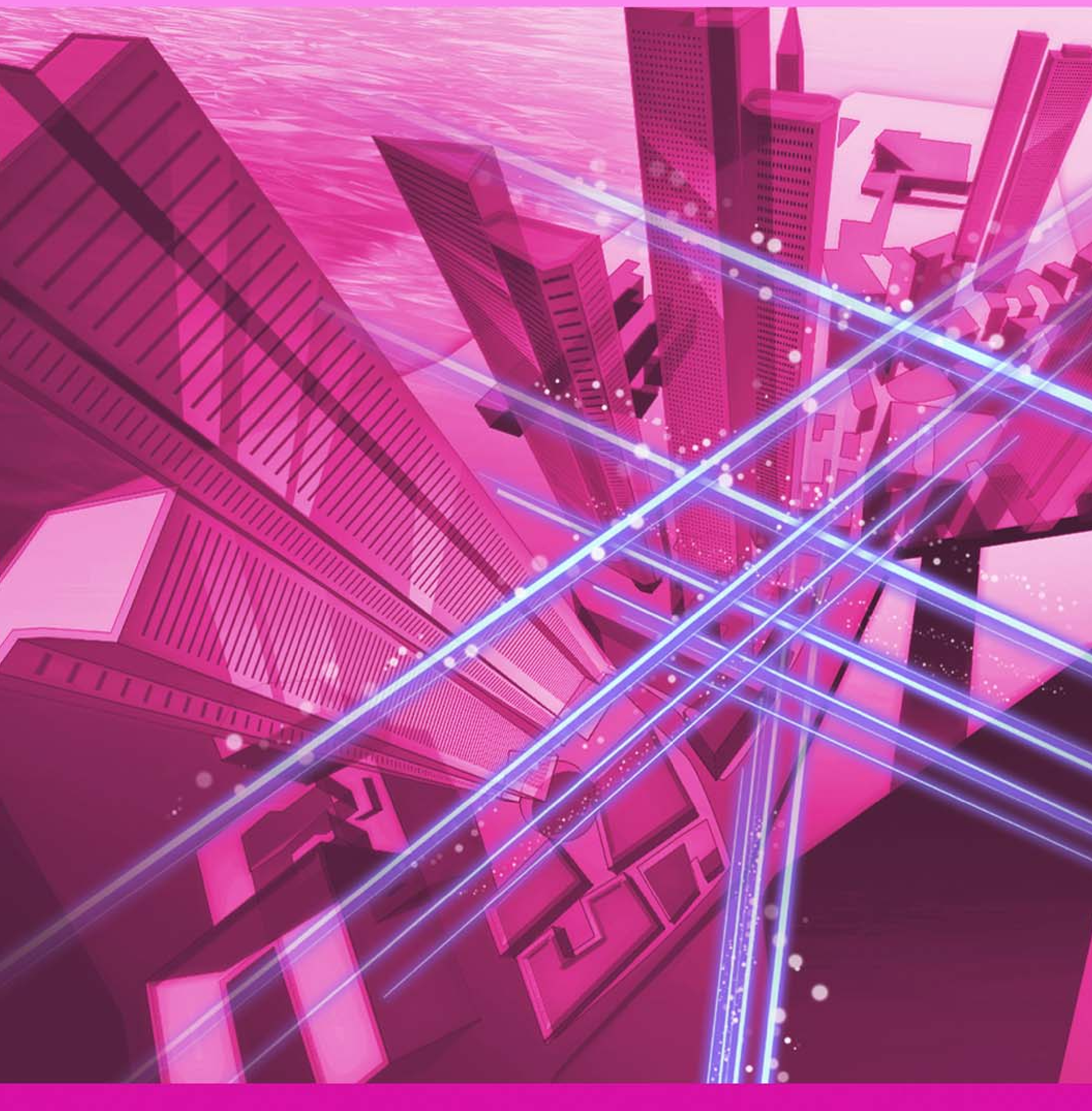
11
2018

NTT Technical Review



November 2018 Vol. 16 No. 11

NTT Technical Review

November 2018 Vol. 16 No. 11

View from the Top

Front-line Researchers

Feature Articles: Further Exploring Communication Science

Regular Articles

Global Standardization Activities

Information

Short Reports

External Awards/Papers Published in Technical Journals and Conference Proceedings

Toward Co-creation of New Value through a Receptive Attitude to Outside Ideas—NTT DOCOMO's "Declaration Beyond" Medium-term Strategy Embodied in Six Declarations

Kazuhiro Yoshizawa
President and Chief Executive Officer,
NTT DOCOMO

Overview

With its sights firmly on the 5G (fifth-generation mobile communications networks) era, NTT DOCOMO has achieved a gigabit-capable mobile network by increasing its effective download speed by 1.5 times over the previous year. Breaking away from its traditional image of a mobile and smartphone company, NTT DOCOMO has strived to improve its long-term corporate value and to create a safe and secure, comfortable, and richer society, for which it has gained a high reputation, including earning a No. 1 ranking in total score in the 2018 Toyo Keizai CSR (Corporate Social Responsibility) Ranking. We asked Kazuhiro Yoshizawa, President and Chief Executive Officer of NTT DOCOMO, about his outlook and aspirations based on the Medium-Term Strategy 2020 known as "Declaration beyond."

Keywords: 5G, artificial intelligence, co-creation

Medium-Term Strategy 2020 "Declaration beyond" aiming for sustainable growth

—Mr. Yoshizawa, NTT DOCOMO is said to be undergoing a business transformation. How are things going?

Business operations are on the right track. Competition in the mobile market is becoming increasingly severe, so we are working day in and day out to gain the favor of our customers. For example, customer satisfaction is high with the "docomo with" billing plan, and our subscriber base is increasing steadily. However, we cannot become complacent with these results. Our Medium-Term Strategy 2020 "Declaration beyond" announced in April 2017 is aimed at establishing a solid business platform while we earnestly address the ever-changing needs of society and return more value to customers (**Fig. 1**). We will continue to challenge new frontiers with an eye

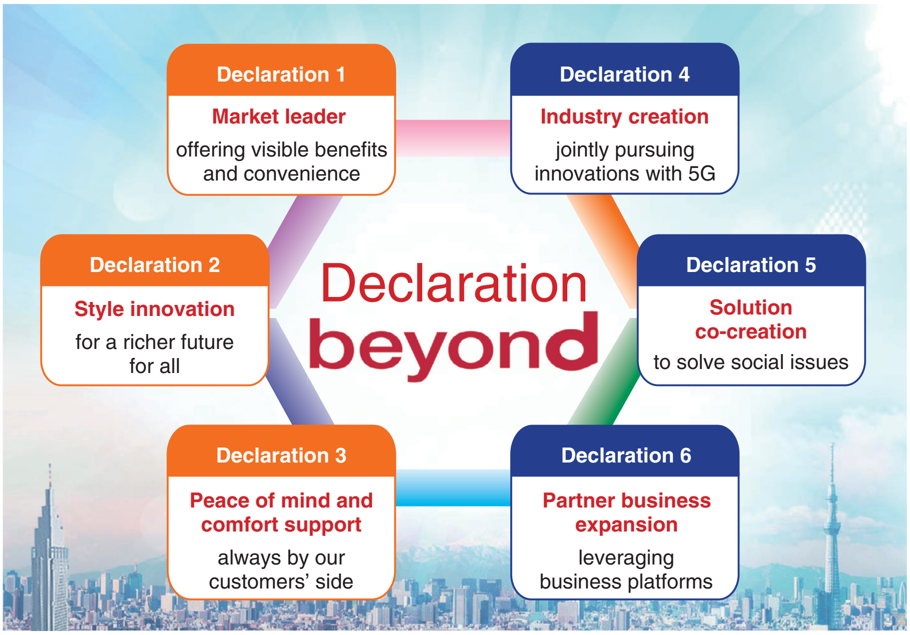

Declaration 1
Market leader
offering visible benefits and convenience
Declaration 4
Industry creation
jointly pursuing innovations with 5G
Declaration 2
Style innovation
for a richer future for all
Declaration beyond
Declaration 5
Solution co-creation
to solve social issues
Declaration 3
Peace of mind and comfort support
always by our customers' side
Declaration 6
Partner business expansion
leveraging business platforms

Fig. 1.   Medium-Term Strategy 2020 "Declaration beyond."

to 2020 and beyond. We aim to create a richer future with fifth-generation mobile communications systems (5G) by providing surprising and exciting services exceeding our customers' expectations and by co-creating new value with our business partners. At present, we are steadily moving into the execution phase.

As its services expand beyond communications, NTT DOCOMO is redefining the meaning of "customer." Up to now, we have focused all of our energy on increasing the number of our mobile subscribers, but with the penetration rate of mobile phones now exceeding 100% in Japan, we are making a shift from a customer foundation based on "subscribers" to one based on "membership." Here, a "member" is someone who uses NTT DOCOMO services regardless of whether that person is an NTT DOCOMO mobile subscriber. In other words, we seek to maximize the value that we can offer, even to customers who may be mobile subscribers of other companies but who use NTT DOCOMO services such as "d POINT" (loyalty points program), "d CARD" (credit card), and "dmarket" (content delivery portal).

Specifically, as stated in "Declaration 1: Market leader" of "Declaration beyond," we aim to lead the market in providing benefits, convenience, and surprise, by integrating and enhancing services, billing plans, and loyalty points programs. Similar ecosystems already exist in other companies' services, but we aim to expand the number of "d POINT" partners to more than 300 companies by fiscal year 2020 and to develop "d POINT" as one of Japan's largest loyalty points programs.

We are also busy with a variety of initiatives in relation to artificial intelligence (AI). For example, in May 2018 we launched the "my daiz" AI agent service—one of the nine challenges proclaimed in "Declaration 2: Style innovation." On assuming the office of president, I announced my goal to achieve the ultimate personal agent that blends in perfectly with people's lives, and I can say that "my daiz" is truly the precursor of this agent. This service *anticipates* behavior using information obtained from a customer's smartphone operations and assists the customer in choosing actions that make for a smoother day. For example, "my daiz" can anticipate and provide information that the customer will likely need such as information on a delay in the running of a train that the customer always uses. Looking forward, we want to be able to anticipate customer behavior to make personalized and optimized proposals. To this end, we continue to refine AI, which takes on the role of a "mind," to provide surprise and excitement beyond the customer's expectations.

In addition, we plan to use 5G with augmented reality, virtual reality, AI, and the Internet of Things (IoT) to achieve "experience innovation," "life style innovation," and "work style innovation." For example, by partnering with content providers and broadcasters, we plan to place multiple cameras at a venue and deliver a live program such as a sports event and concert by 4K/8K video using NTT's "Kirari!" immersive telepresence technology and other video processing technologies. In this way, we hope to provide customers in remote locations with the sense of reality as if they were present at the venue. High-presence, high-resolution, wide angle, and multiple simultaneous transfers constitute an area that makes full use of 5G. We plan to exploit these features to provide customers with truly exciting experiences.

—It appears that many initiatives are being launched that will transform the world.

That's true, and 5G in particular will be a driving force. With the aim of offering the 5G experience sooner, NTT DOCOMO will deliver a "pre-service" at the stadiums that will be used in the Rugby World Cup to be held in September 2019 in Japan. Building on these achievements, we plan to commence commercial services of 5G in the spring of 2020.

As stated in "Declaration 4: Industry creation," NTT DOCOMO will contribute to the further development of society and industry by leveraging the 5G features of high speed and large capacity, low latency, and massive device connectivity to expand the possibilities for business in collaborations with partner companies. I believe that 5G will be a major pillar of our digital transformation, which will lead to (1) improved business operations, that is, improved productivity, (2) radically enhanced user-interface/user-experience designs, and (3) the creation of totally new, revolutionary services. We aim to achieve these transformations through 5G.

I remember well, when 3G and LTE (Long-Term Evolution) services were being launched, it was often said, "Are these high-speed communications really necessary?" But as it turned out, the sudden spread of social media, streaming video, and smartphones have today made those high-speed communications an absolute necessity. The approach taken here was to build the network first and to catch up with services later. In the case of 5G, however, our goal is to provide 5G-era services from day one, that is, to launch both the 5G network and 5G services at the same time. For this reason, *co-creation* with a wide range of partners is essential.

To facilitate this co-creation of new 5G-era services with diverse partners, we provide the DOCOMO 5G Open Partner Program at no charge. More than 1800 companies and organizations have already come to participate in this program, which features workshops that include lectures and lively discussions among partners on creating new 5G services. We also provide the DOCOMO 5G Open Lab technology-testing environment in which 5G experimental base-station equipment can be used without charge. One site was set up in Tokyo in April and another in Osaka in September, and one is scheduled to open in Okinawa in December.

Furthermore, with the aim of deploying specific 5G services, we are conducting a wide variety of trials—more than 70—with partners at 5G Trial Sites that have been set up in the Tokyo Waterfront City district and Tokyo Skytree® neighborhood. We expect that promoting co-creation with diverse partners at an early stage in this way will support the rapid creation of new services in the 5G era.

Let me give you two examples of these trials. The first trial involves the remote operation of construction machinery, and we are conducting this jointly with Komatsu Ltd. In this trial, we mount 5G terminals on construction machinery such as hydraulic excavators and bulldozers and have a remote operator control the machinery while watching video of it via 5G. Through this trial, we are studying the feasibility of performing accurate and efficient on-site work and work management from a remotely located office by checking on-site conditions in real time through NTT DOCOMO 5G connections.

The second trial entails remote medical services that we conducted in collaboration with the Wakayama Prefecture and Wakayama Medical University. In this trial, we first transmitted images of ultrasound-image diagnostic apparatus (echo) and MRI (magnetic resonance imaging) systems, or 4K images of the affected area of the patient's body to the university's Regional Medical Support Center. We then asked a medical specialist at the university to make a diagnosis via 4K teleconference that could support medical treatment at the rural clinic. In this way, we are trying to resolve disparities in the provision of healthcare services, one of the major social issues in the country.

Value creation arises from co-creation with partners

—Changes are also taking place at construction sites, and the workers involved could be changing.

That's right. The on-site working environment in civil engineering and construction has long been described as "3K," referring to the three Japanese words *kitsui*, *kitanai*, and *kiken* (hard, dirty, and risky). I've heard that the driver's seat in some models of bulldozers is not equipped with air conditioning. For on-site work during hot and humid summers and at accident scenes or hazardous zones that people should not enter, it will become possible to use a camera in place of the human eye and transmit high-resolution video by 5G to a remote monitor. This capability will enable remote operation of machinery in real time as if the operator were at the actual site.

We can also consider that the skillful use of big data can be applied to solve social problems. For example, we are implementing the LANDLOG open IoT platform to improve productivity in the construction industry. LANDLOG uses IoT technology to collect data at construction sites on dump trucks that carry earth and on worker movements and other details. It will then analyze that data using AI so that on-site operations can be optimized and work efficiency improved. Our goal here is to achieve process innovation at construction sites.

Additionally, as stated in "Declaration 5: Solution co-creation," we are working to solve social problems and revitalize regional economies in primary industries, education, mobility, and other fields. This is expected to stimulate growth in Japan and create a richer society.

I can give you some examples of applications in primary industries such as agriculture and the fishing industry. In agriculture, we are working to reduce the manual labor associated with the day and night monitoring of rice paddies by applying a system that measures the temperature and level of water in a paddy and visualizes such data on the farmer's smartphone screen. In the fishing industry, we are working on the ICT (information and communication technology) Buoy solution system that measures ocean data such as water temperature and salt concentration and visualizes those data on smartphones. This system makes it possible to reduce the risks of relying only on human experience and know-how and to complement such know-how with data, thereby improving quality and efficiency in the cultivation of seaweed and oysters.

Meanwhile, while it was only a short time ago that services such as self-driving cars were thought to be something only seen in movies, we now know that such unbelievable services may be achievable in the real world through technical innovation. It would not be possible, however, to switch all cars from being ordinary ones to self-driving ones all at once. Whatever the matter may be, it's simply not feasible to make a complete change at the same point in time. Likewise, in our initiatives for construction sites, agriculture, and the fishing industry, it is essential that we apply a plan-do-check-act (PDCA) cycle any number of times while reflecting the best idea at each time. Innovation never ends. I believe that repeating the PDCA cycle is one way of fostering innovation.

—"Repetition" appears to be a straightforward approach.

Yes it is, and management as well undergoes cycles of improvement. Here, one must take a hard look at the results to date, but there is no guarantee that revenues will increase by using the same method as before. New things must always be incorporated, and you yourself must change.

At the same time, even if the economy and social conditions should be undergoing substantial change, the reason for NTT DOCOMO's existence does not change, which I believe to be the continuous provision of new value to customers and society. At present, however, what we can do by ourselves is limited. Accepting ideas and technologies possessed by a variety of partner companies and organizations is what I call *receptivity*, and combining the strengths of all parties involved to create totally new value is what I recognize to be open innovation, or in other words, co-creation.

We call this initiative +d (plus d), which involves an effort to achieve positive results for both a partner and NTT DOCOMO. The +d initiative is already making progress in diverse fields. NTT DOCOMO has extensive business assets such as its mobile network, membership base, and secure settlement system. Combining these assets with our partners' assets should generate a "chemical reaction" from which we can create new business useful to our partners and co-create new social value.

In today's world, value creation will be limited if we rely only on the abilities of individuals or single companies. We must therefore collaborate with partners and bring out ideas. As part of this +d initiative, we are trying to change the procedure for product and service development used in the past, where salespeople meet with customers and listen to what they have to say and pass that information on to those in charge of technology for reflection in development. When salespeople talk to customers nowadays, they frequently need more technical knowledge than they did before. Therefore, research and development (R&D) staff should also participate in the discussions with customers from the very start so that we can directly connect customer needs to development or make proposals that anticipate the customer's needs or wishes.

The program we have developed called "Top Gun" can reduce the time and labor involved in building a bridge from corporate sales and marketing to R&D, thereby enabling us to repeat the PDCA cycle more rapidly. In this way, customers, corporate sales and marketing, and R&D become a three-party collaboration that can solve problems quickly and create new value.

The first project of the Top Gun program was an experimental trial of a child monitoring solution for elementary school students in Kobe City, Hyogo Prefecture. This solution involves attaching a Bluetooth Low Energy tag to students' backpacks and checking whether they had passed a certain location by using sensors installed throughout the city. The R&D team also proposed a mechanism that uses the Bluetooth function of smartphones by asking people in various occupations such as taxi drivers, insurance salespersons, and security guards to install a special application in their smartphones so that position information could be collected whenever such a person and a student passed each other.

This was the system eventually deployed in Kobe City. This made it possible to acquire a more detailed history of movement than with any previous system. We also applied this mechanism to systems to determine the location of baggage carts in airports, baby carriages, and other items. In fact, we launched a service called "Location Net$^{\text{TM}}$" using this mechanism in October 2017. Therefore, the Top Gun program has already been implemented in more than 10 projects, including the ones mentioned here.

Having a receptive attitude unconstrained by one's standing or position

—Mr. Yoshizawa, can you leave us with a message for all NTT DOCOMO employees?

I would be happy to. I am very conscious of receptivity, which I think is a very important attitude to have. I always say, "I would like you to always keep in mind the perspective of others, regardless of whether the person you're talking to is your superior or subordinate." In other words, let's listen carefully to each other, on an equal footing in a humble manner. It's easy to give instructions to a subordinate from the position of a manager, but doing so may narrow the range of that person's work or thoughts or suppress initiative. Listening closely to what someone has to say can help clarify the proposer's logic or ideas. In such cases, I don't hesitate to adopt the other person's idea if it's better than mine. I also don't want to miss out on hearing new viewpoints or ideas by cutting off a conversation saying, "I already know all about this." For this reason, I try to hear proposals directly whenever possible. Of course, I may encourage someone to make it short because of time limitations, but basically speaking, I place importance on listening closely. I associate this attitude with our +d initiative. One's standing or position within the company does not matter. This is the way we should approach our work throughout NTT DOCOMO. Through such an attitude, I am convinced that we can share values with our customers and gain their understanding of NTT DOCOMO's initiatives and efforts.

Nevertheless, even with these efforts, there are few times when business proceeds absolutely smoothly. What happens more often than not is that one's work does not go to plan and proceeds in an undesirable direction. At such times, it is important that you do not avoid something you don't like and that you work to overcome the difficulty. Business is like walking along a narrow ridge with a steep cliff on both sides. If you suddenly come to a stop or even if you fall off, you have to face the situation calmly and think about your next step. Additionally, even if you have been walking well so far, you can still tumble down the cliff if you start to walk aimlessly. Arrogance is your enemy here. This is what I think business is all about. Keep this in mind and approach your work as something that is enjoyable. Pessimism has no place here. Let's do our work filled with optimism!

Interviewee profile

■ Career highlights

Kazuhiro Yoshizawa joined Nippon Telegraph and Telephone Public Corporation (now NTT) in 1979. He has been involved in the mobile phone business since the dawn of the industry. He became NTT DOCOMO Senior Vice President and Managing Director, Corporate Sales and Marketing Department II, in 2007, Senior Vice President and Managing Director of the Human Resources Management Department in 2011, and Executive Vice President, Managing Director of the Corporate Strategy & Planning Department in 2012. He became Senior Executive Vice President in June 2014. He assumed his current position in June 2016. Concurrently, he retains his position as a member of the Board of Directors, which he assumed in 2011.

Pursuing a Dialogue System for Making Society More Harmonious

Ryuichiro Higashinaka
Senior Distinguished Researcher, NTT Media Intelligence Laboratories

Overview

The spread of smartphones and artificial intelligence (AI) systems that speak means that the opportunities for general users to come into contact with AI are increasing. Users are now expecting improved ease of use and more new technologies from research and development. Under these circumstances, NTT has continuously gained attention—both domestically and globally—in announcing pioneering technologies in the fields of question answering functions and language processing. We visited Ryuichiro Higashinaka, Senior Distinguished Researcher at NTT Media Intelligence Laboratories, and asked him about the progress and outlook of his research, his attitude as a researcher, and how the development of AI and a future in which robots and humans can talk smoothly will bring about changes in our society.

Keywords: artificial intelligence, natural language processing, question answering technology

Research on dialogue systems takes steady work

—Tell us about the research that you are currently working on and the initiatives undertaken so far.

I joined NTT in 2001, and since then, I have been researching language processing, artificial intelligence (AI), and dialogue systems. A dialogue system is a technology for interacting with a computer. It is probably easier to understand if imagined as an animated robot character like Doraemon or Astro Boy who can talk smoothly with humans. The idea of “smoothness,” that is, “naturalness,” is the point of my research, and for that reason it is necessary to unravel what kind of elements a person’s conversation is made up of.

As shown by the fact that 60% of human conversation constitutes chat (casual conversation), chat is very important. A relationship between people cannot be established simply by having a conversation about work. Chat serves as cushioning material that makes it possible to know the other person’s personality, thereby encouraging cooperative work. Let’s consider that one person in a conversation is replaced with a dialogue system (i.e., a computer): if the computer does not understand what kind of person the person is, and if the person does not understand what kind of thing the computer is like, the conversation will not go smoothly. In my past research, though, I focused on tasks rather than chat, so I had the idea that a dialogue between a person and a computer should be brief. However, I came to realize the importance of communication and building relationships in achieving natural conversation, so I am currently focusing on chat. In regard to the study of an actual dialogue system, I’m repeating the straightforward work of

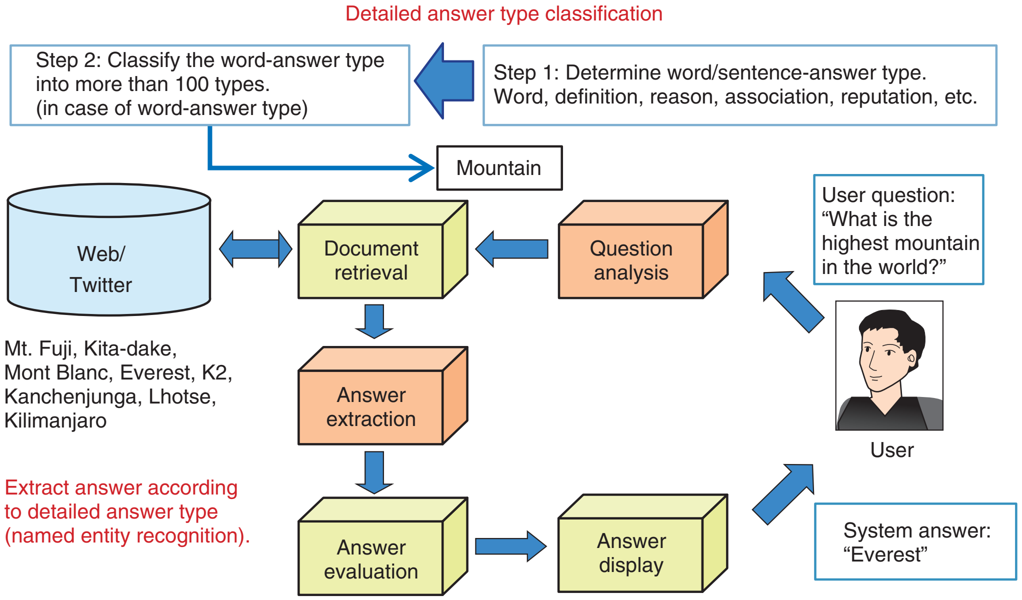

Detailed answer type classification
Step 2: Classify the word-answer type into more than 100 types. (in case of word-answer type)
Step 1: Determine word/sentence-answer type. Word, definition, reason, association, reputation, etc.
Mountain
Web/ Twitter
Document retrieval
Question analysis
User question: "What is the highest mountain in the world?"
Mt. Fuji, Kita-dake, Mont Blanc, Everest, K2, Kanchenjunga, Lhotse, Kilimanjaro
Answer extraction
User
Extract answer according to detailed answer type (named entity recognition).
Answer evaluation
Answer display
System answer: "Everest"

Fig. 1. Question answering logic used by "Shabette Concier."

observing conversations and building the system, getting people to talk with the system and checking unnatural parts of speech, and feeding the results back to the system. In that sense, research on dialogue systems is step-by-step.

—Can you introduce the research that you have worked on specifically?

There is NTT DOCOMO's voice-agent service called "Shabette Concier" (talking concierge), which was launched in March 2012 in Japan. I was in charge of the logic used by the question answering system supporting this service (**Fig. 1**). The question answering system was adopted as a centerpiece function when Shabette Concier was upgraded in June 2012. It recognizes and analyzes a user's words and searches for relevant information on the Internet and finds the best answers within a few seconds. It was designed to work smoothly even if many users are connected at the same time. It took about half a year for us to make the system practical, and as the practical application of question answering technology—which had not been developed on a large scale then—it received much attention from the research community. Incidentally, even though I was directly in charge of the final stage of development of the function, the basic research that formed its foundation spanned over 10 years before that stage. I don't think that we would have been successful without that basic research.

As for AI, I was involved in the project called "Can a robot get into the University of Tokyo?" led by the National Institute of Informatics (**Figs. 2** and **3**). This is a project to get AI to solve problems and answer questions in a university entrance exam. I was in charge of the subject of English in collaboration with joint-research institutes. At that time, we applied language processing technology with the goal of strengthening the English skills of *Torobo-kun* (the name of the AI. *Torobo* stands for the University of Tokyo Robot, and *kun* is an honorific title in Japanese normally used for boys). We had Torobo-kun study English in order to take the National Center Test for University Admissions, but I actually think that was more difficult than studying by oneself. In particular, clarifying what was difficult was problematic. That admission test is made up of a variety of questions such as pronunciation problems, filling-the-gap problems, long-sentence reading comprehension, and listening comprehension. By utilizing dictionaries and big data, we were able to reach a deviation score of 50.5; however, we have not yet reached the acceptance standard of the University of Tokyo. All the researchers involved are pursuing the next step. By solving this admission test, I hope to improve language processing technology and develop more advanced natural dialogue systems.

Also, in collaboration with Professor Hiroshi

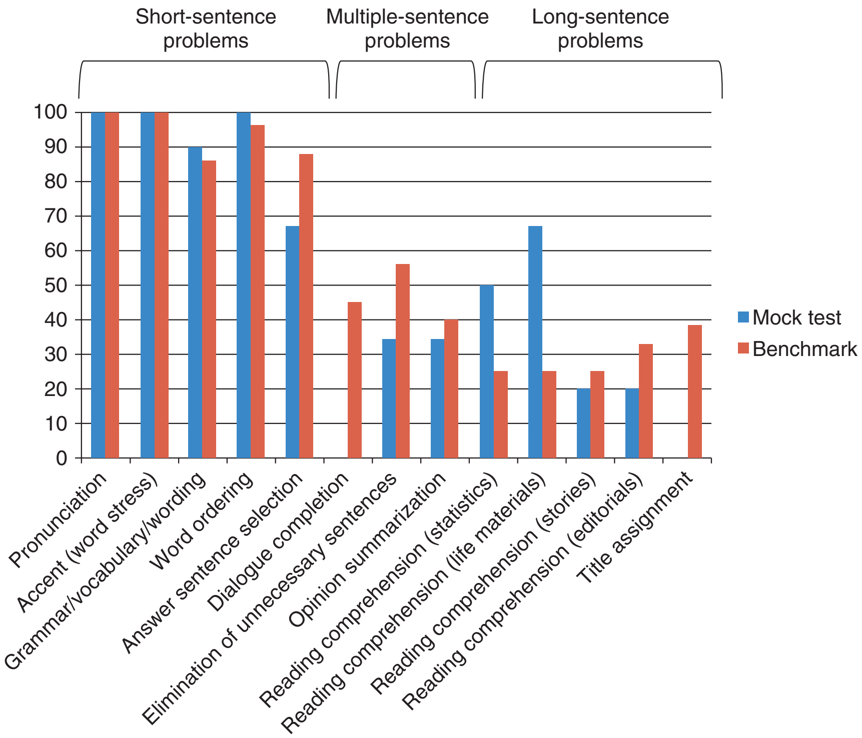


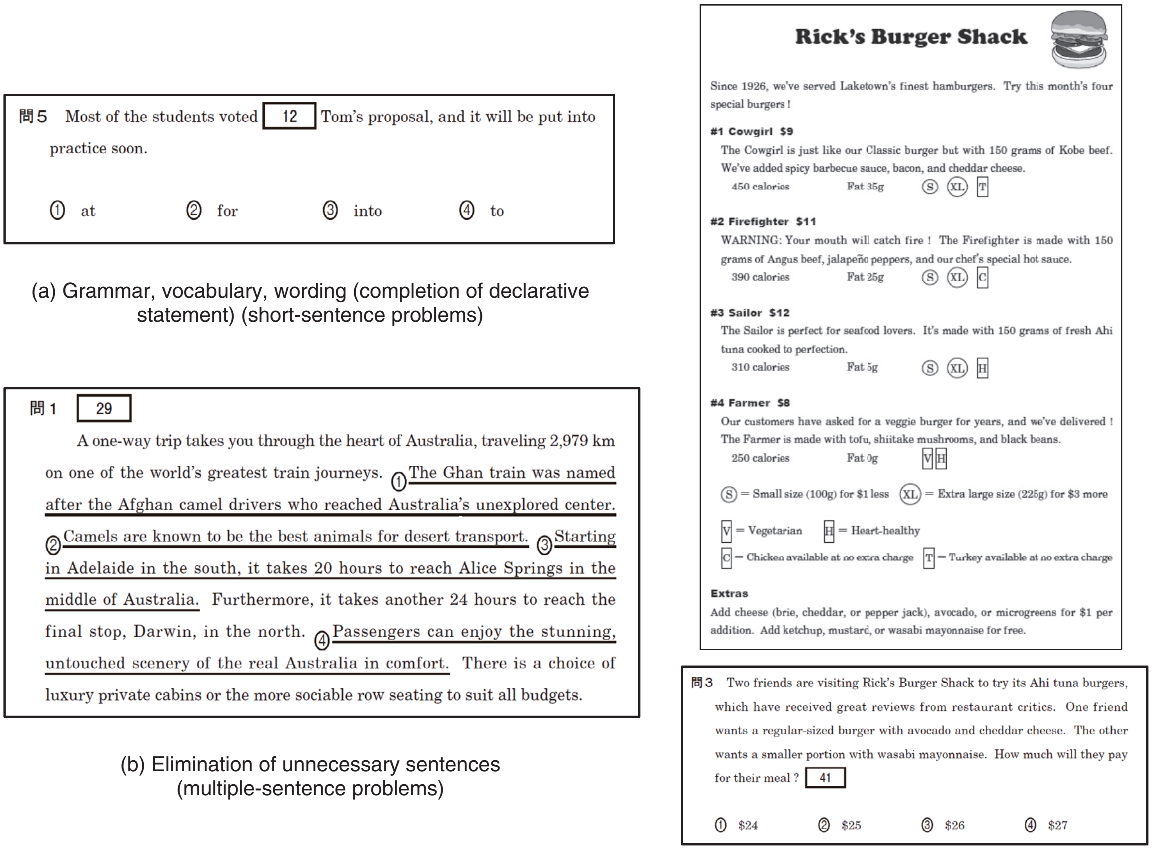
Fig. 2. Scoring rate achieved by Torobu-kun in comparison with benchmark (test data created at laboratory) for various problems in University of Tokyo entrance English test.

Ishiguro of Osaka University, I'm researching an android that can hold a conversation. Professor Ishiguro is at the forefront of research on humanoid robots. He found the point at which human-like robots make us feel uncomfortable and is now pursuing that *something* that makes us humans. Meanwhile, I want to approach human nature by exploring what the essence of dialogue is. In common with these efforts, we have been conducting joint research with Professor Ishiguro such as creating "Matsukoroid," which looks like the famous Japanese television personality Matsuko Deluxe, in 2015.

Moreover, in a collaborative experiment with Dwango Co., Ltd., I have begun work on the development of AI of "Ayase" (a character in a light-novel series called "*My Little Sister Can't Be This Cute*.") to thereby create "Ayase AI" (**Fig. 4**). This project is aimed at developing AI with personality with the user's involvement. A certain amount of data is required to make a computer personality like a person; however, currently there is not enough data in the whole world to express personality, and from the viewpoint of privacy, it is difficult to obtain personal data. Ayase AI started from the idea that if the data are not available, people should make the data. By asking users to become Ayase and talk on the web, we can collect fundamental data and foster the personality of Ayase based on such data. As the amount of data grows, the dialogue system becomes unique. Users can feel as if they're really talking to someone, so I think we are getting closer to finding the essence of dialogue.

—What is the significance of your research?

I think that the significance of my research is to pursue the essence of human beings. We as humans have elements of ourselves that we do not understand. Humans are social creatures that cannot live by themselves, so we have developed communication skills for the purpose of living with others. I believe that if we can clarify that communication on a scientific basis, humans will get closer to understanding each other. I believe that if human mutual understanding progresses, cooperative work will become smoother, we will feel happier, and so on, leading to improved quality of life. I hope to pursue my goal of developing a dialogue system as a shared property that will make



問５　Most of the students voted [12] Tom's proposal, and it will be put into practice soon.

① at　② for　③ into　④ to

(a) Grammar, vocabulary, wording (completion of declarative statement) (short-sentence problems)

問１ [29]

A one-way trip takes you through the heart of Australia, traveling 2,979 km on one of the world's greatest train journeys. ①The Ghan train was named after the Afghan camel drivers who reached Australia's unexplored center. ②Camels are known to be the best animals for desert transport. ③Starting in Adelaide in the south, it takes 20 hours to reach Alice Springs in the middle of Australia. Furthermore, it takes another 24 hours to reach the final stop, Darwin, in the north. ④Passengers can enjoy the stunning, untouched scenery of the real Australia in comfort. There is a choice of luxury private cabins or the more sociable row seating to suit all budgets.

(b) Elimination of unnecessary sentences (multiple-sentence problems)

Rick's Burger Shack

Since 1926, we've served Laketown's finest hamburgers. Try this month's four special burgers!

#1 Cowgirl $9
The Cowgirl is just like our Classic burger but with 150 grams of Kobe beef. We've added spicy barbecue sauce, bacon, and cheddar cheese.
450 calories　Fat 35g　Ⓢ　(XL)　T

#2 Firefighter $11
WARNING: Your mouth will catch fire! The Firefighter is made with 150 grams of Angus beef, jalapeño peppers, and our chef's special hot sauce.
390 calories　Fat 25g　Ⓢ　(XL)　C

#3 Sailor $12
The Sailor is perfect for seafood lovers. It's made with 150 grams of fresh Ahi tuna cooked to perfection.
310 calories　Fat 5g　Ⓢ　(XL)　H

#4 Farmer $8
Our customers have asked for a veggie burger for years, and we've delivered! The Farmer is made with tofu, shiitake mushrooms, and black beans.
250 calories　Fat 0g　V H

Ⓢ = Small size (100g) for $1 less　(XL) = Extra large size (225g) for $3 more

V = Vegetarian　H = Heart-healthy

C = Chicken available at no extra charge　T = Turkey available at no extra charge

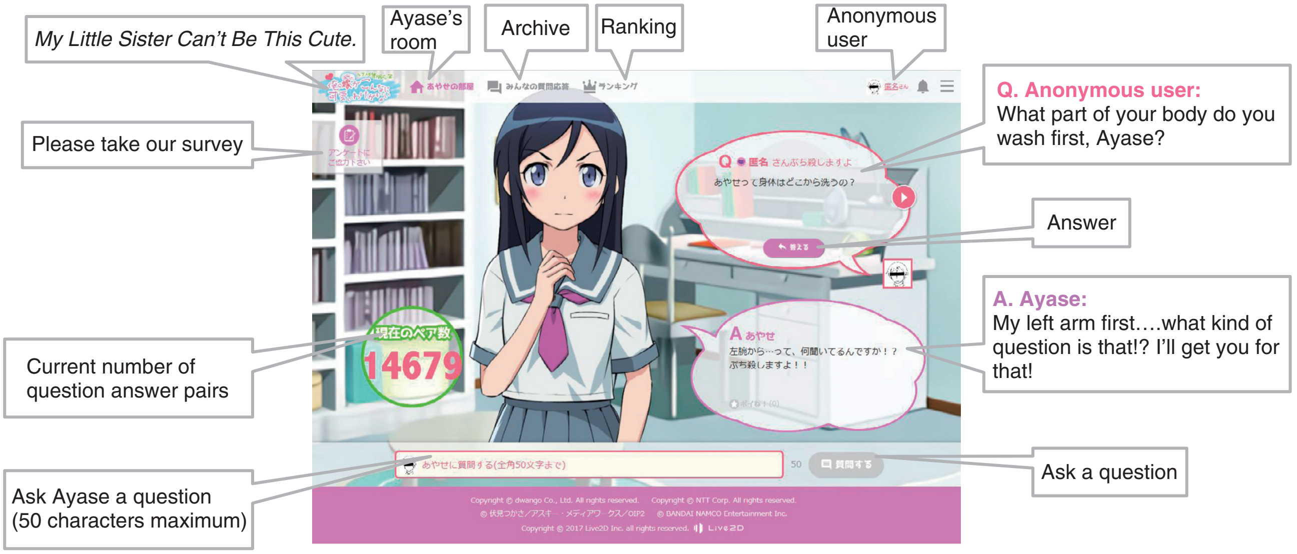
Extras
Add cheese (brie, cheddar, or pepper jack), avocado, or microgreens for $1 per addition. Add ketchup, mustard, or wasabi mayonnaise for free.

問３　Two friends are visiting Rick's Burger Shack to try its Ahi tuna burgers, which have received great reviews from restaurant critics. One friend wants a regular-sized burger with avocado and cheddar cheese. The other wants a smaller portion with wasabi mayonnaise. How much will they pay for their meal? [41]

① $24　② $25　③ $26　④ $27

(c) Reading comprehension (life materials) (long-sentence problems)

Fig. 3. Examples of problems.



Q 匿名 さんぶち殺しますよ
あやせって身体はどこから洗うの？

A あやせ
左腕から…って、何聞いてるんですか！？
ぶち殺しますよ！！

Q. Anonymous user: What part of your body do you wash first, Ayase?

A. Ayase: My left arm first….what kind of question is that!? I'll get you for that!

Copyright © dwango Co., Ltd. All rights reserved.　Copyright © NTT Corp. All rights reserved.
© 伏見つかさ／アスキー・メディアワークス／OIP2　© BANDAI NAMCO Entertainment Inc.
Copyright © 2017 Live2D Inc. all rights reserved.

©Tsukasa Fushimi/ASCII MEDIA WORKS/OIP2 ©BANDAI NAMCO Entertainment Inc. Copyright© 2017 Live2D Inc.

Fig. 4. Screenshot of question answering bulletin board.

society more harmonious and allow us to lead better lives.

Shifting from liberal arts to sciences, and taking time and effort to overcome adversity

—How did you end up on the path to become a researcher?

Actually, my origins lie in the liberal arts. I thought that I would like to study law at university, but I chose a university and faculty leaning towards public policy after many twists and turns. At university, computer education was vigorous, and we used email—which was rare at that time—to exchange coursework and learn programming. I fit in really well in that environment, so I ended up pursuing programming, and when I was in graduate school, I spent a year and a half as a student researcher at IBM Tokyo Research Laboratory (TRL). While at IBM TRL, I encountered natural language processing, so I decided to take the researcher's path and entered the NTT laboratories.

Because I'm also very interested in foreign languages and have studied in the UK, when I joined NTT, I offered to specialize in research on translation. However, at that time, translation was not a statistical process, like it is now, but a rule-based process conforming to rules such as grammar. Since the translation research I wanted to do was shifting to a commercial basis and not being tackled at basic research laboratories, I was assigned to a department responsible for dialogue systems. From that point on, I decided to engage in research on dialogue for the first time. However, that research was really difficult, and even seemingly simple tasks like booking a meeting room by talking to the dialogue system were difficult.

With that difficulty in mind, I became more interested in why people can converse, and gave myself up to research on dialogue. Around 2001, AI was still in a period of winter-like hardship, and dialogue systems were receiving little attention in a field considered to be for diehard researchers only. However, since I began working on dialogue systems back then, I have had many opportunities to present my research—which has been ongoing for 17 years—and this research has started drawing attention.

—What is the driving force behind your research activities?

The driving force behind my research is "curiosity." I consider that everything is interesting, so I take a stance in which I never refuse what comes my way. And since nothing can be done in one leap, I think it is better to do a lot of experiments and find things out one by one. Anyway, I value taking time and effort in my pursuits. In that sense, I think of myself as a craftsman. As well as taking time, research does not necessarily lead to successful outcomes, so I think that it is enough to obtain one or two successes out of 100 attempts.

When I joined NTT, I was coming from a liberal arts background but was with many employees who came from a science background, so I was immersed in things I did not understand. To overcome this hurdle, I worked on problems more carefully than others, and I applied trial and error as much as possible. My experiences of trial and error led to the present and have built confidence in me.

—Do you have anything particularly memorable concerning your research activities?

We participated in a large-scale event held in the USA, called "SXSW" (South by Southwest), which combines a music festival, film festival, and interactive festival, in two consecutive years (2016 and 2017) (**Fig. 5**). In 2016, we showcased the dialogue system of an android that talked in English in Professor Ishiguro's demonstration, and it was well received. In 2017, I was invited as a featured speaker and took the podium with Professor Ishiguro. It was a very honorable occasion, since among Japanese people, only very few have received such an invitation. That year, we wanted to demonstrate more advanced technology than what we did the previous year, so we decided to set up a discussion between humans and robots. We set up a situation in which robots outnumbered humans, based on the assumption that we might create a world in which robots are a majority. We wanted to present a world in which a lot of robots that are smarter than humans exist and to get people to think about what human beings really are.

By the way, the demonstrations in 2016 and 2017 were packed with drama. In 2016, the demo system was only completed the day before the demonstration. The system needed a fairly complicated program, which I continued to write after arrival, not to mention on the plane to the event. We finally got the program to run when the system and the dialogue partner were "face to face," so to speak. Although the situation at that time is now the stuff of legend, it was truly a miracle that the program ran.

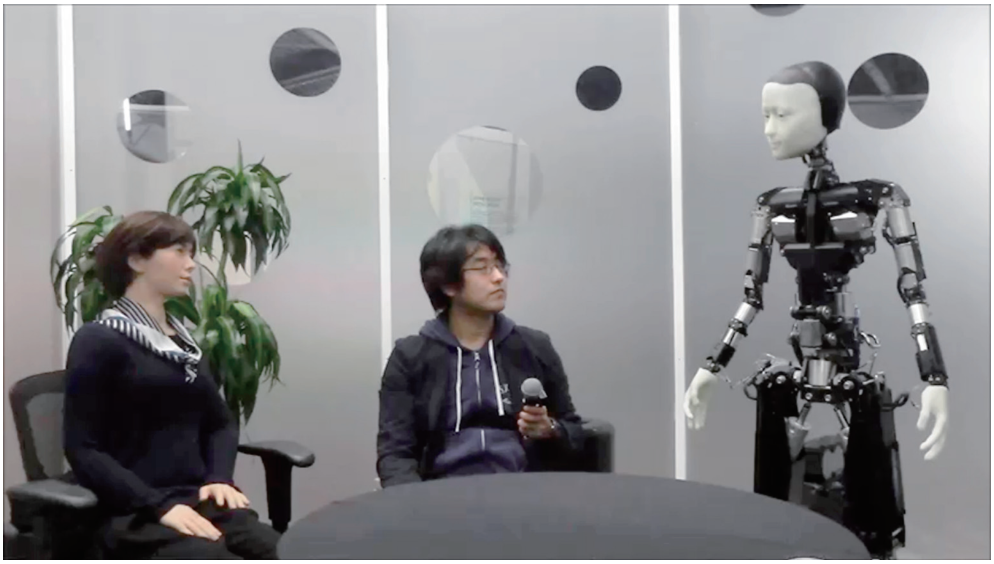


Fig. 5. Demonstration of argumentative dialogue system at SXSW 2017 (in collaboration with Ishiguro Laboratory, Osaka University).

On taking the podium in 2017, the system did not work as we expected. I managed to keep talking for about 30 minutes, and in the meantime, we made various adjustments to the system, and it finally started working. And when it worked, everyone broke into applause. Because I was in front of more than 1000 spectators, I felt frustrated when it didn't work right away. Even so, I was optimistic that it would work at some point. Since I knew that dialogue systems are complicated and don't work that easily, I was probably just keeping my nerve.

Let's find places of interest, things, and other matters, and move forward continuously

—How will research on dialogue systems proceed in the future?

The level of Japanese research in this field is relatively high and has developed rapidly recently. Our research is of course often influenced by the research and development done by major global companies. Nevertheless, many concepts that I thought about have been achieved since I joined the company, and I have a "sense of the future." As for my goal, I'd like the act of speaking with a dialogue system to become an everyday event. Since AI and computers are not good at everything, I think that it is better to divide the tasks that humans will bear responsibility for and the tasks left to computers, and that division will allow us to enter an era in which we can live a richer life. To that end, I want to build a good relationship between humans and computers (robots).

Robots can work longer than people can, and only robots can do advanced simulation calculations. And some tasks are better done if they are not done by humans. From the viewpoint of privacy protection, there are certain occasions, such as counseling, when it is better to leave it to the robot. Recently, dialogue systems have started to be applied in the counseling field. Efforts to introduce dialogue systems in facilities for the elderly and to apply them with the aim of dementia prevention are underway—which is said to be a promising area.

However, I think that whether AI is able to respond to unknown events or whether it could acquire an ego are major challenges. Current AIs do not possess an ego, and discussions on ethics and legal matters in regard to whether an ego should be given to AI are ongoing. Although guidelines concerning those matters are already being created, the industry has not yet reached a consensus. On the other hand, I think that we cannot carry out dialogue, in the true sense of the word, with things that have no ego. In the future, we will also conduct research on giving AI an ego, and I'd like to pursue the concept of "What is a human being?"

—Please say a few words to all young researchers.

I have been very blessed to have good co-workers and leaders. Now that we are in a technological boom, the number of people working on dialogue systems has increased accordingly, but when I started, there were few such researchers. Fortunately, NTT had a group dedicated to research on dialogue, and many employees were engaged in that research. For the time being, I have been concentrating on academic activities in order to publicize the value of the dialogue system and increase the number of like-minded researchers. We organize a symposium called Dialogue System Symposium, and it has expanded to a large-scale event attended by 200 people. Some of our projects have been exhibited at events in the USA and have become internationalized. Through such projects, we have come to share a common sense of value with researchers in various fields and with people in different positions, which is very meaningful.

It is currently difficult for fourth-year undergraduate students aiming to become researchers, postgraduate students in master's programs, and new employees to start research on dialogue systems if there are no colleagues in the same field. I have tried various things to foster the next generation in this field, including publicizing data to reach such people, planning events, and so on. Through these kinds of activities, I'd like to continue to increase the number of fellow researchers.

I think that it is better to share new concepts and ideas to get closer to the truth. It is a good time to be an AI researcher now. The status is also getting better, and compared to when there was no Internet, the amount of data is now abundant, and we have all the tools needed. We are in an era where you can push forward with your own ideas and anyone who has the ability can do anything. Since the world is becoming borderless, aim to be top class and keep the world in mind. Our ability is our output. There are various ways of producing output such as getting papers published, writing programs, and implementing systems, and if you leave your mark through output and results, you will be acknowledged as a fellow researcher. For that reason, trial and error is important. Although there are many different research styles, let's find interesting things that will connect to good outputs.

Acknowledgments

Dr. Higashinaka thanks the National Center for University Entrance Examinations, JC Educational Institute, Inc., Takamiya Academy School Corporation, and Benesse Corporation for providing the problems used in the Torobo Project.

■ Interviewee profile
Ryuichiro Higashinaka

Senior Research Scientist (Senior Distinguished Researcher), Knowledge Media Project, NTT Media Intelligence Laboratories.

He received a B.A. in environmental information, a Master of Media and Governance, and a Ph.D. from Keio University, Kanagawa, in 1999, 2001, and 2008. He joined NTT in 2001. His research interests include building question answering systems and spoken dialogue systems. From November 2004 to March 2006, he was a visiting researcher at the University of Sheffield in the UK. He received the Maejima Hisoka Award from the Tsushinbunka Association in 2014 and the Prize for Science and Technology of the Commendation for Science and Technology by the Minister of Education, Culture, Sports, Science and Technology in 2016. He is a member of the Institute of Electronics, Information and Communication Engineers, the Japanese Society for Artificial Intelligence, the Information Processing Society of Japan, and the Association for Natural Language Processing.

Shift to New Dimensions—Further Initiatives to Deepen Communication Science

Takeshi Yamada

Abstract

NTT Communication Science Laboratories aims to realize communication that *reaches the heart*, from person to person, and between people and computers. We are building fundamental theories in pursuit of the essence of people and of information, and working to create core technologies that will transform society. This article introduces some of our initiatives to push deeper in communication science, in areas including speech and audio processing, dialogue processing, human information science, sports brain science, and machine learning and optimization.

Keywords: artificial intelligence, communication science, corevo

1. Introduction

There have been some remarkable developments in artificial intelligence (AI) recently. Developments in deep learning, in particular, have achieved capabilities approaching those of humans in areas such as speech and image recognition and natural language processing, which were once considered human strengths that could not be matched by computers. At the NTT laboratories, we consider it important and necessary to use these leading-edge technologies and apply them to the issues we are facing. However, as these technologies spread, we must further strive to open new and next-generation areas that are not simple extensions of earlier work, and to make bold changes in our research themes.

We at NTT Communication Science Laboratories (CS Labs) aim to realize communication that *reaches the heart*, from person to person, and between people and computers. We are building fundamental theories in pursuit of the essence of people and of information, and working to create core technologies that will transform society. Our fields of research support the four types of AI that comprise the NTT Group's AI technology called corevo®. These are Agent-AI, Heart-Touching-AI, Ambient-AI, and Network-AI [1]. As such, we have focused mainly on media processing, human information science, and data and machine learning, and have also focused recently on sports brain science (**Fig. 1**). The Feature Articles in this issue introduce some of our initiatives to push deeper in these areas of communication science.

2. Speech and audio processing approaching human capabilities

Agent-AI, a component of corevo, supports interaction between humans and computers. CS Labs is working on speech and audio processing, image recognition, and natural language processing to provide a platform for Agent-AI. Speech recognition under conditions where a single person is speaking into a close-talking microphone has already matured to a level where it is used in everyday life, due to the rapid spread of smartphones and, more recently, devices such as AI speakers.

Research and development (R&D) trends are now shifting toward speech recognition with multiple people speaking freely around a table, some distance from the microphone. This requires increased performance of speech recognition itself, but it is also important to combine it with speech enhancement

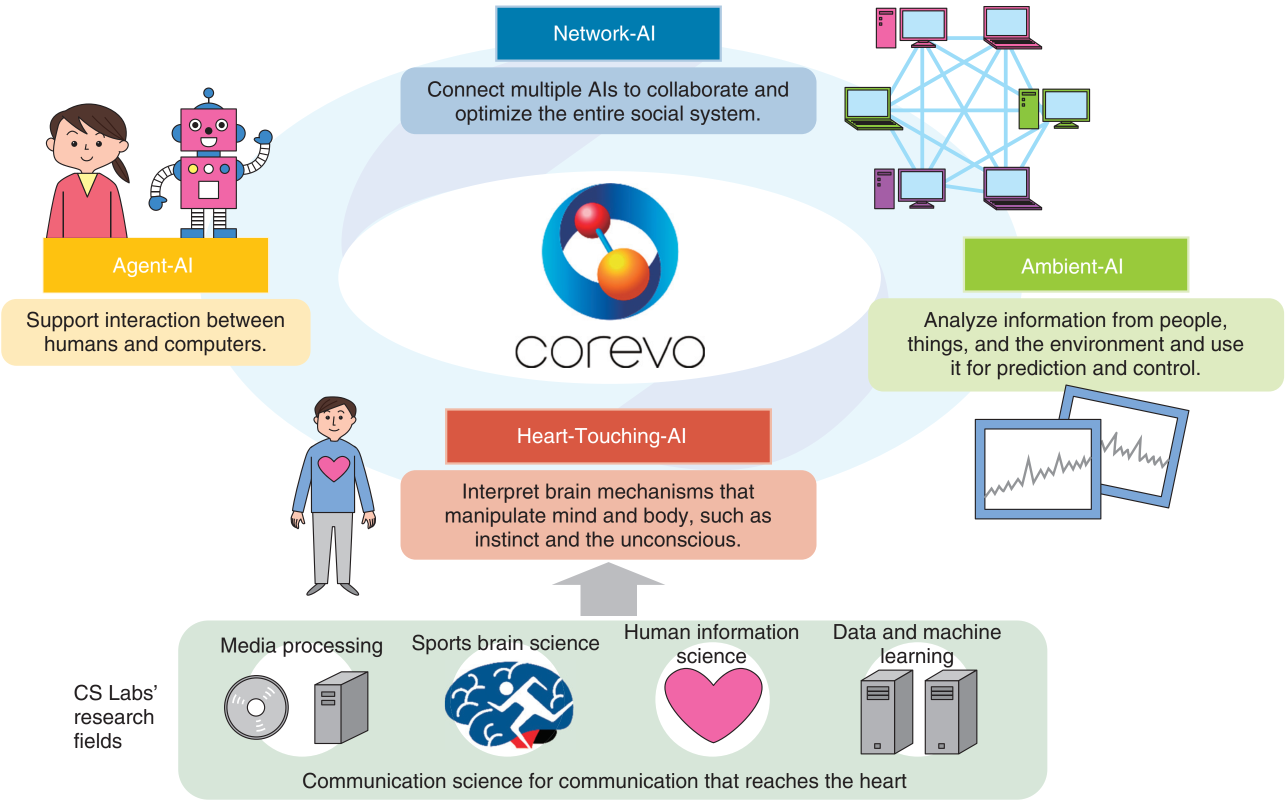

Network-AI
Connect multiple AIs to collaborate and optimize the entire social system.
Agent-AI
Support interaction between humans and computers.
corevo
Ambient-AI
Analyze information from people, things, and the environment and use it for prediction and control.
Heart-Touching-AI
Interpret brain mechanisms that manipulate mind and body, such as instinct and the unconscious.
CS Labs' research fields
Media processing
Sports brain science
Human information science
Data and machine learning
Communication science for communication that reaches the heart

Fig. 1. NTT's four AI directions and communication science.

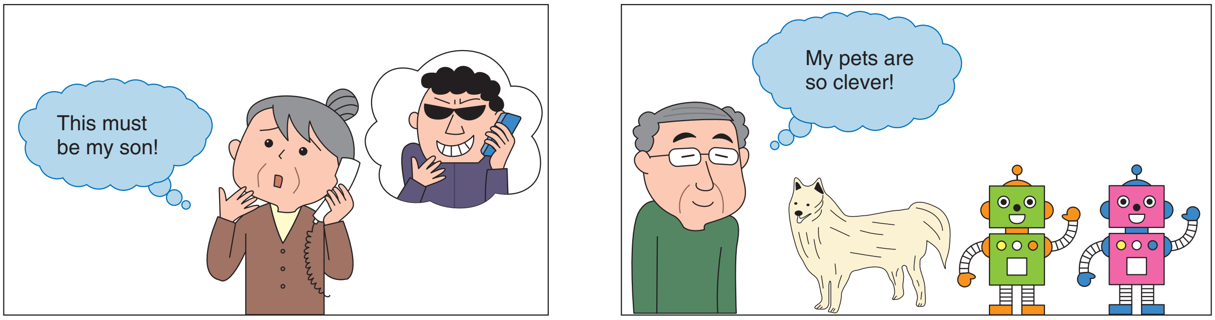
technologies such as denoising to remove background noise in noisy environments, and dereverberation to remove reverberation from the walls and floor of a room. A speech recognition technology from CS Labs combining these capabilities was entered in the 3rd CHiME Speech Separation and Recognition Challenge (CHiME-3) held in 2015, where it placed first among 25 participating organizations [2]. We are working to use these technologies for automatically creating the minutes of a meeting involving several people, while it is still in an experimental stage.

Humans can pick out the voice of a person they want to hear and can understand what that person is saying, in a discussion in a meeting or in a conversation during a party, when there are several people talking or when music is playing in the background. This is called selective listening. In the article in this issue entitled "SpeakerBeam: A New Deep Learning Technology for Extracting Speech of a Target Speaker Based on the Speaker's Voice Characteristics " [3], we introduce a technology that uses deep learning to implement selective listening using computers.

People are also able to visualize a scene in their mind just by listening to a sound. Cross-media scene analysis implements this type of behavior in computers. Deep learning can be applied to various media in a unified framework, so cross-media scene analysis based on deep learning transcends single-media processing, combines multiple types of media, and handles them in a complementary way. Cross-media scene analysis is introduced in the article, "Cross-media Scene Analysis: Estimating Objects' Visuals Only from Audio" [4].

3. Elimination of the human-AI gap

The capabilities of computers are approaching those of humans in certain situations such as those described above, but it will take more time for AI performance to advance beyond the complexity of the human brain. Nevertheless, humans are sometimes easily fooled by telephone payment scams and the like. This is related to certain human cognitive biases. For example, people tend to look only for evidence suited to their existing beliefs, hopes, and suppositions and to interpret circumstances accordingly. This



Confirmation bias: people tend to look only for evidence suited to their existing beliefs, hopes, and suppositions and to interpret circumstances accordingly

ELIZA effect: people have a tendency to arbitrarily anthropomorphize and sense intelligence from behavior that superficially appears to be intelligent

Fig. 2. Examples of cognitive biases.

type of cognitive bias is called a confirmation bias [5], which explains why elderly people easily assume a phone call asking for money is from their child and do not notice any inconsistencies.

People also have a well-known tendency to arbitrarily anthropomorphize and sense intelligence from behavior that superficially appears to be intelligent. This is another cognitive bias called the ELIZA effect, after a computer program developed at MIT (Massachusetts Institute of Technology) in the 1960s [6]. For example, people tend to think their pets are smarter than they really are (**Fig. 2**). Moreover, there are well-known examples of illusions such as the checker shadow illusion, which effectively demonstrates the fact that humans do not see physical quantities as they are, literally, before their very eyes [7].

Effective use of such cognitive biases and illusions in interfaces and for feedback is the key to filling the gap between humans, who are complex and imperfect, and AI, which is currently still immature and limited. Regarding cognitive biases, CS Labs is conducting research on dialogue processing for human-robot interaction, extending the level from casual conversation to serious debate, and at the same time, moving from dialogue with a single robot to that with multiple (two) robots. For conversation with a single person, a single robot may seem to be sufficient as a counterpart. However, multiple robots can be made to appear more intelligent to the human by dividing roles suitably and by utilizing robot-robot interaction that can take advantage of human cognitive biases more effectively. Dialogue with multiple robots seems more natural to the human and can be maintained for longer periods than that with a single robot, even if there are speech recognition mistakes or the dialogue context is lost [8].

CS Labs has also produced interfaces that use illusion, such as Buru-Navi, a device that produces an illusion of being pulled, and Hengento Projection, which produces an illusion of dynamic motion by projecting light onto a printed picture or photograph. The article in this issue, "Ukuzo—A Projection Mapping Technique to Give Illusory Depth Impressions to Two-dimensional Real Objects" [9], introduces a new technique that exploits human visual illusion.

4. Explaining implicit brain activities

To realize communication that reaches the heart, CS Labs is focusing on explaining implicit brain activity related to the basic human senses of vision, hearing, and motion. Research on interfaces as described above, using illusion, is also derived from this approach. This basic research on human information science is the Heart-Touching-AI component of corevo, an AI platform to interpret brain mechanisms that manipulate mind and body, such as instinct and the unconscious, and thereby support humans. Realizing communication that reaches the heart will surely be possible by developing AI that *touches the heart*, namely, Heart-Touching-AI.

Initiatives in sports brain science have recently become a new theme in Heart-Touching-AI [10]. This research makes use of knowledge from brain science that explains implicit brain activity, advanced information and communication technology such as wearable sensors and virtual reality, and machine learning to train the brain to win sports games. This research

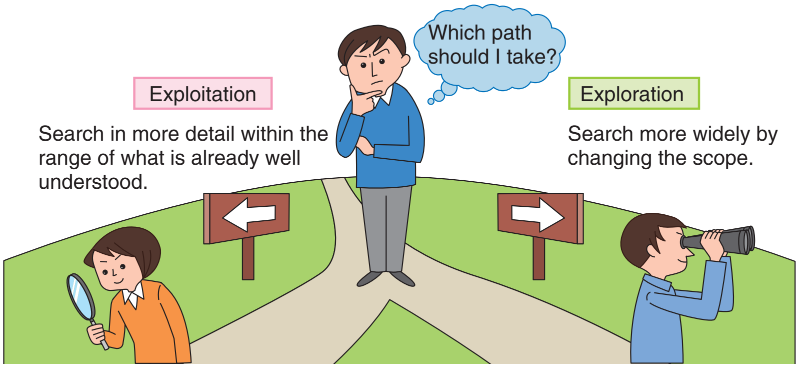

Which path should I take?
Exploitation
Search in more detail within the range of what is already well understood.
Exploration
Search more widely by changing the scope.

Fig. 3. Exploration-Exploitation Dilemma.

is not about making the body stronger but is focused on finding a way to coordinate the body parts optimally, and how to control the human mental state well. For example, on the basis of the fact that hearing has better time resolution than vision, we have proposed a method that provides intuitive feedback by measuring the activity of muscles in certain parts of the body and converting them to sounds in real time.

Beyond sports, it is also important to support people in their daily lives in order to maximize use of the implicit capabilities of their minds and bodies. The article in this issue, “Measuring, Understanding, and Cultivating Wellbeing in the Age of Technology” [11], introduces our challenge to identify design guidelines for human wellbeing, which cannot be grasped qualitatively at a glance, treating it quantitatively from a human-scientific perspective, and improving it.

5. Machine learning and optimization

With machine learning technology, it is possible from large amounts of data to automatically discover patterns that even human experts would not notice. The collective behavior of people is complex, as with a crowd of people walking in a town, but there are definite patterns, as each person has his or her own typical intention. At NTT, we are conducting R&D on technology that can predict the risks of congestion or delays in the near future from real-time data obtained by observing the flow of people. It can then automatically provide near-optimal on-line navigation to avoid such risks as efficiently as possible without impeding people in fulfilling their intentions. Here, we use techniques such as multi-agent simulation and Bayesian optimization [12].

With limited computing resources, it is often difficult to search for the optimal solution with a fine-toothed comb (i.e., a simple exhaustive search) when there are vast numbers of combinations. In such cases, we are obliged to choose whether to prioritize an Exploitation approach, in which we search in more detail within the range of what is already well understood, or an Exploration approach, in which we change the scope to search more widely. This is called the Exploration-Exploitation Dilemma (**Fig. 3**). Bayesian optimization is a method for efficiently narrowing down candidate solutions in a search, with consideration for the balance between Exploitation and Exploration [13]. NTT is conducting this R&D as a component of Ambient-AI, which provides AI as intelligence in the Internet of Things, analyzing information from people, things, and the environment, and using it for prediction and control.

On the other hand, there are also cases when data structures can be devised so that all combinations can be enumerated efficiently and a strictly optimal solution computed, even when a simple exhaustive search is difficult. The article “Network Reliability Optimization by Using Binary Decision Diagrams” [14] in this issue introduces examples of formerly unsolvable large-scale problems that were solved with surprising efficiency by using data structures such as binary decision diagrams.

As Ambient-AI progresses and AI technologies are used more often in networks, multiple AIs will collaborate and optimize overall social systems, forming Network-AI.

6. Future prospects

This article has introduced some new initiatives in communication science at CS Labs. New initiatives also have accompanying risks. We cannot always obtain the scientific results we hope for immediately. In R&D as well, we face the Exploration-Exploitation Dilemma. We need to decide whether we should take an Exploitation approach, using the latest technologies and attempting to apply them skillfully to the problems we are facing, or give priority to Exploration, working to open up new dimensions that are not extensions of earlier work [15]. We at CS Labs will continue to face challenging problems, giving priority to Exploration approaches. To be sure, we need to be careful of cognitive biases in conducting research, so that we do not only look for and interpret evidence that is well suited to our own assumptions [16].

References

[1] T. Yamada, S. Takahashi, F. Naya, T. Ikebe, and S. Furukawa, “Artificial Intelligence Research Activities and Directions in the NTT Group,” NTT Technical Review, Vol. 14, No. 5, 2016.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201605fa1.html

[2] NTT press release, “NTT Achieved Top Performance in a Noisy Speech Recognition International Challenge,” Dec. 14, 2015.
http://www.ntt.co.jp/news2015/1512e/151214a.html

[3] 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

[4] G. Irie, H. Kameoka, A. Kimura, K. Hiramatsu, and K. Kashino, “Cross-media Scene Analysis: Estimating Objects’ Visuals Only from Audio,” NTT Technical Review, Vol. 16, No. 11, pp. 35–40, 2018.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201811fa5.html

[5] M. Akiyama, “Psychology of Deception and Trouble for Seniors,” Kokumin Seikatsu, Vol. 13, pp. 1–4, 2013 (in Japanese).

[6] J. Weizenbaum, “ELIZA—A Computer Program for the Study of Natural Language Communication between Man and Machine,” Communications of the ACM, Vol. 9, No. 1, pp. 36–45, 1966.

[7] Website of Illusion Forum by CS Labs, Checker Shadow (in Japanese).
http://www.kecl.ntt.co.jp/IllusionForum/v/checkerShadow/ja/index.html

[8] NTT press release issued on January 1, 2018 (in Japanese).
http://www.ntt.co.jp/news2018/1801/180131b.html

[9] 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

[10] M. Kashino, “Understanding and Shaping the Athlete’s Brain—NTT Sports Brain Science Project—,” NTT Technical Review, Vol. 16, No. 3, 2018.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201803fa1.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] F. Naya, M. Miyamoto, and N. Ueda, “Optimal Crowd Navigation via Spatio-temporal Multidimensional Collective Data Analysis,” NTT Technical Review, Vol. 15, No. 9, 2017.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201709fa5.html

[13] E. Brochu, V. M. Cora, and N. de Freitas, “A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning,” arXiv:1012.2599, 2010.

[14] M. Nishino, T. Inoue, N. Yasuda, S. Minato, and M. Nagata, “Network Reliability Optimization by Using Binary Decision Diagrams,” NTT Technical Review, Vol. 16, No. 11, pp. 25–29, 2018.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201811fa3.html

[15] J. G. March, “Exploration and Exploitation in Organizational Learning,” Organization Science, Vol. 2, No. 1, pp. 71–87, 1991.

[16] R. Nuzzo, “How Scientists Fool Themselves – and How They Can Stop,” Nature, Vol. 526, No. 7572, pp. 182–185, 2015.

Takeshi Yamada

Vice President and Head of NTT Communication Science Laboratories.

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

SpeakerBeam: A New Deep Learning Technology for Extracting Speech of a Target Speaker Based on the Speaker's Voice Characteristics

Marc Delcroix, Katerina Zmolikova, Keisuke Kinoshita, Shoko Araki, Atsunori Ogawa, and Tomohiro Nakatani

Abstract

In a noisy environment such as a cocktail party, humans can focus on listening to a desired speaker, an ability known as selective hearing. Current approaches developed to realize computational selective hearing require knowing the position of the target speaker, which limits their practical usage. This article introduces SpeakerBeam, a deep learning based approach for computational selective hearing based on the characteristics of the target speaker's voice. SpeakerBeam requires only a small amount of speech data from the target speaker to compute his/her voice characteristics. It can then extract the speech of that speaker regardless of his/her position or the number of speakers talking in the background.

Keywords: deep learning, target speaker extraction, SpeakerBeam

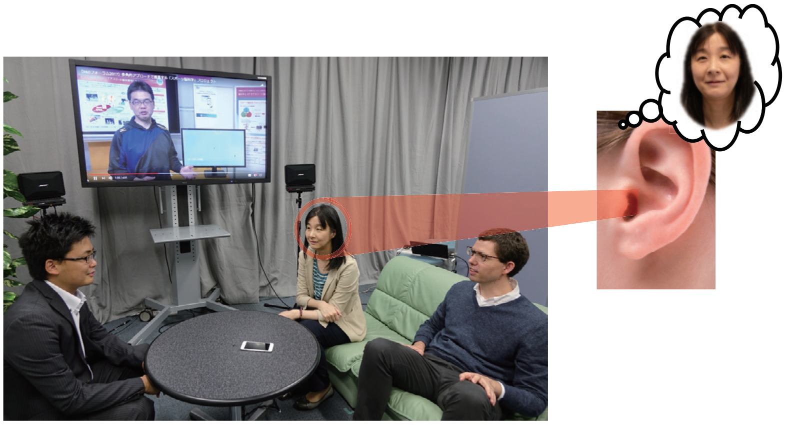
1. Introduction

Automatic speech recognition technology has progressed greatly in recent years, thus enabling the rapid adoption of speech interfaces in smartphones or smart speakers. However, the performance of current speech interfaces deteriorates severely when several people speak at the same time, which often happens in everyday life, for example, when we take part in discussions or when we are in a room where a television is on in the background. The main reason for this problem arises from the inability of current speech recognition systems to focus solely on the voice of the target speaker when several people are speaking [1].

In contrast to current speech recognition systems, human beings have a selective hearing ability (see **Fig. 1**), meaning that they can focus on speech spoken by a target speaker even in the presence of noise or other people talking in the background by exploiting information about the characteristics of the voice and the position of the target speaker.

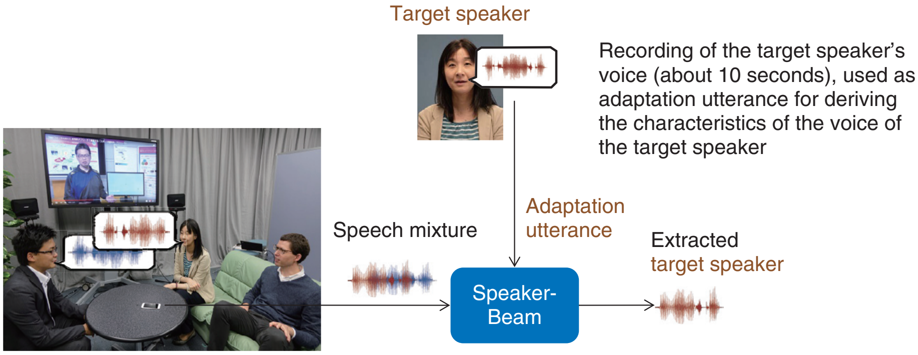
Previous attempts to replicate computationally the human selective hearing ability used information about the target speaker position [1]. With these approaches, it is hard to focus on a target speaker when the speaker's position is unknown or when he/she moves, which limits their practical usage.

We have proposed SpeakerBeam [2], a novel approach to mimic the human selective hearing ability that focuses on the target speaker's voice characteristics (see **Fig. 2**). SpeakerBeam uses a deep neural network to extract speech of a target speaker from a mixture of speech signals. In addition to the speech mixture, SpeakerBeam also inputs the characteristics of the target speaker's voice so that it can extract speech that matches these characteristics. These voice characteristics are computed from an adaptation utterance, that is, another recording (about 10 seconds long) of the target speaker's voice.



Ability to listen only to a target speaker by focusing on the characteristics of his/her voice (pitch, timbre, etc.) and the direction of arrival of the sound

Fig. 1. Human selective hearing ability.


Target speaker
Recording of the target speaker's voice (about 10 seconds), used as adaptation utterance for deriving the characteristics of the voice of the target speaker
Adaptation utterance
Speech mixture
Extracted target speaker
Speaker-Beam

Example recording conditions

Fig. 2. SpeakerBeam's selective hearing capability.

Consequently, SpeakerBeam enables the extraction of the voice of a target speaker based solely on the target speaker's voice characteristics without knowing his/her position, thus opening new possibilities for the speech recognition of multi-party conversations or speech interfaces for assistant devices.

In the remainder of this article we briefly review conventional approaches for selective hearing. We then detail the principles of the proposed SpeakerBeam approach and present experimental results confirming its potential. We conclude this article with an outlook on possible applications of SpeakerBeam and future research directions.

2. Conventional approaches for computational selective hearing

Much research has been done with the aim of finding a way to mimic the selective hearing ability of human beings using computational models. Most of the previous attempts focused on audio speech separation approaches that separate a mixture of speech signals into each of its original components [1, 3]. Such approaches use characteristics of the sound mixture such as the direction of arrival of the sounds to distinguish and separate the different sounds.

Speech separation can separate all the sounds in a mixture, but for this purpose it must know or be able to estimate the number of speakers included in the mixture, the position of all the speakers, and the

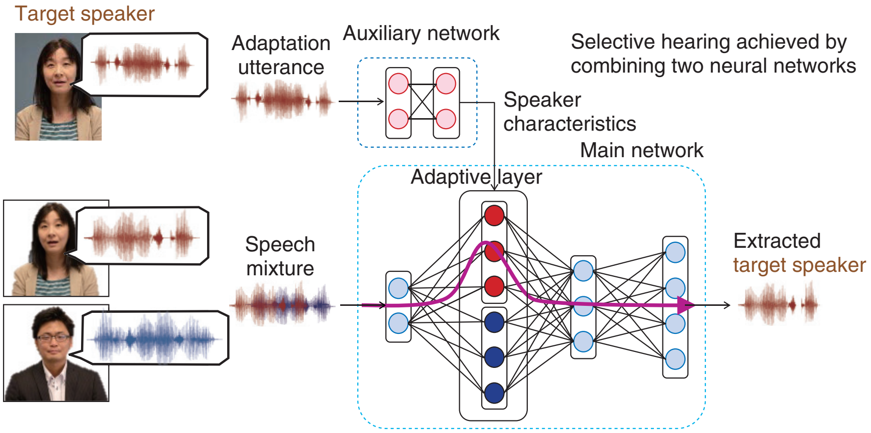


Fig. 3. Novel deep learning architecture developed for SpeakerBeam.

background noise statistics. These conditions often change dynamically, making their estimation difficult and thus limiting the actual usage of the separation methods. Moreover, to achieve selective hearing, we still need to inform the separation system which of the separated signals corresponds to that of the target speaker.

3. Principles of SpeakerBeam

SpeakerBeam focuses on extracting only the target speaker instead of separating all components in the mixture. By focusing on the simpler task of solely extracting speech that matches the voice characteristics of the target speaker, SpeakerBeam avoids the need to estimate the number of speakers, the position, or the noise statistics. Moreover, it can perform target speech extraction using a short adaptation utterance of only about 10 seconds.

SpeakerBeam is implemented by using a deep neural network that consists of a main network and an auxiliary network as described below and shown in **Fig. 3**.

(1) The main network inputs the speech mixture and outputs the speech that corresponds to the target speaker. The main network is a regular multi-layer neural network with one of its hidden layers replaced by an adaptive layer [4, 5]. This adaptive layer can modify its parameters depending on the target speaker to be extracted; namely, it can change its parameters depending on the characteristics of the voice of the target speaker provided by the auxiliary network.

(2) The auxiliary network is a multi-layer neural network that inputs a recording of only the voice of the target speaker (adaptation utterance) that is different from that in the speech mixture. The auxiliary network outputs the characteristics of the voice of the target speaker.

These two networks are connected to each other and trained jointly to optimize the speech extraction performance. Training the auxiliary network jointly with the main network enables the system to learn automatically from data the features that best characterize the target speaker's voice, thus avoiding the complex task of manually engineering features characterizing the target speaker's voice. Moreover, by training the network with a large amount of training data covering various speakers and background noise conditions, SpeakerBeam can learn to achieve selective hearing even for speakers that were not included in the training data. Details of the network architecture and training procedure are explained in our published report [2].

4. Performance of SpeakerBeam

We conducted experiments to evaluate the speech extraction performance of SpeakerBeam and its impact on speech recognition [2]. We used a corpus consisting of sentences read from English newspaper articles and created artificially mixtures of two speakers. Although SpeakerBeam can work with a single microphone, it achieves better performance when using more microphones. In this experiment, we used eight microphones and combined SpeakerBeam with microphone array processing (i.e., beamforming).

An example of processed speech using SpeakerBeam

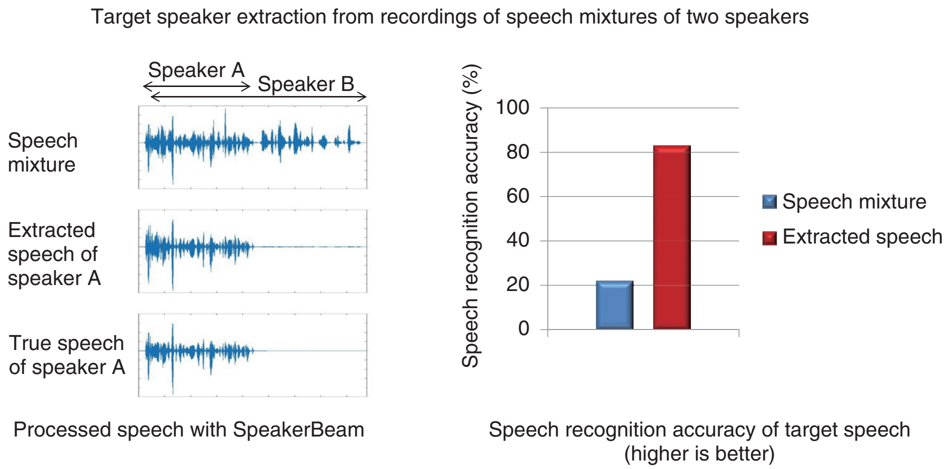


Fig. 4. Evaluation of speech extraction performance and automatic speech recognition with SpeakerBeam.

and the speech recognition accuracy obtained when recognizing mixtures of two speakers with SpeakerBeam (red bar) and without it (blue bar) are shown in **Fig. 4**. We observed a 60% relative improvement in speech recognition performance with SpeakerBeam.

SpeakerBeam can also be employed to improve the audible quality. Interested readers can refer to a video [6] to appreciate the target speaker extraction performance in realistic conditions (real recordings in reverberant conditions with music in the background).

5. Outlook

SpeakerBeam is a novel approach to perform computational selective hearing that offers several advantages compared to previous approaches. For example, it can track a target speaker regardless of the number of speakers or noise sources in the mixture and regardless of the speaker's position. This opens new possibilities for speech recognition of multi-party conversations, speech interfaces for assistant devices such as smart speakers, or for voice recorders and hearing aids that could focus on the speech of a target speaker.

However, there are some issues that need to be addressed before SpeakerBeam can be widely used. For example, speech extraction performance degrades when two speakers with similar voices speak at the same time. To tackle this issue, we plan to investigate improved target speaker characteristics that could better distinguish speakers and to combine target speaker characteristics with location information such as direction-of-arrival features.

Acknowledgment

Part of this development was achieved through a research collaboration with Brno University of Technology in Brno, Czech Republic.

References

[1] T. Hori, S. Araki, T. Nakatani, and A. Nakamura, "Advances in Multispeaker Conversational Speech Recognition and Understanding," NTT Technical Review, Vol. 11, No. 12, 2013.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201312fa3.html

[2] K. Zmolikova, M. Delcroix, K. Kinoshita, T. Higuchi, A. Ogawa, and T. Nakatani, "Learning Speaker Representation for Neural Network Based Multichannel Speaker Extraction," Proc. of the 2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU 2017), Okinawa, Japan, Dec. 2017.

[3] S. Makino, H. Sawada, and S. Araki, "Blind Audio Source Separation Based on Independent Component Analysis," In: M. E. Davies, C. J. James, S. A. Abdallah, and M. D. Plumbley (eds), Independent Component Analysis and Signal Separation—Proc. of ICA 2007, Lecture Notes in Computer Science, Vol. 4666, Springer, Berlin, Heidelberg, 2007.

[4] M. Delcroix, K. Kinoshita, A. Ogawa, S. Karita, T. Higuchi, and T. Nakatani, "Personalizing Your Speech Interface with Context Adaptive Deep Neural Networks," NTT Technical Review, Vol. 15, No. 11, 2017.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201711fa6.html

[5] M. Delcroix, K. Kinoshita, A. Ogawa, C. Huemmer, and T. Nakatani, "Context Adaptive Neural Network Based Acoustic Models for Rapid Adaptation," IEEE/ACM Trans. Audio, Speech and Lang. Proc., Vol. 26, No. 5, pp. 895–908, 2018.

[6] SpeakerBeam video in English, https://www.youtube.com/watch?v=7FSHgKip6vI
SpeakerBeam video in Japanese, https://youtu.be/BM0DXWgGY5A

Marc Delcroix

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

He received an M.Eng. from the Free University of Brussels, Brussels, Belgium, and the Ecole Centrale Paris, Paris, France, in 2003 and a Ph.D. from the Graduate School of Information Science and Technology, Hokkaido University, in 2007. He was a research associate at NTT Communication Science Laboratories from 2007–2008 and 2010–2012 and then became a permanent research scientist at the same lab in 2012. His research interests include robust multi-microphone speech recognition, acoustic model adaptation, integration of speech enhancement front-end and recognition back-end, speech enhancement, and speech dereverberation. He took an active part in the development of NTT robust speech recognition systems for the REVERB and CHiME 1 and 3 challenges, which all achieved best performance results in the tasks. He was one of the organizers of the REVERB challenge 2014 and of the 2017 Institute of Electrical and Electronics Engineers (IEEE) Automatic Speech Recognition and Understanding Workshop (ASRU 2017). He is a member of the IEEE Signal Processing (SP) Society Speech and Language Processing Technical Committee (SL-TC). He is also a visiting lecturer at the Faculty of Science and Engineering of Waseda University, Tokyo. He received the 2005 Young Researcher Award from the Kansai section of the Acoustic Society of Japan (ASJ), the 2006 Student Paper Award from the IEEE Kansai section, the 2006 Sato Paper Award from ASJ, the 2015 IEEE ASRU Best Paper Award Honorable Mention, and the 2016 ASJ Awaya Young Researcher Award. He is a senior member of IEEE and a member of ASJ.

Katerina Zmolikova

Ph.D. student, Brno University of Technology.

She received a Bc. degree in information technology in 2014 and an Ing. degree in mathematical methods in information technology in 2016 from the Faculty of Information Technology, Brno University of Technology (BUT), Czech Republic. Since 2013, she has been part of the Speech@FIT research group at BUT, where she is currently working towards her Ph.D. degree. Her research interests include robust speech recognition, speech separation, and deep learning.

Keisuke Kinoshita

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

He received an M.Eng. and a Ph.D. from Sophia University, Tokyo, in 2003 and 2010. Since joining NTT in 2003, he has been researching speech and audio signal processing. His research interests include single- and multichannel speech enhancement and robust automatic speech recognition. He received a Paper Award from the Institute of Electronics, Information and Communication Engineers (IEICE) in 2006, ASJ Technical Development Award in 2009, an ASJ Awaya Young Researcher Award in 2009, a Japan Audio Society Award in 2010, and the Maejima Hisoka Award in 2017. He is a member of IEEE, ASJ, and IEICE.

Shoko Araki

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

She received a B.E. and M.E. from the University of Tokyo in 1998 and 2000, and a Ph.D. from Hokkaido University in 2007. She joined NTT in 2000 and has since been engaged in researching acoustic signal processing, array signal processing, blind source separation, meeting diarization, and auditory scene analysis. She has been on the organizing committees of several conferences, including ICA 2003 (Fourth International Symposium on Independent Component Analysis and Blind Signal Separation), IWAENC 2003 (2003 International Workshop on Acoustic Echo and Noise Control), and WASPAA (IEEE Workshop on Applications of Signal Processing to Audio and Acoustics) 2007 and 2017, and HSCMA2017 (Fifth Joint Workshop on Hands-free Speech Communication and Microphone Arrays). She has also served as the evaluation co-chair of the Signal Separation Evaluation Campaign (SiSEC) 2008, 2010, and 2011. She has been a member of the IEEE Signal Processing Society Audio and Acoustics Technical Committee (AASP-TC) since 2014, and a board member of ASJ since 2017. She received the 19th Awaya Prize from ASJ in 2001, the Best Paper Award of IWAENC in 2003, the TELECOM System Technology Award from the Telecommunications Advancement Foundation in 2004 and 2014, the Academic Encouragement Prize from IEICE in 2006, the Itakura Prize Innovative Young Researcher Award from ASJ in 2008, the Commendation for Science and Technology by the Minister of Education, Culture, Sports, Science and Technology Young Scientists' Prize in 2014, an IEEE Best Paper Award in 2015, and an IEEE ASRU 2015 Best Paper Award Honorable Mention in 2015. She is a member of IEEE, IEICE, and ASJ.

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 is engaged in research on speech recognition and speech enhancement. He is a member of IEEE and the International Speech Communication Association (ISCA), IEICE, the Information Processing Society of Japan (IPSJ), and ASJ. He received ASJ Best Poster Presentation Awards in 2003 and 2006.

Tomohiro Nakatani

Group Leader and Senior Distinguished Researcher, Signal Processing Research Group, Media Information Laboratory, NTT Communication Science Laboratories.

He received a B.E., M.E., and Ph.D. from Kyoto University in 1989, 1991, and 2002. Since joining NTT as a researcher in 1991, he has been investigating speech enhancement technologies for developing intelligent human-machine interfaces. He was a visiting scholar at Georgia Institute of Technology in 2005. Since 2008, he has been a visiting assistant professor in the Department of Media Science, Nagoya University, Aichi. He received the 2005 IEICE Best Paper Award, the 2009 ASJ Technical Development Award, the 2012 Japan Audio Society Award, the 2015 IEEE ASRU Best Paper Award Honorable Mention, and the 2017 Maejima Hisoka Award. He was a member of the IEEE SP Society AASP-TC from 2009 to 2014 and served as the chair of the AASP-TC Review Subcommittee from 2013 to 2014. He has been a member of the IEEE SP Society SL-TC since 2016. He served as an associate editor of the IEEE Transactions on Audio, Speech and Language Processing from 2008 to 2010, Chair of the IEEE Kansai Section Technical Program Committee from 2011 to 2012, Technical Program co-Chair of IEEE WASPAA-2007, Workshop co-Chair of the 2014 REVERB Challenge Workshop, and as a General co-Chair of the IEEE ASRU. He is a member of IEICE and ASJ.

Network Reliability Optimization by Using Binary Decision Diagrams

Masaaki Nishino, Takeru Inoue, Norihito Yasuda, Shin-ichi Minato, and Masaaki Nagata

Abstract

We introduce an algorithm that can automatically identify communication network topologies that are robust against failures. Robustness is usually assessed by the metric of network reliability. Since communication networks are a critical infrastructure, designing networks with high network reliability values is essential. However, the problem of finding a network topology that offers the maximum network reliability is a computationally difficult problem, and previous methods therefore restrict their application area to very small networks. Our proposed method exploits the novel data structure called binary decision diagrams, which makes it possible to find the most reliable network topology for communication networks with more than 10 times as many nodes (100) than is possible with previous methods.

Keywords: network reliability, optimization, binary decision diagram

1. Network reliability

Communication networks have become a key infrastructure and so must work without failure. However, network components such as links and nodes may fail for several reasons. Since these failures are inevitable, communication networks must be designed so that they continue to function even if these failures occur. How can we design such networks?

An example of a simple communication network is shown in **Fig. 1(a)**. Since the network connects two terminals with one link, the network will fail if the link fails. In contrast, a network with an additional link is shown in **Fig. 1(b)**. This network will continue to work if one of the links fails. Therefore, this network is more reliable than the single-link network.

Network reliability is a measure used for quantifying how robust a communication network is against failures. It is defined as the probability that the network will continue to support communications assuming that the failure of its components follows some probabilistic distributions.

Let us compute the network reliabilities of the networks in Fig. 1. We assume that each link fails independently with a probability of 20%. Since the probability of the network in Fig. 1(a) working equals the probability that the link between terminals works, its network reliability is 80%. In contrast, the network in Fig. 1(b) will continue to work unless both links fail simultaneously. Since the probability of such an event happening is 20% × 20% = 4%, the reliability of the network is 96%. In this way, network reliability can quantify the robustness of a communication network. We note, however, that evaluating network reliability is a computationally difficult problem and becomes infeasible for large networks.

2. Network reliability maximization

If a communication network is assessed to have sufficient reliability, we can continue to use it without modification. If, however, the reliability is insufficient, remedial action is needed. A typical approach is adding links to the network since that always improves the reliability. The task of finding the best way to add links to improve reliability can be formulated and solved as the combinatorial optimization*1 problem called the network reliability maximization problem. In what follows, we make the realistic assumption that the total budget for adding links to

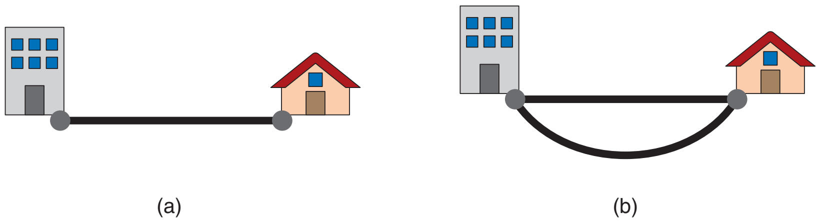
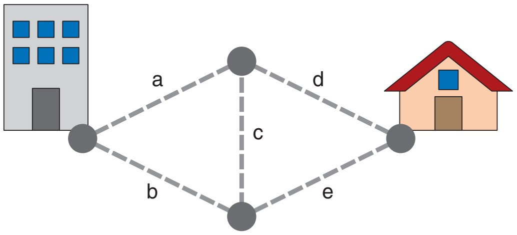


Fig. 1. Communication network examples. Network (b) is more robust against link failures than network (a).

Adding links to maximize network reliability while holding the total cost below 10
The costs and failure probabilities are listed in the table.



Link	Fail probability	Cost
a	5%	2
b	8%	3
c	9%	3
d	7%	3
e	6%	1

Fig. 2. Budget constrained network reliability maximization problem.

the network is limited, and we want to maximize network reliability under this budget constraint. We call this problem the budget constrained network reliability maximization problem.

The input of this optimization problem is a communication network that consists of nodes and links, the total budget, and a set of candidate positions for adding new links. We assume that for each candidate position, the cost for adding a link to the position, and the probability that the added link will fail are known in advance. The output of the problem is the communication network topology that achieves the maximum network reliability score and satisfies the constraint that the total cost incurred in adding links does not exceed the budget. An example of a budget constrained network reliability maximization problem and the set of candidate solutions of the problem are respectively shown in **Figs. 2** and **3**. Of the candidate solutions shown in Fig. 3, the center one is the optimal solution—the communication network topology that achieves the maximum reliability among those satisfying the budget constraint.

We have seen that we can design a communication network that is robust against failures by solving the network reliability maximization problem. However, this problem is known to be computationally hard; it takes a prohibitively long time even if we exploit powerful modern computers.

A straightforward approach to solve the network reliability maximization problem is to first enumerate all candidate network topologies that can be made while satisfying the budget constraint and then evaluating the network reliability of each candidate. However, this simple approach has two potential problems. First, the number of candidate topologies satisfying a constraint may grow exponentially with network size. Second, evaluating the reliability of a candidate solution also takes an exponential amount of time. To evaluate the reliability of a network, we must enumerate all the possible link failure patterns with which the network works. Since the number of such failure patterns grows exponentially with

*1 Combinatorial optimization: The problem of finding the best combination from the set of combinations that satisfies given constraints. Traveling salesman problems and knapsack problems are typical examples of combinatorial optimization problems.

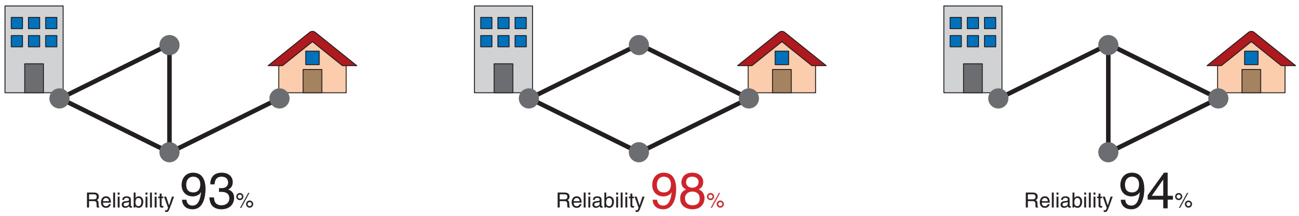
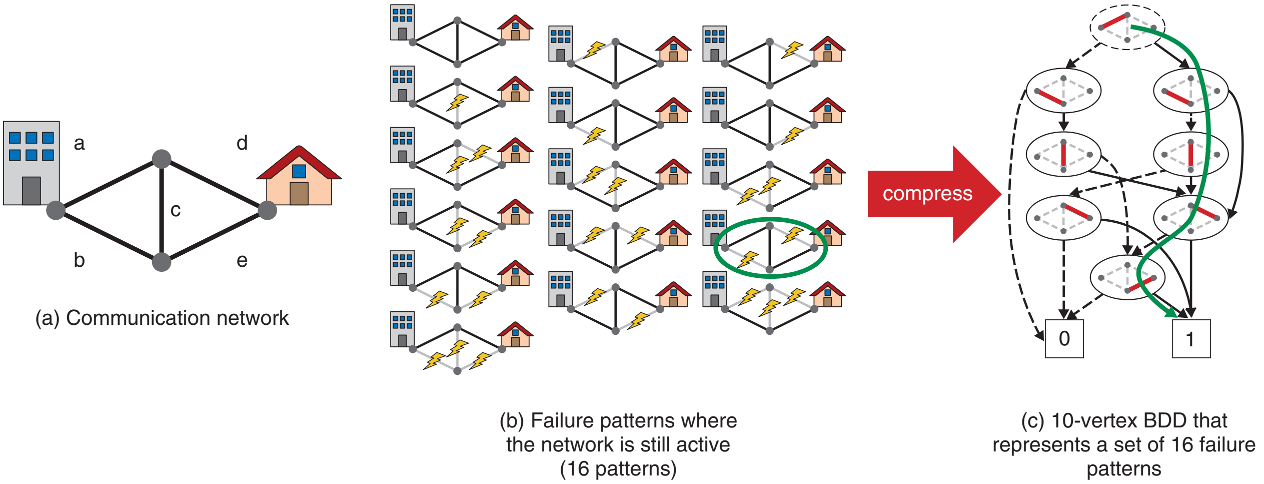


Fig. 3. Example solutions for the problem in Fig. 2. The network in the center is an optimal solution.



(a) Communication network

(b) Failure patterns where the network is still active (16 patterns)

(c) 10-vertex BDD that represents a set of 16 failure patterns

Fig. 4. Using a BDD to represent failure patterns.

network size, the computation time also grows exponentially. Due to these two difficulties, previous methods can find optimal solutions only for very small networks, that is, networks with fewer than 20 nodes.

3. Efficient optimization with binary decision diagrams

We propose an efficient algorithm[*2] for finding the optimal solutions of budget constrained reliability maximization problems [1]. This algorithm can find optimal solutions for communication networks with more than 100 nodes. It can handle networks that are 10 times the size of those possible with previous methods. Moreover, the proposed method works more than 10 thousand times faster than existing methods. The key to our algorithm is that it uses the data structure called binary decision diagrams (BDDs) [2].

BDDs can represent a set of failure patterns in a compressed form. We show in **Fig. 4** a communication network (Fig. 4(a)) and its possible failure patterns where the network still survives (Fig. 4(b)). In Fig. 4(c), we show a BDD representing the set of failure patterns in Fig. 4(b). A BDD is a directed graph that consists of two types of vertices—the circles and rectangles in the figure. Every circle vertex has two arcs—a solid arc and a dashed arc—and is associated with a link of the communication network in Fig. 4(a). Every rectangle vertex has a label of either **1** or **0** and they are placed at the bottom.

In the BDD shown, every failure pattern in Fig. 4(b) corresponds to a directed path in the BDD from the root BDD vertex to the rectangle terminal vertex with label **1**. We can obtain the path that corresponds to a failure pattern in the following way; given a failure pattern, we first select the root BDD vertex (say *v*). Then we check the link associated with the label of *v*.

[*2] Algorithm: A computation procedure for solving a problem. A computer can solve various problems by running algorithms.

If the link works in the failure pattern, we follow the solid arc and update *v* to the node reached. Otherwise, if the corresponding link fails in the current failure pattern, we follow the dashed arc of *v* and update *v* to the node reached.

By repeatedly updating *v* depending on its labels, we obtain a path in the BDD. For example, the failure pattern marked by the green circle in Fig. 4(b) corresponds to the path on the BDD represented by the green arrow. By representing each failure pattern as a path, the BDD can share equivalent partial paths and thus represent the set of paths in a succinct way. For example, the BDD in Fig. 4(c) represents 16 failure patterns as a directed graph with 10 vertices. Since we need 5 vertices for each failure pattern if they are represented as paths, the compression ratio of this BDD is 12.5%. When the input network is large, the compression rate is smaller.

By using a BDD to compress the set of failure patterns, we can accelerate some of the computations needed in solving budget constrained network reliability maximization problems. First, we can accelerate network reliability evaluation. By using a BDD, we can precisely evaluate network reliability in time linear to the number of BDD vertices. If the compression ratio of a BDD is 99%, just 1% of the original computation cost is needed to evaluate network reliability. Second, a BDD can be used to estimate the amount of improvement that can be achieved by adding a link to the network. Such estimations allow us to efficiently discard candidate network topologies that may not achieve high reliability. This can also reduce the computation cost needed to find an optimal solution. In this way, BDDs enable us to optimize the reliability of large communication networks.

Up to this point we have focused on the problem of maximizing network reliability under budget constraints. In the real world, another design goal is to achieve reliability higher than a given threshold value at minimum cost. Our algorithm can be applied to this problem and can efficiently find the minimum cost solution.

4. Conclusion

We have introduced an efficient algorithm that can find network topologies that maximize network reliability. This algorithm can be applied to designing infrastructure networks in addition to communication networks such as road networks, rail networks, and power transmission networks, all of which demand high reliability. Of course, we have to consider several aspects other than reliability when designing communication networks. The most reliable network might not be the best one if other aspects are considered. Future work includes enhancing our optimization method so that it can simultaneously handle multiple aspects, one of which is reliability.

References

[1] M. Nishino, T. Inoue, N. Yasuda, S. Minato, and M. Nagata, “Optimizing Network Reliability via Best-first Search over Decision Diagrams,” Proc. of the IEEE International Conference on Computer Communications (INFOCOM 2018), pp. 1817–1825, Honolulu, HI, USA, Apr. 2018.

[2] R. E. Bryant, “Graph-based Algorithms for Boolean Function Manipulation,” IEEE Transactions on Computers, Vol. C-35, No. 8, pp. 677–691, 1986.

Masaaki Nishino

Research Scientist, 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.

Takeru Inoue

Senior Researcher, NTT Network Innovation Laboratories.

He received a B.E., M.E., and Ph.D. from Kyoto University in 1998, 2000, and 2006. He was an ERATO (Exploratory Research for Advanced Technology) researcher at the Japan Science and Technology Agency (JST) from 2011 through 2013. His research interests widely cover the design and control of network systems. He received the Best Paper Award from the Asia-Pacific Conference on Communications in 2005 and research awards from the Institute of Electronics, Information and Communication Engineers (IEICE) Technical Committee on Information Networks in 2002, 2005, and 2012. He is a member of the Institute of Electrical and Electronics Engineers (IEEE).

Norihito Yasuda

Senior Researcher, NTT Communication Science Laboratories.

He received a B.A. in integrated human studies and an M.A. in human and environmental studies from Kyoto University in 1997 and 1999, and a D.Eng. in computational intelligence and system science from Tokyo Institute of Technology in 2011. He joined NTT in 1999. He also worked as a research associate professor with the Graduate School of Information Science and Technology, Hokkaido University, in 2015. His current research interests include discrete algorithms and natural language processing.

Shin-ichi Minato

Professor, Graduate School of Informatics, Kyoto University.

He received a B.E., M.E., and D.E. in information science from Kyoto University in 1988, 1990, and 1995. He worked in the NTT laboratories from 1990 until 2004. He was a visiting scholar in the Computer Science Department at Stanford University, USA, in 1997. He joined Hokkaido University as an associate professor in 2004 and was promoted to professor in October 2010. He moved to Kyoto University in 2018 (present position). He has also worked as a visiting professor at the National Institute of Informatics since 2015. He was a research director of the JST ERATO Minato Discrete Structure Manipulation System Project from 2009 to 2016. His research interests include efficient representations and manipulation algorithms for large-scale discrete structures such as Boolean functions, sets of combinations, sequences, and permutations. He is a senior member of IEICE and the Information Processing Society of Japan (IPSJ) and a member of IEEE and the Japanese Society for Artificial Intelligence (JSAI).

Masaaki Nagata

Senior Distinguished Researcher, Group Leader, NTT Communication Science Laboratories.

He received a B.E., M.E., and Ph.D. in information science from Kyoto University in 1985, 1987, and 1999. He joined NTT in 1987. His research interests include morphological analysis, named entity recognition, parsing, and machine translation. He is a member of IEICE, IPSJ, JSAI, the Association for Natural Language Processing, and the Association for Computational Linguistics.

Ukuzo—A Projection Mapping Technique to Give Illusory Depth Impressions to Two-dimensional Real Objects

Takahiro Kawabe

Abstract

Ukuzo is a light projection technique that gives illusory depth impressions to two-dimensional real objects by projecting cast shadow images onto them. The technique can not only give depth impressions but also manipulate material impressions of the object with the projected shadow patterns. The technique is promising for enhancing the expression in paper-based advertisements and visual arts.

Keywords: projection mapping, visual illusion, augmented reality

1. Introduction

Painters use a variety of techniques to elaborately express depth. For example, shadowing has been employed as an important technique to increase the realistic impression of objects in paintings. In medieval Europe, it was prohibited to use shadowing as a drawing technique in pictures for religious reasons. In the Renaissance period, Leonardo De Vinci, an eminent painter as well as scientist, reintroduced shadowing as a drawing technique and used it to enhance the realistic impression of his paintings [1]. Shadowing has also recently been used digitally in the graphical user interface of computers. For example, it is used to give illusory depth impressions to virtual objects shown on the computer display.

A vision science study [2] proposed an interesting illusion in which a shadow produces an illusory three-dimensional (3D) layout of an object. This is illustrated in the two pink squares shown in **Fig. 1**. The right square is perceived as being more separate from its background than the left one. In fact, the two squares have identical image information. The difference between them is the spatial relationship between the square and its shadow. The distance between the right square and its shadow is larger than the distance between the left one and its shadow. The previous study reported that as the distance between an object and a shadow increases, the object is perceived to apparently float up from its background to a greater extent [2].

Another study [3] on a light projection technique reported that light projection of a shadow motion can give wobbly motion impressions to a miniature car, indicating that the projection of a cast shadow can give illusory motion impressions to a real object.

We recently developed a light projection technique to give illusory depth perceptions that make it look as if an object in a 2D image was floating upward. This is caused by perceptual effects coming from the projected pattern of a shadow. We call this technique *Ukuzo*.

In this article, the hardware and software systems of Ukuzo are explained, and perceptual experiences that Ukuzo offers to observers are described. The role of light projection techniques in future applications of spatial augmented reality is also explained.

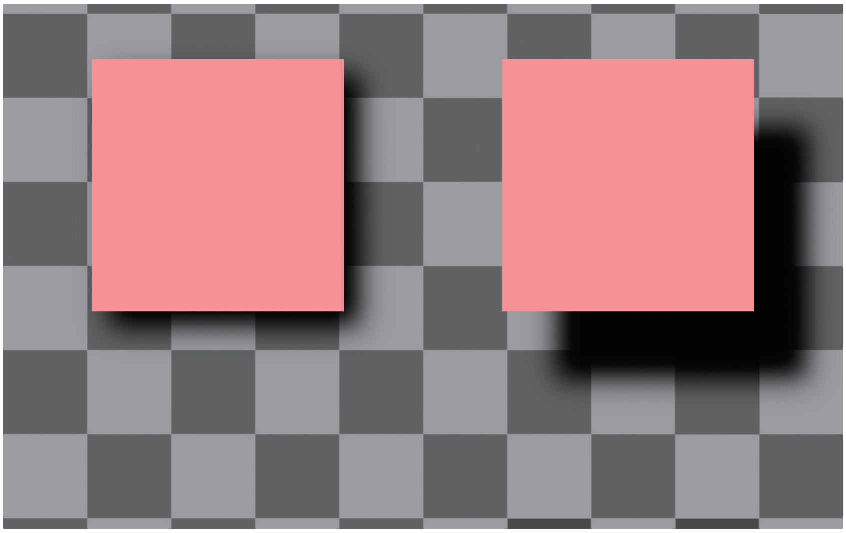


Fig. 1. Effect of a shadow on the depth of an object.

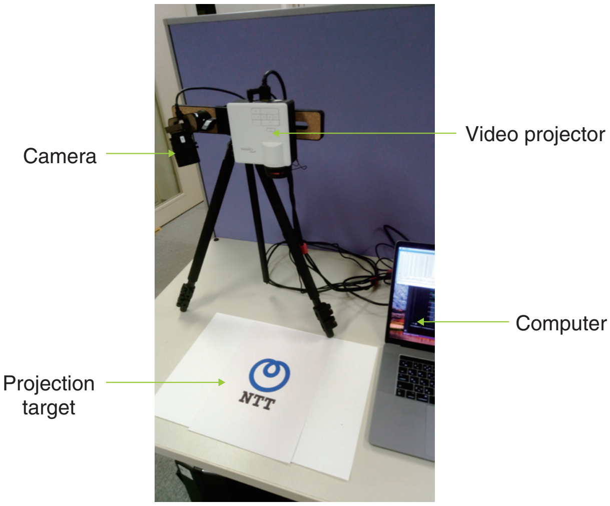


Fig. 2. Photograph of Ukuzo hardware components.

2. Ukuzo system

A photograph of the Ukuzo hardware system is shown in **Fig. 2**. A shadow pattern is emitted onto a projection target from a video projector. Since it is impossible to project black light, the shadow pattern we employed consists of a darker region serving as a shadow against a brighter background. A digital camera is used to capture the image of the projection target and to conduct geometric calibrations. We create a shadow pattern based on the captured image of the projection target. Finally, the created shadow pattern is projected to the spatial vicinity of the projection target so that it seems as if the projection target is casting the shadow pattern.

3. Perceptual effects of Ukuzo

The main visual effect of Ukuzo is to give illusory depth impressions to 2D real objects. A photograph of the Ukuzo effect is shown in **Fig. 3**. A dynamic loop—the NTT logo—is printed on a piece of paper, and Ukuzo conveys a cast shadow pattern to the dynamic loop. The viewer gets the impression that the dynamic loop is floating up from its original position. Moreover, the distance between the dynamic loop and the cast shadow pattern can be dynamically changed. In this way, Ukuzo can give the illusory perception that the dynamic loop is dynamically changing its depth positions. As the previous study [2] indicated, it is possible to change the apparent depth position by manipulating the degree of blurring of a shadow pattern.

Ukuzo can give not only depth impressions but also transparency impressions to an opaque paper object. A photograph of printed materials in which two disk-shaped areas with different colors are overlapping in the middle is shown in **Fig. 4(a)**. In **Fig. 4(b)**, Ukuzo projects a shadow pattern to the greenish disk shape on the left. When we view the effect shown in Fig. 4(b), the left disk appears to consist of a transparent greenish material such as glass or plastic. Similarly, in

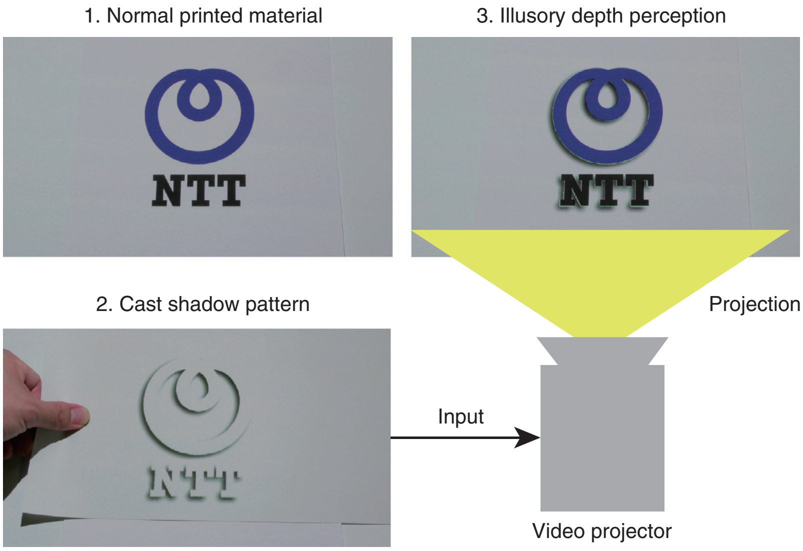
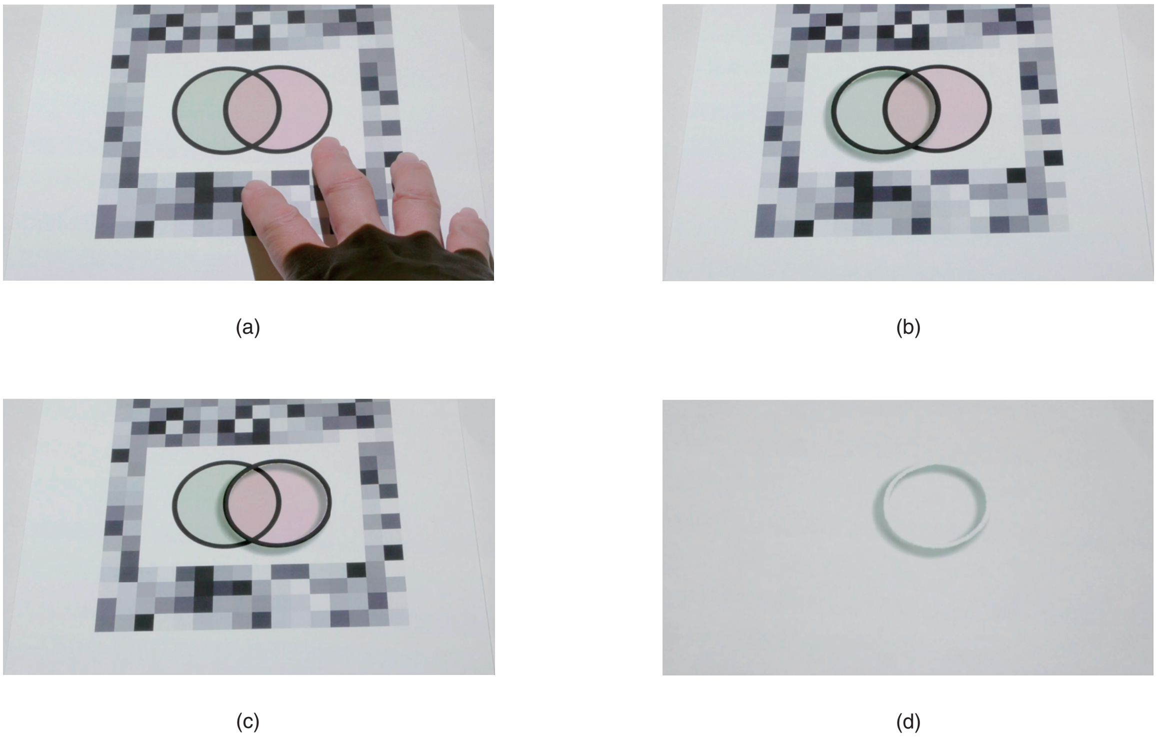


Fig. 3. A dynamic loop (NTT logo) with a cast shadow pattern projected by Ukuzo.


(a)

(b)

(c)

(d)

Fig. 4. (a) Photo of printed materials in which two disk shapes are overlapping. (b) The left disk has a cast shadow pattern projected by Ukuzo. (c) The right disk has a cast shadow pattern projected by Ukuzo. (d) The cast shadow pattern projected on the right disk shape.

Fig. 4(c), Ukuzo projects a shadow pattern to the reddish-colored disk on the right. This gives the effect that the right disk shape is made of a transparent material. The cast shadow pattern used in Fig. 4(c) is

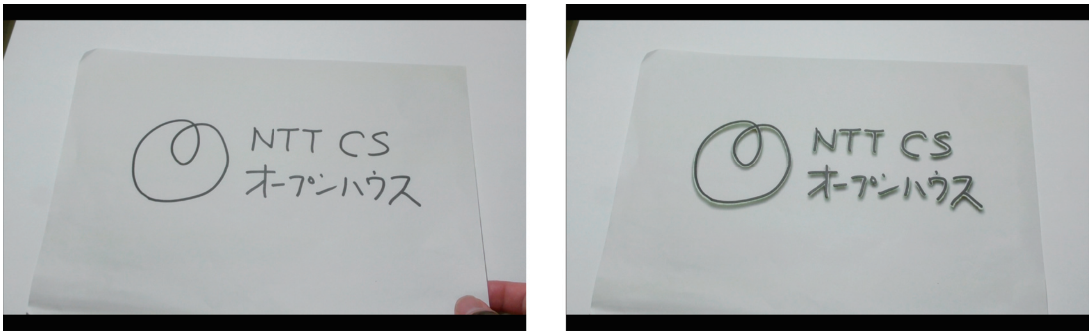


Fig. 5. (Left) Handwritten words. (Right) Cast shadow patterns are projected onto the handwritten words.

shown by itself in **Fig. 4(d)**. Such patterns can be projected onto 2D printed material such as that shown in Fig. 4(a). When the shape of the projected cast shadow pattern is elaborated, Ukuzo can give the impression that opaque paper materials are made of transparent materials such as glass or plastic.

Moreover, Ukuzo can give depth impressions to letters and illustrations that users draw. As shown in Fig. 2, the Ukuzo system has a camera to capture a projection target. The system captures the images of handwritten letters and illustrations, then generates the cast shadow pattern based on the captured images. When Ukuzo projects the cast shadow pattern onto the handwritten letters and illustrations, it gives illusory depth impressions to the handwritten materials (**Fig. 5**).

4. Future of Ukuzo

Ukuzo can be categorized as a kind of spatial augmented reality technique that is generally designed for changing the appearance of real objects by projecting digital images through a video projector. Spatial augmented reality techniques are usually intended to change the appearance of real objects in a physically correct manner. Specifically, if engineers want to change the appearance of a piece of paper from opaque to transparent, they need to calculate and project the desired light transport in order to display the appearance of a transparent sheet. However, in general, the light intensity of a typical projector is not sufficiently strong to achieve the physically correct light transport in ambient lighting.

In contrast, Ukuzo is not aimed at changing the appearance of real objects in a physically correct way. Rather, Ukuzo takes advantage of visual illusions wherein observers recognize a projected dark region as the shadow of an object and uses the shadow to estimate the 3D location of the object. Because it is possible for a standard projector to project the darker region in the vicinity of an object, it is feasible to use Ukuzo under normal ambient lighting.

Ukuzo can be used in various scenes. For example, Ukuzo can modify depth impressions of objects in paper posters that are displayed in stores and transport stations. Paper posters cannot usually be edited after printing. By making use of Ukuzo, designers can highlight the portions of posters on which they want customers to focus. Moreover, when a shadow pattern is projected toward an area of text on the poster that might be overlooked, it is possible to get customers to notice it. Furthermore, Ukuzo may be able to motivate children to draw pictures on the basis of interactions between hand-drawn pictures and illusory depth impressions that Ukuzo can give to pictures. That is, Ukuzo may help to support children's artistic capabilities. Adults can also enjoy the effects of Ukuzo when cast shadow patterns are conveyed to their calligraphy work or woodcut prints. That is, Ukuzo effects can be employed as a technique in future visual arts.

References

[1] F. Fiorani, "The Colors of Leonardo's Shadows," Leonardo, Vol. 41, No. 3, pp. 271–278, 2008.

[2] D. Kersten, D. C. Knill, P. Mamassian, and I. Bülthoff, "Illusory Motion from Shadows," Nature, Vol. 379, No. 6560, p. 31, 1996.

[3] R. Raskar, R. Ziegler, and T. Willwacher, "Cartoon Dioramas in Motion," Proc. of the 2nd International Symposium on Non-Photorealistic Animation and Rendering (NPAR 2002), pp. 284–300, Annecy, France, June 2002.

Takahiro Kawabe

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

He received a Doctor of Psychology from Kyushu University, Fukuoka, in 2005. In 2011, he joined NTT Communication Science Laboratories, where he studies applied aspects of human perception. He received the 2013 JPA Award for International Contributions to Psychology: Award for Distinguished Early and Middle Career Contributions from the Japanese Psychological Association. In 2018, he was also awarded the Young Scientists' Prize of the Commendation for Science and Technology by the Minister of Education, Culture, Sports, Science and Technology. He is a member of the Vision Sciences Society and the Vision Society of Japan.

Cross-media Scene Analysis: Estimating Objects' Visuals Only from Audio

Go Irie, Hirokazu Kameoka, Akisato Kimura, Kaoru Hiramatsu, and Kunio Kashino

Abstract

Human beings can get a visual image of the surrounding environment from sounds they hear. Can we give similar capabilities to computers? In this article, we introduce our recent efforts in cross-media scene analysis applied to estimate the type, location, and visual shape of objects in a scene based only on sound sources recorded with multiple microphones.

Keywords: cross media, scene analysis, deep learning

1. Introduction

The success of deep learning has completely changed the framework of media processing and recognition. Deep learning has already delivered superhuman performance that can be applied to address many problems in image and audio recognition. More important is that although media processing technologies for different types of media (e.g., images, video, audio, sound, and language) had at one time mostly been studied and developed independently, they are now being looked at together as their frameworks are similar.

In this article, we introduce cross-media scene analysis technology that can predict image recognition results only from sound information. As the number of people who are interested in security/safety increases day by day, the importance of surveillance and crime prevention technology is increasing. Most such technologies achieve excellent performance by leveraging advanced image recognition techniques. However, such technologies are probably not applicable in very dark places or rooms that have many blind spots, or in private or public spaces where privacy is prioritized and cameras are prohibited.

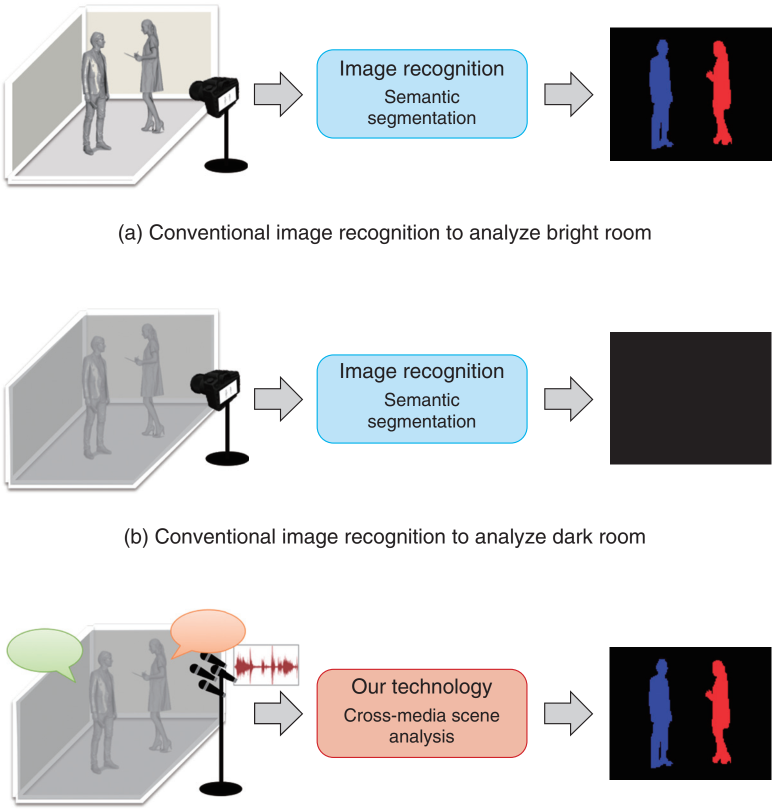
Our technology aims to provide visual recognition functionality without using any camera devices. This will make it possible to provide safe and comfortable monitoring even in those places where normal image recognition technologies cannot be used.

2. Cross-media scene analysis

Human beings recognize and understand their surrounding environments using their eyes and ears. More interestingly, we integrate and use these signals cross-sectionally to estimate the states of a scene. For example, when we are walking on a road and suddenly catch the sound of a car coming from behind us, we can infer how far away it is and possibly even what type of car it is without actually turning around and looking at the car.

What we aim to do here is to equip a computer with this kind of human ability. More specifically, the technology introduced in this article is designed to predict an image recognition result as if it had been photographed and recognized by a camera, using only the sound recorded by microphones.

An example of a use case of our technology is shown in **Fig. 1**. Suppose there are two people in a room and the objective is to automatically recognize them using a computer. The initial idea would be to



(a) Conventional image recognition to analyze bright room

(b) Conventional image recognition to analyze dark room

(c) Our cross-media scene analysis to analyze dark room

Fig. 1. (a, b) Conventional image recognition and (c) our cross-media scene analysis technology.

install a camera in the room and recognize them by using image recognition technology. If semantic segmentation technology is used, it is possible to detect not only the presence or absence of people, but also their locations and postures in silhouette form, as shown in Fig. 1(a). However, this approach would not work if the room was very dark or was a place where cameras are prohibited (Fig. 1(b)).

However, our technology can still be applied in such cases. Our technology uses audio information recorded by multiple microphones to deliver semantic segmentation results, instead of using a camera (Fig. 1(c)). If two people in the room are talking, their voices are recorded by the microphone arrays. With this sound information, our technology directly predicts the expected semantic segmentation result. To achieve this, we need to know (1) what kind of sound is coming from which direction, and (2) what kind of object (and its shape) is generating the sound. These details are respectively determined by signal processing and deep learning.

Imagine that a sound occurs at some location and is captured by multiple microphones. In this case, microphones closer to the source catch the sound earlier than the more distant ones. By analyzing this time difference of arrival, we can extract a directional feature that indicates the direction of the sound source. Furthermore, by analyzing the frequency information of the sound captured by each microphone, we can obtain a tonal feature that is useful to identify the type of the sound source (e.g., whether it sounds like a human voice or a train running on a track). With these features, we can determine the direction and the type of sound source. However, this information is not enough to recover a visual silhouette of the sound source representing the position and shape as compared to the semantic segmentation result.

Therefore, a deep neural network is used to estimate the type, shape, and position of the object. The overall setup is illustrated in **Fig. 2**. The neural network receives directional and tonal features as inputs

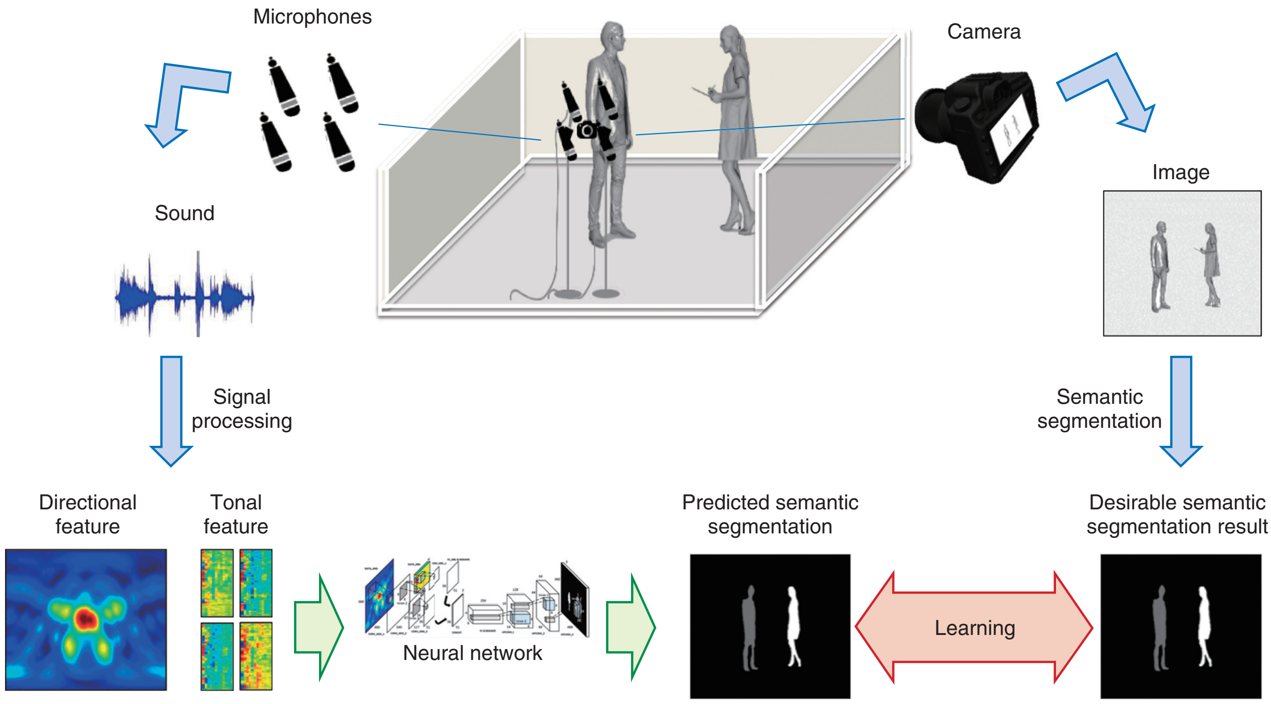


Fig. 2. Overview of cross-media scene analysis setup.

and is trained to output the semantic segmentation result directly from them. This is the core part of our technology that uses deep learning to convert media types from audio to visual. This is why we call our technology cross-media scene analysis. To train the network, we need to have a pair consisting of a desirable semantic segmentation result and corresponding audio features. Of course, a camera is needed to collect such data for training, but it is not necessary for actual recognition. With this process, we can build a basic mechanism to predict semantic segmentation results using only sound.

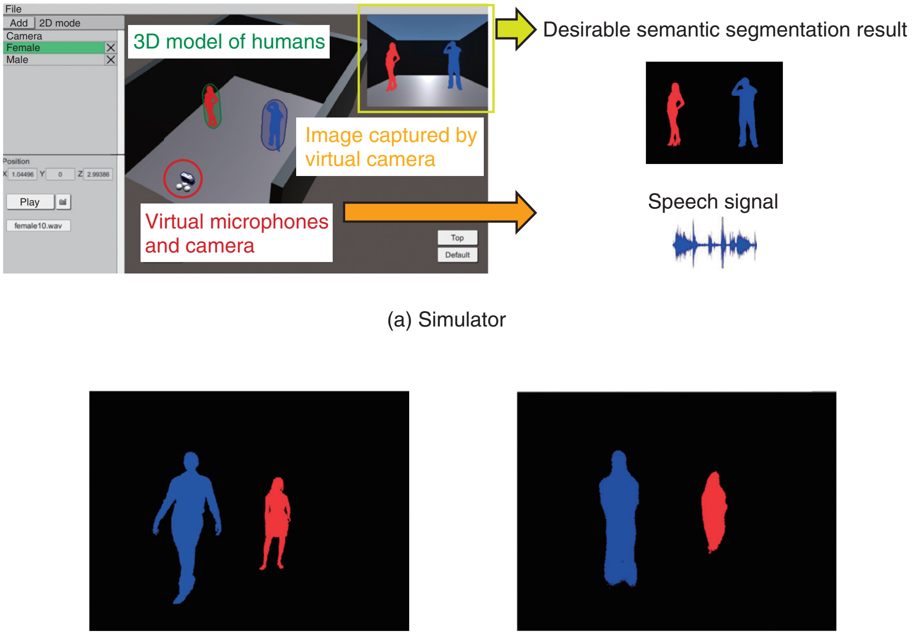
3. Proof of concept

We have conducted various experiments to evaluate how well our technology can predict actual semantic segmentation results from sound. We describe a few examples here.

First, we introduce an experiment using simulation data that was conducted to verify the feasibility of this technology in an ideal situation. We developed a simulator that can reconstruct a virtual room and generate conversational voices of people speaking and their corresponding semantic segmentation result simultaneously (**Fig. 3(a)**). We can freely change the size of the room, arrange people (three-dimensional models), the virtual microphones and camera, and simulate sounds recorded by the microphones taking the reverberation and echo of the room into account. At the same time, we can also simulate the image of the room taken by the camera and obtain the corresponding semantic segmentation result of the scene. Hence, we can train our neural network and measure how accurate the semantic segmentation result predicted by our technology is by comparing it to the desirable one.

We show an example of the desirable semantic segmentation result and the result predicted with our technology in **Fig. 3(b)** and **(c)**, respectively. Although the predicted result does not accurately recover details of posture and shape, we can see that the distance and rough shape can be successfully predicted by our technology.

Next, we describe an experiment done on real sounds. The task was to estimate the position and orientation of a toy train based on its running noise. Our evaluation setup is shown in **Fig. 4(a)**. It consisted of a toy train running on a circumferentially connected rail and four microphones connected to a personal computer in which our technology was installed. The entire setup was covered by a clear


File
Add
2D mode
Camera
Female
Male
Position
Play
female10.wav
Top
Default
3D model of humans
Image captured by virtual camera
Virtual microphones and camera
Desirable semantic segmentation result
Speech signal

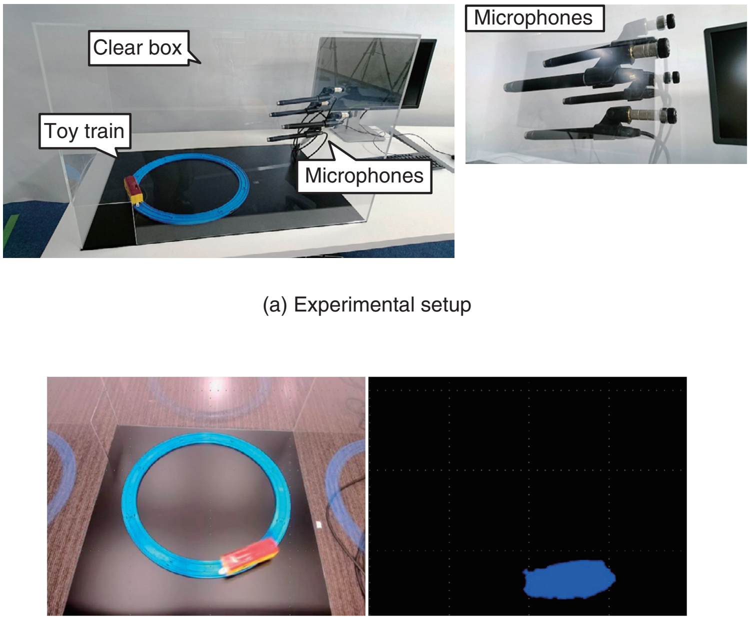
(a) Simulator

(b) Desirable semantic segmentation result

(c) Semantic segmentation result predicted by our technology

3D: three-dimensional

Fig. 3. Our simulator and predicted semantic segmentation result obtained by our technology.



(a) Experimental setup

(b) Actual position and orientation of the toy train

(c) Semantic segmentation result predicted by our technology

Fig. 4. Experimental setup and predicted results by our technology.

acrylic box. An example of the semantic segmentation result predicted by our technology is shown in **Fig. 4(c)**. The blue region represents the silhouette of the train. Compared to the position and orientation of the real train **(Fig. 4(b)**), we can see that the estimated position and orientation are roughly consistent with it, even though it is a little blurry. This shows that our technology can be applied to sounds other than human voices.

4. Future development

We will continue working to improve and demonstrate our technology toward application to a more natural space. To date, we have successfully verified the technology in structured environments such as those reconstructed by our simulator or the clear box. However, we need to make our method more robust in order to apply it in more realistic situations where more complex types of noise exist. The types of objects that can currently be recognized by our method are limited (people and toy trains), so we will work to expand those and test for more diverse categories. We will continue to improve our technology to achieve secure monitoring with arbitrary media information suitable for the situation.

Go Irie

Senior Research Engineer, Media Information Laboratory, NTT Communication Science Laboratories.

He received a B.E. and M.E. in systems engineering from Keio University, Kanagawa, in 2004 and 2006, and a Ph.D. in information science and technology from the University of Tokyo in 2011. He joined NTT in 2006 and has been studying multimedia content analysis, indexing, and retrieval. During 2011–2012, he was a visiting researcher at Columbia University, NY, USA. He is a member of the Institute of Electrical and Electronic Engineers (IEEE) and the Institute of Electronics, Information and Communication Engineers (IEICE).

Hirokazu Kameoka

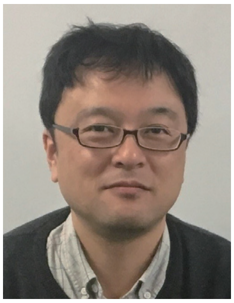
Senior Research Scientist, Media Information Laboratory, NTT Communication Science Laboratories.

He received a B.E., M.S., and Ph.D. from the University of Tokyo in 2002, 2004, and 2007. He is currently a Distinguished Researcher and a Senior Research Scientist with NTT Communication Science Laboratories, and an Adjunct Associate Professor with the National Institute of Informatics. From 2011 to 2016, he was an Adjunct Associate Professor with the University of Tokyo. His research interests include audio and speech processing and machine learning. He has been an Associate Editor for the IEEE/ACM Transactions on Audio, Speech, and Language Processing since 2015 and a member of the IEEE Audio and Acoustic Signal Processing Technical Committee since 2017.

Akisato Kimura

Senior Research Scientist, Research Planning Section, NTT Communication Science Laboratories.

He received a B.E., M.E., and D.E. in communications and integrated systems from Tokyo Institute of Technology in 1998, 2000, and 2007. He has been with NTT Communication Science Laboratories since 2007. He is engaged in work on pattern recognition, machine learning, and data mining. He is a board member of the Japanese Society for Artificial Intelligence (JSAI), a senior member of IEEE and IEICE, and a member of the Association for Computing Machinery (ACM) SIGMM (Special Interest Group on Multimedia) and SIGKDD (Special Interest Group on Knowledge Discovery and Data Mining).



Kaoru Hiramatsu

Senior Manager, NTT Geospace Corporation.*

He received a B.S. in electrical engineering and an M.S. in computer science from Keio University, Kanagawa, in 1994 and 1996, and a Ph.D. in informatics from Kyoto University in 2002. He joined NTT Communication Science Laboratories in 1996, where he worked on the Semantic Web, sensor networks, and media search technology. From 2003 to 2004, he was a visiting research scientist at the Maryland Information and Network Dynamics Laboratory, University of Maryland, USA. He is a member of the Information Processing Society of Japan (IPSJ) and JSAI.

*He was a Senior Research Scientist, Supervisor, and Leader of Recognition Research Group, Media Information Laboratory, NTT Communication Science Laboratories until June 2018.

Kunio Kashino

Senior Distinguished Researcher, Head of Media Information Laboratory, NTT Communication Science Laboratories.

He received a Ph.D. from the University of Tokyo for his pioneering work on music scene analysis in 1995. 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 (NII). He has been working on audio and video analysis, recognition, and search algorithms. He is a senior member of IEEE and IEICE, and a member of ACM.

Measuring, Understanding, and Cultivating Wellbeing in the Age of Technology

Junji Watanabe, Yuuki Ooishi, Shiro Kumano, Monica Perusquía-Hernández, Takashi G. Sato, Aiko Murata, and Ryoko Mugitani

Abstract

The development of information and communication technology (ICT) has brought efficiency and convenience to our daily lives. However, it has also been observed to have a negative impact on the emotional state of users. The concept of *design for wellbeing* is currently drawing a lot of attention. This approach seeks to explore how ICT can support human psychological wellbeing. In this article, we address several questions related to the issue of how to exploit technology and design to improve the wellbeing of humans.

Keywords: wellbeing, human science, self-tracking

1. Introduction

People can now access information anywhere and anytime as a result of the development of information and communication technology (ICT). While the efficiency of intellectual tasks has improved dramatically as a result, there is now concern about the negative impact of ICT on people’s mental state. For example, when users constantly pay attention to smartphone screens and alerts, they cannot relax their mind and body because they are forced to be in a psychological state of agitation. Also, because search engines present content that they surmise users want to see, there is the possibility that users will not come into contact with serendipitous information. In addition, social issues such as the high costs arising from addiction to social online games and bullying in private communication groups have also emerged.

Due to these trends, a paradigm shift in design that calls for technology to be not just efficient but also to contribute to wellbeing (WB) is taking place. Research on WB and experimental evidence that serve as the foundation of this design are being demanded [1]. In fact, WB is one of the critical Sustainable Development Goals (SDGs) for 2030 adopted by the United Nations in 2015. In the field of architecture, the WELL Certification for buildings has already been adopted. This is an environmental certification system that evaluates architectural spaces for how well they incorporate WB. Information technology (IT) companies such as Google practice mindfulness, and business magazines such as Time publish articles on mindfulness and WB. These trends suggest that the issue of WB has reached a level of general awareness.

2. What is WB?

WB means good physical, psychological, and social conditions. However, its exact definition differs depending on the field and purpose. WB is broadly divided into three types. The first is medical WB, in which medically sound functioning of the mind and body is seen as the foundation of our life.

Medical WB can be measured in health checkups that we receive regularly and in questionnaires on mental health. The second is hedonic WB. This concept considers WB as the subjective emotion of happiness at the moment. In general, when people are asked "Are you happy?", the answer tends to indicate a temporary feeling, which is hedonic WB. The third type is eudaimonic WB. Eudaimonic WB is defined as a state of being that can demonstrate the potential capabilities of mind and body, feeling significant, and leading an active life amid relationships with people around oneself. It is also expressed as *flourishing* [2].

From the perspective of eudaimonic WB, because achieving a *vibrant* state over a certain period of time is more important than momentary pleasure, WB in this definition is not necessarily limited to the idea that positive feelings are good and negative feelings are bad. For example, the process of accomplishing something may involve temporary hardship. However, if it brings about the feeling of accomplishment in oneself, it can be included in the process of eudaimonic WB. Thus, as seen above, there are several definitions of WB. Research on WB in recent years has not been limited to hedonic WB; rather, understanding eudaimonic WB through multidimensional factors such as subjective reports and physiological responses over a certain period of time is increasingly the objective.

Like *weather* and *the economy*, eudaimonic WB is a construct that can be understood by identifying specific factors. In the case of weather, quantitative evaluation of factors such as the temperature, level of humidity, barometric pressure, and wind speed can determine whether the weather is *good* or *bad*. Evaluation of WB also requires identification of factors and multidimensional evaluation. In general, factors that increase eudaimonic WB can be divided into several types (**Fig. 1**). Factors focusing on one's present state include engagement, mindfulness, and motivation. Furthermore, factors indicating acknowledgement of self-affirmation are also included. These include self-esteem, self-competence, and self-compassion. There are also factors not contained within the individual, but which involve other people for smooth interpersonal relationships. These include empathy and compassion. In addition, there are also factors that transcend the boundaries of the individual person, such as contributing to society.

Although the achievement of all or some of these factors can contribute to WB, which factors bring about an increase in WB depends on each person. WB also differs depending on one's life stage, such as before and after marriage. Factors of WB in the workplace and in the home are also different. It is therefore important to clarify the time period of the person and the situation when thinking about how to increase WB. It is not possible to create a panacea that benefits all. Rather, it is important to be aware of one's own WB and autonomously choose options to improve WB based on scientific evidence.

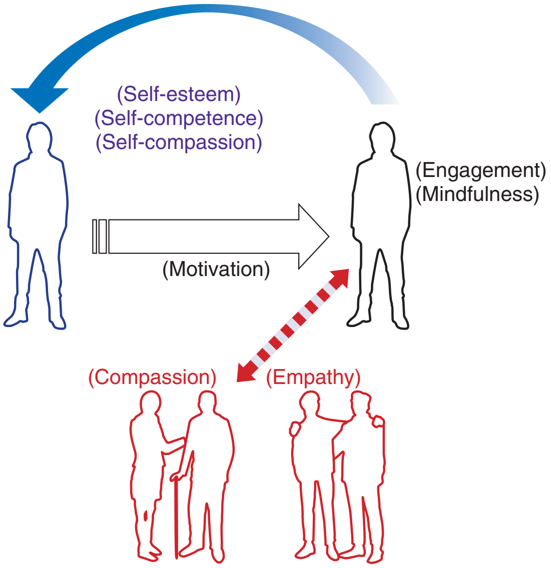


Fig. 1. Examples of factors of WB.

3. Measuring WB

In this section, we describe methodologies on how to measure factors related to eudaimonic WB. However, because there is not yet a unified methodology on what should be measured to evaluate eudaimonic WB and how, we give examples of current measurable phenomena and describe their possibilities. Phenomena that have been determined to be measurable factors of WB can be broadly divided into three categories. The first category consists of physiological responses such as changes in heart rate, respiratory rate, body temperature, and hormonal levels. Measuring these factors enables one to read the basic state of a person such as whether his or her body is in a state of nervousness or relaxation. Changes in the cardiac cycle in particular can be used as indicators to evaluate the activities of the autonomic nervous system's sympathetic nervous system (nervousness) and parasympathetic nervous system (relaxation). Also, a hormone called cortisol, which can be extracted from saliva and other bodily secretions, is used to measure stress.

The second category consists of behavioral information such as facial expressions, vocal intonation, and limb movements. When people are feeling positive, this is clear from their features such as the movements of particular facial expression muscles and the intonation of their voice.

The third category involves listening directly to descriptions. We can have a person express his or her conditions (subjective report), or ask others of their impressions of the person when viewed externally (third-person evaluation). Because the subjective report of the same event can differ if the person is asked to report how he or she feels at that point in time or if he or she is asked to reflect back on it after a certain period of time, care must be taken on how emotions are measured temporally. The content of the report may describe not only a person's bodily health and mental state, but also the subject's own feelings about the group or organization he or she belongs to.

WB in an organization is defined by replacing *the individual* with *the organization* in the definition of eudaimonic WB described above. It then refers to the state of an organization where resources can demonstrate their potential capabilities, carry out socially significant activities, and work actively in relationships with other organizations. It may be considered to be independent from WB for individuals. If that is the case, when we separate personal WB and organizational WB, it is possible to consider a balance of the two WBs and the factors contributing to this balance.

For eudaimonic WB, the factors need to be continually measured and comprehensively evaluated. What is critical is IT. At present, in addition to recording one's own subjective state, various technologies for self-tracking such as recording body temperature, heart rate, blood pressure, and sleep, are being commercialized. Using IT for self-tracking makes it possible to gain an accurate understanding of one's conditions based on data. Another advantage is the ability to compare one's own state at a certain time to the environmental conditions at that time. A simple example is the effect that a change in climate can have on one's state of health.

4. Research on WB at NTT Communication Science Laboratories

Because WB is determined by interpersonal relationships at a variety of levels such as family, friends, and organizations, it is necessary to analyze the mechanisms at each level when researching WB. At NTT Communication Science Laboratories, we have initiated research on WB and are measuring human physiological responses and elucidating empathy mechanisms in communication. We are establishing a foundation to explore evidence-based WB mechanisms in human beings and study design guidelines by integrating these research areas.

Specifically, in the area of human science, we are conducting research to scientifically understand the effects of experiences such as listening to music and practicing meditation, research on third-party evaluation of emotions to understand how a person's emotions are recognized by others, research on understanding the principles of psychological and physical bonds between mothers and their children, research on the effects that being together with others have on emotional response, and research on quantifying crowd sensations such as at an event or stadium by measuring group physiological responses simultaneously. Furthermore, we are conducting research on physiological response-linked tactile presentation technologies and research on intervention.

In this way, by engaging in WB research from a comprehensive and interdisciplinary perspective, we are pursuing the essence of WB. At the same time, we are engaged in design methodologies to realize conditions in which people and organizations can sustain vibrant activities (that is, WB). Specifically, we are creating indicators and developing measurement technologies for quantifying WB, suggesting product designs, service designs, and organizational designs, and studying how to contribute to sustaining interpersonal relationships (child-rearing and relationships in local communities) in a new technological environment.

References

[1] R. Calvo and D. Peters, "Positive Computing - Technology for Well-being and Human Potential," MIT Press, Cambridge, USA, 2014.

[2] M. E. P. Seligman, "Flourish - A Visionary New Understanding of Happiness and Well-being," Atria Books, New York, USA, 2012.

Junji Watanabe

Senior Research Scientist (Distinguished Researcher), Sensory Resonance Research Group, Human Information Science Laboratory, NTT Communication Science Laboratories.

He received a Ph.D. in information science and technology from the University of Tokyo in 2005. He presented his work at technology showcases, science museums, and art festivals.

Yuuki Ooishi

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

He received a B.S., M.S., and Ph.D. in physics from the University of Tokyo in 2003, 2005, and 2008. His research interest is the mechanism of neural dynamics, including auditory-evoked emotion, and autonomic nerve-neuroendocrinological interaction. He is a member of the Japan Neuroscience Society.

Shiro Kumano

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

He received a Ph.D. in information science and technology from the University of Tokyo in 2009 and joined NTT Communication Science Laboratories the same year. His research interests include affective computing, cognitive science and computer vision. He received the ACCV (Asian Conference on Computer Vision) 2007 Honorable Mention Award. He serves as an organizing committee member of International Conference on Affective Computing and Intelligent Interaction (ACII). He is a member of the Institute of Electrical and Electronics Engineers (IEEE) and the Institute of Electronics, Information and Communication Engineers (IEICE).

Monica Perusquía-Hernández

Research Associate, Sensory Resonance Research Group, Human Information Science Laboratory, NTT Communication Science Laboratories.

She received a B.Sc. in electronic systems engineering from Instituto Tecnológico y de Estudios Superiores de Monterrey in 2009, and an M.Sc. in human–technology interaction and a Professional Doctorate in engineering in user–system interaction from the Eindhoven University of Technology, Eindhoven, The Netherlands, in 2012 and 2014. Her research interests include affective computing and biosignal processing.

Takashi G. Sato

Senior Research Scientist, Moriya Research Laboratory, NTT Communication Science Laboratories.

He received a B.S., M.S., and Ph.D. in information science and technology from the University of Tokyo in 2003, 2005, and 2008. His research interests include human interfaces, especially brain-machine, audio, and tactile interfaces using psychophysiological measurements as feedback information. He is a member of IEEE, IEICE, the Acoustical Society of Japan, the Japan Neuroscience Society, and the Society of Instrument and Control Engineers.

Aiko Murata

Research Associate, Sensory Resonance Research Group, Human Information Science Laboratory, NTT Communication Science Laboratories.

She received a Ph.D. in social psychology from Hokkaido University in 2016. She joined NTT Communication Science Labs in 2018. Her research interests include investigating the mechanisms of human empathy.

Ryoko Mugitani

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

She received a B.A. in education in 1999, an M.A. in health science in 2001, and a Ph.D. in arts and sciences in 2004 from the University of Tokyo. She joined NTT Communication Science Labs in 2004. Her interests include development of communication and child wellbeing.

High-speed Avalanche Photodiodes toward 100-Gbit/s per Lambda Era

Masahiro Nada, Toshihide Yoshimatsu,
Fumito Nakajima, Hideaki Matsuzaki,
and Kimikazu Sano

Abstract

We have developed an avalanche photodiode (APD) in an effort to achieve 100-Gbit/s operation with a single wavelength (100-Gbit/s/lambda). The APD features a gap-grading layer and was set between the absorber and avalanche layers. It achieved an operating bandwidth of 42 GHz in low-gain conditions. An optical receiver made with the APD demonstrated 106-Gbit/s 4-level pulse amplitude modulation operation and 40-km optical amplifier-free transmission over a single-mode fiber under assumption of a KP4 forward error correction threshold.

Keywords: avalanche photodiode (APD), 100-Gbit/s PAM4, datacenter, optical receiver

1. Introduction

The interest in optical fiber communications systems in the last few decades has primarily been focused on long-haul systems capable of handling transcontinental communications and country-to-country communications. Such long-haul systems have included advanced optical devices used for optical communications as well as advanced modulation formats and detection mechanisms. One important example is digital coherent systems [1]. Coherent detection techniques used in combination with digital signal processors have achieved transmission distances of several thousand kilometers, and the development of densely integrated optical transmitters and receivers has enabled digital coherent systems to reach a transmission capacity of over 100 Gbit/s.

Recent interest in optical fiber communications systems has been focused on short-reach applications as represented by inter/intra-datacenter networking. In 2010, 100-Gbit/s Ethernet, which utilizes a bit rate per wavelength of 25 Gbit/s, was standardized for such purposes [2], and the required bit rate per wavelength is now approaching 100 Gbit/s (100-Gbit/s/lambda). Unlike digital coherent systems, such short-reach networks are required to be low-cost and thus simple in structure. Consequently, the network systems and optical components used in short-reach networks must achieve larger capacity and longer transmission distances by using simple direct detection mechanisms and high-speed optical components.

An avalanche photodiode (APD) is a key device to meet such requirements. Avalanche multiplication is initiated by electrons and holes in the APD, which provides a built-in first stage of gain of the electrical signals. Therefore, APDs obtain higher responsivity compared with the conventional PIN-PDs (positive-intrinsic-negative photodiodes), leading to higher sensitivity of the optical receivers. Thus, they effectively extend the transmission distance. To date, we have demonstrated 25-Gbit/s operation for NRZ (non-return-to zero) signals and 50-Gbit/s operation for 4-level pulse amplitude modulation (PAM4) signals by using the high-speed APDs we developed [3–5].

These high-speed APDs surpass the results of our previous examples and are aimed at achieving 100-Gbit/s/lambda operations. A gap-grading layer introduced between the absorber and the avalanche layers helped to boost the operation speed of the

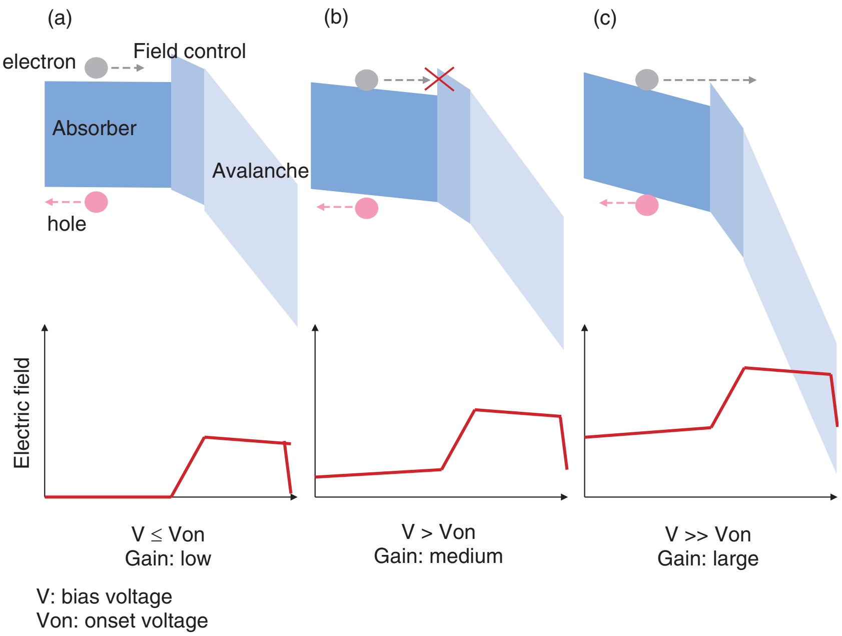


Fig. 1. Schematics of band diagrams of APD for various bias conditions.

APD. An optical receiver fabricated with the APD achieved 106-Gbit/s PAM4 transmission over a 40-km single-mode fiber without using an optical amplifier.

2. Device design and performance

Low-gain operation must be considered when trying to boost the speed of APDs. In principle, there is a strict trade-off between the gain and the speed of the APD according to the gain-bandwidth product (GBP) rule that is governed by the material and thickness of the avalanche layer; that is, the operation speed decreases as the gain increases. Thus, one simple way to obtain high-speed operation is to operate the APD with lower gain. We employ a 90-nm-thick indium aluminium arsenide (InAlAs) avalanche layer, which provides larger GBP as well as low excess noise, in our APD. However, one problem is that a large conduction band offset exists between the indium gallium arsenide (InGaAs) absorber and the InAlAs avalanche layer. This conduction band offset is as large as 0.5 eV, which can be problematic for low-gain and high-speed operation. A schematic band diagram illustrating this issue is shown in **Fig. 1**.

In Fig. 1(a), we show the band diagram around the absorber and the avalanche layer when the bias voltage (V) is lower than the onset voltage (Von). In this condition, the electric field in the absorber is zero, and the potential barrier caused by the conduction band offset remains. The photo-generated electrons and holes do not have drift components; thus, high-speed operation of the APD is not observed in this condition. Additional bias voltage depletes the p-type field control layer, and the electric field in the undoped absorber is invoked, as shown in Fig. 1(b). The photo-generated electrons and holes have drift components in the absorber. However, the remaining potential barrier originating from the conduction band offset prevents electrons from moving toward the avalanche layer. Consequently, high-speed operation cannot be obtained even under the V > Von condition depicted in Fig. 1(b).

To eliminate the potential barrier caused by the conduction band offset, further additional bias voltage is needed (Fig. 1(c)). The additional bias voltage effectively eliminates the potential barrier but simultaneously strengthens the electric field in the avalanche layer, resulting in the increased gain of the APD. Consequently, the speed of the APD is lowered by the GBP limitation.

A gap-grading layer is known to effectively relax the potential barrier [6]. We introduced a 40-nm-thick 1.1-eV indium aluminium gallium arsenide (InAlGaAs) gap-grading layer between the InGaAs absorber and InAlAs avalanche layers to relax the

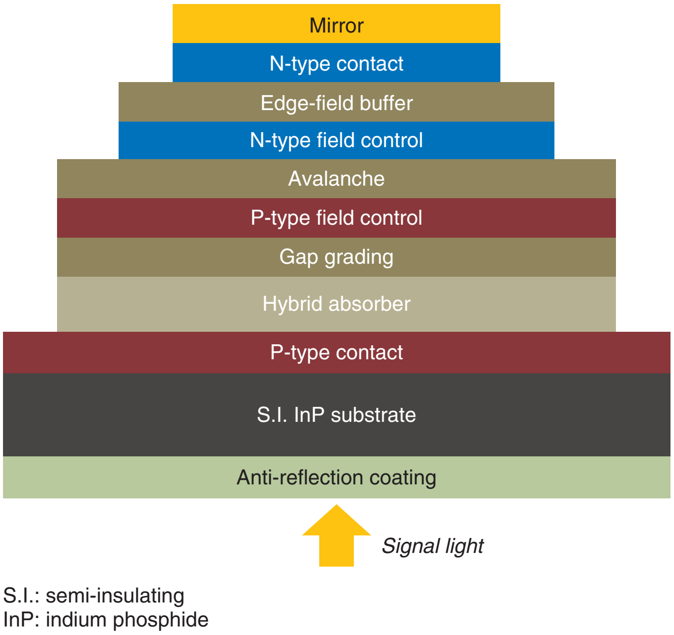


Fig. 2. Schematic cross-sectional view of the fabricated APD.

potential barrier that prevents the electron transport from the absorber to the avalanche layer under lower bias conditions, which can boost the speed of the APD.

A cross-sectional view of the APD we developed for 100-Gbit/s/lambda (100-Gbit/s PAM4) operation is shown in **Fig. 2**. The APD is designed based on an inverted p-down structure [7]. The epitaxial layers, including the p-type contact, p-type absorber, undoped absorber, gap-grading, p-type field control, avalanche, n-type field control, edge-field buffer, and n-type contact layers are grown on a semi-insulating InP (indium phosphide) substrate by metal-organic chemical vapor deposition. After the epitaxial growth, we formed the triple-mesa structure shown in Fig. 2 using a wet-etching technique. Then we formed the electrodes and mirror metal by electron beam evaporation. The anti-reflection film is coated after the front-end process. Thus, the APD has a back-side illumination structure.

As described above, we employed a 90-nm-thick InAlAs avalanche layer. We also used a hybrid absorber consisting of 150-nm p-type and 150-nm undoped InGaAs to reduce the transit time of the electrons and holes [8].

The current-voltage (I-V) characteristics of the fabricated APD with an active diameter of 14 µm are shown in **Fig. 3**. The photocurrent rapidly rises from 5.5 V to 7.5 V and then increases gradually toward a breakdown voltage (Vb) of 25.4 V due to the avalanche gain. The dark current does not show any unexpected increase, indicating that the APD has no edge breakdown or other unexpected breakdown. The estimated responsivity at unit gain is 0.5 A/W against 1300-nm-wavelength optical input.

The frequency characteristics of the fabricated APD are shown in **Fig. 4**. A maximum bandwidth of 42 GHz was observed with a gain as low as 1.5 thanks to the gap-grading layer in the APD. A bandwidth of over 30 GHz, which is sufficient for 100-Gbit/s PAM4 operation, was maintained for the increase in gain to 3.

3. Receiver characteristics

To demonstrate the 100-Gbit/s PAM4 operation of the APD, we mounted the fabricated APD into a butterfly package together with a transimpedance amplifier and conducted a transmission test with the fabricated APD receiver. The transimpedance amplifier had a bandwidth of 33 GHz, which is sufficient for 100-Gbit/s PAM4 operation. The package has a GPPO electrical output. The optical transmitter we used consisted of a 1309.49-nm-wavelength electro-absorption modulator integrated DFB (distributed feedback) laser (EML) with a launch power of +3.5

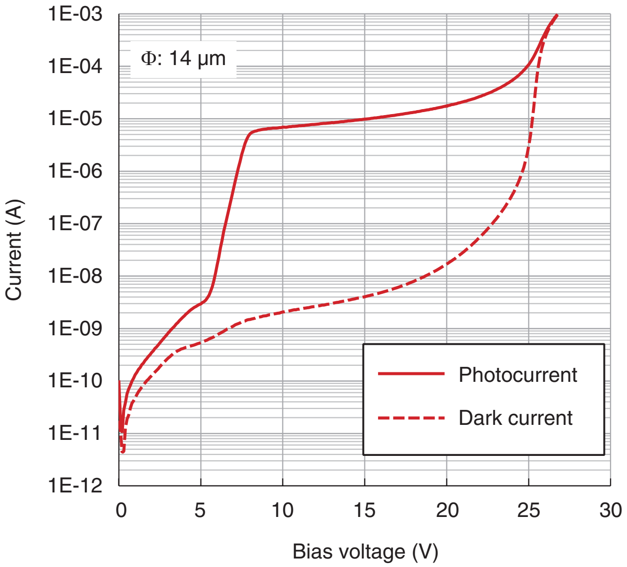


Fig. 3. I-V characteristics of the fabricated APD with an active diameter of 14 μm.

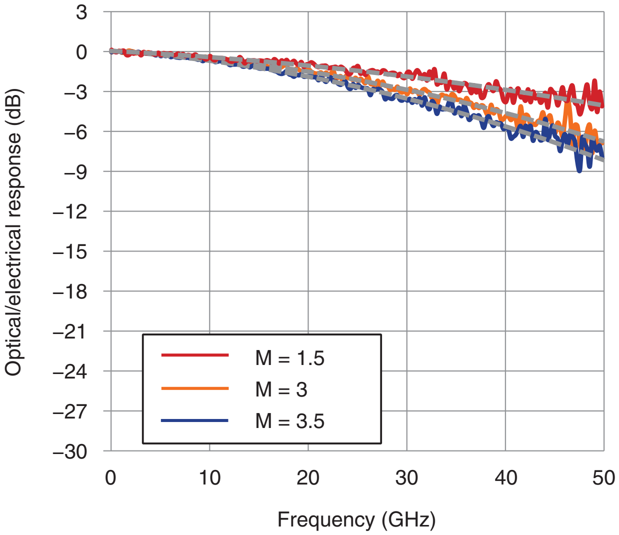
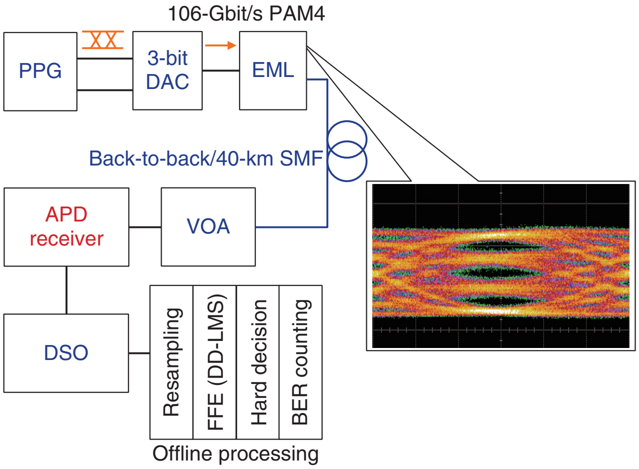


Fig. 4. Frequency characteristics of the fabricated APD.

dBm [9]. The bit rate of the electrical signal that drives the EML was set to be 106-Gbit/s PAM4 by considering the overhead of the forward error correction (FEC). The 106-Gbit/s PAM4, PRBS (pseudo-random binary sequence) signal of 2^{15}-1 was generated with a 3-bit DAC (digital-to-analog converter). The optical output signal from the EML had an extinction ratio at 106-Gbit/s PAM4 operation of about 7 dB. The electrical output signal was captured by a real-time DSO (digital storage oscilloscope) with a 160-GS/s sampling rate and 62-GHz bandwidth.

The transmission tests were performed with offline processing using feed-forward equalization (FFE) with a half-symbol-spaced (T/2-spaced) adaptive equalizer. No pre-emphasis on the transmitter was used. The received power was determined by the VOA (variable optical attenuator) set in front of the receiver. For the 40-km transmission test, we used single-mode fiber (SMF) with a loss of 0.33 dB/km. The measurement setup for the characterization of the APD optical receiver is shown in **Fig. 5**.

The bit error rate (BER) characteristics of the fabricated APD receiver under back-to-back conditions and after 40-km transmission are presented in **Fig. 6(a)**. The number of taps of the FFE was set to be 17 for both conditions. In **Fig. 6(b)** and **(c)**, we respectively show the equalized electrical eye diagrams of the APD receiver for back-to-back and after 40-km transmission conditions. The BER characteristics in Fig. 6(a) indicate that the fabricated APD receiver successfully demonstrated 106-Gbit/s PAM4 operation even for the 40-km transmission over SMF without an optical amplifier, as well as in the back-to-back condition. The 40-km transmission was achieved with an average received power ($P_{avg.}$) of –11.47 dBm at a KP4-FEC limit. The power penalty against the back-to-back condition was 0.32 dB. The eye diagram obtained after 40-km transmission



BER: bit error rate
DD-LMS: decision-directed least mean square
PPG: pulse-pattern generator

Fig. 5. Experimental setup for 100-Gbit/s PAM4 bit error rate test utilizing APD receiver.

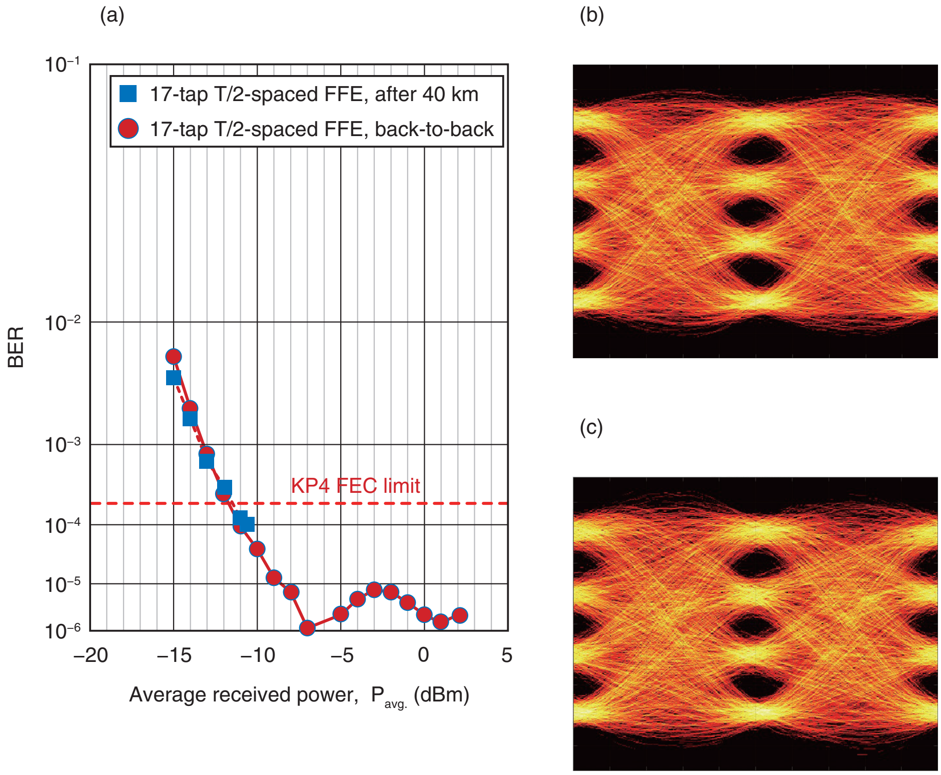


Fig. 6. (a) BER characteristics of the fabricated APD receiver for the back-to-back condition and after 40-km transmission. The T/2-FFE is set to 17 taps. (b) Electrical eye diagram of the APD receiver at back-to-back condition and (c) after 40-km transmission.

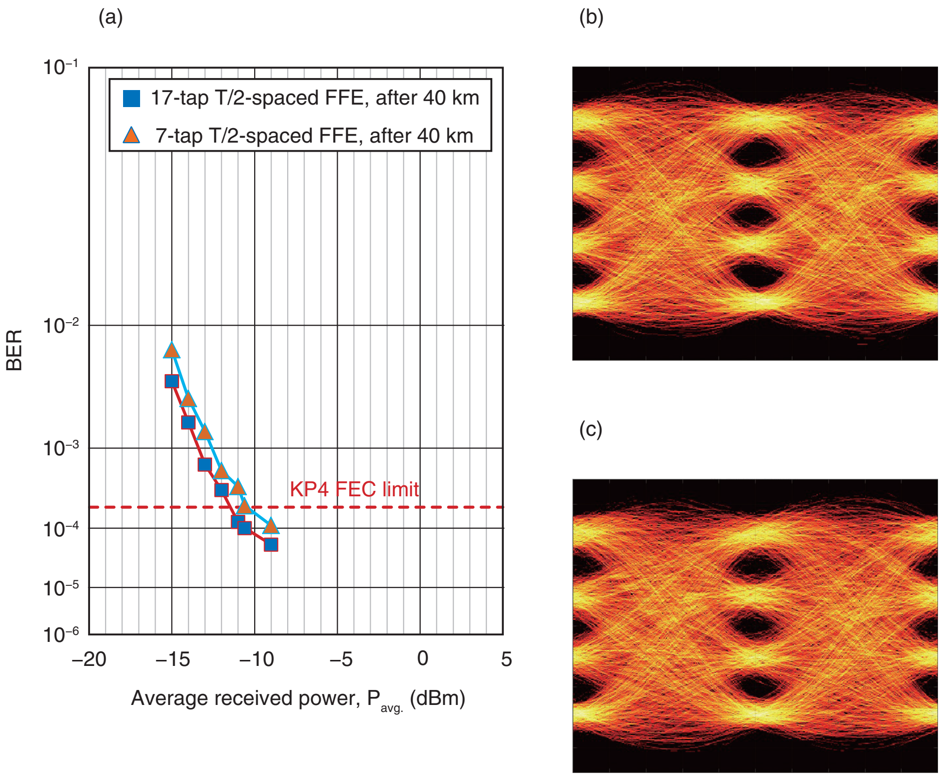


Fig. 7. (a) BER characteristics of the fabricated APD receiver after 40-km transmission for 17-tap and 7-tap conditions, (b) electrical eye diagram of the APD receiver after 40-km transmission with 17 taps, and (c) with 7 taps.

shows no degradation from that of the back-to-back condition.

A smaller number of FFE taps is advantageous in terms of power consumption and latency by signal processing when considering the digital-signal processing used in practical optical transceivers. The BER characteristics of the APD receiver for 17-tap and 7-tap FFE after 40-km transmission are shown in **Fig. 7(a)**. A minimum receiver sensitivity as low as –10.57 dBm was obtained, even with the 7-tap FFE. The equalized eye diagram of the APD receiver with 7-tap FFE is presented in **Fig. 7(b)**. No significant degradation from the 17-tap condition is apparent, which indicates that the 7-tap FFE is also applicable for 40-km transmission without an optical amplifier using the fabricated APD.

4. Conclusion

We developed a high-speed APD designed for use in the 100-Gbit/s per wavelength era. We introduced an optimized gap-grading layer consisting of 1.1-eV InAlGaAs in order to achieve low-gain, high-speed operation. Thanks to the gap-grading layer, the APD exhibited a peak bandwidth of 42 GHz with a responsivity at unit gain of 0.5 A/W. A 30-GHz bandwidth was maintained for a gain of 3. The optical receiver fabricated with the APD successfully achieved 40-km optical-amplifier-free transmission for 106-Gbit/s PAM4 optical signals. These results suggest that the 100-Gbit/s/lambda era can be realized even when extending the transmission distance by using low-power-consumption, cost-effective optical components that high-speed APDs provide.

References

[1] K. Kikuchi, "Digital Coherent Optical Communication Systems: Fundamentals and Future Prospects," IEICE Electronics Express, Vol. 8, No. 20, pp. 1642–1662, 2011.

[2] IEEE P802.3ba 100 Gb/s Ethernet Task Force, http://www.ieee802.org/3/ba/

[3] T. Yoshimatsu, M. Nada, M. Oguma, H. Yokoyama, T. Ohno, Y. Doi, I. Ogawa, and E. Yoshida, "Compact and High-sensitivity 100-Gb/s (4 × 25 Gb/s) APD-ROSA with a LAN-WDM PLC Demultiplexer," Proc. of the 38th European Conference on Optical Communication (ECOC 2012), Th.3.B.5, Amsterdam, The Netherlands, Sept. 2012.

[4] F. Nakajima, M. Nada, and T. Yoshimastu, "High-speed Avalanche Photodiode and High-sensitivity Receiver Optical Subassembly for

100-Gb/s Ethernet," J. Lightw. Technol., Vol. 34, No. 2, pp. 243–248, 2016.

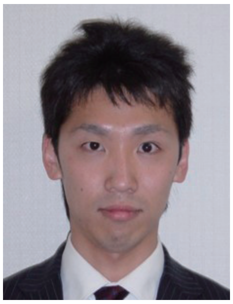
[5] Y. Nakanishi, T. Ohno, T. Yoshimatsu, Y. Doi, F. Nakajima, Y. Muramoto, and H. Sanjoh, "4 x 28 Gbaud PAM4 Integrated ROSA with High-sensitivity APD," Proc. of the 20th Opto-Electronics and Communications Conference (OECC 2015), post-deadline paper, Shanghai, China, June/July 2015.

[6] J. C. Campbell, "Recent Advances in Telecommunications Avalanche Photodiodes," J. Lightw. Technol., Vol. 25, No. 1, pp. 109–121, Jan. 2007.

[7] M. Nada, Y. Muramoto, H. Yokoyama, T. Ishibashi, and H. Matsuzaki, "Triple-mesa Avalanche Photodiode with Inverted P-Down Structure for Reliability and Stability," J. Lightw. Technol., Vol. 32, No. 8, pp. 1543–1548, Apr. 2014.

[8] M. Nada, T. Yoshimatsu, Y. Muramoto, H. Yokoyama, and H. Matsuzaki, "Design and Performance of High-speed Avalanche Photodiodes for 100-Gb/s Systems and Beyond," J. Lightw. Technol., Vol. 33, No. 5, pp. 984–990, 2015.

[9] S. Kanazawa, S. Tsunashima, Y. Nakanishi, Y. Muramoto, H. Yamazaki, Y. Ueda, W. Kobayashi, H. Ishii, and H. Sanjoh, "Equalizer-free 2-km SMF Transmission of 106-Gbit/s 4-PAM Signal Using Optical Transmitter/Receiver with 50 GHz Bandwidth," Proc. of the 21st Opto-Electronics and Communications Conference (OECC 2016), ThD1-2, Niigata, Japan, July 2016.



Masahiro Nada

Research Engineer, NTT Device Innovation Center.

He received an M.E. in applied physics from the University of Electro-Communications, Tokyo, in 2009 and a Ph.D. in engineering from the University of Tokyo in 2017. He joined NTT Photonics Laboratories in 2009 and has since been researching and developing high-speed, high-responsivity APDs and their applications in optical fiber communications systems. Dr. Nada received the Young Researcher's Award from the Institute of Electronics, Information and Communication Engineers (IEICE) in 2014. He is a member of IEICE, the Optical Society (OSA), and the Institute of Electrical and Electronics Engineers (IEEE).

Toshihide Yoshimatsu

Senior Research Engineer, NTT Device Innovation Center.

He received a B.E. and M.E. in applied physics from Tohoku University, Miyagi, in 1998 and 2000. He joined NTT Photonics Laboratories in 2000. He is involved in researching and developing ultrafast opto-electronic devices. He received the International Conference on Solid State Devices and Materials (SSDM) Paper Award in 2004. He is a member of IEICE and the Japan Society of Applied Physics (JSAP).

Fumito Nakajima

Senior Research Engineer, NTT Device Technology Laboratories.

He received a B.E., M.E., and Ph.D. from Hokkaido University in 1998, 2000, and 2003. He joined NTT Photonics Laboratories in 2003. He is involved in the research and development (R&D) of high-speed photodiodes and APDs for optical receivers. He received the Young Researcher's Award in 2006 and the Achievement Award in 2018 from IEICE. Dr. Nakajima is a member of IEICE.

Hideaki Matsuzaki

Senior Research Engineer, Supervisor, NTT Device Technology Laboratories.

He received a B.S. and M.S. in physics from Kyoto University in 1993 and 1995. He joined NTT's Atsugi Electrical Communications Laboratories in 1995. He is currently engaged in R&D of compound semiconductor devices, photodiodes, and laser diodes. He is a member of IEICE.

Kimikazu Sano

Senior Research Engineer, Supervisor, NTT Device Innovation Center.

He received a B.S., M.S., and Ph.D. in electrical engineering from Waseda University, Tokyo, in 1994, 1996, and 2004. He joined NTT in 1996. Since then, he has been designing and evaluating ultrafast integrated circuits (ICs) and optoelectronic ICs. From 2005 to 2006, he was a visiting researcher at the University of California, Los Angeles (UCLA), USA, where he researched a microwave/millimeter-wave sensing system. He was with NTT Electronics from 2012 to 2014, where he developed high-speed analog ICs and packaged modules for coherent optical systems. He is currently developing lasers, photodiodes, and analog ICs for optical metro-access networks. He served as a member of the Technical Program Committee for the IEEE Compound Semiconductor IC Symposium (CSICS) from 2008 to 2010.

Telecom Infra Project—Its Structure and Activities

Takeshi Kinoshita, Hitoshi Masutani, and Katsuhiro Shimano

Abstract

Telecom Infra Project (TIP) was launched in February 2016 through the initiative of Facebook. It is aimed at bringing about innovations in telecom hardware, software, and operations by introducing such technical trends as openness and disaggregation, which have become increasingly important in recent years. TIP is similar to open source projects in that communities play an essential role.

Keywords: TIP, openness, disaggregation

1. Background

In the technological fields involving datacenters and datacenter interconnections, de facto standards are widely used. They come from products and specifications developed by organizations such as the Open Networking Foundation as well as open source projects. What is characteristic is that these standards are developed through making the designs and interfaces open. In many cases, the resulting products are also open source. The trend of *openness* is now seen in not only the development of software but also of hardware. For example, the Open Compute Project (OCP) pushed through openness in the specifications of servers and racks used in datacenters, leading to a reduction in the purchasing cost as well as in the power consumption. In this approach, functions and components were unbundled. It is this *disaggregation* that made it possible for users to choose only the parts necessary to them, eliminating others.

The trend of openness and disaggregation is now beginning to expand to telecom networks. Telecom Infra Project (TIP) was established with the leadership of Facebook, one of the leading companies of OCP, and with the idea that the same approach as OCP's could be applied in telecom networks.

2. Overview of TIP

Since TIP started its activities in February 2016, the number of member companies has increased from several dozen to over 500. NTT joined TIP in 2017. The Board of Directors includes people from telecom operators such as Deutsche Telekom (DT) and SK Telecom, which have held the position since the foundation of the organization, and from Telefónica, Vodafone, and British Telecom (BT). This demonstrates that European operators are actively involved in the activities. Many operators from other regions such as India, South Africa, and Brazil are also participating. For these operators whose service areas include places where communications infrastructure has not been well established, the open-sourced products and the know-how arising from TIP would serve as a solution to address the issue adequately and cost-effectively.

Manufacturers of white box equipment, which use merchant chips and non-proprietary operating systems, and chip vendors are among those who have expanded their presence in concert with the markets of these products. There are also other kinds of participants such as Internet service providers and system integrators, including startup companies. There is a great deal of interest in the latest technical trends and expectations for their implementation among the wide variety of participants.

The organizational structure of TIP is shown in **Fig. 1**. Project Groups (PGs), responsible for creating specifications or developing products, are established after their charters are accepted by the Board of Directors and the Technical Committee. Although some PGs have been dissolved, the number of PGs has increased since TIP was founded, showing flexibility similar to that of other organizations in the field of datacenter technologies.

PGs are classified by their scopes into three technical fields: Access Projects, Backhaul Projects, and Core and Management Projects. The scopes of individual PGs are written in their charters along with goals and deliverables. Products developed by a PG are usually tested in a Community Lab. There are currently six Community Labs, two in the USA and one each in Germany, South Korea, Brazil, and India. Each Community Lab is hosted and operated by a specific company with a policy of allowing the participation of member companies. As is evident in many successful open source projects, creating open communities is recognized as a key to advancing development.

TIP Ecosystem Acceleration Centers (TEACs) provide startup companies with an environment for product development and tests. Currently, BT, DT, Orange, and SK Telecom each host a TEAC.

3. Access Projects

The activities of the PGs whose scope is related to the Access field are listed in **Table 1** [1]. These PGs aim to make Internet access easier by innovating access network technologies. Reducing infrastructure costs and creating new technologies that are applicable where existing ones cannot be easily used are among the examples of the PGs' goals.

The OpenCellular PG is working to produce base stations of Long-Term Evolution (LTE) cellular networks. While open source software is employed for the management of the base stations, chips and other hardware are developed by the participant companies, and their design has been made public as open source hardware. This PG is focusing on the existing mobile technology, rather than the emerging fifth-generation (5G) mobile technology, and it has brought openness into its products, representing the TIP's concept. The products have been used in field trials in India, Pakistan, and other African and South American countries.

The activities of many of the other Access PGs are also related to mobile communications infrastructure.

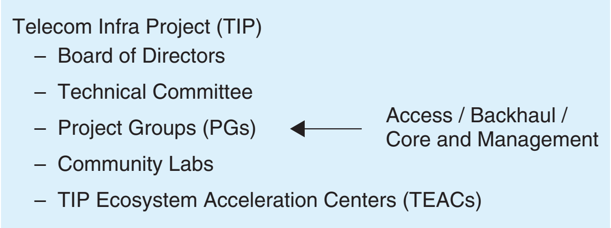

Telecom Infra Project (TIP)
– Board of Directors
– Technical Committee
– Project Groups (PGs)
– Community Labs
– TIP Ecosystem Acceleration Centers (TEACs)
Access / Backhaul / Core and Management

Fig. 1. Organizational structure of TIP.

The vRAN Fronthaul PG is tackling issues on the mobile fronthaul, which is responsible for the connection between antennas and baseband signal processing units of base stations. In developing countries, optical fibers cannot always be used as transmission media due to costs or environmental problems. The PG is addressing this issue by studying how to implement fronthaul transmission using twisted pair cables, coaxial cables, or other non-ideal media.

A power supply can also be an impediment in deploying infrastructure that costs or environmental problems impose. The Power and Connectivity PG aims to establish technologies and business models to deploy power supply infrastructure in rural areas with lower costs, inviting power companies and battery makers as members.

The OpenRAN PG's scope is similar to that of the OpenCellular PG's in that both are aimed at developing LTE base stations. Their focus, however, is on software implementation and its programmability. Because the functionalities are programmed to work with general purpose processors, flexible and timely functional updates are possible.

4. Backhaul Projects

A rapid increase in the amount of communications traffic is being seen in many countries and is a common issue for telecom operators worldwide. Against this background, PGs related to the Backhaul field are trying to innovate technologies in trunk and backbone networks. The activities of the PGs focusing on this field are summarized in **Table 2** [1].

The Open Optical Packet Transport PG is targeting innovative optical fiber transmission techniques. They have released two equipment prototypes named Voyager and Cassini. As well as being optical transponders with integrated packet switching functions, they are characterized by the use of nonproprietary

Table 1. Access Projects.

Project Group	Scope and Targets*	Chair
Edge Computing	The Edge Computing Project Group will focus on lab and field implementations for services/applications at the network edge, leveraging open architecture, libraries, software stacks, and mobile edge computing (MEC), into a platform.	Intel Telefónica
Power and Connectivity	The Power and Connectivity Project Group will initially focus on enabling network operators to deploy connectivity in areas that do not have electricity. It will also provide an ecosystem in which connectivity and electricity providers can collaborate to pilot and scale innovative technology and business models.	Telefónica Facebook
System Integration and Site Optimization	The System Integration and Site Optimization Project Group will address system integration via innovative, cost-effective and efficient end-to-end solutions in order to serve both rural and urban regions in optimal and profitable ways.	Deloitte Facebook
OpenCellular	The OpenCellular Project Group will focus on wireless access platforms (including cellular) and related technologies. It will develop solutions running on the OpenCellular platform, and support an ecosystem of contributors, OEMs, distributors and system integrators.	Facebook
Solutions Integration	The Solutions Integration Project Group will develop an open radio access network (RAN) architecture by defining open interfaces between internal components and focusing on the lab activity with various companies for multi-vendor interoperability.	SK Telecom
OpenRAN	The OpenRAN Project Group's main objective is the development of fully programmable RAN solutions based on general purpose processing platforms (GPPPs) and disaggregated software so they can benefit from the flexibility and faster pace of innovation capable with software-driven development.	Vodafone Intel
CrowdCell	The CrowdCell Project Group will focus on creating a CrowdCell by leveraging GPPPs, software-defined radio and open source designs for both hardware and software to minimize costs with a "one design" flexible platform.	Vodafone
vRAN Fronthaul	The vRAN Fronthaul Project Group will focus on virtualization of the RAN for non-ideal backhaul, in particular, maximizing the performance through optimization of the physical layer, compression, and other methods.	BT Vodafone

* Excerpts from the TIP website

OEM: original equipment manufacturer

Table 2. Backhaul Projects.

Project Group	Scope and Targets*	Chair
Millimeter Wave Network	The Millimeter Wave Network Project Group will define and advance 60-GHz wireless networking solutions to address the growing demand for bandwidth in dense, highly populated cities by delivering gigabits of capacity more quickly, easily, and cost-effectively compared to deploying fiber.	Facebook DT
Open Optical Packet Transport	The Open Optical Packet Transport Project Group will define dense wavelength division multiplexing (DWDM) open packet transport architecture that triggers new innovation and avoids implementation lock-ins. Open DWDM systems include open line system & control, transponder and network management, and packet-switch and router technologies.	Facebook

* Excerpts from the TIP website

components. In Cassini, for example, a merchant digital signal processor is employed to implement coherent optical transmission, which NTT's laboratory has developed with its partners and is now sold by NTT Electronics. The operating system employed in it is also a commercial, nonproprietary one.

This type of equipment, referred to as white box, enables multi-vendor implementation in such a way as to enable specific parts to be replaced with those of another vendor. Although the usage of Voyager or Cassini seems to be limited to a point-to-point interconnection between datacenters, the trend of disaggregation that they embody is likely to penetrate into more complicated optical transmission equipment.

In the field of wireless transmission, the Millimeter Wave Network PG is working to establish an infrastructure technology using 60-GHz radio waves. The frequency falls in an unlicensed band in many countries, including Japan and other Asian countries, Europe, and North and South American countries, reducing hurdles for its use. The wireless infrastructure can also be installed more easily than optical

Table 3. Core and Management Projects.

Project Group	Scope and Targets*	Chair
Artificial Intelligence and Applied Machine Learning	The Artificial Intelligence and Applied Machine Learning Project Group will define and share reusable, proven practices, recipes, models, and technical requirements for applying artificial intelligence and machine learning to reduce the cost to plan and operate telecommunications networks.	DT Telefónica
End-to-End Network Slicing	The End-to-End Network Slicing Project Group will identify end-to-end use cases that can be researched, developed, and demonstrated to help operators overcome many of the key challenges of employing network slicing to support their 5G services.	BT HPE
People and Process	The goal of the People and Process Project Group is to share cultural and process transformation practices that can materially improve operators' key metrics.	Facebook Bell Canada

* Excerpts from the TIP website

fiber. With these advantages, the PG is planning a field trial to obtain practical expertise to apply the technology in city areas where capacity demands tend to increase in a short period.

5. Core and Management Projects

The goal of PGs focusing on the Core and Management field is to innovate network operations. Their activities are summarized in **Table 3** [1].

The End-to-End Network Slicing PG deals with virtualized networks, or network slices, that are built on physical infrastructure. Their goal is to achieve operation that involves configuration and management of network slices spanning multiple operators' domains. To that end, they study interface specifications that enable cooperation of multiple orchestrators, each of which manages network slices and their constituent physical resources in a specific domain. Their scope is aligned with emerging 5G mobile technologies.

The Artificial Intelligence and Applied Machine Learning PG is working to introduce artificial intelligence (AI) in network operations. They study how to streamline operation processes or how to optimize networks by utilizing AI where human judgment currently intervenes. Although their work is in an early phase, it could attract attention as a new approach for introducing innovation.

6. Outlook

The advancement of processor chips' capabilities and software engineering techniques and the trend of openness have brought various innovations, disaggregation being an example. Disaggregation, or unbundling functions or parts that used to be integrated, sometimes as a black box, was first introduced in datacenters.

Because telecom networks contain a wide variety of equipment in terms of both functionality and size, it is not likely that the same techniques will be directly introduced into telecom networks immediately. However, it is almost certain that the approach will have a significant influence on the research and development of telecom networks.

NTT has been participating in several PGs of TIP not just to observe their activities but also to make contributions. For example, in the Open Optical Packet Transport PG, we have proposed a packet switching and forwarding architecture called Multi-Service Fabric (MSF) [2, 3] that enables highly reliable, multi-vendor implementations. Through the activities at TIP, we hope to keep engaging ourselves in the newest technical trends and use them in our research and development.

References

[1] Excerpts from TIP website, https://telecominfraproject.com/

[2] K. Takahashi, H. Yoshioka, K. Ono, and T. Iwai, "Promoting the MSF Architecture for Flexible Networks," NTT Technical Review, Vol. 14, No. 10, 2016.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201610fa6.html

[3] Multi-Service Fabric (MSF), https://github.com/multi-service-fabric/msf

Trademark notes

Takeshi Kinoshita

Research Engineer, Media Innovation Laboratory, NTT Network Innovation Laboratories.

He received a B.E. and M.E. in nuclear engineering from Kyoto University in 1994 and 1996. After joining NTT Optical Network Systems Laboratories in 1996, he studied optical fiber communications systems in access networks and their management. He then moved to the R&D center of NTT WEST, where he was responsible for developing and introducing commercial services, including broadband Internet access, Internet protocol (IP) v4/v6 virtual private networks, and wide area Ethernet. He also worked on reducing energy consumption of datacenter components such as servers, communications equipment, and air-conditioning systems. His current research involves network softwarization technologies including software-defined networking (SDN) and network virtualization.

Hitoshi Masutani

Senior Research Engineer, Media Innovation Laboratory, NTT Network Innovation Laboratories.

He received a B.E. in communication engineering in 1999 and an M.E. in electrical, electronic, and information engineering from Osaka University in 2001. After joining NTT Network Innovation Laboratories in 2001, he studied multicast networking and SIP (session initiation protocol)-based home networking. In 2005, He moved to the Visual Communication Division of NTT Bizlink, where he was responsible for developing and introducing visual communication services, including an IP-based high-quality large scale video conferencing system and a real-time content delivery system on IPv6 multicast. He also worked on developing their service order management system and network management system for video conference services. He has recently been developing high performance network function software for telecom carrier networks based on network functions virtualization at NTT Network Innovation Laboratories.

Katsuhiro Shimano

Senior Manager, NTT Network Innovation Laboratories.

He received a B.S. in physics from Waseda University, Tokyo, in 1991 and an M.S. in physics from the University of Tokyo in 1993. Since joining NTT in 1993, he has studied optical networks and network management and related areas such as optical network management systems, GMPLS (generalized multiprotocol label switching), and traffic engineering. He also spent time at the headquarters working on NGN (Next Generation Network) network architecture from the first phase of construction. He has recently been leading the research on SDN and network virtualization at NTT Network Innovation Laboratories.

Event Report: NTT Communication Science Laboratories Open House 2018

Tomoharu Iwata, Tomohiro Amemiya,
Hideharu Nakajima, Yoshinari Shirai,
and Yoshifumi Shiraki

Abstract

NTT Communication Science Laboratories Open House 2018 was held in Keihanna Science City, Kyoto, on May 31 and June 1, 2018. Over 1600 visitors enjoyed 6 talks and 29 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 industries, universities, and research institutions who are engaged in research, development, business, and education.

This year, the Open House was held at the NTT Keihanna Building in Kyoto on May 31 and June 1, and over 1600 visitors attended it over the two days. We prepared many hands-on exhibits to allow 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. We also organized an invited talk. 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 "Shift to new dimensions—Further initiatives to deepen Communication Science" (**Photo 1**).

Dr. Yamada mentioned the remarkable progress

Photo 1. Dr. Takeshi Yamada delivering keynote speech.

Photo 2.   Dr. Yasuhiro Takahashi giving research talk.

Photo 3.   Dr. Junji Watanabe giving research talk.

being achieved in deep learning, including recent developments in artificial intelligence (AI) technologies, and said that CS Labs must not only consist of a group of masters who use these leading technologies at a high level but also a group of researchers who are constantly taking on challenges that will open up new dimensions. He then introduced recent efforts in human and information science research at CS Labs. He discussed how to cope with risks and uncertainties accompanying these new initiatives by explaining the Exploration-Exploitation dilemma, which is well known in the field of AI research. The dilemma involves whether to do deeper research in a field we have already studied (i.e., Exploitation), or to explore a field about which we have less experience and do research widely (i.e., Exploration). He declared that every researcher at CS Labs who is responsible for long-term fundamental research will seek to discover and explore new, unknown expanses of knowledge boldly and resolutely.

3. Research talks

Four 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) "The real worth of quantum computers with elementary operations—Analysis of computational power of gate-based quantum computers," by Dr. Yasuhiro Takahashi, Media Information Laboratory

Dr. Takahashi focused on a gate-based quantum computer, which has theoretical evidence of being more powerful than today's computers. There are many problems in realizing gate-based quantum computers equipped with a sufficient amount of computational resources, but he nevertheless introduced two of his recent results demonstrating that gate-based quantum computers are still powerful even under the more realistic condition where computational resources are limited (**Photo 2**).

(2) "Role of haptics in improving wellbeing—Science and design of touch can enhance human flourishing," by Dr. Junji Watanabe, Human Information Science Laboratory

Dr. Watanabe introduced a recently launched study on wellbeing. Wellbeing is gathering attention because research on designing mind-enriching information technology has been pursued recently from a standpoint that is different from the one aiming to merely achieve greater efficiency. He introduced several topics including measuring and specifying factors of wellbeing and modifying cognitive and empathetic attitudes towards oneself and others with haptic experiences (**Photo 3**).

(3) "Beyond combinatorial explosion—Enumeration and optimization with Binary Decision Diagrams," by Dr. Masaaki Nishino, Innovative Communication Laboratory

Dr. Nishino focused on a phenomenon called

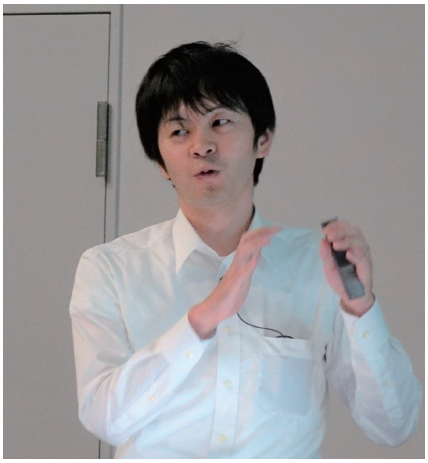


Photo 4. Dr. Masaaki Nishino giving research talk.

Photo 5. Dr. Sadao Hiroya giving research talk.

a combinatorial explosion, in which the number of possible combinations increases exponentially with the number of elements, which we have often experienced in a grouping or travel routes search. He explained how the number of combinations created by the combinatorial explosion can be counted by algorithms based on the data structure called a binary decision diagram (**Photo 4**).

(4) "Speech production and perception share common brain pathways—Investigation of the mechanisms of speech communication by speech conversion and brain imaging," by Dr. Sadao Hiroya, Human Information Science Laboratory

Dr. Hiroya introduced research on neural mechanisms underlying speech communication and showed the existence of the common pathways in the brain between speech production and perception, even though it has been considered that the pathways are different. He also discussed the difference between human brain mechanisms and speech recognition and synthesis techniques (**Photo 5**).

4. Research exhibits

The Open House featured 29 exhibits displaying CS Labs' latest research results. We categorized them into the following four areas: "Science of Machine Learning," "Science of Communication and Computation," "Science of Media Information," and "Science of Humans."

Each exhibit was housed in a booth and employed techniques such as slides on a large-screen monitor or hands-on demonstrations, with researchers explaining the latest results directly to visitors (**Photos 6** and **7**). The following list, taken from the Open House website, summarizes the research exhibits in each category.

4.1 Science of Machine Learning

- Finding similar voice recordings in big data—Graph index-based audio similarity search
- Learning feature combinations from multiple tasks—MOFM: low-rank regression for learning common factors
- Where do they come from? Where are they going?—"Data assimilation and navigation learning for crowd"
- Datafying cities—Event analysis by environmental sensing and machine learning
- Memory efficient deep learning for mobile devices—Quantized neural networks for model compression
- Optics makes machine learning much faster—Photonic reservoir computing for high-speed machine learning
- Interpreting deep learning from network structure—Detecting communities in trained layered neural networks

4.2 Science of Communication and Computation

- Can I borrow your quantum memory?—High-speed quantum computations with uninitialized qubits

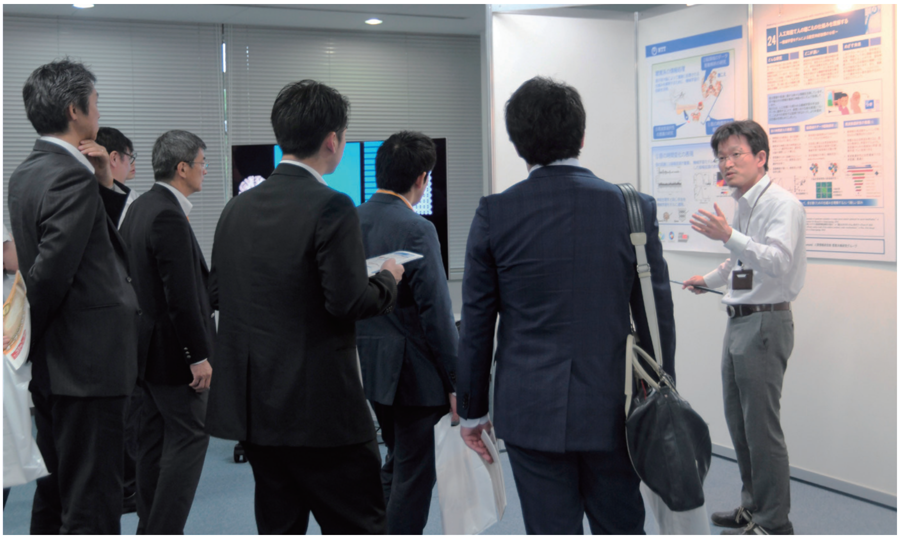


Photo 6. The latest research results were exhibited.

Photo 7. Researchers explaining a demonstration.

- Designing fault-tolerant networks—Maximizing network reliability via binary decision diagrams
- Can computer translate considering context?—Context understanding tests for neural machine translation
- Early vocabulary development in late talkers—Collecting and analyzing data from pediatric medical fields
- Chatting with robots broadens your knowledge—Integration of chat and QA based on two-robot coordination
- Anytime, anywhere, we can speak like a native!—Speech rhythm conversion by mobile application

- Sharing enthusiasm between remote sites—Applause coding for bi-lateral immersive sharing

4.3 Science of Media Information

- Illumination-based color saturation control—Spectral operation using color enhancement factors
- Pay attention to the speaker you want to listen to—Computational selective hearing based on deep learning
- Solving two-choice questions makes AI clever—Deep pairwise comparison model for ASR hypothesis selection
- Estimating objects' visuals only from audio—Cross-media scene analysis
- Cast shadows add dimensions—Projection mapping giving depth illusion to real objects
- Converting impression and intelligibility of speech—Speech attribute conversion using deep generative models
- Creating favorite images with selective decisions—Hierarchical image analysis and synthesis with DTLC-GAN

4.4 Science of Humans

- Predicting attention to the ears by the eyes—Auditory spatial attention revealed as pupillary response
- Measuring, understanding, and empowering wellbeing—Cross-disciplinary research toward "eudaimonic wellbeing"
- Understanding human hearing with AI—Analyzing auditory neural mechanisms with machine learning
- How do excellent batters hit the ball?—Cognitive processes revealed by body movements in batting
- How do excellent batters look at the ball?—Cognitive processes revealed by eye movements in batting
- The sooner you decide, the better you can localize—Visual motion processing in the perception and action
- Let's FEEL shape and action by a force—Can we receive environmental information by Buru-Navi4?
- Feeling bumps on a flat sheet—Magnetic haptic printing technology

5. Invited talk

This year's event also featured an invited talk by Dr. Hiroshi Nakagawa, group director, Artificial Intelligence in Society Research Group, RIKEN Center for Advanced Intelligence Project. The title of his talk was "AI, ethics and social impact." He explained ethics guidelines for developing AI systems proposed in domestic and international organizations, personal data protection in big data utilization, and the proper way to have a relationship with AI. He also discussed the hot topic of whether AI will take away all our jobs. He pointed out that natural language processing will be one of the most important research fields for future AI technologies and explained the importance of redefining what our jobs are by ourselves from a higher viewpoint.

6. Concluding remarks

Just like last year, many visitors came to NTT CS Labs Open House 2018 [1, 2] 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 2018 (in Japanese).
http://www.kecl.ntt.co.jp/openhouse/2018/index.html

[2] Website of NTT Communication Science Laboratories Open House 2018 (in English).
http://www.kecl.ntt.co.jp/openhouse/2018/index_en.html

Tomoharu Iwata

Distinguished Researcher/Senior Research Scientist, Ueda Research Laboratory, NTT Communication Science Laboratories.

He received a B.S. in environmental information from Keio University, Kanagawa, in 2001, an M.S. in arts and sciences from the University of Tokyo in 2003, and a Ph.D. in informatics from Kyoto University in 2008. In 2003, he joined NTT Communication Science Laboratories. From 2012 to 2013, he was a visiting researcher at University of Cambridge, UK. His research interests include data mining and machine learning.

Tomohiro Amemiya

Distinguished Researcher/Senior Research Scientist, Sensory and Motor Research Group, 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 biomedical information science from Osaka University in 2008. He joined NTT in 2004. He was an honorary research associate at the Institute of Cognitive Neuroscience, University College London, in 2014–2015. He researches human-computer interfaces applying sensory illusions and haptic perception.

Hideharu Nakajima

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

He received a Ph.D. in science from Waseda University, Tokyo, in 2010. His research interests include prosodic/linguistic/pragmatic analysis of spoken/written messages, spoken language processing (speech recognition, speech synthesis), speech communication with robots/agents, and educational technology.

Yoshinari Shirai

Senior Research Scientist, Learning and Intelligent Systems Research Group, Innovative Communication Laboratory, NTT Communication Science Laboratories.

He received a Ph.D. in engineering from the Graduate School of Engineering at the University of Tokyo. His research interests include history-enriched environments, ubiquitous computing, and interaction design.

Yoshifumi Shiraki

Research Scientist, Moriya Research Laboratory, NTT Communication Science Laboratories.

He received a Ph.D. in engineering from Tokyo Institute of Technology in 2015. His research focuses on signal processing in sensor networks and visible light communication.

Ultrahigh-speed Integrated Circuit Capable of Wireless Transmission of 100 Gbit/s in the 300-GHz Band

1. Introduction

NTT and Tokyo Institute of Technology have jointly developed an ultrahigh-speed integrated circuit (IC) for wireless front-end that operates on a terahertz frequency band, and they have succeeded in developing the world's fastest 100-Gbit/s wireless transmission data rate in the 300-GHz band.

Unused terahertz waves* are expected to be applicable to high-speed wireless transmission since a wide frequency band can be secured. In our research, we implemented a mixer circuit that applied a unique proprietary high isolation design technology with an Indium phosphide high electron mobility transistor (InP-HEMT). This enlarged the transmission bandwidth, which is a problem in the conventional 300-GHz-band wireless front end. It also improved the signal-to-noise ratio (SNR). In addition, we used this circuit to develop a 300-GHz-band wireless front-end module, and we achieved wireless transmission of 100 Gbit/s.

In this research, we achieved 100-Gbit/s wireless transmission with one wave (one carrier), so in the future, we can extend this idea to multiple carriers by making use of the wide frequency band of 300 GHz and using spatial multiplexing technology such as multiple-input multiple-output (MIMO) and orbital angular momentum (OAM).

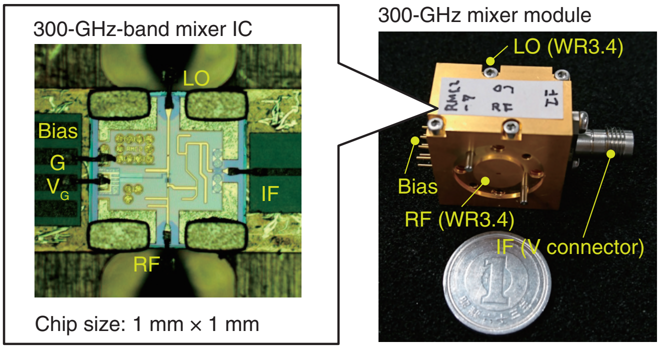
Ultrahigh-speed IC technology is expected to enable high-capacity wireless transmission of 400 Gbit/s. This is about 400 times that of the current LTE (Long-Term Evolution) and Wi-Fi, and 40 times that of 5G (fifth-generation mobile communications system) technology. Ultrahigh-speed ICs are also expected to open up utilization of the unused terahertz wave frequency band in the communications field and in non-communications fields.

2. Research background

High-capacity wireless transmission technology of 100 Gbit/s has attracted worldwide attention with the spread of broadband networks. There are three ways of further increasing the capacity of wireless transmission—expanding the transmission bandwidth, increasing the modulation multi-level number, and increasing the spatial multiplexing number. To realize future large capacity wireless transmission technology from a level of 400 Gbit/s to that of one terabit per second (Tbit/s), it is necessary to expand both the transmission bandwidth and the modulation multi-level number simultaneously in one wave (one carrier) and to increase the number of spatial multiplexing transmissions by superimposing them multiple times.

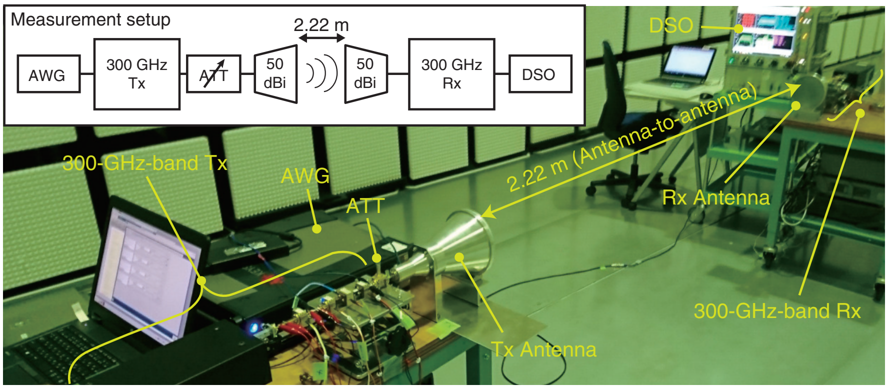
The transmission bandwidth is limited in the carrier frequencies from 28 GHz to 110 GHz that are currently being researched and developed. Thus, researchers are studying the use of frequencies that make it easier to expand the transmission band area, from the 300-GHz band to the terahertz wave frequency band. The 300-GHz band has a frequency that is 10 times or more higher than the 28-GHz band that is being studied for 5G, which will be the next generation mobile communications technology. With the 300-GHz band, it will be easier to secure a wide transmission bandwidth. However, at high frequencies, leakage of unnecessary signals tends to occur between the ports inside the IC and the mounting, and

* Terahertz wave: Just as we use the term *kilo* to mean 10^3, so we use the term *giga* to mean 10^9 and *tera* to mean 10^{12}. Hertz (Hz) is a unit of a physical quantity called frequency. It indicates how many times alternating electric signals and electromagnetic waves change polarity (plus and minus) per second. That is, one terahertz (1 THz = 1000 GHz) is the frequency of the electromagnetic wave in which the polarity changes 1×10^{12} times per second. In general, a terahertz wave often indicates an electromagnetic wave of 0.3 THz to 3 THz.



IF: intermediate frequency port
LO: local oscillation frequency port
RF: radio frequency port

Fig. 1. Mixer IC and module.



ATT: attenuator
AWG: arbitrary waveform generator
DSO: digital storage oscilloscope
Rx: receiver
Tx: transmitter

Fig. 2. Transmission experiment.

so far, it has been impossible to obtain a sufficiently high SNR. For this reason, even if a 300-GHz band is used, it is impossible to obtain both a wide transmission bandwidth and a high modulation multi-level value at the same time, and so wireless transmission up to now has remained at the rate of several tens of gigabits per second.

3. Research results

In this research, we devised our own high isolation design technology and applied this technology to a mixer circuit, which is a key component responsible for frequency conversion in the 300-GHz-band wireless front end. We also developed an IC with an InP-HEMT. By applying high isolation design technology, we succeeded in suppressing leakage of unnecessary signals inside each IC and between ports in the IC. We also succeeded in improving signal noise reduction and expanding the bandwidth, which had been issues confronting the use of conventional 300-GHz-band wireless front-end technology until now. With these technologies, we also developed a 300-GHz-band wireless front-end module (**Fig. 1**), and we confirmed reception of a good 16QAM (quadrature amplitude modulation) signal in back-to-back transmission. In the 300-GHz band, we also confirmed transmission at a speed of 100 Gbit/s, the first time in the world this speed has been attained (**Fig. 2**).

4. Future prospects

In our research, we achieved 100-Gbit/s transmission with one wave (one carrier), so in the future we will expand our research to use multiple carriers or spatial multiplexing technology such as MIMO and OAM together with this result. With this combination, we expect to achieve ultrahigh-speed IC technology that enables high-capacity wireless transmission of over 400 Gbit/s. In addition, we expect that our approach will be applicable in various fields such as imaging and sensing in which terahertz waves are expected to be used. Through collaboration with partners, NTT aims to create new services and new industries using ultrahigh-speed ICs, and also aims to further develop ultrahigh-speed IC technology.

For Inquiries

NTT Science and Core Technology Group
http://www.ntt.co.jp/news2018/1806e/180611a.html

External Awards

Research Encouragement Award

Winner: Hitoshi Niigaki, Ken Tsutsuguchi, and Tetsuya Kinebuchi, NTT Media Intelligence Laboratories
Date: June 22, 2018
Organization: The Institute of Image Electronics Engineers of Japan

For "Detection of Tree Trunk from 3D Point-cloud Measured by Mobile Mapping System –Feature Extraction for Sweep Object Detection–."

Best Paper Award

Winner: Kyoko Sudo, Toho University; Kazuhiko Murasaki, Tetsuya Kinebuchi, Media Intelligence Laboratories; Shigeko Kimura, Kayo Waki, and Kazuhiko Ohe, The University of Tokyo
Date: July 15, 2018
Organization: Multimedia on Cooking and Eating Activities, Human Communication Group, the Institute of Electronics, Information and Communication Engineers (IEICE)

For "Intuitively Estimating the Healthiness of Meals from Their Images."
Published as: K. Sudo, K. Murasaki, T. Kinebuchi, S. Kimura, K. Waki, and K. Ohe, "Intuitively Estimating the Healthiness of Meals from Their Images," Proc. of the Joint Workshop on Multimedia for Cooking and Eating Activities and Multimedia Assisted Dietary Management, Stockholm, Sweden, July 2018.

IEEE Photonics Society Distinguished Lecturer

Winner: Hiroki Takesue, NTT Basic Research Laboratories
Date: July 18, 2018
Organization: The Institute of Electrical and Electronics Engineers (IEEE) Photonics Society

For "Coherent Ising Machine: A Photonic Ising Model Solver Based on Degenerate Optical Parametric Oscillator Network."

Suzuki Memorial Award

Winner: Shota Orihashi, NTT Media Intelligence Laboratories
Date: August 30, 2018
Organization: The Institute of Image Information and Television Engineers (ITE)

For "A Study on Fast Block Partitioning in HEVC Using Convolutional Neural Network" (in Japanese).
Published as: S. Orihashi, S. Kudo, M. Kitahara, and A. Shimizu, "A Study on Fast Block Partitioning in HEVC Using Convolutional Neural Network," ITE Annual Convention, 21B-3, Tokyo, Japan, Aug./Sept. 2017.

IPSJ Yamashita SIG Research Award

Winner: Hiroyuki Kirinuki, NTT Software Innovation Center
Date: August 30, 2018
Organization: Information Processing Society of Japan (IPSJ)

For "Automatic Locator Repair on GUI Test Scripts."
Published as: H. Kirinuki, H. Tanno, and K. Natsukawa, "Automatic Locator Repair on GUI Test Scripts," Proc. of IPSJ/SIGSE Software Engineering Symposium (SES2017), Tokyo, Japan, Aug./Sept. 2017.

Papers Published in Technical Journals and Conference Proceedings

Interactive Proofs with Polynomial-time Quantum Prover for Computing the Order of Solvable Groups

F. Le Gall, T. Morimae, H. Nishimura, and Y. Takeuchi

Proc. of the 43rd International Symposium on Mathematical Foundations of Computer Science (MFCS 2018), Vol. 117, No. 1, pp. 26:1–26:13, Liverpool, UK, August 2018.

In this paper we consider what can be computed by a user interacting with a potentially malicious server, when the server performs polynomial-time quantum computation but the user can only perform polynomial-time classical (i.e., non-quantum) computation. Understanding the computational power of this model, which corresponds to polynomial-time quantum computation that can be efficiently verified classically, is a well-known open problem in quantum computing. Our result shows that computing the order of a solvable group, which is one of the most general problems for which quantum computing exhibits an exponential speed-up with respect to classical computing, can be realized in this model.

Novel Optimizing Technique for Linear Optical Mach-Zehnder Modulator and Its Experimental Verification Using PAM-8 Signal

H. Kawakami, H. Yamazaki, and Y. Miyamoto

Proc. of the 44th European Conference on Optical Communication (ECOC 2018), We2.13, Rome, Italy, September 2018.

We propose an optimizing technique for a linear optical modulator for selecting whether to linearize the optical power or electric field response. With this technique, a PAM-8 signal was successfully generated with a driving signal having a swing voltage of 0.7Vpi.

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

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

arXiv:1809.07552 [quant-ph], September 2018.

Measurement-based quantum computing is one of the most promising quantum computing models. Among various universal resource states proposed so far, the Union Jack state is the best in the sense that it requires only Pauli-X, Y, and Z basis measurements. It was open whether only two Pauli bases are enough for universal measurement-based quantum computing. In this paper, we construct a universal hypergraph state that only requires X and Z-basis measurements. We also show that the fidelity between a given state and our hypergraph state can be estimated in polynomial time using only X and Z-basis measurements, which is useful for the verification of quantum computing. 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.

Verification of Many-qubit States

Y. Takeuchi and T. Morimae

Proc. of the 18th Asian Quantum Information Science Conference (AQIS 2018), p. 48, Nagoya, Aichi, Japan, September 2018.

Verification is a task to check whether a given quantum state is close to an ideal state or not. In this paper, we show that several many-qubit states can be verified with only single-qubit Pauli measurements. Specifically, we introduce protocols for ground states of Hamiltonians, quantum states generated by some quantum circuits, and all polynomial-time-generated hypergraph states. Importantly, we do not make any assumption that the i.i.d. copies of the same states are given. Our protocols work even if entanglement is created among copies in any artificial way. As an application, we consider the verification of the Bremner-Montanaro-Shepherd-type IQP circuits.

Power of Uninitialized Qubits in Shallow Quantum Circuits

Y. Takahashi and S. Tani

Proc. of AQIS 2018, p. 49, Nagoya, Aichi, Japan, September 2018.

We study the computational power of shallow quantum circuits with $O(\log n)$ initialized and $n^{O(1)}$ uninitialized ancillary qubits, where n is the input length and the initial state of the uninitialized ancillary qubits is arbitrary. First, we show that such a circuit can compute any symmetric function on n bits that is computable by a uniform family of polynomial-size classical circuits. Then, we regard such a circuit as an oracle and show that a polynomial-time classical algorithm with the oracle can estimate the elements of any unitary matrix corresponding to a constant-depth quantum circuit on n qubits. Since it seems unlikely that these tasks can be done with only $O(\log n)$ initialized ancillary qubits, our results give evidences that adding uninitialized ancillary qubits increases the computational power of shallow quantum circuits with only $O(\log n)$ initialized ancillary qubits. Lastly, to understand the limitations of uninitialized ancillary qubits, we focus on sub-logarithmic-depth quantum circuits with them and show the impossibility of computing the parity function on n bits.

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

Y. Takeuchi

AQIS 2018 Kyoto satellite workshop on quantum computing, Kyoto, Japan, September 2018.

In measurement-based quantum computing, checking whether correct graph states are generated or not is essential for reliable quantum computing. Several verification protocols for graph states have been proposed, but none of these are particularly resource efficient: Many copies are required in order to extract a single state that is guaranteed to be close to the ideal graph state. For example, the best protocol currently known requires $O(n^{15})$ copies of the state, where n is the size of the graph state. In this talk, we propose a significantly more resource-efficient verification protocol for graph states that needs only $O(n^5 \log n)$ copies. The key idea that achieves such a drastic improvement is to employ Serfling's bound, which is a probability inequality in classical statistics. Utilizing Serfling's bound also enables us to generalize our protocol for qudit graph states and continuous-variable weighted hypergraph states. This talk is based on joint work with Atul Mantri, Tomoyuki Morimae, Akihiro Mizutani, and Joseph F. Fitzsimons. The detail is given in arXiv:1806.09138.