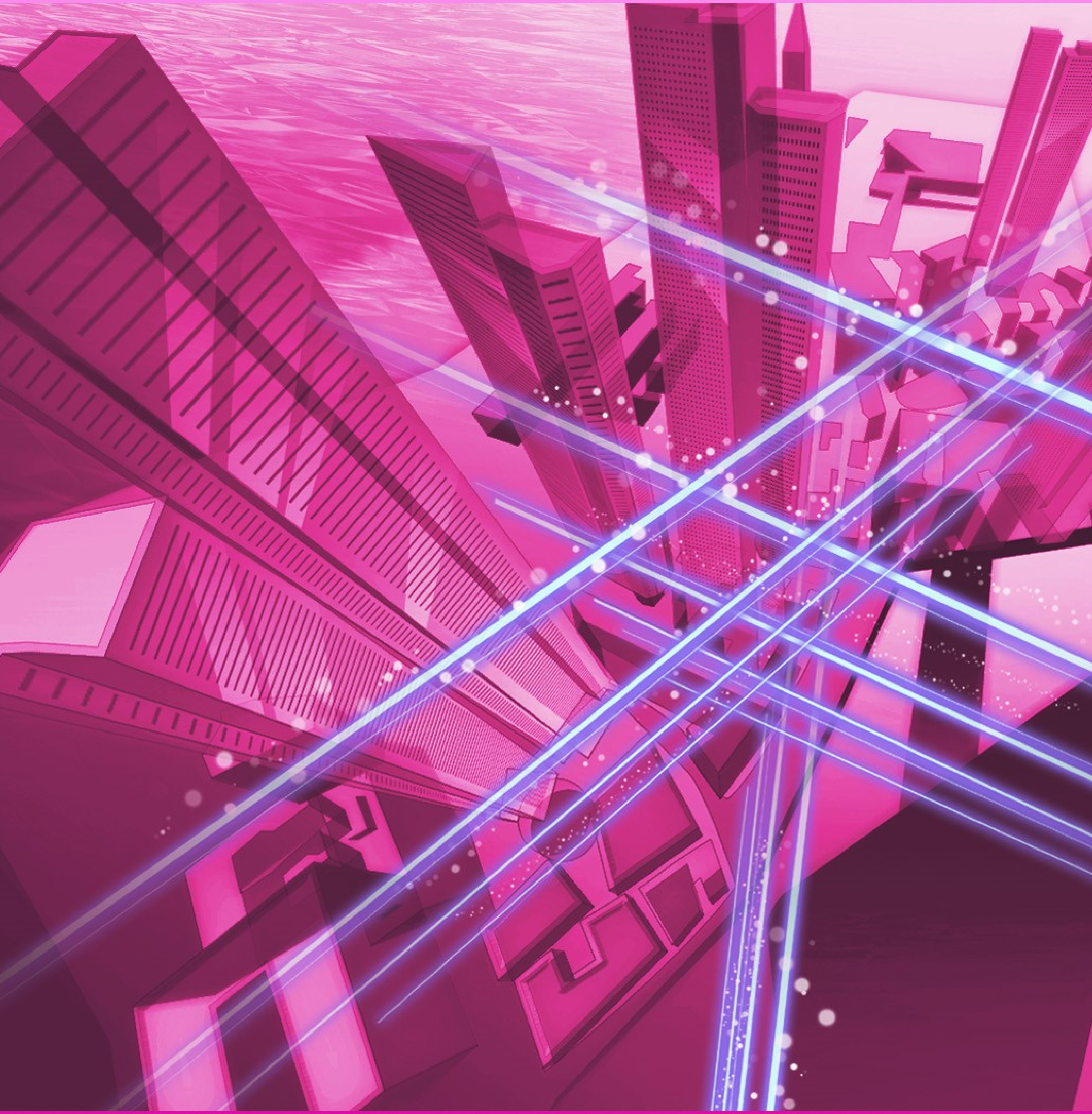
NTT Technical Review

11
2019



November 2019 Vol. 17 No. 11

NTT Technical Review

November 2019 Vol. 17 No. 11

View from the Top

Feature Articles: Communication Science for Achieving Coexistence and Co-creation between Humans and AI

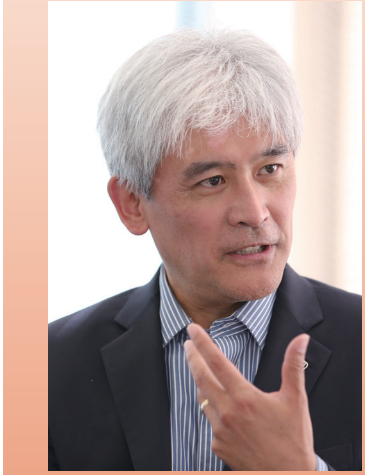
Regular Articles

Global Standardization Activities

Information

External Awards/Papers Published in Technical Journals and Conference Proceedings

At the Epicenter of Innovation—How an Open Mind Fundamentally Changes How We Define “The Next Big Thing”



Kazuhiro Gomi
President and CEO, NTT Research, Inc.

Overview

NTT Research, Inc. was established in Palo Alto, California—in the heart of Silicon Valley, USA—in April 2019. It comprises three laboratories, 1) Quantum science and computing, 2) Medical and health informatics, and 3) Cryptography and information security. The goal is to build a new human-resource ecosystem in the areas of advanced basic research that can fundamentally change the way we live and the way we work.

At the opening event held in July 2019, substantial encouragement and support were evident among the many notable guests from academia and the business worlds. Kazuhiro Gomi, President and Chief Executive Officer of NTT Research, Inc., a leader who thoroughly understands the importance of human networks, was asked what the prospects are for NTT Research, Inc.

Keywords: NTT Research, Inc., research and development, digital twin computing

A grand opening encouraged by both academia and business

—First, congratulations on the success of the grand-opening reception. I felt the enthusiasm for your endeavor.

I’d like to express my sincere appreciation to those guests who came to the opening event, including representatives from Stanford University and other leading institutes, as well as renowned business leaders—many of whom are NTT’s clients. This reception confirms my belief that NTT Research, Inc. will excite and inspire both academia and industry.

At the event, I received wonderful feedback. Many guests expressed their appreciation for NTT’s long-term commitment to work that can fundamentally address various social problems. We’ll do this by empowering researchers and revitalizing their research, which is not necessarily an approach that many other companies would dare to take. I do not wish to be incautious, but there are clearly significant expectation and excitement surrounding our efforts.

—Tell us about your management policy.

Our core mission is to create intellectual properties related to novel technologies that will fundamentally change the way society works. I want to approach this by promoting a sort of “chemistry” between the existing

NTT Group's research and development (R&D) activities and the global talents here in the Silicon Valley.

As a corporate research institute that conducts R&D using the funds allocated from each NTT Group operating company, NTT Research, Inc. should increase the value of NTT Group as a whole through basic and fundamental research activities. By sponsoring the world's leading researchers and scientists, whose great vision and insights will deliver cutting-edge technologies and inventions, NTT Research, Inc. represents the depth of NTT Group's commitment. Of course, the outcomes from those highly accomplished and motivated researchers are what will ultimately define our core values.

—Why did you choose to be based in Palo Alto?

Palo Alto, located in the heart of Silicon Valley, is the ideal location for NTT Group and NTT Research, Inc. Having Stanford University and an entire ecosystem for innovation, Palo Alto is the birthplace of innumerable new technologies. It's a unique atmosphere that makes it easy for people to come together and develop new ideas. I feel that the "air" here is different. We have many potential collaborators and partners here who are the best researchers, scientists, engineers, and visionaries in the world—all within a business environment that supports their efforts. Having said that, it seems natural to expand our activities to a place like Silicon Valley.

Three research laboratories to strengthen our global business competitiveness

—NTT Research, Inc. has three laboratories. Tell us what role each of them plays and what each one is developing.

The Physics & Informatics Laboratories explores the interdisciplinary field between physics and informatics, in which NTT laboratories have accumulated our own technologies. At this laboratory, we conduct basic physics research, especially on quantum theories concerning dissipative systems, as well as research to establish completely new theories on applications of quantum theories to information processing. As for basic research on cryptography and information security to build a safe and secure future, the Cryptography & Information Security Laboratories studies cryptography for advanced functions and safety theories concerning scattered environments (such as those involving blockchain). In addition, the Medical & Health Informatics Laboratories are applying technologies such as artificial intelligence for analyzing biological information in a manner that leads to precision medicine. The directors of these three laboratories are world-renowned and proven experts. Indeed, we are leveraging their human networks to form new research teams for collaboration with NTT researchers in Japan. By adding other outstanding researchers around the world to the mix, we'll empower the unique technologies accumulated by NTT laboratories.

Recently, NTT R&D released a visionary theme called Innovative Optical and Wireless Network, or IOWN for short. IOWN is the overall concept of innovative optical-based networks and related technologies that NTT is currently investigating. One example of what is possible with IOWN is digital twin computing (DTC), which accurately reproduces the physical world in cyberspace. This permits various simulations and testing that make it possible to precisely predict the state of physical substances. For example, if you apply this concept in the medical field, DTC will allow us to predict the types of diseases that a particular individual may develop if they stay on the same diets, habits, sleep patterns, etc—all based on analysis of their digital twin. In other words, the precise information associated with a real individual will be used to model a duplicate of that individual in cyberspace. When such models are effectively created, they allow us to precisely prescribe the correct medical interventions for that individual.

This approach is a bit like a human version of the weather forecast. Today, weather predictions have become more accurate, even to the point of knowing when rain will begin to fall—within a range of just minutes—in a particular geography. This precision is achieved by gathering information from a number of sensors at the national and global levels and utilizing massive computing power to analyze the information via sophisticated models. The digital twin metaphor I mentioned earlier is similar to modern weather forecasting. The only difference is that we apply this approach to human bodies. This will ultimately improve the health and life expectancy of all humankind. To achieve such a world, a huge amount of processing power is required. The issue of privacy must be seriously addressed along the way. Moreover, we need to figure out how to collect and accumulate medical information and how to model human bodies in the cyberspace.

Apart from DTC and its application in the medical field, new types of cryptography are attracting attention. Today, encryption is used to prevent information theft or to prevent stolen information from being read (i.e., encryption). However, as we exchange more and more data throughout our daily lives, the meaning of encryption will change. Access to information may no longer revolve around a single “key.”

Different kinds of keys may begin to allow targeted access to predefined portions of information—all based on the type of key provided. It will no longer be an “all or nothing” scenario in terms of accessing information. One can create special keys based on the profiles of different readers. This allows adjusted access per individual. Researchers at NTT Research, Inc. are currently at work on this type of encryption. As such technologies mature, I expect significant impact to our society.

We foresee quantum computing, next-gen encryption technologies, and medical and health informatics as crucial to realizing DTC. Of course, these areas of focus—studied by our three laboratories—are challenging. Therefore, we are required to make long-term commitments before they yield meaningful outcome. Also, some hurdles may not be overcome just by applying the results of our own basic research. However, combining these three themes will surely get us closer to achieving the final goal of Bio DTC. The people who attended our opening event are excited about our work and stated direction, so I feel we are on the right track.

—With such vison in mind, tell us what should we expect from R&D.

My first assignment when joining NTT was R&D (voice recognition related technology development). However, I have been on the business side for the last 18 years. Rather than conducting or leading R&D activities, I was in a position to leverage the results coming from R&D within our business practices. The business-side expectations for technologies, products, and services coming out of R&D are vast. In other words, I recognize the results of our research as something that provides NTT with substantial advantage in the world of business competition. Also, the reputation of NTT as a company with unequaled research capabilities and long-term vision helps NTT sustain the trust and respect of our partners and customers.

Today, information technology (IT) is probably the most important source of innovation for almost all companies, regardless of what market they play in. Missteps or ignorance on IT may, with high probability, create fatal issues for a company. For many corporate leaders, therefore, selecting the most trustworthy IT partner is not only critical but strategic.

But what kind of partner is “trusted?” From a customer’s point of view, a trusted partner is a player you want to deal with for a long time, whose value grows rather than diminishes with each passing day. To that end, we *need* to be a company with a robust R&D team. We *need* to have vision and execution capabilities. I want NTT Research, Inc. to become a key propeller of the sense of trust NTT Group enjoys in the market. To gain and sustain that reputation, a solid vision with foresight—while broadening our field of view through outstanding researchers and R&D teams—is essential. As importantly, we must produce excellent results.

Incidentally, when I was interviewed by magazines several years ago, I called myself a “technology geek." Although I believe I still am, when I see presentations by the researchers at NTT Research, Inc. or I talk with them about their work, I feel they are a hundred times more enthusiastic than I have ever been. On one hand, I am thrilled that we are able to launch NTT Research, Inc. with such an exciting and talented group of people. On the other hand, I feel embarrassed a bit having called myself a technology geek (laugh).

Being open minded: start by accepting anything

—You have faced competitors not only in Japan and the United States but also around the world. Are there any lessons learned from these previous experiences?

We are living in an ever more globalized society. What I've been feeling while doing business in the United States is that many companies in this market are represented by people from many different nationalities and ethnic groups. I like soccer, so let's take soccer as a metaphor. When playing against a team populated by the best players from many different nations, Japan's national team, even if it represents the best players in all of Japan, may fall short of victory. I feel strongly that we have to form a team by selecting the best players from all over the world. Coaches or managers of such teams must know how to assemble these players, to inspire them and unify their focus. To win games, our managers must be prepared to analyze, judge, and lead, while remaining open-minded to the diverse perspectives and unique talents of every player.

Needless to say, if you work based solely on a general idea of "common sense" for Japan, your teammate from a different background may not be as happy (or effective) using that approach as you are. In other words, because your teammate comes from a different culture with a different concept of common sense, your proposal, opinion, expression, etc., may not be taken as you might have intended. If you feel uncomfortable with your teammate or that something is strange, you should not be dismayed or defensive. Rather, take a deep breath and pause. Determine why what was said was said and find the real intention behind the words. Maintaining an open mind to differences is crucial to teamwork. Look for the potential and the positive in alternative points of view.

I learned this lesson while working in the United States, where I've lived for a long time. There, society as a whole feels more open to differences, and those who are socially successful are usually most open-minded. I believe this is an insight we must value within any diverse culture.

—What can we do to encourage researchers to develop this open-minded perspective?

First of all, I recommend getting out of the laboratory in Japan. Go work with people of various cultures and backgrounds. Please seize any opportunity, especially as young researchers, to step outside of your box even for a short period of time.

Outside of their labs, I'd like them to actively interact with people who make them a bit uncomfortable. At the same time, you have to be cautious. I myself haven't succeeded in everything at every occasion. Rather, I've had bitter experiences that I don't like to talk about even now. You should not follow other people blindly; trust your own instincts. This is especially true in the US business scene where contracts are crucial. We must represent our own perspectives with respect but also with strength—to strive for a sense of balance between trusting and doubting.

Throughout my own experiences, setting goals for myself has helped me grow. Finding a mentor can be invaluable, someone who gives sensible advice from time to time. Ideally, a mentor should be readily accessible. However, even if your mentor is far away, it's important to have someone you can model yourself after, someone who inspires you to become your best self. For me, former IBM chairman Louis V. Gerstner was such a role model. I still remember listening to his speech at Telecom 1999 in Switzerland. When I heard his speech, I said to myself, "I want to be able to deliver a message as strongly and convincingly as he does." That was 20 years ago.

I want the world to say, "NTT Research, Inc. is amazing!"

—Your speech at the opening event was very impressive.

Thank you for saying so. As for my presentation, I tried to make a speech like Mr. Gerstner would have done. The presentation that inspired me back in 1999 was about the e-business that IBM was promoting at the time. He didn't use any slides; instead, he went on stage and just talked without looking at any materials. Talking about e-business for almost an hour, he mesmerized everyone at the venue. At that time, I thought "This is cool; this is it!" That's why at our opening event, I presented in the "Gerstner style."

—I felt that you had already established your "Kazu-style presentation." Now, NTT Research, Inc., which was introduced in your Kazu-style presentation, is finally in full swing. How will you proceed from now onwards? Tell us your aspirations.

"Kazu-style"? Thanks for the compliment! As of July 9, NTT Research, Inc. was officially launched with about 30 people. Among those members, approximately 20 are researchers, all of whom have doctorates. More than half are star players with the title of professors. While I use their titles, such as doctors and professors, as a means to describe the superiority of NTT Research, Inc. members, this is not just about titles. They all have outstanding talent and track records. I am glad that we are starting this new endeavor with these talented people. In a sense, I feel that I am very lucky. In return, though, to meet the expectations of these ambitious researchers, I'm determined to build NTT Research, Inc. with a clear vision in mind. Currently, we're conducting research in the three areas I mentioned earlier from our base in Silicon Valley. However, it is necessary to expand our research focus and the footprint of our activities in the future. To that end, I will create an environment that permits researchers and other staff to work together comfortably and enthusiastically. I want to accelerate this important work and produce results so that the world will say "NTT Research, Inc. is amazing!"

Interviewee profile

■ Career highlights

Kazuhiro Gomi joined NTT in 1985. Before taking up his current post in April 2019, he served as Vice President (VP) of the Global Business Department of NTT Communications from 2001 to 2004, after which he served as VP of the Global IP Network Business Unit of NTT America (2004 to 2009), Chief Operating Officer of NTT America (2009 to 2010), and President and Chief Executive Officer of NTT America (2010 to 2019).

Processing Like People, Understanding People, Helping People—Toward a Future Where Humans and AI Will Coexist and Co-create

Takeshi Yamada

Abstract

Artificial intelligence (AI) has been making remarkable progress in recent years and has even been approaching the level of human performance for certain functions, but it still has its limitations. In contrast, human beings are highly advanced and complex, which is why they are also imperfect and prone to mistakes as reflected by their vulnerability to bias and illusions. This article introduces NTT initiatives in communication science to bring AI technology closer to a human level and to develop an even deeper understanding of human beings with the aim of closing the gap between AI and humans and achieving AI that can help people.

Keywords: artificial intelligence, communication science, brain science

1. Introduction

Recent developments in artificial intelligence (AI) have been truly remarkable. In the beginning, computers were especially good at performing batch processing of large amounts of data that humans could not process and at performing high-speed processing on behalf of humans for tasks that humans were weak at. However, thanks to recent advances in deep learning, computers are approaching—and surpassing in some cases—human abilities in areas where they have long been behind, for example, speech and image recognition and natural language processing that humans are inherently good at. In the future, we can expect progress in AI to accelerate in this area of media processing.

Nevertheless, neural processes are complex, with many of them still unexplained. It is said that a level of AI performance exceeding the abilities of the complex human brain still lies somewhere in the future. In contrast, humans, though being very advanced and complex creatures, appear at first glance to be imperfect since they can make mistakes under the influence of cognitive bias and be tricked by illusions into thinking that something that does not exist is real.

With the above in mind, the mission of NTT Communication Science Laboratories, which incorporates the words *communication science* in its name, is to connect and close the gap between computers (AI), which will continue to develop rapidly within a limited range, and humans, whose complexity also makes them imperfect (**Fig. 1**). Specifically, we look to build a theoretical foundation and develop innovative technologies toward person-to-person and person-to-computer *heart-touching communication* [1].

As one straightforward example of building a theoretical foundation, we have proposed a highly efficient coding method for sending and receiving messages up to the limit of coding efficiency (Shannon limit). This is described in “Transmission of Messages

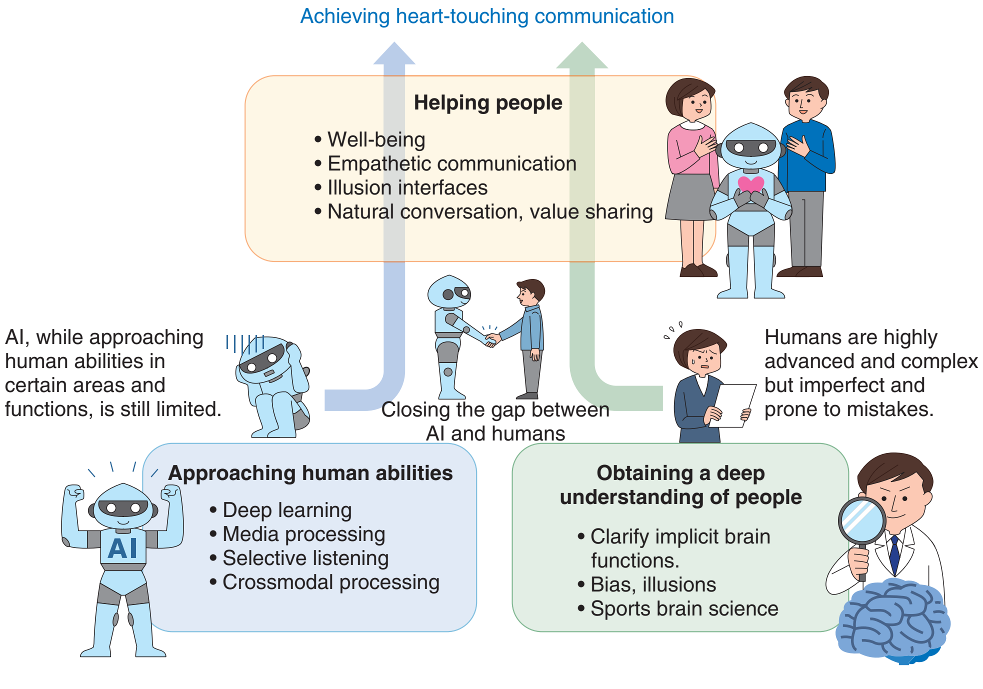

Achieving heart-touching communication
Helping people
• Well-being
• Empathetic communication
• Illusion interfaces
• Natural conversation, value sharing
AI, while approaching human abilities in certain areas and functions, is still limited.
Closing the gap between AI and humans
Humans are highly advanced and complex but imperfect and prone to mistakes.
Approaching human abilities
• Deep learning
• Media processing
• Selective listening
• Crossmodal processing
AI
Obtaining a deep understanding of people
• Clarify implicit brain functions.
• Bias, illusions
• Sports brain science

Fig. 1. Mission of communication science.

to the Efficiency Limit—Implementation of Tractable Channel Code Achieving the Shannon Limit" in the Feature Articles in this issue [2].

Furthermore, to truly achieve heart-touching communication, we must, of course, study technology that can approach human abilities with a focus on media processing. It is also important, however, that we explain human functions and characteristics and obtain a deeper understanding of people overall with the aim of developing technology that can truly help people.

2. Technology approaching human abilities

There are still many processes today that are difficult for computers to accomplish but that humans do exceptionally well. Of course, the accuracy of machine translation has been improving by leaps and bounds, and it has even become possible for an AI system to correctly answer to some extent fill-in-the-blank questions in the English portion of a Japanese university entrance exam [3]. Nevertheless, computers have yet to reach the level at which they can deeply understand the meaning of a sentence or exhibit commonsense.

It is also true, however, that computers have approached the level of human abilities in specific areas such as image recognition and speech recognition through the use of deep learning technology. Take, for example, a meeting or party where it is common for more than one person to be talking at the same time or for music to be playing in the background. Despite such noisy conditions, a human is able to zero in on the voice characteristics of the person he or she wants to listen to and to understand what that person is talking about. This is a distinctive feature of human hearing known as selective listening, which is a typical example of the broader concept of selective attention.

Computers have traditionally been weak at selective listening, but at NTT Communication Science Laboratories, we have applied proprietary deep learning techniques to develop technology that enables computers to catch only the words of the target speaker based on the voice characteristics of that person—much like humans do—and have begun to roll out this technology [4].

The key to enabling such media processing technologies to progress even further toward human abilities is crossmodal processing, which refers to processing that can cross the boundary of a single modality such as speech, video, or text. In the past, the conventional approach was to research media such as speech, video, and text separately using different analysis techniques. Today, however, thanks to the advent of deep learning that has taken up the role of a common language, recognition, generation, and conversion across multiple modalities are becoming possible.

Humans, on the other hand, have always made use of multiple senses (the five senses) in perceiving the outside world. For example, just by hearing a sound, humans are capable of imagining to a certain extent the situation associated with that sound at that place. For people, this type of crossmodal processing is commonplace in everyday life. In addition, the phenomenon of sensory substitution is well known as a means of replacing a sensory function that has been lost due to an injury or other reason with another functioning sense, as is done by visually impaired people who use their fingertips to read Braille printing.

For humans, it is natural on seeing a photo of a person's face to imagine to some extent the voice associated with that face. Could computers be made to do the same? At NTT Communication Science Laboratories, we have taken up the challenge of enabling computers to actually perform such crossmodal processing. For example, the aim of crossmedia scene analysis technology is to use sound to recognize all active events in a scene even at a location situated in a camera's blind spot. The latest crossmodal processing technologies now being pursued at NTT Communication Science Laboratories are described in "See, Hear, and Learn to Describe—Crossmodal Information Processing Opens the Way to Smarter AI" [5].

3. Technology for obtaining a deep understanding of people

In the above way, computers are approaching human abilities in specific areas if not surpassing them, but it appears that more progress is needed if AI performance is to exceed the complexity of the human brain. Humans, on the other hand, are sometimes swayed by cognitive bias or fooled by illusions that lead to completely unexpected mistakes, as reflected by the ease at which some people are taken in by bank transfer scams. The Illusion Forum website managed by NTT Communication Science Laboratories provides information on a variety of illusions that can make a person doubt one's own eyes or ears [6].

In a famous experiment conducted by Christopher Chabris and Daniel Simons [7], subjects are shown a video of six players in white and black shirts passing around basketballs to each other and instructed to count the number of times that the players in white pass one of the balls to each other. Here, at nine seconds into the video, a gorilla walks into the scene, faces the camera, pounds his chest majestically, and finally exits. Nevertheless, about half of the subjects are so engrossed in counting that they never notice the gorilla. In this way, a human turning his/her attention to something fails to notice other things that are happening in the same scene. In other words, the flip side of the remarkable human characteristic of selective attention is selective inattention. In addition, a person focusing on something does not even notice that this is happening. Thus, it is not only the elderly who get taken in by bank transfer scams.

In this way, the complex nature of humans also makes them imperfect as reflected by their tendency to be fooled by bias or illusions. In contrast, AI, while limited at present, is steadily advancing. To therefore close the gap between humans and AI and achieve coexistence and co-creation between them, it is essential that we obtain a deeper understanding of human beings in all their complexity before believing—without careful consideration—the idea that AI will one day surpass the human brain.

To this end, NTT Communication Science Laboratories is expending effort to clarify and understand implicit brain functions related to the basic human senses of seeing, hearing, and sense of movement. Here as well, illusions can provide important clues to understanding such implicit brain functions.

Understanding implicit brain functions is a challenging task. The brain activity patterns, for example, may vary greatly across individuals. Focusing on top-ranking athletes as subjects, we are working to explain the outstanding abilities of these individuals from the viewpoint of brain science and to find out how mind, technique, and body in humans are interrelated as part of our efforts in sports brain science.

For example, we have taken up the challenge of clarifying the mechanism of how a top hitter in baseball can judge whether the incoming ball is slow or fast and adjust the timing of his swing accordingly all within a very short period of time of about 0.1 second [8]. Sports brain science can be regarded as a new technology and an ambitious undertaking that departs from conventional sports science and sports analysis techniques that are mainly concerned with body training.

Incidentally, crossmodal processing as mentioned above takes place on a variety of levels within the brain. For example, when people view an ordinary video, the brain initially processes the information on color, form, and movement separately and integrates that information later. As a result, any inconsistencies that might exist among those different modalities of information will be corrected in the integration process.

This brain processing mechanism was used to devise Hengento at NTT Communication Science Laboratories [9, 10]. When experiencing Hengento, the user obtains color and form from a static object while obtaining movement from monochrome video projected on that object. Since color and form are static here, a spatial inconsistency occurs with movement. However, the brain, which attempts to see an object in a consistent manner, will correct for this inconsistency when integrating movement, color, and form. Consequently, in the Hengento experience, the user notices no inconsistencies among movement, form, and color and falls under the illusion that the color and form of the static object are actually moving. The name Hengento is derived from a Japanese word meaning illusory transforming lamps.

4. Technology for helping people

The results obtained in sports brain science research are not limited to sports. They can also be used to bring implicit mental and physical abilities into full play in the everyday life of human beings. That is to say, they can be used as knowledge for improving well-being in people. Improving human well-being is a qualitatively elusive problem, so we are tackling it in a quantitative manner from the viewpoint of human science and establishing design guidelines to enhance the sense of well-being.

One example of this approach is our work in measuring the effects of empathetic communication that occurs when a number of people come to share the same space [11]. Additionally, given the eye-straining effects of display devices such as televisions and smartphones that surround us in our modern society, we are proposing a method for self-checking the state of one's eyes on a routine basis using a general-purpose tablet device in a game format. This method is described in "Measuring Visual Abilities in an Engaging Manner" [12] in this issue.

At the same time, while illusions play an important role as clues to explaining implicit brain functions, they also hold the key to filling in the gap between humans and AI and to facilitating interfaces and feedback designed to help people in their daily lives. At NTT Communication Science Laboratories, we have developed a device called Buru-Navi that generates the illusion of being pulled by some force as an interface that exploits human illusions. We are also working on a means of making a sitting person feel as if he/she is actually walking. This development is described in "Creating a Walking Sensation for the Seated—A Sensation of Pseudo-walking Expands Peripersonal Space" [13].

In fact, we have announced a series of interesting interfaces in this area, including the aforementioned Hengento that makes a printed picture or photograph appear to move simply by projecting light on it, Hidden Stereo that enables a viewer to enjoy three-dimensional (3D) video while wearing 3D glasses and vivid 2D video when not wearing them, Ukuzo, an optical projection technique that gives 2D objects such as those on printed matter a 3D floating effect by projecting shadow-like patterns onto them [14], and Danswing papers, which gives an impression of motion to static paper objects and was selected as a top 10 finalist for the 2018 Best Illusion of the Year contest [15].

In future research, we plan to work on new types of interfaces that use illusions while simultaneously investigating the possibility of novel forms of perceptual expression that exploit illusions to create experiences that cannot be achieved by physical means.

To achieve natural dialog between humans and robots or AI, dialog-processing technologies such as speech recognition and natural language processing will be important, and it would seem at first that human biases and illusions are unrelated. However, AI has yet to reach the point of being able to understand the full meaning of a sentence or exhibit commonsense the way humans do, so dialog between humans and AI is mostly limited to a one-question/one-answer format at present. As a result, if an inconsistency arises in what is being said with what was said shortly before, a glitch in the process will quickly be exposed, and the dialog will be short-lived. It is therefore necessary to learn how to make effective use of such limited abilities and to exploit human biases and illusions so that AI appears *smart* to humans.

At NTT Communication Science Laboratories, we have achieved dialog processing that enables natural ongoing conversation even in a one-question/one-answer format by skillfully dividing up tasks between two robots. Furthermore, with the aim of breaking away from the one-question/one-answer format, we focused on the fact that much of what users talk about concerns event-related information and therefore proposed a technique for structuring and understanding user utterances in units of events. This technique is described in "Chat Dialogue System with Context Understanding" [16]. In this way, context understanding can be improved, and simulated experiences of systems that match events can be shared, which should result in dialog that can truly help people such as by inducing empathy between people and robots.

5. Conclusion

As described above, human beings are advanced and complex creatures, while AI, though approaching the level of human performance in certain areas and functions, is still limited. Achieving intelligence that surpasses that of humans is not that simple. Humans, on the other hand, are complex and imperfect; they can be taken in by bank transfer scams, be mistaken about cause-and-effect relationships, be vulnerable to bias, and be prone to mistakes. It is also known from optical illusions that humans may not always be observing physical quantities for what they really are. It is therefore important to close the gap between humans and AI and achieve AI that can help people by refining AI technology to approach the level of human abilities while simultaneously deepening our knowledge of human characteristics. This is the mission of NTT Communication Science Laboratories as we work toward achieving heart-touching communication.

References

[1] T. Yamada, "Shift to New Dimensions—Further Initiatives to Deepen Communication Science," NTT Technical Review, Vol. 16, No. 11, pp. 14–18, 2018.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201811fa1.html

[2] J. Muramatsu, "Transmission of Messages to the Efficiency Limit—Implementation of Tractable Channel Code Achieving the Shannon Limit," NTT Technical Review, Vol. 17, No. 11, pp. 34–39, 2019.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201911fa6.html

[3] R. Higashinaka, H. Sugiyama, H. Isozaki, G. Kikui, K. Dohsaka, H. Taira, and Y. Minami, "Taking the English Exam for the 'Can a Robot Get into the University of Tokyo?' Project," NTT Technical Review, Vol. 13, No. 7, 2015.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201507ra2.html

[4] 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, 2018.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201811fa2.html

[5] K. Kashino, "See, Hear, and Learn to Describe—Crossmodal Information Processing Opens the Way to Smarter AI," NTT Technical Review, Vol. 17, No. 11, pp. 12–16, 2019.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201911fa2.html

[6] Illusion Forum (in Japanese), http://www.kecl.ntt.co.jp/IllusionForum/

[7] C. Chabris and D. Simons, "The Invisible Gorilla,"
http://www.theinvisiblegorilla.com/videos.html

[8] D. Nasu, "Timing Adjustment of Baseball Batters Determined from Motion Analysis of Batting," NTT Technical Review, Vol. 16, No. 3, 2018.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201803fa3.html

[9] T. Kawabe, T. Fukiage, M. Sawayama, and S. Nishida, "Deformation Lamps: A Projection Technique to Make Static Objects Perceptually Dynamic," ACM Transactions on Applied Perception, Vol. 13, No. 2, Article 10, 2016.
https://dl.acm.org/citation.cfm?id=2874358&dl=ACM&coll=DL

[10] Press Release issued by NTT on February 17, 2015,
https://www.ntt.co.jp/news2015/1502e/150217a.html

[11] J. Watanabe, Y. Ooishi, S. Kumano, M. Perusquía-Hernández, T. G. Sato, A. Murata, and R. Mugitani, "Measuring, Understanding, and Cultivating Wellbeing in the Age of Technology," NTT Technical Review, Vol. 16, No. 11, pp. 41–44, 2018.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201811fa6.html

[12] K. Maruya, K. Hosokawa, and S. Nishida, "Measuring Visual Abilities in an Engaging Manner," NTT Technical Review, Vol. 17, No. 11, pp. 17–22, 2019.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201911fa3.html

[13] T. Amemiya, "Creating a Walking Sensation for the Seated—A Sensation of Pseudo-walking Expands Peripersonal Space," NTT Technical Review, Vol. 17, No. 11, pp. 23–27, 2019.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201911fa4.html

[14] 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

[15] T. Kawabe, "Danswing Papers,"
http://illusionoftheyear.com/2018/10/danswing-papers/

[16] H. Narimatsu, H. Sugiyama, M. Mizukami, T. Arimoto, and N. Miyazaki, "Chat Dialogue System with Context Understanding," NTT Technical Review, Vol. 17, No. 11, pp. 28–33, 2019.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201911fa5.html

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 the 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 and a 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.

See, Hear, and Learn to Describe—Crossmodal Information Processing Opens the Way to Smarter AI

Kunio Kashino

Abstract

At NTT Communication Science Laboratories, we are researching information processing across different types of media information such as images, sounds, and text. This is known as crossmodal information processing. The point of crossmodal information processing is to create a common space, which is a place where multiple types of media data are associated. The common space enables us to realize new functions that have never existed before: new transformations between different media—such as creating images and descriptions from sound—and the acquisition of concepts contained in media information.

Keywords: crossmodal information processing, AI, concept acquisition

1. Introduction

The driving force behind the recent development of artificial intelligence (AI) is deep learning technology. Deep learning, when it is applied to object recognition, for example, involves preparing a large amount of data in which images of various objects are photographed, and the names of objects (class labels) such as *apple* and *orange* are combined (training pairs). This kind of learning is known to achieve recognition of objects in images with high accuracy.

While deep learning has been researched and used in various fields due to its excellent classification performance, we are particularly interested in its ability to find correspondence between different types of media information (for example, images and sounds). The different kinds of information such as images, sounds, and text are called modalities, and the correspondence of information across different modalities is called crossmodal information processing. In this article, we introduce the concept of crossmodal information processing and its implications.

2. New information conversion

One advantage of crossmodal information processing is that it makes it possible to transform information in a way that was not previously thought possible through a common space, which is a place where different kinds of media information are related (**Fig. 1**).

2.1 Creating an image from sounds

For example, our research team is working on the task of estimating images from sounds. We humans can imagine the visual scene from the sound around us even when we close our eyes. This may imply that it could be possible to create an image of the scene from the sound picked up by microphones. For example, several microphones can be placed in a room to record several people talking there. If you use four microphones, you will get four sound spectrograms representing the frequency components of the sound captured by each microphone, and an angular spectrum representing the directions of arrival of the sounds. The system takes these as inputs. It then processes these pieces of information using neural networks and maps them into a low-dimensional space.

Another neural network then uses this information

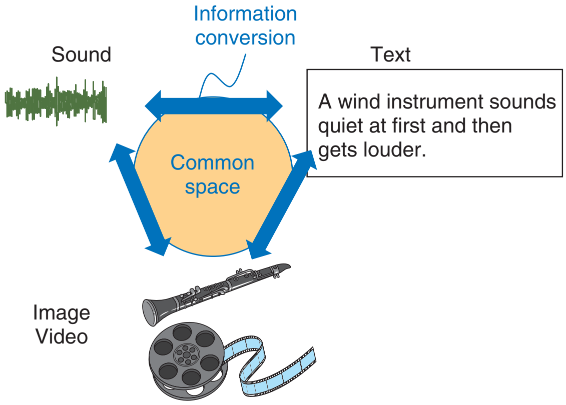

Information conversion
Sound
Text
A wind instrument sounds quiet at first and then gets louder.
Common space
Image
Video

Fig. 1. Conceptual diagram of information conversion by crossmodal information processing.

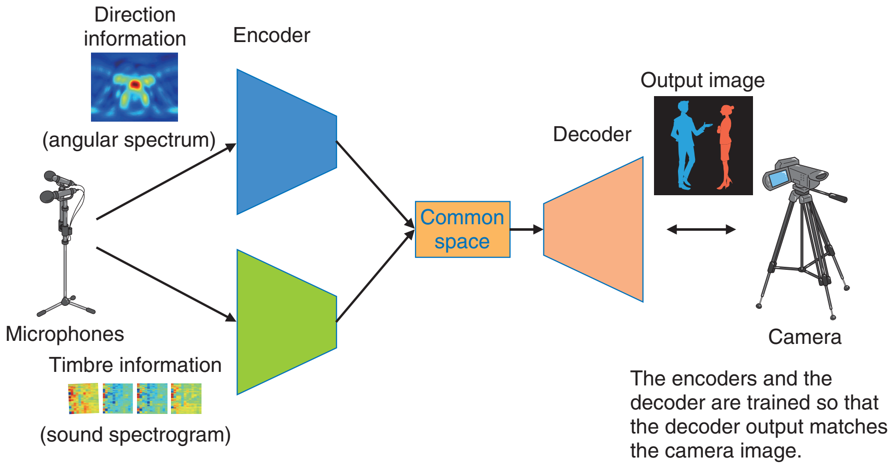


Fig. 2. Generating an image from sound.

to create an image. This image shows what kind of people are speaking and where they are in the room, so we can get a rough idea of what is going on in the room (**Fig. 2**). The process of mapping (encoding) an input into a low-dimensional space and generating (decoding) high-dimensional information from it in this way is generally called the encoder-decoder model, and it can be constructed using deep learning by providing input/output pairs as training data.

Up to now, we have conducted simulation experiments and experiments using actual sound-producing objects to confirm that it is actually possible to show what is where by means of an image under certain conditions [1]. This kind of sound to image conversion is a new information processing method that has never been tried before, to the best of our knowledge. As this technology develops, we believe that it will be useful for confirming the safety of people and property in places where cameras are not allowed or in situations where the camera cannot capture images very well such as in shadows and darkness.

2.2 Explaining sounds in words

Another example of heterogeneous information conversion is from sound to text. Conventional speech recognition systems can convert spoken words into text but cannot convert sounds other than spoken words into appropriate text. However, we have developed a technology to generate onomatopoeic words to express a sound and descriptions of the sound as a full sentence from a sound picked up by a microphone [2].

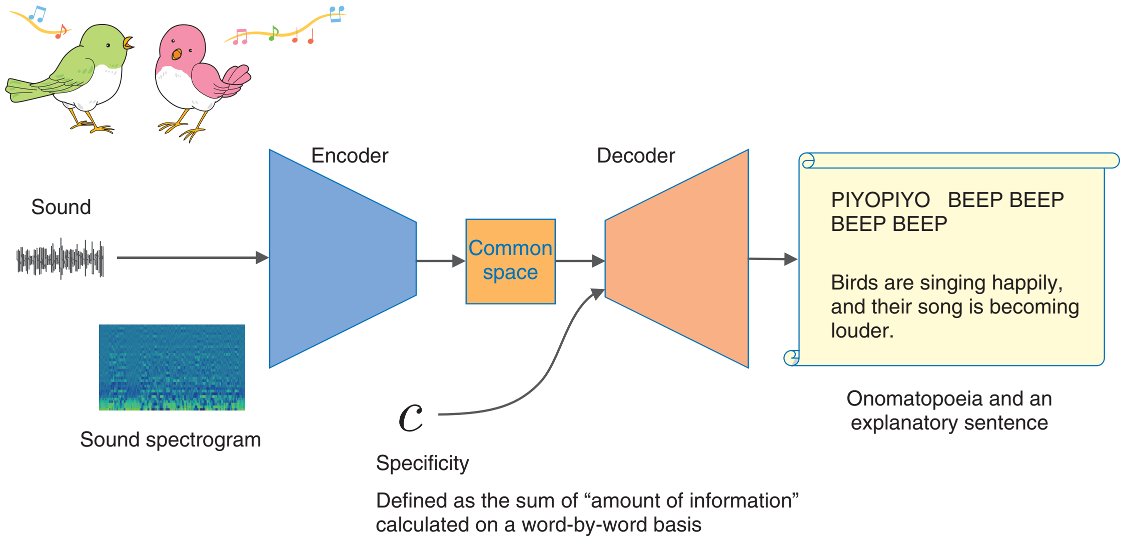

Sound
Sound spectrogram
Encoder
Common space
Decoder
c
Specificity
Defined as the sum of “amount of information” calculated on a word-by-word basis
PIYOPIYO BEEP BEEP BEEP BEEP
Birds are singing happily, and their song is becoming louder.
Onomatopoeia and an explanatory sentence

Fig. 3. Generation of explanatory statements from sounds: the conditional sequence-to-sequence caption generation method.

This method, called conditional sequence-to-sequence caption generation, or CSCG, is also based on the encoder-decoder model (**Fig. 3**). It is used here to convert one sequence to another sequence. First, the features extracted from the input acoustic signal are encoded as a time series by a recurrent neural network and mapped in the low-dimensional space. Another recurrent neural network then decodes the phoneme sequence (onomatopoeia) or the word sequence (description) from the information.

When a description is generated, the kind of description that is appropriate depends on the case, and you cannot specify a single, generic correct answer. For example, a scene may need to be expressed simply in a short sentence, such as “A car is approaching; it is dangerous,” or a scene may need to be expressed in detail in terms of subtle nuances of engine sounds according to the type of vehicle and vehicle speed.

To achieve this, we controlled the function of the decoder using an auxiliary numerical input called *specificity* and made it possible to adjust the detail of expression. The specificity value is defined as the sum of the amounts of information contained in words in a sentence. Consequently, smaller specificity values produce shorter descriptions, while greater ones produce more specific and longer descriptions. Experiments under certain conditions show that our technology can generate onomatopoeic words that are more receptive than the onomatopoeic words given by humans, and can also effectively generate explanatory texts according to the designated specificity.

We believe that this technology is useful for creating subtitles for video and real environments and for searching media. Traditionally, attempts have been made to assign known class labels to sounds, such as *gunshot*, *scream*, *the sound of a piano*, and so on. However, with sound, the correspondence between the sound signal and the name of the sound source is not always obvious, and it is quite common to encounter sounds that we cannot identify. In such cases, the effectiveness of classification alone is limited.

This technology makes it possible to search for sounds based on the description by linking the sound with the description. In fact, it is possible to directly measure the distance between sounds and onomatopoeic words or descriptions in the common space, and therefore, to search for sounds using onomatopoeic words or descriptions. In such cases, one may want to specify in detail the nuances of the desired sound in the description. With this technology, it is possible to specify not only class labels such as *car* and *wind* but also the pitch, size, and temporal changes of the sound. To our knowledge, this is the first method that can generate descriptions of sounds in the form of a full sentence.

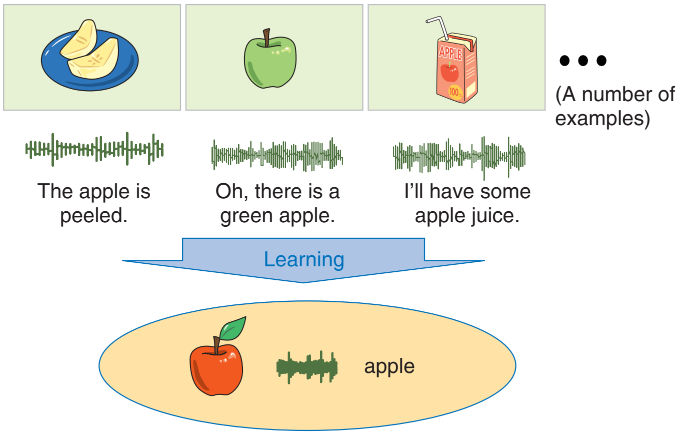


Fig. 4. Obtaining concepts from crossmodal information.

3. Learning new concepts by itself

Another advantage of crossmodal information processing is that it is possible to acquire concepts by finding the correspondence between different kinds of information in the common space. Preparing the large amount of data required for deep learning is often no simple task. It is often difficult to collect sufficient data for rare events, for example, and in many cases, it is also difficult to define classes in advance. We are therefore working on research that aims to automatically acquire a set of concepts contained in media information and use these concepts for recognition and retrieval.

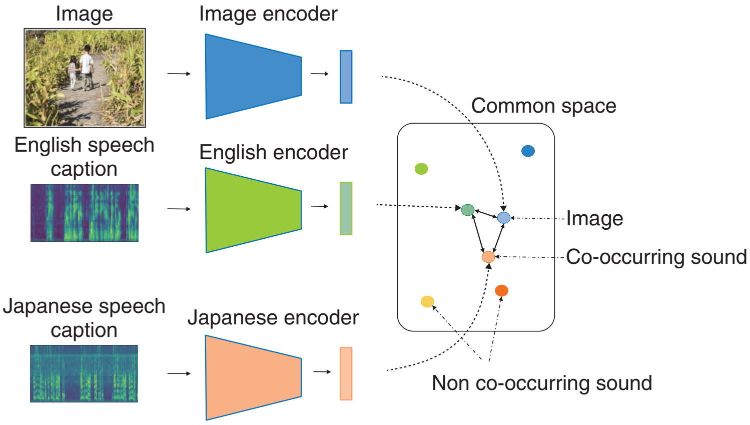
The co-occurrence of different types of media information, that is, the appearance of different types of media information originating from the same thing in the real world with specific spatiotemporal relationships instead of random relationships, makes it possible to pair media data through a common space without manually pairing media data.

For example, in daily life, it is not unusual to hear the spoken word "apple" when we see an image of an apple somewhere around us. This phenomenon means that it is no longer necessary to provide a pair of apple images and class labels in advance. Just seeing and hearing images and sounds makes the system learn their association (**Fig. 4**). In addition, the system will learn how to feel and behave, according to its circumstances. That is, if people always call an apple *ringo* (Japanese word for apple), then the system will learn it as ringo. This can be compared to the process in which we learn various things in our daily lives as we grow up.

In fact, we have confirmed that it is possible to associate words in multiple languages with objects in a photograph. As artificial co-occurrences, we prepared a set of 100,000 photographic images and their descriptions in spoken words in English and Japanese. We used them to show that the system can automatically obtain knowledge for translation between the languages regarding the objects that frequently appear in the images [3] (**Fig. 5**).

4. Future development

This article introduced crossmodal information processing. A common objective in our studies is to separate the appearance-level representation of various media information such as sound, images, and text, from the underlying common space, that is, the intrinsic information that does not depend on any specific modalities, to fully utilize both of them. If progress continues to be made in such research, it may be possible to develop AI that lives with human beings and learns by itself while sharing how to feel and behave with us humans. Such an AI could be a friendlier partner with us.



Encoder learning is performed so that co-occurring information sets are arranged close to each other in a common space, while non co-occurring information is not. Doing this for many image-speech caption pairs enables the system to extract the relationships between certain parts of the sounds and the images.

Fig. 5. Building common space by multiple encoders for concept acquisition.

References

[1] G. Irie, M. Ostrek, H. Wang, H. Kameoka, A. Kimura, T. Kawanishi, and K. Kashino, "Seeing through Sounds: Predicting Visual Semantic Segmentation Results from Multichannel Audio Signals," Proc. of the 2019 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Brighton, UK, May 2019.

[2] S. Ikawa and K. Kashino, "Generating Sound Words from Audio Signals of Acoustic Events with Sequence-to-sequence Model," Proc. of ICASSP 2018, Calgary, Canada, Apr. 2018.

[3] Y. Ohishi, A. Kimura, T. Kawanishi, K. Kashino, D. Harwath, and J. Glass, "Crossmodal Search Using Visually Grounded Multilingual Speech Signal," IEICE Tech. Rep., Vol. 119, No. 64, PRMU2019-11, pp. 283–288, 2019.

Kunio Kashino

Senior Distinguished Researcher, NTT Communication Science Laboratories.

He received a B.S. in 1990 and a Ph.D. in 1995, both from the University of Tokyo. Since joining NTT in 1995, he has been leading research projects on robust multimedia search and recognition. He is also an adjunct professor at the Graduate School of Information Science and Technology, the University of Tokyo, and a visiting professor at the National Institute of Informatics. He served as head of the Media Information Laboratory at NTT Communication Science Laboratories from 2014 to 2019.

He received the Commendation for Science and Technology from the Minister of Education, Culture, Sports, Science and Technology of Japan in 2007 and 2019. He is a senior member of the Institute of Electrical and Electronics Engineers and a Fellow of the Institute of Electronics, Information and Communication Engineers, Japan.

Measuring Visual Abilities in an Engaging Manner

Kazushi Maruya, Kenchi Hosokawa, and Shin'ya Nishida

Abstract

The human visual system differs considerably from person to person, and its ability varies with the context, task, and circumstances. To grasp the variability of visual ability in daily circumstances, we created two test batteries to easily measure visual abilities.

One is a simple visual test called a Tablet Test, which can be performed using conventional measurement methods. The other is a visual test battery called Shikaku no Mori that involves short video games. The gamification improves the enjoyability of the test compared with the conventional experimental method. The measurement time is approximately three minutes, and the accuracy of the test is comparable to that obtained in laboratory experiments, in which it often takes several hours to acquire data. The proposed test batteries would be useful for research in vision sciences as a method to investigate the diversity of visual ability and early detection of eye disease.

Keywords: vision test, self-check, gamification

1. Introduction

The incidence rate of eye disease increases with age. It can therefore safely be said that in an aging society, the number of people suffering from eye disease will increase. In addition to the aging problem, there are problems associated with display devices. We are surrounded by various display devices in our daily lives, and the effects of these devices on our visual functions are not yet clear. The types of devices have rapidly increased in number, and the range of choices has broadened. It is therefore important that we understand our own visual characteristics in advance when choosing a device.

In addition, on the device design side, knowledge of the variability of visual functions among individuals—referred to as visual diversity—is often required. However, people often forgo tests at a hospital or clinic in their busy daily lives, especially when they do not notice any abnormalities in their vision. If there were a way to self-check visual functions and skills more easily, more people would be able to better ascertain the characteristics of their own visual functions.

However, there are various problems in applying the methods used in eye examinations at hospitals, clinics, or in vision science experiments to self-check scenarios. For example, conventional vision tests require high measurement accuracy and take a considerable amount of time to complete. In addition, the equipment and measurement kits require professional operators. Furthermore, the tasks in those kits are usually not very enjoyable and are unsuitable for use by people who do not have any concerns about their eye functions.

Researchers at NTT Communication Science Laboratories have conducted various experiments in visual science over many years and acquired the know-how for appropriate data acquisition. In this research project, we considered ways to make use of our accumulated know-how and tried to create test batteries that could be used easily.

2. Two test batteries for measuring visual functions

We propose two new test batteries for visual function measurement (**Fig. 1**). One is a simple visual test

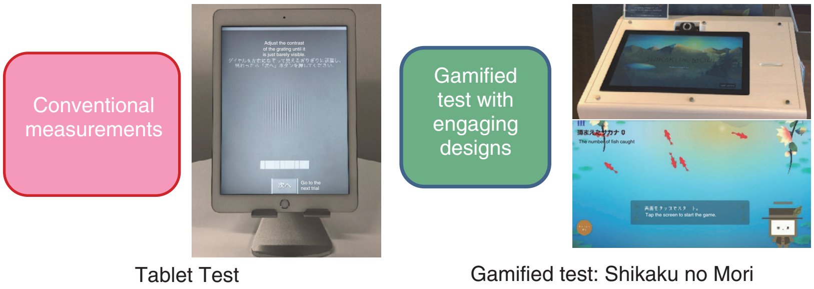
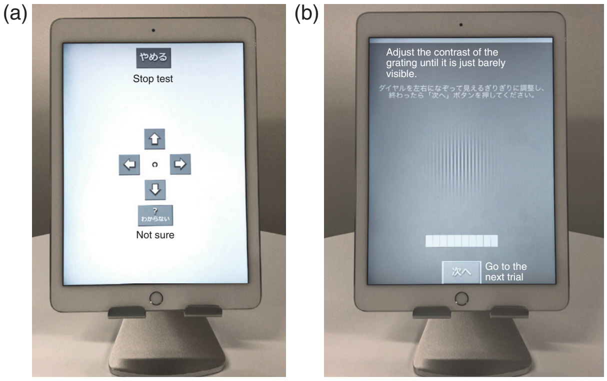

Conventional measurements
Gamified test with engaging designs
Tablet Test
Gamified test: Shikaku no Mori

Fig. 1. Images of screens used in test batteries.



Adjust the contrast of the grating until it is just barely visible.

ダイヤルを左右になぞって見えるぎりぎりに調整し、終わったら「次へ」ボタンを押してください。

Fig. 2. Screen images from Tablet Test.

called a Tablet Test, which can be performed using conventional measurement methods. The other is a visual test battery called Shikaku no Mori that involves short video games. The test battery employs graphics, directions, and task settings designed to motivate users and promote repeated use in everyday life [1].

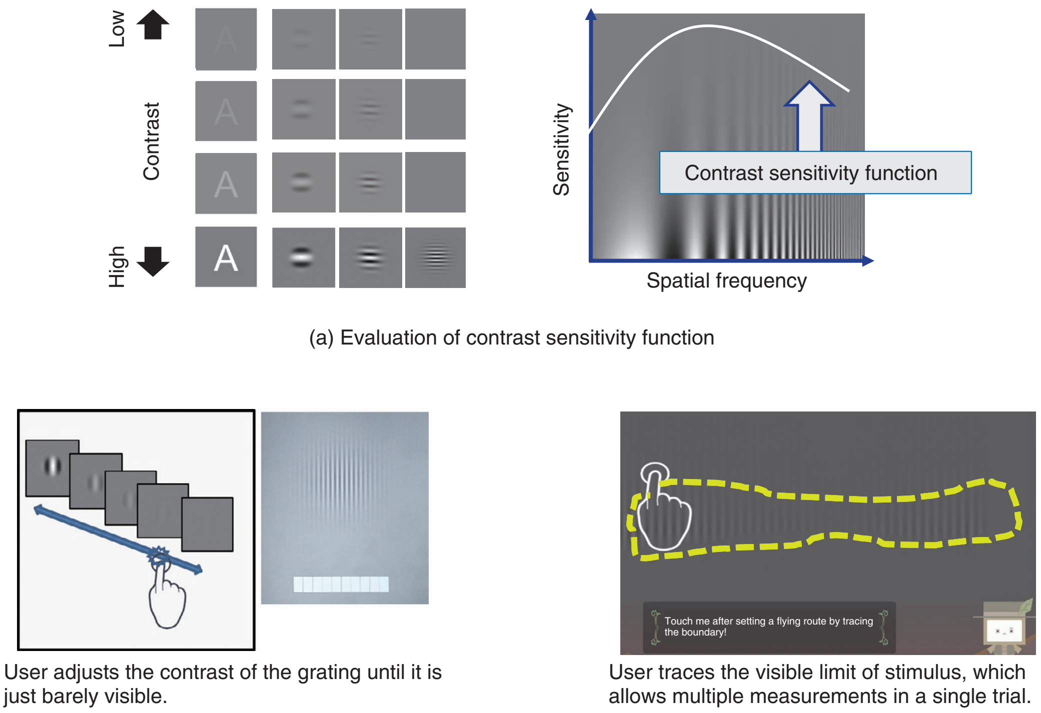
2.1 Tablet Test

In developing the Tablet Test, we considered the possibility of using it for the early detection of serious eye diseases such as cataracts, glaucoma, and age-related macular degeneration. With the cooperation of Dr. Satoshi Nakadomari from the Kobe Eye Center, we selected several test items for the test (**Fig. 2**). We started with the development of a test that can measure vision and contrast sensitivity, which can reveal the basic characteristics of the user's visual functions.

One common test used in measuring vision is performed by having the testee report the direction of the gap in a figure with the form of a "C," which is called a Landolt ring (or Landolt C) (Fig. 2(a)). For determining contrast sensitivity, the ability to see subtle differences in the color (or luminance) of black and white stripes of various widths is measured (Fig. 2(b) and **Fig. 3(a)**). Various eye diseases can reduce contrast sensitivity [2], so its measurement can be used to check for them. Moreover, contrast sensitivity data provide important information about a person's basic visual ability. In the Tablet Test, the user measures the contrast sensitivity curve by adjusting the stimulus's contrast to almost invisible levels with stripes of various widths (**Fig. 3(b)**). One can also use the Tablet Test to check if there is a field of view that is significantly reduced in sensitivity near the center of the visual field.

2.2 Gamified test: Shikaku no Mori

The other test set introduces a video game format that includes well-designed graphics and test methods that differ significantly from conventional ones in measuring visual functions (**Fig. 4**). In this test set, it is assumed that even people who do not notice anything


Low
Contrast
High
A
Sensitivity
Contrast sensitivity function
Spatial frequency
Touch me after setting a flying route by tracing the boundary!

(a) Evaluation of contrast sensitivity function

User adjusts the contrast of the grating until it is just barely visible.

User traces the visible limit of stimulus, which allows multiple measurements in a single trial.

(b) Measurements in Tablet Test

(c) Measurements in gamified test

Fig. 3. Measuring contrast sensitivity.

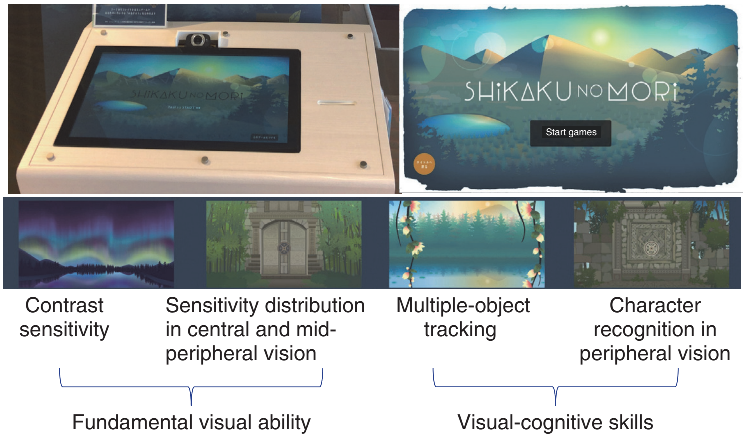

SHiKAKU NO MORi
Start games
Contrast sensitivity
Sensitivity distribution in central and mid-peripheral vision
Multiple-object tracking
Character recognition in peripheral vision
Fundamental visual ability
Visual-cognitive skills

Fig. 4. Gamified test: Shikaku no Mori.

unusual in their own visual functions will use their spare time to assess them. When game elements and forms are incorporated into tasks that originally are not intended for entertainment, various positive effects can sometimes be observed in the behavior of persons who perform them. These include improved motivation, a better understanding of the issues, and a sharper focus on the issues.

In fact, in the field of vision research, games have been developed that are designed to improve the visual functions of low-vision patients, especially children [3]. It follows then that it would be effective to introduce game elements into other visual function measurements. This test set consists of four short video games. Each can easily measure the level of contrast sensitivity, sensitivity distribution in central and mid-peripheral vision, character recognition ability in the peripheral visual field, and the ability to track multiple moving objects. The former two items are also included in the Tablet Test and can be used to check basic vision functions. The latter two items can be used to check functions that include visual cognition processing at higher levels in the visual system.

For the three items other than contrast sensitivity, we measure the functions that have already been recognized as being important in research on the relationship between video games and visual functions. When these are combined with the measurement of contrast sensitivity (**Fig. 3(c)**), which is considered important in the field of visual science, it is possible to check visual functions within a wider range of the processing stream in the visual system. After they have finished playing, the players can view the measurement results in a graph on the display screen. In addition to this graph, detailed results are recorded in a QR (Quick Response) code that is also displayed on the screen. The QR code is encrypted, and special data decryption software is provided to read the data and display the data graphically.

2.3 Test performance

The tests we created were implemented as applications that run on a web browser, under the assumption that the browser is used on a general-purpose device. The measurements use simplified methods. However, we included several technical functions to make the measurements as accurate as possible. One is a capability of accurate graphic drawing, which we developed by using JavaScript and WebGL (Web Graphics Library) with precise time control [4]. For example, the color control in general equipment is genuinely 8-bit, 256 steps. In the proposed software, color control of pseudo-12-bit, 4096 steps is achieved by utilizing a method called space-time dithering. This contributes to enhancing the accuracy of contrast sensitivity measurements and visual field examinations [5].

In addition, the proposed tests preserve the essence of measurement obtained using conventional methods. However, procedures and measurement conditions that do not critically influence the results are omitted. Furthermore, we implemented a relatively new measurement method that utilizes the characteristics of tablet computers and reduced the measurement time. We have devised these approaches to improve the performance of the test as much as possible, even with short measurement times.

We conducted experiments to examine whether these ideas were actually reflected in the performance of the test. The results showed that gamification improved the enjoyability of the test compared with the conventional experimental method. The measurement time was approximately three minutes, and the accuracy of the test was comparable to that in laboratory experiments, in which it often takes several hours to acquire data [1]. We have also confirmed that other tests can be performed with the expected accuracy.

3. Possibilities and tasks for future development of our simple vision test batteries

The circumstances under which the two test batteries are expected to be used are different (**Fig. 5**). For example, we assume that the proposed gamified tests will be used repeatedly for short durations in everyday situations, including at home. Through repetitive use, users will naturally come to know the range of game scores based on their visual ability. A continual decrease in the scores would indicate that something may have happened to the user's eyes. In that case, the user can be directed to perform measurements with the Tablet Test, which uses a task similar to conventional examinations. If the user senses any abnormality during the Tablet Tests, he or she would then go to a hospital or clinic to have a detailed examination done by an ophthalmologist. Thus, the proposed test batteries would be useful for early detection and treatment of eye disease and rehabilitation outside the hospital.

Furthermore, the test set proposed here may be useful for research in visual science as a method to investigate the diversity of visual ability. The gamified test could provide data on many healthy people or people with mild abnormalities because it offers a way to check visual ability with a general-purpose device repeatedly in daily life. The Tablet Test might provide data from various groups, including patients with eye disease, when utilized by specialists in hospitals. By combining the large amount of diverse data with deep data accumulated from precise experiments in previous visual science research, we can elucidate the diversity of visual ability and the factors behind it.

To realize this possibility, it is necessary to make the proposed test batteries accessible to many people. We are now conducting trial experiments at the Kobe Eye Center and at various events. In addition to these

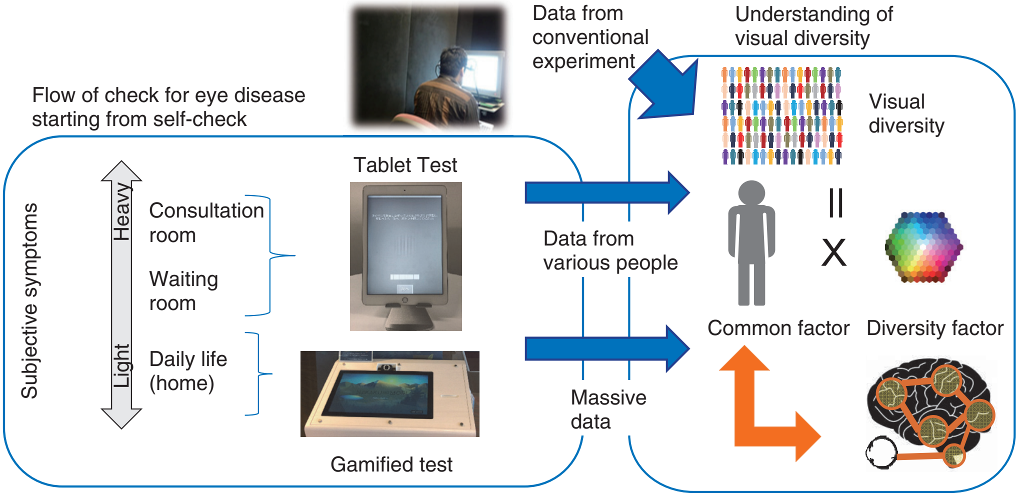


Fig. 5. Diagram outlining future use of the proposed vision test batteries.

trial experiments, we are preparing to have more people try our test batteries through the Internet.

Acknowledgments

Part of this study was achieved through research collaboration with the Kobe Eye Center. We express our sincere thanks to Dr. Satoshi Nakadomari (Kobe Eye Center) for his advice.

References

[1] K. Hosokawa, K. Maruya, S. Nishida, M. Takahashi, and S. Nakadomari, "Gamified Vision Test System for Daily Self-check," Proc. of IEEE Games, Entertainment, and Media Conference (GEM) 2019, New Haven, CT, USA, June 2019 (in press).

[2] A. Atkin, I. Bodis-Wollner, M. Wolkstein, A. Moss, and S. M. Podos, "Abnormalities of Central Contrast Sensitivity in Glaucoma," Am. J. Ophthalmol., Vol. 88, No. 2, pp. 205–211, 1979.

[3] C. Gambacorta, M. Nahum, I. Vedamurthy, J. Bayliss, J. Jordan, D. Bavelier, and D. M. Levi, "An Action Video Game for the Treatment of Amblyopia in Children: A Feasibility Study," Vision Res., Vol. 148, pp. 1–14, 2018.

[4] K. Hosokawa, K. Maruya, and S. Nishida, "Testing a Novel Tool for Vision Experiments over the Internet," Journal of Vision, Vol. 16, 967, 2016.

[5] R. Allard and J. Faubert, "The Noisy-bit Method for Digital Displays: Converting a 256 Luminance Resolution into a Continuous Resolution," Behav. Res. Methods, Vol. 40, No. 3, pp. 735–743, 2008.

Trademark notes

Kazushi Maruya

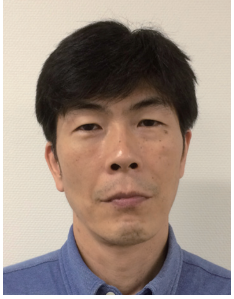
Senior Research Scientist, Sensory Representation Group, Human Information Science Laboratory, NTT Communication Science Laboratories.

He received a Ph.D. in psychology from the University of Tokyo in 2004. He joined NTT Communication Science Laboratories in 2008, where he studies human visual perception and human-computer interactions. He is a member of the Vision Sciences Society and the Vision Society of Japan.

Kenchi Hosokawa

Research Assistant, Human Information Science Laboratory, NTT Communication Science Laboratories.

He received a Ph.D. in psychology from the University of Tokyo in 2015. He joined NTT Basic Research Laboratories in 2015. He has been studying depth perception from motion parallax. His current research interests include the diversity of visual abilities among the general population and the development of a method for conducting Internet-based psychological experiments. He is a member of the Japanese Psychonomic Society and the Vision Society of Japan.



Shin’ya Nishida

Research Professor, Sensory Representation Group, Human Information Science Laboratory, NTT Communication Science Laboratories.

He received a B.S., M.S., and Ph.D. in psychology from Kyoto University in 1985, 1987, and 1996. He joined NTT in 1992. He is an expert in psychophysical research on human visual processing, in particular, motion perception, cross-attribute/modality integration, time perception, and material perception. He served as president of the Vision Society of Japan and was an editorial board member of the Journal of Vision and Vision Research.

Creating a Walking Sensation for the Seated—A Sensation of Pseudo-walking Expands Peripersonal Space

Tomohiro Amemiya

Abstract

Body action such as walking is known to extend the subjective boundaries of peripersonal space (PPS; the space immediately surrounding our body) and to facilitate the processing of audio-tactile multisensory stimuli presented within the PPS. However, it is unclear whether the boundaries change when a sensation of walking is induced with no physical body motion. In this study, we presented several vibration patterns on the soles of the feet of seated participants to evoke a sensation of walking, together with an approaching sound toward the body. We measured reaction times for detecting a vibrotactile stimulus on the chest, which was taken as a behavioral proxy for the PPS boundary. Results revealed that a cyclic vibration consisting of lowpass-filtered walking sounds presented at the soles that clearly evoked a sensation of walking reduced the reaction times, indicating that the PPS boundary was expanded forward by inducing a sensation of walking.

Keywords: sensory illusion, bodily sensation, walking sensation

1. Introduction

In recent years, virtual reality (VR) technology has been attracting the most attention in the field of games and entertainment due to the emergence of high-performance, inexpensive devices such as head-mounted displays (HMDs). VR technology can also be used in various industries such as surgical training in the medical field, worker training at over-the-counter stores, and safety education at construction sites. However, most recent VR systems only provide visual information presentation using HMDs, which is far from our actual experience in everyday life. It is very important to integrate multimodal information such as vision, audition, touch, or a sensation of body motion to produce high-quality and realistic experiences.

It is challenging to create an artificial sensation of physical walking in VR environments. To provide a sensation of walking in unlimited VR space despite having limited actual space for walking, some techniques physically cancel out the spatial displacement caused by walking. Other techniques use a visuohaptic interaction to modify the human spatial perception. With these techniques, both the brain's motor commands and the user's proprioceptive information from body movements can be utilized since the users actually move their bodies in space.

One major problem using these techniques, however, is that these systems require users to actually walk, which means that it is not possible to apply these systems to people who have difficulties walking. We therefore proposed a technique to apply multisensory cues, including proprioceptive or vestibular stimuli, to evoke a sensation of walking in seated users [1]. Our technique can be used, for example, in the living room without the subjects having to actually walk. This article introduces one of the techniques to create a pseudo-walking sensation using multisensory cues and a way to evaluate the sensation (**Fig. 1**).

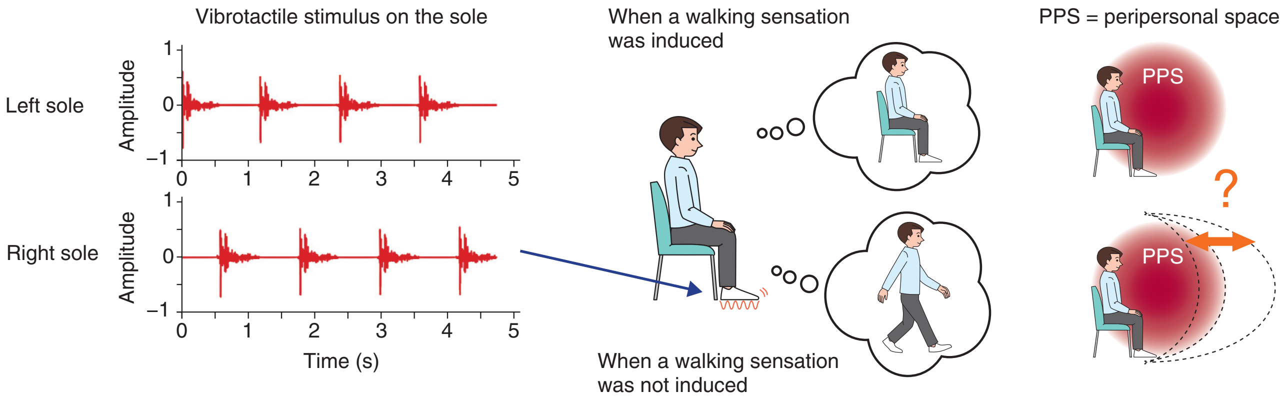

Vibrotactile stimulus on the sole
Left sole
Right sole
Amplitude
Time (s)
When a walking sensation was induced
When a walking sensation was not induced
PPS = peripersonal space
PPS
?

Fig. 1. Possible relationship between vibration on the soles of the feet and the size of peripersonal space.

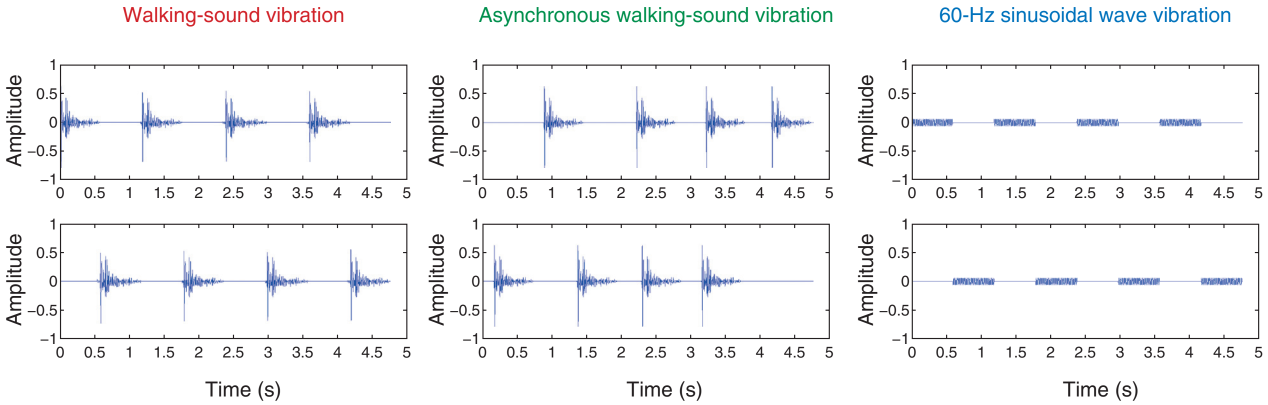


Fig. 2. Vibrotactile stimulus patterns presented on the soles of the feet in the user study.

2. Vibrations on the soles

We found that vibrations on the soles of the feet play one of the most important roles in evoking the sensation of walking. This seems reasonable because the sole is an interface between the body and the ground, and other studies have provided evidence of the importance of identifying a floor's materials while walking or maintaining body posture. Moreover, there seems to be a clear link between the tactile input from the soles hitting the ground and sensory feedback during walking such as the primitive reflex seen in newborns.

In our system, a vibration pattern consisting of a recorded footstep sound, processed through a low-pass filter at 120 Hz was presented to the heel of the foot using voice-coil vibrators (**Fig. 2**). We also presented an asynchronous pattern of the vibration, which was the same footstep sound, but the onset of each step was asynchronized and randomized, or we presented a pattern of 60-Hz sinusoidal wave vibrations that differed from the footsteps but was synchronized to the onset of each step. Note that the average amplitudes of the vibration patterns were set to be identical.

3. Expansion of peripersonal space

In addition to the subjective scale, we confirmed the change in the boundary of peripersonal space (PPS) when a sensation of pseudo-walking was induced. PPS is the space immediately surrounding our body where we mediate physical or social interactions with others or the external world.

A number of studies have provided evidence of the existence of dedicated neurophysiological, perceptual,

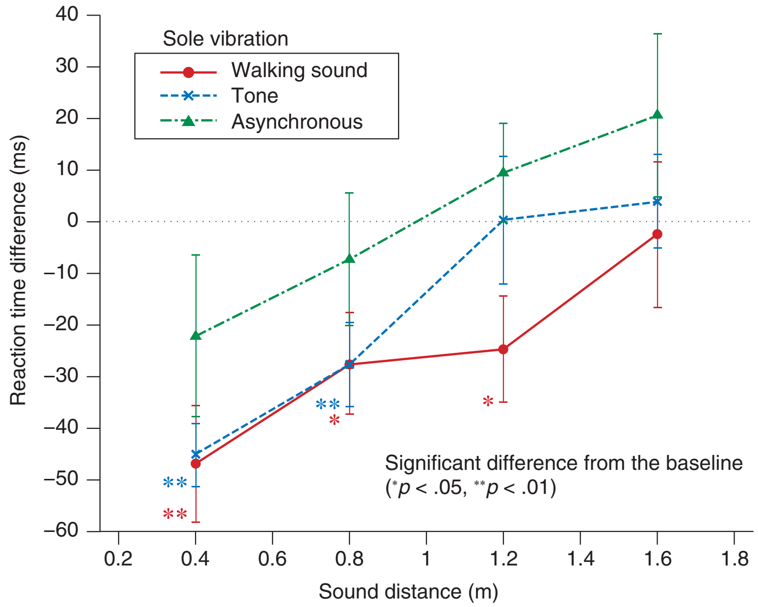


Fig. 3. Mean RT difference for tactile stimulus on the chest between the no-vibration baseline and sole vibration conditions as a function of sound distance from the participant.

or behavioral functions in PPS. Previous studies have shown that a moving sound that gives an impression of a sound source approaching the body boosts tactile reaction times (RT) when it is presented close to the stimulated body part, that is, within and not outside the PPS [2]. Therefore, we applied the method to examine PPS boundaries by measuring the distance from the participant's body where approaching sounds affect the tactile RT. We found that vibration on the soles evoking a walking sensation facilitates the RT to detect a tactile stimulus near the body when one is listening to approaching sounds [1].

Specifically, we asked the participants to detect the vibrotactile stimulus (suprathreshold, frequency of 150 Hz) on the chest as quickly as possible. Approaching white-noise sounds were presented as a change in the sound intensity of a white noise sound source by simulating the distance from the body. The tactile stimulus was given at four different temporal delays from the sound onset to imply that tactile information on the chest was processed when the sound was perceived at four different distances from the participant.

Our results showed that the vibrations on the sole boosted the processing of a vibrotactile stimulus on the chest when the position of the sound source was within a limited distance from the body. The tactile RTs in the walking-sound vibration condition were shorter than those in the asynchronous walking-sound or sinusoidal vibration conditions and in the baseline condition when the approaching sound was located at a distance of 1.2 m from the participant's body (**Fig. 3**). Our finding indicates that the PPS boundaries in the walking-sound condition expanded forward compared to other vibration conditions. In addition, the walking-sound condition received the highest rating of the subjective walking-sensation (**Fig. 4**). We speculate that there is a synergy between the subjective ratings and the multisensory facilitation.

4. Toward a higher realistic sensation

At the NTT Communication Science Laboratories' Open House 2019 held in May 2019, we conducted a demonstration in which subjects sat in a motorized chair, and vibrations were conveyed to the soles of their feet to create a very realistic sensation of pseudo-walking (**Fig. 5**). We received feedback from a number of participants who said that they felt a clear walking sensation.

5. Conclusion and future work

To the best of our knowledge, this is the first demonstration aside from those done in previous studies by our group to create a sensation of walking using

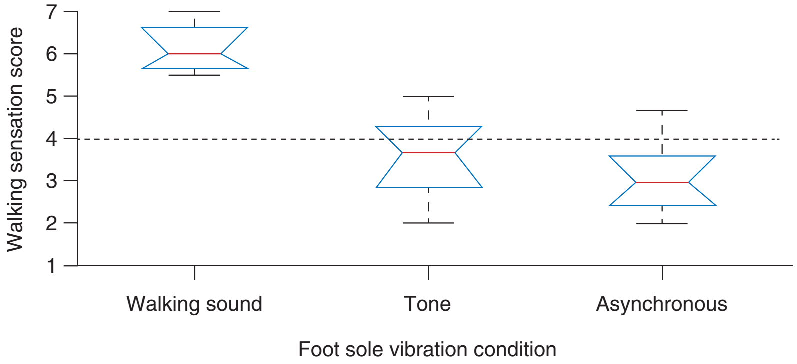


Fig. 4. Box plots of subjective walking scores of vibrotactile stimulus on the soles.

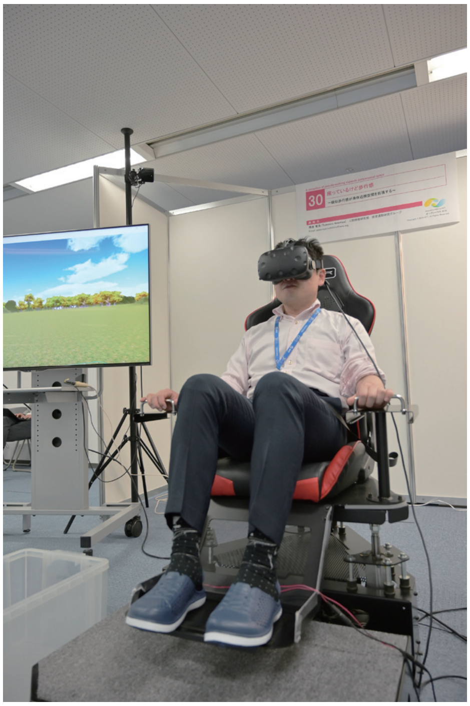


Fig. 5. Pseudo-walking sensation created using a motorized chair and vibration shoes.

multisensory stimulation without any actual body action. Such research on human characteristics is ongoing at NTT Communication Science Laboratories as part of efforts to investigate various issues [3–5]. This kind of technology will open the door to providing rich VR experiences, which is a relatively new field.

The characteristics of human perception have been extensively investigated in developing conventional video and audio systems. In the future, various kinds of sensory information will be considered in order to develop sophisticated interactive systems. Thus, we will continue to focus not only on understanding the mechanisms of sensorimotor processing in the brain, but also on finding prerequisites for developing interactive and natural user-friendly interfaces.

Acknowledgment

This research was conducted in part with Tokyo

Metropolitan University and Toyohashi University of Technology. This research was supported by a JSPS (Japan Society for the Promotion of Science) Grant-in-Aid for Scientific Research (B) #18H03283 (Principal Investigator: Tomohiro Amemiya).

References

[1] T. Amemiya, Y. Ikei, and M. Kitazaki, "Remapping Peripersonal Space by Using Foot-sole Vibrations Without Any Body Movement," Psychol. Sci., Vol. 30, No. 10, pp. 1522–1532, 2019.

[2] J.-P. Noel, P. Grivaz, P. Marmaroli, H. Lissek, O. Blanke, and A. Serino, "Full Body Action Remapping of Peripersonal Space: The Case of Walking," Neuropsychologia, Vol. 70, pp. 375–384, 2015.

[3] T. Amemiya, "Perceptual Illusions for Multisensory Displays," Proc. of the 22nd International Display Workshops, Vol. 22, pp. 1276–1279, Otsu, Japan, Dec. 2015.

[4] T. Amemiya, S. Takamuku, S. Ito, and H. Gomi, "Buru-Navi3 Gives You a Feeling of Being Pulled," NTT Technical Review, Vol. 12, No. 11, 2014.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201411fa4.html

[5] T. Amemiya, "Haptic Interface Technologies Using Perceptual Illusions," Proc. of 20th International Conference on Human-Computer Interaction (HCI International 2018), pp .168–174, Las Vegas, NV, USA, July 2018.

Tomohiro Amemiya

Senior Research Scientist, Human Information Science Laboratory, NTT Communication Science Laboratories.*

He received a B.S. and M.S. in mechano-informatics from the University of Tokyo in 2002 and 2004, and a Ph.D. in information science and technology (biomedical information science) from Osaka University in 2008. He was a research scientist at NTT Communication Science Laboratories from 2004 to 2015 and a distinguished researcher from 2015 to 2019. Since 2019, He has been an associate professor in the Graduate School of Information Science and Technology, the University of Tokyo. He was concurrently an honorary research associate at the Institute of Cognitive Neuroscience, University College London (UCL), UK, in 2014–2015. His research interests include haptic perception, tactile neural systems, wearable interfaces, and assistive technologies. He is a director of the Virtual Reality Society of Japan and the Human Interface Society. He has received several academic awards including the Grand Prix du Jury at the Laval Virtual International Awards (2007) and the Best Demonstration Award (Eurohaptics 2014).

*Current affiliation and position: Associate Professor, the University of Tokyo

Chat Dialogue System with Context Understanding

Hiromi Narimatsu, Hiroaki Sugiyama, Masahiro Mizukami, Tsunehiro Arimoto, and Noboru Miyazaki

Abstract

Many difficulties arise in developing a dialogue system that can perform conversation in the manner of humans, even for casual conversations. Recent research on chat conversation has led to the development of dialogue systems that can respond to users in a wide range of topics, which was the first major challenge of chat conversation. However, it is still difficult to construct dialogue systems that can properly respond to user utterances according to the dialogue context, and this has often made users feel that the system did not understand what they said. In this article, we introduce our work on a chat dialogue system that has the ability to understand the dialogue context.

Keywords: chat dialogue, context understanding, natural language processing

1. Toward development of a chat dialogue system

Conversations between humans and machines have been increasing with the growing use of agents in smartphones and artificial intelligence (AI) speakers (smart speakers). Most dialogue systems in commercial use are mainly used for executing tasks by giving verbal instructions such as “Call Mr. A” or “Tell me today’s weather,” but there are high expectations for dialogue systems that can chat with humans as a conversational partner. Chatting is said to have many beneficial effects such as helping to organize one’s memory and to improve communication skills. Research has been underway at NTT Communication Science Laboratories on chat dialogue systems from the early stages of dialogue system development.

Unlike with task-oriented dialogue systems, the development of chat dialogue systems is especially challenging because the system must respond to a wide range of topics in user utterances, and the dialogue scenario cannot be designed in advance. With specific tasks such as a restaurant reservation, it is possible to determine in advance the information necessary to make the reservation, such as the date and time or the reserving person’s name and telephone number. In casual conversation, on the other hand, it is impossible to predict the information contained in a user’s utterance. Therefore, it is difficult to make the system respond properly to a variety of user utterances.

Our research group has been working on techniques to develop chat dialogue systems that can respond to utterances in a wide range of topics. One typical technique is to prepare a large number of utterance pairs, such as questions and responses, and use them as training data for machine learning methods. Another technique is to select utterances similar to the user utterance by calculating the similarity between utterances using a dataset of utterance pairs. With the results of previous research, it has become possible to respond to an utterance close to a user’s utterance intention in a one question/one answer format.

However, to make a conversational partner that is more human-like, the system needs to be able to appropriately respond to user utterances according to the context. We introduce here our latest attempts to meet these challenges.

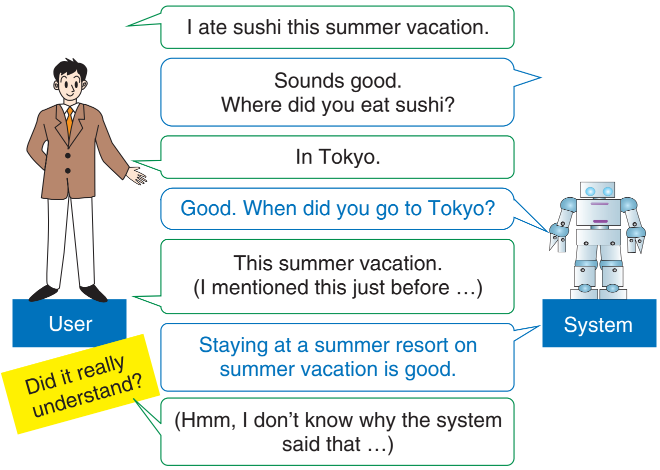

I ate sushi this summer vacation.
Sounds good. Where did you eat sushi?
In Tokyo.
Good. When did you go to Tokyo?
This summer vacation. (I mentioned this just before …)
Staying at a summer resort on summer vacation is good.
(Hmm, I don't know why the system said that …)
User
System
Did it really understand?

Fig. 1. Conversation example between user and conventional dialogue system using utterance pairs.

2. Problems with dialogue systems using single-question/single-answer utterance pairs

In a conventional dialogue system based on utterance pairs, that is, one question and one answer, the approach used is to select a similar response to a user utterance from a large number of prepared utterance pairs [1]. As a result, the system often responds to an aberrational utterance or an utterance that does not match the user's previous dialogues, and this makes the user feel as if the system does not understand them, even with just a brief dialogue. For example, conversations like that shown in **Fig. 1** are often seen in communication between humans and dialogue systems. In this dialogue, although the user said, "I ate sushi this summer vacation." in the first utterance, the system asks, "When did you go?" in the fourth utterance. This makes the user think, "I just said 'this summer vacation,' but the system didn't understand ..." Furthermore, the utterance "Staying at a summer resort on summer vacation is good." suddenly drifts away from the topic of sushi, and it confuses the user, who thinks, "I don't know why the system said that."

Such utterances indicating that the response (a) *does not match the dialogue context* and (b) *does not explain why the system said that* may cause users to feel that the system does not understand what they said or that they do not know what the system is trying to say and may thus lead them to give up conversing with the system. Consequently, the dialogue system would be viewed not only as an unskilled conversational system but also a system that does not work as a communication partner with people.

3. Development of a conversational partner

For a dialogue system to at least be recognized as a conversational partner, the problems described in the previous section need to be resolved. The psychologist H. P. Grice also stated a condition for establishing a dialogue, which was to avoid utterances that (a) referred to irrelevant matters (postulate of relevance) or (b) involved unsubstantiated and inappropriate claims (postulate of quality), since these types of utterances lead to the breakdown of dialogue [2]. Therefore, to produce a system that could give substantiated utterances according to the dialogue context while avoiding the above problems, we investigated ways of understanding the dialogue context as well as two utterance-generation methods, that is, utterance generation according to the dialogue context and utterance generation based on evidence. The details are described in the following sections.

4. Understanding the dialogue context

How should a dialogue system understand and maintain context information? We focused on the fact that the user's experience can often be described using 5W1H (Who, What, When, Where, Why, and How) + impressions, and we considered how understanding could be achieved and how to use the information of 5W1H + impressions as context. The 5W1H framework is very simple, and the simple strategy of asking 5W1H questions is often used in

Table 1. Comparison between location phrases extracted by conventional method and by proposed method.

User utterance (Red: location phrase)	Named entity extractor	Our phrase extractor
I went to Italy this summer vacation.	Italy	Italy
I went to the park near Kyoto Station and saw cherry blossoms.	Kyoto Station	the park near Kyoto Station
I often go to electronics stores.	–	electronics stores

human-human conversations and counseling dialogues. These strategies are often seen in daily life, for example, when we talk about travel or about eating delicious food, the questions "Where did you go?", "When did you go?", and "How was it?" are naturally asked in human-human conversations.

How can we develop a system that understands 5W1H + impression information through conversation? Information on time and place, taken from 5W1H information, has been the extraction target in the field of named entity recognition. For example, for the given sentence "I went to Tokyo yesterday," *yesterday* is extracted as the entity of time, and *Tokyo* as the entity of location. The extraction targets of the named entity recognition are proper nouns and specific expressions of date and time. However, is the information extracted as named entities enough for a system to understand human casual conversation? We examined the phrases that people understand as time or location in actual human conversations and found that phrases other than proper nouns accounted for the majority of location phrases. Specifically, about 70% of location phrases are not named entities.

Therefore, we developed a phrase extractor to extract phrases corresponding to 5W1H + impressions contained in the user's utterance. We developed the extractor by using the sequence-labeling methods that are effective for named entity recognition. The most representative model is CRF (conditional random field) [3], but methods using deep neural networks have also been proposed recently. First, we manually annotated the words or phrases that people understand as items of 5W1H + impressions to actual conversation between humans. Then we developed the extractor by having it train a model with the annotated conversation dataset [4].

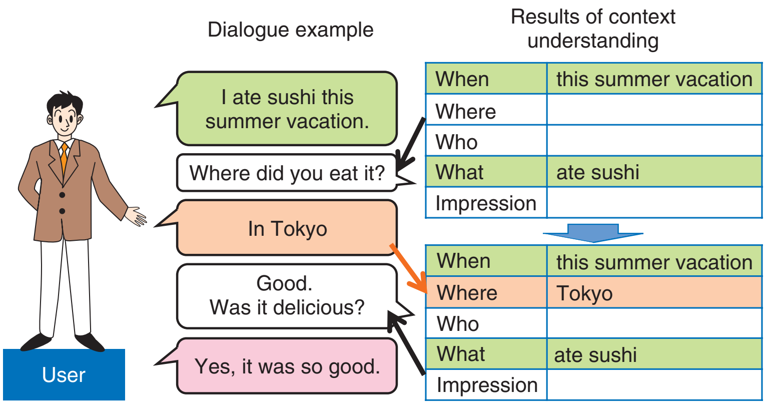
As a result, new types of phrases can be extracted as the target; "the park near Kyoto Station" is extracted as a location, even if it is not a formal proper name, and "I ate sushi" is extracted as a What item. In comparing the results extracted by the conventional named entity extractor and those by our proposed phrase extractor (**Table 1**), we found that phrases including both proper nouns and common nouns could be extracted by our extractor. With this technique, we can develop a system that can understand the context by filling in the 5W1H + impression frames through conversation.

5. Utterance generation aligned with dialogue context

With the results of the contextual understanding described in the previous section, it is easier to generate questions and utterances corresponding to the dialogue context. For example, if the system takes a conversational strategy of asking 5W1H + impressions, this technique prevents the system from asking a question whose answer has already been mentioned by a user (**Fig. 2**). Moreover, this technique helps the system to generate utterances that are appropriately relevant to the dialogue context. For example, if the utterances "I went on a trip during summer vacation" and "I went sightseeing in Tokyo" exist in the dialogue context, our technique helps associate the information of the two utterances and prompts a response such as "Tokyo is hot in summer, isn't it?" This utterance can be considered more appropriate than the utterance "There is Tokyo Tower in Tokyo," which is generated by the conventional dialogue system.

6. Utterance generation based on evidence

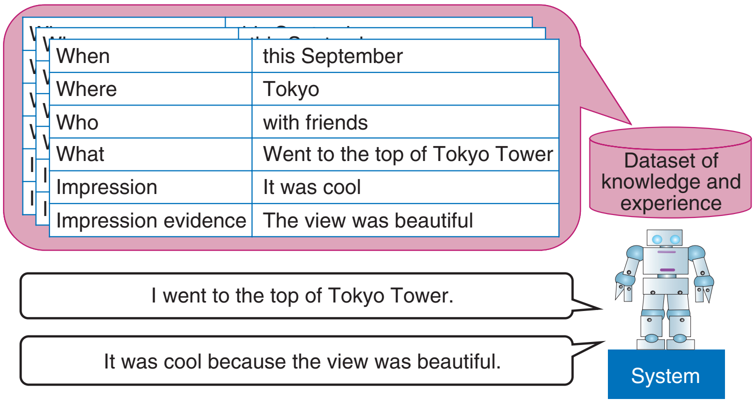
Simply appropriating context does not necessarily result in generation of utterances that show a clear correspondence, or evidence, to why the system produces a particular utterance. Therefore, we proposed a system that provides additional information on the reason the system says the utterance. Here, we introduce an approach using two examples. In the first example, the reason the system asks the question when it does is mentioned. When the system asks, "Can I enjoy it there when I go in summer?", it provides a reason as supporting information such as "I plan to go there during summer vacation, so I want to



When	this summer vacation
Where	
Who	
What	ate sushi
Impression	

When	this summer vacation
Where	Tokyo
Who	
What	ate sushi
Impression	

Fig. 2. Question generation based on results of context understanding.



When	this September
Where	Tokyo
Who	with friends
What	Went to the top of Tokyo Tower
Impression	It was cool
Impression evidence	The view was beautiful

Dataset of knowledge and experience
I went to the top of Tokyo Tower.
It was cool because the view was beautiful.
System

Fig. 3. Utterance generation based on knowledge and experience of the system.

know this." This additional statement tells the user why the system asked the question [5]. In the second example, the system mentions the reason the system thinks so as the evidence of empathic feelings or impressions when the system expresses such feelings. When the system says, "It was cool," it adds the reason for the impression such as "It was cool because the view was beautiful." We took a simple approach using the structured experience dataset [6], as shown in **Fig. 3** and an utterance template as "I also did [what] and had [impression] because [impression reason]." The utterance "I also went to the top of Tokyo Tower. It was cool because the view was beautiful." is generated by filling each item in the utterance template from an experience dataset. This utterance makes users feel more empathic than the simple utterance, "It was cool." since the system expresses the empathic feelings based on the system's experience and knowledge.

Combining the contextual understanding and context-aligned utterances described previously enables us to add further evidence to context-aligned utterances. This enables the system to produce a dialogue that makes the user think, "This system understands me" (**Fig. 4**).

7. Future work

Through the efforts made in this study, we have developed a dialogue system that is able to understand the context and generate appropriate questions and grounded utterances. This is a major step toward changing an interactive dialogue system that in the past has had users thinking "This system and I do not understand each other" into one that enables them to interact with understanding. If users had a system that

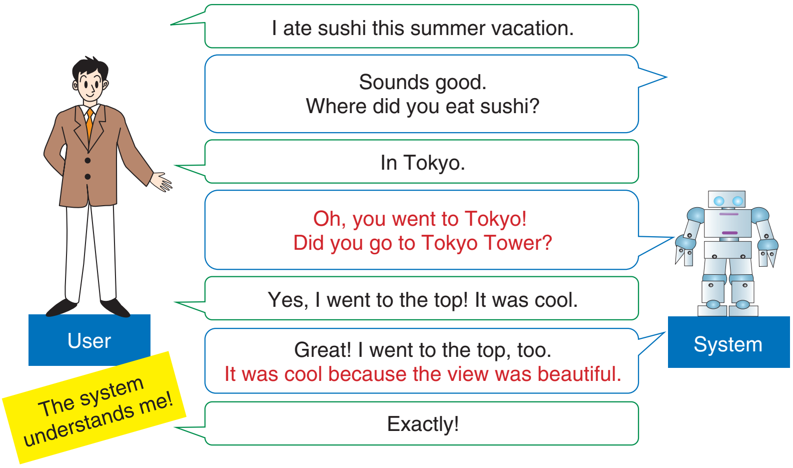

I ate sushi this summer vacation.
Sounds good.
Where did you eat sushi?
In Tokyo.
Oh, you went to Tokyo!
Did you go to Tokyo Tower?
Yes, I went to the top! It was cool.
Great! I went to the top, too.
It was cool because the view was beautiful.
Exactly!
User
System
The system understands me!

Fig. 4. Dialogue example between user and proposed system.

understood what they were saying, they would talk with it in the manner of a conversation between humans. This would also promote the use of dialogue systems in various applications such as communication training and consultation.

However, to achieve this, it is necessary to effectively design the flow of the dialogue and to manually create data that can be used as the knowledge of the system. Moreover, it is not the case that anyone can easily create a similar system. In the future, we will work on a method to automatically generate data through the web or actual conversations between the system and humans, rather than using manually generated data.

References

[1] H. Sugiyama, R. Higashinaka, and T. Meguro, “Towards User-friendly Conversational Systems,” NTT Technical Review, Vol. 14, No. 11, 2016.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201611fa4.html

[2] H. P. Grice, “Logic and Conversation,” Syntax and Semantics, Vol. 3, Speech Acts, P. Cole and J. Morgan (eds.), pp. 41–58, 1975.

[3] J. Lafferty, A. McCallum, and F. C.N. Pereira, “Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data,” Proc. of the Eighteenth International Conference on Machine Learning (ICML), pp. 282–289, Williamstown, MA, USA, June 2001.

[4] H. Narimatsu, H. Sugiyama, and M. Mizukami, “Detecting Location-indicating Phrases in User Utterances for Chat-oriented Dialogue Systems,” Proc. of the Fourth Linguistic and Cognitive Approaches to Dialog Agents Workshop (LaCATODA), Stockholm, Sweden, July 2018.

[5] H. Sugiyama, H. Narimatsu, M. Mizukami, and T. Arimoto, “Empirical Study on Domain-specific Conversational Dialogue System Based on Context-aware Utterance Understanding and Generation,” SIG-SLUD, Vol. B5, No. 02, 2018 (in Japanese).

[6] M. Mizukami, H. Sugiyama, and H. Narimatsu, “Event Data Collection for Recent Personal Questions,” Proc. of LaCATODA, Stockholm, Sweden, July 2018.

Hiromi Narimatsu

Research Scientist, Interaction Research Group, Innovative Communication Laboratory, NTT Communication Science Laboratories.

She received an M.E. and Ph.D in engineering from the University of Electro-Communications, Tokyo, in 2011 and 2017. She joined NTT in 2011. Her research interests include natural language processing, spoken dialogue systems, and mathematical modeling. She is a member of the Institute of Electrical and Electronics Engineers (IEEE), the Institute of Electronics, Information and Communication Engineers (IEICE), Information Processing Society of Japan (IPSJ) and the Japanese Society for Artificial Intelligence (JSAI).

Hiroaki Sugiyama

Research Scientist, Interaction Research Group, Innovative Communication Laboratory, NTT Communication Science Laboratories.

He received a B.E. and M.E. in information science and technology from the University of Tokyo in 2007 and 2009, and a Ph.D. in engineering from Nara Institute of Science and Technology. He joined NTT Communication Science Laboratories in 2009 and studied chat-oriented dialogue systems and language development of human infants. He is a member of IEEE and JSAI.

Masahiro Mizukami

Researcher, Interaction Research Group, Innovative Communication Laboratory, NTT Communication Science Laboratories.

He received a Ph.D. in engineering from Nara Institute of Science and Technology in 2017. His research interests include dialogue systems.

Tsunehiro Arimoto

Researcher, Interaction Research Group, Innovative Communication Laboratory, NTT Communication Science Laboratories.

He received a B.E., M.E., and Ph.D. in engineering from Osaka University in 2013, 2015, and 2018. He joined NTT Communication Science Laboratories in 2018. His research interests include artificial intelligence, human-robot interaction, and dialogue systems.

Noboru Miyazaki

Senior Research Engineer, Cognitive information Processing Laboratory, NTT Media Intelligence Laboratories.

He received a B.A. and M.E. from Tokyo Institute of Technology in 1995 and 1997. He joined NTT Basic Research Laboratories in 1997. His research interests include speech processing and spoken dialogue systems. He is a member of IEICE, the Acoustical Society of Japan, and JSAI.

Transmission of Messages to the Efficiency Limit—Implementation of Tractable Channel Code Achieving the Shannon Limit

Jun Muramatsu

Abstract

This article introduces CoCoNuTS, a technology for implementing an error correcting code (channel code) that achieves the efficient transmission limit known as the Shannon limit. It was once believed that a huge time complexity was necessary to achieve the Shannon limit for a given channel. However, practical channel codes that achieve the Shannon limit have recently been developed, but these codes achieve the Shannon limit only for a restricted class of channels. We have proven mathematically that we can construct codes achieving the Shannon limit with our CoCoNuTS technology. Furthermore, we have confirmed experimentally that the implemented codes outperform conventional codes for a channel where it is impossible to achieve the Shannon limit using the conventional codes.

Keywords: information theory, channel code, CoCoNuTS

1. Introduction

To achieve effective communication, messages must be transmitted correctly over noisy environments. Channel codes, also known as error correcting codes, represent technology that provides error-free transmission. They are applied to transmissions over optical fiber and radio environments as well as to computers, recorders such as hard/optical discs, and two-dimensional codes. It is not too much to say that they are implemented in almost all communication devices.

For a given noisy environment (channel), there is a transmission efficiency limit to achieve error-free transmission. This is called the Shannon limit[*1] after the computer scientist C. E. Shannon, who presented this limit in 1948. However, his code construction is impractical in the sense that it requires a huge amount of time. The information theory community has been studying construction of practical codes designed to achieve the Shannon limit for around 70 years.

Low-density parity check (LDPC) codes[*2] and polar codes have recently been developed as practical codes that achieve the Shannon limit. They are implemented in fifth-generation (5G) mobile communication technology. However, their codes do not achieve

[*1] The Shannon limit: A generic name representing the fundamental performance limit of transmission (source codes, channel codes, and codes for information-theoretic security) derived by information theory. The name refers to the computer scientist C. E. Shannon, who was the founder of information theory. In the context of a channel code (error correcting code), this limit is also called the channel capacity, where higher speed transmission is possible when the limit is increased. When the efficiency of the codes reaches the limit by increasing the number of transmitted signals, we say that these codes have achieved the Shannon limit.

[*2] LDPC codes: A class of codes that are tractable and that achieve the Shannon limit for a particular class of channels. A sparse matrix (where almost all the elements are zero) is used for practical decoding. It was introduced by computer scientist R. G. Gallager in 1962 but was not practically implemented because of the limited computer power at that time. It was re-evaluated in the 1990s and is implemented in wireless local area networks, satellite digital broadcasting, and 5G mobile communication technology.

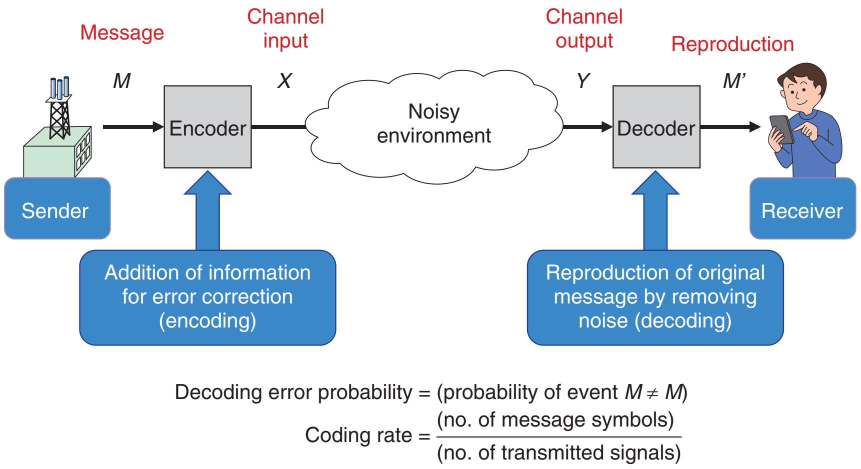

Message
Channel input
Channel output
Reproduction
M
X
Y
M'
Encoder
Noisy environment
Decoder
Sender
Receiver
Addition of information for error correction (encoding)
Reproduction of original message by removing noise (decoding)
Decoding error probability = (probability of event M ≠ M)
Coding rate = (no. of message symbols) / (no. of transmitted signals)

Fig. 1. Communication system realized using channel codes.

the Shannon limit for all channels but only for a particular class.

2. Research accomplishment

CoCoNuTS[*3], a coding technology that achieves the Shannon limit, was developed at NTT Communication Science Laboratories. We can apply this technology to construct channel codes as well as source codes and codes for information-theoretic security that are tractable and achieve the fundamental limit of transmission. In this article, we describe how we applied this technology to channel codes and proved mathematically that we can use them to achieve the Shannon limit [1–3]. Furthermore, we confirmed experimentally that our codes outperform LDPC codes in a channel where the Shannon limit is not achievable with LDPC codes.

3. Coding technology

In this section, we briefly describe channel coding. We also explain the Shannon limit and introduce our newly developed CoCoNuTS technology.

3.1 Communication system achieved using channel codes

A communication system achieved using channel codes is shown in **Fig. 1**. In this figure, the sender is a wireless station that is sending messages, and the receiver is a user who has a smartphone. The encoder converts a message M to a signal X called a channel input. It is transmitted as a modulated radio wave, where it is assumed that noise is added to the transmission. The decoder converts a reproduction M' from a received signal Y called a channel output. The transmission is successful when $M = M'$ is satisfied, and the decoding error probability is defined as the probability of events satisfying $M \neq M'$.

The transmission efficiency, hereafter called the coding rate, is defined as the number of message symbols divided by the number of transmitted signals. A higher coding rate means higher speed transmission. However, a coding rate that is too high makes it impossible to achieve a decoding error probability close to zero. Our goal is to construct an encoder and decoder pair where the decoding error probability is close to zero and the coding rate is close to the fundamental limit.

3.2 Example of channel coding

An example of channel coding of the message “01” is shown in **Fig. 2**. The encoder converts the message to the channel input signal “01011,” which is a twofold repetition of the message and a 1-bit parity check, which is the sum (exclusive-or) of two message bits. This operation is called *encoding*. In this example, the decoder observes the channel output signal “01111.” The decoder determines the position of the noise by using the encoding rule and reproduces the original message. This operation is called

[*3] CoCoNuTS: The name of a technology developed by NTT to design codes that achieve the Shannon limit. We call this technology CoCoNuTS (Code based on Constrained Numbers Theoretically achieving the Shannon limit) because it uses constrained-random-number generators to achieve the fundamental limit.

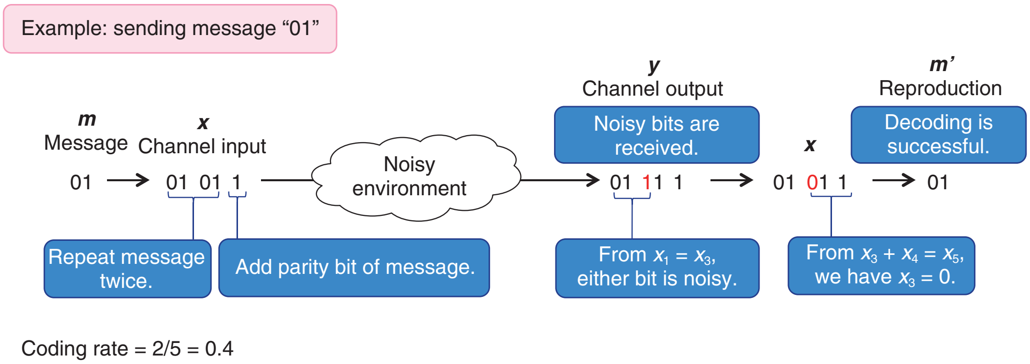


Fig. 2. Example of channel coding.

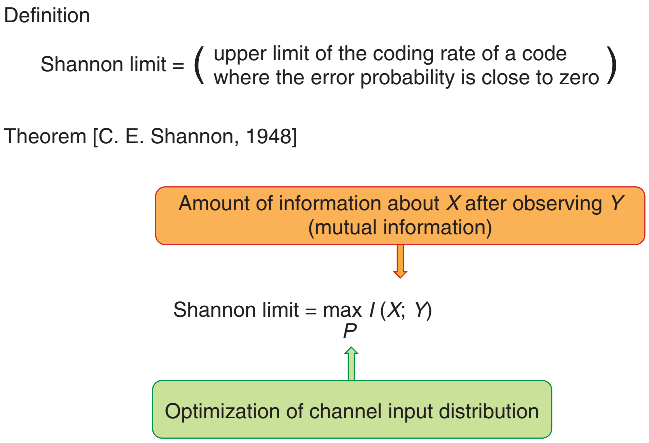

Definition
Shannon limit = (upper limit of the coding rate of a code where the error probability is close to zero)
Theorem [C. E. Shannon, 1948]
Amount of information about X after observing Y (mutual information)
Shannon limit = $\max_P I(X; Y)$
Optimization of channel input distribution

Fig. 3. The Shannon limit.

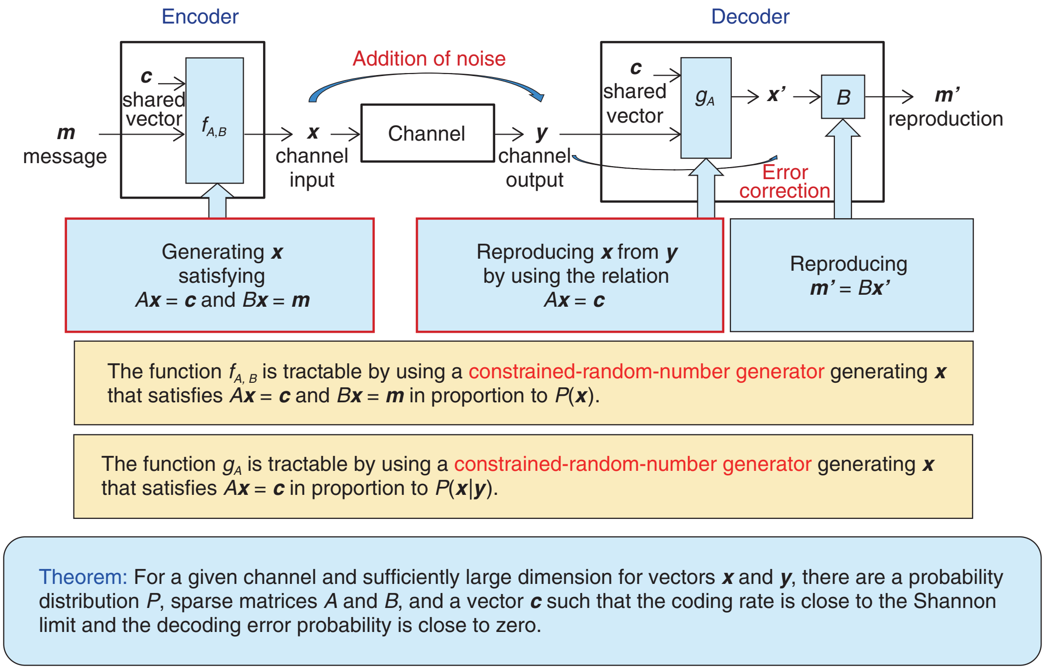
decoding. In this example, the message is successfully reproduced, but a noisier output may prevent us from reproducing the original message. In Fig. 2, a 2-bit message is encoded in a 5-bit channel input, where the coding rate is 2/5 = 0.4. By increasing the number of repetitions and parity check bits, the number of noisy bits that the decoder can specify is increased while the coding rate is decreased. For high-speed error-free transmission, it is necessary to specify all noisy bits with a probability close to one and obtain the highest possible encoding rate.

3.3 Shannon limit

The Shannon limit (channel capacity) is defined in **Fig. 3**. As shown in Fig. 1, the decoding error probability of a code must be close to zero. However, a more efficient code has a higher coding rate. The Shannon limit is the optimum coding rate of codes, where the decoding error probability is close to zero. Shannon derived the limit illustrated in Fig. 3, where the optimum is achievable by letting the number of transmitted signals reach infinity. It is theoretically impossible to construct codes with a rate beyond the Shannon limit. It should be noted that we have to optimize the channel input distribution P, which appears in the maximum operator on the right-hand side of the equality.

3.4 Proposed method CoCoNuTS

The construction of a channel code using CoCoNuTS is shown in **Fig. 4**. With the proposed method, a code is constructed by using two sparse matrices (almost all elements of a matrix are zero) A and B, and a vector $\boldsymbol{c}$. Functions $f_{A,B}$ and g_A (red squares in Fig. 4) are implemented by using tractable constrained-random-number generators [1–3]. We can


Encoder
Decoder
Addition of noise
$\boldsymbol{c}$
shared vector
$\boldsymbol{m}$
message
$f_{A,B}$
$\boldsymbol{x}$
channel input
Channel
$\boldsymbol{y}$
channel output
$\boldsymbol{c}$
shared vector
g_A
$\boldsymbol{x'}$
B
$\boldsymbol{m'}$
reproduction
Error correction
Generating $\boldsymbol{x}$ satisfying $A\boldsymbol{x} = \boldsymbol{c}$ and $B\boldsymbol{x} = \boldsymbol{m}$
Reproducing $\boldsymbol{x}$ from $\boldsymbol{y}$ by using the relation $A\boldsymbol{x} = \boldsymbol{c}$
Reproducing $\boldsymbol{m'} = B\boldsymbol{x'}$

The function $f_{A,B}$ is tractable by using a constrained-random-number generator generating $\boldsymbol{x}$ that satisfies $A\boldsymbol{x} = \boldsymbol{c}$ and $B\boldsymbol{x} = \boldsymbol{m}$ in proportion to $P(\boldsymbol{x})$.

The function g_A is tractable by using a constrained-random-number generator generating $\boldsymbol{x}$ that satisfies $A\boldsymbol{x} = \boldsymbol{c}$ in proportion to $P(\boldsymbol{x}|\boldsymbol{y})$.

Theorem: For a given channel and sufficiently large dimension for vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, there are a probability distribution P, sparse matrices A and B, and a vector $\boldsymbol{c}$ such that the coding rate is close to the Shannon limit and the decoding error probability is close to zero.

Fig. 4. Proposed method CoCoNuTS.

achieve the Shannon limit by employing an optimal channel input distribution P.

In contrast, a generator matrix is used as the encoder of the conventional LDPC codes, where the channel input distribution is approximately uniform. This implies that the LDPC codes achieve the Shannon limit if the maximum in Fig. 3 is achieved with a uniform input distribution P, and otherwise they cannot achieve the limit.

3.5 Experimental results

In **Fig. 5**, the proposed codes are compared with the conventional LDPC codes. The horizontal axis represents the coding rate, where a larger value implies better performance. The vertical axis represents the decoding error probability, where a smaller value implies better performance. The graph shows that the proposed codes outperform the LDPC codes. For example, when they are compared on the horizontal line at a decoding error probability of 10^{-4}, the proposed codes can transmit 60 bits of messages more than the LDPC codes per 2000 bits of transmitted signals.

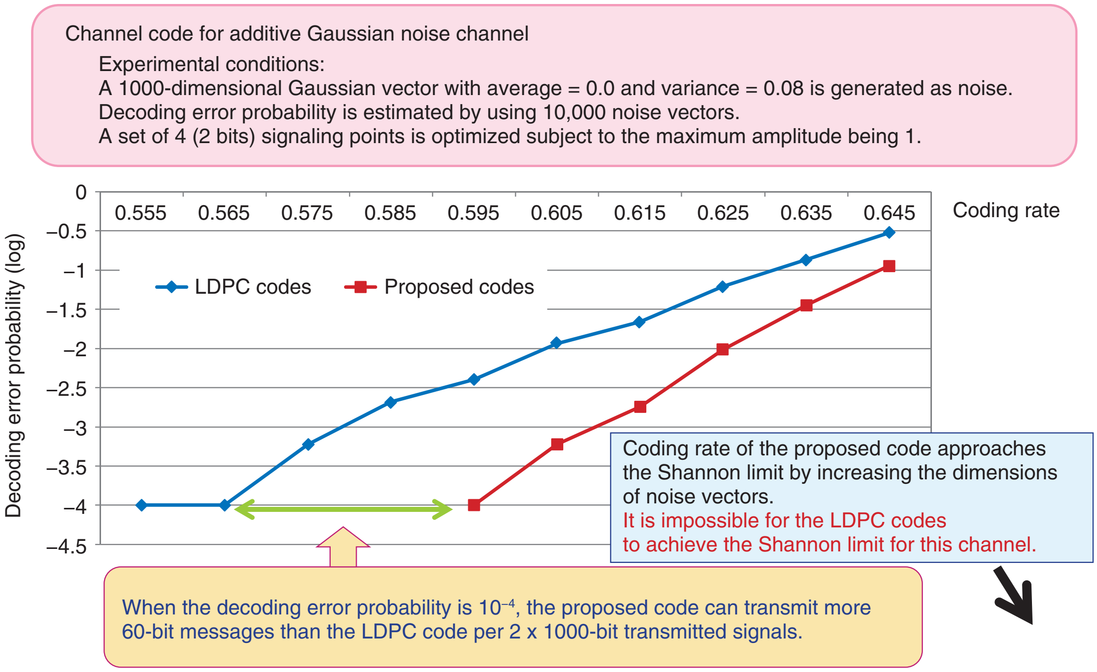
A comparison at the same coding rate of 0.6 is shown in **Fig. 6**. The black dots represent the decoding error events, where encoding and decoding are repeated 1200 times under the conditions described in Fig. 5. There is no decoding error event with the proposed code, while there are seven decoding error events with the LDPC code. This result suggests that the proposed code is more reliable than the conventional code in the same noisy environment and with the same coding rate.

A comparison when a compressed image file (JPEG)[*4] is transmitted, where it is assumed that there is no decoding error, is shown in **Fig. 7**. In this situation, a decoding error destroys the file, and most of the original image cannot be reproduced. When the coding rate is 0.5, both the proposed code and the LDPC code can reproduce the original image. However, the LDPC code cannot reproduce the original image at a coding rate of 0.6. This implies that the coding rate limit of the LDPC code is between 0.5 and 0.6, which is less than that of the proposed code.

4. Future work

Our goal is to achieve future high-speed digital communication by establishing related peripheral technologies. We will continue our research in this area and report our results.

[*4] JPEG: JPEG is a standard format for image compression developed by the Joint Photographic Experts Group.



Channel code for additive Gaussian noise channel

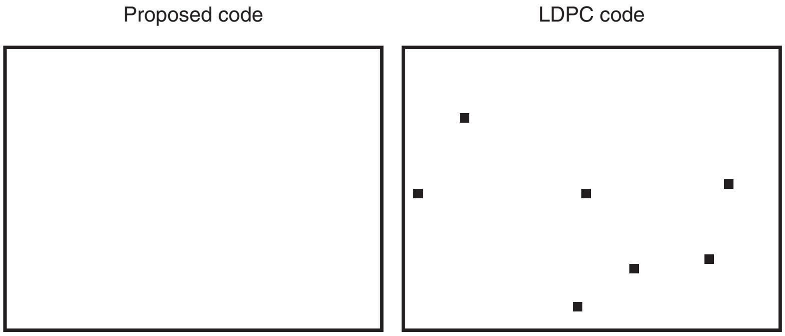
Experimental conditions:
A 1000-dimensional Gaussian vector with average = 0.0 and variance = 0.08 is generated as noise.
Decoding error probability is estimated by using 10,000 noise vectors.
A set of 4 (2 bits) signaling points is optimized subject to the maximum amplitude being 1.

Coding rate of the proposed code approaches the Shannon limit by increasing the dimensions of noise vectors.
It is impossible for the LDPC codes to achieve the Shannon limit for this channel.

When the decoding error probability is 10^{-4}, the proposed code can transmit more 60-bit messages than the LDPC code per 2 x 1000-bit transmitted signals.

Fig. 5. Experimental results.

Experimental conditions:
The coding rate of both codes is 0.6.
Decoding error probability is estimated by using 30 x 40 = 1200 noise vectors.
The conditions for a noise vector are as described in Fig. 5.



No decoding error observed (0/1200).

Decoding errors observed (7/1200).

A black dot represents a decoding error event, where there is at least a 1-bit difference between an original 1200-bit message and its reproduction.

Fig. 6. Visualization of decoding error events at a coding rate of 0.6.

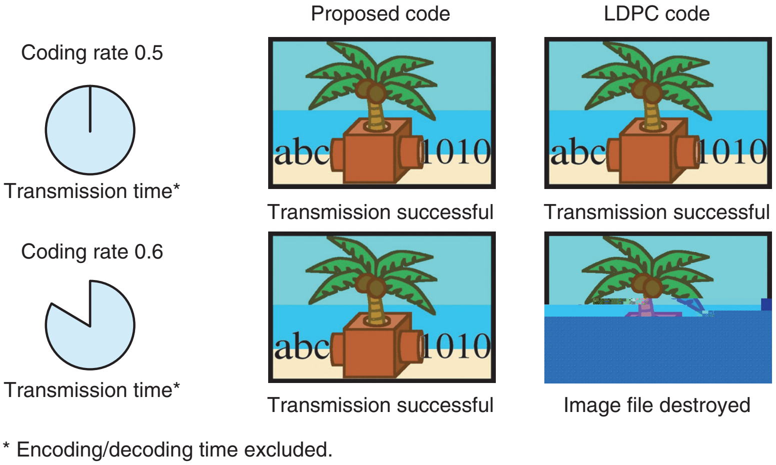


Fig. 7. Transmission of compressed image at a coding rate of 0.6.

References

[1] J. Muramatsu and S. Miyake, "Concept of CoCoNuTS," Proc. of the 10th Asia-Europe Workshop on Information Theory, p. 4, Boppard, Germany, June 2017.

[2] J. Muramatsu, "Channel Coding and Lossy Source Coding Using a Constrained Random Number Generator," IEEE Trans. Inf. Theory, Vol. 60, No. 5, pp. 2667–2686, May 2014.

[3] J. Muramatsu and S. Miyake, "Channel Code Using Constrained-random-number Generator Revisited," IEEE Trans. Inf. Theory, Vol. IT-65, No. 1, pp. 500–508, Jan. 2019.

Jun Muramatsu

Research Scientist, NTT Communication Science Laboratories.

He received a B.S. and M.S. in mathematics and a Ph.D. from Nagoya University, Aichi, in 1990, 1992, and 1998. He joined NTT Transmission Systems Laboratories in 1992 and moved to NTT Communication Science Laboratories in 1995. He has been conducting research on information theory. From February 2007 to February 2008, he was a visiting researcher at ETH Zurich, Switzerland. From 2006 to 2010, he was an associate editor of the Institute of Electronics, Information and Communication Engineers (IEICE) Transactions on Fundamentals of Electronics, Communications and Computer Sciences. He has been Chair of the IEICE Technical Committee on Information Theory since 2018. He received the Young Researcher Award from SITA (the Society of Information Theory and Its Application) in 2003 and the 63rd Best Paper Award from IEICE in 2007. He is a member of IEICE and the IEEE (Institute of Electrical and Electronics Engineers) Information Theory Society.

Noninvasive Glucose Measurement Using Electromagnetic Waves: Photoacoustic Spectroscopy and Dielectric Spectroscopy

Masahito Nakamura, Yujiro Tanaka, Takuro Tajima, and Michiko Seyama

Abstract

Noninvasive glucose measurement without needle pricking is anticipated as a novel medical and healthcare application. We introduce here our research on the use of near-infrared photoacoustic spectroscopy and microwave dielectric spectroscopy for noninvasive glucose measurement using electromagnetic waves. We also present our recent work involving *in vivo* measurement. Both measurement techniques are based on optical and wireless components and system integration, which have been investigated in telecommunication system development at NTT.

Keywords: noninvasive glucose measurement, photoacoustic spectroscopy, dielectric spectroscopy

1. Introduction

The number of diabetes sufferers has increased greatly throughout the world in recent years [1]. Effective management of blood glucose level for diabetes care relies on having accurate glucose monitors such as self-monitoring blood glucose monitors, continuous glucose monitors, and flash glucose monitors (FGMs). There is a huge demand for noninvasive glucose testing to enable patients to check their own glucose level without the need for reagents or the use of needles to take blood samples. Furthermore, such a testing method could also possibly be used in novel Internet of Things devices for people other than diabetics. They would be able to acquire time series data of changes in their glucose level as effectively as done with conventional wearable devices such as acceleration sensors and heartbeat sensors.

A noninvasive testing method would measure individual glucose metabolism and is therefore a potential tool for predicting the risk of diabetes and improving one's quality of life. Various noninvasive glucose monitors for patients with diabetes mellitus have been studied in the last few decades [2]. The noninvasive concept was first described more than 30 years ago. Nevertheless, most of the current noninvasive technologies are still in their early stages of development. A list of noninvasive methods that have been developed is given in **Fig. 1** [3].

Optical techniques were initially investigated, for example, optical transmission spectroscopy, diffuse reflectance spectroscopy, optical coherence tomography, and photoacoustic spectroscopy (PAS). Non-optical techniques such as bioimpedance spectroscopy, microwave spectroscopy, ultrasound, and heat propagation were investigated later. Some of these techniques have been proven to work in *in vivo* studies. However, there is still no approved noninvasive testing method. For proof of applicability of such a method, it needs to be tested with a large number of people in actual usage situations, and the measurement instability arising from human physiological

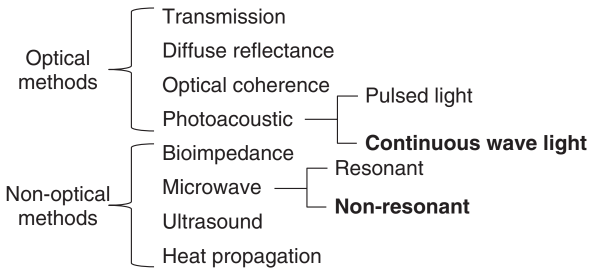


Fig. 1. Noninvasive glucose sensing methods.

characteristics needs to be resolved.

We describe here our recent research on noninvasive glucose monitoring, focusing in particular on near-infrared (NIR) PAS and microwave dielectric spectroscopy (DS). These technologies are advantageous in that they enable compact equipment designs that can be applied to wearable devices in the future because they can potentially use commercially available optical and wireless components used for telecommunication systems. We explain the sensing principles and the results of physiological-range glucose detection in *in vitro* samples using PAS and DS. Finally, we present *in vivo* results of each method compared with a commercial FGM.

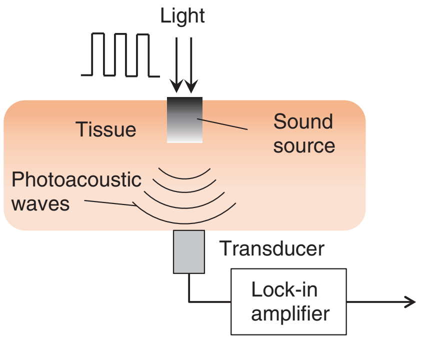


Fig. 2. PAS using continuous wave light.

2. PAS

In diffused reflectance optical spectroscopy, the repeatability of calibration models can be severely affected by skin conditions because transmitted light is affected by the variations in light scattered in skin and tissue. In contrast, PAS involves direct detection of acoustic waves propagating through tissue without them being scattered. It thus has potentially better tolerance. It has been demonstrated with nanosecond pulsed excitations in the visible, NIR, and mid-infrared regions [3, 4].

We focus here on NIR PAS. We investigated the use of semiconductor lasers of the sort found in optical telecommunications because they make it possible to produce sensors that are both portable and reliable. The concept is shown in **Fig. 2** [4]. When tissue is irradiated with two intensity modulated lasers in antiphase, a differential photoacoustic wave is generated. The traveling acoustic waves are detected with a transducer. We used laser lights at wavelengths of 1.38 and 1.61 μm. As shown in **Fig. 3**, the water absorption is equal at the two wavelengths, but the glucose absorption is slightly different. Therefore, the differential photoacoustic source mainly depends on the glucose concentration.

Photoacoustic wave generation involves light-tissue interaction and successive thermoelastic conversion processes. In light-tissue interaction, photons are instantly absorbed by molecules in the tissue and transformed into the kinetic energy of the molecules. Because the laser lights used in biomedical applications are low in power, other photon-molecular interactions such as radiative emission and chemical reactions are negligible. The signal-to-noise ratio (SNR) is determined by system parameters including device performance and signal processing parameters. To achieve a better SNR, the experimentally determined modulation frequency at which the transducer has the highest responsivity is used for each measurement. In addition, because the cell length L corresponds to the n-th acoustic longitudinal resonance mode in closed ends ($n = \omega L/\pi \nu$), the frequency also corresponds to the fifth longitudinal mode of the cell cavity [5]. This results in an SNR of more than 30 dB.

The results of *in vitro* measurement of glucose in aqueous solutions are shown in **Fig. 4**. We obtained an excellent linear response with correlation

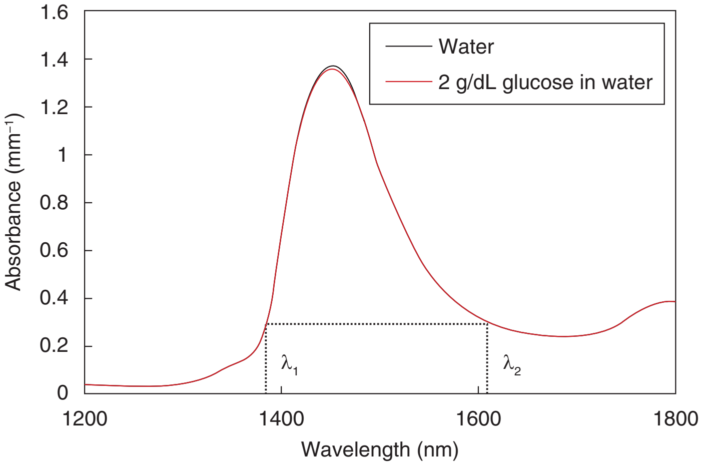


Fig. 3. NIR spectrum for pure water and 2 g/dL glucose concentration in water.

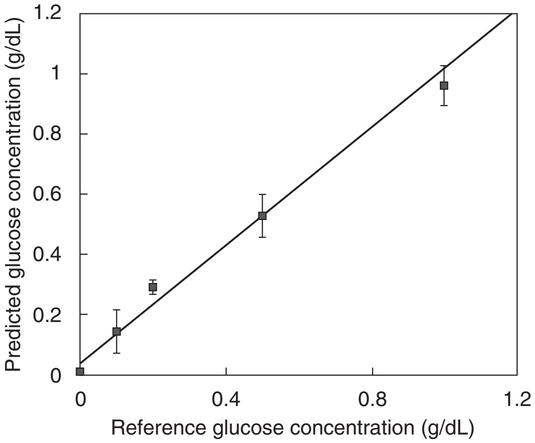


Fig. 4. Results of *in vitro* measurement of glucose using PAS.

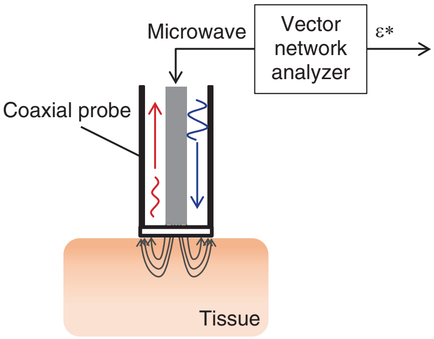


Fig. 5. DS using coaxial probe method.

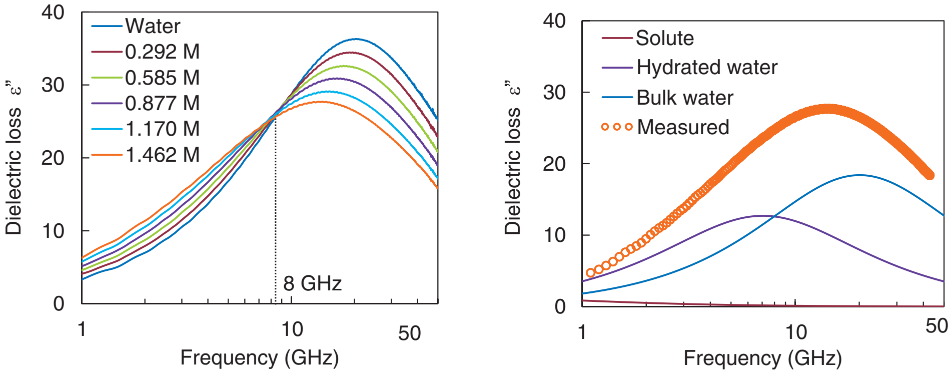
coefficient of 0.998 across the range of 100 mg/dL to 1.5 g/dL.

3. DS

Microwave glucose sensing has advantages over the use of optical sensors due to its lower energy per photon and smaller scattering from tissue. Two approaches to capturing the glucose footprint have been investigated. One involves measuring the frequency shift of the resonance in an LC (inductor-capacitor) resonator that corresponds to the change in permittivity at a specific frequency. The other involves measuring the complex permittivity over a wide frequency range by using the conventional methodology [6]. The former approach generally provides higher sensitivity to glucose thanks to its utilization of a high-Q (quality factor) resonator. However, it lacks selectivity to glucose.

The effects of other components need to be compensated for *in vivo* and under practical conditions. With blood components, large molecules such as proteins may degrade accuracy because they exclude larger volumes of water that are dominant components in the dielectric properties. We focus on the spectroscopy-based sensors that can potentially provide high selectivity using preprocessing based on spectroscopic analysis.

The experimental setup for carrying out DS of aqueous solutions using a coaxial probe and vector network analyzer is illustrated in **Fig. 5**. A one-port scattering parameter is measured, and the dielectric



(a)

(b)

Fig. 6. (a) Dielectric spectra of water and aqueous solution including glucose. (b) Peak assignments of 1.462 M of glucose concentration in water.

constant is calculated using measured data of calibration standards, air, electrical short circuits, and pure water.

The imaginary parts of measured broadband complex dielectric spectra for different glucose concentrations (0–1.462 M) are shown in **Fig. 6(a)**. Microwave DS is a potential tool for investigating the hydration dynamics of aqueous solutions to reveal the functionality of biomolecular systems. A comparison of the measured data versus the fitted curve based on a linear combination of Debye's relaxation model is shown in **Fig. 6(b)** [7]. As the glucose concentration increases, the glucose-hydration water absorption at around 12 GHz increases. In contrast, bulk water absorption at around 20 GHz decreases. As a result, the peak of the dielectric spectrum shifts to a lower frequency and is broader than that of pure water.

The results of *in vitro* measurement of glucose in aqueous solutions using DS in the frequency range of 500 MHz to 50 GHz are presented in **Fig. 7**. We obtained a linear response with a linear regression correlation coefficient of 0.987 over the range of 50 mg/dL to 1.0 g/dL. One difficulty with taking measurements on a living body is signal drift due to changes in temperature, water content, and other factors. We proposed a drift correction technique that has been standardized using the dielectric constant at 8 GHz [8]. Since the frequency point is the isosbestic point of the aqueous solution of glucose, the effect of water content on measurement is expected to be suppressed. We conducted *in vivo* measurements using only two frequency points in view of the expected future miniaturization of the system.

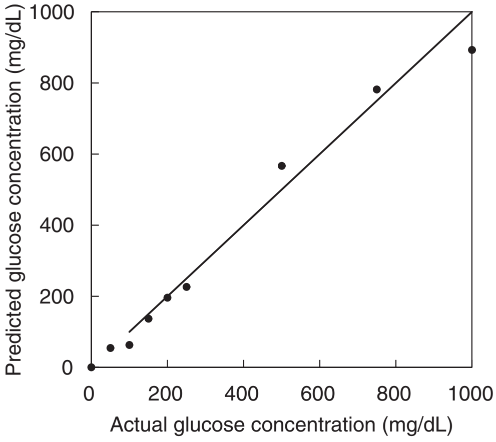
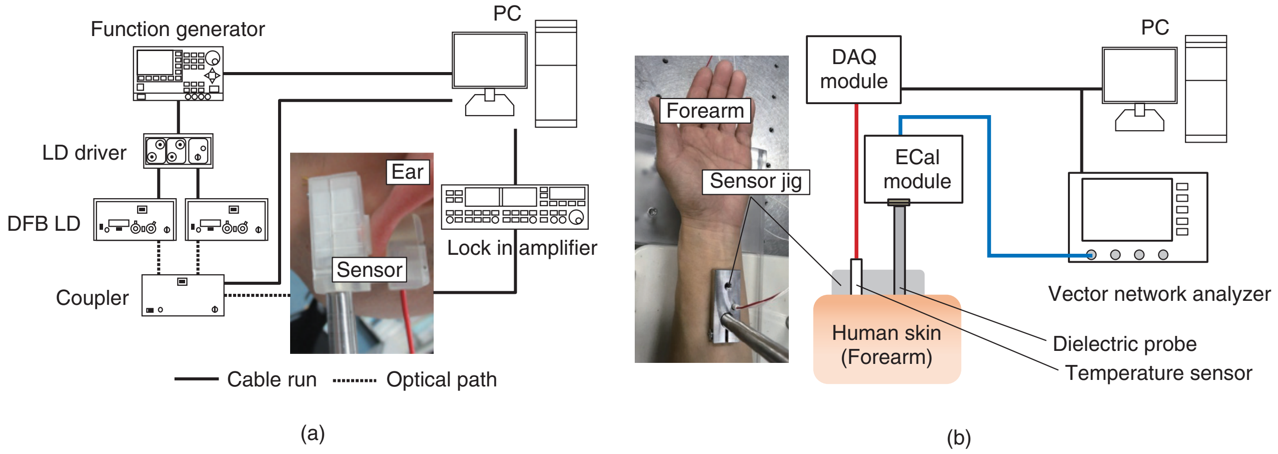


Fig. 7. Results of *in vitro* measurement of glucose using DS.

4. *In vivo* measurement and results

Ethical approval was obtained from the ethics committee at the University of Tokyo for *in vivo* measurements. All experiments were performed in accordance with relevant guidelines and regulations. Volunteers were healthy males and females aged from 21 to 31, and informed consent was obtained. We conducted oral glucose tolerance tests (OGTTs) to compare the results of PAS and DS with commercially available invasive glucose sensors. The glucose level during the OGTTs was monitored with a commercial FGM (FreeStyle Libre, Abbott Co.) [9]. A needle sensor was inserted into tissue under the skin of the left



DAQ: data acquisition
DFB: distributed feedback
ECal: electronic calibration
LD: laser diode
PC: personal computer

Fig. 8. Schematic image of experimental configuration of (a) PAS and (b) DS.

upper arm, and the glucose level in tissue was measured every five minutes.

The experimental setup of the noninvasive glucose monitor is shown in **Fig. 8** [8, 10]. The sensor interfaces of the PAS and DS were attached to the left earlobe (Fig. 8(a)) and forearm (Fig. 8(b)), respectively. For the OGTTs, liquid glucose containing 75 g of glucose (TRELAN G75, AY PHARMACEUTICALS CO., LTD.) was taken orally by the volunteers 30 minutes after the start of the measurement of PAS and DS to determine the base line. The signals were collected every minute, and the total duration of the test was three hours.

The temperature of the sample is uncontrollable in *in vivo* measurement, and the effect of tissue temperature on the measurement results is unknown. Temperature sensors were therefore integrated with each system, and the measured signals were corrected using the temperature data. The results of min-max scaling of PAS, DS, and FGM signals are plotted in **Fig. 9**. Our proposed measurement techniques show good traceability to the FGM. The respective correlation coefficients of PAS and DS were 0.94 and 0.84. The signals of noninvasive measurement appeared to be delayed by 10–20 min compared with FGM. We considered that this delay was caused by a difference in the penetration depth. FGM generally has a 10–15 min time lag in showing changes in the glucose level because of the time taken for blood components to travel from blood vessels to interstitial fluid. The time lag seen in Fig. 9 is thought to be due to the traveling time. However, the specific cause of this phenomenon is not clear. We will continue researching this to elucidate a measurement mechanism.

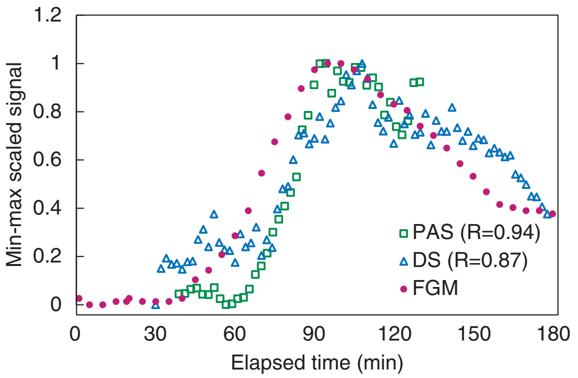


Fig. 9. Results of *in vivo* measurement: comparison of scaled signals for PAS, DS, and FGM.

5. Conclusion

We introduced a means of noninvasive glucose measurement using PAS and DS. The measurement techniques were proposed to suppress the instability from biomedical samples, including the human body itself, for example, water absorption and signal drift during measurement. The results of *in vivo* measurement showed a high correlation between both techniques and the FGM. This result indicates the feasibility of using a noninvasive glucose sensor to visualize changes in glucose level. In the future, we plan to work on developing a calculation algorithm to determine absolute values and to miniaturize the system to a size small enough to be used as a wearable device.

Acknowledgment

We thank Dr. Kayo Waki and Ms. Ritsuko Nagatomo of the University of Tokyo for their help in conducting the *in vivo* measurements and for their valuable advice on glucose metabolism and glucose monitoring equipment.

References

[1] International Diabetes Federation, "IDF Diabetes Atlas," 8th edition, 2017.

[2] W. V. Gonzales, A. T. Mobashshe, and A. Abbosh, "The Progress of Glucose Monitoring—A Review of Invasive to Minimally and Non-invasive Techniques, Devices and Sensors," Sensors, Vol. 19, No. 4, 800, 2019.

[3] T. Tajima, M. Nakamura, Y. Tanaka, and M. Seyama, "Advances in Noninvasive Glucose Sensing Enabled by Photonics, Acoustics, and Microwaves," Int. J. of Autom. Technol., Vol. 12, No. 1, pp. 64–72, 2018.

[4] T. Tajima, Y. Okabe, Y. Tanaka, and M. Seyama, "Linearization Technique for Dual-wavelength CW Photoacoustic Detection of Glucose," IEEE Sensors J., Vol. 17, No. 16, pp. 5079–5086, 2017.

[5] Y. Tanaka, T. Tajima, and M. Seyama, "Acoustic Modal Analysis of Resonant Photoacoustic Spectroscopy with Dual-wavelength Differential Detection for Noninvasive Glucose Monitoring," IEEE Sens. Lett., Vol. 1, No. 3, 3500704, 2017.

[6] M. Nakamura, T. Tajima, K. Ajito, and H. Koizumi, "Selectivity-enhanced Glucose Measurement in Multicomponent Aqueous Solution by Broadband Dielectric Spectroscopy," IEEE MTT-S International Microwave Symposium, San Francisco, CA, USA, May 2016.

[7] K. Shiraga, T. Suzuki, N. Kondo, T. Tajima, M. Nakamura, H. Togo, A. Hirata, K. Ajito, and Y. Ogawa, "Broadband Dielectric Spectroscopy of Glucose Aqueous Solution: Analysis of the Hydration State and the Hydrogen Bond Network," J. Chem. Phys., Vol. 142, 234504, 2015.

[8] M. Nakamura, T. Tajima, M. Seyama, and K. Waki, "A Noninvasive Blood Glucose Measurement by Microwave Dielectric Spectroscopy: Drift Correction Technique," 2018 IEEE International Microwave Biomedical Conference, Philadelphia, PA, USA, June 2018.

[9] Abbott Diabetes Care (in Japanese), http://myfreestyle.jp/

[10] Y. Tanaka, T. Tajima, M. Seyama, and K. Waki, "In-vivo Study on Resonant Photoacoustic Spectroscopy Using Dual CW Light Wavelengths for Non-invasive Blood Glucose Monitoring," 2018 IEEE Sensors, New Delhi, India, Oct. 2018.

Masahito Nakamura

Researcher, NTT Device Technology Laboratories and NTT Biomedical Informatics Research Center.

He received a B.E. and M.E. in electrical and electronic engineering from Tokyo Metropolitan University in 2011 and 2013. He joined NTT Microsystem Integration Laboratories in 2013. His research interests include microwave, millimeter-wave, and terahertz wave measurement for nondestructive and noninvasive applications. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE).

Yujiro Tanaka

Researcher, NTT Device Technology Laboratories and NTT Biomedical Informatics Research Center.

He received a B.S. and M.S. in engineering from Tohoku University, Miyagi, in 2011 and 2013. He joined NTT Microsystem Integration Laboratories in 2013. He has been with NTT Device Technology Laboratories since 2014, where he is currently researching photoacoustic technology for biosensing. He is a member of the Japan Society of Applied Physics (JSAP).

Takuro Tajima

Senior Research Engineer, NTT Device Technology Laboratories.

He received a B.S. and M.E. from the University of Tokyo in 2000 and 2002, and a Ph.D. from Tokyo Institute of Technology in 2017. He joined NTT Telecommunication Energy Laboratories in 2002, where he has been engaged in the research and development of laser photoacoustic spectroscopy for biomedical sensing, sub-terahertz antenna-in packages using substrate integrated waveguide technology, and a broadband dielectric spectroscopic system using photonic integration technologies. His current research involves multi-modal biomedical analysis using millimeter waves, terahertz waves, optics, and ultrasonics. He is a member of IEICE and JSAP.

Michiko Seyama

Senior Research Engineer, Supervisor, NTT Device Technology Laboratories and NTT Biomedical Informatics Research Center.

She received a B.S., M.E., and Ph.D. from Waseda University, Tokyo, in 1995, 1997, and 2004. She joined NTT in 1997 and worked on an odor sensing system based on plasma-polymerized organic film for environmental contamination and biogas, and a one-dimensional-surface plasmon resonance biosensor combined with a microfluidic device for on-site immunological sensing. Her current interest is the development of biosensing platforms for blood components and blood coagulation including both invasive and noninvasive technologies. She is a member of JSAP, the Electrochemical Society of Japan, and the American Chemical Society.

An Update on Open Source Communities Engaged in SDN/NFV, with a Focus on the Open Networking Foundation

Dai Kashiwa and Wenyu Shen

Abstract

The Open Networking Foundation (ONF) was founded in 2011, and since then, numerous open source communities focusing on software-defined networking (SDN) and network functions virtualization (NFV) have been launched and are becoming more active. This article first gives an overview of the open source communities concerned with SDN/NFV technologies and then describes the latest activities of the ONF and the initiatives undertaken by the NTT Group. The activities of the MEF (Metro Ethernet Forum) are also explained as an example of work pursued by the NTT Group.

Keywords: open source community, SDN/NFV, Open Networking Foundation

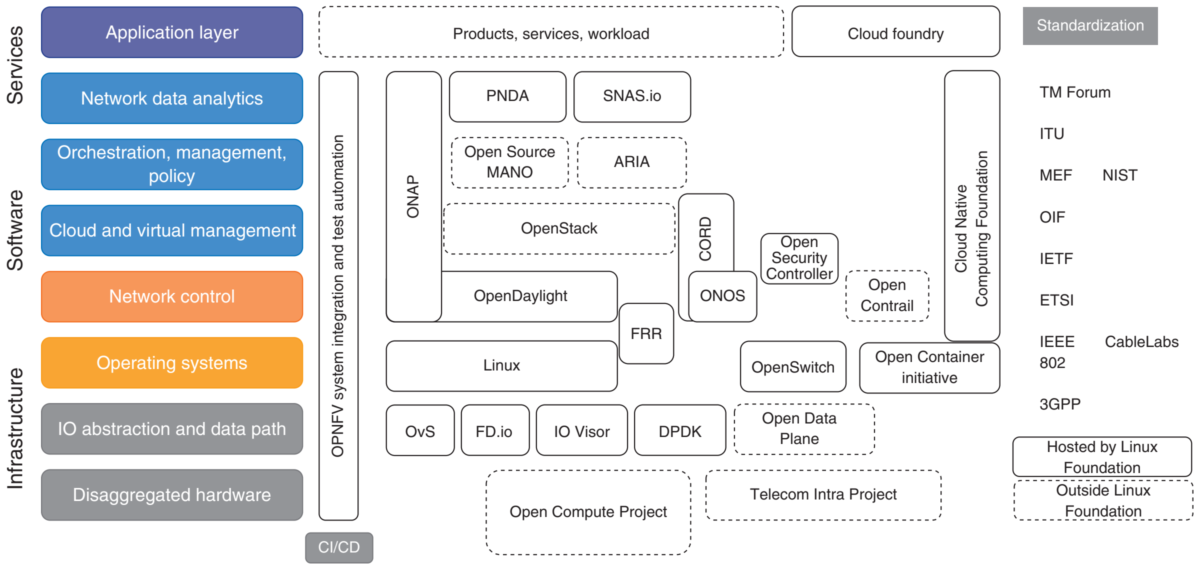
1. Introduction

The application of software-defined networking (SDN) and network functions virtualization (NFV) technologies is expanding, and in line with this trend, a variety of open source development projects and standardization projects have been launched. The results from these projects are being introduced in the construction of service systems by communication carriers, cloud providers, and enterprise users. These movements started with the development of SDN controllers such as the OpenDaylight and ONOS (Open Network Operating System) projects, after which they expanded to include the development of high-speed data planes, as represented by Open vSwitch and DPDK (Data Plane Development Kit), and the development of NFV infrastructures such as OPNFV (Open Platform for NFV).

These open source development activities have been linked with the standardization efforts of organizations such as ETSI (European Telecommunications Standards Institute) and the TMF Forum, creating a cycle of activities whereby standardization has influenced implementation, and the work of implementation has been fed back into the standardization process. Today, there is intense interest in orchestration platforms, as represented by ONAP (Open Network Automation Platform), and container management systems, as represented by Kubernetes.

The mappings of major networking-related projects, as compiled by the Linux Foundation (LF) [1], are shown in **Fig. 1**. There are projects in each of the three layers: the infrastructure layer consisting of devices, data transfer, and operating systems; the software layer consisting of network control, cloud platforms, and orchestration frameworks; and the service layer consisting of data analytics and applications. As the use of 5G (fifth-generation mobile communications), Internet of Things, and artificial intelligence technologies becomes more and more widespread, attention has been drawn to edge computing. As a result, many edge-related projects have been launched under the LF.

In January 2019, a new organization called LF Edge was established to promote collaboration between different projects. It consolidated those



Source: Created from document presented by Arpit Joshipura of Linux Foundation at 2018 OCP Summit

ARIA: Agile Reference Implementation of Automation
CI/CD: continuous integration/continuous delivery
CORD: Central Office Re-architected as a Datacenter
FD.io: Fast Data - input/output
FRR: FRRouting (Free Range Routing)
IEEE: The Institute of Electrical and Electronics Engineers
IETF: The Internet Engineering Task Force
IO: input/output
ITU: The International Telecommunication Union
MANO: management and orchestration
MEF: Metro Ethernet Forum
NIST: National Institute of Standards and Technology
OIF: Optical Internetworking Forum
OvS: Open vSwitch
PNDA: Platform for Network Data Analytics
SNAS: Streaming Network Analytics System
3GPP: 3rd Generation Partnership Project

Fig. 1. Open source networking landscape.

projects that developed software programs to be used in edge computing. Five projects are in progress under the umbrella of LF Edge: Akraino Edge Stack, EdgeX Foundry, Home Edge Project, Open Glossary of Edge Computing Project, and EVE (Edge Virtualization Engine) [2].

2. Open Networking Foundation (ONF)

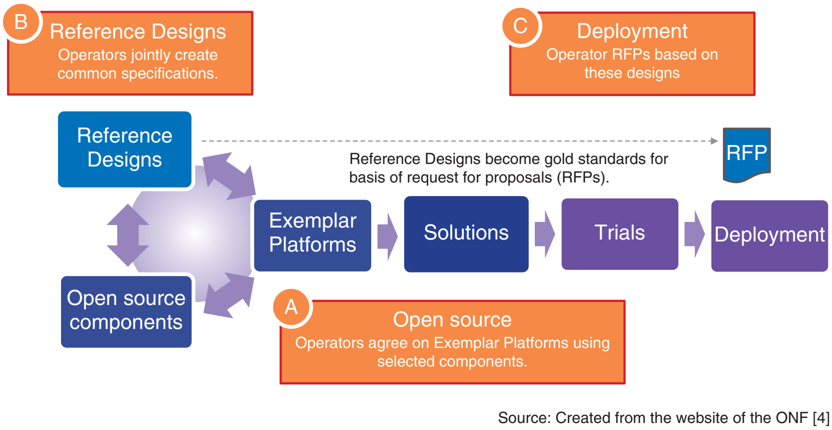
In October 2016, the ONF, which mainly worked on standardization, was merged with ON. Lab (Open Networking Lab), which was involved in open source development. Since then, the new ONF has been working on open source development and also specifying the design information obtained in the development process by publishing standards documents [3]. In March 2018, it announced the ONF Strategic Plan. This plan declares that operators (AT&T, Deutsch Telekom, NTT Group, etc.), participating as partners, will define common requirements, and that the ONF will work with ONF members, including vendors and system integrators, to create Reference Designs (RDs) and Exemplar Platforms (EPs), which are both needed for implementing these requirements (**Fig. 2**).

An RD defines software components needed for implementing specific use cases and also specifies the functional requirements for these components and inter-component interfaces. RD documents are released to ONF members. An EP is an aggregation of open source software programs that implement the RD and is core software used for the development of commercial products. The ONF aims to provide operators with systems that are based on open technologies (open source software, white boxes, and network disaggregation) by creating a cycle of design, development, deployment, and maintenance founded on the RDs and EPs [4].

As of June 2019, the ONF handles five use cases: SDN Enabled Broadband Access (SEBA), Trellis, Open and Disaggregated Transport Network (ODTN), next-generation SDN (NG-SDN), and Converged Multi-Access and Core (COMAC).

(1) SEBA

This use case virtualizes the functions of the optical line termination, broadband network gateway, and other components and implements access network


B
Reference Designs
Operators jointly create common specifications.
C
Deployment
Operator RFPs based on these designs
Reference Designs
Reference Designs become gold standards for basis of request for proposals (RFPs).
RFP
Exemplar Platforms
Solutions
Trials
Deployment
Open source components
A
Open source
Operators agree on Exemplar Platforms using selected components.

Source: Created from the website of the ONF [4]

Fig. 2. ONF's strategic plan.

technologies such as passive optical networks and G. Fast by combining open source programs. It also achieves a high-speed and seamless connection with the backhaul. This work is led by AT&T and Deutsche Telekom. The ONF devotes the greatest proportion of its resources to this use case.

(2) Trellis

The objective of this use case is to develop a leaf-spine fabric for NFV using open technologies. It implements routing (Border Gateway Protocol and Segment Routing), Q-in-Q control, and a dual-homing function.

(3) ODTN

This use case is aimed at disaggregating transmission network devices and achieving interactions between components and controllers through an open application programming interface (API). This is described in more detail in the following section.

(4) NG-SDN

This use case is designed to make the data plane programmable by using P4-language-based white box switches, network operating systems, and controllers.

(5) COMAC

This use case is intended to terminate and manage mobile and wireline access networks seamlessly and to provide network slices. It provides the data plane with programmability based on the P4 language. It also aims at providing customer management, a mobility management entity, home subscriber server, and other functions as a common platform.

3. NTT Group's initiatives for collaboration with open communities

NTT Group companies have been collaborating with various open source communities for many years. Some of these collaborations are described in more detail in this section.

3.1 Collaboration with ONF

NTT Communications has participated in the ONF since the inception of the latter in 2011. It has been actively involved in ONF activities and holds the status of a partner (the highest level of participation, serving as a board member and thus being involved in the ONF's decision-making). The entire NTT Group has been a partner since 2017, thereby further increasing its involvement. NTT Communications has launched and led the ODTN project as part of its effort to implement a transport network using open technologies.

With the aim of making the data plane programmable, NTT EAST and NTT Network Service Systems Laboratories are participating in the NG-SDN project and driving its technical studies, with a focus on P4. NTT WEST also takes part in the NG-SDN project. NTT Access Network Service Systems Laboratories is involved in the SEBA project and contributes to the formulation of the RD and implementation of parts of the EP.

3.2 ODTN

Transport networks comprise optical transmission systems such as transponders and ROADM (reconfigurable optical add/drop multiplexers). Most of

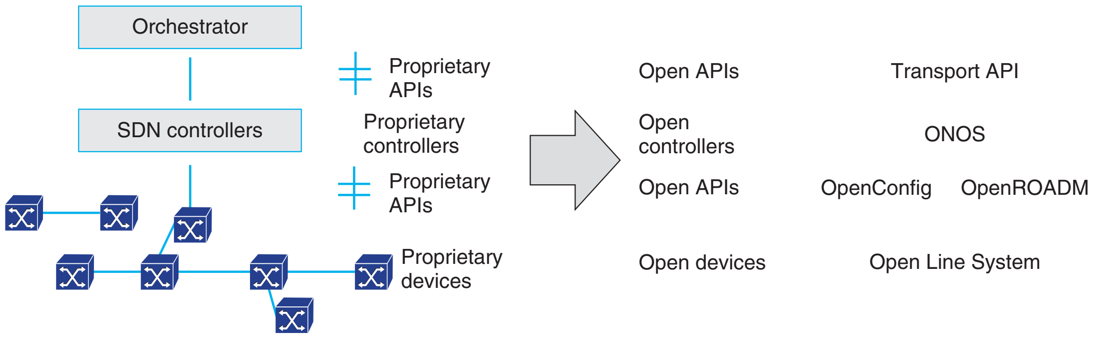


Fig. 3. Configuration of ODTN.

these network devices and control software programs, for example, element management systems and network management systems, are implemented by vendors in a proprietary manner with vertical integration, which hinders interoperability or open architecture. The results are strong vendor lock-in and long update cycles, making it difficult to introduce leading-edge technologies and products at short intervals.

The idea of disaggregating functions has emerged as a way to solve these problems. Several initiatives are in progress. The Open Line System (OLS) [5] separates transponders from transport systems—which have conventionally integrated the functions of transponders, multiplexers, demultiplexers, and amplifiers—so that multi-vendor transponders can coexist in a single transport system. The OpenConfig [6] working group defines the data model and API for open optical transport, making it possible to control and manage compliant devices in an integrated manner, and to monitor and collect data from these devices using telemetry.

The ODTN is a technical development project aimed at innovating transport networks end-to-end by collaborating in the above-mentioned activities, using an aggregation of open technologies and open source software, including controllers [7]. The correspondence between component devices and open technologies used is shown in **Fig. 3**. The ODTN project plans to integrate the devices, API definitions, and controllers shown in Fig. 3, conduct a technical verification to demonstrate technical feasibility, and provide the related initiatives with detailed design values and feedback on the problems and requirements identified through technical evaluation. To date, it has completed Phase 1, in which an initial technical verification is carried out for the use case of a point-to-point connection using transponders and OLS. It is now endeavoring to enhance quality and expand use cases.

3.3 Collaboration with the Metro Ethernet Forum (MEF)

The MEF is a not-for-profit organization established in 2001 that currently has a membership of 220 companies. It formulated the device specifications needed for providing Carrier Ethernet services. It has now added elements of SDN and NFV to the specifications and announced life service orchestration (LSO), a concept model for building a network that features high agility. The MEF defines the functional requirements for LSO and APIs that support them in order to provide for end-to-end orchestration between the networks of different operators. Together with the NTT laboratories, NTT Communications has begun activities to use these APIs to achieve interconnections and collaboration between mobile and fixed network slices as well as activities to use these APIs for interconnection between software-defined wide area networks (SD-WANs).

More than 40 companies are currently producing individual solutions in the rapidly growing SD-WAN market. Although the number of operators that handle multiple solutions are on the rise, they face serious problems related to delivery and operations. For example, since the type of customer premises equipment (CPE) varies from solution to solution, these operators need to have a sales and delivery organization specific to each solution. Also, each time a new solution is introduced, they need to make a large investment to develop new peripheral systems such as operation support systems and business support systems.

The MEF is standardizing an SD-WAN data model

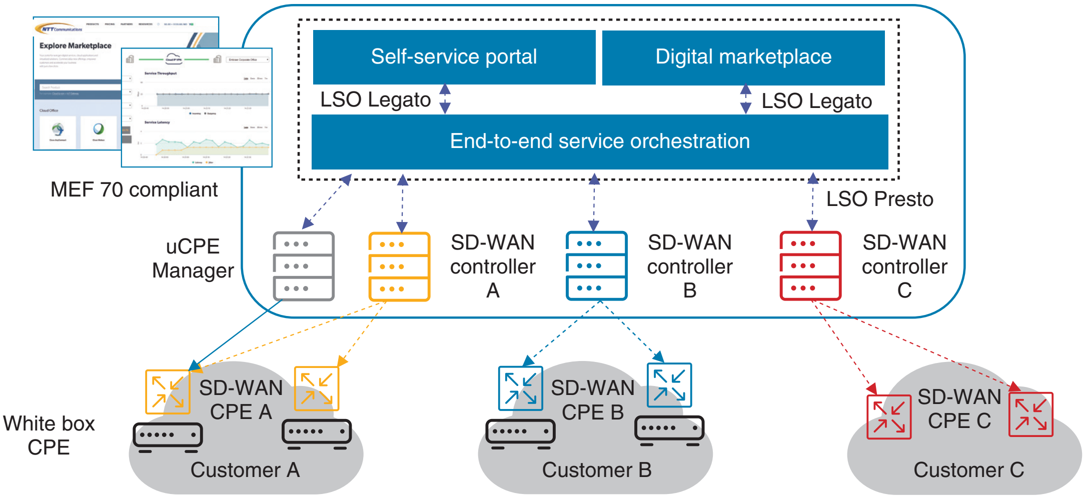


Fig. 4. Proof-of-concept configuration of multi-vendor SD-WAN service based on white boxes.

and various interfaces, thereby offering the possibility to rectify the above problems. NTT Communications started a project called a *multi-vendor SD-WAN service based on white boxes*. The aim of the project is to evaluate and verify a universal CPE platform that enables multiple SD-WAN solutions to be implemented on white boxes; MEF SD-WAN Presto, an interface that enables multiple SD-WANs to be controlled in an integrated manner using the same portal; and SD-WAN services that conform to MEF 70 standards (**Fig. 4**). A proof of concept will be demonstrated at an MEF 19 event to be held in November 2019 [8].

4. Future outlook

This article described open source community activities in the area of SDN/NFV technologies, the latest developments in the ONF, and initiatives taken by the NTT Group. Community activities will reinforce the trend toward developing software components through open innovation, which in turn will invigorate community activities. There have been many cases recently where the subject areas of different communities overlap. It is expected that communities will increase their efforts to collaborate or merge with other communities to allocate responsibilities appropriately.

References

[1] A. J. Weissberger, “OCP – Linux Foundation Partnership Accelerates Megatrend of Open Software Running on Open Hardware,” The IEEE ComSoc Technology Blog, Mar. 2018. https://techblog.comsoc.org/2018/03/26/ocp-linux-foundation-partnership-accelerates-megatrend-of-open-software-running-on-open-hardware/

[2] The Linux Foundation, “The New Open ‘Edge’,” https://www.lfedge.org/wp-content/uploads/2019/06/LF-Edge-web-june.pdf

[3] The ONF, https://www.opennetworking.org/softwaredefined-standards/overview/

[4] Reference Designs, ONF, https://www.opennetworking.org/reference-designs/

[5] G. Bennett, “Open Line Systems and Open ROADM: How Open Is Your Line System?”, https://tnc18.geant.org/getfile/4520

[6] OpenConfig, http://www.openconfig.net/

[7] ODTN, https://www.opennetworking.org/odtn/

[8] MEF, https://www.mef.net/mef-3-0-sd-wan

Trademark notes

Dai Kashiwa

Vice President of SDN/NFV Technology Development, NTT Communications Corporation.

He received a Ph.D. in engineering in 2003. After joining NTT in 1997, he was involved in researching network management systems, network security, and active networks. He has been with NTT Communications since 2004, where he has been developing business network services including video broadcasting, dynamic VPN (virtual private network) and SDN services. He currently leads a number of incubation and software development projects using SDN/NFV technologies.

He has been a board member of ONF since 2015 and is leading the ODTN project, an operator-led initiative to build datacenter interconnects using disaggregated optical equipment, open and common standards, and open source software. He was successively appointed to the ONOS/CORD board and the OpenDaylight user advisory board.

Wenyu Shen

Manager of SDN/NFV Technology Development, NTT Communications Corporation.

He has over 10 years of experience in the telecommunications industry and currently serves as a technology development manager at NTT Communications. In this role, he drives NTT Communications' SDN/NFV strategy and leads a team developing a next generation SD-WAN and NFV service platform. He is also actively involved with open source communities and standardization bodies including ONF and MEF. As a representative of the NTT Group, he is currently a member of the technical leadership team at ONF.

Prior to joining NTT Communications, he was with NTT Network Innovation Laboratories, where he oversaw many core research projects including a European FP7 project, covering generalized multiprotocol label switching, operation support systems/business support systems, and network virtualization. He holds more than 10 patents and is the author of numerous papers and presentations on network architecture and design.

Event Report: NTT Communication Science Laboratories Open House 2019

Atsunori Ogawa, Xiaomeng Wu, Masaaki Nishino, Mathieu Blondel, and Takemi Mochida

Abstract

NTT Communication Science Laboratories Open House 2019 was held at Keihanna Science City, Kyoto, on May 30 and 31, 2019. Around 1500 visitors enjoyed 5 talks and 30 exhibits, which included our latest research efforts in the fields of information and human sciences.

Keywords: information science, human science, artificial intelligence

1. Overview

NTT Communication Science Laboratories (CS Labs) aims to establish technologies that enable *heart to heart* communication between people and people, and between people and computers. We are thus working on a fundamental theory that approaches the essence of human beings and information, as well as on innovative technologies that will transform society.

NTT CS Labs Open House has been held annually with the aim of introducing the results of the CS Labs' basic research and innovative leading-edge research to both NTT Group employees and visitors from business industries, universities, and research institutions who are engaged in research, development, business, and education.

Open House 2019 was held at the NTT Keihanna Building in Kyoto on May 30 and 31, and around 1500 visitors attended it over the two days. This year, we invited the former athlete, Deportare Partners Representative, Mr. Dai Tamesue, and held a special talk with NTT Fellow Dr. Makio Kashino. We also tried an outdoor demonstration exhibition for the first time. We prepared many hands-on exhibits to enable visitors to intuitively understand our latest research results and to share a vision of the future where new products based on the research results are widely used. This article summarizes the event's research talks and exhibits.

2. Keynote speech

The event started with a speech by the Vice President and head of NTT CS Labs, Dr. Takeshi Yamada, entitled "Processing like people, understanding people, helping people—Toward the future where humans and AI will cohabitate and co-create" (**Photo 1**).

Dr. Yamada pointed out that CS Labs places particular importance on basic research not only toward the development of technology that can approach human abilities but also technology that can be used to elucidate human functions and characteristics and to understand what it means to be human, and technology to help people in their daily lives. In Japan, 2019 has turned out to be both a year marking Reiwa, a new name in the traditional Japanese era system, and a year on the verge of holding the Olympic and Paralympic Games. Dr. Yamada introduced the latest

Photo 1. Dr. Takeshi Yamada delivering keynote speech.

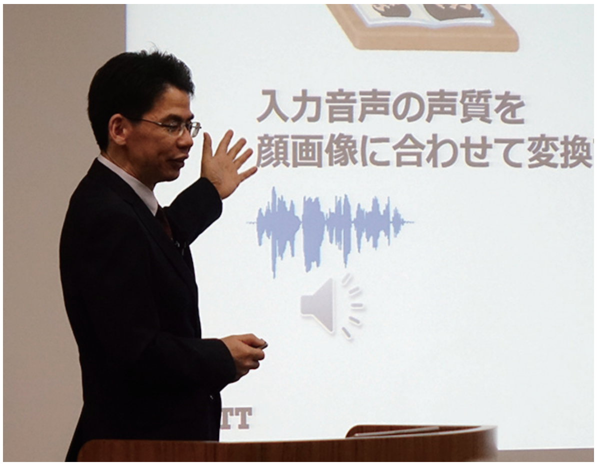


Photo 2. Dr. Kunio Kashino giving research talk.

artificial intelligence (AI) technologies developed by CS Labs at this transition point between eras and declared that they will boldly and tenaciously undertake new challenges with a focus on technologies that carry out processing the way people do, and that also understand and help people.

3. Research talks

Three research talks were given, as summarized below, which highlighted recent significant research results and high-profile research themes. Each presentation introduced some of the latest research results and provided some background and an overview of the research. All of the talks were very well received.

(1) "See, hear, and learn to describe—Crossmodal information processing opens the way to smarter AI," by Dr. Kunio Kashino, Media Information Laboratory

Recent advances in AI and machine learning research are breaking down the barriers between modalities such as language, sounds, and images, which have been studied separately so far. There are various implications for this dramatic change. Most importantly, AI is now achieving a new breakthrough—its ability to learn concepts on its own based on multimodal inputs. The key is to acquire, analyze, and utilize common representations that can be shared among those multiple modalities. In this research talk, Dr. Kashino called this approach crossmodal information processing, introduced its concept, and discussed how it will help people in their future lives (**Photo 2**).

(2) "Measuring multiple visual abilities in daily circumstances—Towards establishment of daily selfcheck for eye health," by Dr. Kazushi Maruya, Human Information Science Laboratory

The human visual system has considerable interpersonal differences, and its ability varies with the context, task, and circumstances. To grasp the variability in visual ability in daily circumstances, a novel set of eye health-check tests are proposed to measure visual abilities. Each test can be finished in a short time (around 3 min), and some tests are gamified so that users can check their visual ability in an enjoyable way. In this research talk, Dr. Maruya explained the details of this self-check test battery and clarified problems that should be overcome in order to establish daily self-checks of eye health (**Photo 3**).

(3) "Like likes like strategy: search suitable for various viewpoints—Picture book search system 'Pitarie' with graph index based search," by Mr. Takashi Hattori, Innovative Communication Laboratory

A novel method is proposed for finding similar objects in a large-scale database, based on a graph index. Each vertex corresponds to an object, and two similar vertices are likely to be connected. The graph index, constructed using a *like likes like* strategy, shows small-world behavior: any two vertices can be connected by a small number of steps. A graph index search terminates quickly and is applicable to many types of media, including text, images, and audio. In this research talk, Mr. Hattori introduced Pitarie, an

Photo 3. Dr. Kazushi Maruya giving research talk.

Photo 4. Mr. Takashi Hattori giving research talk.

application of this proposed method that enables searching for similar picture books by both text and images (**Photo 4**).

4. Research exhibits

The Open House featured 30 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 Human.

Each exhibit was housed in a booth and employed techniques such as slides presented on a large-screen monitor or hands-on demonstrations, with researchers explaining the latest results directly to visitors (**Photos 5** and **6**). The following list, taken from the Open House website [1, 2], summarizes the research exhibits in each category. (Abbreviations in the titles have been defined.)

4.1 Science of Machine Learning

- Learning and finding congestion-free routes—Online shortest path algorithm with binary decision diagrams
- Efficient and comfortable AC (air conditioning) control by AI—Environment reproduction and control optimization system
- Recover urban people flow from population data—People flow estimation from spatiotemporal population data
- Improving the accuracy of deep learning—Larger capacity output function for deep learning
- Which is cause? Which is effect? Learn from data!—Causal inference in time series via supervised learning
- Forecasting future data for unobserved locations—Tensor factorization for spatio-temporal data analysis
- Search suitable for various viewpoints—"Pitarie": Picture book search with graph index based search

4.2 Science of Communication and Computation

- We can transmit messages to the efficiency limit—Error correcting code achieving the Shannon limit
- New secrets threaten past secrets—Vulnerability assessment of quantum secret sharing
- Analyzing the discourse structure behind the text—Hierarchical top-down RST (rhetorical structure theory) parsing based on neural networks
- When children begin to understand hiragana—Emergent literacy development in Japanese
- Measuring emotional response and emotion sharing—Quantitative assessment of empathic communication
- Touch, enhance, and measure the empathy in crowd—Towards tactile enhanced crowd empathetic communication
- Robot understands events in your story—Chat-oriented dialogue system based on event understanding

4.3 Science of Media Information

- Voice command and speech communication in

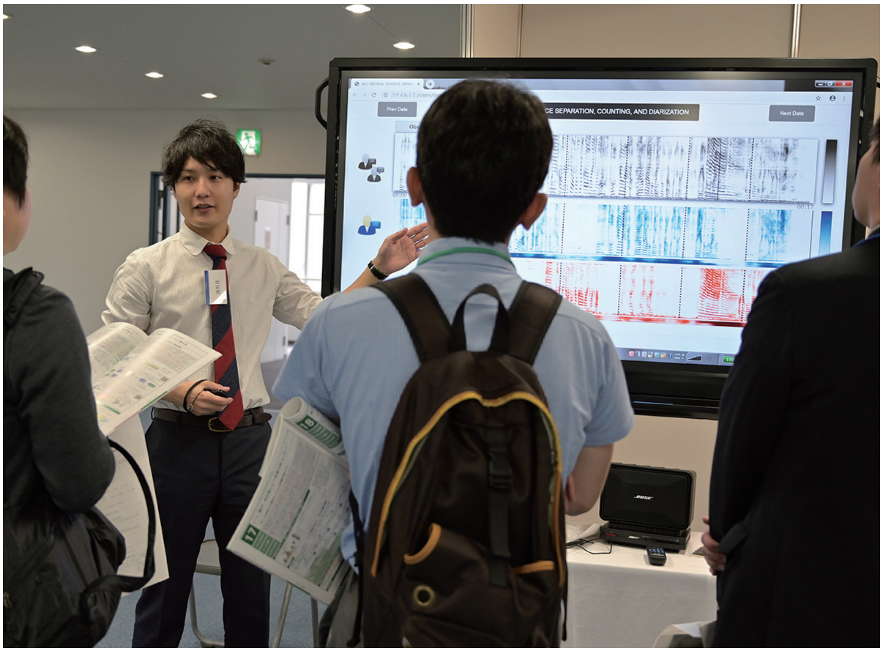


Photo 5. Researcher explaining a demonstration.

Photo 6. The latest research results were exhibited.

car—World’s best voice capture and recognition technologies

- Learning speech recognition from small paired data—Semi-supervised end-to-end training with text-to-speech
- Who spoke when & what? How many people

were there?—All-neural source separation, counting and diarization model
- Changing your voice and speaking style—Voice and prosody conversion with sequence-to-sequence model
- Face-to-voice conversion and voice-to-face conversion—Crossmodal voice conversion with deep generative models
- Learning unknown objects from speech and vision—Crossmodal audio-visual concept discovery
- Neural audio captioning—Generating text describing non-speech audio
- Recognizing types and shapes of objects from sound—Crossmodal audio-visual analysis for scene understanding

4.4 Science of Human

- Speech of chirping birds, music of bubbling water—Sound texture conversion with an auditory model
- Danswing papers—An illusion to give motion impressions to papers
- Measuring visual abilities in a delightful manner—Self eye-check system using video games and tablet PCs (personal computers)
- How do winners control their mental states?—Physiological states and sports performance in real games
- Split-second brain function at baseball hitting—Instantaneous cooperation between vision and action
- Designing technologies for mindful inclusion—How sharing caregiving data affects family communication
- Real-world motion that the body sees—Distinct visuomotor control revealed by natural statistics
- Creating a walking sensation for the seated—A sensation of pseudo-walking expands peripersonal space

5. Invited talk

This year, NTT Fellow Dr. Makio Kashino invited the former athlete, Deportare Partners Representative, Mr. Dai Tamesue and held a special talk entitled “Sports in the future and human potentiality.” A wide range of topics was addressed, including *mental* control, record-breaking discontinuities, collaborative behavior, limit of growth due to premature optimization, language and sports skill communication, and balance of consciousness and unconsciousness. As an athlete who represented Japan in three Olympic Games and as a scientist who is studying brain functions underlying the amazing skills of athletes, respectively, Mr. Tamesue and Dr. Kashino talked about the essence of human nature revealed by the athletic practice and scientific research of sports, and shared their predictions of how science and technology will change sports and humans in the future. The audience engagement was high thanks to the vivid discussion that included a lot of physical gestures.

6. Concluding remarks

Just like last year, many visitors came to NTT CS Labs Open House 2019 and engaged in lively discussions on the research talks and exhibits and provided many valuable opinions on the presented results. In closing, we would like to offer our sincere thanks to all of the visitors and participants who attended this event.

References

[1] Website of NTT Communication Science Laboratories Open House 2019 (in Japanese).
http://www.kecl.ntt.co.jp/openhouse/2019/index.html

[2] Website of NTT Communication Science Laboratories Open House 2019 (in English).
http://www.kecl.ntt.co.jp/openhouse/2019/index_en.html

Atsunori Ogawa

Senior Research Scientist, Signal Processing Research Group, Media Information Laboratory, NTT Communication Science Laboratories.

He received a B.E. and M.E. in information engineering and a Ph.D. in information science from Nagoya University, Aichi, in 1996, 1998, and 2008. He joined NTT in 1998. He has been engaged in research on speech recognition and speech enhancement at NTT Cyber Space Laboratories (now, NTT Media Intelligence Laboratories) and NTT Communication Science Laboratories. His research interests include speech recognition, speech enhancement, and spoken language processing.

Xiaomeng Wu

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

He received a B.S. in energy and power engineering from the University of Shanghai for Science and Technology, China, in 2001, and an M.S. and Ph.D. in information science and technology from the University of Tokyo in 2004 and 2007. He joined NTT Communication Science Laboratories in 2013. His research interests include image processing, image retrieval, and pattern recognition.

Masaaki Nishino

Distinguished Researcher, Linguistic Intelligence Research Group, Innovative Communication Laboratory, NTT Communication Science Laboratories.

He received a B.E., M.E., and Ph.D. in informatics from Kyoto University in 2006, 2008, and 2014. He joined NTT in 2008. His current research interests include data structures, natural language processing, and combinatorial optimization.

Mathieu Blondel

Distinguished Research Scientist, Ueda Research Group, NTT Communication Science Laboratories.

He received an engineering diploma from Telecom Lille, France, in 2008 and a Ph.D. in engineering from Kobe University, Hyogo, in 2013. He joined NTT Communication Science Laboratories in 2013. His current research interests include machine learning, mathematical optimization, the design of efficient machine learning software, and the application of these areas to real-world applications.

Takemi Mochida

Senior Research Scientist, Kashino Diverse Brain Research Laboratory, NTT Communication Science Laboratories.

He received a B.S and M.S. in engineering from Waseda University, Tokyo, in 1992 and 1994 and a Ph.D. in systems information science from Future University Hakodate in 2011. He joined NTT in 1994. His research interests include sensorimotor mechanisms in skilled human behavior.

External Awards

Information and Systems Society Excellent Paper Award

Winner: Ryo Masumura, NTT Media Intelligence Laboratories; Taichi Asami, NTT DOCOMO, INC.; Takanobu Oba, NTT Media Intelligence Laboratories; Hirokazu Masataki, NTT TechnoCross Corporation; Sumitaka Sakauchi, NTT Media Intelligence Laboratories (now with NTT EAST); Akinori Ito, Tohoku University; Satoshi Takahashi, NTT TechnoCross Corporation
Date: June 25, 2019
Organization: The Institute of Electronics, Information and Communication Engineers (IEICE)

For their three consecutive papers: "Investigation of Combining Various Major Language Model Technologies including Data Expansion and Adaptation," "N-gram Approximation of Latent Words Language Models for Domain Robust Automatic Speech Recognition," and "Domain Adaptation Based on Mixture of Latent Words Language Models for Automatic Speech Recognition."
Published as: R. Masumura, T. Asami, T. Oba, H. Masataki, S. Sakauchi, and A. Ito, "Investigation of Combining Various Major Language Model Technologies including Data Expansion and Adaptation," IEICE Trans. Inf. & Syst., Vol. E99.D, No. 10, pp. 2452–2461, 2016.
R. Masumura, T. Asami, T. Oba, H. Masataki, S. Sakauchi, and S. Takahashi, "N-gram Approximation of Latent Words Language Models for Domain Robust Automatic Speech Recognition," IEICE Trans. Inf. & Syst., Vol. E99.D, No. 10, pp. 2462–2470, 2016.
R. Masumura, T. Asami, T. Oba, H. Masataki, S. Sakauchi, and A. Ito, "Domain Adaptation Based on Mixture of Latent Words Language Models for Automatic Speech Recognition," IEICE Trans. Inf. & Syst., Vol. E101.D, No. 6, pp. 1581–1590, 2018.

IEVC2019 Best Paper Award

Winner: Kazuhiko Murasaki, NTT Media Intelligence Laboratories; Chihiro Kazato, NTT Access Network Service Systems Laboratories (now with NTT EAST); Shingo Ando, NTT Media Intelligence Laboratories; and Atsushi Sagata, NTT Media Intelligence Laboratories (now with NTT Service Evolution Laboratories)
Date: August 24, 2019
Organization: The Institute of Image Electronics Engineers of Japan (IIEEJ)

For "N-AUC: Maximization of the Narrow Area Under the ROC Curve for Recall-oriented Abnormality Detection."
Published as: K. Murasaki, C. Kazato, S. Ando, and A. Sagata, "N-AUC: Maximization of the Narrow Area Under the ROC Curve for Recall-oriented Abnormality Detection," The 6th IIEEJ International Conference on Image Electronics and Visual Computing (IEVC2019), Kuta Bali, Indonesia, Aug. 2019.

Papers Published in Technical Journals and Conference Proceedings

3-bit Digitized-RoF Retransmission of 12ch 16APSK Broadcast Signals with Improved Nonlinear Compression

R. Shiina, T. Fujiwara, and S. Ikeda

Proc. of the 44th European Conference on Optical Communication (ECOC 2018), pp. 114–116, Rome, Italy, September 2018.

We achieved 3-bit retransmission of 12ch multiplexed 16APSK (amplitude phase shift keying) broadcast satellite signals having extremely wideband, 38.36 MHz/ch, for 4K/8K-UHD (ultrahigh definition) using a DRoF (digitized radio-over-fiber)-based video distribution system with a new compression technique that offers a 40% cut in the transmission rate.

Privacy-preserving Network BMI Decoding of Covert Spatial Attention

T. Nakachi, H. Ishihara, and H. Kiya

Proc. of the 12th International Conference on Signal Processing and Communication Systems (ICSPCS 2018), Cairns, Australia, December 2018.

The brain-machine interface (BMI) has attracted much attention in the fields of biomedical engineering and ICT (information and communication technology) human communications. Of particular interest, neural decoding methods have rapidly developed over the last decade in neuroscience, allowing us to estimate the contents of human perception and subjective mental states by capturing brain activity patterns. However, the development of neural decoding will generate significant concern about privacy violation. In this manuscript, we propose a secure network BMI decoding method based on sparse coding for a covert spatial attention task. It is shown that secure sparse coding enables us to not only protect observed EEG (electroencephalography) signals, but also achieve the same estimation performance as that offered by sparse coding with unprotected observed signals.

Participating-domain Segmentation Based Delay-sensitive Distributed Server Selection Scheme

A. Kawabata, B. C. Chatterjee, and E. Oki

IEEE Access, Vol. 7, pp. 20689–20697, February 2019.

This paper proposes a participating-domain segmentation based server selection scheme in a delay-sensitive distributed communication approach to reducing the computational time for solving the server selection problem. The proposed scheme divides the users' participation domain into a number of regions. The delay between a region and a server is a function of locations of the region and the server. The length between the region and the server is considered based on conservative approximation. The location of the region is determined regardless of the number of users and their participation location. The proposed scheme includes two phases. The first phase uses the server finding process and determines the number of users that are accommodated from each region by each server, instead of actual server selection, to reduce the computational complexity. The second phase uses the delay improvement process and determines the overall delay and the selected server for each user. We formulate an integer linear programming problem for the server selection in the proposed scheme and evaluate the performance in terms of computation time and delay. The numerical results indicate that the computational time using the proposed scheme is smaller than that of the conventional scheme, and the effectiveness of the proposed scheme improves as the number of users increases.

VLC/RF Channel Switching Process Adapting to User Mobility in Coexistence Architecture

R. Shiina, K. Hara, T. Taniguchi, T. Nakahira, T. Murakami, and S. Ikeda

Proc. of the 24th OptoElectronics and Communications Conference (OECC 2019), WA3-3, Fukuoka, Japan, July 2019.

We propose a seamless channel switching process that considers user mobility in the WiSMA (strategy management architecture for wireless resource optimization)-based VLC/RF (visible light communication/radio frequency) coexistence architecture. Experiments and theoretical evaluations confirm the feasibility of the proposal over an expected range of user velocities.

Observation of Intracellular Protein Localization Area in a Single Neuron Using Gold Nanoparticles with a Scanning Electron Microscope

T. Goto, N. Kasai, R. Filip, K. Sumitomo, and H. Nakashima

Micron, Vol. 126, 102740, August 2019.

The localization areas of intracellular proteins in rat cortical neurons were visualized using a scanning electron microscope (SEM) coupled with a focused ion beam (FIB) system. To obtain a clear contrast in the SEM images, gold nanoparticles (GNPs) were bound to specific intracellular proteins by antigen-antibody reactions. By obtaining a cross section of the desired location of the neurons by FIB milling under the SEM imaging condition, it was possible to observe the proteins inside the cells as clear bright spots. When a neuron was stained with antitau and anti-histone H1 antibodies, the bright spots were localized in the cross section of the axon and the nucleus, respectively. It was confirmed that targeted proteins in a single neuron on a substrate could be successfully identified. The development of FIB/SEM observation with immunological GNP staining will offer important information for the stable growth of neurons on various substrate structures, since the elongation and turning of axons on the substrates are activated by the redistribution of intracellular proteins.

Quantum Key Distribution with Simply Characterized Light Sources

A. Mizutani, T. Sasaki, Y. Takeuchi, K. Tamaki, and M. Koashi

The 9th International Conference on Quantum Cryptography (QCrypt 2019), Montreal, Canada, August 2019.

In general, in order to show that a QKD (quantum key distribution) protocol is secure, some assumptions on light sources and/or measurement apparatuses are necessary. In this work, we have relaxed the assumptions.

Resource-efficient Verification of Quantum Computing Using Serfling's Bound

Y. Takeuchi, A. Mantri, T. Morimae, A. Mizutani, and J. F. Fitzsimons

QCrypt 2019, Montreal, Canada, August 2019.

In order to verify the correctness of measurement-based quantum computation, which is one of the universal quantum computing models, it is important to check whether the target graph state is faithfully prepared. In this work, we have proposed a resource-efficient verification protocol for graph states. Our protocol can be applied to blind quantum computation.

Impossibility of Blind Quantum Sampling for Classical Client

T. Morimae, H. Nishimura, Y. Takeuchi, and S. Tani

Quantum Information and Computation, Vol. 19, No. 9&10, pp. 793–806, August 2019.

Blind quantum computing enables a client, who can only generate or measure single-qubit states, to delegate quantum computing to a remote quantum server in such a way that the input, output, and program are hidden from the server. It is an open problem whether a completely classical client can delegate quantum computing blindly (in the information theoretic sense). In this paper, we show that if a completely classical client can blindly delegate sampling of subuniversal models, such as the DQC1 (deterministic quantum computation with one quantum bit) model and the IQP (instantaneous quantum polynomial time) model, then the polynomial-time hierarchy collapses to the third level. Our delegation protocol is the one where the client first sends a polynomial-length bit string to the server and then the server returns a single bit to the client. Generalizing the no-go result to more general setups is an open problem.

Quantum Computational Universality of Hypergraph States with Pauli-X and Z Basis Measurements

Y. Takeuchi, T. Morimae, and M. Hayashi

Scientific Reports, Vol. 9, 13585, September 2019.

Measurement-based quantum computing is one of the most promising quantum computing models. Although various universal resource states have been proposed so far, it was open whether only two Pauli bases are enough for both universal measurement-based quantum computing and its verification. In this paper, we construct a universal hypergraph state that only requires X and Z-basis measurements for universal measurement-based quantum computing. We also show that universal measurement-based quantum computing on our hypergraph state can be verified in polynomial time using only X and Z-basis measurements. Furthermore, in order to demonstrate an advantage of our hypergraph state, we construct a verifiable blind quantum computing protocol that requires only X and Z-basis measurements for the client.

Verifying Commuting Quantum Computations via Fidelity Estimation of Weighted Graph States

M. Hayashi and Y. Takeuchi

New Journal of Physics, Vol. 21, No. 9, 093060, September 2019.

The instantaneous quantum polynomial time (IQP) model is a promising model to demonstrate a quantum computational advantage over classical computers. If the IQP model can be efficiently simulated by a classical computer, an unlikely consequence in computer science can be obtained (under some unproven conjectures). In order to experimentally demonstrate the advantage using medium or large-scale IQP circuits, it is inevitable to efficiently verify whether the constructed IQP circuits faithfully work. There exist two types of IQP models, each of which is the sampling on hypergraph states or weighted graph states. For the first-type IQP model, polynomial-time verification protocols have already been proposed. In this paper, we propose verification protocols for the second-type IQP model. To this end, we propose polynomial-time fidelity estimation protocols of weighted graph states for each of the following four situations where a verifier can (i) choose any measurement basis and perform adaptive measurements, (ii) only choose restricted measurement bases and perform adaptive measurements, (iii) choose any measurement basis and only perform non-adaptive measurements, and (iv) only choose restricted measurement bases and only perform non-adaptive measurements. In all of our verification protocols, the verifier's quantum operations are only single-qubit measurements. Since we assume no independent and identically distributed property on quantum states, our protocols work in any situation.

Using Seq2Seq Model to Detect Infection Focusing on Behavioral Features of Processes

S. Tobiyama, Y. Yamaguchi, H. Hasegawa, H. Shimada, M. Akiyama, and T. Yagi

Journal of Information Processing, Vol. 27, pp. 545–554, September 2019.

Sophisticated cyber-attacks intended to earn money or steal confidential information, such as targeted attacks, have become a serious problem. Such attacks often use specially crafted malware, which utilizes the art of hiding such as by process injection. Thus, preventing intrusion using conventional countermeasures is difficult, so a countermeasure needs to be developed that prevents attackers from reaching their ultimate goal. Therefore, we propose a method for estimating process maliciousness by focusing on process behavior. In our proposal, we first use one Seq2Seq model to extract a feature vector sequence from a process behavior log. Then, we use another Seq2Seq model to estimate the process maliciousness score by classifying the obtained feature vectors. By applying Seq2Seq models stepwise, our proposal can compress behavioral logs and extract abstracted behavioral features. We present an experimental evaluation using logs when actual malware is executed. The obtained results show that malicious processes are classified with a highest Areas Under the Curve (AUC) of 0.979 and 80% TPR even when the FPR is 1%. Furthermore, the results of an experiment using the logs when simulated attacks are executed show our proposal can detect unknown malicious processes that do not appear in training data.

Beyond Fourier Analysis: Recipes for Nonlinear Oscillator Expansion

F. Ishiyama

Proc. of the SICE Annual Conference 2019, pp. 1643–1648, Hiroshima, Japan, September 2019.

We are developing a method for time series analysis, which is a superset of Fourier analysis. We piecewisely quantize given time series into five dimensional particles, and we obtain a series expansion with general complex functions. We outline our method and present various recipes to show what we can do with our method. The recipes are less-than-a-cycle analysis, instability detection, trend-mode analysis, nonlinear analysis, and signal-noise separation.

DRoF-based Optical Video Re-transmission of Commercial RF Broadcast Signals

R. Shiina, T. Fujiwara, T. Taniguchi, and S. Ikeda

Journal of Optical Communications and Networking, Vol. 11, No. 11, pp. 559–567, November 2019.

We propose a novel digitized radio-over-fiber (DRoF)-based optical video re-transmission system. This system digitally transmits radio frequency (RF) broadcast signals via the communication network (NW) while maintaining the existing RF interface at both ends of the NW by using DRoF technology. Directly digitizing the RF signal makes the required transmission rate of the optical NW impractically large. In order to resolve this issue, we also propose an improved nonlinear quantization (INL) method that combines amplitude clipping and nonlinear quantization. This paper experimentally evaluates the feasibility of optical re-transmission by using a commercial 9-channel multiplexed digital terrestrial television broadcasting signal. The results of the experiment and our theoretical analysis show that it is necessary to re-transmit with 7-bit quantization at the transmitter in order to achieve the required signal quality. However, we show that re-transmission can be realized while satisfying the required quality with just 5 bits by using our proposed INL method. The resulting 2-bit reduction afforded by the INL proposal reduces the transmission rate by 28.6%.