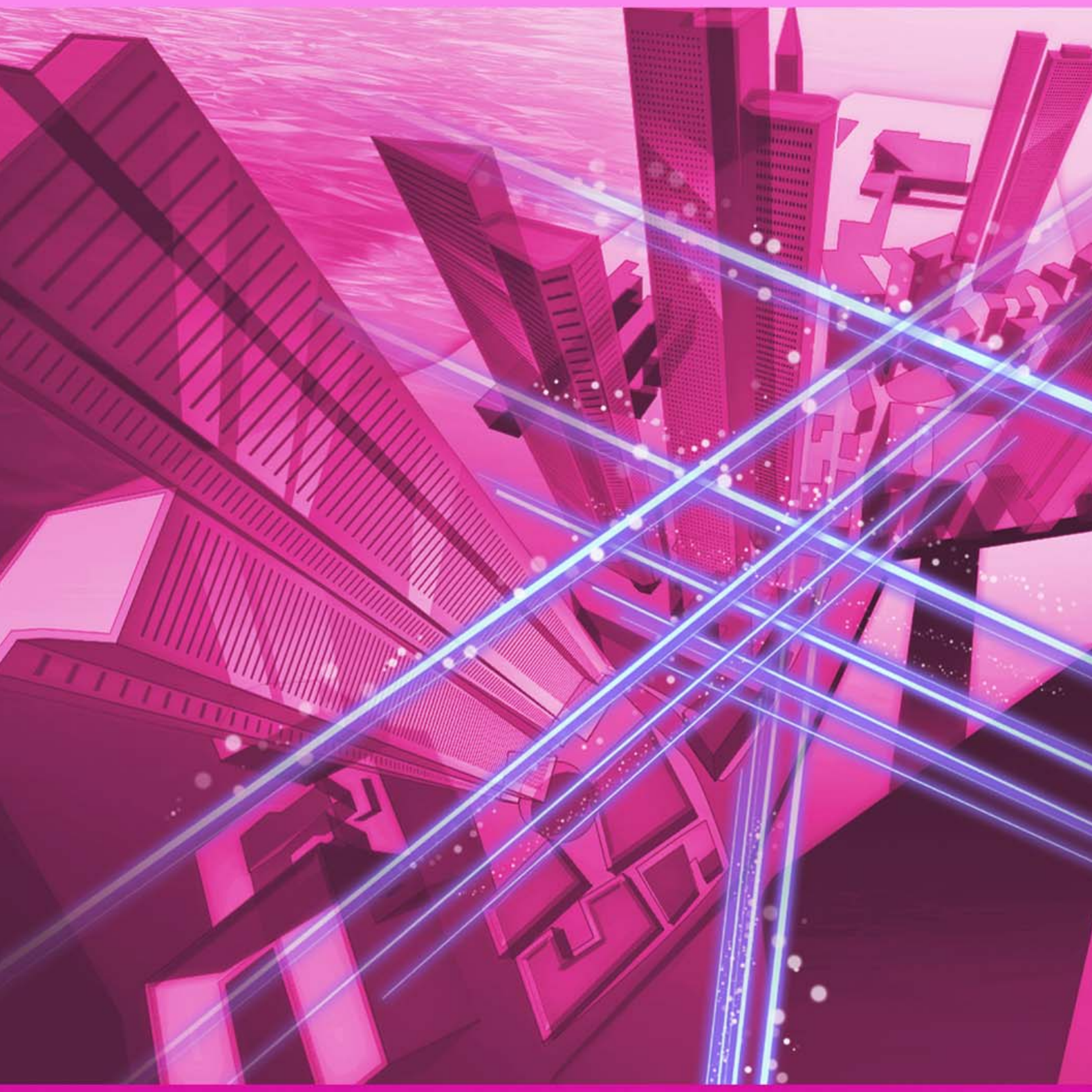
# NTT Technical Review

11
2020



November 2020 Vol. 18 No. 11

**NTT Technical Review**

# November 2020 Vol. 18 No. 11

# Listening to Employees' Thoughts and Wishes Then Make Them a Reality

*Yoshihiro Kuroda*
*President, NTT InfraNet*

## Overview

In 2019, NTT InfraNet celebrated its 20th anniversary. Based on its original mission of providing total management solutions for infrastructure facility operations, the company is reshaping its business to accord with the current times and undertaking initiatives such as responding to natural disasters, reforming work style by improving productivity and efficiency, and solving social issues. We asked Yoshihiro Kuroda, President of NTT InfraNet, about his leadership style and smart infrastructure management solutions that use the technological capabilities required of telecommunication infrastructure professionals.

*Keywords: infrastructure facility management, digital transformation, smart maintenance*

### Reshaping the infrastructure facility management business to go forward

*—It has been four years since your last appearance in "View from the Top." Please tell us about NTT InfraNet from your new position as president.*

NTT InfraNet was established in 1999 as a subsidiary of NTT EAST. As a group of experts who have fully and uniformly inherited the business and technical expertise related to telecommunication infrastructure facilities such as underground conduits and manholes, we have engaged in the management of the infrastructure facilities of NTT Group companies such as NTT EAST, NTT WEST, NTT DOCOMO, and NTT Communications. In July 2019, to create a business that can be provided to companies other than NTT Group with the *Smart Infra Platform* by centralizing infrastructure data in a three-dimensional (3D) manner, we became a wholly owned subsidiary of NTT holding company, changing ownership from NTT EAST. Moreover, we took over the civil-engineering infrastructure and business of NTT Communications, acquired NTT Geospace and AIREC Engineering as subsidiaries in December 2019 and March 2020, respectively, and merged with NTT Geospace in July 2020.

Although the operation and maintenance of infrastructure facilities within the NTT Group has already been unified, the requirements of each company slightly differ. Accordingly, since taking up my post in June 2019, I have been working on creating a system for efficiently operating and managing these infrastructure facilities. The infrastructure possessed by the NTT Group includes about 670,000 km of underground conduits, through which cables run, about 690,000 manholes, and about 654 km of tunnels called service tunnels. The volume of these facilities is about the same as that of the domestic water supply. Since the total length of all domestic

gas pipes is about 270,000 km, it is easy to see that the length of our conduits far exceeds that length. Development of a database is indispensable for efficiently operating and managing these infrastructure facilities, and using the database in more-advanced ways will lead to the establishment of the Smart Infra Platform.

It has been 40 to 60 years since the peak of construction of these infrastructure facilities in Japan, and the civil engineers who maintain them are, naturally, aging, and the working population will continue to decline. It is necessary to maintain efficient operation and management of many infrastructure facilities. Information about infrastructure facilities had mainly been managed on paper; however, we want to respond to aging facilities and shortage in the workforce by speeding up digital transformation (DX), such as digitizing information about infrastructure facilities and using it for tasks on devices such as tablets and making inspections and other tasks smarter by using information and communication technology (ICT). In July 2020, we launched the Smart Infra Business Concept to create the Smart Infra Platform by building a centralized database of infrastructure facilities using ICT, which will also be used by infrastructure companies other than the NTT Group. Through these initiatives, we want to contribute to society while increasing our profits.

*—What specific changes will DX bring to your daily work?*

Managing 3D and 4D facility data on the Smart Infra Platform makes it possible to not only improve the efficiency of our work but also smoothly coordinate work for constructing roads, for example. As with communication cables, electric-power pipelines and cables, gas pipes, water pipes, and so on are generally buried under roads. When constructing roads, it is necessary to check the status of these facilities underground and visit the offices of their owners and users individually to discuss the method of construction (excavation). By storing the facility data of each company on the Smart Infra Platform, one can check the status of these facilities in one step, thereby simplifying the procedure and further reducing the number of on-site observations. In the future, when it will be possible to search for infrastructure-facility information on the Internet, we will be able to help solve social issues such as reducing overlapping investments due to repeated excavation and backfilling by individual construction works for electricity, gas, water supply, and telecommunications and relieving traffic congestion by shortening construction periods.

We are promoting *smart maintenance*, which uses ICT for the purpose of efficient equipment maintenance, as our solution concerning *triple I+P* (Infrastructure, IT, Innovation Platform). In particular, smart maintenance is used for more efficient

management of facility inspection and for proposals of facility-renewal plans by creating a real-time *mash up* of external information published by government offices, local governments, other companies, and other sources in addition to internal information. Regarding our Mobile Mapping System (MMS) solution, images of manholes, etc. are captured with a camera or laser mounted on a car and subjected to high-precision 3D measurement, and we are thinking of using these measurement data for tasks such as diagnosing facility deterioration, observation, construction, inspection, and repair work. Smart maintenance and MMS have already been introduced for the management of facilities and inspection of utility poles in operations of access-system facilities of NTT EAST and NTT WEST.

NTT InfraNet has conducted civil-engineering solutions and gained the trust of the Ministry of Land, Infrastructure, Transport and Tourism and local governments through long-term relationships. From the viewpoints of countermeasures against traffic accidents, landscape protection, and safety measures against damage such as fallen utility poles due to natural disasters, we received many requests to bury utility poles in the ground and provide geospatial data management systems. I feel that there are expectations for us to tackle social issues using ICT. However, we cannot do everything on our own, so we will collaborate with other NTT Group companies to tackle these issues.

## Creating a story that benefits all stakeholders

*—What is the key to maximizing resources while working together?*

To make the best use of collaboration and resources, it is important that we harmonize with the goals of each company. To take the flow in the direction you want to go, create a story that matches the goals of each company. Of course, it is also necessary to think and act in a give-and-take manner, and if we do not interweave these elements, the story will stop halfway. Consensus building is necessary even if it takes time. To that end, we must work in good faith. We cannot only pursue our own interests; in other words, how to create a story that benefits all stakeholders is important. In some cases, you may have to bear the costs or take a loss at the beginning, but I think that if you take the time to work in good faith, you will gain trust. Within the company, trust relationships are also of the utmost importance. One of the basics is to have one goal and gain faith.

Even if we proceed with our business in this manner, various problems and conflicts will sometimes occur. The requests and contract details sent to us from each company are not the same. Some companies present a comprehensive annual construction budget in an effort to improve efficiency, while other companies determine the budget for each process. Some companies want to entrust maintenance and operation of not only infrastructure facilities but also cables. From our viewpoint, we want to apply the same standard for contracts, but that won't happen overnight, so I think the only way to get such standardization right is to develop standards little by little over time. At the end of the day, it is important for each company to make a profit and recognize the benefits of our services such as cost reduction and quality and for NTT InfraNet to become an indispensable partner for them.

*—It is important to gain trust.*

Regardless of the size of the work, there are mountains to be climbed, and a relationship of trust is essential to climb them. The reorganization of NTT that I experienced was a big mountain. It was my first experience of an organizational design. At the stage of considering and coordinating the transfer of the

mission and organization to each company after the reorganization, various departments made requests based on their own circumstances and reasoning, such as wanting to move a certain unit to the holding company or wanting to send certain people to certain departments. At that time, NTT InfraNet was also established, becoming an independent company. Anyway, I had to take into account the various thoughts of each organization and form them into a concrete plan. I remember that it was quite difficult to reconcile everyone's interests because the scale of the measures was large and covered many fields.

No matter how difficult it is, I want to get the job done properly. There was no question that I was not allowed to leave the task incomplete, but I've been paying the most attention to how I can proceed toward the results, not to mention the accomplishments. That is still something I'm always thinking about. When you are young, you may not be able to complete your tasks and you may get stuck. At such times, I'm telling young people to go to see other people. It is valuable to have various people teach you things, and those who listen and talk will stand by you. It is exactly killing two birds with one stone that will increase your knowledge and friends. I'm sure there are a lot of things in the company life of young engineers that they don't understand. In such cases, solutions won't suddenly appear; I think the shortest way is to take it steadily. There are many problems you won't be able to solve no matter how much you think at your desk but can solve by visiting and meeting people. By the way, you don't have to talk much. Have other people teach you. I was the same. Start by having people teach you, then, if there is something that seems to be useful for you, go with it even little by little. So why not try working in that order?

## Self-realization through work

*—If you listen to people more, you will be able to work better.*

I think it's quite difficult to listen to people. It's harder to listen than to talk. When listening to opinions, including those of a subordinate, never lose your temper, interrupt their talking, or negate their words midway through a point. It is important to create an atmosphere in which people can talk freely. When people seem to feel uneasy about speaking out, they tend to hold back on saying something important. In any case, whenever someone comes to see me, I give them my top priority and listen to what they have to say. If you can't do that, you won't be able to get information. Since taking office, I've visited each of our locations around the country and exchanged opinions with employees. On those occasions, I repeatedly said, "Tell me anything," listened to anything they had to say, and gave them proper feedback. Even if I couldn't grant all requests, on returning to the head office, I always tried to find a way to grant one

or two of the requests that came up in each area I visited. I consider work is the main point of contact between a person and society. Since humans are social beings, we cannot live alone without social contact. I think that work is a means to self-realization and the most-familiar social activity in our lives.

*—It's true that work is an activity that occupies most of the day, and you naturally show your own personality.*

Some people emphasize the importance of clearly separating their work and private lives in the manner of the work-life balance, namely, harmonizing their work and private lives; however, for me, work and private lives are inextricably linked. I think that the most important thing is to be connected to society, both at work and in private. In addition, going into the field and listening to people's stories will continue to be a consistent part of my work. Nothing is easy, so I'll continue to do my best.

However, to prevent the spread of the novel coronavirus, the situation is not going the way I want. Those who suffer most are new employees. It is relatively easy to switch to remote work after you have built some relationships with colleagues, but to work remotely straight off the bat is difficult. Immediately after joining the company, new employees are stuck at home and do not get enough training, and they do not know where their fellow employees who joined the company at the same time are assigned. Even when they go to the office, their faces are partly hidden by a mask. It was also a big issue for me. I can send information via the intranet, but the opposite (such as collecting feedback, opinions, and information) is very difficult. Naturally, communication is insufficient. I think there are many things they want to say, but I can't fully grasp those thoughts. To somehow overcome this situation, we have therefore started meetings to exchange ideas on an area-by-area basis. We limit the number of people attending face-to-face meetings and connect people with other employees at other offices via web conferences so they can exchange ideas. I visited all departments in 2019, but in 2020, under the declaration of a state of emergency due to the coronavirus pandemic, I haven't been able to move as I expected. While keeping a close eye on the situation, I resumed office visits after the state of emergency was lifted in May, and I think I'll be able to visit all our major sites by the beginning of next year.

**Interviewee profile**

■ Career highlights

Yoshihiro Kuroda joined Nippon Telegraph and Telephone Public Corporation (now NTT) in 1981. After serving at NTT Communications as director of Consumer & Office Users Business Division, Strategy Planning Department and director of Human Resource Development Group, he became senior director of NTT General Affairs Department in 2008, director of NTT WEST Hiroshima office and general manager of NTT WEST Chugoku regional headquarters in 2011, executive director of NTT WEST Network Department in 2014, and senior executive vice president of NTT WEST in 2016. He took up his current position in June 2019.

# Do Your Best to Be Second to None in Achieving What You Want to Do

*Takahiro Kawabe*
*Senior Distinguished Researcher, NTT Communication Science Laboratories*

**Overview**

Many people would be surprised if they looked at a poster or signboard that they thought was a still image but was in fact moving. NTT Communication Science Laboratories is conducting research to acquire a scientific understanding of human sensory-information processing. Research on visual illusions has made it possible to develop an information-presentation technique that provides a unique and eye-catching experience that had not existed before. We asked Takahiro Kawabe, a senior distinguished researcher at NTT Communication Science Laboratories, who is engaged in this research, about the current progress of his research and his attitude as a researcher.

*Keywords: visual illusion, Poisson effect, pseudo-tactile sensation*

## Applying illusions in the real world

*—Please tell us about the research you are currently conducting.*

My current research theme is information presentation using illusions. *Illusion* refers to the perceptual characteristic of humans in perceiving something different from what is happening in the real world. Such an illusion is generally thought of as undesirable because we think it is preventing us from accurately seeing the real world. However, thinking that using an illusion makes it possible to create a new way of expression and provide a sensory experience that is unlikely to occur in the real world, I and other researchers at NTT Communication Science Laboratories developed a technique called Hengento (deformation lamp) [1] (**Fig. 1**). This technique was also the trigger for initiating my current research theme.

Using projection-mapping technology, Hengento makes it possible to give an illusion in which a stationary paper object appears to be moving. Hengento projects an image of light and dark calculated to create an illusion that exactly overlaps the target. The mechanism for detecting movement in the human brain works even in the case of low contrast, so the image projected by Hengento may also have low contrast.

Human motion detectors are sensitive to brightness but insensitive to color; therefore, a change in brightness is sufficient for the image projected by Hengento. Since only a faint light-and-dark image is projected, it is possible to give the illusion of movement only to the target without losing almost any of the color or texture of the projected target. This technique was commercialized in collaboration with NTT Communications and Dai Nippon Printing Co., Ltd. and is being used in various places such as supermarkets

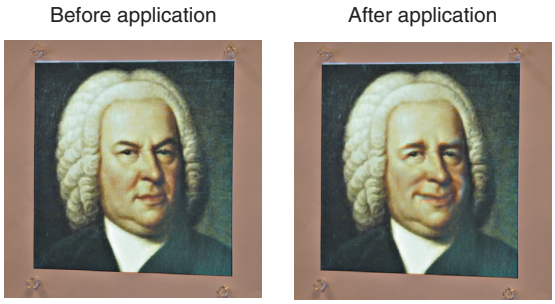



Fig. 1. Hengento.

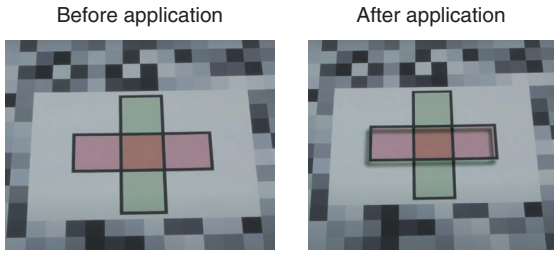



Fig. 2. Ukuzo.

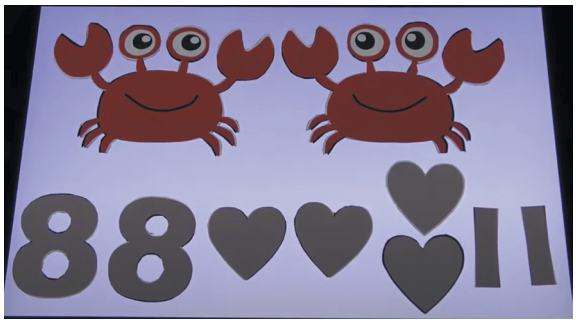


Fig. 3. Danswing Papers.

and museums.

I then developed a technique called Ukuzo (**Fig. 2**) [2]. By projecting a shadow pattern onto a real object, Ukuzo gives an illusion as if the real object is floating in the air. It has long been known that an object appears to float up from its background when a shadow is cast onto it, and this phenomenon is used in applications ranging from computer interfaces, comic books, and animation. By simply capturing an image of the target object with a camera, Ukuzo automatically generates and projects a shadow pattern of the object onto the object and provides an illusory depth impression of a real object.

I also worked on Danswing Papers (*danswing* is a combined word of dance and swing) [3] (**Fig. 3**). It has been known for a long time that if light and dark lines are added to the outline of a stationary object, and the object is presented on a background with temporarily changing brightness, the illusion of the object moving is created. However, it has only been applied to displays. I have reported that adding a light and dark outline to a picture on a paper object and placing the object on a display that switches between light and dark gives the impression of the picture moving. That report was highly evaluated in the Best Illusion of the Year Contest at an international illusion contest in 2018 [4]. I am currently working on the commercialization of this technique.

*—All three are unique and exciting techniques. What are you currently focusing on?*

For about a year and a half, I have been researching illusion techniques that convey visual "softness." Softness is mainly perceived through the sense of touch, and how humans perceive softness only from images is not completely clear.

With this issue in mind, I focused on a physical phenomenon called the Poisson effect. The Poisson effect is a physical phenomenon by which a material contracts vertically when pulled horizontally (or expands vertically when squeezed horizontally). It can be described by a numerical value called Poisson's ratio. For example, Poisson's ratio of rubber is close to 0.5, but that of cork is almost zero. In other words, if rubber is stretched horizontally, it will shrink vertically, but if cork is stretched horizontally, it will hardly shrink vertically. Poisson's ratio is thus a good indicator of such a physical property. It has been shown that Poisson's ratio for everyday materials (excluding metamaterials) does not exceed 0.5.

I am researching how Poisson's ratio contributes to the appearance of softness of objects. First, I investigated how humans perceive Poisson's ratio and found that humans do not feel uncomfortable with an object when its Poisson ratio exceeds 0.5. As a result of repeated investigations, I also found that humans judge discomfort due to the Poisson effect based on the change in the area of an image before and after the image is deformed.

More recently, I showed that the Poisson effect affects the human *pseudo-tactile* sensation. Pseudo-tactile sensation refers to the haptic illusion generated

in a user's mind when a visual image is manipulated. For example, if the cursor suddenly slows while you are moving your computer mouse, you get the impression that it is heavy. The impression is considered a kind of pseudo-tactile sensation. I created an experimental setting that enables an object to extend laterally on a display while a user tries to horizontally extend an imaginary object in the air. I then tried to manipulate Poisson's ratio. As a result, I found that the higher the Poisson's ratio, the softer the object is perceived. I think in the future this technique can be applied in scenes in which the softness of objects in remote areas can be conveyed and controlled.

## The trigger was a movie that I watched on a flight back from an academic conference

*—Could you tell us what inspired you to carry out these studies?*

When I attended an academic conference held overseas in 2018, the idea leading to this technique came to me in a flash when I watched a movie on my return flight. I think the movie was "Iron Man," and in the movie, the main character was operating a three-dimensional hologram. In the real world, holograms are just images, so you can't touch them; even so, I imagined that the hero had the sensation of touching the hologram and was operating it. Therefore, I started my research thinking that the texture of an object could be enhanced by giving it a pseudo-tactile sensation.

When I think back to my childhood, I felt like I could see through my hand when I held my hand over a miniature bulb, and I wondered if it was an extrasensory perception. As I grew older, I gradually became aware of reality, and realized things that do not actually exist may look like as if they existed. That realization might be the original experience that leads to current research activities. Some time after that, I majored in psychology at university and studied vision science. Moreover, as a fan of video games, I felt the impact when I crashed into a wall in the car racing game, so I wondered if it would be possible to transmit information using such a visual illusion.

*—Have you personally changed since becoming a senior distinguished researcher? Please tell us what you have in mind when looking for a research theme.*

It's only been two months since I became a senior distinguished researcher, and I've been working from home to help prevent the spread of the novel coronavirus, so I haven't yet felt a change in my role. I did begin to think about the viewpoint I should have. When I was working at a university 10 years ago, I became determined to contribute to science by becoming a science *building block*. However, I started thinking that I could pursue research differently at NTT than at a university. After starting to conduct research at NTT on the application of visual illusions in the real world and having people use those illusions, I felt that my research can not only contribute to the foundation of science but also to society. As a senior distinguished researcher at NTT, I'd like to be a person who can provide valuable research results to both science and society.

I think that finding research themes is one of the aspects that researchers will struggle with. Even now, I sometimes wonder, "What should I do next?" I remember when I was a graduate student, I was worried about finding a research theme, and I was dressed down by a senior student, who is now a university teacher, saying, "You cannot decide what to do because you are not making enough effort." Certainly, if you don't structurally understand the issues in your research area, you won't know what to research. Thus, the resulting words "You're not trying hard enough." uttered by my senior were apt. To understand the issues in my research area and what I do not know, I try to read four or five papers on themes in that area during my spare time every day.

I also try to imagine things in the manner of the song of the Japanese cartoon character "Doraemon," namely, "That looks fun; I wish I could do it." For example, first, I imagine that it would be great if we could use illusions to do such wonderful things. I then think whether it would be feasible and what state-of-the-art knowledge and technology are needed to achieve this. When I ponder all this, I can lay out specific research processes, and the papers that I've been reading daily will be useful. In this way, I think that proceeding with research from the *bottom up* (namely, understanding what is unknown in a particular research area) while continuing to think from the *top down* (namely, thinking what we can achieve in the world) will make it easier to find issues and research themes.

Certain events that happened at graduate school influenced how I settled on research themes and what approaches I took. When I was worried about my career path, a professor in the laboratory next to mine said, "Mr. Kawabe, you can do as much research as you like, not as much as you *need* to do, but as much

as you *want* to do." After hearing those words, I realized that someone else will probably do what I need to do, so I've been continuing researching what I want to research without losing out to anyone. Those words are the reason my research got off to a good start.

I'm constantly thinking that I shouldn't limit myself to a particular research field. To solve problems in the real world, knowledge of one research field is rarely enough; instead, in most cases, knowledge of multiple research fields is required. In such a case, I will not easily give up solving the problem even if it's not my field; instead, I'll try to acquire knowledge by reading papers in other fields. However, it is also true that it is impossible to surpass experts, so I try to listen actively to people in other fields. I think that the accumulation of knowledge will facilitate discussions and future collaborations.

## Take up challenges with bottom-up and top-down approaches

*—Please say a few words to junior researchers.*

In my case, since I still have more seniors than juniors, I'd like to say what is important to us as I stand in the same shoes as my juniors. I'd like you, our juniors, to keep in mind bottom-up and top-down approaches that I talked about earlier. Taking these two approaches, you will naturally see how you can contribute to your research area as well as to society. I think it's fine to focus on research themes that no one else has tackled if you want to. There is always someone to appreciate your desire if you try. My hope is that many creative and outstanding researchers will be nurtured in doing so.

What's more, I think it's probably not so easy for researchers to link their basic research to direct output. For example, we put Hengento into practical use after receiving feedback from someone who looked at our research results and suggested we could use it in a certain way. I also think that researchers who take a top-down approach and have aspirations can lead to output. In this sense, NTT provides good opportunities such as NTT R&D Forum, so I think it is important for our juniors to actively take those opportunities to obtain the opinions of others. With that in mind, I am actively exhibiting my research results at every opportunity.

*—How will you move forward? Please tell us about your challenges and aspirations.*

My future challenge is to expand the real-world implementation of illusions. Illusions occur not only with regard to vision but also to other senses such as hearing and touch. There are also illusions that occur when multiple senses combine. I think that we can change the scope and depth of technology that uses sensations by incorporating illusions that occur with various sensations into that technology. For example, I believe that we can conduct research from various perspectives under the keyword "softness," namely, expressing softness through a combination of visual and tactile senses, manipulating the softness of real objects by using light projection, and judging softness in virtual space.

Now that I'm at NTT laboratories, and I'm often asked why I'm doing certain research. I feel that my research stance differs from that when I was researching at a university—where I had little contact with the outside world. I have come to think that my research is often questioned because researching at NTT laboratories may contribute to society. I feel that it is important to make not only scientific contributions but also social contributions by using the technologies that I have researched. After the state of emergency was declared in Japan to prevent the spread of the novel coronavirus, I wondered what kind of social contribution I could make to help stop the spread. For example, aiming to promote non-contact technology, thereby prevent infection, we are trying to develop a display that uses illusion to give the user the sensation of touching the buttons of a cash machine, elevator, or the like.

I think it is vital to solve these social problems one by one and anticipate problems that may occur in society. I had been thinking about the idea of a non-contact display even before the coronavirus crisis and was attempting to propose a rich and profound information-presentation technique by using illusions. Such non-contact displays could help prevent the spread of infectious diseases.

I imagine that people might not understand exactly what researchers like me are doing; however, I also believe that people think researchers are trying to improve society by gathering wisdom. I don't know anything other than my work as a researcher, and I've never wanted to be anything else. I'll continue researching what I like by keeping the two approaches of top down and bottom up in mind.

## References


[1] Video of Hengento, https://www.youtube.com/watch?v=OmdSxdSs-w4

[2] T. Kawabe, "Ukuzo—A Projection Mapping Technique to Give Illusory Depth Impressions to Two-dimensional Real Objects," NTT Technical Review, Vol. 16, No. 11, pp. 30–34, 2018. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201811fa4.html

[3] T. Kawabe, "Danswing Papers," exhibition panel of NTT Communication Science Laboratories Open House 2019, http://www.kecl.ntt.co.jp/openhouse/2019/exhibition24/index_en.html

[4] Video of Danswing Papers, Top 10 finalist of the 2018 Best Illusion of the Year Contest, http://illusionoftheyear.com/2018/10/danswing-papers/



■ **Interviewee profile**

**Takahiro Kawabe**

Senior Distinguished Researcher, Sensory Representation Group, Human Information Science Laboratories, NTT Communication Science Laboratories.

He received a Doctor of Psychology from Kyushu University, Fukuoka, in 2005. In 2011, he joined NTT Communication Science Laboratories, where he studies human material recognition and cross-modal perception. He received the 2013 JPA Award for International Contributions to Psychology: Award for Distinguished Early and Middle Career Contributions from the Japanese Psychological Association. He is a member of the Vision Sciences Society and the Vision Society of Japan.

# I Want to Learn More about You: Getting Closer to Humans with AI and Brain Science

*Takeshi Yamada*

**Abstract**

NTT Communication Science Laboratories aims to achieve communication that *reaches the heart* by pursuing innovative technologies that approach and exceed human abilities such as media processing, data analysis, and machine learning as well as studying cognitive neuroscience and brain science for obtaining a deeper understanding of people. It also aims to deliver concrete results to society through collaboration with its business partners. This article introduces the efforts to achieve these aims.

*Keywords: artificial intelligence, communication science, brain science*

## 1. Introduction

In 1985, the Nippon Telegraph and Telephone Public Corporation was privatized, which led to the founding of NTT. Before privatization, each home only had a traditional black telephone, which was rented from the corporation and mainly located in the entrance hall. After privatization, however, one could freely buy a telephone of different colors and with more functions such as cordless reception that enabled having a receiver in each room. The popular love song titled "I want to learn more about you"* focused on an intimate conversation between a girl and her (possibly) boyfriend over the phone with her asking "where are you now and what are you doing" and achieved estimated sales of 391,000 copies. Today, various forms of social media have almost entirely subsumed this role and even been extended to learning about people who are not close friends and basically strangers. A smartphone obtains a great deal of information about its user throughout the day. It is so smart that it may obtain more details about its user than what the user knows about him/herself. On the other hand, the traditional black telephone, while simple and not smart, might be more comfortable to use than current smartphones. As technology continues to develop, how will communication change? It is all the more important to identify the *core of communication* in this era when the physical distance between people must be maintained to prevent the spread of the novel coronavirus while avoiding social disconnection from our close friends and community.

Next year, NTT Communication Science Laboratories (NTT CS Labs), which was founded in Keihanna Science City, Kyoto, in 1991, will mark its 30-year anniversary. Since its founding, it has undertaken fundamental research based not only on the principle of conveying information accurately and efficiently but also on deepening mutual understanding, sharing feelings, and making genuine contact. Even though the research in the beginning focused on person-to-person communication, its aim today is sincere and *heartfelt* communication in both person-to-person and person-to-computer contexts. To this end, we at NTT CS Labs are pursuing innovative technologies for approaching and exceeding human abilities such as media processing, data analysis, and machine learning as well as studying cognitive neuroscience and brain science for obtaining a deeper understanding of people and their cognitive flexibility and diversity

* "I want to learn more about you" was NTT's "TALK ON THE PHONE" campaign song that was released immediately after NTT's privatization in 1985.

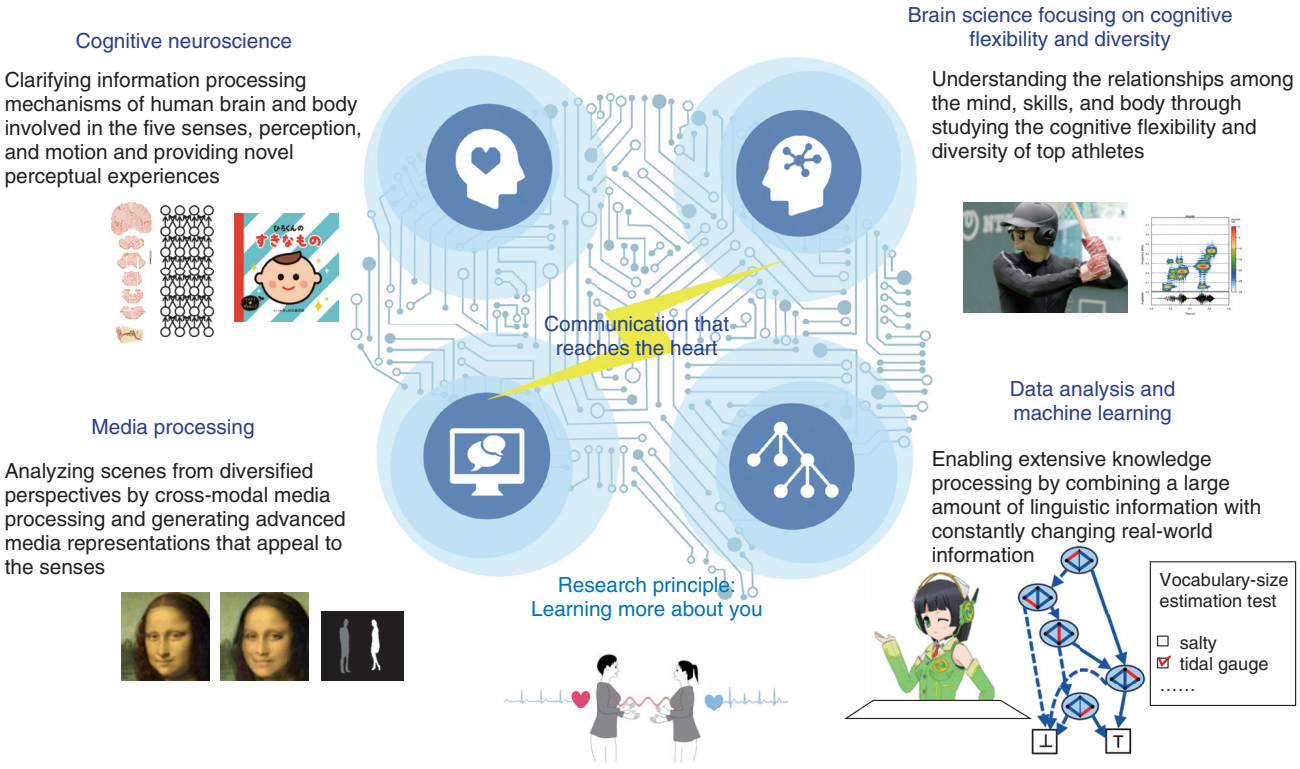



Fig. 1. Research areas of NTT CS Labs.

(**Fig. 1**) [1]. We are also continuing to work on the development of more fundamental mathematical theories [2]. Our primary mission is to carry out these basic research topics with unlimited curiosity to learn more about us humans. We are also delivering concrete results to society through collaboration with our business partners. Several examples of these research endeavors are introduced below.

## 2. Technologies that approach and exceed human abilities

Communication is first and foremost the recognition and understanding of spoken words. On our Illusion Forum website, which provides information on a variety of visual and auditory illusions, noise-vocoded speech (so-called mosaic speech) is introduced [3]. Mosaic speech, which is analogous to visual mosaic images, is a distorted speech sound where the detailed spectral information in the original clean speech sound is manipulated and destroyed. When one listens to it, he or she can somehow understand what is being said, even though the sound is unnatural and hard to hear. In this way, humans can hear the content of speech sounds even if their fine temporal structure is substantially degraded. One may fail to understand the sound at first, but if he or she listens multiple times, or listens to the original sound then listens to the converted one again, he or she will understand it. One has just experienced the depth of the human auditory system and sound recognition ability.

Mammals, including humans, have developed brain functions that recognize such natural sounds as speech and environmental from evolution. In the mammalian brain, the auditory system converts the physical properties of a sound stimulus to neural activities and processes them through a cascade of brain regions. Interestingly, the characteristics of amplitude modulation (AM) tuning in temporal and rate coding are transformed systematically along the processing stages from the periphery to the cortex. The AM rate at which neurons are synchronized gradually decreases and the number of neurons that perform rate coding gradually increases. We have discovered that when a deep neural network (DNN) is trained to classify raw sound data using raw waveforms that are directly input, the trained DNN resembles the

brain's auditory system throughout the entire cascade [4]. Such similarity to the auditory system increases as the classification accuracy improves. These results suggest that AM tuning in auditory systems emerged through evolution as a result of adapting to sound recognition in the real world.

Through evolution, humans have acquired the ability to select the specific voice of a person they want to hear and understand what he or she is saying in a meeting or at a party when several people are talking or when music is playing in the background. This skill is called *selective listening*. We have applied a proprietary DNN and developed technology called SpeakerBeam that enables computers to perform selective listening [5]. However, extracting a target voice is difficult if speakers with similar voices are included in the data. Therefore, in addition to voice, lip movements are used as features to distinguish among speakers even with similar voices. We are also developing DNN-based voice-conversion technology that makes it possible to freely change such voice features as voice quality and intonation while preserving the speech's linguistic content. Further development of these technologies will enable natural communication that overcomes disabilities or age-related decline in speaking or hearing functions and support conversations in foreign languages [6].

We are also investigating the language-acquisition mechanism in children, who basically acquire language by talking with their parents. Language and language-based oral communication are basic human functions that have evolved over the past 150,000 years. Since written language emerged relatively recently, only 50,000 years ago, the ability to read is not an inherent function of the human brain, which did not evolve for reading. Reading is achieved through a flexible combination of basic brain functions such as vision, audition, language, and cognition, which is called *neural recycling* [7].

To understand the language-acquisition mechanism, we constructed the Child Vocabulary Development Database by conducting a large-scale survey of what words children can understand and say at what particular development times and modeling the results [8]. With this database, we created personalized educational picture books in collaboration with NTT Printing Corporation for encouraging children to read. The books' contents were customized to the vocabulary development of each child. In cooperation with Onna Village in Okinawa and Tokushima City, we worked with NTT Printing Corporation to deliver these books to children and to encourage them to read from an early age [9, 10].

As an approach to understanding the advanced human abilities of both language and knowledge processing, we are participating in the artificial intelligence (AI) project "Todai Robot Project—Can a robot get into the University of Tokyo?" led by the National Institute of Informatics (NII). The project is researching the extent to which AI can solve the same problems that humans can solve. A team made up of members of the NTT CS Labs. and other project members took up the challenge of developing an AI system that can take and pass the English written exam administered by the National Center for University Entrance Examinations. The AI system achieved an extremely high score of 185 out of 200 points (64.1 T-score) in the exam's 2019 version [11]. English exams contain problems that require integrating both natural language processing and knowledge processing to solve. We exploit the knowledge gained in tackling these problems for our conversational AI research, which involves chatting with people (in chat-oriented dialogue systems) and providing information and guidance (in task-oriented dialogue systems) to achieve more natural and mutually understandable conversations between AI and humans.

Based on research on conversational AI, NTT is developing a role-play-based conversational AI called Narikiri AI, which reproduces the behaviors of a celebrity or a character in a novel or a game. Narikiri AI's dialogue data are collected through sets of questions and answers posted by online users, where one user asks a certain character a question, and another user mimics its personality and answers it. A few such Narikiri AIs have been constructed by NTT in cooperation with NTT DOCOMO and DWANGO Co., Ltd. Recently, we have begun a new Narikiri AI project in cooperation with Seika Town, Kyoto, involving its official public-relations character (mascot) Kyomachi Seika; thus, the project is called Narikiri AI Kyomachi Seika [12]. The concept of Kyomachi Seika is that of "a young agent is dispatched from the future to protect the world from a time paradox" [13]. It has attracted many fans through social networking services and is a perfect match with Narikiri AI.

We are conducting basic research on technologies to achieve natural human-like conversations by smoothly switching between chat-oriented and task-oriented dialogues. To demonstrate this technology, we collected dialogue data, including a large amount of knowledge and experience about Seika Town, through cooperation with its residents. We will build

a dialogue system that provides appropriate information and guidance regarding tourism and town administration, understand user's intensions, and flexibly answers user questions and requests through Kyomachi Seika.

## 3. Studying cognitive neuroscience and brain science for a deeper understanding of people

AI development is also intensifying the importance of obtaining a deeper understanding of people. For example, when searching on the Internet, advertisements suddenly appear that match the search words. Sometimes users click on such links and impulsively make purchases. Many users might argue that they purchased the product completely voluntarily, downplaying any third-party manipulation. As AI technology expands, the risk increases for such manipulation, which resembles an AI version of subliminal effects.

To avoid such risks, it is important to obtain a deeper understanding of the preconceptions held by people and how and when they will behave in a certain way. We are conducting research to clarify the information-processing mechanisms of the human brain and body involving the five senses, perception, and motion and providing novel perceptual experiences [14, 15]. We are also carrying out brain science research to understand the relationships among the mind, skills, and body of top athletes through studying their cognitive flexibility and diversity such as how they obtain and judge information from the outside world. Regarding baseball, for example, we are exploring the differences between great and mediocre hitters by investigating how well good hitters actually see the ball or whether a fastball really travels in a straight path. The plan is to use the knowledge gained from our research to provide feedback to athletes to sharpen their brain functions.

As societies worldwide are forced to coexist with the novel coronavirus, establishing profound contact has become more complicated. The economist and philosopher Adam Smith wrote that "we often derive sorrow from the sorrow of others" in his discussion of the importance of sympathy [16]. Sharing sentiment and emotions with others and feeling their experiences as one's own is the essence of sympathy and leads to genuine interaction. In collaboration with Tatsuya Kameda Laboratory at the University of Tokyo, we are studying how pain is shared during face-to-face interactions [17]. In one of the experiments carried out through this collaboration, a pair of strangers (participants) were exposed to identical pain-provoking (thermal) stimuli. The blood-volume pulse of both participants is recorded to measure their acute sympathetic responses while they both simultaneously experienced the stimuli. Under a face-to-face condition, participants with weaker reactions elevated their physiological reactivity to the stimulus based on their partner's reactions but not under a shielded, non-face-to-face condition. These results suggest that during face-to-face interactions, sharing such negative emotions as pain occurs at the unconscious physiological level. Another experiment of a two-player eSports competition showed that the heart rates of both players were synchronized in high-level close battles but not in intermediate one-sided battles.

## 4. Future prospects

The Japanese word for "happiness" is "shi-awase," where "shi" means "action" and "awase" means "match." This combination of meanings suggests that one way to achieve individual happiness is through successful interaction or communication with another person [18]. In other words, the synergistic effect of sincere and *heartfelt* contact improves human happiness, or in more modern terms, the well-being of people. In Japanese, "shi-ai" denotes games such as baseball or tennis, and the heart-rate synchronization in the eSports experiment seems to exhibit such a "shi-awase" or "action-matching" state. The mission of NTT CS Labs is to pursue sincere and *heartfelt* communication that literally achieves the following lyric from "I want to learn more about you"*: "Even if I'm far away, my heart is with you."

## References


[1] T. Yamada, "Processing Like People, Understanding People, Helping People—Toward a Future Where Humans and AI Will Coexist and Co-create," NTT Technical Review, Vol. 17, No. 11, pp. 6–11, Nov. 2019.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201911fa1.html

[2] G. Kato, "Quantum Information Processing via Indirect Quantum Control," NTT Technical Review, Vol. 18, No. 11, pp. 32–35, Nov. 2020.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202011fa5.html

[3] Website of Illusion Forum by NTT CS Labs, Noise Vocoded Speech (in Japanese), http://www.kecl.ntt.co.jp/IllusionForum/a/noise_vocodedSpeech/ja/index.html

[4] Press release issued by NTT, "A Deep Neural Network Trained for Sound Recognition Acquires Sound Representation Similar to that in Mammalian Brains," July 10, 2019.
https://www.ntt.co.jp/news2019/1907e/190710a.html

[5] M. Delcroix, K. Zmolikova, K. Kinoshita, S. Araki, A. Ogawa, and T. Nakatani, "SpeakerBeam: A New Deep Learning Technology for

Extracting Speech of a Target Speaker Based on the Speaker's Voice Characteristics," NTT Technical Review, Vol. 16, No. 11, pp. 19–24, Nov. 2018.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201811fa2.html
[6] K. Tanaka, T. Kaneko, N. Hojo, and H. Kameoka, "Communication with Desired Voice," NTT Technical Review, Vol. 18, No. 11, pp. 27–31, Nov. 2020.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202011fa4.html
[7] S. Dehaene, "Reading in the Brain: The New Science of How We Read," Penguin Putnam Inc, 2010.
[8] S. Fujita, "Measuring Textual Difficulty—Estimating Text Readability and a Person's Vocabulary Size," NTT Technical Review, Vol. 18, No. 11, pp. 36–42, Nov. 2020.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202011fa6.html
[9] Press release issued by NTT Printing, "A Three-party Experiment Using a Personalized Educational Picture Book Has Started!", Feb. 6, 2020 (in Japanese).
https://www.nttprint.com/company/itemid419-000048.html
[10] Press release issued by NTT Printing, "Conclusion of the Joint Experiment Agreement on Supporting Picture Book Reading Activities at Home Using Personalized Educational Picture Books with Tokushima City and NTT," July 2, 2020 (in Japanese).
https://www.nttprint.com/company/itemid419-000053.html
[11] Press release issued by NTT, "AI Achieved a Score of 185 on the English Written Exam of the National Center Test for University Admissions in 2019," Nov. 18, 2019.
https://www.ntt.co.jp/news2019/1911e/191118a.html
[12] Press release issued by NTT, "Starting a Joint Experiment with Seika Town on a Conversational AI System," July 3, 2020 (in Japanese).
https://www.ntt.co.jp/news2020/2007/200703a.html
[13] What is Kyomachi Seika (in Japanese). https://kyomachi-seika.jp/profile/
[14] S. Kuroki, "Towards Understanding Human Skin Sensations," NTT Technical Review, Vol. 18, No. 11, pp. 16–20, Nov. 2020.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202011fa2.html
[15] S. Itoh, "Brain-information Processing for Quick and Stable Human Movements—Stretch-reflex Regulation Based on Visually Updated Body Representation," NTT Technical Review, Vol. 18, No. 11, pp. 21–26, Nov. 2020.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202011fa3.html
[16] A. Smith, "Theory of Moral Sentiments," 2 ed., Strand & Edinburgh: A. Millar; A. Kincaid & J. Bell, 1761.
[17] A. Murata, H. Nishida, K. Watanabe, and T. Kameda, "Convergence of Physiological Responses to Pain during Face-to-face Interaction," Scientific Reports, Vol. 10, 450, Jan. 2020.
[18] S. Genyu, "Zen Happiness Theory," KADOKAWA SSC Books, 2010 (in Japanese).

**Takeshi Yamada**
Vice President and Head of NTT Communication Science Laboratories.
He received a B.S. in mathematics from the University of Tokyo in 1988 and a Ph.D. in informatics from Kyoto University in 2003. He joined NTT Electrical Communication Laboratories in 1988. He was a visiting researcher at the School of Mathematical and Information Sciences, Coventry University, UK, from 1996 to 1997. He was a group leader of the Emergent Learning and Systems Research Group from 2006 to 2009 and an executive manager of Innovative Communication Laboratory from 2012 to 2013 at NTT Communication Science Laboratories. His research interests include data mining, statistical machine learning, graph visualization, metaheuristics, and combinatorial optimization. He is a Fellow of the Institute of Electronics, Information and Communication Engineers, senior member of the Institute of Electrical and Electronics Engineers, and a member of the Association for Computing Machinery and the Information Processing Society of Japan.

# Towards Understanding Human Skin Sensations

*Scinob Kuroki*

**Abstract**

Human fingertips are amazingly sensitive. They can easily detect a single strand of hair that has fallen onto a desktop and distinguish between metal surfaces that have been polished to different degrees. Our group uses perceptual psychology methods to study tactile sensory mechanisms. This article introduces tactile illusions that illustrate how the tactile sensory system processes information and provides stable perception. This line of research is expected to yield efficient information-presentation techniques.

*Keywords: tactile, haptics, perception*

## 1. Tactile × perceptual psychology

The hand is a tool and sensor. The manual dexterity of humans has contributed to the remarkable progress of our species. To skillfully use tools, it helps to be able to measure and process sensory information quickly and correctly. We (myself and other researchers at NTT Communication Science Laboratories) are interested in this processing of information through the skin. Tactile references often appear in metaphors that allude to a deep understanding of things. For example, someone with the "common touch" understands how ordinary people think, someone with the "magic touch" is exceptionally skilled, and someone with the ability to retrieve information instantly is said to have it "at their fingertips." However, surprisingly little research has been done on the information-processing aspects of how tactile sensations actually work, and the mechanisms with which they are processed are not well understood. Since tactile sensations originate from the boundary between oneself and the outside world, it has also been technically difficult to conduct studies such as directly measuring or artificially reproducing phenomena that take place when touching a surface. With recent advances in technology, progress is gradually being made in the clarification of tactile sensing, which is catching up with research on visual and auditory sensing.

Perceptual psychology is a field of study in which humans are regarded as systems and which seeks to identify their internal processes. It provides a means of determining what types of perception a human observer obtains as output when provided with input stimuli such as images or sounds. For example, people can perceive changes in brightness when a light is slowly switched on and off, but as the switching-speed increases, the flickering becomes less visible. At higher speeds (like the frequency of fluorescent lights), this flickering becomes imperceptible. By making small changes to input stimuli in this manner, we can examine the ranges of various physical quantities that affect our perception.

This article introduces two illusions that offer inroads into reasoning about the mechanisms and characteristics of tactile perception. These illusions are phenomena in which, for example, identical physical inputs are perceived as being different or different physical inputs are perceived as being the same. These illusions allow us to investigate what type of information the brain is sensitive and not sensitive to. They are also useful for considering the means in which information can be presented more effectively.

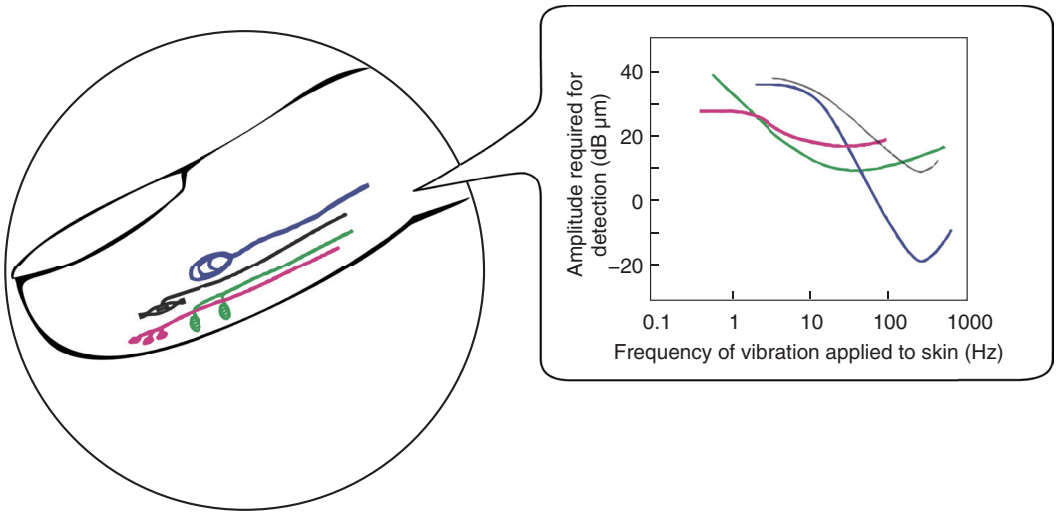



Fig. 1. Response characteristics of mechanoreceptors and the nerves.

## 2. The illusion of mixed frequency from two vibrations

When you move your fingers across the surface of an object, it causes vibrations in the skin of your fingertips. These vibrations allow you to distinguish between different textures, such as rough and smooth surfaces. Our perception of these surfaces is based on vibration frequencies as encoded by tactile sensors, i.e., mechanoreceptors, in the skin. The skin includes different types of sensors in much the same manner that the retina includes different sensors that are sensitive to different colors. Each sensor has a different shape and distribution. Some are small and concentrated at shallow parts of the skin surface, while others are large and scattered deep within the skin. As a result, each type has its own idiosyncrasies. For example, some sensors respond to slow deformations, while others respond to high-frequency vibrations such as received vibrations (**Fig. 1**). How are frequencies encoded using the signals of these different sensors?

The difference in frequency sensitivities across sensors is observed not only in tactile sensing but also in visual and auditory sensing. For example, when we see red light (which has a long wavelength) and green light (which has a slightly shorter wavelength) at the same time, we perceive yellow light (which has a wavelength between those of red and green). It is not possible to perceive the red and green lights separately. This is described by the well-known theory of the three primary colors. However, when we hear the musical notes C and E played at the same time, we recognize them as two separate notes forming a major third, instead of hearing the note D that lies between them. Which of these processes is exhibited by the tactile system?

We conducted a series of experiments to ascertain whether the tactile system detects frequency in a manner analogous to the three primary colors by testing whether an intermediate frequency is perceived when multiple vibrations with different frequencies are presented [1]. If the skin is simply subjected to a synthetic vibration obtained by adding two frequencies together, this vibration may be physically attenuated by the time it reaches the sensor due to the elastic nature of the skin. We therefore simulated a situation of seeing red and green at the same time by applying two vibrations of different frequencies to two different fingers (the index and middle fingers) at the same time. In these experiments, we applied synthetic stimuli consisting of two sinusoidal vibrations at two different frequencies on two different fingers of each participant and compared the perceived frequency with comparative stimuli consisting of two sinusoidal vibrations of the same frequency applied to the two fingers (**Fig. 2**). (See [1] for details about this method.) We found that the perceived frequency of the synthetic stimuli was an intermediate frequency between the two presented frequencies of the synthetic stimuli. Based on this result, it appears that the perception of vibrations is similar to the perception of

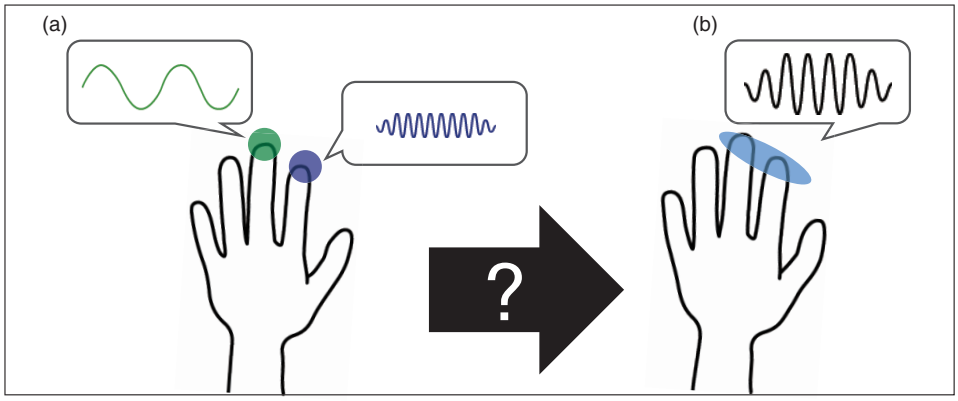



Fig. 2. The illusion of mixed frequency from two vibrations.

the three primary colors of light. We also found that this phenomenon occurred not only when the synthetic stimuli were applied to two fingers on the same hand but also when applied to fingers on the left hand and fingers on the right hand. Even when vibrations with completely different frequencies were applied to clearly different parts, i.e., fingers of the left and right hands, the participants seemed to perceive as if the two vibrations had the same frequency.

However, we also found that the synthetic stimuli do not mix completely. When touching synthetic stimuli with two fingers, it is possible to feel differences (e.g. in vibration strength) between the two fingers. Also, the synthetic stimuli feel noisier than simple sinusoidal vibration at a matched frequency. Unlike in vision, there is no guarantee of perceptual equivalence between synthetic stimuli and matched vibrations in touch.

In summary, when you experience multiple vibrations at different frequencies via the skin, it is possible to perceive multiple vibrations, which shows that the encoding of tactile vibrations differs from that of visual colors. It also differs from the encoding of auditory sounds because the tactile system cannot encode the two frequencies separately. It seems that tactile-frequency information is processed through a strategy that lies somewhere between the strategies of sight and hearing.

## 3. The illusion of textures that look different but feel similar

Everyday objects have a variety of surface irregularities and textures. Since it had been technically difficult to artificially design tangible textures exactly as intended, tactile texture research has been conducted either by creating a simple experimental stimulus or using existing materials without alteration. To bridge the gap between simple experimental stimuli and complex real-world textures, we use laser cutters and three-dimensional (3D) printers to create surface textures with complex spatial patterns [2].

For texture design, we used images as height maps (**Fig. 3(a)**). Lighter areas correspond to bumps and darker areas correspond to hollows. Using images like these, we can easily adjust statistics, such as the spatial frequency and bandwidth, of surface irregularities. In this experiment, each image was 3D printed as a 4 × 4 cm sample piece with a maximum engraving depth of 2 mm to create a set of surface textures. The five images shown in Fig. 3(a) have central frequencies represented numerically (similar to grit numbers on sandpaper) and are modulated with relatively simple statistical quantities. In the 3D-printed sample stimuli based on these images, the central frequency increased from left to right, resulting in surfaces ranging from bumpy to grainy. With these stimuli, texture discrimination is quite easy. Humans are very sensitive to central frequency modulation. Therefore, how well can human observers discriminate between 3D-printed stimuli based on completely different images? Five natural scene images obtained from an image database are shown in **Fig. 3(b)**. Although the brightness values of these images were matched to have the same average and variance, there are differences in the spatial arrangement of black and white that are clearly visible when viewed with the eye. However, it was not easy to distinguish between the corresponding 3D-printed textures by touching with the hands. It was particularly

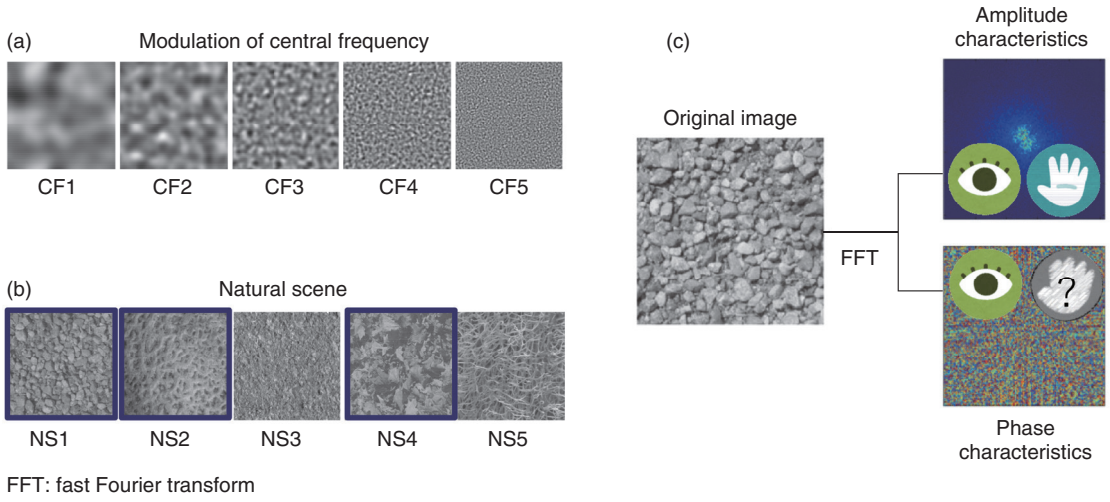



Fig. 3. The illusion of textures that look different but feel the same.

impossible to distinguish between the stone, coral, and leaf textures shown in Fig. 3(b) (the images outlined in dark blue), even though they looked completely different. This illusion shows that the tactile and visual systems use different mechanisms to process spatial textures.

How does information processing differ between the eyes and hands? To answer this, we analyzed the original images used as templates for the 3D-printed surfaces. A characteristic of natural images is that the amplitude decreases at higher spatial frequencies. We calculated the Fourier transform of each of the five images shown in Fig. 3(b) and compared the amplitude and phase spectrum of each transform (amplitude indicates the intensity of the light and shade and phase conveys information about the distribution of light and shade). Although the amplitude spectra were very similar for all five images, their phase spectra were completely different. The visual system processes both the amplitude and (some of) the phase characteristics of an image after Fourier decomposition. It seems that the tactile system mostly processes the amplitude characteristics without using much of the phase characteristics (**Fig. 3(c)**). This finding can be exploited to freely create visually dissimilar textures that feel similar when touched. The calculation of phase is computationally more expensive than the calculation of amplitude. Our tentative understanding is that the tactile system does not expend many resources on processing phase information and may instead have evolved to be sensitive to small differences in amplitude characteristics.

As mentioned above, we are using perceptual psychology methods to study tactile sensory mechanisms. Originally, most research on the human senses focused on vision and audition, and few considered the tactile sense. However, the number of tactile studies has recently increased due to advances in the tools that can be used for conducting research. Although the five senses involve many similar processes, the fact that these sensory organs evolved in different forms suggests that they may be processing information in different ways. In other words, visual information is processed by visual rules, auditory information by auditory rules, and tactile information by tactile rules. We hope that by accumulating an understanding of these rules of perception, it will be possible to gain an understanding of the tactile sense while at the same time increasing the possibility of multimodal information presentation.

## References

[1] S. Kuroki, J. Watanabe, and S. Nishida, "Integration of Vibrotactile Frequency Information beyond the Mechanoreceptor Channel and Somatotopy," Sci. Rep., Vol. 7, 2758, 2017.
[2] S. Kuroki, M. Sawayama, and S. Nishida, "Haptic Metameric Textures," bioRxiv, 2019. doi: https://doi.org/10.1101/653550.

**Scinob Kuroki**
Research Scientist, Sensory Representation Research Group, Human Information Science Laboratory, NTT Communication Science Laboratories.
She received a Ph.D. in information science and technology from the University of Tokyo in 2011. Her research is focused on human tactile processing, particularly frequency perception, time perception, and motion perception.

# Brain-information Processing for Quick and Stable Human Movements—Stretch-reflex Regulation Based on Visually Updated Body Representation

*Sho Ito and Hiroaki Gomi*

**Abstract**

From the extraordinary performances of athletes to our most common everyday actions, human body movements are heavily governed by unconscious and automatic information processing of sensory-motor flow, some of which is known as a reflex. Our resent research has demonstrated that the brain flexibly regulates stretch reflexes, using body information shaped by integrating multiple sensory cues, including vision. This finding suggests that there is sophisticated brain processing underlying reflex control, in contrast to the classical view of reflexes as simple and primitive.

*Keywords: stretch reflex, motor control, multisensory integration*

## 1. Introduction

When we move our bodies, various sensory-motor processes are unconsciously executed in our brain. Reflex is one of these mechanisms, which evokes actions without conscious thought in response to stimuli such as visual or proprioceptive* inputs. Since a reflex can evoke quicker action than voluntary movements (movements performed via conscious thought), it enables an instantaneous reaction in response to an opponents' movement, for example, during competitive sports. Moreover, a wide range of actions in our everyday life cannot be performed without the support of unconscious motor control mediated by the reflex system.

NTT Communication Science Laboratories has investigated how the reflex system is regulated in our brain and contributes to our action execution [1–3]. This article briefly explains the function of the stretch reflex (a reflex driven by proprioceptive inputs) then reviews our recent study [4] regarding information processing for generating the stretch reflex.

## 2. Mechanisms and functions of the stretch reflex

The stretch reflex is involuntary muscle contraction induced by a sudden passive muscle extension. This sensory-motor mechanism plays an important role in maintaining posture stability (**Fig. 1**). For instance, a sudden change in arm posture can be caused by unexpected contact with an external object. Muscle spindles, receptors that sense changes in muscle length, detect the sudden stretch of the muscle occurring with the postural change and send that signal to the central nervous system. This signal then causes contraction of the stretched muscles after it is instantaneously

* Proprioception: The sensation of body position or of movements that originates from a group of mechanosensory receptors called proprioceptors located in muscles, tendons, and joints. Due to proprioception, we can perceive our own body posture or self-movements even without visual information.

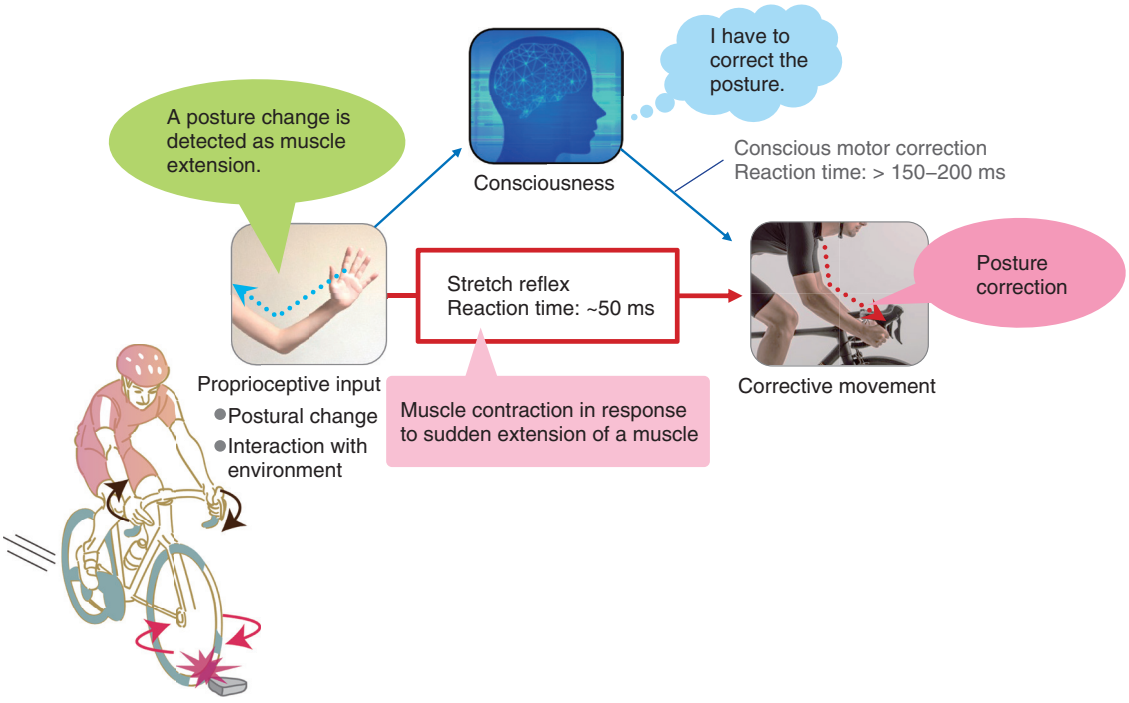



Fig. 1. Function of the stretch reflex.

processed by specific brain areas. There are two responses of the stretch reflex: (1) short-latency stretch reflex generated in a spinal circuit and (2) long-latency stretch reflex generated through broader neural circuits including a sensory-motor area in the cerebral cortex [2]. Importantly, even the long-latency stretch reflex generates muscle contraction in quite a short time of 50 ms from the stimulus onset, which is shorter than the response latency of voluntary reactions (100–150 ms). Accordingly, stretch reflexes enable us to react and compensate for postural changes more quickly than the fastest voluntary movements.

Previous research has shown that stretch reflex responses are not constant but can be modulated in response to changes in the tasks or environment [3, 5]. Thanks to modulation of the reflex system, the brain can flexibly adapt motor control to various contexts. However, the details of the information processing underlying this functional reflex modulation remains unclear. For instance, it is unknown whether adjustments of the stretch reflex depend on only proprioceptive information or body states obtained by combining multiple sensory information, including vision. To examine this point, we investigated if visual information affects stretch reflexes by manipulating visual cues through a series of experiments.

## 3. Contributions of multisensory integration to reflex control

There are two possible hypotheses accounting for the information process underlying the regulation of the stretch reflex (**Fig. 2**). The first hypothesis is that the regulation of the reflex simply depends on unimodal information. Since the stretch reflex is triggered by proprioceptive inputs, this hypothesis proposes that proprioception alone is the source for stretch-reflex regulation. The second hypothesis, on the other hand, proposes that stretch-reflex modulation depends on the body information provided by multiple sensory sources, including vision. This assumes that the brain uses multisensory integration to obtain reliable estimates of the current body state, even for reflex control. To dissociate these two possibilities, we recently conducted experiments in which we manipulated a visual cue representing body posture and tested if this visual manipulation affected the stretch reflex [4].

Using the setup shown in **Fig. 3(a)**, we asked participants to perform reaching movements toward a visual target by flexing the wrist (**Fig. 3(b)**). In some

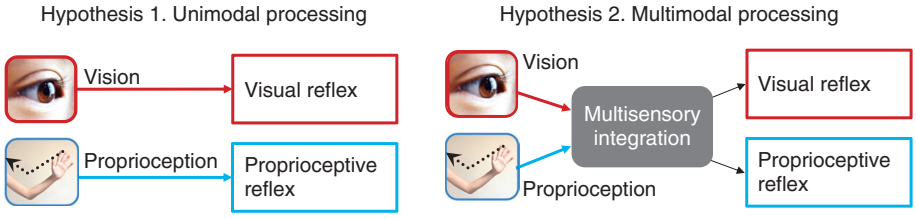



Fig. 2. Hypotheses of sensory process for reflex control.

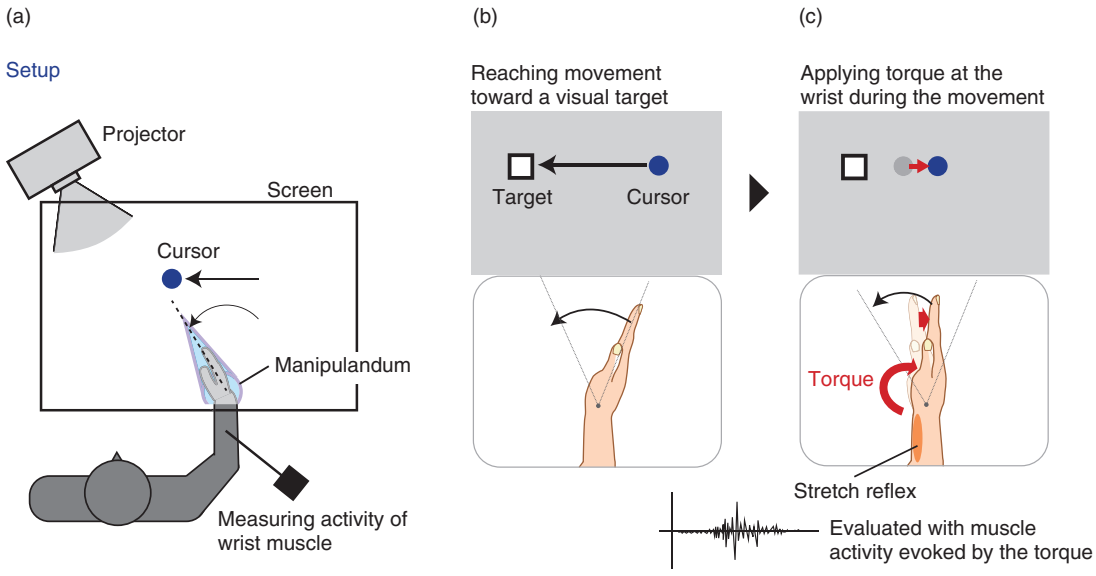



Fig. 3. Experimental setup.

trials, a manipulandum (a robotic device for measuring hand movements) applied an instantaneous mechanical torque to extend the wrist (**Fig. 3(c)**). This sudden and rapid wrist extension evokes stretch-reflex response in the wrist extensor muscle. We evaluated the size of the stretch-reflex response by measuring muscle activities with electrodes attached to the skin. As an experimental manipulation, we introduced a mismatch between visual information and actual body motion by giving rotation in the direction of cursor movements (**Fig. 4(a)**). Notably, the displayed location of the target and the starting point were also rotated so that actual hand movements were identical regardless of the visual rotation. We compared the amplitude of the stretch-reflex response under different visual rotation angles to investigate the effect of visual information on stretch-reflex regulation. As shown in **Fig. 4(b)**, we found a decrease in the amplitude of the long-latency stretch reflex as the visual rotation angle increased, which clearly demonstrates the impact of visual manipulation on the stretch-reflex response. This result supports the hypothesis that multiple sensory information, including vision, contributes to stretch-reflex regulation. Accordingly, even for reflex control, our brain seems to use reliable body-state estimates obtained by multisensory integration, as it does for voluntary motor control.

## 4. Effect of uncertainty in estimating body states on stretch-reflex control

The first experiment showed that rotation of visual feedback reduces the amplitude of the stretch-reflex response. One possible reason of this modulation is that uncertainty in estimating body states, due to

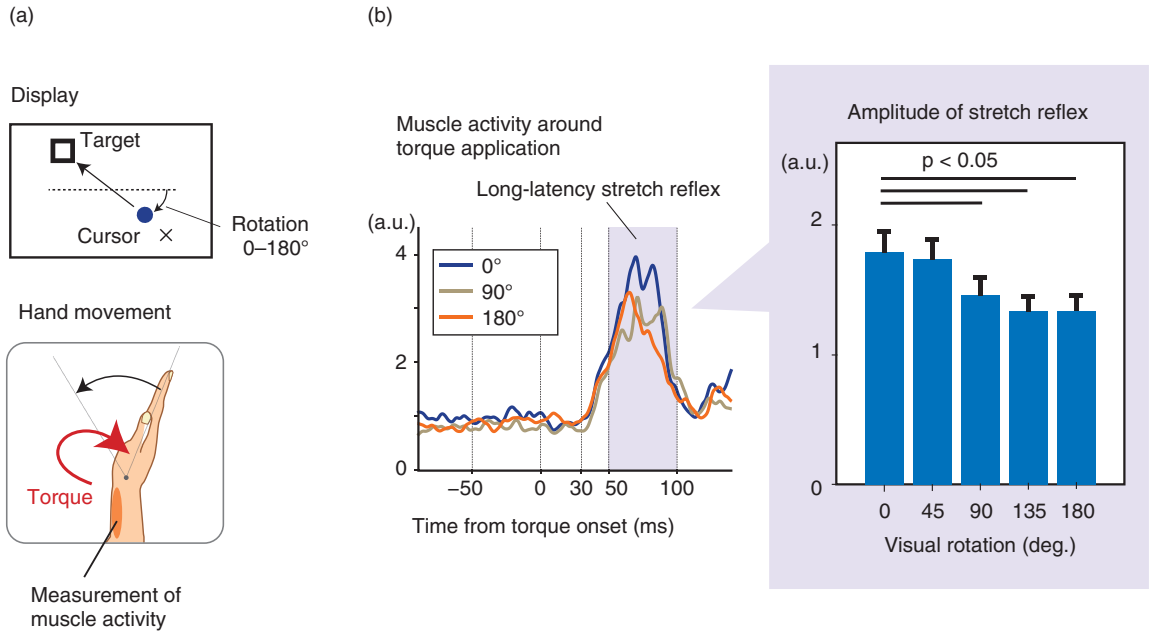



Fig. 4. Effect of inconsistent visual feedback on stretch reflex.

visual rotation, led to the attenuation of responses. To examine this hypothesis, we eliminated the visual cursor representing hand position during wrist movements (**Fig. 5(a)**) in the second experiment. By varying locations to eliminate the cursor, we controlled the presentation time of the cursor at four levels to add uncertainty in visual information. We found that movement endpoints became more variable as the presentation time decreased (**Fig. 5(b)**). This increase in movement variability indicates that the elimination of the cursor increased uncertainty in body-state estimates during the hand movement. We also found a decrease in amplitude of the long-latency stretch reflex as the presentation time of the cursor shortened (**Fig. 5(c)**). This result suggests that the stretch reflex is down-regulated depending on the uncertainty of the body state during movements. An interpretation of this result is that the brain reduces the risk of generating a large response in an incorrect direction, due to erroneous body-state estimation, by attenuating reflex response when the sensory information is not reliable.

## 5. Summary and future direction

Our recent study [4] revealed that visual information relating to the body state contributes to the regulation of the stretch reflex. Our experiments showed a decrease in amplitude of the stretch-reflex response under the conditions of (1) a mismatch between visual feedback and actual movement and (2) elimination of visual feedback. These results indicate that the brain uses body representation containing visual information to regulate the stretch reflex. By changing the presentation time of the visual cursor, we also found a negative relationship between the size of the reflex response and variability of movement endpoints. This is interpreted as occurring because the stretch reflex is regulated depending on the uncertainty of body-state estimates during movement. These findings suggest that more complex brain computation underlie stretch-reflex regulation than has conventionally been thought. In the future, we aim to further elucidate the information processing in the brain that regulates the reflex system. Particularly, our interests are to reveal the neural basis of reflex modulation and build a theoretical framework to account for the functional regulation of reflexes. We expect that the motor-control-related insights obtained through this reflex study will also be beneficial to a range of applications involving human movements such as analyzing the performance of athletes or developing effective methods for physical training. In addition, our study demonstrated the essential role of multisensory integration in movement execution. Understanding how the brain uses multiple forms of

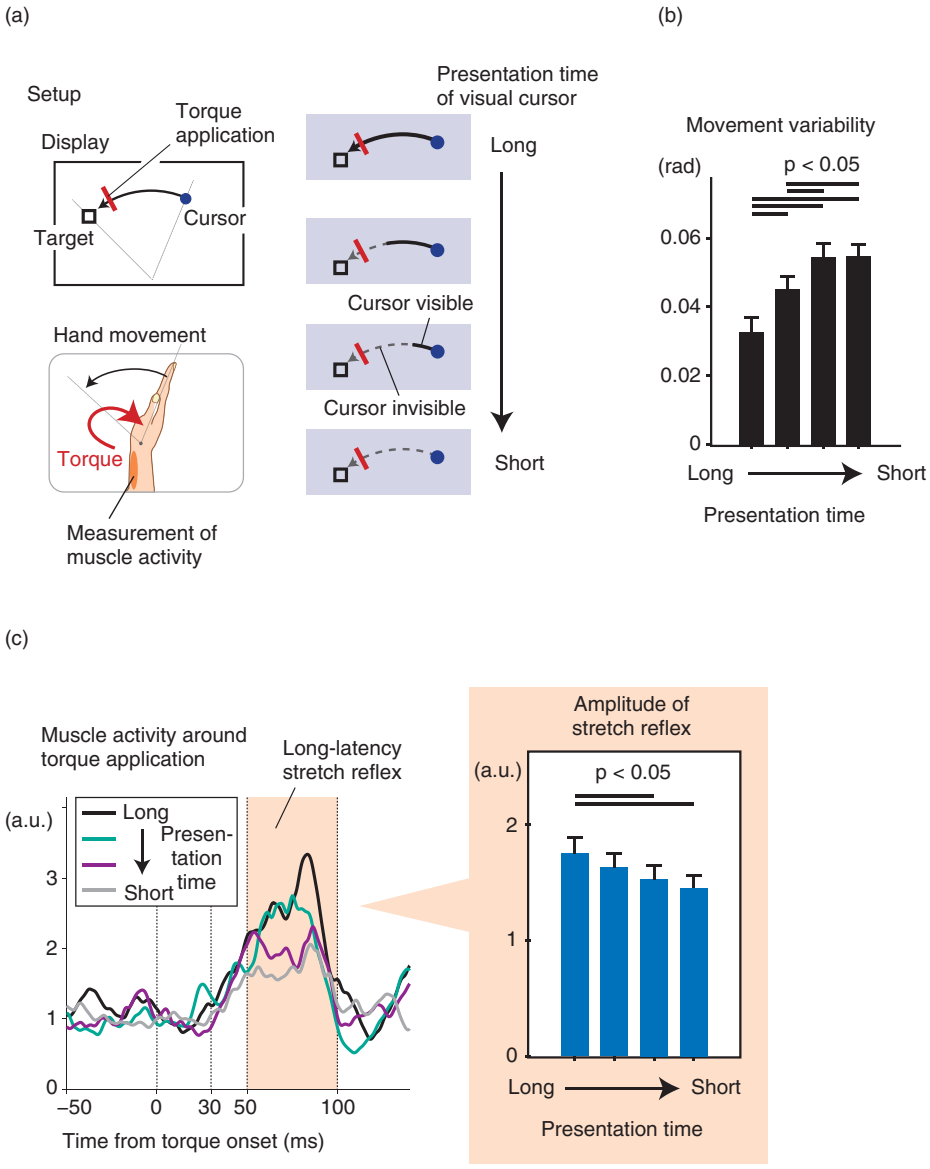



Fig. 5. Effect of uncertainty in body states on stretch reflex.

sensory information in action is important to designing human-machine interfaces accompanying body movements. We will accumulate further knowledge on sensory-motor processing for practical application, including the development of virtual-reality systems.

## Acknowledgments

This work was supported by Grants-in-Aid for Scientific Research (JP16H06566) from Japan Society for the Promotion of Science.

## References

[1] N. Saijo, I. Murakami, S. Nishida, and H. Gomi, "Large-field Visual Motion Directly Induces an Involuntary Rapid Manual Following Response," J. Neurosci., Vol. 25, No. 20, pp. 4941–4951, 2005.
[2] T. Kimura, P. Haggard, and H. Gomi, "Transcranial Magnetic Stimulation over Sensorimotor Cortex Disrupts Anticipatory Reflex Gain Modulation for Skilled Action," J. Neurosci., Vol. 26, No. 36, pp. 9272–9281, 2006.
[3] T. Kimura and H. Gomi, "Temporal Development of Anticipatory Reflex Modulation to Dynamical Interactions during Arm Movement," J. Neurophysiol., Vol. 102, No. 4, pp. 2220–2231, 2009.
[4] S. Ito and H. Gomi, "Visually-updated Hand State Estimates Modulate the Proprioceptive Reflex Independently of Motor Task Requirements," Elife, Vol. 9, e52380, 2020.

[5] J. A. Pruszynski and S. H. Scott, “Optimal Feedback Control and the Long-latency Stretch Response,” Exp. Brain Res., Vol. 218, No. 3, pp. 341–359, 2012.

**Sho Ito**

Research Scientist, Sensory and Motor Research Group, Human Information Science Laboratory, NTT Communication Science Laboratories.

He received a B.S. and M.S. in engineering from Kyoto University in 2008 and 2010. He joined NTT Communication Science Laboratories in 2010, where he studies human motor control. He is a member of the Japan Neuroscience Society.

**Hiroaki Gomi**

Senior Distinguished Research Scientist, Leader of Sensory and Motor Research Group, NTT Communication Science Laboratories.

He received a B.E., M.E., and Ph.D. in mechanical engineering from Waseda University, Tokyo, in 1986, 1988, and 1994. He was involved in biological motor control research at ATR (Advanced Telecommunication Research Labs., Kyoto) from 1989 to 1994, where he developed computational models of human motor control, novel robot learning mechanisms (demonstration learning), and a manipulandum system for investigating human arm movement. He was an adjunct lecturer of Waseda Univ. (1995–2001) and adjunct associate professor (2000–2003) and adjunct professor (2003–2004) of Tokyo Institute of Technology. He was also involved in the CREST (1996–2003, 2010–2015) and ERATO (2005–2010) projects of Japan Science and Technology Agency. He served as a committee member of the neuro-computing technical group of the Institute of Electronics, Information and Communication Engineers (IEICE) (1997–2000), as its vice chair (2006) and chair (2007), committee member of Japanese Neural Network Society (JNNS) (2012–2018, 2020), and chair of ‘Brain and Mind Mechanism workshop’ (2015–2020). His current research interests include computational and neural mechanisms of implicit human sensorimotor control and interaction among sensory, motor, and perception, and the development of tactile interfaces. He is now involved in the ‘Correspondence and Fusion of Artificial Intelligence and Brain Science’ Project (2016–2020). He is an IEICE fellow and member of the Society for Neuroscience, the Society for the Neural Control of Movement, the Japan Neuroscience Society, JNNS, and the Society of Instrument and Control Engineers.

# Communication with Desired Voice

*Kou Tanaka, Takuhiro Kaneko, Nobukatsu Hojo, and Hirokazu Kameoka*

## Abstract

We convey/understand our intentions/feelings through speech. We also change the impression we want others to have of us by controlling our voice, including intonation, speaking characteristics, and rhythm. Unfortunately, the types of voice and voice controllability an individual can generate are limited. In this article, we introduce the challenges with conventional voice-conversion techniques and introduce improved voice conversion techniques with the following question in mind, "What can be done when the voice is combined with deep learning, which has been advancing in recent years?" Finally, we look at the future of deep learning and speech generation and conversion.

*Keywords: deep learning, signal processing, voice conversion*

## 1. Non/para-linguistic information conversion

Speech is one of the fundamental methods for conveying not only linguistic information but also non/para-linguistic information. We can control the impression we want to give to others of us by changing our voice properties, including intonation, speaking characteristics, and rhythm. However, the expression of the speech generated by an individual is limited due to physical or psychological constraints. Voice-conversion techniques help us go beyond this limitation and express ourselves as we wish by converting our voice into the desired voice (**Fig. 1**). There are a wide variety of applications of voice conversion, i.e., speaker identity conversion, assistance for people with vocal disabilities, emotion conversion, and pronunciation/accent conversion for language learning. These applications have several requirements according to the usage scenarios. We particularly focus on the quality of generated speech, the amount of training data, non-parallel data* training, real-time conversion, and not only voice timbre but also suprasegmental features conversion.

## 2. Research on speech × deep generative model

A typical voice-conversion technique is statistical voice conversion based on Gaussian mixture models [1]. By using the time-aligned acoustic parameters as the training data, we train a model describing the joint probability of acoustic parameters of the source and target speech to obtain a mapping function from the source acoustic parameters to target acoustic parameters. Within the framework that requires the parallel data described above, several methods using a neural network and an example-based method, such as non-negative matrix factorization, have recently been proposed. Although these methods improve conversion quality and controllability, there are still disadvantages, i.e., 1) speech-sample pairs have the same content as the training data, 2) convertible acoustic parameters are limited to the voice timbre, and 3) it is too easy to distinguish the converted speech from normal speech because classical vocoders are used for generating speech waveforms from the given acoustic parameters.

Variational auto-encoder (VAE), generative adversarial networks (GANs), and sequence-to-sequence model (Seq2Seq) have been proposed for tasks such as image recognition and natural-language processing. The auto-regressive model, which is a part of Seq2Seq, VAE, and GANs, is treated as the main deep generative model and has been proven to be

* Non-parallel data: Non-parallel data does not restrict utterance content of input speech and target speech.

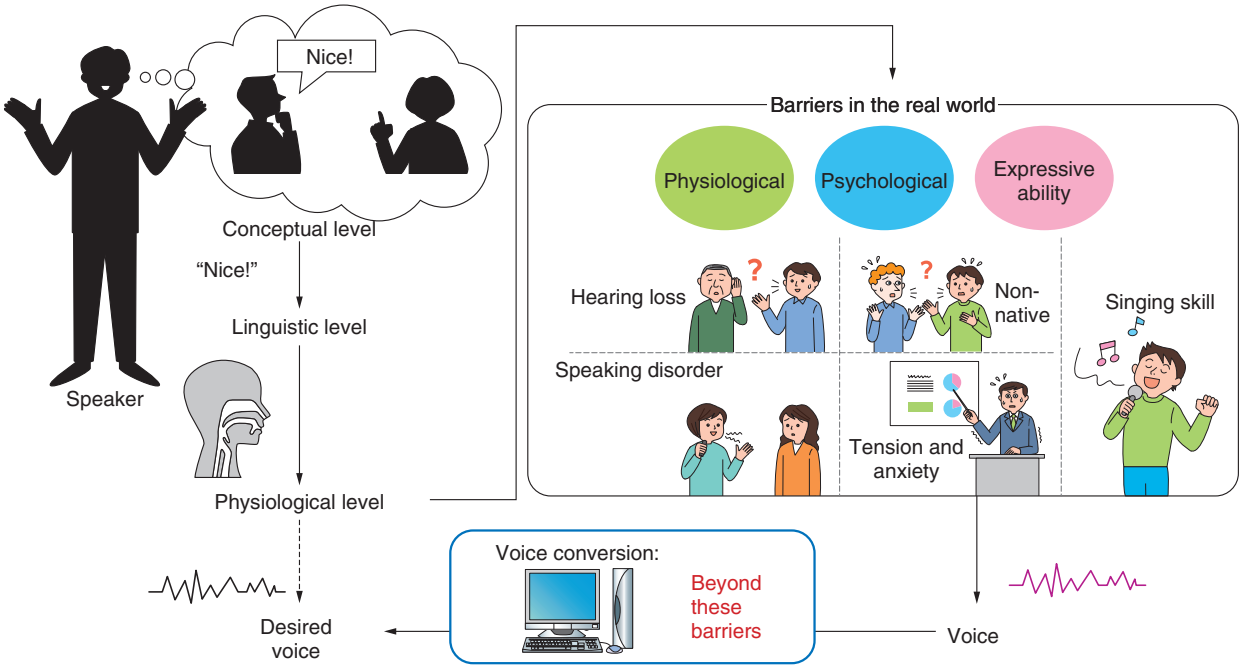



Fig. 1. Speech-generation process and voice conversion.

outstanding in various research tasks. In the research field of the machine translation in mid-2015, an attention mechanism [2] was proposed and has attracted attention due to its effectiveness.

At NTT Communication Science Laboratories, to develop a versatile voice-conversion system that can be flexibly applied to various usage scenarios and overcome the problems with the conventional voice-conversion techniques mentioned above, we proposed the following functions based on the main deep generative models: 1) voice-series conversion for converting not only voice timbre but also prosody and accent, which are long-term dependent features, 2) non-parallel voice conversion trained using non-parallel training data with no restrictions on the utterance content, and 3) waveform post-filter for directly modifying synthetic speech waveforms to real voice waveforms not on the acoustic parameters space but on the waveform space. With these functions, we achieved high-quality, high-efficiency, and real-time voice conversion. In addition, as wider research, we also developed a cross-modal voice-conversion function for converting voice by using the face image of the target speaker.

## 3. High-quality and stable training-voice conversion based on Seq2Seq

We briefly introduce our voice-series conversion function [3]. Unlike words (symbols) that are treated as discrete values in natural-language processing, speech is observed as a series of continuous values. While conventional voice-conversion techniques using series conversion can be expected to significantly improve the quality of generated voice, the difficulty in learning when input and output are continuous value series is an issue. To address this issue, the mainstream approach with conventional speech-sequence conversion is to combine automatic speech recognition (ASR) and text-to-speech synthesis. This approach uses ASR to recognize symbols such as words from given speech, converts the recognized symbol sequence into a symbol sequence of the target speaker, and synthesizes the desired speech from the converted symbol sequence. Since conversion is executed using symbols that are discrete values, the training is relatively stable. However, not only speech but also text labels are required to train the model. In addition, it is not easy to convert laughter, which makes it difficult to convert speech into text.

NTT Communication Science Laboratories has achieved voice-sequence conversion that enables

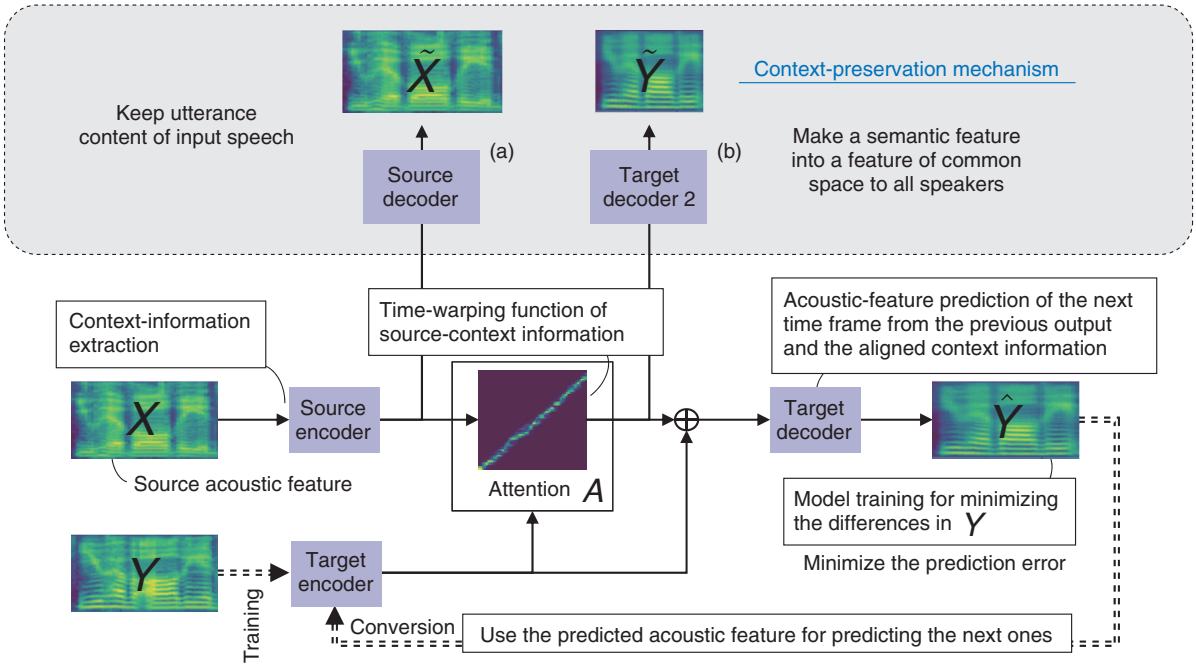



Fig. 2. Seq2Seq voice conversion.

stable training without using text labels. We developed a context-preservation mechanism, shown in **Fig. 2** and described later, to train all modules from only audio data. Instead of text labels, we use this mechanism for adequate model training to convert utterance context. In our context-preservation mechanism, the left module (Fig. 2(a)) is used to restore the input speech. Namely, the utterance content of the input speech is kept in the conversion process. This is called auto-encoding. The right module (Fig. 2(b)) is used to predict the acoustic parameters of the target speech and converts the semantic features into predictable features for both the input voice and the target voice. In other words, it is a constraint that the input speech is mapped to the common space for all speakers. Unlike speech recognition, the important semantic features are automatically determined from the training data. Therefore, richer semantic features rather than the symbols are extracted from the input speech, enabling more accurate conversions.

## 4. Waveform modification for converting synthetic speech to real speech

Another issue with conventional voice-conversion techniques is that they are affected by the accuracy of the waveform synthesizer when synthesizing waveforms from acoustic parameters. Therefore, we may be able to determine that voices have been artificially synthesized. We aim to eliminate such synthetic-voice likeness and directly modify waveforms to normal voice waveforms with higher sound quality and real-voice likeness.

There are two main difficulties in modifying a speech waveform in deep learning. One involves the sampling rate. For example, 16-kHz sampling audio has 16,000 samples in one second. It is easy to understand the difficulty in alignment, which finds the correspondence point of the synthesized speech waveform and real speech. The other involves phase information of speech. In conventional voice-conversion techniques, it is standard practice to discard phase information and handle only amplitude information.

NTT Communication Science Laboratories is tackling these problems by directly modifying speech waveforms using deep learning [4]. A converter, which is the core module in the model, is trained according to two criteria, as shown in **Fig. 3**. One is minimization of discrimination error, which is a criterion for adversarial training. The discriminator recognizes the difference between processed speech and non-processed speech that is difficult to quantify, and the converter is trained to eliminate as much of

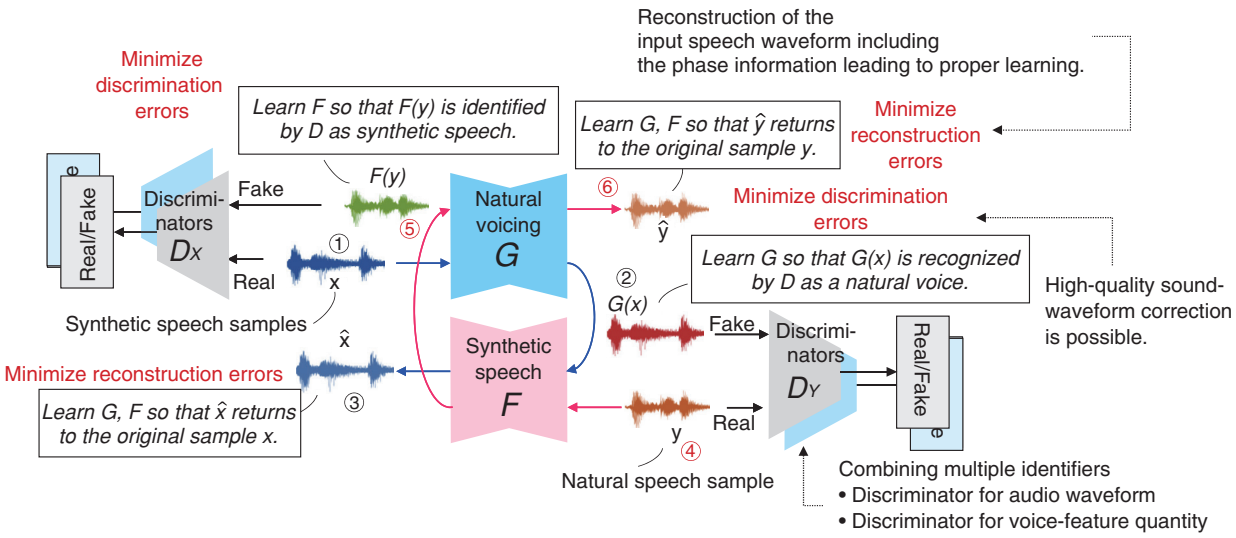



Fig. 3. Synthetic-to-natural speech-waveform modification.

the difference as possible. We also proposed using multiple classifiers instead of one classifier for recognizing the difference among various views and enabling high-quality modification.

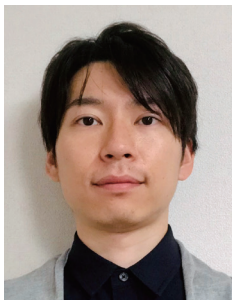
The other criterion is minimization of reconstruction error, which is a criterion for cyclic models. By converting synthetic voice into natural voice then converting it into synthetic voice again, the input synthetic voice has to be restored. The key is that the amplitude information as well as phase information have to be restored. Namely, it is possible to properly process the phase information and make the training process stable. Moreover, this cyclical adversarial model does not require parallel data of synthetic speech and normal speech. By appropriately collecting synthetic speech and natural speech, we can train the desired model.

## 5. Future work

NTT Communication Science Laboratories will continue to improve and demonstrate technology for converting more diverse voices. To date, we have succeeded in high-quality conversion when we have a specific target speaker, e.g., "I want to speak with a target speaker's voice." However, perceptual score conversion is still challenging, e.g., "I want to speak with a cute voice" and "I want to speak with a sterner voice." To enable such conversion, it is necessary to model the perceptual space of speech and interpolate latent variables in that space. This is why perceptual score conversion is challenging. We will promote research and development toward a versatile voice-conversion system that can flexibly work on various usage scenarios and respond to all user requirements.

## References

[1] T. Toda, A. W. Black, and K. Tokuda, "Voice Conversion Based on Maximum Likelihood Estimation of Spectral Parameter Trajectory," IEEE Trans. TASLP, Vol. 15, No. 8, pp. 2222–2235, 2007.
[2] D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," Proc. of the 3rd International Conference on Learning Representations (ICLR 2015), San Diego, CA, USA, May 2015.
[3] K. Tanaka, H. Kameoka, T. Kaneko, and N. Hojo, "AttS2S-VC: Sequence-to-sequence Voice Conversion with Attention and Context Preservation Mechanisms," Proc. of 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, May 2019.
[4] K. Tanaka, T. Kaneko, N. Hojo, and H. Kameoka, "Synthetic-to-natural Speech Waveform Conversion Using Cycle-consistent Adversarial Networks," 2018 IEEE Spoken Language Technology Workshop (SLT), Athens, Greece, Dec. 2018.



**Kou Tanaka**
Research Scientist, NTT Communication Science Laboratories.

He received a B.E. from Kobe University in 2012 and an M.E. and D.E. from the Nara Institute of Science and Technology in 2014 and 2017. He was a research fellow of the Japan Society for the Promotion of Science from 2015 to 2017. His research interests include speech-signal processing with a strong focus on deep generative models.

**Takuhiro Kaneko**
Distinguished Researcher, NTT Communication Science Laboratories.

He received a B.E. and M.S. from the University of Tokyo in 2012 and 2014. He has been pursuing his Ph.D. degree at the University of Tokyo from 2017. His research interests include computer vision, signal processing, and machine learning. He is currently working on image generation, speech synthesis, and voice conversion using deep generative models. He received the ICPR2012 Best Student Paper Award at the 21st International Conference on Pattern Recognition in 2012.

**Nobukatsu Hojo**
Research Scientist, NTT Communication Science Laboratories.

He received a B.E. and M.E. from the University of Tokyo in 2012 and 2014. He joined NTT Media Intelligence Laboratories in 2014, where he engaged in the research and development of speech synthesis. He is a member of the Acoustical Society of Japan, the International Speech Communication Association, and the Institute of Electronics, Information and Communication Engineers of Japan.

**Hirokazu Kameoka**
Senior Distinguished Researcher, NTT Communication Science Laboratories.

He received a B.E., M.S., and Ph.D. from the University of Tokyo in 2002, 2004, and 2007. He is currently an adjunct associate professor at the National Institute of Informatics. From 2011 to 2016, he was an adjunct associate professor at the University of Tokyo. His research interests include audio, speech, and music signal processing and machine learning. He has been an associate editor of the IEEE/ACM Transactions on Audio, Speech, and Language Processing since 2015, a member of IEEE Audio and Acoustic Signal Processing Technical Committee since 2017, and a member of IEEE Machine Learning for Signal Processing Technical Committee since 2019. He received 17 awards, including the IEEE Signal Processing Society 2008 SPS Young Author Best Paper Award. He is the author or co-author of about 150 articles in journal papers and peer-reviewed conference proceedings.

# Quantum Information Processing via Indirect Quantum Control

*Go Kato*

**Abstract**

Quantum information processing will enable us to solve problems that conventional computers cannot or provide informationally secure cryptography. However, there are still difficulties in achieving quantum information processing. This article introduces *indirect quantum control* as one method to overcome these difficulties and shows that indirect quantum control has excellent properties for achieving quantum information processing.

*Keywords: quantum computer, quantum indirect control, noise*

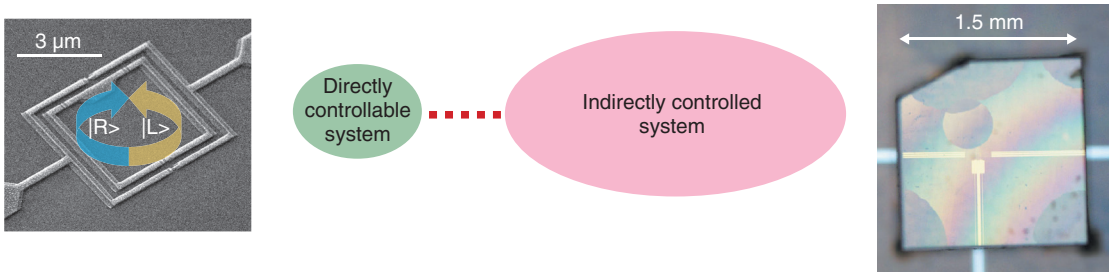
## 1. Current status of quantum information processing

A quantum computer is a quantum-information-processing device that uses the properties of quantum mechanics, which allows two states to exist at the same time. As a result, a quantum computer can naturally perform parallel processing to do things that are impossible with today's computers. For example, it can easily decode any encrypted code currently in use, which is one of the reasons the implementation of quantum computers is greatly anticipated. Another type of quantum information processing is quantum cryptography. This type of cryptography uses a property of quantum mechanics that makes it impossible to measure certain quantities at the same time and is secure in principle. Quantum cryptography is expected to complement the modern cryptography techniques currently in use.

In terms of practical application, devices that implement quantum cryptography have been developed. Although these devices are not currently available to the public, they can be used in special cases. It is only a matter of time before they are introduced into the real world. In fact, the governments of China, the EU, and the United States are conducting field experiments using quantum cryptography.

However, we are still a bit far from the application of quantum computers. An article authored by Google researchers showed how an implemented quantum computer outperforms conventional computers, and IBM offers a service to use a quantum computer on the cloud. Quantum computers made up of a few tens of qubits have been constructed. However, we are not yet ready to implement a quantum computer of a scale and level of accuracy that could change the nature of society. For example, to crack the ciphers currently in use on the Internet, we need thousands of ideal qubits. If just a little noise is left on the qubits, we would need millions of them to crack the ciphers. In other words, the scale of current quantum computers is still an order of magnitude too small to have any kind of significant impact on society.

There are currently two problems that need to be overcome to increase the scale of quantum computers. The first problem is the increase in noise. No matter how small a quantum system, which is composed of quantum computers and other quantum devices, is, it is impossible to completely eliminate noise in quantum computers. Moreover, as the scale of quantum computers increases, the effect of noise becomes greater. This is why we need to find ways to avoid the increase in noise as we increase the scale of such computers. Quantum error-correcting codes are the most promising to overcome this problem. However, they are in the research phase and not yet ready for practical use. The other problem is that the mechanisms needed to control large-scale quantum




(a) Superconducting qubit (b) NV center

Fig. 1. Indirect quantum control.

computers are difficult to implement. In a typical quantum computer model, the number of control mechanisms is proportional to the number of qubits that make up the quantum computer. In other words, as the number of qubits in a quantum computer increases, a vast number of control mechanisms—in the millions—is required. This makes device design very difficult. Conventional computers have overcome the same problem with the concept of a bus, but a quantum structure for what one might call a quantum bus has not yet been developed for large-scale quantum computers.

## 2. Indirect quantum control as a means of overcoming the problems in developing quantum computers

Although various studies have used well-known techniques in attempts to overcome the above problems, it is also important to explore solutions from a new perspective. The idea of indirect quantum control will be just such a new solution.

An indirect quantum protocol is a means of controlling a quantum system via direct manipulation of a part (system S) of the system. That is, the rest of the parts (system E) are controlled indirectly via the interaction between them and system S.

An example of system S is a qubit system that is easy to manipulate. It is composed of superconducting qubits (one such qubit is shown in **Fig. 1(a)**) and has superconducting material arranged in a ring shape. A nitrogen-vacancy (NV) center is an example of system E composed of one qubit. It is made from a lattice defect caused by nitrogen in diamonds (**Fig. 1(b)**). Simply placing a diamond plate containing NV centers on top of a superconducting qubit connects the two quantum systems through their interaction. Thus, a structure for indirect quantum control can be created. Experiments to implement a one-qubit quantum memory system with little noise [1] have been carried out using this structure.

This indirect quantum control structure is generally effective in overcoming the first problem mentioned above, namely the increase in noise. It can be understood as follows. To manipulate a quantum system, it is necessary to place physically implemented manipulators in it. However, if such manipulators are connected, they also allow noise pathways to enter the system (**Fig. 2**). If we want to reduce noise ingress, we have to sacrifice operability by reducing such manipulators. In other words, there is a trade-off between ease of operation and noise reduction in the quantum system. The current implementation of quantum computers relies on the high operability of the entire quantum system. This imposes a limitation on reducing the noise-intrusion paths. Therefore, if it is possible to have complete controllability via limited operability, we will be able to reduce the noise ingress paths and achieve a high level of noise reduction.

The ability to control the entire quantum system by indirect quantum control is a solution to the second problem mentioned above. When a quantum system is completely and indirectly controlled, any quantum information in system E must merge into system S at some point. Such information is manipulated on system S at that particular time and returned to system E. This is abstractly the function of the quantum bus. In other words, the capability of perfect control via large-scale indirect quantum control includes the function of a large-scale quantum bus.

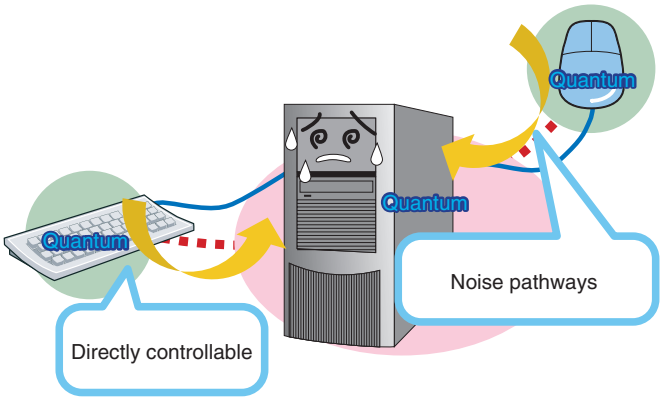



Fig. 2. Noise ingress.

## 3. Challenges in implementing quantum computers with indirect quantum control

As mentioned above, we can overcome the two main problems to constructing large-scale quantum computers by using indirect quantum control. Nevertheless, indirect quantum control has not been considered as a means of large-scale quantum information processing. Though it is used for some small-scale examples of the quantum memory system mentioned above, it has not yet been sufficiently researched. This is because even a small difference in the interaction between systems S and E can significantly affect the type of control that can be implemented. It is not easy even to discuss the ability to process information in such a situation. Any quantum device made up of quantum computers must be completely controlled. However, we have not been able to clearly answer even the basic question of which type of indirect quantum control could meet this requirement.

The research results presented in our recent paper [2] answer this basic question. Given a quantum system, the set of controls performed on it can be identified using a dynamical Lie algebra[*1], which is equal to a linear space made from a set of skew-Hermitian matrices[*2]. Therefore, to answer the question of whether a given quantum system has sufficient controllability, it is sufficient to evaluate its dynamical Lie algebra. For example, in the case of typical quantum computers as currently conceived, the corresponding dynamical Lie algebra contains all skew-Hermitian matrices, which means that the quantum system can be completely controlled.

Regarding general quantum systems, however, there are many possible dynamical Lie algebras. This is why it has been difficult to give answers to even the above basic question in indirect quantum control. However, our recent study [2] has shown that, when operation is done by indirect quantum control, the dynamical Lie algebra must possess a Jordan algebra[*3]. It is well known that a Jordan algebra is much less diverse than a dynamical Lie algebra. Using this fact, we fully classified the dynamical Lie algebra that can be identified in the case of indirect quantum control and found that indirect quantum control has properties that make it a useful means of implementing quantum information processing.

According to the obtained classification of a dynamical Lie algebra, its structure deeply depends on the dimension[*4] of the quantum system (system S) that can be directly manipulated. When the dimension of system S is three or more, the dynamical Lie algebra always contains arbitrary skew-Hermitian matrices, which act on the space that can be affected by manipulating system S. In other words, if the dimension of system S is three or more, a quantum system behind system S, which may be a part of system

*1 Dynamical Lie algebra: A set of elements that are closed with respect to the commutator and form a linear space. In other words, this is a set that satisfies the condition that AB-BA, xA, and A+B are also in this set, where A and B are elements of the set and x is a real number.

*2 Skew-Hermitian matrix: A matrix that makes itself negatively signed by complex conjugation plus transposition.

*3 Jordan algebra: A set of elements that are closed with respect to the anti-commutator and form a linear space. In other words, this is a set that satisfies the condition that AB+BA, xA, and A+B are also in this set, where A and B are elements of the set and x is a real number.

*4 Dimension: The radix in standard information processing. For example, a quantum system consisting of three qubits is an eight-dimensional space.

E, can be completely controlled. This means that indirect quantum control is a universally sufficient means of implementing quantum information processing.

## 4. Summary

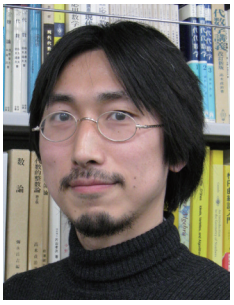
By studying indirect quantum control mathematically, we can see that it is promising as a means of achieving quantum information processing. Though this will not directly enable us to construct large-scale quantum computers with sufficiently low noise tomorrow, we theoretically proved that indirect quantum control, which has not been used before, is a promising means to overcome serious problems in constructing large-scale quantum computers. This is a significant step forward in itself.

By further developing the theories that explain the properties of dynamical Lie algebras identified in the case of indirect quantum control [2], we will be able to evaluate the execution time of control or plan procedures that implement desired controls. This will further significantly contribute to the construction of quantum computers.

## References

[1] X. Zhu, S. Saito, A. Kemp, K. Kakuyanagi, S. Karimoto, H. Nakano, W. J. Munro, Y. Tokura, M. S. Everitt, K. Nemoto, M. Kasu, N. Mizuochi, and K. Semba, "Coherent Coupling of a Superconducting Flux Qubit to an Electron Spin Ensemble in Diamond," Nature, Vol. 478, pp. 221–224, 2011.

[2] G. Kato, M. Owari, and K. Maruyama, "Algebra and Hilbert Space Structures Induced by Quantum Probes," Ann. Phys., Vol. 412, 168046, 2020.



**Go Kato**

Senior Research Scientist, Computing Theory Research Group, Media Information Laboratory, NTT Communication Science Laboratories.

He received a doctor of science degree in physics from The University of Tokyo in 2004. He joined NTT Communication Science Laboratories in the same year and has been engaged in the theoretical investigation of quantum information. He is especially interested in mathematical structures emerging in the field of quantum information. He is a member of the Physical Society of Japan.

# Measuring Textual Difficulty—Estimating Text Readability and a Person's Vocabulary Size

*Sanae Fujita*

**Abstract**

Whether it feels difficult or easy to read text depends on both the readability of that text and the amount of knowledge possessed by the reader. If text readability and the amount of knowledge of the reader (vocabulary size) can be easily and accurately estimated to automatically estimate the difficulty of text, it should be possible to recommend text having a suitable level of difficulty, thereby support the learning process. This article introduces methods we, NTT Communication Science Laboratories, devised for estimating text readability and vocabulary size.

*Keywords: vocabulary size, word familiarity, readability*

## 1. What does it mean to measure difficulty?

Taking, for example, a child who is just learning to read, one can imagine how a parent might select a picture book for the child to read on his or her own but end up reading the book to the child who simply finds it too difficult. Additionally, one can imagine how English text that may have been a struggle for a Japanese student to read in the first year of junior high school seems very simple when he/she is a college student. Given the same text, whether it feels difficult or easy depends on the amount of knowledge the reader has.

If recommendations can be made for picture books, novels, textbooks, or even such books in English that are perfectly readable or readable with a little effort, it should be possible to increase the reader's knowledge without difficulty. However, *determining perfectly readable* or *readable with a little effort* is not a trivial task. This is because both a person's amount of knowledge and text readability must be appropriately estimated.

## 2. Measuring a person's vocabulary size

One type of knowledge that a person needs is vocabulary. At NTT Communication Science Laboratories, researchers have been surveying and estimating the vocabulary size of people in various age groups for over 20 years.

In a survey targeting infants, their vocabulary is not that large, so it is not impossible to investigate all words that can be understood and spoken. We, NTT Communication Science Laboratories, constructed the Child Vocabulary Development Database (CVD) by collecting data on when infants learn and speak what types of words from over 1500 parent-infant pairs.

It is difficult to survey all the words a person knows from elementary school onward. In this case, the approach adopted for estimating vocabulary size is to ask a person whether he or she knows certain presented words. The more words that are presented, the more accurate the estimation becomes, but vocabulary size can still be estimated with only several dozen words.

A key point of this estimation method is hypothesizing how many words a person knows when answering that he or she knows a certain word. For example, given the words "salty" and "tidal gauge," there would probably be less people who know "tidal gauge." The assumption can therefore be made that a

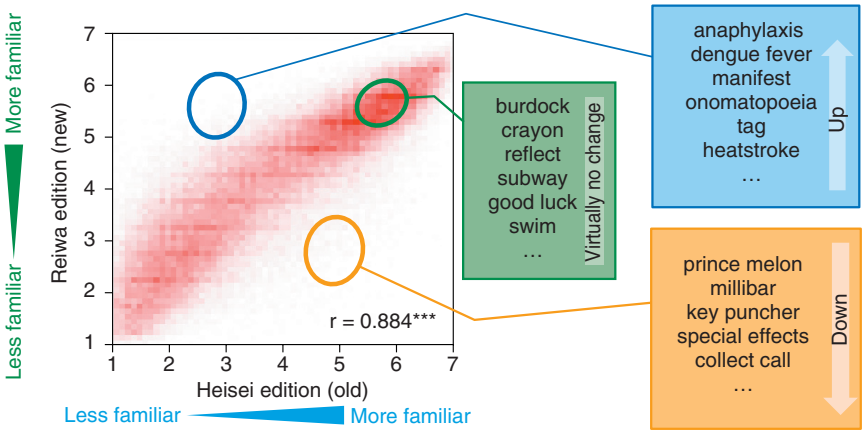



Fig. 1. Change in word familiarity: from Heisei to Reiwa.

person who knows the words "tidal gauge" has a larger vocabulary than a person who only knows "salty." Given that a person is knowledgeable of "tidal gauge," how many words can we assume that person knows? The basis for such an assumption is *word familiarity*.

## 3. Word-familiarity database

The quantification of the degree of word recognition through rating experiments is called *word familiarity*. A word assigned a large number corresponds to a word that many people are familiar with, and a word assigned a small number corresponds to a word that many people are unfamiliar with.

We have been constructing fundamental language resources, such as a word-familiarity database, for over 20 years. The Heisei edition of such a database (*Heisei* is the name of the previous Japanese era from Jan. 8, 1989 to Apr. 30, 2019) consisting of about 77,000 entries was released in 1999 as part of the NTT Database Series "Lexical Properties of Japanese." This series of databases has been widely used as a basic reference in such fields as psychology, language education, and speech and language therapy. However, over 20 years have passed since the initial surveys, which opened up the possibility that word familiarity has changed with the times. There is also the problem that new words (such as "Internet" and "convenience store") are not included.

In the face of these issues, we re-examined the relevance of words appearing in the Heisei edition, took up the inclusion of new words, and constructed the largest word-familiarity database consisting of about 163,000 entries as the Reiwa edition (*Reiwa* is the name of the current Japanese era that began on May 1, 2019) [1]. We also investigated changes in the word-familiarity database from the Heisei edition and found that a strong correlation exists between the two editions with no major change in familiarity for many words after more than 20 years. However, some words did in fact undergo significant change in familiarity, and we clarified what types of words underwent such change (**Fig. 1**).

## 4. Estimating vocabulary size using word familiarity

To estimate vocabulary size using word familiarity, we set up questions by sampling words from high to low degrees of familiarity (**Fig. 2**). Specifically, given a word for which a participant's answer is "Yes, I know this word," we assume that that participant knows all words with a degree of familiarity higher than that of that word and estimate vocabulary size accordingly. For example, if a participant's knowledge extends up to "salty," we assume the vocabulary size to be about 1500 words, but if extending up to "tidal gauge," we assume a vocabulary size of about 139,000 words. We consider, however, that variation likely exists as to what is known near the boundaries between known and unknown words. We therefore apply a logistic regression curve to the participant's answers and take the vocabulary size for which the probability of knowing is exactly 50% as the estimation result (**Fig. 3**).

With this method, the vocabulary size of a participant can be easily estimated by simply having the

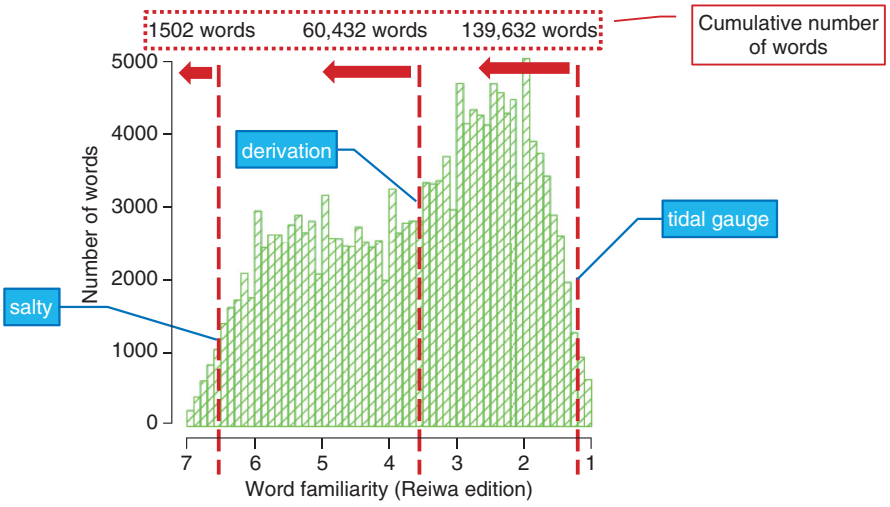



Fig. 2. Histogram of word familiarity: sampling words form high to low degrees of familiarity.

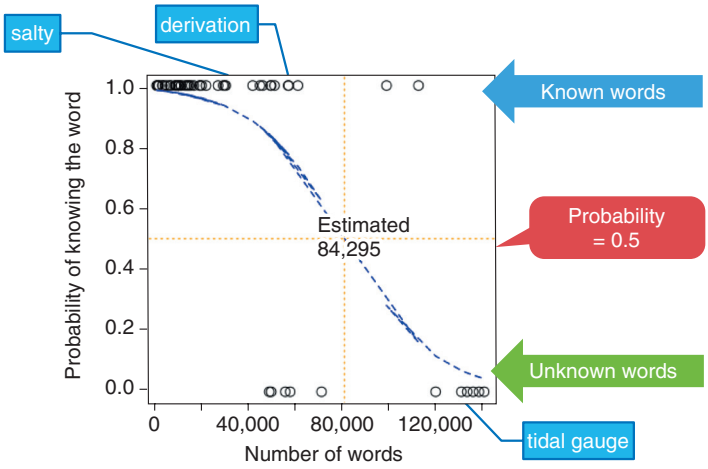



Fig. 3. Method of vocabulary-size estimation.

participant check a small number of sampled words. In addition, since there are words that theoretically have the same degree of familiarity, any of those words may be posed as a question, which means that changing presented words is easy and tests with a certain amount of variation can be prepared. Of course, the more words a participant is asked to check, the greater the estimation accuracy.

The upper limit of vocabulary size that can be estimated with this method depends on the size of the word-familiarity database, but construction of the Reiwa edition of the word-familiarity database significantly raised the upper limit of the vocabulary size that can be estimated, thereby enhancing the versatility of the test.

## 5. Large-scale survey of vocabulary size

There has essentially been no large-scale vocabulary-size surveys targeting children and students, but we used the method described above to carry out a vocabulary-size survey of 4600 individuals including more than 2800 public-school students from elementary school to high school.

The results revealed that vocabulary size increased rapidly for elementary-school students (6–7 to 11–12

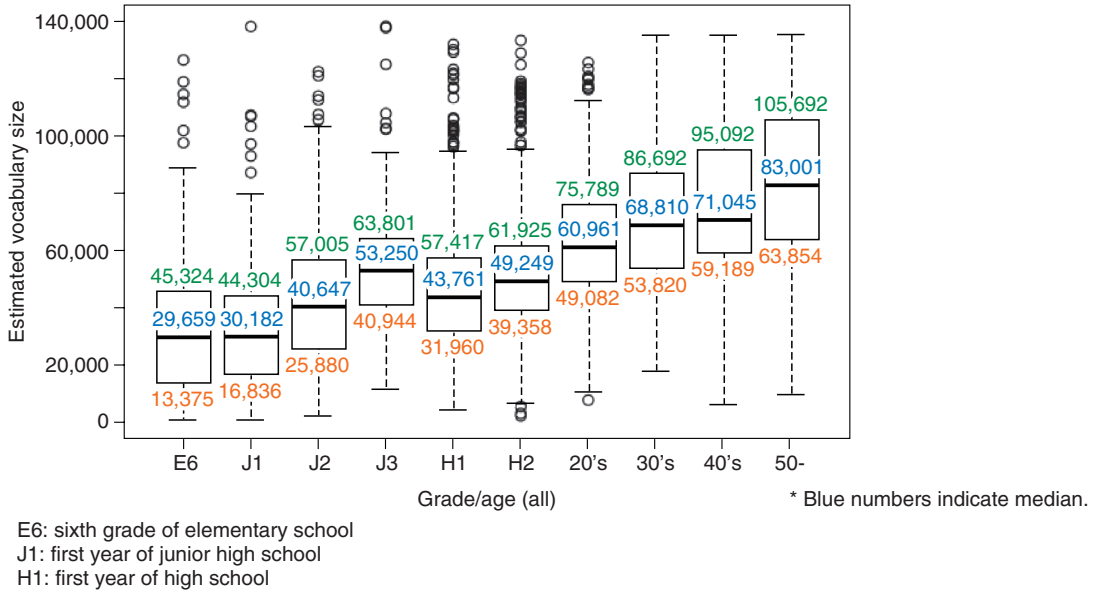



Fig. 4. Estimated results for each grade/age.

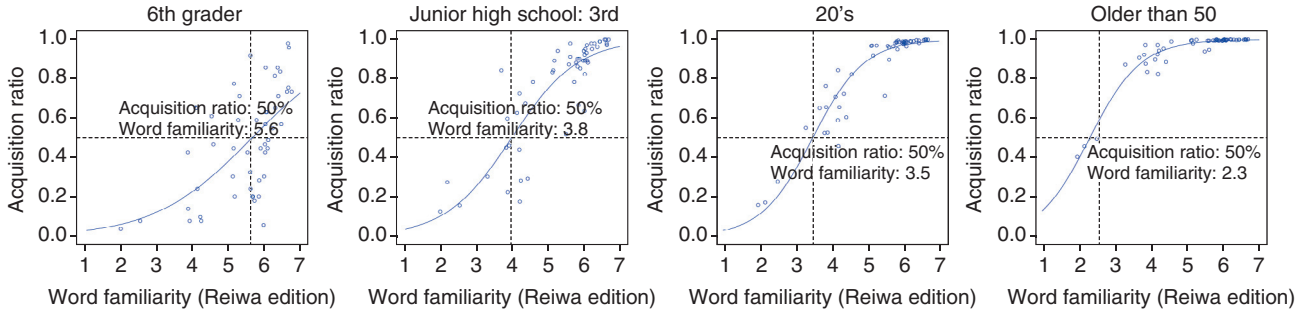



Fig. 5. Relationship between word familiarity and acquisition ratio for each grade/age.

years old) and junior-high students (12–13 to 14–15 years old) and increased for adults along with age (**Fig. 4**). These results also indicate large variation in vocabulary size even among same-year students, so such survey results should be useful in identifying students in need of assistance [2].

We also analyzed the relationship between word familiarity and vocabulary acquisition (percentage of individuals knowing that word) for various school years and age groups, as shown in **Fig. 5**. This figure shows the percentage of individuals answering "Yes, I know this word" for words of various degrees of familiarity (acquisition ratio) for various school years and age groups. These results indicate that the percentage of individuals answering in the affirmative tends to increase for higher degrees of familiarity for any school year or age group and that this trend becomes especially clear with age. However, compared with adults, there is much individual difference and variation regarding knowing or not knowing among elementary-school students and junior-high students, even for words that have a relatively high degree of familiarity. Based on this analysis, it should be possible to use word familiarity as a clue in identifying vocabulary that children and students should prioritize from then on or vocabulary that should be acquired.

We also released a vocabulary-size estimation test based on word familiarity (Reiwa edition) on the web for use by the general public in conjunction with NTT

Communication Science Laboratories Open House (released on June 4, 2020, http://www.kecl.ntt.co.jp/icl/lirg/resources/goitokusei/). About ten days after its release, more than 30,000 individuals have taken the test. We encourage everyone to give it a try.

## 6. Measuring the difficulty of text

We now introduce a method for estimating text readability. We began this research by estimating the readability of a picture book. It was thought that combining this method with research on child vocabulary development conducted at NTT Communication Science Laboratories might contribute to clarifying and supporting vocabulary development.

Text readability is influenced by both the difficulty of the vocabulary used and the difficulty of sentence structure. In the case of picture books, this requires an accurate analysis of *hiragana* (characters that represent sounds in written Japanese). For example, the readability of the hiragana characters "とうさん" (pronounced *tousan*) varies greatly depending on whether they correspond to "父さん" (also pronounced *tousan* but meaning "father") or "倒産" (likewise pronounced *tousan* but meaning "bankruptcy"), where "父" and "倒産" are *kanji* (Chinese) characters. To improve the accuracy of such *hiragana* analysis, we used features that reflect the difficulty of the vocabulary used but referring, for example, to the CVD, and features that reflect the difficulty of sentence structure such as sentence length. We were able to estimate text readability with a level of accuracy as high as 87.8% in terms of classifying picture books into target readership, namely, the four age groups of 0–2, 3, 4, and 5 [3]. However, even if a certain picture book is described as "recommended for children 3 years of age," for example, it may in fact be appropriate for 3-year olds closer to 2 years of age or 3-year olds closer to 4 years of age. In short, estimating the target age at a year before or after the stated recommended age should not present much of a problem. If we therefore consider that such approximate estimations are acceptable, the level of accuracy that we achieved would in fact be 96.7%. It can therefore be said that the proposed method is robust and highly reliable in estimating text readability with practically no instances of greatly mistaking the target age. This method is currently being used to search for picture books suitable for children of various ages as part of NTT's *Pitarie* picture book search system [4].

Much research on estimating text readability is focused on evaluating whether the school year targeted by individual textbooks can be estimated. With this in mind, we also applied this method to estimating the target year of textbooks and found that it has a high level of accuracy above 98% for nine categories from the first year of elementary school to the third year of junior high school. This result indicates that our method for improving the accuracy of estimating the readability of picture books is also useful in estimating the readability of text for elementary and higher students.

## 7. NTT corpus of picture books and children's books

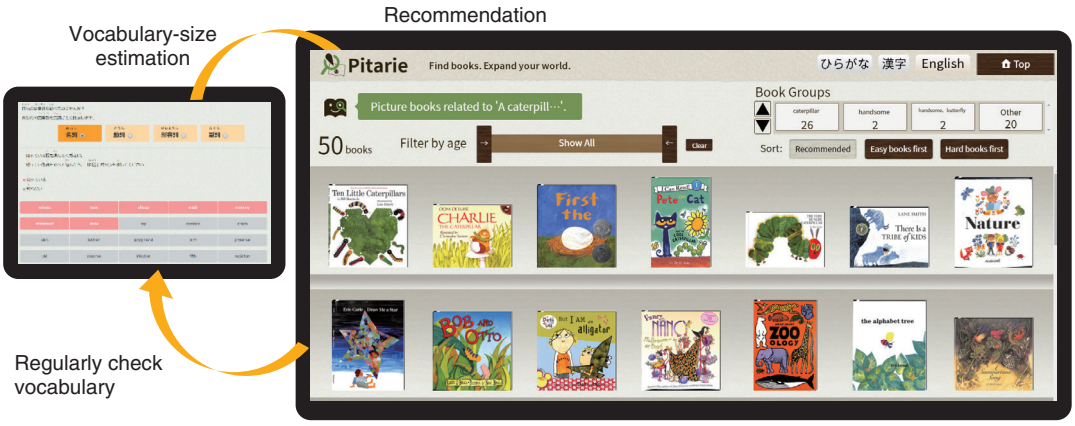
Estimating the text readability described above uses digitized picture-book-text data (corpus). However, no picture book corpus existed, so we had to begin our research by constructing such a corpus. This part of our research was the most arduous and time consuming. The pictures in picture books often include characters, which makes character recognition by optical character recognition infeasible. In the end, most text had to be entered manually. This painstaking work paid off in the form of an NTT picture book corpus consisting of more than 6000 volumes of picture books in Japanese and more than 2500 volumes of picture books in English, a corpus of unprecedented size. There are plans to expand this corpus even further.

We are also actively engaged in many new research projects using this picture-book corpus and children's books such as investigating the relationship between picture books and child vocabulary and emotional development.

## 8. Future developments

Estimating a person's vocabulary size and text readability have up to now been conducted independently. However, combining the two should make it possible to recommend text that is *perfectly readable* or *readable with a little effort* for any individual while periodically checking that person's vocabulary size. In fact, we are also researching the estimation of English vocabulary size and English text readability and beginning an initiative to recommend picture books in English appropriate to a person's vocabulary level for use in English language education at Japanese schools (**Fig. 6**) [5].

Going forward, we aim to provide child-rearing and educational support in both Japanese and English for every person from elementary through high school as




*Information on the picture books shown here is given after the reference section.

Fig. 6. Content recommendation suitable for vocabulary size.

well as adults while collecting evidence on its effectiveness.

## References


[1] S. Fujita and T. Kobayashi, "Reexamination of Word Familiarity and Comparison with Past Examination," Proc. of the 26th Annual Meeting of the Association for Natural Language Processing, pp. 1037–1040, Mar. 2020.

[2] S. Fujita, T. Kobayashi, T. Yamada, S. Sugawara, T. Arai, and N. Arai, "Vocabulary Size of Elementary, Junior High and High School Students and Analysis of Relationship with Word Familiarity," Proc. of the 26th Annual Meeting of the Association for Natural Language Processing, pp. 355–358, Mar. 2020.

[3] S. Fujita, T. Kobayashi, Y. Minami, and H. Sugiyama, "Target Age Estimation of Texts for Children," Journal of Japanese Cognitive Science, Vol. 22, No. 4, pp. 604–620, 2015.

[4] S. Fujita, T. Hattori, T. Kobayashi, Y. Okumura, and I. Aoyama, "Picture-book Search System 'Pitarie'—Finding Appropriate Books for Each Child—," Journal of Natural Language Processing, Vol. 24, No. 1, pp. 49–73, 2017.

[5] S. Fujita, T. Hattori, T. Kobayashi, and F. Naya, "Estimation Method of Vocabulary Size of English Learners," The 34th Annual Conference of the Japanese Society for Artificial Intelligence, June 2020.


**Information on the picture books shown in Fig. 6**

Top row (from left):

B. Martin Jr (author), L. Ehlert (illustrator), "Ten Little Caterpillars," Beach Lane Books, 2011.

D. DeLuise (author), C. Santoro (illustrator), "Charlie the Caterpillar," Aladdin, 1990.

L. Vaccaro Seeger (author and illustrator), "First the Egg," Roaring Brook Press, 2007.

J. Dean, "Pete the Cat and the Cool Caterpillar," Series: I Can Read! 1, HarperCollins, 2018.

E. Carle (author and illustrator), "The Very Hungry Caterpillar," Philomel Books, 1969.

L. Smith (author and illustrator), "There Is a Tribe of Kids," Two Hoots, 2018.

A. Grée, (author and illustrator), "Nature," Button Books, 2012.

Bottom row (from left):

E. Carle (author and illustrator), "Draw Me a Star," Philomel Books, 1992.

R. O. Bruel (author), Nick Bruel (illustrator), "Bob and Otto," Roaring Brook Press, 2007.

L. Child (author), "But I Am an Alligator," Series: Charlie and Lola, Grosset & Dunlap, 2008.

J. O'Connor (author), R. Preiss Glasser (illustrator), C. Bracken (illustrator), "Fancy Nancy Halloween...or Bust!," Series: Fancy Nancy, Harper Festival, 2009.

J. Jolivet (author and illustrator), "Zoo Ology," Roaring Brook Press, 2002.

L. Lionni (author and illustrator), "The Alphabet Tree," Alfred A. Knopf, 1968.

I. Haas (author and illustrator), "A Summertime Song," Margaret K. McElderry Books, 1997.

**Sanae Fujita**
Senior Research Scientist, Linguistic Intelligence Research Group, Innovative Communication Laboratory, NTT Communication Science Laboratories.

She joined NTT Communication Science Laboratories in 1999 and received a Ph.D. in engineering from Nara Institute of Science and Technology in 2009. She is currently researching natural language processing, especially word disambiguation, text-readability estimation, and vocabulary-size estimation. She received the 2013 Best Paper Award of Journal of Natural Language Processing and 2018 NTT President Award. She is a member of the Association for Natural Language Processing and the Information Processing Society of Japan.

# Prediction of Hydrogen Embrittlement of Reinforcing Steel Bars in Concrete Poles

*Takuya Kamisho, Ryuta Ishii, and Masayuki Tsuda*

**Abstract**

Concrete poles are key components of the infrastructure facilities that support telecommunication services. Although the reinforcing steel bars in concrete poles rarely deteriorate under normal operating conditions, the combination of various adverse conditions may lead to hydrogen embrittlement. To enable safer and more economical maintenance of concrete poles, we have been investigating how to predict hydrogen embrittlement of reinforcing steel bars in concrete poles by means of accelerated tests to determine how tensile stress, hydrogen content on the surface of the reinforcing steel bars, and temperature affect the fracture probability and time to fracture. This article presents an overview of the technology for predicting hydrogen embrittlement, including the evaluation results of fracture probability, time to fracture, and their prediction accuracies.

*Keywords: prediction of deterioration, concrete poles, hydrogen embrittlement*

## 1. Introduction

Japan's social infrastructure was intensively constructed during the period of high economic growth in the 1960s and 70s, and as many of these structures are now aging, the cost of maintenance is increasing. The number of skilled maintenance personnel is also decreasing yearly due to the decline in the working-age population stemming from Japan's declining birthrate and aging population. These factors are poised to make the maintenance of social infrastructure quite difficult in the future. To solve this social problem, it is necessary to develop technologies to maintain infrastructure facilities safely and economically.

Concrete poles are key components of the infrastructure facilities that support telecommunication services in Japan. NTT owns about 7 million concrete poles throughout the country, and they are a familiar infrastructure in our daily lives. If an accident such as breakage of a concrete pole due to deterioration occurs, serious damage may result, so all concrete poles are meticulously maintained. The maintenance of concrete poles is currently very expensive, so a safer and more economical maintenance technology for concrete poles is required.

A concrete pole is a hollow reinforced concrete structure with a tensile strength guaranteed with the reinforcing steel bars in the concrete. Prestressed concrete, where compressive stress is applied by reinforcing steel bars to which tensile stress has been applied in advance, is used for concrete poles to suppress cracking. Although these reinforcing steel bars rarely deteriorate under normal operating conditions, the combination of various adverse conditions may lead to hydrogen embrittlement.

Hydrogen embrittlement is a phenomenon in which hydrogen penetrates a metal and reduces its strength under tensile stress, leading to cracking and fracture. The phenomenon known as "delayed fracture" occurs when hydrogen penetrates a metal under constant load and causes fracture to occur after a certain period. It is difficult to predict hydrogen embrittlement of reinforcing steel bars in concrete poles because the mechanism underlying it is not yet fully understood. Therefore, we have been investigating methods of

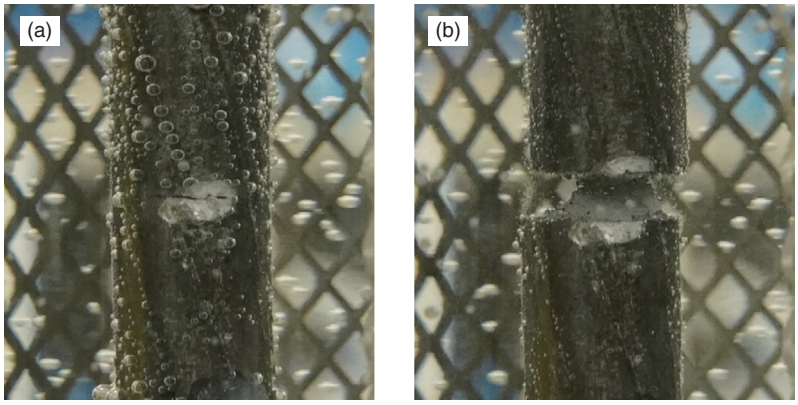



Fig. 1. Fracture of reinforcing steel bar from accelerated test of hydrogen embrittlement: (a) immediately before and (b) immediately after fracture.

predicting such hydrogen embrittlement to enable safer and more economical maintenance. This article presents an overview of the technology for such prediction.

## 2. Prediction of hydrogen embrittlement

Hydrogen embrittlement of reinforcing steel bars in concrete poles occurs when hydrogen is generated by corrosion of the bars and penetrates them with excessive tensile stress. Since concrete is usually an alkaline environment, corrosion of the bars does not occur in most cases. However, when concrete poles are deflected by excessive load, the concrete cracks and becomes neutralized, then the bars corrode. This deflection also applies excessive tensile stress to some of the reinforcing steel bars. The risk of hydrogen embrittlement increases when hydrogen is generated due to the corrosion of reinforcing steel bars with excessive tensile stress.

The occurrence of hydrogen embrittlement is determined by the tensile stress on and hydrogen content and temperature of the reinforcing steel bars. We can estimate the magnitude of the tensile stress on the bars by measuring the amount of deformation of the concrete pole, and the temperature can be determined from meteorological data. If the hydrogen content in a reinforcing steel bar is known, it should be possible to predict the hydrogen embrittlement of that bar from the deformation data of the concrete pole and meteorological data. In this study, we predicted the hydrogen embrittlement of reinforcing steel bars in concrete poles by deriving the relational expressions of the fracture probability and time to fracture obtained from an accelerated test. We can use these values in place of those in the actual environment.

### 2.1 Accelerated test of hydrogen embrittlement

Hydrogen embrittlement of reinforcing steel bars is expected to take ten years or more to occur in an actual environment. Since this length of time is impractical for an actual experiment, we conducted an accelerated test to generate hydrogen embrittlement fractures in a shorter time by increasing the hydrogen content in the reinforcing steel bars (**Fig. 1**). We then evaluated the relationship between hydrogen embrittlement and each variable.

The accelerated test was carried out by applying a predetermined tensile stress to a reinforcing steel bar using a tensile testing machine and charging hydrogen into the reinforcing steel bar. The bar was immersed in a test solution using a test cell, and the hydrogen was charged using a cathode charging method in which hydrogen was generated by electrolysis through the current flow. We used a test solution of 1 mol/L $NaHCO_3$ with a predetermined amount of $NH_4SCN$, which increased the amount of hydrogen penetration. The hydrogen content in the reinforcing steel bar was controlled by the amount of $NH_4SCN$ added. The temperature was controlled by circulating water at a predetermined temperature in the outer layer of the test cell.

### 2.2 Prediction of fracture probability

Hydrogen embrittlement of reinforcing steel bars may or may not occur under the same environmental conditions, so we express whether a reinforcing steel bar will fracture in the future as the fracture probability.

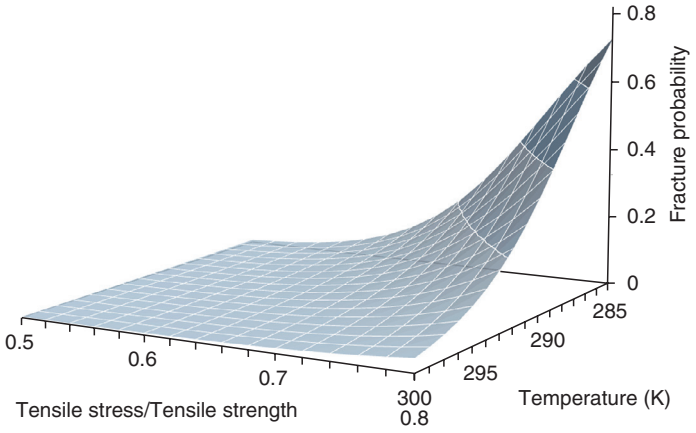



Fig. 2. Relationships between fracture probability, tensile stress, and temperature.

The fracture probability increases when both the tensile stress applied to a bar and the hydrogen content in the bar increase. The fracture probability becomes almost zero when the tensile stress and hydrogen content are low enough. In addition, we assume that the fracture probability is affected by temperature since the mechanical properties of reinforcing steel bars and the diffusion behavior of hydrogen change as temperature changes.

We used logistic regression analysis to quantitatively evaluate the relationship between the fracture probability and each variable. This type of analysis is typically conducted to evaluate the probability when the objective variable is binary (e.g., fracture or not fracture). We can determine how the fracture probability changes depending on each variable of hydrogen embrittlement by defining the fracture of a reinforcing steel bar from the accelerated test as 1 and the non-fracture as 0 then conducting logistic regression analysis.

The relationships between the fracture probability, tensile stress, and temperature at a certain hydrogen content on the surface of a reinforcing steel bar after repeated accelerated tests under various conditions are shown in **Fig. 2**. We can see that the fracture probability increased with increased tensile stress and decreased temperature, and the risk of hydrogen embrittlement increased at high tensile stress and low temperature. In the future, it should be possible to predict the fracture probability in an actual environment by substituting the values of each variable in the environment with the relational expression of the fracture probability obtained from the accelerated tests.

### 2.3 Prediction of time to fracture

When a reinforcing steel bar fractures due to hydrogen embrittlement, the time to fracture changes depending on the environmental conditions. It is clear that the time to fracture decreases with increasing tensile stress and hydrogen content, and that the time to fracture is affected by temperature as well as the fracture probability, but the interactions among the variables, including temperature, are not yet clear.

Regression analysis using a simple linear model cannot be used to effectively evaluate the relationship between the time to fracture and each variable because of the complex interactions among variables. Therefore, we evaluated this relationship by applying variable transformation analysis and maximum likelihood estimation to the time-to-fracture data obtained from the accelerated test.

The relationships between the time to fracture, tensile stress, and temperature at a certain hydrogen content on the surface of a reinforcing steel bar after repeated accelerated tests under various test conditions are shown in **Fig. 3**. We can see that the time to fracture decreased with increased tensile stress and decreased temperature. This indicates that there is a risk of fracture in a short time at high tensile stress and low temperature. In the future, it should be possible to predict the time to fracture in an actual environment by substituting the values of each variable in the environment with the relational expression of the time to fracture obtained from the accelerated tests.

### 2.4 Evaluation of prediction accuracy

To apply the relationship between the fracture probability and time to fracture to the maintenance of

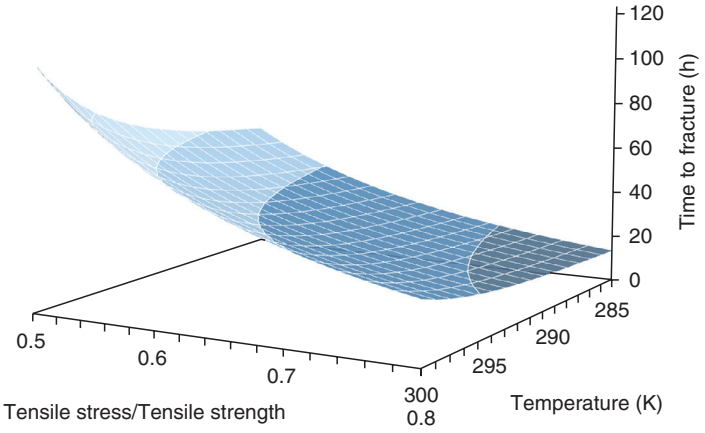



Fig. 3. Relationships between time to fracture, tensile stress, and temperature.

concrete poles, we need to know the prediction accuracy.

We evaluated the accuracy of the fracture probability as determined from the percentage of correct predictions, defining the predicted fracture probability ≥ 0.5 as fracture and < 0.5 as no fracture. The evaluation results indicate that the prediction accuracy of the fracture probability was 0.87, which demonstrates that sufficient prediction accuracy was obtained.

Next, we checked the accuracy of the time to fracture by examining the predicted vs. measured plot of the time to fracture. This plot is a graph in which the predicted value is plotted on the horizontal axis and the measured value on the vertical axis. The closer the plot is to the diagonal line, the better the prediction accuracy. The results are shown in **Fig. 4**, where the plots of the measured values represent the mean of the measured values and the error bars represent ± standard deviation. We can see that the plots were gathered around the diagonal line, which indicates that sufficient prediction accuracy was obtained. We also examined another quantitative measure of accuracy, the mean absolute percent error (MAPE), and found that the MAPE of the mean value of the time to fracture was 14%, which confirms that sufficient prediction accuracy was obtained.

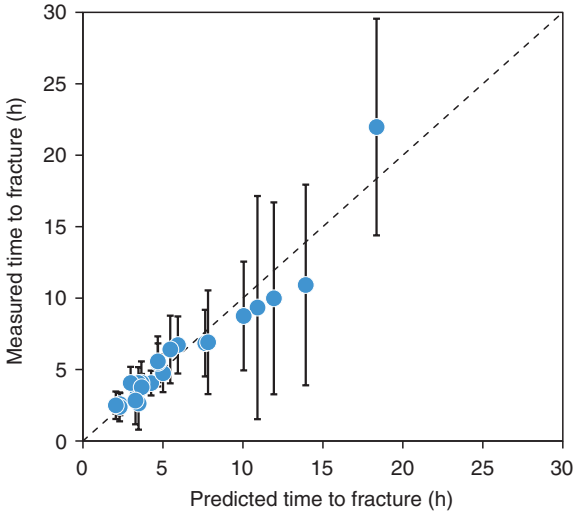
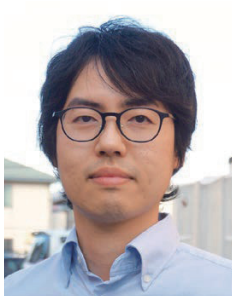



Fig. 4. Predicted vs. measured plot of time to fracture (plots represent mean of measured values, and error bars represent ± standard deviation).

## 3. Conclusion

To enable safer and more economical maintenance of concrete poles, we examined how to predict hydrogen embrittlement of reinforcing steel bars in concrete poles by means of accelerated tests to determine how tensile stress, hydrogen content on the surface of the bars, and temperature affect the fracture probability and time to fracture. We derived the relational expression of the fracture probability and time to fracture of these reinforcing steel bars and found that sufficient prediction accuracy was obtained. In the future, it should be possible to predict the fracture probability and time to fracture in an actual environment by substituting the values of each variable in the environment with the relational expressions of the fracture probability and time to fracture obtained from the accelerated tests.



**Takuya Kamisho**

Research Engineer, Advanced Infrastructure Maintenance Technology Research Group, NTT Device Technology Laboratories.

He received a B.S. and M.S. in meteorology from Hokkaido University in 2008 and 2010. He joined NTT Environment and Energy Research Laboratories in 2010 and has been with NTT Device Technology Laboratories since 2018, where he is currently working on the development of prediction technology for hydrogen embrittlement.

**Ryuta Ishii**

Researcher, Advanced Infrastructure Maintenance Technology Research Group, NTT Device Technology Laboratories.

He received a B.S. and M.S. in condensed matter physics from Osaka University in 2012 and 2018. He joined NTT Device Technology Laboratories in 2018, where he is currently working on the development of prediction technology for hydrogen embrittlement.

**Masayuki Tsuda**

Senior Research Engineer, Supervisor, and Group Leader, Advanced Infrastructure Maintenance Technology Research Group, NTT Device Technology Laboratories.

He received an M.E. and Ph.D. in mechanical engineering from Tokyo Institute of Technology in 1998 and 2005. He joined NTT Integrated Information and Energy Systems Laboratories in 1998, where he developed materials for lithium-ion batteries and environmental impact assessment of information and communication technology services. From 2013 to 2018, he worked at the NTT Environment Promotion Office. He was transferred to NTT Device Technology Laboratories in 2018, where he is currently working on the development of technology to visualize infrastructure deterioration.

# Technology for Understanding Service Impact Using Network Resource Management Technology that Is Independent of Network Type

*Masataka Sato, Shohei Nishikawa, Kenji Murase, and Kenichi Tayama*

**Abstract**

In large-scale and diverse networks of telecommunications carriers, it is difficult to quickly understand the effects on communication services due to damaged communications buildings and cables. NTT Access Network Service Systems Laboratories is researching and developing network resource management technology that enables unified management of various networks. In this article, we introduce technology for understanding service impact during a large-scale disaster using this network resource management technology.

*Keywords: network resource management, understanding service impact caused by disasters, management model*

## 1. Introduction

During and immediately after large-scale disasters, such as earthquakes and typhoons, it becomes difficult to provide communication services due to damage to communication buildings and cables. Communication services are vital infrastructures for modern society, so those damaged by a disaster must be quickly repaired.

The larger the scale and wider the area of a disaster, the more communication equipment will be damaged. It is also necessary to quickly restore important communication services such as lines between government offices responsible for disaster recovery and lines used by many users. However, the networks of telecommunications carriers combine various communication technologies to provide services nationwide, so damage to facilities in one area affects services in other areas as well as other types of services. Therefore, it is difficult to quickly understand the total service impact caused by disasters. In such a situation, there is a demand for technology to enable quick understanding of the service impact caused by disaster-affected equipment and prioritize equipment to be restored from the vast amount of disaster-affected equipment.

## 2. Overview

We have been researching and developing network resource management technology called Network Operation Injected Model (NOIM), which enables unified management of various networks [1]. NOIM manages network information in a versatile data format that does not depend on network type, such as network termination points and connectivity, and enables various networks to be managed centrally. It is possible to quickly understand the service impact caused by communication-equipment failure in a complicated carrier network that combines many

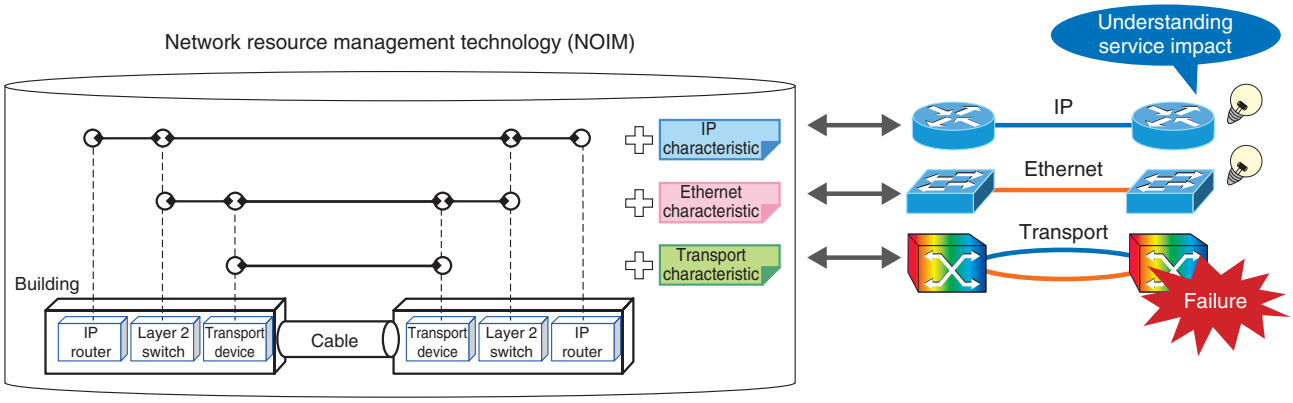



Fig. 1. Overview of NOIM.

communication technologies.

In addition to the example of storing transport/Ethernet/Internet protocol (IP) network information shown in **Fig. 1**, NOIM can store a wide range of network information such as leased lines and mobile networks and help in understanding the service impact during a disaster on those networks.

## 3. Purpose of NOIM

NOIM adopts the entity specified in the information framework called shared information/data model (SID) discussed in the TM Forum [2] and enables unified management of network information in a versatile data format. We adopted three types of representative entities among the SID's logical resources. Termination point encapsulation (TPE), representing termination point of forwarding; network forwarding domain (NFD), representing the opportunity to configure the potential for forwarding; and forwarding relationship encapsulation (FRE), representing all aspects of forwarding of information across a network on NFD. Similarly, NOIM is expressed as a versatile entity with regard to physical resources such as communication devices: physical device (PD), optical fibers: physical link (PL), communication buildings: physical structure (PS), and cables: aggregate section (AGS). Combining these versatile entities makes it possible to express the management information necessary for managing the multi-layer communication protocol (**Fig. 2**).

Furthermore, when retaining this versatile entity information, NOIM has a mechanism to externally define different information for each communication technology. Since the conventional resource management technology that manages networks has a database (DB) specialized for holding information to be managed for each target network, a management function needs to be added to add or change the managed network. In NOIM, the management function is implemented as versatile logic that does not depend on the network type based on the versatile entity information, so there is no need to modify the management function when adding or changing the network type.

**Figure 3** compares the conventional technology and NOIM. Consider the case of managing an IP network and Ethernet network as an example.

The conventional technology (Fig. 3(a)) has a DB specialized for an IP + Ethernet network and a management function specialized for those networks. If a transport network is added to the management target, it is necessary to change the DB or modify the management function so that the transport network can be managed.

In NOIM (Fig. 3(b)), even when a transport network is added, the data of the transport network can be stored by externally defining their characteristics. The target network can be easily expanded. Similarly, regarding the management function, the logic is based on a versatile entity, and it is not necessary to change the function by adding the transport network.

## 4. Technology for understanding service impact

With NOIM, we have developed a technology for quickly understanding the service impact caused by disasters. Using the versatile data format entity described above, the information of various communication networks constructed by telecommunications

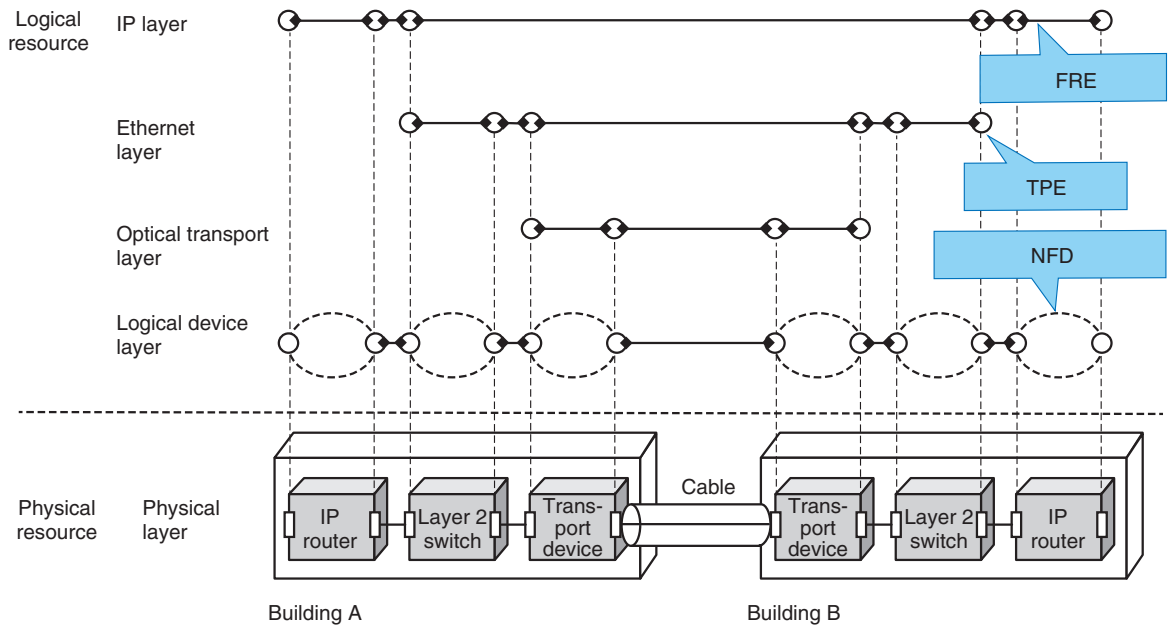



Fig. 2. Data format of NOIM.

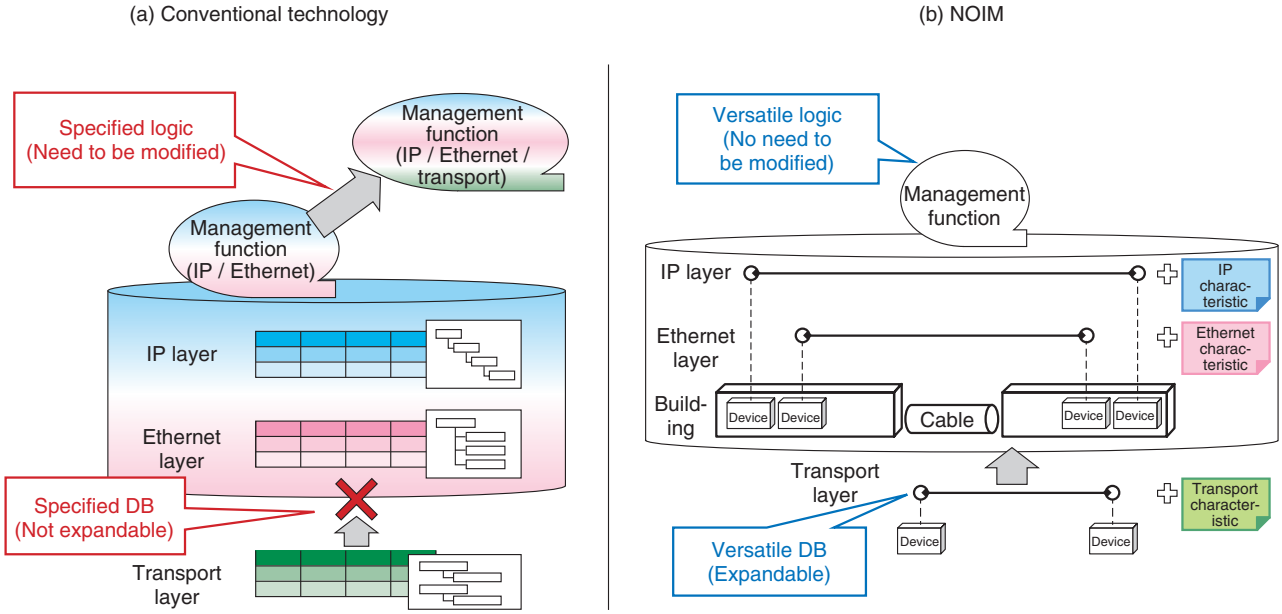



Fig. 3. Purpose of NOIM.

carriers is stored in unified data (**Fig. 4**).

This technology accepts physical equipment (buildings and cables) damaged by a disaster as an input, searches for logical resource entities related to the damaged equipment, and outputs these entities. It is possible to understand the impact on the communication service that occurs due to propagation damage from the communication equipment. Figure 4 shows an image of searching for the impact of propagation from a physical resource entity to logical resource

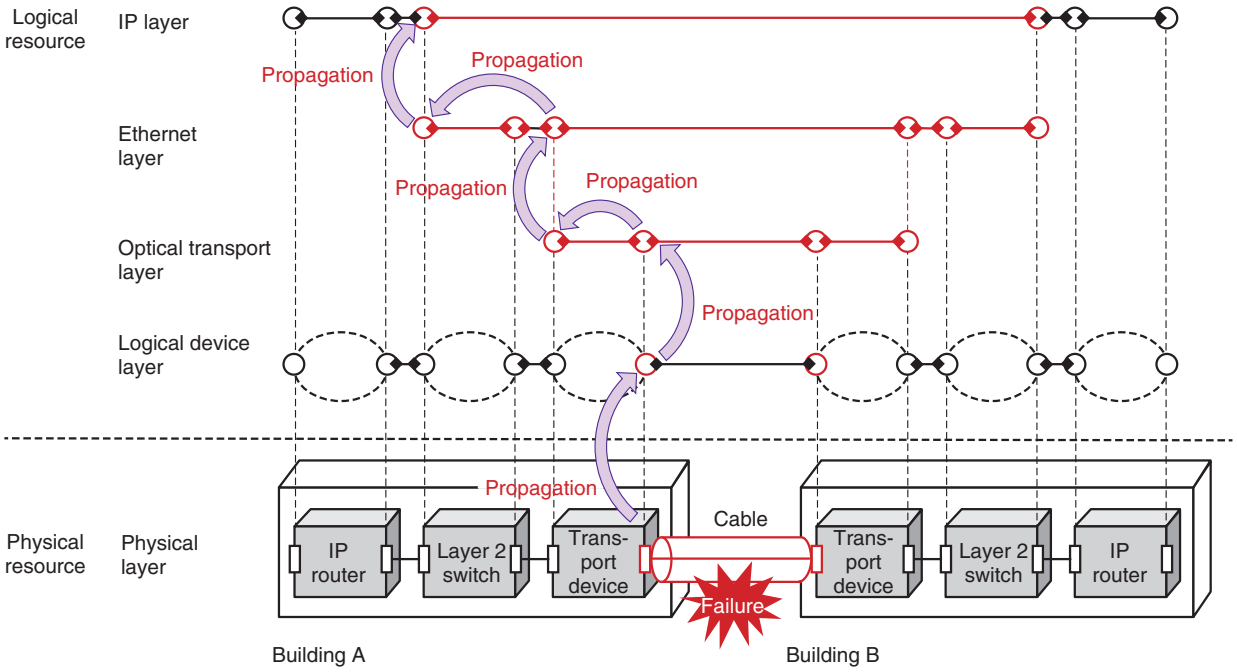



Fig. 4. Procedure of technology for understanding service impact.

entities, assuming that the communication cable is damaged. All physical/logical resource entities have information on the hierarchical relationship between layers and the end point of the transfer path, and it is possible to search for the impact across layers using this information.

This technology with NOIM does not depend on network type due to tracing the relationship between the entities (termination points, transfer paths, etc.) expressed in the versatile data format and their layers [3].

## 5. Future plans

We will implement our technology for understanding service impact during a large-scale disaster by using NOIM through commercial trials. We will also continue to research and develop technology that further enhances and improves the efficiency of disaster response operations.

## References

[1] M. Sato, K. Akashi, S. Horiuchi, and T. Kotani, "Study of Variable Management Architecture for Diverse Networks," IEICE Technical Report, Vol. 116, No. 324, ICM2016-31, pp. 37–42, Nov. 2016.
[2] TM Forum site, http://www.tmforum.org/
[3] K. Fukami, K. Murase, M. Sato, and K. Tayama, "Study on Method of Identifying Service Influence Occurred by Network Fault," IEICE Technical Report, Vol. 118, No. 483, ICM2018-51, pp. 13–18, Mar. 2019.

**Masataka Sato**
Senior Research Engineer, NTT Access Network Service Systems Laboratories.
He received a B.E and M.E. in engineering from Tokyo Institute of Technology in 2002 and 2004. He joined NTT Access Network Service Systems Laboratories in 2004 and has been researching and developing access network operation systems. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE).

**Shohei Nishikawa**
Engineer, NTT Access Network Service Systems Laboratories.
He received a B.E and M.E. in engineering from Keio University, Kanagawa, in 2011 and 2013. He joined NTT Comware in 2013 and has been researching and developing access network operation systems.

**Kenji Murase**
Senior Research Engineer, NTT Access Network Service Systems Laboratories.
He received an M.E. in engineering from Waseda University, Tokyo, in 2006. He joined NTT Communications the same year and is currently engaged in developing operation support systems of access networks.

**Kenichi Tayama**
Senior Research Engineer, Supervisor, NTT Access Network Service Systems Laboratories.
He received a B.E. and M.E. in electrical engineering from the University of Electro-Communications, Tokyo, in 1993 and 1995. He joined NTT Optical Network Systems Laboratories in 1995. He also worked at NTT EAST's IT Innovation Department and NTT-ME's Network Operation Center, where he researched and developed network operations support systems. He is a member of IEICE.

# Standardization Trends in 3GPP Related to IP Interconnect Specifications

*Haruka Eitoku*

**Abstract**

NTT Network Service Systems Laboratories has been working over a decade on the standardization of the specifications of migrating from public switched telephone networks to Internet protocol (IP) networks and inter-operator IP interconnection of telephone services due to the expanding use of voice over IP. Thanks to this effort, IP interconnection in Japan will start in 2021. In this article, we describe the trends and NTT's activities in international and domestic standardization of IP interconnection.

*Keywords: SIP, IMS, IP interconnection*

## 1. Introduction

To enable users served by different telecommunication operators to talk with each other, the network of each operator needs to be interconnected with one another. With conventional voice over Internet protocol (VoIP) services, each telecommunication operator's network is connected to the public switched telephone network (PSTN), and phone calls between VoIP users are provided through the PSTN. However, telecommunication operators in Japan and the Ministry of Internal Affairs and Communications are planning to change the via-PSTN interconnections to direct VoIP interconnections due to the inefficiency of IP-to-STM (Synchronous Transport Module) conversions and maintenance limits of some PSTN nodes (**Fig. 1**).

The conventional interconnection through the PSTN is achieved with the mature signaling protocol called Integrated Services Digital Network User Part (ISUP), but direct IP interconnection needs to be achieved with the Session Initiation Protocol (SIP), which is excessively complex for telephone services and has a wide variety of usage. NTT Network Service Systems Laboratories has worked on developing new SIP common interface specifications to ensure interoperability of IP telephone networks through domestic and international standardization activities.

## 2. Standards developing organizations for IP interconnection

Today, SIP is widely used as a protocol for VoIP session control (e.g., connecting and disconnecting communications with the other party). The basic specification of SIP is defined in Request for Comment (RFC) 3261 developed by the Internet Engineering Task Force (IETF), an organization that develops technical standards for the Internet. In addition to the basic specification of RFC 3261, many extensions to the basic SIP have also been developed. RFC 5411, which is a guide for understanding SIP-related RFCs, indicates that there were over 100 related RFCs even in 2009. Telecommunication operators need to select technical specifications among the numerous RFCs to be implemented for providing their VoIP services. To solve this problem, international standards developing organizations, such as the International Telecommunication Union - Telecommunication Standardization Sector (ITU-T) and the 3rd Generation Partnership Project (3GPP), have deliberated service requirements and the network architecture for telecommunication operators to provide their services and published profiles of

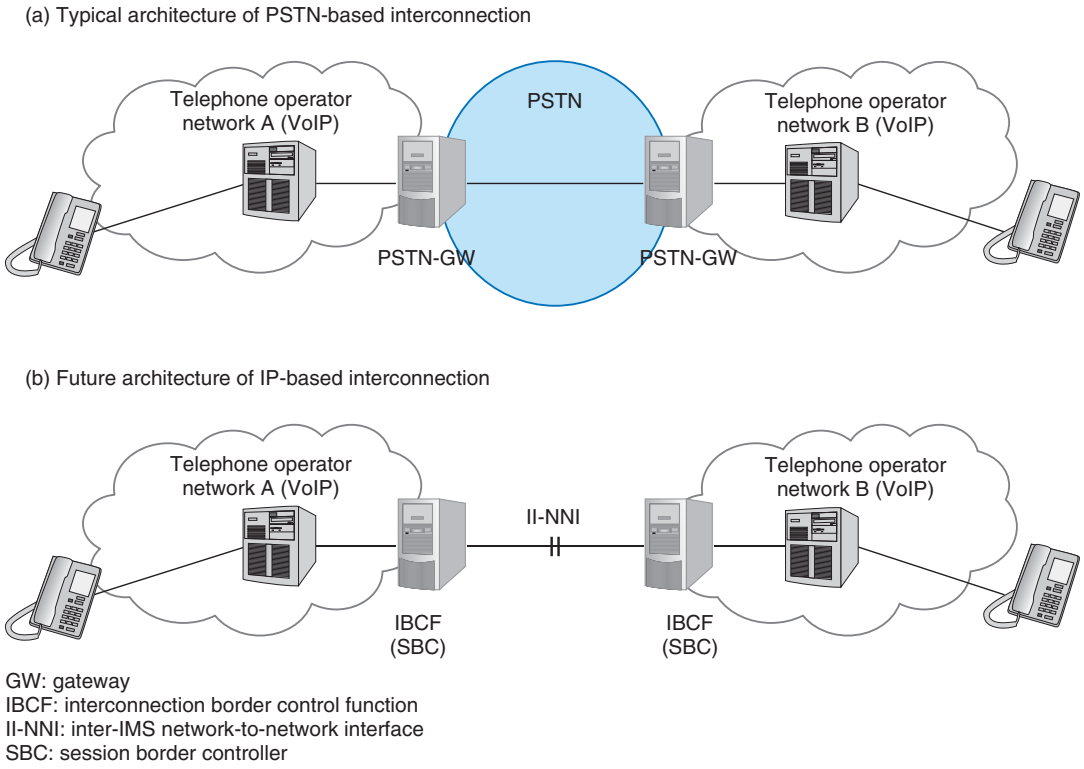



Fig. 1. Architecture of interconnection for VoIP services.

signaling protocols as international standards. Currently, the IP Multimedia Subsystem (IMS) specified by 3GPP is the global standard for telecommunication operator networks using SIP.

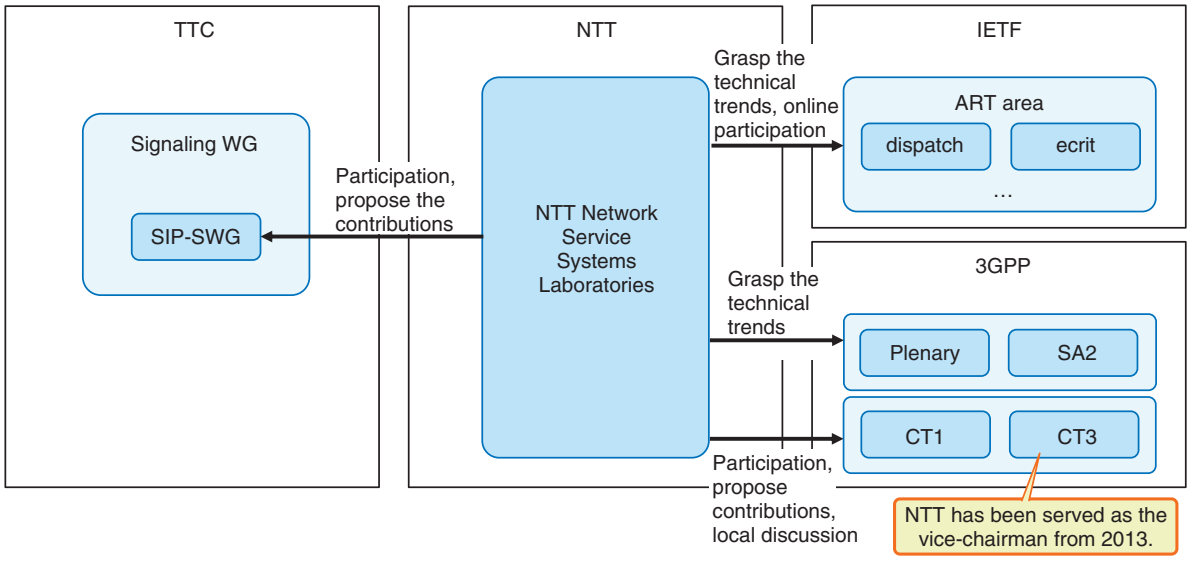
## 3. Latest trends in 3GPP standardization

3GPP was organized to promote the standardization of 3rd-generation and later mobile communications and related technologies (e.g., 3G, 4G, 5G, IMS). The 3GPP specifications are implemented by mobile operators.

3GPP is now vigorously developing technical specifications related to 5G and studies the progress in units called Releases. The goal of Release 15 was developing the basic specification for 5G (Phase 1), that of Release 16 was the completion of the 5G specification (Phase 2), and that of Release 17 is additional enhancements related to 5G. The "5G" referred to here consists of the 5G radio network and 5G core network (5GC), and the IMS nodes are interconnected with the 5GC. The IMS nodes are interconnected with the 4G core network using the Diameter protocol, so achieving interconnection between the IMS nodes and 5GC using this protocol was the goal of Release 15. Achieving interconnection between the IMS nodes and 5GC using HTTP (Hypertext Transfer Protocol) was defined in Release 16 to follow the latest technical trends (e.g., cloud computing). In Release 17, 3GPP is working on the use of 5GC specific features (e.g., network slicing and mobile edge computing) for IMS. As well as 5G enhancements, 3GPP is continuing to introduce new IMS supplementary services. For example, restricted local operator services that enable unauthenticated devices to be given a specific service and multi-device and multi-identity services have been specified. We are proposing many contributions based on the analysis of the impact of these new features and specifications to incorporate them into existing national services and domestic standards with no negative effects.

## 4. IP interconnection standardization trends

We have worked on standardization at 3GPP, IETF, and the Telecommunication Technology Committee (TTC) (**Fig. 2**), and the outputs are reflected in




ART: Applications and Real-Time Area
CT: Core Network and Terminals
ecrit: Emergency Context Resolution with Internet Technologies
SA: Service and System Aspect

Fig. 2. Standards developing organizations in which NTT participates.

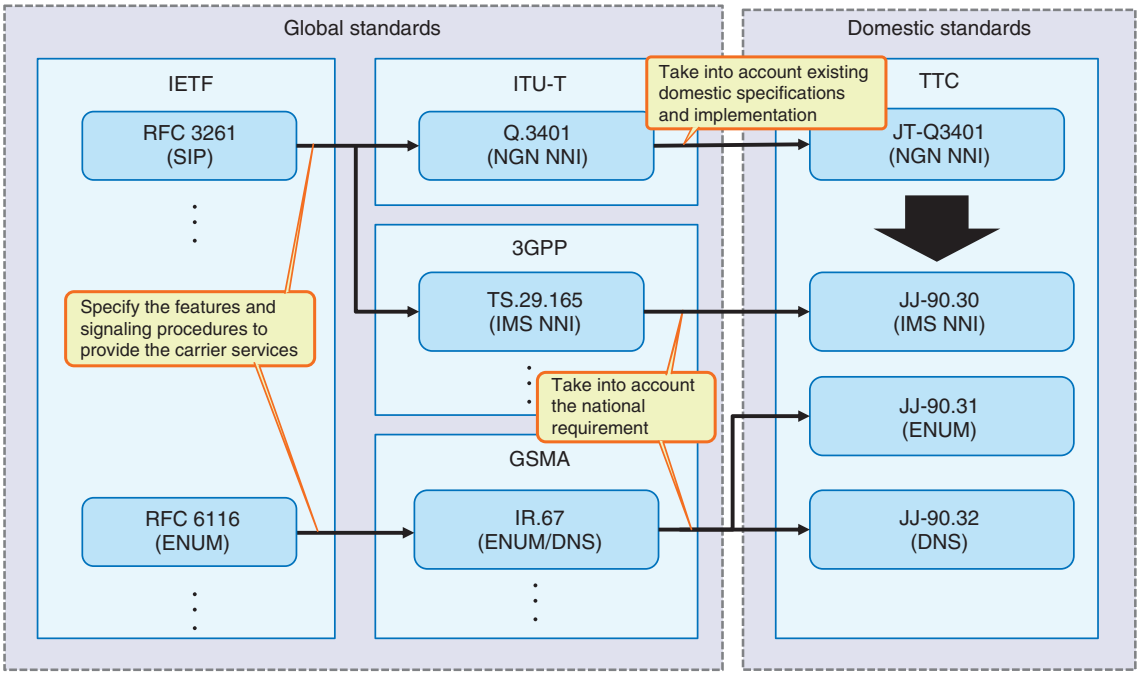
domestic standards for IP interconnection (e.g., TTC JJ-90.30) and global standards (e.g., 3GPP TS 29.165) endorsed by the domestic ones (**Fig. 3**).

The interface of NTT's Next Generation Network (NGN) was originally designed based on the NGN specifications of ITU-T. The domestic network-to-network interface/user network interface (NNI/UNI) specifications of NGN (TTC JT-Q3401/JT-Q3402) were published based on ITU-T Q.3401/Q.3402 considering the national requirements and services in 2007. 3GPP created a new NNI specification between IMS networks in 2008, which is known as Technical Specification (TS) 29.165. We expect that TS 29.165 will become the basis of future domestic NNI specifications for IP interconnection instead of ITU-T ones because the interconnection between fixed and mobile networks needs to be addressed. We began analyzing the differences between ITU-T specification JT-Q3401 and 3GPP specification TS 29.165 and came to the conclusion that it would be feasible to make TS 29.165 compatible with JT-Q3401. We then began work in 2010 on incorporating all stipulations included in JT-Q3401 but not in TS 29.165 and that are not Japan-specific into TS 29.165.

We also proposed two work items to 3GPP as the rapporteur for clarifying the setting condition of the header fields/parameters on inter-IMS NNI (II-NNI) and option items facilitating the inter-operator agreements on the interface, which were accomplished in 2013. This means we completed the preparation for migration from JT-Q3401 to the new domestic II-NNI standard based on TS 29.165 [1]. Moreover, we have successfully reflected the emergency call service [2], which was a Japan-specific requirement, in 3GPP documents, so making a specification of domestic emergency calls based on 3GPP standards also became feasible.

We are proposing contributions related to the feedback from the developments of session border controller (SBC) or other IMS nodes for IP interconnection and commercial operations, e.g., addition of the II-NNI condition for new supplementary services, clarification of the methods of restoration detection on the SIP layer, correction in setting conditions of the SIP header fields used for inter-operator charging, and amendments of descriptions related to the handling of calling party number in ISUP-SIP interworking. Restoration detection on the SIP layer is explained below as an example of our proposals.

Operators will set up gateway nodes called SBCs, which are equivalent to the interconnection border control function (IBCF) in the IMS architecture, at the point of interconnection of their IP telephone networks. In domestic standards, a method using the




DNS: domain name system
ENUM: E.164 number to uniform resource identifier mapping
GSMA: GSM Association

Fig. 3. Relation between standard documents of IP interconnection.

OPTIONS request is specified for detecting restoration of SBCs on the SIP layer after detecting faults in the SBCs. With this method, an SBC detecting the fault of the opposite SBC sends OPTIONS requests periodically and detects the restoration by receiving a successful response from the peer. This method uses the OPTIONS request only in the inter-SBC hop; however, 3GPP specifications define only end-to-end procedures for the OPTIONS request. Because the restoration detection method in domestic standards is useful for reducing downtime and operational errors and is being used for private IP interconnection in some operators, we proposed that the inter-SBC method be allowed explicitly. The proposal was agreed, so the TTC standards could maintain consistency with the 3GPP standards.

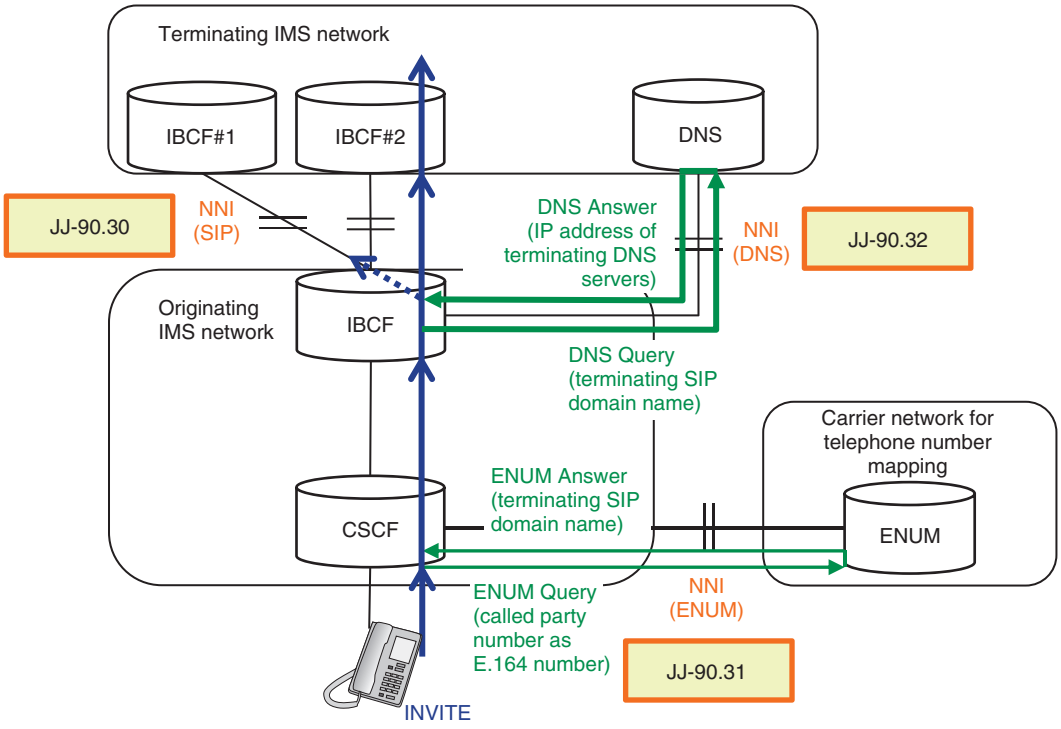
NTT has also contributed to the 3GPP Core Network and Terminals (CT) Working Group (WG) 3, which considers interworking with external networks, as a vice chair since 2013. The activities have resulted in the increasing presence of NTT.

## 5. Overview of NNI technical specifications in Japan

TTC has published JJ-90.30 as the common interconnection interface specification among IMS operator networks in Japan. This document specifies the II-NNI SIP signaling conditions based on 3GPP TS 29.165. In addition to JJ-90.30, TTC has published JJ-90.31 as the common interconnection interface specification for carrier ENUM (E.164 number to uniform resource identifier mapping), JJ-90.32 as the common interconnection interface specification for SIP domain name resolution with domain name systems (**Fig. 4**), JJ-90.27 as the interconnection interface specification for call transfer and diversion, and JJ-90.28 as the interconnection interface specification for emergency calls. JJ-90.31 and JJ-90.32 are specified based on GSM Association IR.67, and JJ-90.27 and JJ-90.28 are based on 3GPP specifications.

## 6. Future prospects

Since current TTC standards refer to 3GPP Release




CSCF: call session control function

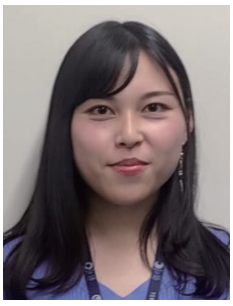
Fig. 4. Relation between domestic standard documents of IP interconnection.

15, we are working on updating the reference to 3GPP Release 16. The progress of 3GPP Release 16 was delayed due to the impact of the COVID-19 pandemic but was finally completed and the specification was frozen in July 2020. We have examined all the changes in Release 16 and applied necessary updates to TTC standards including JJ-90.30. The revised TTC standards will be published in November 2020. Our work at 3GPP is reflected in domestic standards such as this process. Inter-operability testing (IOT) of IP interconnection between domestic operators based on the revised TTC standards is planned to start at the end of this year, and commercial IP interconnection will start in 2021. We will continue to clarify the specifications and make them more practical by taking into account the feedback from IOT.

## References

[1] S. Sakaya, K. Arai, and S. Kanegae, "Standardization Efforts in IP Interconnect Specifications" NTT Technical Review, Vol. 12, No. 9, Sept. 2014.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201409gls.html

[2] M. Ogawa and T. Jocha, "Standardization Trends of the Emergency Call and Standardization Efforts for IP Interconnection," NTT Technical Journal, Vol. 30, No. 2, pp. 57–61, Feb. 2018 (in Japanese).
https://www.ntt.co.jp/journal/1802/files/JN20180257.pdf



**Haruka Eitoku**
Researcher, NTT Network Service Systems Laboratories.

She received an M.S. in physics from Kyushu University, Fukuoka, in 2017. In the same year, she joined NTT Network Service Systems Laboratories, where she researched interconnection between IMS networks. She has been contributing to standardization efforts in 3GPP CT WG1 and CT WG3 since 2018. She is currently researching interconnection between IMS networks.

# Event Report: NTT Communication Science Laboratories Open House 2020

*Koji Kamei, Naotoshi Abekawa, Tomoki Ookuni, Jun Muramatsu, Hiroshi Ueda, and Masataka Sawayama*

**Abstract**

NTT Communication Science Laboratories Open House 2020 was held online, the content of which was published on the Open House 2020 web page at noon on June 4th, 2020. Within the first two days, around 2000 visitors enjoyed 5 talks and 31 exhibits, which included our latest research efforts in information and human sciences.

*Keywords: information science, human science, artificial intelligence*

## 1. Overview

NTT Communication Science Laboratories (CS Labs) aims to establish cutting-edge technologies that enable heartfelt communication between people and people and between people and computers. We are thus working on a fundamental theory that approaches the essence of being human and information science, as well as on innovative technologies that will transform society. NTT CS Labs' Open House is held annually with the aim of introducing the results of our basic research and innovative leading-edge research with many hands-on intuitive exhibits to those who are engaged in research, development, business, and education.

Open House 2020 was, however, held via our website considering the recent situation against the spread of the novel coronavirus. The latest research results were published with recorded lecture videos on the Open House 2020 web page at noon on June 4th when the event was originally planned to start [1]. The content attracted many views in a month not only from NTT Group employees but also from businesses, universities, and research institutions. The event content is still available.

This article summarizes the event's research talks and exhibits.

## 2. Keynote speech

Dr. Takeshi Yamada, vice president and head of NTT CS Labs, presented a speech entitled "I want to know more about you – Getting closer to humans with AI and brain science –," in which he looked back upon the history of NTT and establishment of NTT CS Labs then introduced present and future cutting-edge basic research and technologies (**Photo 1**).

Dr. Yamada pointed out the mission of NTT CS Labs to promote heartfelt communication between people and people or people and computers, by placing particular importance on pursuing artificial intelligence (AI) technologies that "approach and exceed human abilities" and science research to "obtain a deep understanding of people" and "make heartfelt contact." Regarding today's situation in which physical contact and face-to-face communication are significantly restricted, the talk pointed out that it has become even more important to identify the essence of emotional contact in heartfelt communication. The talk introduced NTT CS Labs' recent AI technologies


"I want to know more about you – Getting closer to humans with AI and brain science –"

Photo 1. Keynote speech (Dr. Takeshi Yamada).

from three perspectives of *getting closer to humans with voice and acoustic processing, understanding humans with language*, and *heartfelt communication with empathy and happiness (shi-awase)* then declared to continue pursuing the essence of communication and tackle new challenges boldly and persistently.

## 3. Research talks

The following three research talks highlighted recent significant research results and high-profile research themes. Each talk introduced some of the latest research results and provided some background and an overview of the research.

(1) "Communication with desired voice – Deep generative model opens the way to innovative speech transformation –," by Dr. Kou Tanaka, Media Information Laboratory

Dr. Tanaka introduced many technologies that apply deep learning, which is a common concept in AI, to speech information processing and various speech-to-speech conversions with high quality. NTT CS Labs considers speech information as an important tool for communication as well as the following four points as particularly important requirements for its conversion technology: 1) high quality, 2) learning from a small amount of data, 3) convertible in real time, and 4) covering various feature conversions such as voice characteristics, prosody, and accent. The research and development of such speech-to-speech conversion technology targets real-world application such as assistance for those with vocal disabilities, conversion of speaking style including emotion, and conversion of pronunciation and accent in language learning (**Photo 2**).


Photo 2. Research talk (Dr. Kou Tanaka).

(2) "Towards understanding human skin sensations – How the brain integrates spatio-temporal information in touch –," Dr. Scinob Kuroki, Human Information Science Laboratory

Dr. Kuroki explained information processing of human skin sensations. She introduced a psychophysical experiment in which a human is presented with a tactile stimulus and asked the timing and position of the stimulus. A mysterious phenomenon

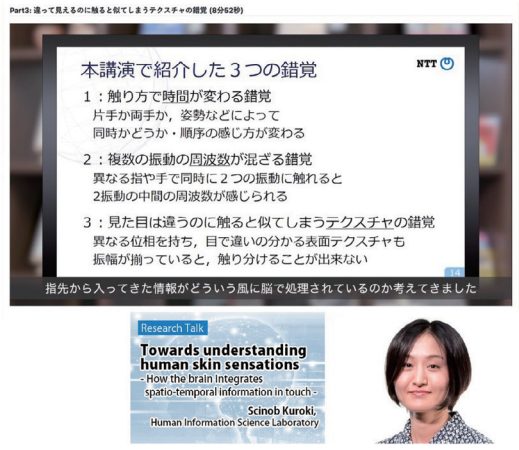



Photo 3. Research talk (Dr. Scinob Kuroki).


Photo 4. Research talk (Dr. Sanae Fujita).

occurred in which the answers differed from the stimulus given. The brain causes illusions by interpreting physical input stimuli, but by carefully unraveling such phenomena and illusions, she explored the mechanism by which the brain processes tactile information. She also introduced advanced approaches such as designing and generating tactile stimuli using a three-dimensional (3D) printer based on a method inspired by human visual information processing systems and a new direction of haptic science (**Photo 3**).

(3) "Which word is more difficult for you, 'car' or 'vehicle'? – Estimation of text readability and human vocabulary size –," Dr. Sanae Fujita, Innovative Communication Laboratory

Dr. Fujita introduced studies of two estimation methods, text readability and human vocabulary-size estimation. The estimation of human vocabulary size is based on the word familiarity database that NTT CS Labs has been creating for over 20 years. The database has recently been updated with more than 160,000 words and the Reiwa version of the vocabulary-size estimation test was published and made available as a demonstration [2]. She also introduced efforts aimed at tailor-made education support such as recommending books suitable for individual learners by combining vocabulary-size estimation and text-difficulty estimation (**Photo 4**).

## 4. Research exhibition

The Open House featured 31 exhibits displaying NTT CS Labs' latest research results. We categorized them into four areas: *Science of Machine Learning*, *Science of Communication and Computation*, *Science of Media Information*, and *Science of Humans*. Each exhibit prepared presentation slides with recorded explanation and published on the event web page (**Photo 5**). Several provided online demonstrations or demo videos to make them closer to direct demonstrations. The following list, taken from the Open House website, summarizes the research exhibits in each category.

### 4.1 Science of Machine Learning

- People of the WWW, give us your computation! – Generating datasets using people and information of the WWW –
- Presenting a quick solution to system failures – Generating recovery-command sequences by neural networks –
- Refining spatially aggregated data from cities – Multivariate Gaussian processes for spatially aggregated data –
- Fast inference of accurate anomaly detector – Transfer anomaly detection for unseen datasets –
- Anomaly detection with low false-positive rate – Semi-supervised learning for maximizing partial AUC –
- Is the data really biased? – Testing combinatorial correlation by decision diagrams –

### 4.2 Science of Communication and Computation

- What happens if every player rushes selfishly


Photo 5. Exhibition web page.

– Equilibrium computation of congestion games –

- Handle a huge quantum world through a tiny window
  – Investigation of the ability of indirect quantum controls –
- Tuning machine translation with small tuning data
  – Domain adaptation with JParaCrawl, a large parallel corpus –
- Assessing children's emotional development
  – Investigating developmental changes via multiple cues –
- Creating a personalized picture book
  – Support for parent-child picture book interaction –
- How many words do you know?
  – Vocabulary-size test, Reiwa edition –
- Kyomachi Seika will guide you!
  – Training the role-playing AI with community cooperation –
- What does he/she think in this situation?
  – Sentiment text generation based on personality –

### 4.3 Science of Media Information

- Can you guess the age from the voice?
  – Deep speaker attribute estimation with speaker clustering –
- More wireless microphones are available in a room
  – BRAVE: bit-error-robust low-delay audio and voice encoding –
- Pay attention to the speaker you want to listen to (II)
  – Neural selective hearing with audio-visual speaker clues –
- Controlling voice expression using face expression

– Crossmodal voice expression control –
- Learning to search like humans
  – Adaptive spotting for efficient object search –
- Deep learning without data aggregation from nodes
  – Asynchronous consensus algorithm for ML (machine learning) over various NWs (networks) –
- Cardiac model that makes it heart
  – Gaussian process with physical laws for 3D cardiac modeling –
- Listening carefully to your heart beat
  – Cardiohemodynamical analysis based on stethoscopic sounds –

### 4.4 Science of Humans

- Make natural-looking illusions by perceptual model
  – Adaptive motion retargeting for illusion-based projection AR (augmented reality) –
- Tiny eye movements reflect cognitive states
  – Relation of eye-movement dynamics with cognition and pupil –
- Haptic metameric textures
  – Direct control of perceived texture of 3D printed stimuli –
- What causes emotional change?
  – Monitoring emotion in experimental settings and daily life –
- Special cognitive abilities of e-sports experts
  – Performance, physiological state, and brain activity –
- Realizing harmony in rugby scrum
  – Easy assessment of player coordination with wearable sensors –
- What is a "straight" ball?
  – Physical and perceptual attributes of a pitched ball –
- Body representation for quick and skillful action
  – Uncertainty of hand-state estimate regulates stretch reflex –
- Unconscious is smarter than conscious
  – Environmental dependency in visuomotor responses –

## 5. Special lecture

We asked Professor Noriko Osumi, vice president of Tohoku University, to give a special lecture entitled "A challenge to scientifically understand 'individuality.'" It is well known that people with autistic spectrum disorder sometimes make considerable achievements in fields such as art and scientific research. Her research group has regarded atypical development as individuality, and investigated the genetic and non-genetic mechanisms underlying such atypicality using mice as model animals. In particular, she found that paternal aging affected the early development of vocal communication of pups and introduced cutting-edge research topics of molecular mechanisms with changes in epigenome information. "Science of individuality" must be required in a diversity society. The lecture finished with providing a grand perspective from "ideal form of society" to "human evolution."

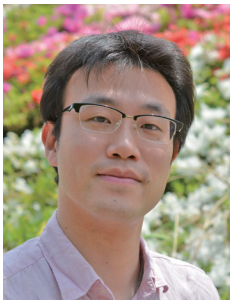
## 6. Concluding remarks

Unlike last year, Open House 2020 was not held as an event at a physical venue but instead to present our latest results on a website. The lecture videos were viewed more than 10,000 times in June from various segments of users. Though it was difficult to have lively discussion, the long-term growth of visitors encourages us in further research activities. In closing, we would like to offer our sincere thanks to all the participants of this online event.

## References

[1] Website of NTT Communication Science Laboratories Open House 2020, http://www.kecl.ntt.co.jp/openhouse/2020/index_en.html
[2] A Japanese Word Familiarity Database (in Japanese), http://www.kecl.ntt.co.jp/icl/lirg/resources/goitokusei/

**Koji Kamei**
Senior Research Scientist. Learning and Intelligent Systems Research Group, Innovative Communication Laboratory, NTT Communication Science Laboratories.
He received a B.E. and M.E. in electronic engineering from Kyoto University in 1995 and 1997. He joined NTT in 1997 and has been engaged in research on cooperative systems and information sharing and recommendation. He was also engaged in research on cooperative robotic services at Advanced Telecommunications Research Institute International (ATR) from 2009 to 2019.

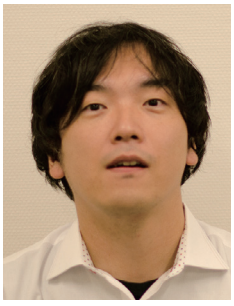


**Naotoshi Abekawa**
Distinguished Researcher, Human and Information Science Laboratory, NTT Communication Science Laboratories.
He received a B.E. from Tokyo Metropolitan University, an M.E. from Tokyo Institute of Technology, and a Ph.D. from Kyoto University, in 2003, 2005, and 2013. He joined NTT in 2005 and has been engaged in research on human information processing. His research interests include human sensorimotor mechanisms, especially visuomotor control properties. He is a member of the Society for Neuroscience, the Japan Neuroscience Society, the Japanese Neural Network Society, and the Institute of Electronics, Information and Communication Engineers (IEICE).

**Tomoki Ookuni**
Senior Research Scientist, Research and Planning Section, NTT Communication Science Laboratories.
He received a B.A. in economics from Kyoto University in 1990 and an M.A. in marketing science from Osaka Prefecture University in 2004. He joined NTT in 1990 and has been engaged in managing research and development at both NTT WEST and NTT since 1998. He joined NTT Communication Science Laboratories in 2019.

**Jun Muramatsu**
Research Scientist, Learning and Intelligent Systems Research Group, Innovative Communication Laboratory, NTT Communication Science Laboratories.
He received a B.S., M.S., and Ph.D. in mathematics from Nagoya University, Aichi, in 1990, 1992, and 1998. He joined NTT in 1992 and has been engaged in research on information theory. From Feb. 2007 to Feb. 2008, he was a visiting researcher at ETH, Zurich, Switzerland. From 2006 to 2010, he was an associate editor of IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences. He received the Young Researcher Award of SITA (the Society of Information Theory and Its Application) in 2003 and the 63rd Best Paper Award of IEICE in 2007. He is a member of IEICE and the IEEE Information Theory Society

**Hiroshi Ueda**
Research Scientist, Kashino Diverse Brain Research Laboratory, NTT Communication Science Laboratories.
He received a B.S. from University of British Columbia in 2008, an M.S. from Tokyo Institute of Technology in 2011, and a Ph.D. from The University of Tokyo in 2014. He joined NTT Communication Science Laboratories in 2015. His current research interests include various topics related to human visuomotor control.



**Masataka Sawayama**
Research scientist, Human and Information Science Laboratory, NTT Communication Science Laboratories.
He received a Ph.D. in experimental psychology from Chiba University in 2013 and joined NTT Communication Science Laboratories the same year. He is an expert in psychophysical research on human visual processing, in particular color and material perception. He is a member of the Vision Sciences Society and the Vision Society of Japan.

# External Awards

### MEF 2020 Top Contributor Award
**Winner:** Hiroki Baba, NTT Network Technology Laboratories
**Date:** July 27, 2020
**Organization:** MEF (Metro Ethernet Forum)

For his contribution to MEF work, including MEF 3.0 Proofs of Concept and a white paper "Slicing for Shared 5G Fronthaul and Backhaul."

### Outstanding Research Award
**Winners:** Shinobu Saito and Yukako Iimura, NTT Software Innovation Center; Emad Aghayi and Thomas D. LaToza, George Mason University
**Date:** September 3, 2020
**Organization:** Information Processing Society of Japan (IPSJ) Special Interest Group on Software Enginering

For "Can Microtask Programming Work in Industry?".
**Published as:** S. Saito, Y. Iimura, E. Aghayi, and T. D. LaToza, "Can Microtask Programming Work in Industry?", arXiv:2009.05207, Sept. 2020.

### JSAP Fellow
**Winner:** Akira Fujiwara, NTT Basic Research Laboratories
**Date:** September 8, 2020
**Organization:** The Japan Society of Applied Physics (JSAP)

For his study on ultimate control of electrons using silicon nanodevices.

# Papers Published in Technical Journals and Conference Proceedings

### Influence of Condensation Due to Temperature Difference on Corrosion Process of Exposed Reinforcement Structure
H. Kasahara, N. Fujimoto, and Y. Okamura
Journal of the Society of Materials Science (Zairyo), Vol. 69, No. 7, pp. 539–546, July 2020.

The environment in a reinforced concrete communications manhole is always very humid, and the exposed steel bars can easily become corroded. However, since the depth below ground of these manholes is relatively shallow and the annual temperature change is modest compared to that for above-ground structures, it is assumed that the factors controlling the progress of steel bar corrosion are different than those for other structures. Moreover, manholes basically represent an isolated environment, and there are no additional substances such as sea salt present compared to the outside air. In this relatively static environment, it is thought that the amount of water vapor is the main cause of change in corrosion rate. Therefore, in this study, in order to clarify the supply mechanism of water that promotes steel bar corrosion in manholes, the relationship among the actual manhole temperature, humidity, and the corrosion rate was investigated. The results show that in winter, the temperature of the manhole ceiling is lower than the dew point temperature, so condensation occurs and steel bar corrosion progresses. Based on repeated temporal experiments, we found that the increase in condensation due to the difference between the dew point temperature and the temperature of the ceiling must be considered in order to explain the steel bar corrosion rate in manholes.

### Luminance Distribution and Monocular Depth Perception by Smooth Motion Parallax in Visually Equivalent Light Field 3D Display Using Linear Blending Technology
R. Kamada, H. Mizushina, M. Date, S. Shimizu, and S. Suyama
IMID (International Meeting on Information Display) 2020 DIGEST, Vol. 9, p. 1598, August 2020.

To obtain a realistic three-dimensional (3D) display with a high degree of realism without the need for 3D glasses, it is necessary to represent continuous motion parallax caused by changes in the viewing position. However, to present continuous motion parallax by using a multi-view display, a sufficiently narrow viewpoint interval and large amount of parallax image data are required. To solve this problem, the visually equivalent light field 3D (VELF3D) display with linear blending technology has been proposed. We estimated the precise luminance distribution of the VELF3D display and evaluated monocular depth perception by motion parallax.

### Vocal-tract Spectrum Estimation Method Affects the Articulatory Compensation in Formant Transformed Auditory Feedback
Y. Uezu, S. Hiroya, and T. Mochida
Acoustical Science and Technology, Vol. 41, No. 5, pp. 720–728, September 2020.

Auditory feedback has a crucial role in stably controlling speaking

and singing. Formant transformed auditory feedback (TAF) is used to investigate the relationship between perturbation to the formant frequency and the compensatory response to clarify the mechanism of auditory-speech motor control. Although previous studies on formant TAF applied linear predictive coding (LPC) to estimate formant frequencies, LPC estimates false formants for high-pitch voice. In this paper, we investigated how different vocal-tract spectrum-estimation methods in real-time formant TAFs affect the compensatory response of formant frequencies to perturbations. A phase-equalization-based autoregressive exogenous model (PEAR) is applied to the TAF system as a formant estimation method that can estimate the formant frequency more accurately and robustly than LPC. Fifteen Japanese native speakers were asked to repeat the Japanese syllables /he/ or /hi/ while receiving feedback sounds whose formants F1 and F2 were transformed. From the results for the /he/ condition, the F1 compensatory response for PEAR was significantly larger than that of LPC, and the compensation error in the F1–F2 plane for PEAR was less than that for LPC. Our results suggest that PEAR can increase both the accuracy of formant frequency estimation and the naturalness of the transformed speech sound.

**Classification of Formant Estimation Methods in Transformed Auditory Feedback Experiments Using Convolutional Neural Networks**

F. Taguchi1, S. Hiroya, Y. Uezu, and T. Mochida

Acoustical Science and Technology, Vol. 41, No. 5, pp. 800–803, September 2020.

We classified formant transformed auditory feedback experimental data directly comparing phase-equalization-based autoregressive exogenous model and linear predictive coding by using a convolutional neural network. We found that the average identification rate was significantly high in the case of two-channel input of Utter (the speech signals from the microphone) and Trans (the transformed speech signals). This suggests that speech prediction is important in determining the naturalness of feedback speech.