thus cannot detect large objects. In contrast, in the whole process, although the detection resolution is low, a wide receptive field can be obtained. Therefore, with our technology, the division process with high detection resolution and small receptive field and the whole process with low detection resolution and wide receptive field are combined, which enables object detection with high detection resolution and wide receptive field. Concretely, when an image with 1920 × 1080 pixels is input, the minimum gird size without our technology (YOLOv3 only) is 60 × 34 pixels because S is 32. In contrast, with our technology, the minimum grid size becomes 30 × 17 pixels when there are four divisions, and higher detection resolution can be provided.

Our technology is suitable for hardware implementation because all object-detection processes can be executed in parallel. The same weight coefficients can also be used among all object-detection processes. In addition, any AI inference scheme for object detection can be applied, and re-training for applying our technology is not necessary. Moreover, the computational complexity with our technology is proportional to the number of divisions. This means that our technology can reduce the grid size with less complexity than increasing the number of grids in the CNN because the complexity of the CNN increases with the square of the number of grids.

2.2 Details of combination process

In the combination process, the undetected objects in the shrunk whole image are selected from the object-detection results obtained with the divided images. The detected objects in the shrunk whole image and the selected objects are output as the final object-detection results.

This selection requires determining whether an object detected in the divided image is an undetected one in the shrunk whole image. The combination process calculates two indicators: a multiplicity between the detected object in the divided image and that in the shrunk whole image and the ratio between the area size of the detected object in the divided image and that in the shrunk whole image.

When both indicators are high, the process determines that the detected object in the divided image is the same as that in the shrunk whole image and excludes it from the objects to be selected. In contrast, when either of these indicators is low, the process determines that the detected object in the divided image is different from that in the shrunk whole image. This is executed for all detected objects in the divided images in combination with all detected

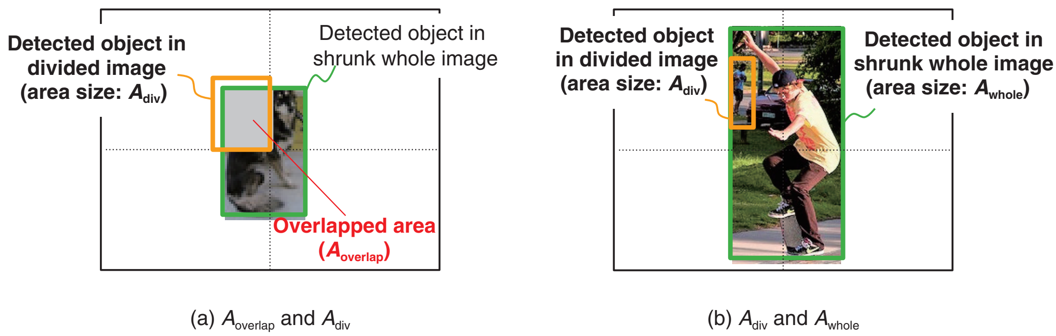


Fig. 3. Examples of A_{div}, A_{whole}, and $A_{overlap}$.

objects in the shrunk whole image, and the non-excluded detected objects in the divided images are selected as the undetected objects in the shrunk whole image. Moreover, the object with the class detected only in the divided images is selected when its object size is sufficiently small that it cannot be detected in the whole shrunk image, e.g., its object size is less than 1/100 of the input image size.

The above-mentioned multiplicity $I_{multiplicity}$ is a dedicated indicator for enabling comparison between the divided and whole objects and is expressed as

$$I_{multiplicity} = A_{overlap} / A_{div}, \quad (1)$$

where $A_{overlap}$ is the overlapped area size between the divided object detected in the divided image and the whole object detected in the shrunk whole image, and A_{div} is the area size of the divided object, as shown in **Fig. 3(a)**. Although intersection over union (IoU) is often used as multiplicity, it is unsuitable to use to compare a divided object and a whole object because it assumes comparison between whole objects. Therefore, we newly define $I_{multiplicity}$. When $I_{multiplicity}$ is larger than or equal to a threshold α, the multiplicity is determined to be high. Under the high multiplicity condition, the divided object is likely to be the same as the whole object.

Even if the multiplicity is high, the detected object in the divided image may be different from that in the shrunk whole image when the area-size ratio between those objects is high, as shown in **Fig. 3(b)**. The above-mentioned size ratio I_{ratio} is used for determining this condition. The I_{ratio} is given by

$$I_{ratio} = A_{div} / A_{whole}, \quad (2)$$

where A_{whole} is the area size of the detected object in the shrunk whole image. Under the high multiplicity condition, when I_{ratio} is larger than or equal to a threshold β, the detected object in the divided image is determined to be same as that in the shrunk whole image.

With these indicators in the same class, the undetected objects in the shrunk whole image are selected from the detected objects in the divided images, and both large and small objects can be detected while avoiding duplicate detection of the same object.

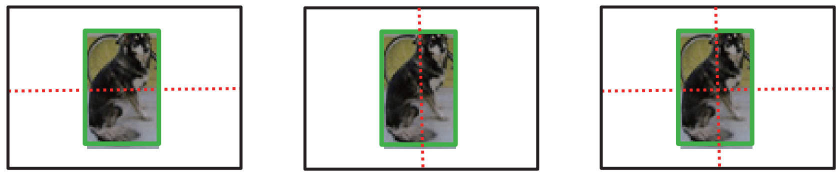
3. Object-detection performance

We applied our technology to the standard YOLOv3 model with 608 × 608 pixels and evaluated object-detection performance. The same weight coefficients are used between all object detectors in our technology, as described in Section 2.1. We used the weight coefficients published in [11]. These coefficients are pre-trained with the Microsoft Common Objects in Context (MS COCO) dataset [12], which is a widely used object-detection dataset with 80 classes.

3.1 Optimization of pre-set parameters

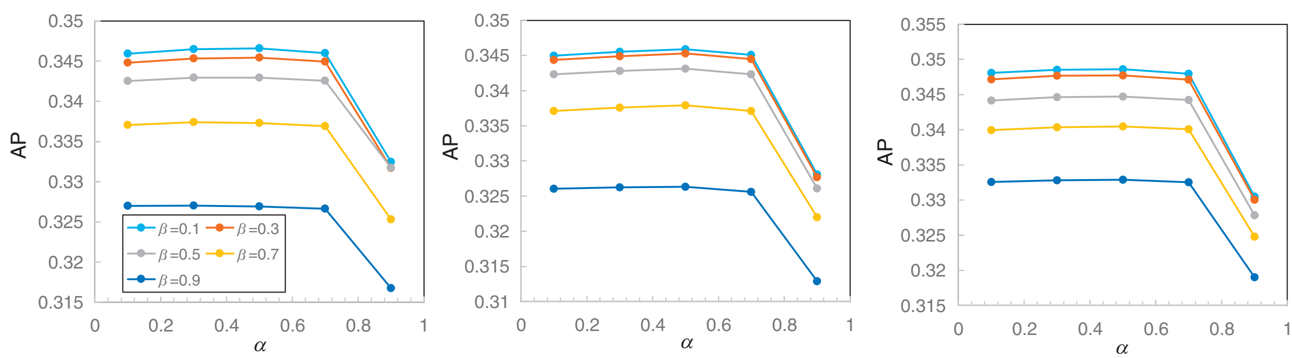
With our technology, α and β are pre-set for the combination process. An object detected in the divided image is more likely to be determined as undetected in the shrunk whole image when these thresholds are higher and more likely to be determined as the same when these thresholds are lower. Thus, α and β should be optimized to execute the combination process properly.

Object-division patterns can be mainly classified into horizontal division (**Fig. 4(a)**), vertical division (**Fig. 4(b)**), or cross-shaped division (**Fig. 4(c)**). When there are two divisions, only the vertical or horizontal division pattern can occur. In contrast, when there are more than two divisions, all division patterns can occur, and the division pattern for the



(a) Horizontal division (b) Vertical division (c) Cross-shaped division

Fig. 4. Division patterns.



(a) Horizontal division (b) Vertical division (c) Cross-shaped division

Fig. 5. Relationship between AP, α, and β.

object varies in accordance with the position of that object in the image. For this reason, commonly available α and β need to be obtained for various division patterns.

We searched for the optimized α and β where the average precision (AP) is maximized. The AP is obtained from precision and recall. In this search, α and β were varied from 0.1 to 0.9 in increments of 0.2. We used the images in the MS COCO dataset val2017 because the objects included in these images are large and the division patterns shown in Fig. 4 can be reliably generated.

Figure 5 shows the measured relationship between AP, α, and β. The AP decreased as β increased from 0.1 in all division patterns. The AP also reached its maximum when α was 0.5 in all division patterns. The AP gradually decreased as α increased from 0.5. This is because an object detected in the divided image is more likely to be determined as undetected in the shrunk whole image as α becomes higher. In other words, the optimized α and β are 0.5 and 0.1, respectively, for all division patterns. Therefore, the combination process with our technology can be properly executed by pre-setting α to 0.5 and β to 0.1.

3.2 Object-detection performance

We conducted evaluations to determine the effectiveness of our technology. On the basis of the optimized result described in the previous section, α and β were set to 0.5 and 0.1, respectively. There are two divisions.

We first conducted a basic evaluation using 5000 images in the MS COCO dataset val2017 and measured the AP for each class. Although the images included in the MS COCO dataset are standard definition (SD) images, such as 600 × 400 pixels, large objects account for a higher percentage of the objects in the images; thus, we can determine the image-division penalty with our technology by comparing the APs with and without our technology. The measured AP is that averaged by 10 IoU thresholds in 0.05 increments from 0.5 to 0.95.

Figure 6 shows the measurement results. The AP improved in almost all classes and by a maximum of 1.2 times. This means that our technology can suppress the image-division penalty for large objects and

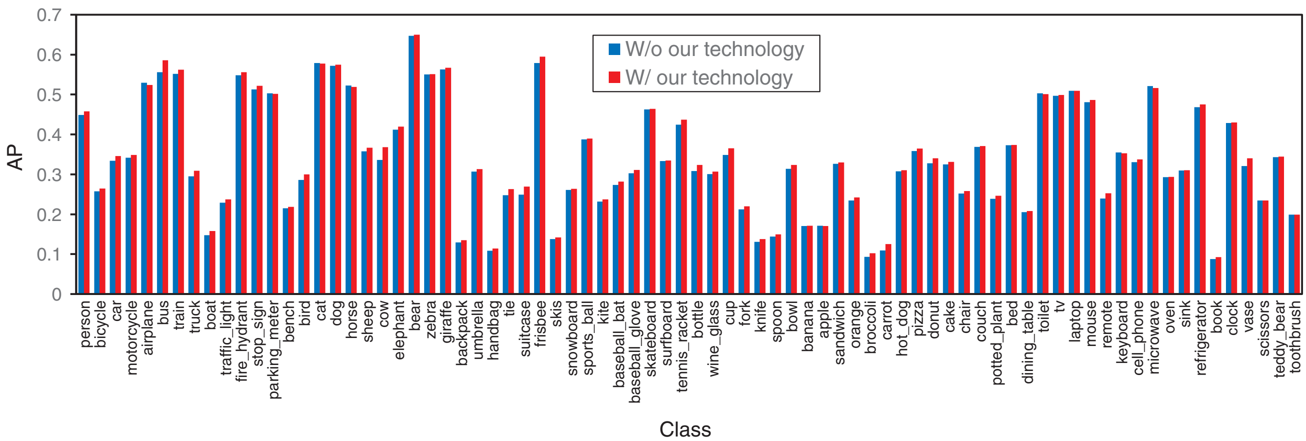


Fig. 6. AP of each class (image size: about 600 × 400 pixels).

Table 1. Summary of the detection results (image size: 1920 × 1080 pixels).

Class	W/o our technology (standard YOLOv3 model only)		W/ our technology	
	Precision [%]	Total number of objects detected in 150 images	Precision [%]	Total number of objects detected in 150 images
Person	62.5	176	62.8	298
Car	83.9	1426	81.5	2049
Bicycle	50.0	6	54.5	11
Truck	17.1	111	19.3	176
Bus	33.3	6	15.0	20
Motorbike	61.5	13	59.5	37

improve object-detection performance even in SD images.

We then evaluated object-detection performance with full HD images. We selected 150 full HD (1920 × 1080) images from the VisDrone2019 dataset [13] and evaluated precision and the number of detected objects. The VisDrone dataset is a large-scale drone-captured dataset with ten classes and includes full HD images. For the weight coefficients, we used the same coefficients pre-trained using the MS COCO dataset as above. To calculate precision, we remapped people and pedestrian labels to person label and van label to car label in the VisDrone dataset, and only the common class labels between MS COCO and VisDrone were evaluated.

Table 1 summarizes the detection results, and **Figure 7** shows the example images obtained in this evaluation. From Table 1, our technology enabled the number of detected objects to be increased while maintaining precision. For example, the number of detected objects in the person class was 1.7 times higher with our technology than without it. Across all evaluation classes, it was 2.1 times higher on average. As shown in Fig. 7, small objects such as passersby and distant cars could be detected with our technology but could not be detected without it.

Figure 8 shows the size distribution of detected objects. Our technology could detect much smaller objects. Specifically, the minimum width size was halved from 12 pixels without our technology to 6 pixels with it. This is because $2S$ (width) × S (height) grids are achieved with our technology when there are two divided images in the division process, as described in Section 2.

These results indicate that our technology can improve object-detection performance in not only SD images but also HD images by suppressing the image-division penalty.

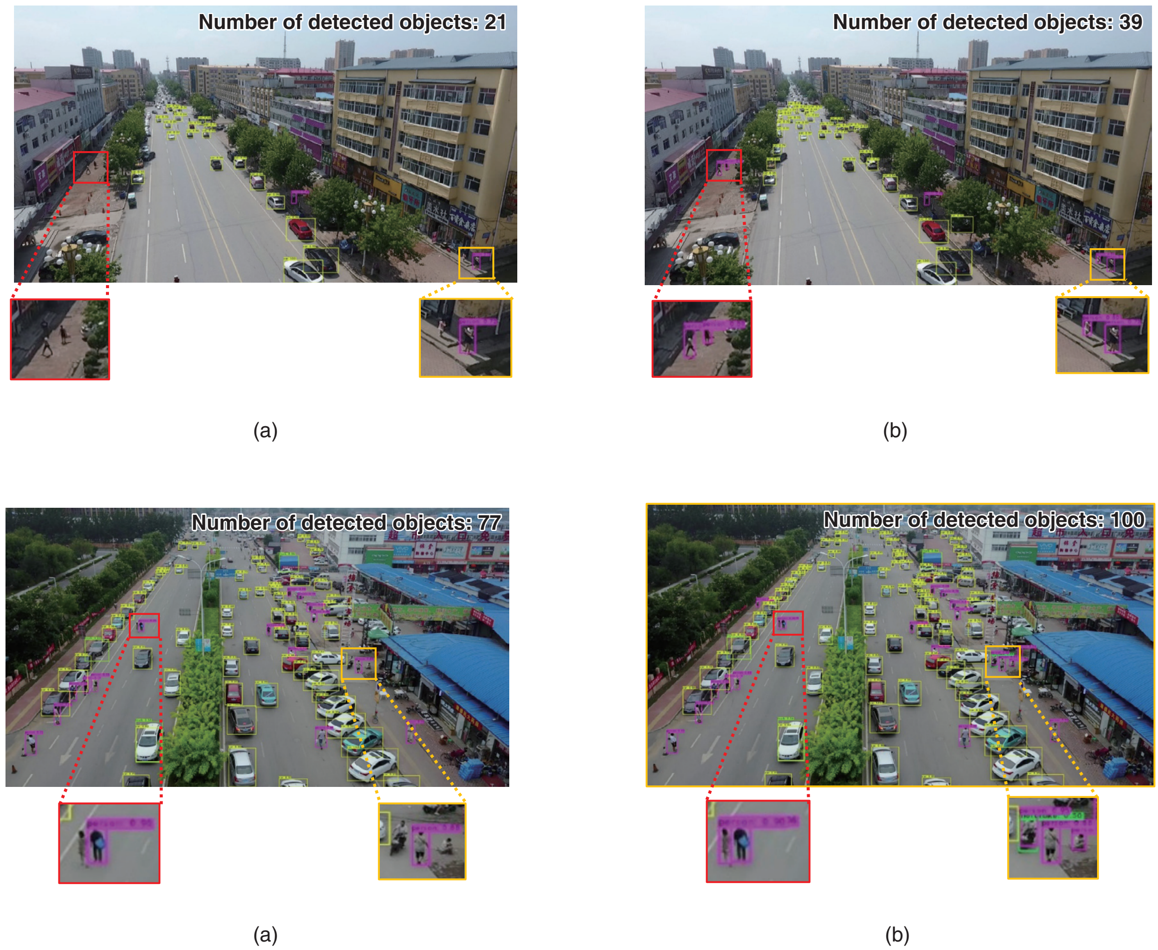


Fig. 7. Examples of detection results: (a) without our technology (standard YOLOv3 model only) and (b) with our technology.

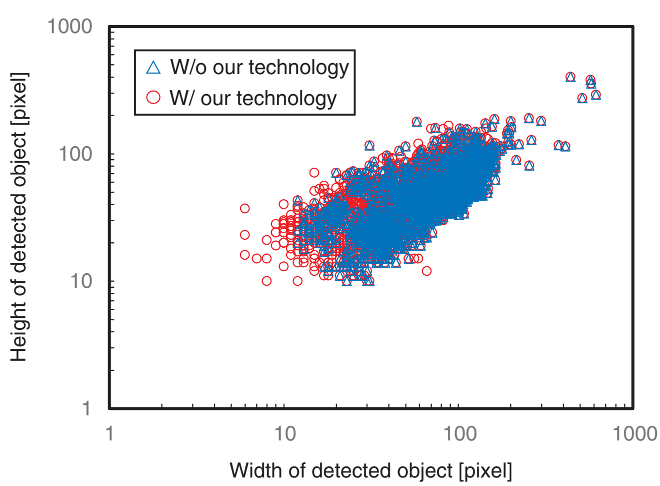


Fig. 8. Size distribution of detected objects in high-definition images.

4. Conclusion

This article introduced our technology that enables the detecting of a wide range of objects with one high-definition camera at one time for edge/terminal AI applications. It detects objects with a high detection resolution and wide reception field by combining the division process with multiple divided images and the whole process with the shrunk whole image.

We applied our technology to the standard YOLOv3 model with 608 × 608 pixels and evaluated object-detection performance. The evaluation results indicate that our technology can improve object-detection performance in not only standard definition images but also high-definition images while suppressing the image-penalty.

This technology is suitable for hardware implementation because all object detectors can be executed in parallel for real-time operation. Any AI inference scheme for object detection can be applied, and re-training for applying our technology is not necessary. This will facilitate its application to various edge/terminal AI applications.

References

[1] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," arXiv:1804.02767, 2018. https://doi.org/10.48550/arXiv.1804.02767

[2] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single Shot MultiBox Detector," arXiv:1512.02325, 2015. https://doi.org/10.48550/arXiv.1512.02325

[3] V. Růžička and F. Franchetti, "Fast and Accurate Object Detection in High Resolution 4K and 8K Video Using GPUs," Proc. of 22nd Annual IEEE High Performance Extreme Computing Conference (HPEC 2018), Waltham, MA, USA, Sept. 2018. https://doi.org/10.1109/HPEC.2018.8547574

[4] G. Plastiras, C. Kyrkou, and T. Theocharides, "Efficient ConvNet-based Object Detection for Unmanned Aerial Vehicles by Selective Tile Processing," Proc. of the 12th International Conference on Distributed Smart Cameras (ICDSC 2018), Eindhoven, The Netherlands, Sept. 2018. https://doi.org/10.1145/3243394.3243692

[5] D. Vorobjov, I. Zakharava, R. Bohush, and S. Ablameyko, "An Effective Object Detection Algorithm for High Resolution Video by Using Convolutional Neural Network," Advances in Neural Networks, LNCS, Vol. 10878, pp. 503–510, 2018. https://doi.org/10.1007/978-3-319-92537-0_58

[6] R. Bohush, S. Ablameyko, S. Ihnatsyeva, and Y. Adamovskiy, "Object Detection Algorithm for High Resolution Images Based on Convolutional Neural Network and Multiscale Processing," Proc. of the 4th International Workshop on Computer Modeling and Intelligent Systems (CMIS-2021), Zaporizhzhia, Ukraine, Apr. 2021. https://doi.org/10.32782/cmis/2864-12

[7] M. Gao, R. Yu, A. Li, V. I. Morariu, and L. S. Davis, "Dynamic Zoom-in Network for Fast Object Detection in Large Images," Proc. of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2018), Salt Lake City, UT, USA, June 2018, pp. 6926–6935. https://doi.org/10.1109/CVPR.2018.00724

[8] C. Tang, Y. Ling, X. Yang, and W. Jin, "Multi-view Object Detection Based on Deep Learning," Applied Sciences, Vol. 8, No. 9, p. 1423, 2018. https://doi.org/10.3390/app8091423

[9] H. Uzawa, S. Yoshida, Y. Iinuma, S. Hatta, D. Kobayashi, Y. Omori, K. Nakamura, S. Takada, H. Toorabally, and K Sano, "High-definition Object Detection Technology Based on AI Inference Scheme and Its Implementation," IEICE Electronics Express, Vol. 18, No. 22, P. 20210323, 2021. https://doi.org/10.1587/elex.18.20210323

[10] H. Uzawa, S. Yoshida, Y. Iinuma, S. Hatta, D. Kobayashi, Y. Omori, Y. Horishita, K. Nakamura, S. Takada, H. Toorabally, K. Nitta, K. Yamazaki, and K. Sano, "High-definition Technology of AI Inference Scheme for Object Detection on Edge/Terminal," IEICE Electronics Express, Vol. 20, No. 13, p. 20232002, 2023. https://doi.org/10.1587/elex.20.20232002

[11] Website of YOLO, https://pjreddie.com/darknet/yolo/

[12] T. Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár, "Microsoft COCO: Common Objects in Context," Computer Vision – ECCV 2014, 2014, LNCS, Vol. 8693, pp. 740–755. https://doi.org/10.1007/978-3-319-10602-1_48

[13] P. Zhu, L. Wen, X. Bian, H. Ling, and Q. Hu, "Vision Meets Drones: A Challenge," arXiv:1804.07437, Apr. 2018. https://doi.org/10.48550/arXiv.1804.07437

Hiroyuki Uzawa

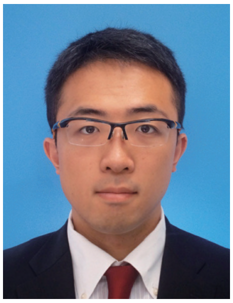
Senior Manager, NTT Device Innovation Center.

He received a B.E. and M.E. in electrical engineering from Tokyo University of Science in 2006 and 2008. In 2008, he joined NTT Microsystem Integration Laboratories, where he was engaged in research and development of the design techniques for network system on a chip (SoC). Since 2019, he has been engaged in research and development of a high-definition AI inference engine at NTT Device Innovation Center. He received the 2012 IEICE Young Engineer Award and the 2021 IEICE ELEX Best Paper Award. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) of Japan.

Shuhei Yoshida

Research Engineer, NTT Device Innovation Center.

He received a B.E. and M.E. in computer science and systems engineering from Kobe University, Hyogo, in 2014 and 2016. In 2016 he joined NTT Device Innovation Center and has been engaged in research and development on hardware design methodology, field-programmable gate array (FPGA)-based network-traffic-monitoring systems, and a high-definition AI inference engine. He is a member of IEICE.



Yuukou Iinuma

Engineer, NTT Device Innovation Center.

He received a B.E., and M.E. in information science and technology from the University of Tokyo in 2019 and 2021. He joined NTT in 2021 and is currently engaged in research and development of an AI inference engine for high-definition video and object-detection algorithms in 8K video.

Saki Hatta

Senior Manager, NTT Device Innovation Center.

She received a B.S. and M.S. in material engineering from Tokyo Institute of Technology in 2009 and 2011. In 2011, she joined NTT Microsystem Integration Laboratories, where she was engaged in the research and development of the design techniques for a network SoC. She is currently with NTT Device Innovation Center and engaged in research and development of a network-traffic-monitoring system. She was a recipient of the IEEE CEDA 2018 Young Researcher Award. She is a member of IEICE.

Daisuke Kobayashi

Senior Manager, NTT Device Innovation Center.

He received a B.E. and M.E. in information and communication engineering from the University of Electro-Communications, Tokyo, in 2007 and 2009. He joined NTT Cyber Space Laboratories in 2009 and since then has been engaged in research and development on high quality video coding and transmission. Since 2020, he has been engaged in research and development of an AI inference hardware engine.

Yuya Omori

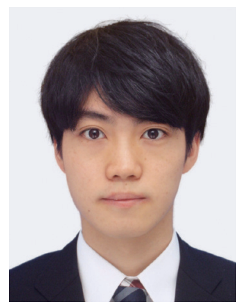
Research Engineer, NTT Device Innovation Center.

He received a B.E. and M.E. from the University of Tokyo in 2012 and 2014. In 2014, he joined NTT Media Intelligence Laboratories and has been engaged in research and development on parallel processing architecture and a high efficiency algorithm of video-processing hardware.

Yusuke Horishita

Senior Research Engineer, NTT Device Innovation Center.

He received a B.E. and M.E. in science and technology from Keio University, Tokyo, in 2010 and 2012. In 2012, he joined Canon Inc., where he was engaged in the development of SoCs for digital cameras and multi-function printers. Since 2022 he has been involved in research and development of an AI processor at NTT Device Innovation Center. He is a member of IEICE.



Tomoki Onoe

Researcher, NTT Device Innovation Center.

He received a B.S. and M.S. from Kyushu University, Fukuoka, in 2021 and 2023. In 2023, he joined NTT Device Innovation Center and has been engaged in research and development on video AI inference large-scale integration (LSI).

Ken Nakamura

Director, NTT Device Innovation Center.

He received a B.E. and M.E. in instrumentation engineering from Keio University, Kanagawa, in 1995 and 1997, and received a Ph.D. in information sciences from Tohoku University, Miyagi, in 2023. In 1997, he joined NTT Human Interface Laboratories and has been engaged in research and development on video coding LSI and systems. He is a member of IEICE and the Information Processing Society of Japan (IPSJ).

Koji Yamazaki

Director, NTT Device Innovation Center.

He received a B.S. in environmental information science from Keio University, Kanagawa, in 2000 and received a Master of Arts and Sciences (information studies) in interdisciplinary information science from the University of Tokyo in 2004. In 1998, he founded a venture company in Tokyo, Japan. In 2004, he joined NTT Microsystem Integration Laboratories. He has been engaged in research and development of LSIs and related FPGA accelerated systems.

Standardization Trends of Northbound APIs in 3GPP

Haruka Eitoku

Abstract

The 3rd Generation Partnership Project (3GPP) has been working on the standardization of the application programming interface (API) framework to enable 3rd party providers (e.g., application providers) to use the functionalities of an operator's network since Release 15 for collaboration with various industries. The API is called "northbound API" in 3GPP since it can be sketched as a northbound (upward) interface provided by the operators to 3rd party providers. This article provides an overview of the method for obtaining user consent when a 3rd party provider uses the northbound APIs exposed by an operator, which is standardized in Release 18 in expectation of the expansion of 5th generation (5G)/5G advanced/6G services and introduces NTT's standardization activities in 3GPP.

Keywords: 3GPP, northbound API, OAuth 2.0

1. Use cases of API invocation based on user's consent

An example in which a 3rd party provider accesses an application programming interface (API) on the basis of user consent is one of the new use cases in the expectation of the development of 5th generation (5G)/5G advanced/6G services. The service flow for the quality of service (QoS) setting during an online game is illustrated in **Fig. 1** as an example.

This example is a situation in which a user who has a contract with an operator plays a time-sensitive game using an application on the user equipment. The user may require higher quality and lower latency for their service experience. For the server of a 3rd party game provider to invoke the QoS API provided by the operator on the basis of the user request, the server needs to obtain the authorization from the operator to invoke the API. If additional fees are required for accessing the QoS API, the server needs to obtain the user's consent for the additional fees.

To enable use cases such as the above example, the 3rd Generation Partnership Project (3GPP) is working on the standardization of Resource owner-aware Northbound API Access (RNAA).

2. RNAA overview

2.1 OAuth 2.0

The procedure to obtain user consent on the basis of OAuth 2.0 is specified for RNAA. OAuth 2.0 is a mechanism widely used on websites to protect resources such as when the API is invoked. It is specified in RFC 6749 and other related documents produced by Internet Engineering Task Force (IETF). Several grant types are applicable for OAuth 2.0, and three grant types, i.e., client credentials grant, authorization code grant, and proof key for code exchange (PKCE), are supported for RNAA in 3GPP. In this section, the OAuth 2.0 flow (**Fig. 2**) based on the authorization code grant is described on the basis of clauses 1.2 and 4.1 of RFC 6749.

A client that requests access to the resource (e.g., API) owned by the resource owner (e.g., end user) initiates the flow by directing the resource owner's user agent (e.g., web browser) to the authorization endpoint inside the authorization server for obtaining the access rights to the API. The authorization server authorizes the resource owner via the user agent, and the resource owner notifies the authorization server to grant the client-access rights. This step for granting the access rights corresponds to the procedure for

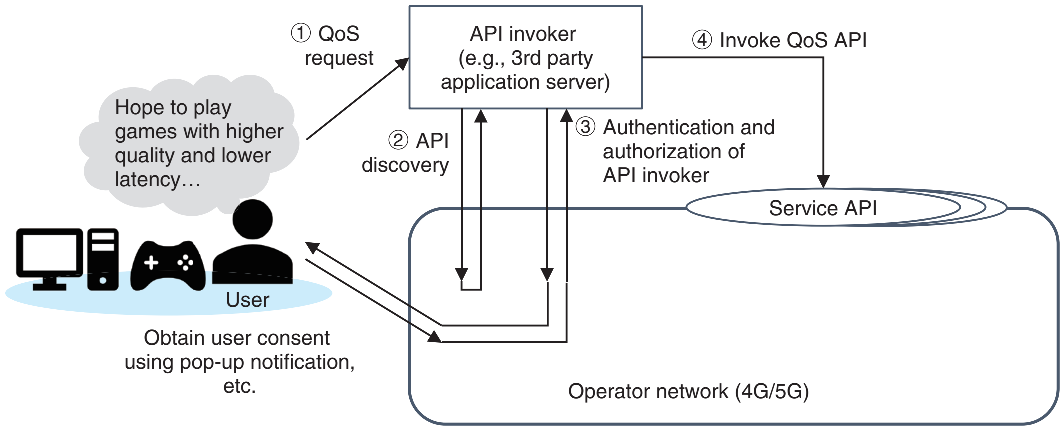


Fig. 1. New use cases of API usage based on user consent.

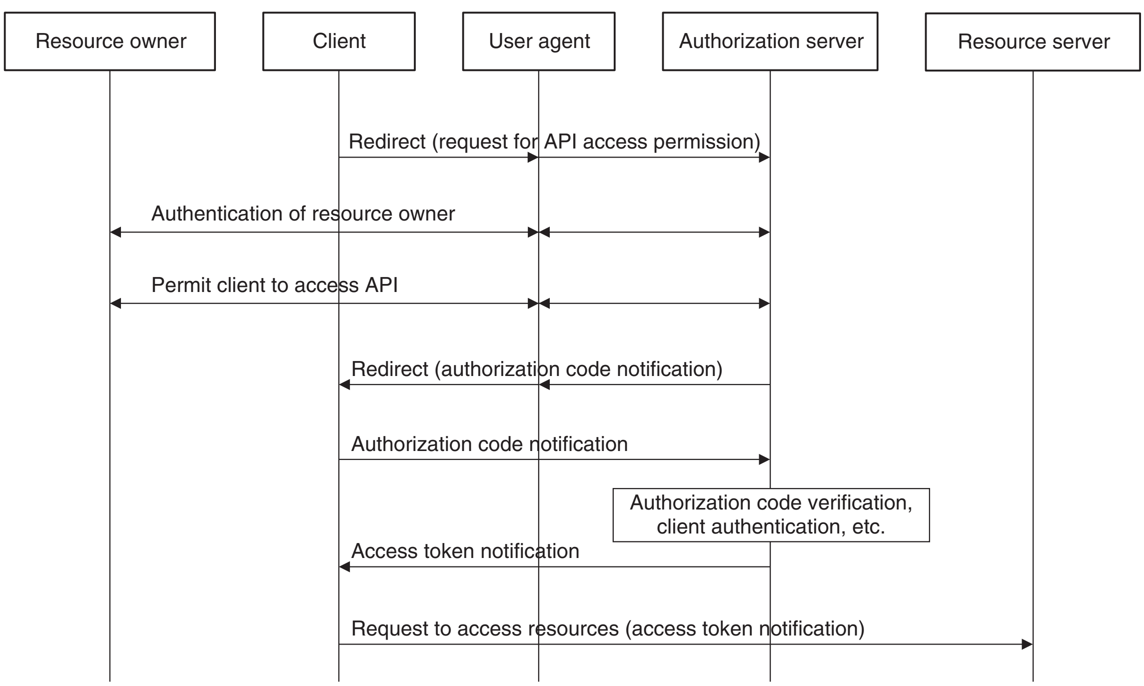


Fig. 2. OAuth 2.0 flow.

obtaining user consent in RNAA. The authorization server then redirects the user agent to the client. The client receives an authorization code, a one-time credential, from the user agent as an indication of the resource owner's permission to grant the access request and sends the code to the authorization endpoint. The authorization server verifies the authorization code and validates the client, and if the result is valid, the authorization server generates an access token and sends it to the client. The client requests accessing the resource server for accessing the resources with the access token. The resource server verifies the access token, and if the request is valid, the client can access the resource on behalf of the resource owner.

2.2 RNAA architecture

3GPP is considering extending the specifications of

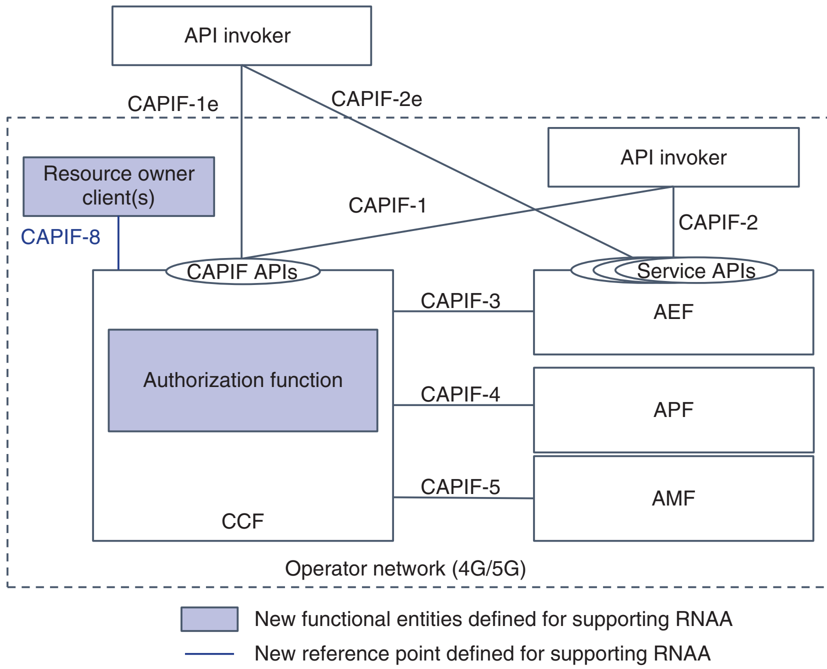


Fig. 3. RNAA architecture.

the Common API Framework (CAPIF) [1–3] for RNAA to enable new use cases for API accessing based on user consent.

The CAPIF allows 3rd party providers to use the functionalities of an operator's network and has been standardized by 3GPP since Release 15. The main capability of the CAPIF is discovering appropriate northbound APIs. The northbound APIs used by 3rd party operators are provided by the service capability exposure function (SCEF) and network exposure function (NEF) in the Evolved Packet Core (EPC) and 5G Core (5GC) networks. The SCEF/NEF is connected to the network functions of the EPC/5GC of carriers via the southbound interface, and the SCEF/NEF can access the EPC/5GC functionalities (e.g., QoS setting) via this interface on the basis of the request for API call from the 3rd party providers.

RNAA extends the CAPIF specifications by introducing new functional entities, called resource owner client and authorization function, and the related reference points (**Fig. 3**). The authorization function is inside the CAPIF core function (CCF) in Release 18, and the resource owner client resides on the user equipment.

The API invoker is the source of an API call, which can be either an operator or 3rd party provider. The API exposing function (AEF) exposes the northbound API called the service API, and the SCEF/NEF corresponds to the AEF. The API publishing function (APF) publishes information of the service APIs to the CCF. The API management function (AMF) manages the published service APIs (e.g., auditing the service API invocation logs received from the CCF). The CCF is the key functional part of the CAPIF and capable of authenticating API invokers, maintaining service API information, and discovering service APIs on the basis of the information. The API invoker can invoke the CAPIF API exposed by the CCF to execute common procedures such as authentication and authorization of the API invoker for invoking the service APIs and discovering APIs.

Although the CAPIF assumes cases in which the CCF and AEF/APF/AMF are provided by different providers, the current RNAA specification assumes that the CCF and AEF/APF/AMF are all provided by the same operator.

2.3 RNAA procedure

In the procedure [3] to obtain user consent for invoking the service API between the API invoker and AEF, the API invoker first calls the CAPIF API exposed by the CCF to authenticate itself to the CCF, which discovers the appropriate API after the API invoker's authentication. The API invoker then calls the CAPIF API again to negotiate the method for authentication and authorization (i.e., security method) between the API invoker and AEF. If user consent is required, the authorization method used to obtain user consent is expected to be negotiated at this time. The AEF authenticates the API invoker using the negotiated security method, and the authorization server authorizes the API invoker using one of the grant types supported by RNAA (i.e., client credentials grant, authorization code grant, PKCE).

3. Future prospects

3GPP has been working on the standardization of detailed procedures for RNAA since 2023 in Technical Specification Group (TSG) Core Network and Terminals Working Group 3 (CT3)[*1], and TSG Service and System Aspects Working Group 3 (SA3)[*2]. The specifications provided by CT3 and SA3 follow the service requirements and architecture specified in SA1[*3] and SA6[*4]; NTT is responsible for the work in CT3. The normative work of RNAA in Release 18 is expected to be completed (at the time of writing this article) in March 2024. NTT will continue to contribute for the collaboration with various industries and the creation of new services through standardization activities in 3GPP.

References

[1] 3GPP TS 23.222: Common API Framework for 3GPP Northbound APIs, https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3337

[2] 3GPP TS 29.222: Common API Framework for 3GPP Northbound APIs, https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3450

[3] 3GPP TS 33.122: Security aspects of Common API Framework (CAPIF) for 3GPP northbound APIs, https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3420

Haruka Eitoku

Researcher, NTT Network Service System Laboratories.

She received an M.S. from Graduate School of Science, Kyushu University, Fukuoka, in 2017. Since joining NTT in 2017, she has been contributing to standardization efforts in 3GPP for interconnection between IMS networks and for immersive real-time communication technologies. She received the Telecommunication Technology Committee (TTC) Distinguished Service Award in 2000.

*1 CT3: Group studying interworking between networks and protocols in 3GPP.

*2 SA3: Group studying architectures and protocols for security and privacy in 3GPP.

*3 SA1: Group studying service requirements.

*4 SA6: Group studying application enabler.

Precautions for Installing and Repairing Powder-coated Messenger Wires

Technical Assistance and Support Center, NTT EAST

Abstract

This article presents precautions to be taken when installing and repairing powder-coated messenger wires to prevent corrosion in salt- and sulfur-affected areas. This is the eighty-first article in a series on telecommunication technologies.

Keywords: messenger wire, powder coating, corrosion protection

1. Introduction

NTT's telecommunication equipment and facilities are installed throughout Japan in a variety of natural environments. The metals, concrete, and plastics used as materials for these facilities and equipment deteriorate over time due to effects from the surrounding environment, such as ultraviolet rays, rainwater, and seawater.

Most outdoor metal equipment and facilities are hot-dip galvanized for corrosion protection. However, in severe natural environments, such as coastal areas that suffer salt damage (due to salt blown inland from the ocean) and hot-spring areas that suffer sulfur damage (due to the sulfur-rich air), the galvanization is severely degraded in a manner that rapidly exposes and corrodes the underlying metal to form red corrosion (highlighted by the red frame in **Fig. 1**).

To protect telecommunication equipment and facilities in such corrosive environments, applying a thick and durable powder coating (**Fig. 2**) over hot-dip galvanized equipment can extend the lifespan of the equipment, as described in the June 2022 issue of this journal [1].

The powder-coating material developed by Technical Assistance and Support Center (TASC), NTT EAST adheres well to hot-dip-galvanized surfaces and is applicable to structures of various shapes. This article presents precautions to be taken when installing and repairing powder-coated messenger wires to prevent corrosion in salt- and sulfur-affected areas.

Fig. 1. Red corrosion on a messenger wire.

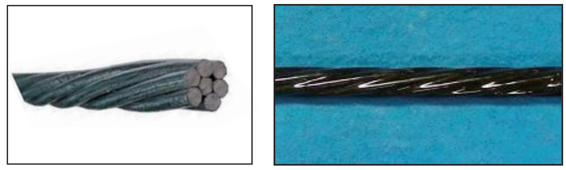


Fig. 2. Messenger wire with and without powder coating.

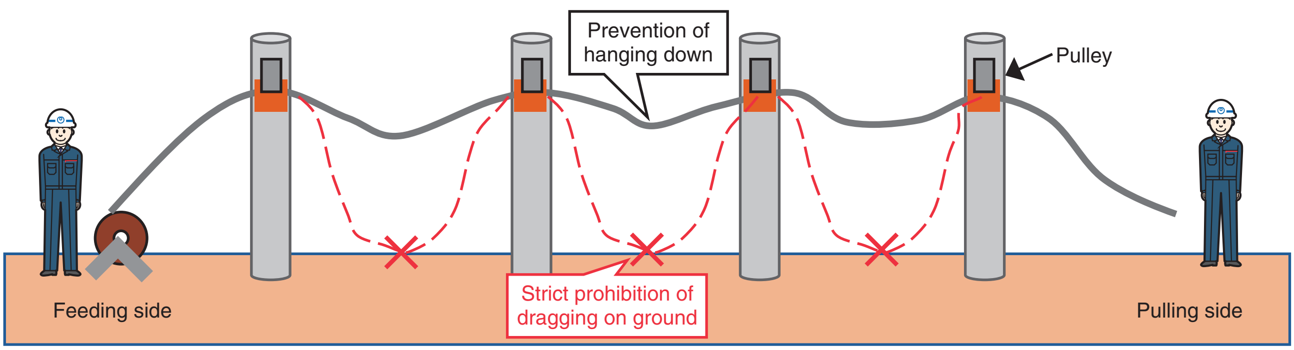


Fig. 3. Measures to prevent damage when installing powder-coated messenger wires.

2. Precautions during design and installation

Powder coating is a method of forming coating films by melting a powder at temperatures of approximately 250°C or higher at factories. As it is difficult to apply powder coating to installed equipment on site, equipment that has been powder coated is either repaired or newly installed. The following section presents precautions to be taken when designing and installing powder-coated messenger wires.

2.1 Precautions during design

When a messenger wire is being newly installed, metal fixtures such as universal bands[*1], dead-end grips[*2], and suspension clamps[*3] are also installed around the messenger wire. Most fixtures can be powder coated before use in the same manner as messenger wires. However, when designing the installation of messenger wires, it is important to avoid using components [2] such as messenger-wire crossing fixtures[*4] that may put stress on the coating film and cause deterioration such as peeling. Since a powder-coated messenger wire has a larger diameter than a non-coated one, it cannot be fitted into a suspension clamp, so basically both ends of the wire are fixed to utility poles. However, the current messenger wire with minimum diameter (cross-sectional area of 30 mm^2) can be fitted into suspension clamps.

2.2 Precautions during installation

The following precautions must be considered when installing powder-coated messenger wires on site. These precautions can be divided into those that focus on mechanical properties and those that focus on electrical properties.

2.2.1 Precautions focusing on mechanical properties

When powder-coated messenger wires are installed, if the coating is peeled off by dragging on the ground or contact with an obstacle, deterioration will progress from that contact point. As shown in **Fig. 3**, to prevent peeling of the coating, a pulley should be installed at the messenger wire string point on each

*1 Universal band: A metal band that wraps around a utility pole and is used as a base for attaching various metal fixtures.
*2 Dead-end grip: A twisted steel rope that wraps around a messenger wire when it is strung to a utility pole.
*3 Suspension clamp: A metal fixture attached to a universal band to grip and secure a messenger wire.
*4 Messenger-wire crossing fixture: A fixture for securing messenger wires that cross in mid-air.

utility pole, and the tow rope should be pulled through the entire section to prevent dragging on the ground or coming in contact with obstacles due to loosening.

2.2.2 Precautions focusing on electrical properties

A ground clamp and ground wire are attached to a messenger wire to ensure electrical safety. However, the powder coating on the messenger wire is made from an insulating resin, so it is not possible to establish conductivity with the wires under the coating simply by attaching a ground clamp. It is therefore necessary to remove the coating with a tool, such as pliers (recommended) or a wrench, and then attach the ground clamp to the exposed wire (**Fig. 4**). To lower the grounding resistance by connecting as many wires as possible, it is appropriate to install two ground clamps side by side.

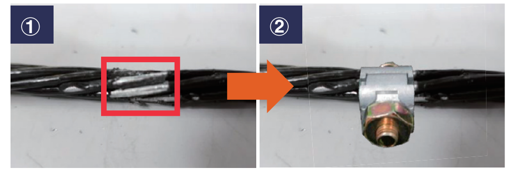


Fig. 4. Installing a ground clamp.

3. Precautions during repair

3.1 Necessity of repair and repair materials

The powder coating and hot-dip-galvanizing applied to messenger wires and surrounding fixtures can become scratched or peel off due to various factors during installation. Damage to the powder coating and galvanized plating will shorten the lifespan of the entire messenger wire, so it is necessary to repair the damaged area by using a method appropriate for the area to restore corrosion resistance.

One repair method is to apply an anti-corrosion coating to the damaged area of the messenger wire. With on-site application of anti-corrosion coating, it is difficult to ensure an even distribution of coating thickness over long distances, and there is a risk of paint splattering, but it can be used as a localized repair method. Two types of coatings are available: liquid [3] and spray [4]. Although the liquid type has excellent corrosion resistance, it must be applied with a brush, so it is difficult to sufficiently apply it to confined spaces. For that reason, it is recommended to repair by applying the liquid type with a brush for open areas, such as the surface of messenger wires where no fixtures are attached, and by using the spray type in confined spaces, such as the fastening points of fixtures.

3.2 Precautions for each repair location

Example locations of repairs to a powder-coated messenger wire are shown in **Fig. 5**. The repair methods and precautions for each repair location are explained as follows.

3.2.1 Cut surface of messenger wire and end of dead-end grip

The cut surface of messenger wire (Fig. 5(A)) and the end of the dead-end grip (Fig. 5(B)) are prone to damage to the coated and plated surfaces owing to impact during cutting and metal-to-metal contact. Since these locations are open areas and relatively easy to repair, repair by applying a liquid-type coating with a brush is recommended.

3.2.2 Fastening point of ground clamp

As explained in Section 2.2.2, when a ground clamp (Fig. 5(C)) is attached to a messenger wire, the coating is removed with a tool such as pliers to ensure conductivity. After the ground clamp has been fastened, the area where the coating has been removed must be repaired to prevent corrosion. For confined spaces such as the gaps between clamps, where it is difficult to apply a liquid-type coating with a brush, it is recommended to apply a spray-type coating. When repairing by spray, it is advisable to fasten the bolt of the ground clamp and spray from upwind to prevent the person who is spraying from inhaling the spray. Measures must also be taken to prevent the spray from dispersing into private houses or dripping onto the ground.

3.2.3 Scratches during attachment of universal band

When attaching a universal band (Fig. 5(D)), an operator uses a hammer to fit the band to the pole, but it is recommended to use a resin hammer instead of a normal (metal) hammer because a metal hammer may damage the coated surface of the band. As shown in **Fig. 6**, even a resin hammer can peel of the coating due to the impact of striking, so any peeled areas should be repaired by re-applying the coating by brush.

3.2.4 Fastening points of wire gripper, branching fixture, and end clamp

At the fastening points of a wire gripper (Fig. 5(E)), branching fixture (Fig. 5(F)), and end clamp (Fig. 5(G)), the tightening force will cause scratches and

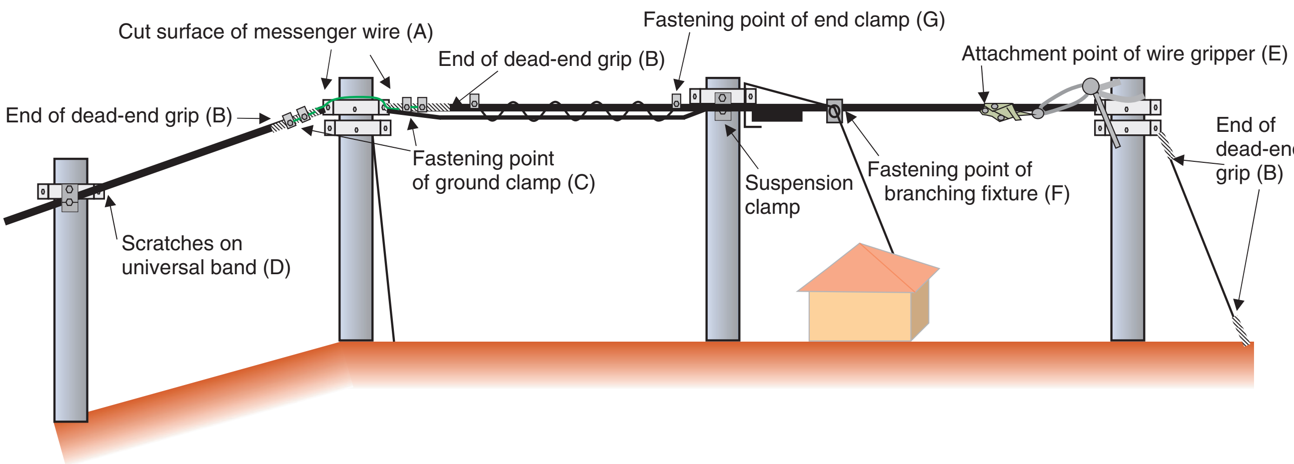


Fig. 5. Example of locations where powder-coated messenger wire is repaired.

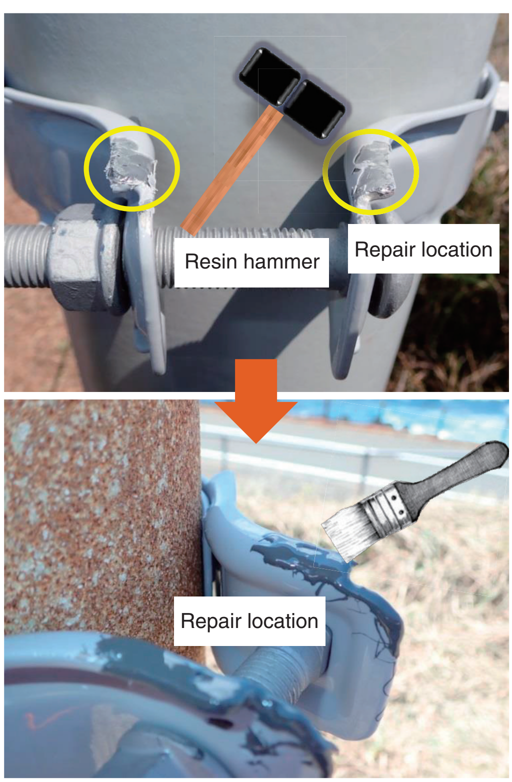


Fig. 6. Method of repairing universal band.

dents on the coated surface of the messenger wire. To repair that damage, it is recommended to apply a liquid-type coating to the area where the wire gripper was installed and apply a spray-type coating to other areas (i.e., confined spaces) in the same manner as for the ground clamp.

4. Conclusion

To protect telecommunication equipment and facilities from corrosion caused by environmental effects such as salt damage, it is effective to apply a strong corrosion-prevention powder coating; however, it is necessary to consider certain points of note in the design and installation of powder-coated equipment that do not apply to normal processes. By paying attention to these points and carrying out appropriate installation and repair, the excellent anti-corrosion performance of a coated surface will be manifested, and the lifespan of the coated equipment will be extended to the full.

TASC will continue to engage in technical cooperation to solve problems in the field, such as issues related to equipment and facility deterioration due to salt damage and other types of corrosion, and contribute to improving the quality and reliability of telecommunication services.

References

[1] Technical Assistance and Support Center, NTT EAST, "Salt Damage in RT-BOXes: Investigation and Countermeasures of Salt-damage Environments," NTT Technical Review, Vol. 20, No. 6, pp. 52–56, June 2022. https://doi.org/10.53829/ntr202206pf1
[2] Website of Zentsukyo, NTT Specification Product List (in Japanese). https://www.zentsukyo.or.jp/product-list/product-ntt-list
[3] Website of ACCESS, RUSTAFF 2100 Series (in Japanese). http://www.access-orbit.co.jp/anticorrosion/index_f.html
[4] Website of ROVAL, ROVAL Cold Galvanizing Compound. https://rovalworld.com/products/roval/

External Awards

Jury Best Demo Honorable Mention/People's Choice Best Demo Honorable Mention

Winner: Kentaro Yasu, NTT Communication Science Laboratories
Date: November 2, 2023
Organization: The 36th Annual ACM Symposium on User Interface Software and Technology (UIST 2023)

For "Demonstrating SuperMagneShape: Interactive Usage of a Passive Pin-based Shape-changing Display."
Published as: K. Yasu, "Demonstrating SuperMagneShape: Interactive Usage of a Passive Pin-based Shape-changing Display," Proc. of UIST 2023, Article no. 59, San Francisco, CA, USA, Oct./Nov. 2023.

Excellent Presentation Award

Winners: Kohei Takahashi, NTT Human Informatics Laboratories; Narimune Matsumura, NTT Human Informatics Laboratories; Hitoshi Seshimo, NTT Human Informatics Laboratories; Shinkuro Honda, NTT Human Informatics Laboratories
Date: January 23, 2024
Organization: The Society of Instrument and Control Engineers (SICE)

For "Examination of a Tele-nursing System on the Subject of Remote Sputum Suction."
Published as: K. Takahashi, N. Matsumura, H. Seshimo, and S. Honda, "Examination of a Tele-nursing System on the Subject of Remote Sputum Suction," Proc. of the 24th SICE System Integration Division Conference (SI2023), 1B5-01, Niigata, Japan, Dec. 2023.

Excellent Presentation Award

Winners: Taichi Kanada, NTT Human Informatics Laboratories; Daisuke Satou, NTT Research and Development Planning Department; Masato Miyahara, NTT Human Informatics Laboratories; Narimune Matsumura, NTT Human Informatics Laboratories; Hitoshi Seshimo, NTT Human Informatics Laboratories
Date: January 23, 2024
Organization: SICE

For "Operator Response Detection Methods for Semi-autonomous Teleoperation Systems."
Published as: T. Kanada, D. Satou, M. Miyahara, N. Matsumura, and H. Seshimo, "Operator Response Detection Methods for Semi-autonomous Teleoperation Systems," Proc. of SI2023, 3E1-05, Niigata, Japan, Dec. 2023.

Best Presentation Award

Winners: Ryota Imai, NTT Human Informatics Laboratories; Mitsuhiro Goto, NTT Human Informatics Laboratories; Kenji Esaki, NTT Human Informatics Laboratories; Hitoshi Seshimo, NTT Human Informatics Laboratories
Date: January 27, 2024
Organization: The Institute of Electronics, Information and Communication Engineers (IEICE) Technical Group on Media Experience and Virtual Environment (MVE)

For "A Control Method for Bodily Sensory Reproduction Player of Windsurfing Considering Frequency Characteristics of Athlete's Swaying/Tilting Sensation."
Published as: R. Imai, M. Goto, K. Esaki, and H. Seshimo, "A Control Method for Bodily Sensory Reproduction Player of Windsurfing Considering Frequency Characteristics of Athlete's Swaying/Tilting Sensation," IEICE Tech. Rep., Vol. 123, No. 359, MVE2023-37, pp. 25–30, Jan. 2024.

Papers Published in Technical Journals and Conference Proceedings

Coherent Response of Inhomogeneously Broadened and Spatially Localized Emitter Ensembles in Waveguide QED

L. Ruks, X. Xu, R. Ohta, W. J. Munro, and V. M. Bastidas
Physical Review A, Vol. 109, 023706, Feb. 2024.

Spectrally and spatially varying ensembles of emitters embedded into waveguides are ever-present in both well-established and emerging technologies. If control of collective excitations can be attained, a plethora of coherent quantum dynamics and applications may be realized on-chip in the scalable paradigm of waveguide quantum electrodynamics (WQED). Here, we investigate inhomogeneously broadened ensembles embedded with subwavelength spatial extent into waveguides employed as single effective and coherent emitters. We develop a method permitting the approximate analysis and simulation of such mesoscopic systems featuring many emitters, and show how collective resonances are observable within the waveguide transmission spectrum once their linewidth exceeds the inhomogeneous line. In particular, this allows for near-unity and tailorable non-Lorentzian extinction of waveguide photons overcoming large inhomogeneous broadening present in current state-of-the-art implementations. As a particular illustration possible in such existing experiments, we consider the classic emulation of the cavity QED (CQED) paradigm here using ensembles of rare-earth ions as coherent mirrors and qubits and demonstrate the possibility of strong coupling given existing restrictions on inhomogeneous broadening and ensemble spatial extent. This work introduces coherent ensemble dynamics in the solid state to WQED and extends the realm to spectrally tailorable emitters.