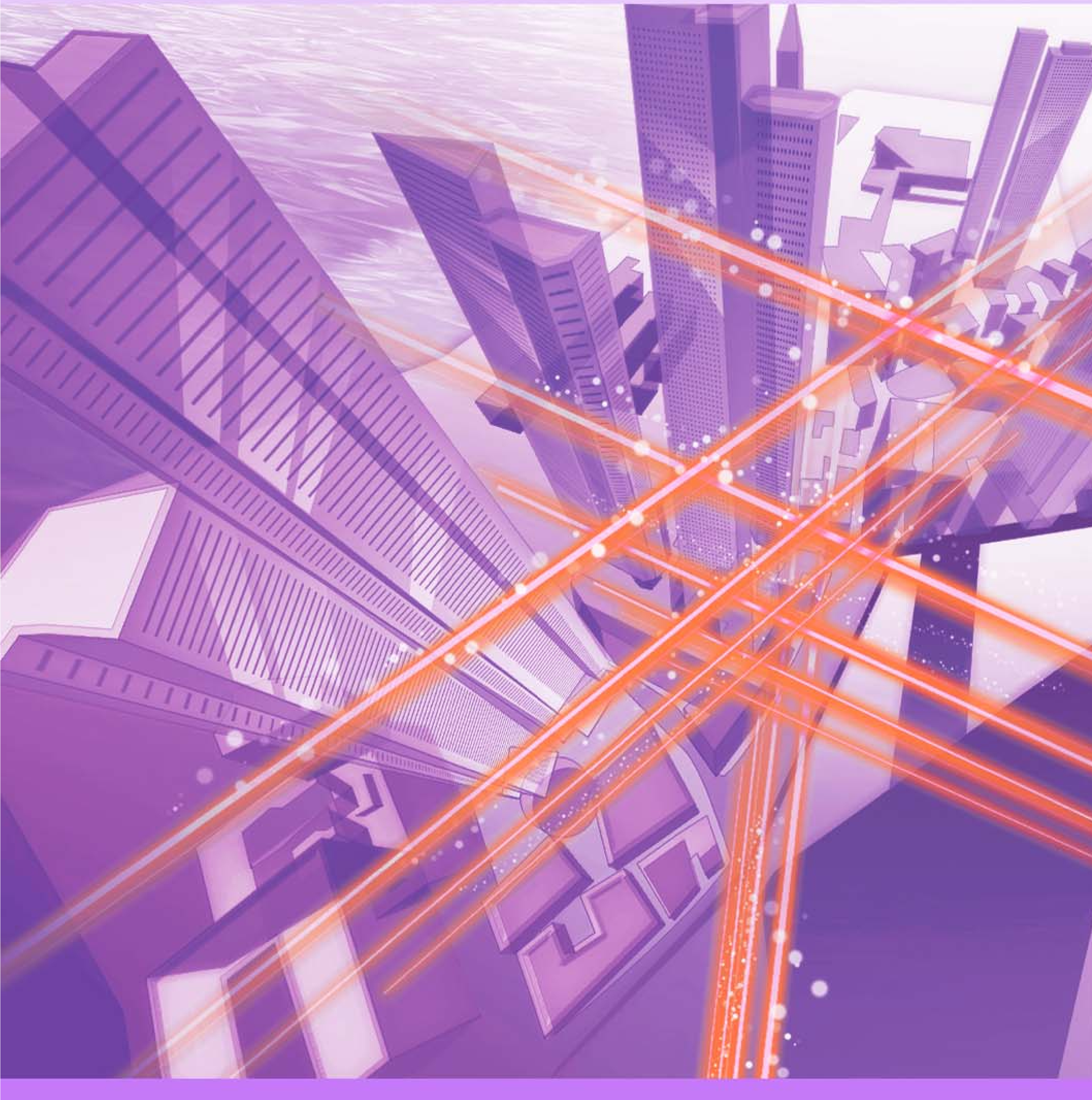
NTT Technical Review

8
2018



August 2018 Vol. 16 No. 8

NTT Technical Review

August 2018 Vol. 16 No. 8

External Awards/Papers Published in Technical Journals and Conference Proceedings

- External Awards/Papers Published in Technical Journals and Conference Proceedings

Changing the World through Basic Research on Light Demands a Pioneering Spirit and Results beyond the Ordinary

Hiroki Takesue
Senior Distinguished Researcher, NTT Basic Research Laboratories

Overview

The pace of achieving further advancements in digital computers is slowing down, but research and development of new types of computers is accelerating throughout the world. NTT has achieved a coherent Ising machine that enables high-speed computation based on phase transition phenomena in laser oscillators. Dr. Hiroki Takesue, Senior Distinguished Researcher at NTT Basic Research Laboratories, received the Nishina Memorial Prize in 2017, a prize awarded to researchers with outstanding achievements in physics and its applications. We asked him about his pioneering research results that could lead to solutions to a variety of problems facing modern society and his mind frame as a researcher.

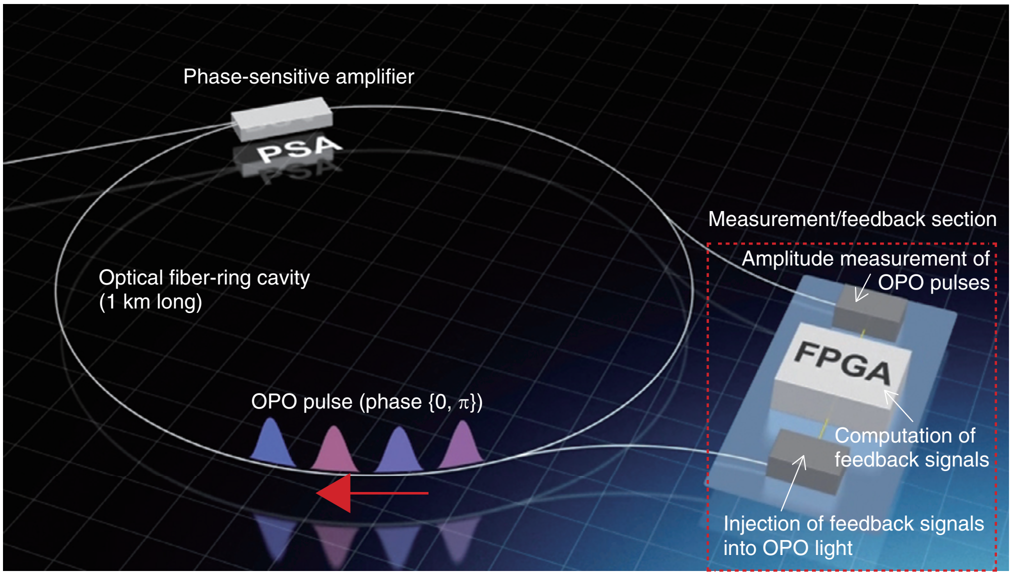
Keywords: Ising model, combinatorial optimization problem, quantum neural network

Research crossing many fields

—Dr. Takesue, please accept our congratulations on receiving the Nishina Memorial Prize. Can you tell us something about the research that resulted in you earning this award?

Thank you very much. I was quite surprised to learn that I would be receiving this prestigious prize, which is usually given out for research already recognized by many people as having great value. I didn't think that our research that is now just starting to blossom would be a target of this award. I am truly humbled by being selected, but I look upon this prize as an expression of encouragement from the research community that will help our team to carry on with great resolve. At the same time, I feel it is my duty to uncover the true essence of this research and to determine what kind of value it can provide.

The research theme that led to this prize being awarded was centered on achieving a large-scale coherent Ising machine (CIM). This is a type of computer used to solve through physical experiments the problem of determining the minimum energy state in an Ising model, which is a theoretical model describing a group of spins that mutually interact with each other. We have achieved a CIM that can compute this Ising model at high speed using a special laser light



FPGA: field programmable gate array
OPO: optical parametric oscillator

Fig. 1. Configuration of CIM.

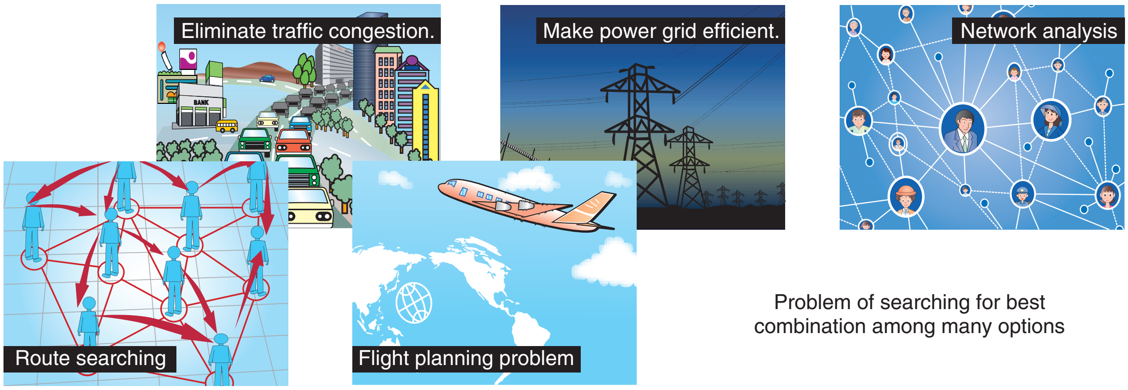
instead of spins (**Fig. 1**).

A *spin* is a physical quantity that we can think of as a tiny magnet that can take on either of two values: up or down. A CIM simulates a spin by a laser light that assigns phase zero to a spin-up state and phase π to a spin-down state. However, ordinary lasers take on any phase value from 0 to 2π, so using such a laser to represent spins requires precise phase control, which is not an easy task. Therefore, we decided to use a special type of laser called an optical parametric oscillator (OPO). When oscillating, an OPO can take on only a phase of 0 or π, which enables stable representation of spins. By using this OPO to generate 2000 optical pulses within an optical fiber-ring cavity whose length is as long as 1 km and applying feedback to those pulses based on the measurement results, we achieved a computer that could instantaneously solve a combinatorial optimization problem (**Fig. 2**) using 2000 artificial spins that exhibit complex interactions with each other.

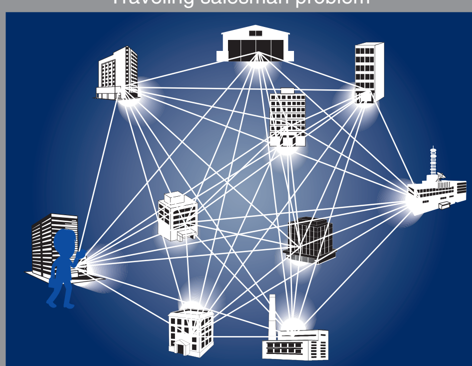
Modern society is becoming increasingly networked in complicated ways and thus provides various situations that correspond to combinatorial optimization problems. The aim here is to derive an optimal or a near-optimal combination from a massive number of options originating in various types of elements and information. Typical examples include selection of an optimal route for a delivery truck and optimal allocation of channels in wireless communications. A classic example of this type of problem is the traveling salesman problem, in which the objective is to find the route that minimizes the distance a salesman has to travel when seeking to visit each of a number of cities only once and return to the starting point. If there are only 5 cities to visit, there will be 12 combinations, but for 10 cities, the number of combinations balloons to an enormous number of about 180,000.

If we attempt to solve this problem by brute-force computations of multiple combinations, the number of combinations will increase in an exponential manner as the number of options becomes larger, which means that those computations would require a huge amount of time. It is known that many such combinatorial optimization problems can be converted to the problem of searching for the minimum energy state in the Ising model, a well-known theoretical model in statistical physics. We kept this in mind when we adopted the approach of “experimentally solving” this Ising model problem using an actual physical system, which would therefore enable us to solve at high speed a variety of combinatorial optimization problems.

We have demonstrated that high-speed solution searching of an Ising model problem consisting of 2000 spins can be performed using a CIM [1]. The



Traveling salesman problem



Number of cities	Number of route combinations
4	3
5	12
6	60
10	181,440
20	60,822,550,204,416,000
30	4,420,880,996,869,850,000,000,000,000,000
60	≅Total number of observable atoms in space (= 10^{80})

Fig. 2. Combinatorial optimization problems.

results of this experiment were well received by the academic community and were presented in the American journal *Science* in 2016 [2]. Incidentally, I would like to point out that CIM operation was still unstable as of 2016, which meant that computations had to be constantly adjusted by a researcher. In November 2017, however, we succeeded in achieving long-term stabilization of a CIM for the QNN (Quantum Neural Network) cloud [3] that enables general users to experience computations by CIM for themselves. This service was achieved in collaboration with the National Institute of Informatics (NII), the University of Tokyo, and the Japan Science and Technology Agency (JST) (**Fig. 3**).

—Do you remember the moment you obtained results with your CIM?

Yes, I remember it well. The first time we generated a large-scale OPO pulse group, we just wanted to see what would happen, but we were surprised at the results we obtained, saying: “It’s working!” We did not expect to have progressed that far in our work at that point.

Then, when we first obtained computational results, we really didn’t know how to go about evaluating them. These were results that were outside the bounds of fields like quantum optics and photonics that we had been researching up to then. We therefore spent several months discussing these results with researchers in other fields such as computer science, and once we were able to confirm that they represented a solution, I was truly relieved. Although this research was part of a national project, we were taking a great risk not knowing whether we could actually obtain a solution, and I felt responsible for the outcome. On the other hand, I remember it as a time in which we realized that we took up the challenge of interdisciplinary research including the establishment of an evaluation index.

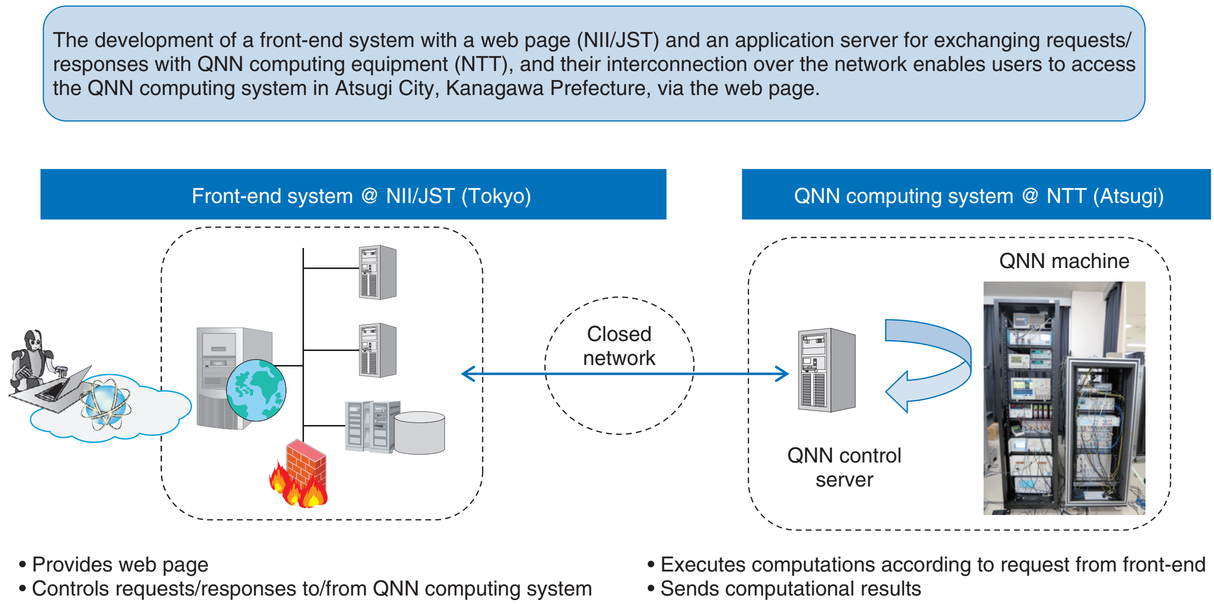


Fig. 3. Website for experiencing QNN [3].

—Where do you think the significance of this research lies?

Information technology (IT) has been progressing by leaps and bounds these last 20 to 30 years supported by an exponential jump in computer performance. Today's computers are nothing like the computers of 30 years ago, but further evolution according to Moore's law appears to be coming to an end, suggesting that computer development has peaked. Meanwhile, IT networks such as the social networking services that we have come to use daily are becoming increasingly complex, and to make them more efficient, we need to find good solutions from combinations of all sorts of elements and options. Obtaining solutions to such problems using conventional computers is extremely difficult, so if a path to solving them by another approach can be found, it might be possible to break through the limits of Moore's law.

Building a bridge between needs and seeds

—Your research seems to be on the path toward revolutionizing society. To that end, how will you approach this research and how do you see it developing going forward?

We are currently faced with two main research problems. One is clarifying the essence of a CIM—what exactly is it? That is, I would like to clarify for ourselves the nature of a machine that couples thousands of laser pulses and solves complex problems and what, in any, are its limits.

The other problem is how to make not only academic contributions but contributions to society as well. As a new technology, the significance of CIM is still up in the air, so there are many challenges that we must face. The number of researchers in this field is small even from a global perspective, and even though we are striving to make this machine useful to the world, we must realize that, in the end, it will be society that decides the value of this research.

Another way of describing this approach is "seeds oriented." "Needs," of course, are also important, but much of basic research is pursued without knowing how useful it might be to society, so it truly is a seeds-oriented endeavor. Basic research may include many failures, but it holds the possibility of creating something new. Indeed, an innovative development can be thought of as something previously unknown to the world. Something that we try out even though we don't understand it or something that suddenly appears much like a mutation may end up changing the world. I myself work on the seeds side, but since I have experienced the importance of both, I can also see myself as building a bridge between seeds and needs.

Yet, when asked about the significance of pursuing combinatorial optimization problems, I am being totally honest when I say that they may hold the key to solving the problems plaguing modern society. This reason, however, is more of an afterthought since my research is initially driven by something that is "interesting" to me. Performing optical experiments is what I like best. I also feel that researching light is a way of opening my "window to the world."

Personally, it's not just about science and physics. Through research, I experience a somewhat vague feeling of wanting to know more about the world and human beings. Actually, I sometimes feel that knowledge obtained through optics research is somehow connected to that obtained in completely different fields. This is when I sense that my research relates to learning about people and the world.

Learning from the environment and making decisions

—Can you tell us what put you on the path toward the life of a researcher?

It was the fascination I felt for the dual nature of light, that is, the contradictory wave nature and particle nature of light, in my fourth year of university. This feeling came about through an interference experiment I conducted using photons. I remember the excitement I felt in the pit of my stomach as I said to myself "This is it—the photon/quantum nature of light!" I had this feeling that light could serve to vividly depict the essence of physics, and I developed a burning desire to become a researcher. Actually, when I entered university, I initially had no vision for the future and was simply a student more interested in clubs and extracurricular activities than studies. What transformed me was my encounter with Prof. Tetsuro Kobayashi, a student supervisor, and Prof. Masahiro Kitagawa, an NTT alumnus, who were involved in an optical measurement experiment using lasers in the laboratory that I had been assigned to. I'm the type of person that can become completely absorbed in one thing, and for me, that was laser experiments, and it was at that time that I resolved to pursue the life of a researcher.

I later joined NTT with an eye to becoming a researcher, but at that time, graduates of master's programs were first assigned to an NTT branch office for training and an aptitude assessment, which is what happened to me. While working at that branch office, I longed for the day that I would become a researcher.

On completing my training, I was told that I would be assigned to a research laboratory, which I was extremely glad to hear. However, my joy was short lived, as I was assigned to a laboratory in Tokai, Ibaraki Prefecture, which was different from what I had hoped for, and to a position involving optical fiber measurements and the like. At first, I felt dissatisfied with my work, but before long, I came to appreciate the abilities of the researchers around me. I was also blessed with great superiors, and I changed my way of thinking, saying to myself: "I'm going to make an effort to follow in their footsteps!" I became highly motivated in my third year there.

I later moved to a research laboratory in Tsukuba, where it was decided that I would be involved in the development of a wavelength multiplexing access network called Tsukuba WAN (wide access network). This was a cutting-edge, epoch-making undertaking for its time, but given my strong desire for basic research, I again started out with a somewhat negative attitude, thinking: "What am I doing here?" However, on working with specialists from other NTT laboratories on this project, I started to notice the high professional level of their work.

All in all, it was a joint effort between optical-network researchers and specialists, in which the former had to understand all aspects of the technology and design the network in detail, while the latter had to grasp the entire project, organize many people, push the project forward, and draw up specifications. These were people with completely different specialties who were each very competent in their work. Compared with them, I felt quite mediocre in what I was doing and made up my mind to pursue basic research in my own way. Looking back, I think that if I had been directly assigned to NTT Basic Research Laboratories and started out on a researcher's path from the start, my present research style would not have developed the way that it did thanks to these experiences.

—So instead of being a detour, it was a period that cultivated your present research style and individuality. Dr. Takesue, what would you say to young researchers based on your experiences?

I believe that basic research is a field that requires "results beyond the ordinary." Visitors to the NTT laboratories enjoy hearing about the new seeds that we have uncovered in our research. Seeing their reaction to our achievements makes me realize that researchers in this field must have a pioneering spirit ready for unforeseen results.

This is what I would like to say to young people based on my experiences and character: even if you should find yourself in a less than satisfactory situation, you should be able to learn something from that environment and the people around you. I would like you to persevere in such a situation and take the time to look around you. If you notice nothing positive at all, you may think it's best to quit, but the NTT laboratories have no shortage of places where you can learn many things.

At the same time, I would like young researchers to have the courage to make decisions and take action. In my seventh year at NTT, the opportunity to move to NTT Basic Research Laboratories—as I had long desired—finally arrived, but I also felt some fear of making a change. At that time, I associated NTT Basic Research Laboratories with very high-level research, and with only a master's degree level of knowledge in quantum physics, I wondered if I could fit in, and I thought that I would certainly be treated as a freshman despite being over 30. I was therefore filled with a lot of anxiety. Perhaps such fear is normal when making a life change, but thankfully, curiosity and a desire to take on a new challenge finally won out. As a result, I now feel that I can research freely to my heart's desire in a new field. When a yearning to make a change springs up in you, I think you will obtain better results by being honest with yourself than by falling victim to your fears.

Finally, I would ask young researchers in a new field to welcome challenges, since they can be great opportunities for success. Researchers in a field having a long history and a proven track record may have a sense of security, but there will be many forerunners in that field with great achievements, which means that unfinished work will be extremely limited. A new field, in contrast, can be thought of as a "broad ocean" with few researchers who have already uncovered value. A new field may be risky, but it should also present many opportunities for young researchers to succeed.

References

[1] H. Takesue, T. Inagaki, K. Inaba, and T. Honjo, "Quantum Neural Network for Solving Complex Combinatorial Optimization Problems," NTT Technical Review, Vol. 15, No. 7, 2017.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201707fa2.html

[2] T. Inagaki, Y. Haribara, K. Igarashi, T. Sonobe, S. Tamate, T. Honjo, A. Marandi, P. L. McMahon, T. Umeki, K. Enbutsu, O. Tadanaga, H. Takenouchi, K. Aihara, K. Kawarabayashi, K. Inoue, S. Utsunomiya, and H. Takesue, "A Coherent Ising Machine for 2000-node Optimization Problems," Science, Vol. 354, No. 6312, pp. 603–606, 2016.

[3] QNNcloud, https://qnncloud.com/

■ Interviewee profile

Hiroki Takesue

Senior Distinguished Scientist (Senior Distinguished Researcher), Group Leader, Quantum Optical State Control Research Group, Optical Science Laboratory, NTT Basic Research Laboratories.

He received a B.E., M.E., and Ph.D. in engineering science from Osaka University in 1994, 1996, and 2002. In 1996, he joined NTT, where he engaged in research on lightwave frequency synthesis, optical access networks using wavelength division multiplexing, and quantum optics. He is currently pursuing research on communication and computation using quantum optics technologies. He is the recipient of several awards, including the ITU-T (International Telecommunication Union - Telecommunication Standardization Sector) Kaleidoscope Conference 2nd Best Paper Award in 2008, The Commendation for Science and Technology by the Minister of Education, Culture, Sports, Science and Technology of Japan (The Young Scientists' Prize) in 2010, and the Nishina Memorial Prize in 2017. He was a visiting scholar at Stanford University, California, USA, from 2004 to 2005, and a guest researcher at the National Institute of Standards and Technology, Colorado, USA, in 2014. He is a Guest Professor in the Graduate School of Engineering Science, Osaka University, and a member of the Institute of Electrical and Electronics Engineers and the Japan Society of Applied Physics.

Developing Artificial Intelligence Services that Satisfy Customer Demands: Moving forward with Social Implementation of corevo® Technologies

Hideaki Ozawa

Abstract

Artificial intelligence (AI) technologies are playing a steadily increasing role in supporting human work and social activities, such as in contact center operations and smart speaker implementations. This article introduces NTT's initiatives in developing AI technologies that are essential in providing services needed by customers in the real world, including the technologies needed for effectively utilizing AI in the real world, such as AI applications for household voice communication systems and medicine. Technologies for accelerating processing speed to collect more extensive data from the real world and improve the accuracy of AI are also introduced.

Keywords: corevo, AI, DNN

1. Introduction

The NTT Group is pursuing technology development, service proposition, and social implementation of its artificial intelligence (AI) technologies called corevo® based on the following four concepts:

(1) Agent-AI: Supporting interaction between humans and computers such as speech recognition and natural language processing

(2) Heart-Touching-AI: Generating added value by leveraging the internal human mental state such as emotions and the ways humans understand surrounding events and phenomena

(3) Ambient-AI: Generating added value by collecting information from our social environment such as factories, vehicles, and roads, and analyzing that information, for example, by applying AI technologies that detect abnormalities in machinery and control the flow of people in crowded areas

(4) Network-AI: Generating knowledge by connecting AI in various locations—similar to the concept of *two heads are better than one*—and enabling stable operations by finding failures in the network itself

The past year has seen rapid progress in AI initiatives among the NTT Group companies, with more than 60 press releases (in Japanese) related to corevo issued. In particular, new Agent-AI services have been developed based on natural language processing and speech recognition to generate new value by enabling computers to understand information produced by humans. For example, computers process information from users of machine translation systems to return translation results, analyze calls in a contact center system to rapidly comprehend the issue described by the customer, or act as reception robots that talk with customers and provide guidance information (**Fig. 1**).

Likewise, new Ambient-AI services have also been

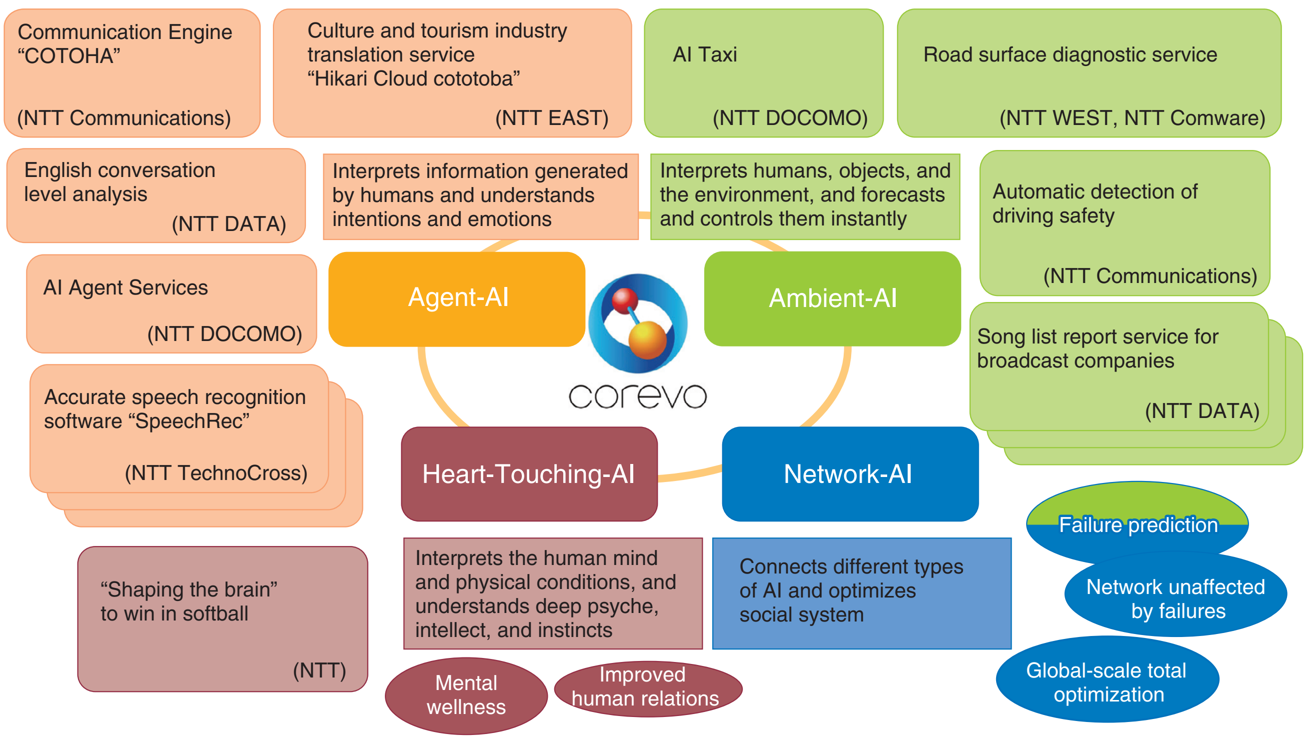


Fig. 1. AI technologies and services provided by the NTT Group.

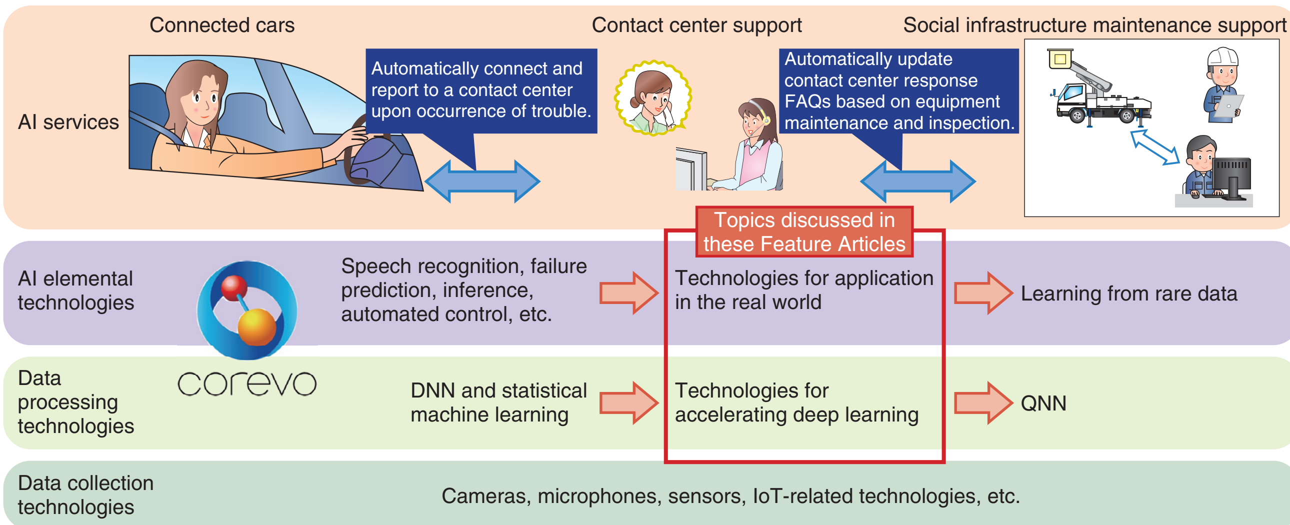
developed to generate added value through the observation of various objects and phenomena in the social environment. These include the prediction of factory and plant failures by analyzing images from cameras and sounds collected by sensors or microphones, discovery of mechanical failures by detecting abnormal sounds produced by machinery, and measurement of holes and cracks in roads and the level of wear of manholes.

AI performance has been reported to surpass that of humans in the world of games, such as in *go* and *shogi* (Japanese chess). However, the NTT Group is aiming for co-creation between humans and AI of services and applications and has therefore endeavored to develop AI technologies that support human social activities in the real world, as AI that extends—rather than competes with—human capabilities. These initiatives have also revealed certain issues that need to be addressed in applying AI technologies in the real world.

2. Shift to AI that takes an active role in the real world from AI that merely processes information

AI technologies developed recently are capable of collecting a vast amount of data and classifying the data based on internal data features through a method called machine learning. They also widely use deep neural networks (DNNs) and other deep learning methods for determining input-output correlations to produce the most probable output answer for newly input data. This is done through mechanisms for automatically learning relationships between input data and output data of correct answers provided by humans. These methods select the best solution from one type of combinatorial problem, making them very effective for selecting the best solution from among an extremely large number of combinations based on a fixed set of rules, such as in go or shogi.

Applying AI to the real world to address various problems that can potentially be solved using AI by simply collecting a large amount of data and finding the best state or solution is sometimes inefficient, though, because it requires a tremendously long computation time. For example, speech recognition technologies collect a large amount of human voice data and use DNN-based learning systems to correlate voice data with text transcribed from the voice data. This can be approached as an optimization problem for correlating human voice output with transcribed



FAQ: frequently asked question
IoT: Internet of Things

Fig. 2. Architecture of AI technologies as conceived by the NTT Group.

text such as by using AI to learn human voice data collected in a low-noise environment and identifying variations in sound between male and female voices or finding differences in sounds that are easy to pronounce for certain people using speech recognition in a low-noise environment.

The real-world environment where speech recognition is used, however, requires distinguishing the human voice amidst various types of noise such as sounds from the television or stereo when using robots in living rooms, and from stereo music, the sound of air conditioning, and road noise when using speech recognition inside vehicles. Since the types and levels of noise vary depending on the environment where speech recognition is used, simply collecting data and analyzing combinations would require an enormous amount of data and computing power. It is therefore necessary to perform learning by taking into account the possible sources of noise in the particular environment where speech recognition is used such us by controlling road noise or other sound disturbances. This will effectively increase the accuracy of speech recognition.

In addition, learning more data from the real world within a short period will make it possible to deliver AI systems that will become increasingly more useful for customers every day. The NTT Group, therefore, is also focusing on developing methods to accelerate the process of learning massive amounts of data. In addition to developing technologies to optimize learning through DNN, which is currently one of the main AI technologies, we are also developing quantum neural network (QNN) technologies for solving combinatorial optimization problems at speeds faster than that of DNN methods.

To offer the best services to users of AI services in the real world, we will continue to strive to make AI wiser by creating a mechanism for accurately finding relevant learning data and rapidly processing a larger amount of information. The Feature Articles in this issue introduce examples of AI technologies being developed by the NTT Group for solving customer problems in the real world and also present the current status of our initiatives to accelerate learning using DNN methods (**Fig. 2**).

3. AI elemental technologies for the consumer sector

Smart speakers and other agent devices are now becoming particularly popular in the consumer sector. Users can listen to music, turn on electronic appliances, and enjoy various other services by giving voice instructions to an agent device. The Feature Article in this issue entitled “Efforts to Enhance Far-field Speech Recognition” [1] introduces a speech recognition technology that accurately captures the user’s voice and is a key technology for agent devices.

Healthcare is another promising sector for targeting consumers. The NTT laboratories are engaged in research to better understand daily physical activity states and health conditions based on sensor information and medical and health data, as well as research to support disease prevention and health promotion based on the analysis of physical activity and health conditions. Another article in this issue, "Biosignal Processing Methods Targeting Healthcare Support Services," [2] introduces a technology for accurately estimating vital data from biosignals obtained from sensors, as a way to more thoroughly understand health conditions. The article entitled "Artificial Intelligence-based Health Management System: Unequally Spaced Medical Data Analysis" [3] introduces AI technologies that support blood sugar management based on clinical data from a university hospital, as part of research to support disease prevention and health promotion.

4. AI elemental technologies for the business sector

The first thing that comes to mind in applying AI for business is optimization of operations. NTT is engaged in research on complementing and drawing out human capabilities using AI.

For example, maintenance inspection of communication facilities and other infrastructures owned by the NTT Group is very labor-intensive and therefore urgently requires optimization. To address this issue, the NTT laboratories are pursuing research and development of ways to optimize maintenance and inspection operations. As an example, the article entitled "Automatic Degradation Estimation of Manhole Covers for Efficient Inspection via Vehicle-mounted Cameras" [4] introduces a technology to estimate the extent of manhole degradation based on images from vehicle-mounted cameras.

Also, the recent increase in inbound visitors to Japan has led to a surge in demand for translation of tourist and other information. The article entitled "Efforts toward Service Creation with Neural Machine Translation" [5] introduces initiatives to improve translation accuracy through actual field applications of AI-based automatic translation technologies.

Another highly promising field for AI business applications is data analytics. The NTT laboratories are working towards developing technologies for predicting and detecting phenomena that cannot be observed in advance, based on imperfect data. The article entitled "People Flow Prediction Technology for Crowd Navigation" [6] introduces a means of spatio-temporal prediction for simultaneously analyzing time and space phenomena.

5. Platform technologies underpinning AI

AI elemental technologies have become possible through complex and massive computational processing. The NTT laboratories are also developing the technologies underpinning these processing operations. Platform technologies are divided into the frameworks related to learning and analysis, hardware, and the technologies for improving the operability and convenience of the frameworks and hardware. The Feature Articles in this issue introduce technologies related to AI learning.

Although learning based on massive data is necessary for AI to derive the right answer, the process entails a number of problems such as the extremely high computation cost of learning and the lack of guarantee of convergence of learning results. The article entitled "Advanced Learning Technologies for Deep Learning" [7] introduces initiatives aimed at resolving these issues inherent to DNN learning processes.

6. Future directions

To further enhance the lifestyles and business operations of many customers through corevo technologies, the NTT Group will continue to work with different partners in pursuing the applications of corevo technologies in the real world.

References

[1] K. Matsui, T. Moriya, H. Itou, T. Fukutomi, S. Saito, Y. Shinohara, S. Kobashikawa, Y. Yamaguchi, and N. Harada, "Efforts to Enhance Far-field Speech Recognition," NTT Technical Review, Vol. 16, No. 8, pp. 12–17, 2018.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201808fa2.html

[2] K. Eguchi, R. Aoki, S. Shimauchi, A. Chiba, and N. Asanoma, "Biosignal Processing Methods Targeting Healthcare Support Services," NTT Technical Review, Vol. 16, No. 8, pp. 29–36, 2018.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201808fa5.html

[3] H. Kurasawa, A. Fujino, and K. Hayashi, "Artificial Intelligence-based Health Management System: Unequally Spaced Medical Data Analysis," NTT Technical Review, Vol. 16, No. 8, pp. 24–28, 2018.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201808fa4.html

[4] K. Murasaki, S. Ando, and T. Kinebuchi, "Automatic Degradation Estimation of Manhole Covers for Efficient Inspection via Vehicle-mounted Cameras," NTT Technical Review, Vol. 16, No. 8, pp. 18–23, 2018.

https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201808fa3.html
[5] S. Yoshida, M. Mizushima, and K. Tanaka, "Efforts toward Service Creation with Neural Machine Translation," NTT Technical Review, Vol. 16, No. 8, pp. 42–46, 2018.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201808fa7.html
[6] D. Sato, H. Shiohara, M. Miyamoto, and N. Ueda, "People Flow Prediction Technology for Crowd Navigation," NTT Technical Review, Vol. 16, No. 8, pp. 47–52, 2018.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201808fa8.html
[7] Y. Ida, S. Kanai, Y. Fujiwara, S. Yagi, and Y. Iida, "Advanced Learning Technologies for Deep Learning," NTT Technical Review, Vol. 16, No. 8, pp. 37–41, 2018.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201808fa6.html

Hideaki Ozawa

Vice President, NTT Media Intelligence Laboratories.

He received a Ph.D. in engineering from the Graduate School of Science and Technology, Keio University, Kanagawa, in 1992. After joining NTT in 1991, he engaged in research and practical application of multimedia processing technologies at NTT Human Interface Laboratories and Cyber Solution Laboratories, and provision of local multimedia information at Walkerplus, Inc. He also worked on the development and management of "goo" internet services including the search engine and healthcare services and was involved in establishing new services at NTT Resonant Inc. In addition to his current position at NTT Media Intelligence Laboratories, he has also served since May 2013 as president of NTT Resonant Technology, Inc., which develops software for Android/iOS.

Efforts to Enhance Far-field Speech Recognition

Kiyoaki Matsui, Takafumi Moriya, Hiroaki Itou, Takaaki Fukutomi, Shoichiro Saito, Yusuke Shinohara, Satoshi Kobashikawa, Yoshikazu Yamaguchi, and Noboru Harada

Abstract

We collected more than 15,000 hours of pseudo data to enhance training data for robust speech recognition by adding various impulse responses and noises to clean data and re-recording the clean data under actual environments. As a result, we reduced the rate of errors by 43% or more in both close and distant conditions. In this article, we introduce the various approaches that we researched, with the focus being on countering the distant conditions.

Keywords: far-field speech recognition, data augmentation, acoustic modeling

1. Far-field speech recognition

The use of speech recognition systems is increasing rapidly, and in line with this trend, communication agents such as smart speakers and voice dialog robots are being rapidly adopted, as are voice search services on smartphones. With search tasks using smartphones, utterances are often made in the immediate vicinity of the microphone, but when talking to robots or smart speakers, we must assume utterances will be made one to three meters away from the microphone. Furthermore, the speakers will exhibit individual differences in voice volume, and when a speaker speaks softly from some distance in a noisy environment, it is very difficult for current systems to correctly recognize the speech.

We researched various approaches to counter the noise environment of the home and to achieve robust speech recognition under both close and distant conditions regardless of noise or voice volume. Our aim was to improve system performance by enhancing training data and evaluation data by generating pseudo data and improving playback recording (or, pseudo recording), adjusting training parameters, and improving speech section detection (**Fig. 1**). Tests conducted using these approaches indicated that an error reduction rate of 43% or more was achieved in both close and distant conditions. These approaches are explained in the following sections.

2. Data augmentation by generation of pseudo data and speech samples

Having sufficient training data is necessary to improve system performance, but it is not always easy or practical to obtain the right kind of data. Data augmentation is a common approach to obtain a sufficient amount of data.

2.1 Augmentation of training data

The characteristics of the sound captured by the microphone will vary greatly with the speaker's proximity to the microphone. With close utterances, since the direct sound of the speaker dominates any echoes or noise, recognition can be performed with high accuracy regardless of the ambient noise environment. In contrast, with distant utterances, not only is the target utterance of the speaker affected by factors

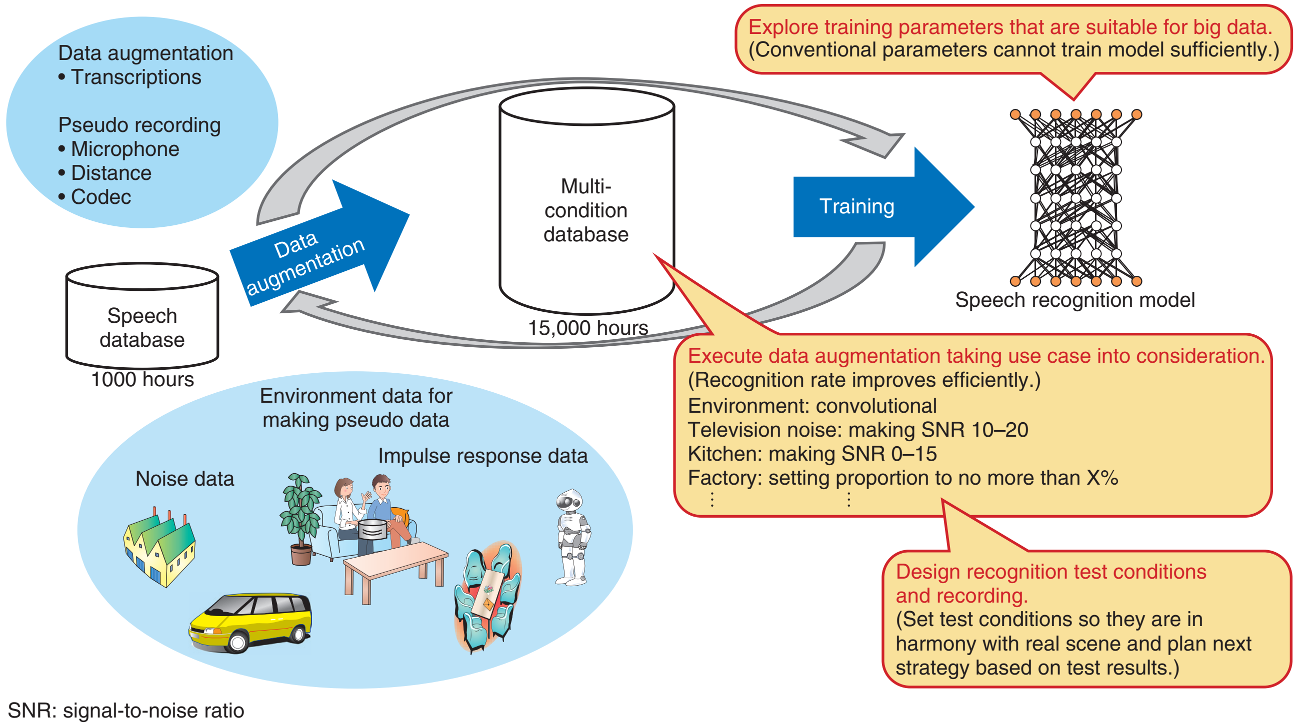


Fig. 1. Overview of our speech recognition enhancement approach.

such as the reverberation characteristics of the room and the attenuation due to distance, but it is also more likely that one or more noise sources that are closer to the microphone than the speaker will drown out the utterances, making recognition very difficult.

The mainstream approach at present is to build an acoustic model using deep neural network (DNN) technology, and it is possible to construct a robust acoustic model if adequate amounts of training data captured in various environments are available. Therefore, to construct acoustic models that are robust against various noises and speakers, it is important to cover as many diverse environments as possible. Normally, data are reinforced by transcribing speech recorded in various environments, but speech transcription takes too long, so gathering sufficient training data is problematic. One promising solution is to add various reverberation effects and noise to existing clean speech data, which yields a sufficient variety of pseudo data. This method of generating pseudo data is important not only for speech recognition but also for DNN training in all fields.

However, increasing the pseudo data will not necessarily yield higher accuracy. For example, with a communication agent that is to be used in the home, we can predict that situations where speaking occurs at distances exceeding 3 m are unlikely to occur. Also, because homes include furniture and sound-absorbing materials such as carpets and curtains, reverberation in the room will not be very strong. The noise environment can be expected to include the noise of television (TV), cooking sounds from the kitchen, and air conditioning. However, modern air conditioners are extremely quiet, and are therefore unlikely to drown out the speech. Thus, we can optimize speech recognition in the target environment by properly reacting to the reverberation strength, type of noise, and volume of noise so as not to generate unrealistic pseudo data.

In addition to generating pseudo data, we also make pseudo recordings by creating corrupted speech as it would be captured by the microphone. This method is very effective when the impulse response cannot be captured due to specifications of the microphone, or when it is necessary to create a more complex reverberation/noise environment. The playback recording configuration is shown in **Fig. 2**. The height, position, and angle of the speaker used for pseudo recording is set considering the practical environment. For example, if the user is assumed to speak while sitting on a

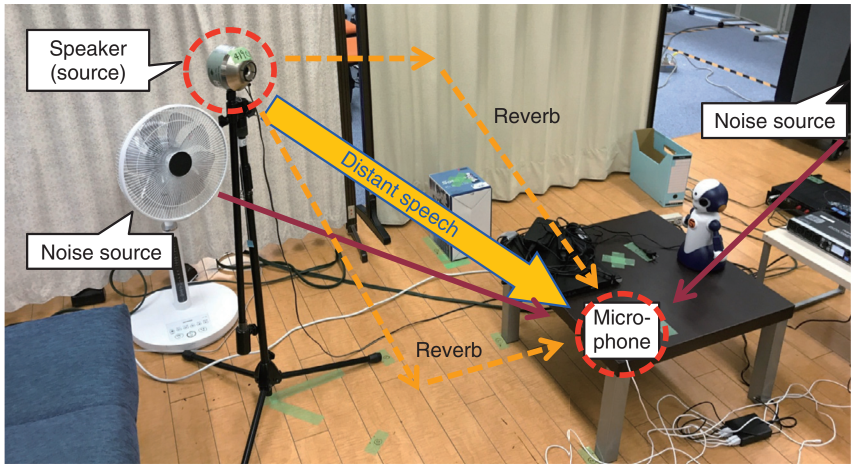


Fig. 2. Pseudo recording setup.

sofa, the speaker's height should match the height of the user's mouth. In addition, the speaker for noise playback is set to replicate the position of the actual noise source. The microphone used to capture the pseudo recording should be set at multiple positions at the same time, as recording in parallel enables more efficient data enhancement. To capture different characteristics, the microphone position can be set at not only different distances to the speaker, but also at different angles with respect to the speaker, position in the room, and other details. Pseudo recording yields training data of higher quality than simulated data because the results are captured after the sounds actually traverse space.

In this way, in addition to the data provided by normal transcription, the training data are reinforced by generating the pseudo data and pseudo recording, and it is possible to increase the training sound data, which originally amounted to only about 1000 hours, to more than 15,000 hours. In addition, at the first stage of speech recognition, utterance sections are accurately extracted by speech segment detection, which greatly improves recognition efficiency. The models used at this time were reinforced by similar data enhancement.

2.2 Creation of evaluation data

After the model is created from training data, evaluation data are needed to evaluate the model's performance. The ideal evaluation data consist of real data captured in the user's environment, but obtaining such ideal data is impractical, and in most cases evaluation data obtained in a nearly equal environment will be used. Accordingly, pseudo recording is an effective solution to the problem of creating evaluation data.

For this research, we prepared and evaluated a total of 36 patterns created by pseudo recordings using the speech of 20 subjects and setting different conditions of reverberation, distance, voice volume, and noise. The resulting data make it possible to assess system performance in environments for which evaluation data are unavailable. The resulting data revealed the strong and weak points of the trained model, which will make training the next model much more efficient.

3. Determining training parameters

The accuracy of the DNN acoustic model is greatly affected by the training parameters used, and these parameters are very difficult to set to maximize system performance. If the training set is very large scale, say more than 15,000 hours, parameters that are conventionally appropriate will not always work, and it is necessary to review what constitutes the optimal parameter setting.

There are two goals with parameter adjustment: speed enhancement and higher accuracy. The former can be achieved by increasing the number of parallel GPUs (graphics processing units) used to implement training and increasing the number of samples used to calculate each process in training (batch size). However, if the batch size is increased too much, any sample that is used will be close to some entries when the average is determined for each process. This

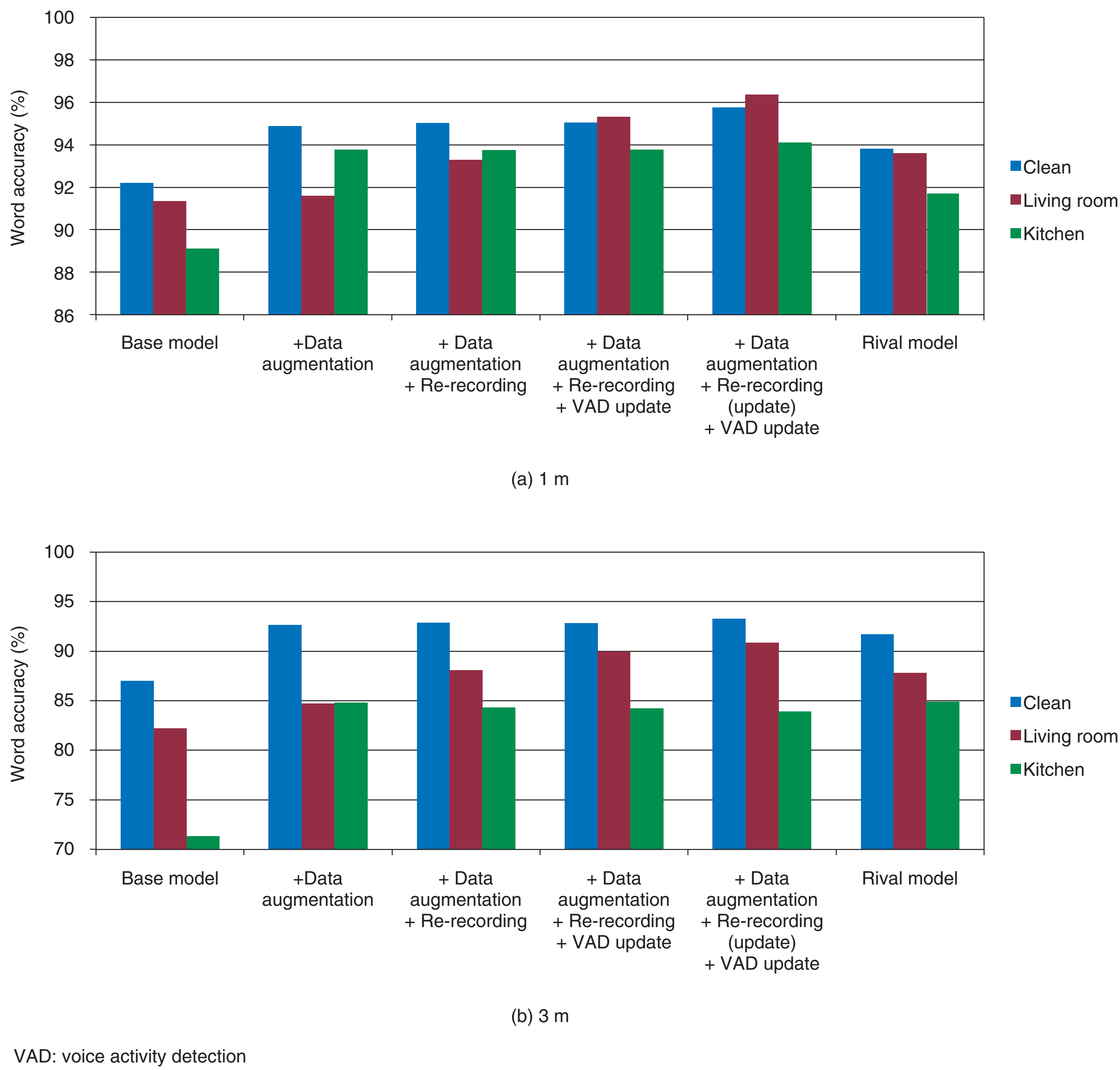


Fig. 3. Effectiveness of data augmentation.

means that convergence may occur even if training is insufficient.

As for the latter, to incorporate data as diverse as possible into training, it is necessary to increase the amount of data processed in each epoch (one training iteration) or make it easier to select specific condition data from the corpus used for training, which is another of our advances.

By adjusting the training parameters, we were able to develop a more accurate acoustic model while keeping the training time the same as it was when only 1000 hours of speech were used for training.

4. Recognition accuracy under different noise and distance values

In this section, we describe how well our data enhancement and model training techniques increase system performance. The performance results of the base model and the model using the proposed techniques are shown in the graph in **Fig. 3**. The training parameters were optimized for each condition. As can be seen in this graph, pseudo data generation greatly improves the recognition rate in clean and kitchen environments with distances to the microphone of 1 m and 3 m, which confirms the value of data reinforcement. In contrast, in the living room environment, no significant improvement in accuracy was seen. This is most likely because the TV in the living room environment puts out voices, and it is very difficult to judge whether such extraneous speech is noise or not.

Our technique of pseudo recording is very effective in coping with this living room noise, as it enables training to include more practical environmental information and strengthens the detection of speech segments. The result is a strong improvement in recognition rate in the living room environment while maintaining the excellent recognition performance in the clean and kitchen environments. The techniques enable us to build an acoustic model that is significantly better than all conventional alternatives.

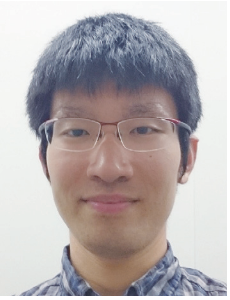
5. Future work

Our research has yielded acoustic models that offer high speech recognition accuracy regardless of the presence or absence of noise in the close and distant speaker conditions. However, in environments where the speaker's voice is mixed with the speech of others, erroneous speech recognition results are generated, so that the system reacts incorrectly or recognizes the user's utterance wrongly. We believe that this problem can be tackled by using direction-based sound collection technology such as the intelligent microphone [1], which will enable robust recognition of the target speaker's voice even in very noisy environments.

This study focused on quite short utterances such as those used for conducting search tasks. Many tasks remain if we are to reliably recognize natural speech as encountered at conferences and call centers. We will continue working to enhance the performance of our techniques to cover these natural utterances.

Reference

[1] T. Oba, K. Kobayashi, H. Uematsu, T. Asami, K. Niwa, N. Kamado, T. Kawase, and T. Hori, "Media Processing Technology for Business Task Support," NTT Technical Review, Vol. 13, No. 4, 2015. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201504fa6.html



Kiyoaki Matsui

Researcher, Audio, Speech, and Language Media Project, NTT Media Intelligence Laboratories.

He received a B.E. and M.E. from Tohoku University, Miyagi, in 2013 and 2015. Since joining NTT in 2015, he has been researching speech recognition, including voice activity detection and speech enhancement. He is a member of the Acoustical Society of Japan (ASJ).

Takafumi Moriya

Researcher, Audio, Speech, and Language Media Project, NTT Media Intelligence Laboratories.

He received a B.E. from Chiba Institute of Technology in 2014, and an M.E. from Tokyo Institute of Technology in 2016. Since joining NTT in 2016, he has been researching speech recognition, including acoustic modeling. He is a member of ASJ.

Hiroaki Itou

Researcher, Audio, Speech, and Language Media Project, NTT Media Intelligence Laboratories.

He received a B.E. and M.E. from Nagoya University, Aichi, in 2007 and 2009. Since joining NTT in 2009, he has been researching acoustic signal processing for sound reproduction and speech recognition. He was awarded the Awaya Prize by ASJ in 2011. He is a member of ASJ and the Institute of Electronics, Information and Communication Engineers (IEICE).

Takaaki Fukutomi

Research Engineer, Audio, Speech, and Language Media Project, NTT Media Intelligence Laboratories.

He received a B.E. and M.E. in design from Kyushu University, Fukuoka, in 2006 and 2008. Since joining NTT in 2008, he has been researching speech recognition, including voice activity detection and speech enhancement. He received the Technical Development Award from ASJ in 2014. He is a member of ASJ.

Shoichiro Saito

Research Engineer, Audio, Speech, and Language Media Project, NTT Media Intelligence Laboratories.

He received a B.E. and M.E. from the University of Tokyo in 2005 and 2007. Since joining NTT in 2007, he has been researching acoustic echo cancellers and hands-free telephone terminals. He is a member of the Institute of Electrical and Electronics Engineers (IEEE), IEICE, and ASJ.

Yusuke Shinohara

Senior Research Engineer, Audio, Speech, and Language Media Project, NTT Media Intelligence Laboratories.

He received a B.S. and M.S. in mechano-informatics from the University of Tokyo in 2002 and 2004. From 2004 to 2017, he was a research scientist at Toshiba Corporate Research and Development Center, Kanagawa. Since joining NTT in 2017, he has been involved in research and development of speech recognition, in particular, acoustic modeling with various types of deep neural networks. His research interests include speech recognition, computer vision, and machine training. He received the IEICE Pattern Recognition and Media Understanding (PRMU) Young Researcher Award in 2004, and the ASJ Awaya Prize Young Researcher Award in 2013. He has served as a committee member (2013–2015, 2017–present) and a secretary (2015–2017) for the Information Processing Society of Japan Special Interest Group (IPSJ SIG) on Spoken Language Processing. He is a member of IEEE, IEICE, ASJ, and IPSJ.

Satoshi Kobashikawa

Senior Research Engineer, Audio, Speech, and Language Media Project, NTT Media Intelligence Laboratories.

He received a B.E., M.E., and Ph.D. from the University of Tokyo in 2000, 2002, and 2013. Since joining NTT in 2002, he has been researching speech recognition and spoken language processing. He received the Kiyasu Special Industrial Achievement Award in 2011 from IPSJ, the 58th Maejima Hisoka Award from the Tsushinbunka Association in 2012, and the 54th Sato Paper Award from ASJ in 2013. He is a member of ASJ, IPSJ, and IEICE.

Yoshikazu Yamaguchi

Senior Research Engineer, Audio, Speech, and Language Media Project, NTT Media Intelligence Laboratories.

He received a B.E. and M.E. in electrical engineering from Osaka Prefecture University in 1993 and 1995. Since joining NTT in 1995, he has been researching automatic speech recognition technologies.

Noboru Harada

Senior Research Scientist, Supervisor, Audio, Speech, and Language Media Project, NTT Media Intelligence Laboratories.

He received a B.S. and M.S. from the Department of Computer Science and Systems Engineering of Kyushu Institute of Technology, Fukuoka, in 1995 and 1997, and he later received a Ph.D. from University of Tsukuba. Since joining NTT in 1997, he has been researching lossless audio coding, high-efficiency coding of speech and audio, and their applications. He is an editor of the International Organization for Standardization/International Electrotechnical Commission (ISO/ IEC) 23000-6:2009 Professional Archival Application Format, ISO/IEC 14496-5:2001/ Amd.10:2007 reference software MPEG-4 Audio Lossless Coding and ITU-T (International Telecommunication Union - Telecommunication Standardization Sector) G.711.0, and has contributed to 3GPP (3rd Generation Partnership Project) Enhanced Voice Services. He is a member of IEICE, ASJ, IPSJ, the Audio Engineering Society, and IEEE.

Automatic Degradation Estimation of Manhole Covers for Efficient Inspection via Vehicle-mounted Cameras

Kazuhiko Murasaki, Shingo Ando, and Tetsuya Kinebuchi

Abstract

Automatic inspection of infrastructure using artificial intelligence and robot technology has increased in recent years because of the aging infrastructure and a lack of personnel with expertise in this area. NTT is developing various technologies for use with infrastructure equipment that the NTT Group owns. In this article, we describe a technique for inspecting manhole covers and a method for automatically estimating degradation of the covers via vehicle-mounted cameras for safe and efficient inspection.

Keywords: smart maintenance, image processing, deep learning

1. Problems in manhole cover inspection

Manholes are positioned to enable maintenance and management of the water supply, sewerage, network transmission lines, and telephone lines. Many manhole covers are exposed on roads all over the world. There are more than 14 million manholes in Japan alone, and a large number of them are driven on by vehicles every day [1]. Manhole covers are typically made of iron. The iron plate is slippery compared to the asphalt road surface, so many iron covers are engraved with a concavo-convex pattern for anti-slip properties. The manhole covers used by NTT are shown as an example in **Fig. 1(a)**. A deep concavo-convex pattern is engraved on the surface, which helps to prevent slipping accidents.

However, the pattern on the cover gradually wears as automobiles pass over the cover, so manhole covers that have been in operation for a long time will potentially lose the effect of slip prevention due to wear, as shown in **Fig. 1(b)**. In addition to the wear of the surface pattern, there may be a step difference between the cover and the frame supporting it due to aging deterioration, as in **Fig. 1(c)**.

All companies that manage manholes, including the NTT Group, regularly inspect the degree of deterioration of manhole covers to prevent the occurrence of accidents caused by wear or step differences and then replace deteriorated covers as necessary. However, this requires a lot of manpower and time. Skilled workers must inspect the covers carefully, which sometimes requires part of the road to be blocked off in order to safely inspect them. This often results in traffic delays and inconvenience to road users.

At NTT, we have been developing a way to estimate the degree of deterioration of manhole covers using a vehicle-mounted camera. The objective is to construct a system to automatically inspect manhole covers simply by driving over them with the camera-equipped vehicle. We have combined image recognition technology and three-dimensional (3D) geometric transformation technology to achieve robust degradation degree estimation under various photographic conditions.

(a) NTT's manhole cover (b) Wearing of the cover surface (c) Step between cover and frame

Fig. 1. Deterioration of manhole covers.

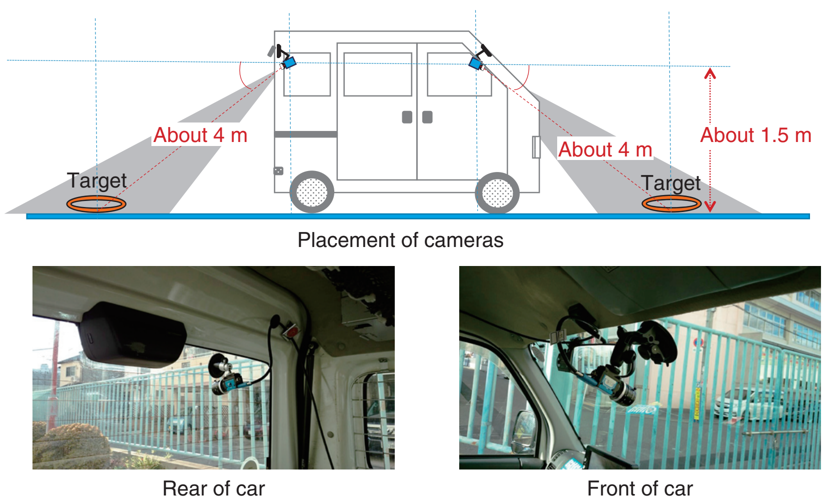


Fig. 2. Inspection system using vehicle-mounted cameras.

2. Manhole cover inspection system using vehicle-mounted cameras

The inspection system we propose is shown in **Fig. 2**. Cameras are installed in the front and rear of the vehicle used for the inspections, and these cameras continuously take pictures of the road while the vehicle is traveling. When a manhole cover is detected from the camera image, the deterioration degree estimation is applied, and the degree of wear of the target cover and the level difference from the supporting frame or road surface are obtained.

Various methods exist to estimate the shape of an object by remote sensing other than image processing, for example, 3D sensing. However, 3D sensing using a laser range finder is very costly as it involves a high unit price for the sensor and special equipment needed to capture 3D data.

In contrast, the proposed inspection system uses relatively inexpensive cameras and an ordinary car. Therefore, this inspection system can be simply and economically constructed using existing vehicles and deployed anywhere the manhole covers are in use all over Japan. In addition, since the deterioration degree can be estimated from just one image, and no special photographic technique is required, it is possible to inspect manhole covers using vehicles intended for non-inspection related work simply by equipping them with cameras.

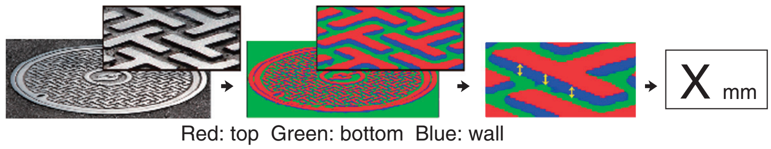


Fig. 3. Process flow of deterioration degree estimation of manhole covers.

3. Estimation of manhole cover degradation by image recognition

The process flow of the proposed method of manhole cover deterioration estimation is shown in **Fig. 3** [2]. First, we extract the pattern wall surface area for the entire input image. Artificial intelligence is used to recognize the wall surfaces learned beforehand. Then we allocate three plane classes consisting of the top surface, bottom surface, and wall surface of the step to each pixel of the input image. The recognition process results in a map of the plane classes, as shown in Fig. 3(b). Next, the height of the step indicated by each wall from the obtained plane class map is estimated. With the focal length of the camera and the diameter of the manhole cover, it is possible to obtain the step height from the wall height in the image by applying 3D geometry. The height of the step is calculated for each of the recognized wall surfaces. The step height indicates the degree of wear of the anti-slip pattern on the cover and the difference in levels between the edge of the cover and the supporting frame or the road surface.

The approach used in the conventional inspection method [3] involves directly learning the relationship between image features and the degree of degradation and then estimating the degree of degradation for each image region. In this approach, however, it is necessary to prepare learning data covering a large variety of images. The appearance of manhole covers changes depending on the illumination conditions, the level of contamination of the covers, and the angle of the camera. The conventional approach thus requires training data obtained in a variety of illumination conditions and shooting angles.

In contrast, with the proposed method, efficient learning is achieved with less data by effectively combining 3D geometry and machine learning, which is used to classify only the three plane classes previously mentioned. By estimating the image capturing angle for each image by 3D geometry and by estimating the step height according to the image, we can not only suppress any effects due to a difference in the image capturing angle but also estimate the step height outside the range of training data. It is not easy to collect a lot of degraded manhole covers for learning, so another advantage of our method is that it is possible to recognize the degradation state accurately even when there are few degraded samples contained in the training data.

4. Wall surface area extraction by deep learning

We utilize a convolutional neural network, the type of deep neural network used most often for image recognition, to extract the wall area. In particular, we use a network that recognizes each pixel of the input image [4] and classifies which kind of plane each pixel belongs to. The network outputs labels for each pixel based on the image features around the target pixel. Its parameters are trained by combinations of the input image and the map of the plane class label.

We used three classification labels: top, bottom, and wall, which make up the step shape, and constructed appropriate training data manually. The map of class labels used for training is shown in **Fig. 4**. The red parts correspond to the top area of the step, the green correspond to the bottom area, and the blue correspond to the wall area. With a sufficient quantity of this kind of data, the model can be trained to extract walls under various conditions. When a new image is classified, the area of the manhole cover is extracted from the photographed image, and the trained model outputs an appropriate wall area map from it.

5. Size estimation of wall surface

After the wall area map is obtained by extracting the wall area, the step height of manhole covers actually in use is estimated for each wall surface area in the wall map. We select a pixel at the top of the wall

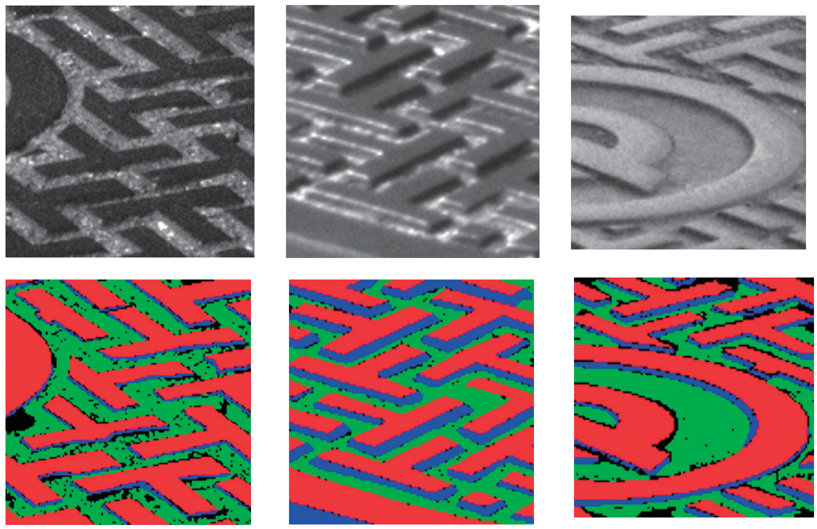


Fig. 4. Examples of label maps for training.

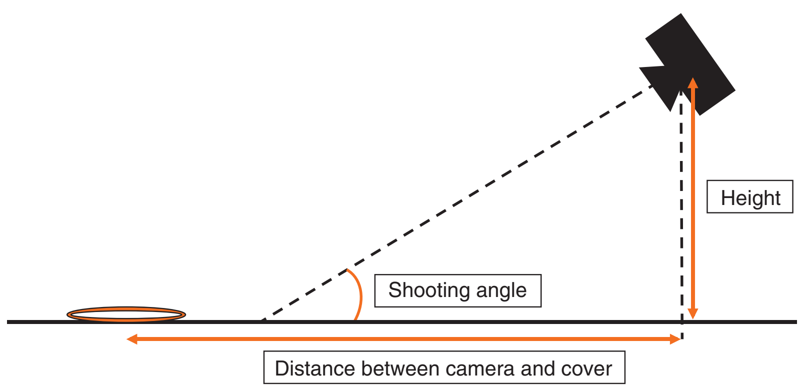


Fig. 5. Computable geometric relationship between the camera and cover.

area—the starting point of height estimation—and find the pixels that continue in the wall area along the line perpendicular to the plane of the manhole cover from the starting point. The pixels indicate the line segment showing the height of the wall area. To calculate the actual length of this line, it is necessary to find the geometric relationship between the camera and the cover. Utilizing the fact that the focal length of the camera and the diameter of the cover are known, and the parameter of the ellipse formed by projecting the cover onto the image is easily calculated, we can obtain the angle of the camera with respect to the plane including the manhole cover and the height of the camera position. The parameters determined by 3D geometry are shown in **Fig. 5**. By estimating these parameters for each image, we can robustly estimate the height of the step even if the position of the camera and the installation condition of the cover are changed.

6. Application to real images of manhole covers

We verified the accuracy of the deterioration degree estimation method using actual NTT manhole covers installed in roads.

The results of wall area extraction for manhole cover images in different illumination conditions are shown in **Fig. 6**. We can see that the wall area is robustly extracted even when the illumination condition differs or when part of the cover is shaded. The

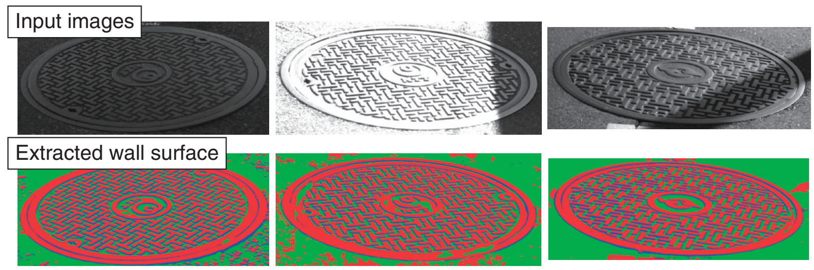


Fig. 6. Examples of wall surface extraction results in various lighting conditions.

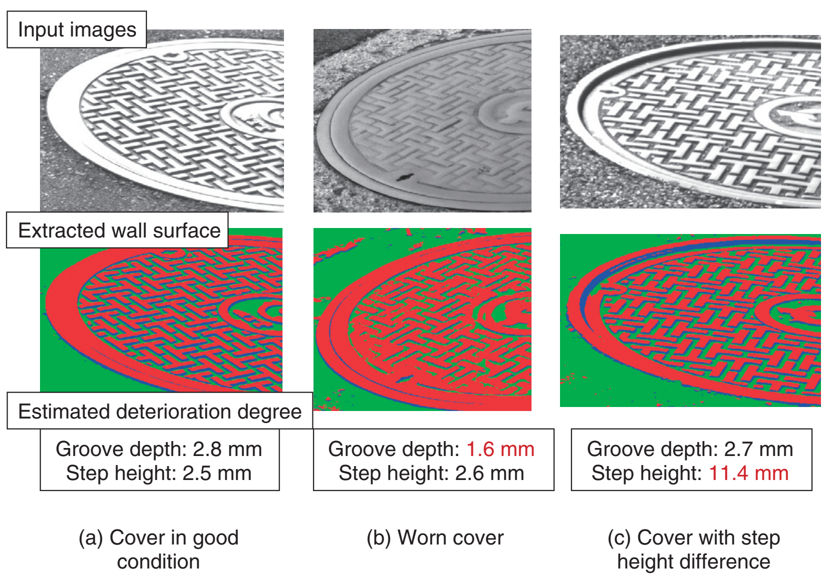


Fig. 7. Examples of deterioration degree estimation.

results estimated using the proposed method for deteriorated manhole covers are shown in **Fig. 7**. It is clear that the worn cover has a smaller extracted wall area, and the estimated depth of the grooves is shallower. Additionally, a large wall area is extracted from the cover where a gap appears relative to the supporting frame, and the estimated step height is also high.

We compared the results of manually measuring the groove depth of the surface pattern of covers using calipers with estimated results using our proposed method for 170 cover images captured with a fixed camera. An average error of 0.55 mm compared to caliper measurement was achieved with our method.

Although it is inevitable that estimation errors will sometimes occur in degradation degree estimation by image recognition, we can still use this method for screening purposes in actual inspections. This involves extensively extracting covers that may have some deterioration from the many manhole covers to be inspected so as not to overlook any and then having skilled workers conduct a detailed visual inspection of them. With this method, an excessive number of deteriorated covers are extracted, but the possibility of missing any deteriorated covers is greatly reduced.

We conducted experiments to carry out screening using the proposed method for 1000 covers installed in roads. The experimental condition (deterioration criteria) was set to extensively extract worn covers so as not to miss any degraded ones. The results indicated that about 40% of the covers that were not actually deteriorated were judged to be deteriorated by the proposed method, while about 60% of the covers were accurately judged to have no deterioration and were thus removed from the inspection work target.

7. Future work

We have established a technology to robustly estimate the degree of manhole cover degradation by recognizing plane classes that make up the concavo-convex shape of manhole covers from images photographed under various conditions. In the future, we aim to implement this method at an actual inspection site, evaluate the accuracy of degradation recognition at the inspection site, and work on solving any problems in operation.

References

[1] The Association for the Promotion of Next-generation High-quality Ground Manholes (in Japanese), http://www.kouhinigm.jp/gm/ul.html

[2] K. Murasaki, S. Ando, and T. Kinebuchi, "Manhole Cover Wearing Estimation via Vertical Plane Segmentation," IEICE Tech. Rep., Vol. 117, No. 210, pp. 205–212, 2017.

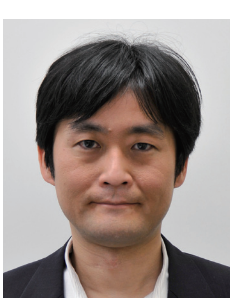
[3] K. Murasaki, K. Sudo, and Y. Taniguchi, "Manhole Cover Wearing Detection by Photo-taking," Trans. of the Society of Instrument and Control Engineers, Vol. 51, No. 12, pp. 814–821, 2015.

[4] A. Bansal, B. Russell, and A. Gupta, "Marr Revisited: 2D-3D Alignment via Surface Normal Prediction," Proc. of 29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016), pp. 5965–5974, Las Vegas, USA, June 2016.

Kazuhiko Murasaki

Researcher, Visual Media Project, NTT Media Intelligence Laboratories.

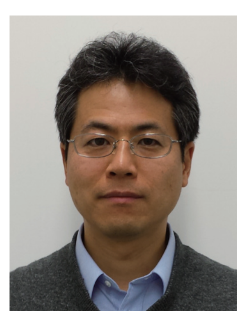
He received a B.E. in 2009 and a Master of Information Science and Technology in 2011 from the University of Tokyo. He joined NTT in 2011 and has been studying image recognition, particularly semantic segmentation, boundary detection, and degradation detection of infrastructure. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) of Japan.



Shingo Ando

Senior Research Engineer, Supervisor, Visual Media Project, NTT Media Intelligence Laboratories.

He received a B.E. in electrical engineering from Keio University, Kanagawa, in 1998 and a Ph.D. in engineering from Keio University in 2003. He joined NTT in 2003. He has been engaged in research and practical application development in the fields of image processing, pattern recognition, and digital watermarks. He is a member of IEICE and the Institute of Image Information and Television Engineers.



Tetsuya Kinebuchi

Senior Research Engineer, Supervisor, Visual Media Project, NTT Media Intelligence Laboratories.

He received an M.S. in physics from Tohoku University, Miyagi, in 1997. He joined NTT in 1997. His research interests include pattern recognition and image processing. He is a member of IEICE.

Artificial Intelligence-based Health Management System: Unequally Spaced Medical Data Analysis

Hisashi Kurasawa, Akinori Fujino, and Katsuyoshi Hayashi

Abstract

We are developing an artificial intelligence (AI)-based health management system that suggests efficient and effective interventions for keeping people healthy based on the potential risk of disease predicted by using AI. We proposed a new feature extraction model for unequally spaced medical data that improves the disease risk prediction. We introduce the model in detail in this article and also describe an application of the model that demonstrates the improved prediction accuracy.

Keywords: unequally spaced medical data, diabetes, poor glycemic control

1. Importance of lifestyle-related disease prevention

Lifestyle-related diseases such as type 2 diabetes, dyslipidemia, and hypertension are defined as diseases largely caused by factors such as an unhealthy life style, lack of physical activity and sleep, and excessive alcohol intake. In Japan, treatment of patients with lifestyle-related diseases accounts for 30% of medical expenses, and the mortality rate from such diseases is 60%. Therefore, prevention of lifestyle-related diseases is one of the most important issues for extending the human health span, which refers to the length of time a person is *healthy*—not just alive.

It is well known that interventions to patients at an early stage of a disease are effective in preventing the onset and progression of the disease, and many intervention programs such as Specific Health Checkups and Specific Health Guidance [1] have been implemented in Japan. However, such programs incur large costs for the government and health insurance providers. Therefore, more efficient and effective interventions for keeping people healthy are needed.

2. Artificial intelligence (AI)-based health management system

We are developing an AI-based health management system that suggests efficient and effective interventions for keeping people healthy based on their risk of disease predicted using AI. The concept of the system is illustrated in **Fig. 1**. The system first gathers health data on a user, for example, electronic health records (EHRs) obtained from a clinic and self-monitoring records measured at home. Then the system predicts the risk of each disease using the records and AI. Finally, it prepares a plan for efficient and effective intervention based on the prediction results and suggests the intervention via clinicians, wearable devices, or robots.

NTT's accumulated knowledge in data science was used to full advantage in developing the system. For example, NTT, in collaboration with the University of Tokyo Hospital, made use of knowledge in the area of human behavior analysis to address patients' treatment behavior and successfully predict possible missed scheduled clinical appointments [2]. Some analysis technologies used in the system have grown from basic research in various fields outside medical

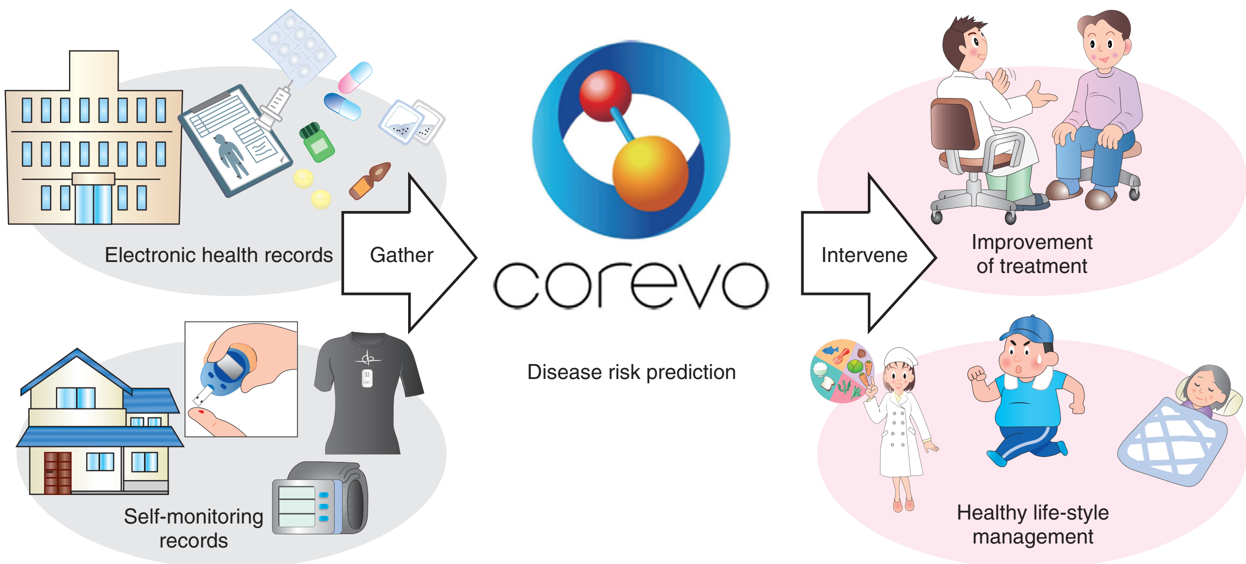


Fig. 1. AI-based health management system.

science.

In this article, we introduce a feature extraction model for unequally spaced medical data. This model improves the disease risk prediction of the system using EHRs and self-monitoring records with unequal intervals. We also describe an application where the model improved the prediction accuracy of poor glycemic control of patients with diabetes.

3. Challenge of analyzing unequally spaced medical data

Medical data often consist of unequally spaced values. In general, lab tests are done and prescriptions are issued during clinical visits. The intervals of clinical visits depend on the disease condition and the adherence to treatment and therefore often vary. When users forget to take their self-monitoring records, these intervals also vary.

The general method of disease risk prediction using unequally spaced values consists of four processes: quantization, completion, feature extraction, and classification as shown in **Fig. 2**.

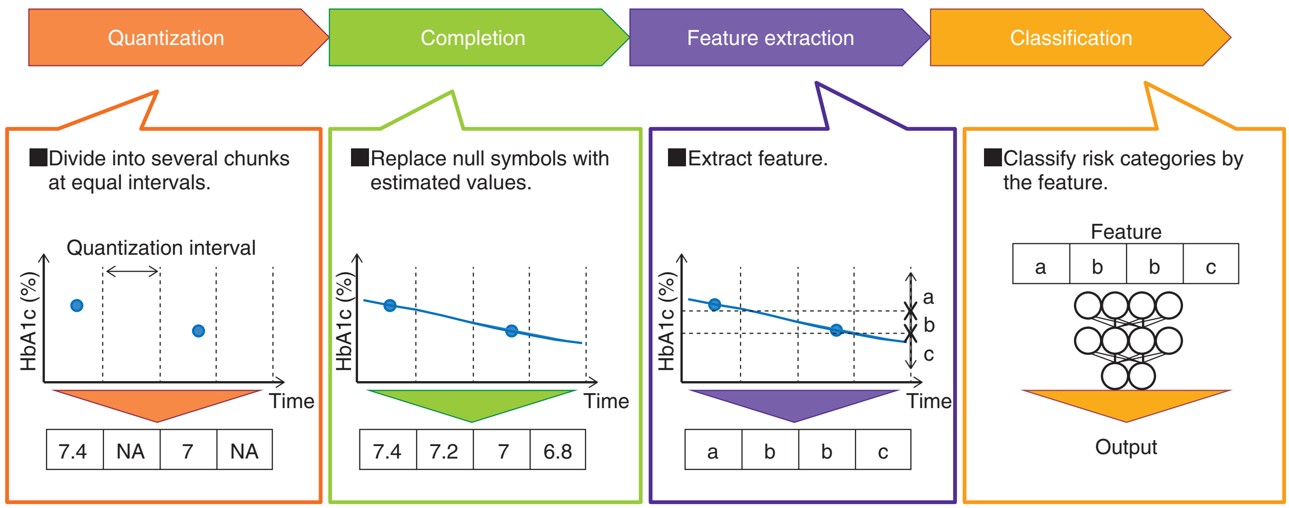
In the quantization process, unequally spaced data are divided into several chunks at equal intervals and converted into a quantized vector. Each chunk that includes values is filled with a representative value such as the average. In contrast, each chunk with no value is filled with a null symbol. For example, when data are divided into one-day intervals, chunks for days with no self-monitoring records are filled with null symbols.

In the completion process, each null symbol is replaced with an estimated value by using an interpolation method, and the quantized vector is converted into an estimated equally spaced vector. The linear interpolation is used for estimating values when the trends in values are approximately expressed by a linear function.

In the feature extraction process, features representing the estimated equally spaced data are extracted. For example, SAX (symbolic aggregate approximation) is used for transforming time series data into a character sequence. Matrix decomposition is used for transforming a vector into a lower-dimensional vector.

In the classification process, a model that receives the features and outputs the disease risk is constructed. There are many classification methods, such as SVM (support vector machine) and logistic regression.

However, a problem arises in the sequence from the completion to the feature extraction when using the general method. The feature does not represent the original quantized vector but rather, the estimated equally spaced vector. Any noise and errors in the completion process may be included in the feature, which can result in lower prediction accuracy. Feature extraction from the original quantized vector therefore involves the challenge of analyzing



HbA1c: Hemoglobin A1c is a marker used to measure blood sugar (glucose) levels over the previous three months.

Fig. 2. General method of analyzing unequally spaced data.

unequally spaced medical data. A method for tackling this challenge is described in the next section.

4. Feature extraction model for unequally spaced medical data

We developed a feature extraction model that skips over the completion process and directly outputs a feature from the quantized vector, as shown in **Fig. 3**. The model uses an autoencoder, which is a data compression technique using a neural network where the difference between the input and output vectors measured by a loss function is minimized by an optimizer, and the vector in the middle layer is used as a feature.

The proposed model modified the loss function of the autoencoder. The model minimizes the modified loss function $L(w)$ and learns the parameters w.

Here, $L(w)$ is defined as:

$$L(w) = b_n \cdot (v_n - o_n(w)),$$

$$o_n(w) = f(b_n, v_n; w),$$

where b is a Boolean vector representing the i-th value of the quantized vector, v is a null symbol (b_i = 0) or not (b_i = 1), and f is a function by the neural network. The loss function is designed to exclude the null symbols. After the minimization, the vector in the middle layer is used as a feature.

The proposed model outputs a feature with a uniform dimension even if the number of null symbols in the quantized vectors varies.

5. Application: prediction of poor glycemic control

Type 2 diabetes is one of the most common lifestyle-related diseases, and there are approximately 10 million diabetic patients in Japan. It is vital for diabetic patients to control their blood glucose level to avoid the complications of severe diabetes. We predicted poor glycemic control in patients with diabetes who needed more interventions in collaboration with the University of Tokyo Hospital [3]. We used the feature extraction model for unequally spaced medical data for the prediction.

We constructed a prediction model using EHRs from the University of Tokyo Hospital that included over 7000 diabetic patients. The intervals of lab tests varied among patients. The average interval of HbA1c (hemoglobin A1c) tests was 5.9 weeks with a standard deviation of 2.6 weeks. The ROC AUC (area under the receiver operating characteristic curve) of the prediction without the feature extraction model was 0.72. This value increased to 0.80 by using the feature extraction model. We therefore confirmed that the model improved prediction accuracy.

6. Future development

We are promoting the development of the AI-based health management system to contribute to efficient and effective interventions that help keep patients healthy. We will improve the core technologies in the

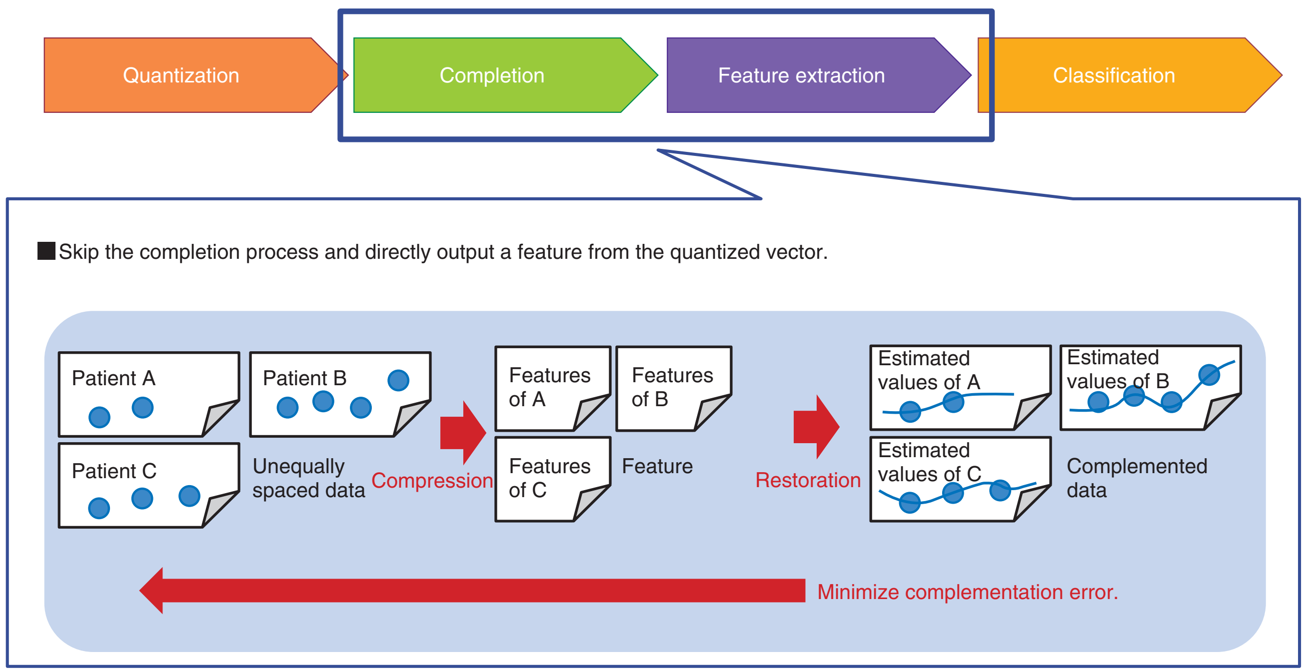


Fig. 3. General method of analyzing unequally spaced data.

system by utilizing standardized medical data stored in the EHR systems of hospitals and measured by wearable devices. We will also take advantage of the knowledge of robotics NTT has accumulated [4] and develop a novel intervention method through networked robots and devices.

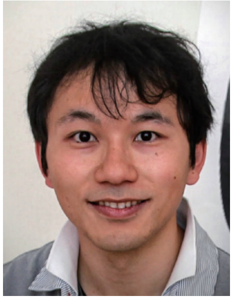
References

[1] Ministry of Health, Labour and Welfare, "Specific Health Checkups and Specific Health Guidance."
http://www.mhlw.go.jp/english/wp/wp-hw3/dl/2-007.pdf

[2] H. Kurasawa, A. Fujino, and K. Hayashi, "Predicting Patients' Treatment Behavior by Medical Data Analysis Using Machine Learning Technique," NTT Technical Review, Vol. 15, No. 8, 2017.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201708fa7.html

[3] H. Kurasawa, K. Hayashi, A. Fujino, T. Haga, K. Waki, T. Noguchi, and K. Ohe, "Machine Learning-based Prediction of Poor Glycemic Control in Patients with Diabetes," Proc. of the 37th Joint Conference on Medical Informatics, Osaka, Japan, Nov. 2017 (in Japanese).

[4] T. Yamada and H. Yoshikawa, "Cloud-based Interaction Control Technologies for Robotics Integrated Development Environment (R-env^{TM})," NTT Technical Review, Vol. 14, No. 5, 2016.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201605fa3.html



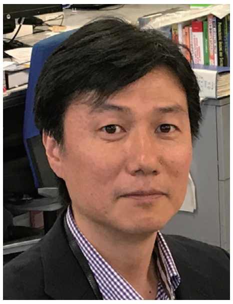
Hisashi Kurasawa
Research Engineer, Networked Robot & Gadget Project, NTT Service Evolution Laboratories*.

He received a B.E., M.E., and Ph.D. in information science and technology from the University of Tokyo in 2006, 2008, and 2011. He joined NTT Network Innovation Laboratories in 2011, where he studied sensor data mining. He joined a Medical AI research project in 2013 and studied clinical condition prediction models. He is currently studying disease risk models at NTT Service Evolution Laboratories. He is a member of the Institute of Electronics, Information and Communication Engineers, the Information Processing Society of Japan, the Japan Association for Medical Informatics, and the Database Society of Japan.

*He moved to NTT DOCOMO in July 2018.

Akinori Fujino
Senior Research Scientist, Learning and Intelligent Systems Research Group, NTT Communication Science Laboratories.

He received a B.E. and M.E. in precision engineering from Kyoto University in 1995 and 1997, and a Ph.D. in informatics from Kyoto University in 2009. He joined NTT in 1997. His current research interests include machine learning and knowledge discovery from complex data.



Katsuyoshi Hayashi
Senior Research Engineer, Social Device Technology Laboratory, Device Technology Laboratories and Research and Development Planning Department, NTT.

He received a B.E., M.E., and Ph.D. in applied chemistry from Waseda University, Tokyo, in 1996, 1998, and 2005. He joined NTT Basic Research Laboratories in 1998, where he researched electrochemical biosensing devices to detect biomolecules including neurotransmitters and hormones in the brain and blood. During 2007–2008, he was a visiting scientist in the Biomedical Engineering Department, University of Wisconsin-Madison, USA, where he studied cell biology with microfluidic devices. He launched a Medical AI research project with Drs. Kurasawa and Fujino in 2013 and is supporting the promotion of the project. He is a member of the Chemical Society of Japan, the Electrochemical Society of Japan, and the Japan Association of Chemical Sensors.

Biosignal Processing Methods Targeting Healthcare Support Services

Kana Eguchi, Ryosuke Aoki, Suehiro Shimauchi, Akihiro Chiba, and Naoki Asanoma

Abstract

NTT is pursuing research and development (R&D) of Heart-Touching-AI—artificial intelligence that will help better our understanding of human physical and mental states that are only dimly perceived by the individual. We provide here an overview of recent NTT R&D in biosignal processing for estimating physical and mental states using shirt-type wearable electrocardiogram (ECG) devices such as a "hitoe" shirt. We also explain how we dealt with the inherent problems in shirt-type wearable ECG devices.

Keywords: wearable, "hitoe," biosignal processing

1. Wearable devices supporting everyday life

Wearable devices now make it easy to continuously measure biosignals of subjects and monitor their condition in daily life. To properly support people's daily activities, it is important to inform them of their physical and mental states, which they may not be aware of, based on data measured from wearable devices. When people perform tasks that do not involve much physical exertion such as office work or driving, it is difficult to accurately gauge the actual workload from only the amount of physical activity. It is therefore important to encourage people to take breaks or other kinds of relaxation to improve their work efficiency or health management based on the actual workload. Attention also needs to be paid to the mental state of workers, which can also be affected by their workload.

2. Estimation of physical and mental states by heart rate variability analysis

Heart rate variability (HRV) analysis is a well-known method to assess a person's physical and mental states that involves analyzing the variations between heartbeats. It can shed light on the static balance between sympathetic nerves and parasympathetic nerves that form the autonomic nervous system [1, 2]. In the electrocardiogram (ECG) waveform shown in **Fig. 1**, the QRS complex corresponds to the depolarization of the ventricles of the heart; the time interval between two adjacent R waves is called the R-R interval (RRI). RRI time series data are known to have fluctuations called HRV, which could possibly reflect autonomic nervous system activity under specific conditions [1].

We focused on these physiological characteristics in developing some estimation algorithms targeting sleep stage [3, 4] or mental fatigue [5, 6]. These algorithms estimate the target state through the HRV analysis process shown in **Fig. 2**. We can use shirt-type wearable ECG devices such as a "hitoe" shirt[*1]

*1 "hitoe" shirt: A shirt containing embedded measurement electrodes made of "hitoe," a new functional material [7] developed in collaboration with textile manufacturer Toray Industries, Inc. By including a transmitter with a built-in analog/digital converter and wireless data transfer unit in the "hitoe" shirt, we can measure ECG waveforms and calculate heartbeats or RRIs from the measured ECG waveforms. Note that measurement electrodes in the "hitoe" shirt should be in direct contact with the wearer's skin.

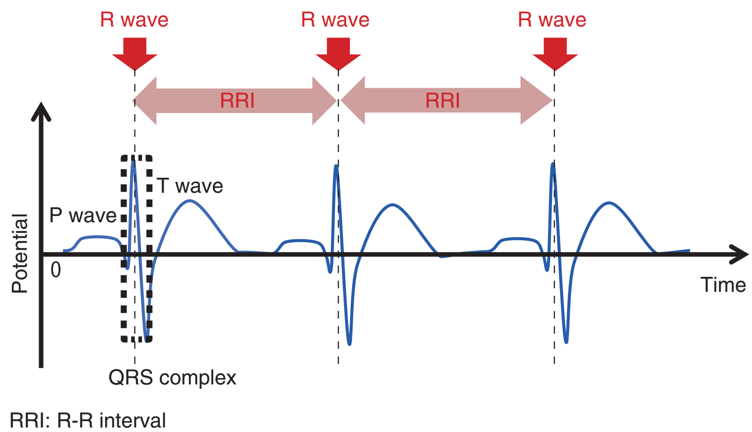


Fig. 1. Schematic diagram of ECG waveform and RRI.

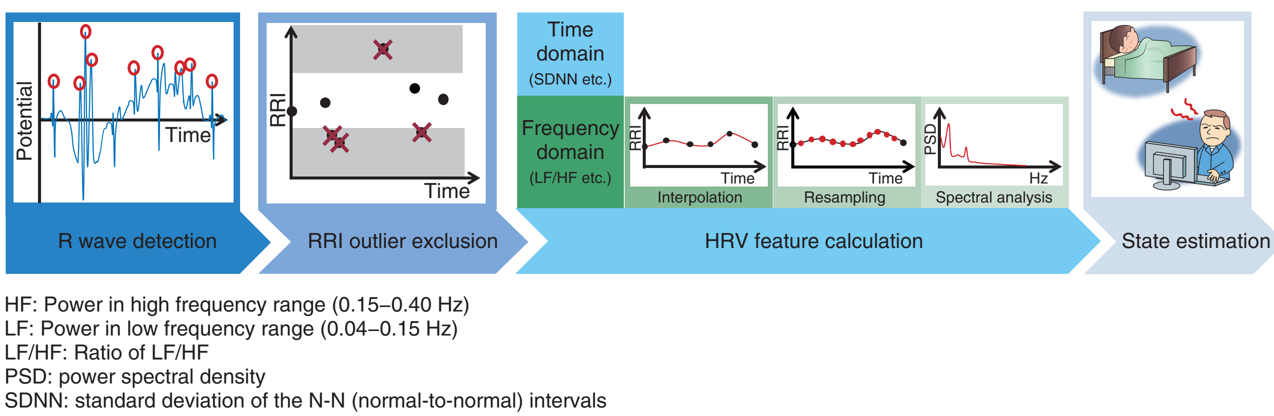


Fig. 2. HRV analysis process.

to measure the target ECGs before carrying out this process.

NTT Group companies are involved in developing and implementing practical products and services in collaboration with outside corporate partners. The "hitoe" shirt, developed in collaboration with Toray Industries, Inc., is used in one such practical application of a driver health management service [6]. This service is based on the fatigue state*2 estimation model that focuses on the relationship between various HRV features obtained through HRV analysis and the wearer's fatigue state. A transmitter attached to a "hitoe" shirt worn by a driver measures ECG, calculates RRIs, and then sends those RRI time series data to the cloud in order to estimate the driver's fatigue state by using a cloud-implemented fatigue state estimation model with parameters calibrated in advance. The model is thus able to continually monitor the fatigue level of the driver simply by having the driver wear the "hitoe" shirt.

Note that the state estimation algorithms using HRV analysis described above generally need to apply an estimation model based on the relationship between HRV features and the target state. This means that those estimation algorithms basically require long-term continuous RRI time series data even when using shirt-type wearable ECG devices such as a "hitoe" shirt. However, ECG measurement electrodes embedded with shirt-type wearable ECG devices stretch along with the shirt itself when the

*2 Fatigue state: In this article, fatigue state means the fatigue of the central nervous system including the brain and spine, measured by the flicker test.

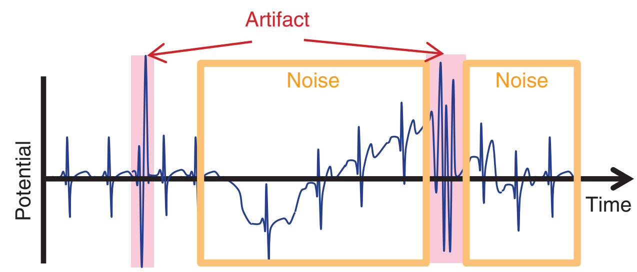


Fig. 3. Schematic diagram of ECG waveform with measurement faults due to stretching of electrodes.

user moves about, and this may cause measurement faults[*3] as illustrated in **Fig. 3**. This is most often observed when the user twists around. Since measurement faults make it impossible to obtain accurate RRIs, conventional methods have excluded these data from HRV analysis to avoid obtaining inaccurate HRV features. In real life, however, people contort their bodies throughout the day; we toss and turn in bed, twist our torso as we turn the steering wheel, and stretch our body during breaks.

To allow for these situations without discarding a significant amount of measured data when conducting HRV analysis in daily life, a new data analysis scheme is required that incorporates advanced outlier processing. We explain in the following sections two new methods that enable more accurate HRV analysis with the goal of enhancing self-state awareness of users through the use of shirt-type wearable ECG devices like the "hitoe" shirt: ECG waveform analysis based on non-orthogonal wavelet expansion [8] and RRI outlier processing [9–11].

3. ECG waveform analysis based on non-orthogonal wavelet expansion

The primary goal of our ECG waveform analysis based on non-orthogonal wavelet expansion (hereafter, NCWE) [8] is to improve the accuracy of R wave detection in HRV analysis, which is the very first step shown in Fig. 2, even when analyzing ECG measured by shirt-type wearable ECG devices. Thus far, ECGs have generally been measured in the presence of physicians or medical technicians, who are responsible for discarding certain portions of ECGs containing measurement faults, so that even simple R wave detection algorithms can detect R waves. However, there are in general no medical personnel present to discard faulty results when ECGs are measured using shirt-type wearable ECG devices. Consequently, measurement faults inevitably make their way into the target ECGs. Moreover, measurement electrodes embedded with shirt-type wearable ECG devices may be attached at different points in accordance with the physique or body movements of users, and this may induce measurement faults or apparently different ECG waveforms.

Many of the conventional simple R wave detection algorithms are often unable to follow these differences in ECG waveforms or to detect R waves buried in measurement faults accurately, and this may result in R wave misdetection or non-detection. This leads to RRI miscalculations, which make it virtually impossible to accurately calculate HRV features. Thus, when HRV features are calculated under everyday environments using a shirt-type wearable ECG device, it is critically important to employ a reliable R wave detection algorithm that is capable of handling measurement faults caused by body movements in order to minimize the instances of miscalculated RRIs (i.e., R wave misdetection and non-detection).

NTT brought together experts from different NTT laboratories to develop a more accurate method of ECG waveform analysis. The members involved were from NTT Service Evolution Laboratories, where they have biosignal processing know-how, and from NTT Media Intelligence Laboratories, where they have audio/acoustic signal processing expertise. The ECG waveform analysis method they developed is based on NCWE and can detect more accurate R waves even when using shirt-type wearable ECG devices such as a "hitoe" shirt.

A schematic overview of how our algorithm detects R waves from target ECGs is shown in **Fig. 4**. The

*3 Measurement faults: States when we cannot measure precise ECGs due to changes in contact between measurement electrodes and skin caused by certain body movements. This can cause noise or artifacts as shown as Fig. 3 in the measured ECG and impede HRV analysis including R wave detection.

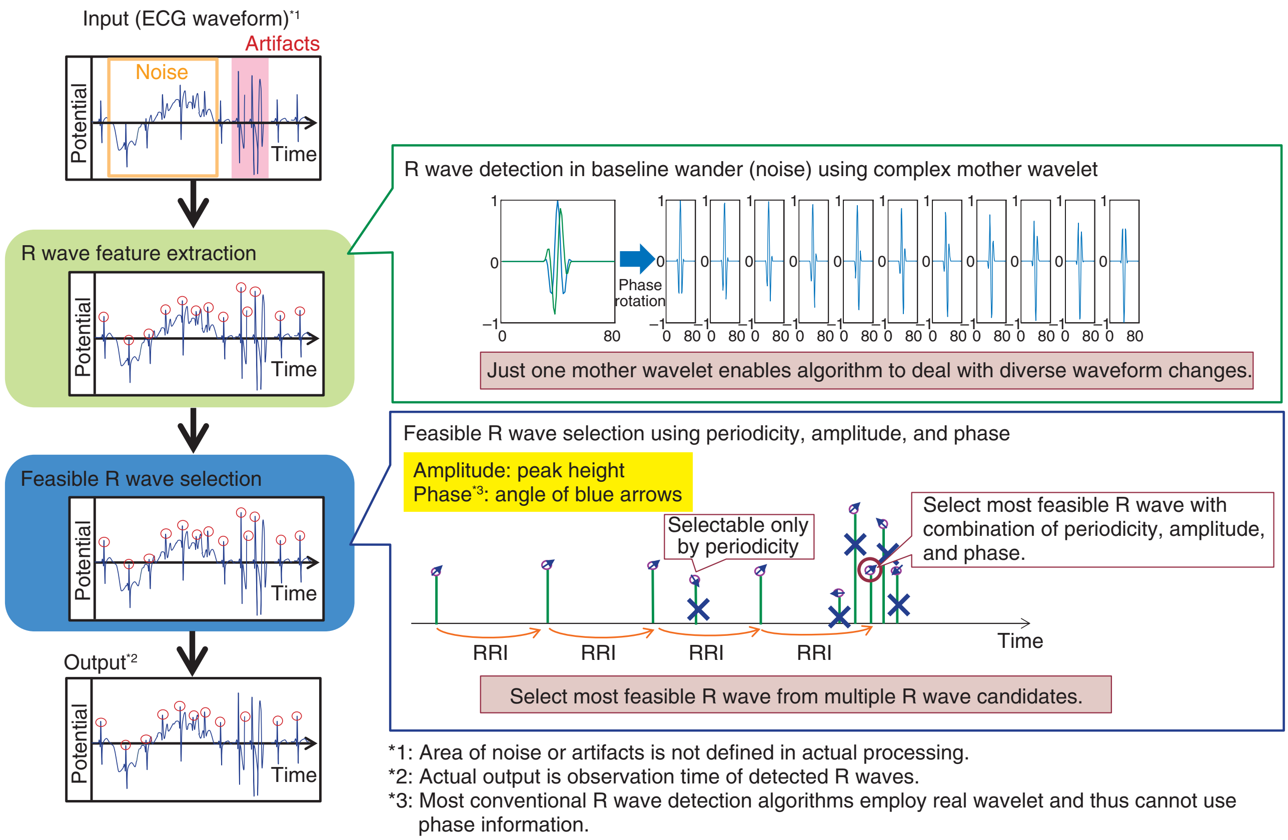


Fig. 4. Overview of ECG waveform analysis based on NCWE.

algorithm puts a general wavelet transform used in audio/acoustic signal processing or ECG waveform analysis into practice to reduce the effects of measurement faults derived from body movements such as low-frequency noise and to detect R waves by following apparently diverse waveform changes of R waves. Furthermore, we reduce the number of misdetected R waves caused by measurement faults such as artifacts not by simply using all detected R waves as they are, but by using only feasible R wave candidates in RRI calculation; we select certain R wave candidates as feasible ones based on formerly measured R waves.

One key in this analysis scheme is a mother wavelet; we employed a complex wavelet as a mother wavelet for primary R wave detection, which makes it possible to detect R waves from ECGs with noise or artifacts measured by shirt-type wearable ECG devices. Most conventional R wave detection algorithms employ a real wavelet as a mother wavelet. This makes it hard to follow diverse changes of R waves by using just one mother wavelet. Our use of a complex wavelet instead of a real wavelet enables the algorithm to follow apparent R wave waveform changes by employing just one mother wavelet with phase rotation. Moreover, this phase information is utilized to select feasible R wave candidates. We select such candidates by making use of three different perspectives—periodicity, amplitude, and phase—all of which we calculate through ECG waveform analysis using NCWE.

To verify the effectiveness of our ECG waveform analysis using NCWE under actual environmental conditions, we compared RRIs calculated using our approach with those resulting from a conventional R wave detection algorithm. We targeted the same ECGs measured with shirt-type wearable ECG devices while the user carried out normal exercises and body movements including stretching, bending, and twisting, all of which typically induce measurement faults.

The experimental results shown in **Fig. 5** indicate that our algorithm successfully calculated stable RRIs, in contrast to the conventional algorithm, even

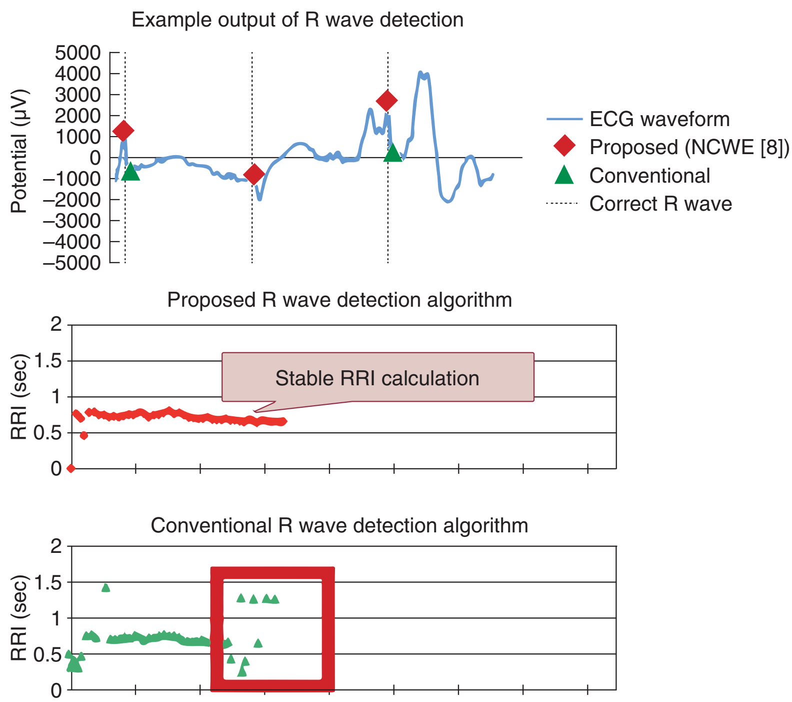


Fig. 5. Experimental results of RRI calculation.

in the presence of measurement faults. This result indicates that our algorithm can potentially improve the accuracy of a range of processes including state estimation based on HRV using shirt-type wearable ECG devices such as a "hitoe" shirt by improving the accuracy of RRIs as well as HRV features calculated from RRIs.

4. RRI outlier processing

To calculate more accurate RRIs and HRV features, our RRI outlier processing technique [9–11] focuses on misdetected R waves contained in the RRIs. This process should be conducted between R wave detection and HRV feature calculation in the HRV analysis (Fig. 2). A schematic diagram of our RRI outlier processing is illustrated in **Fig. 6**. Although there are many R wave detection algorithms as described above, it is almost impossible to completely suppress misdetected R waves due to artifacts, whose frequency characteristics are quite similar to those of R waves. Because HRV analysis only requires accurate RRIs without any outliers, we should exclude the misdetected RRIs derived in artifacts as outliers. However, conventional HRV analyses mostly target ECGs without any measurement faults and only exclude the RRIs as outliers whose duration characteristics deviate from those of the majority of RRIs. Therefore, we cannot exclude certain RRIs including misdetected R waves as outliers, especially when the duration characteristics of those RRIs are close to those of the other RRIs.

Another problem can occur when we calculate frequency domain measures of HRV (hereafter, fHRV) where we should conduct two additional preprocessing steps before feature calculation to appropriately analyze frequency characteristics of RRI time series data. These steps are data resampling using an interpolation function and power spectral density calculation by spectral analysis [1]. When the gap in missing RRIs is too long, we may fail to accurately calculate fHRV due to oscillation of the interpolation function to the outliers or due to overestimation of specific frequency components in data resampling using the interpolation function.

To accurately calculate HRV features under these situations, our RRI outlier processing method conducts two additional processes based on inherent HRV analysis: RRI outlier exclusion [9] and missing RRI complementation [10, 11]. In RRI outlier exclusion, we exclude all RRIs that possibly include misdetected R waves based on the calculated measurement

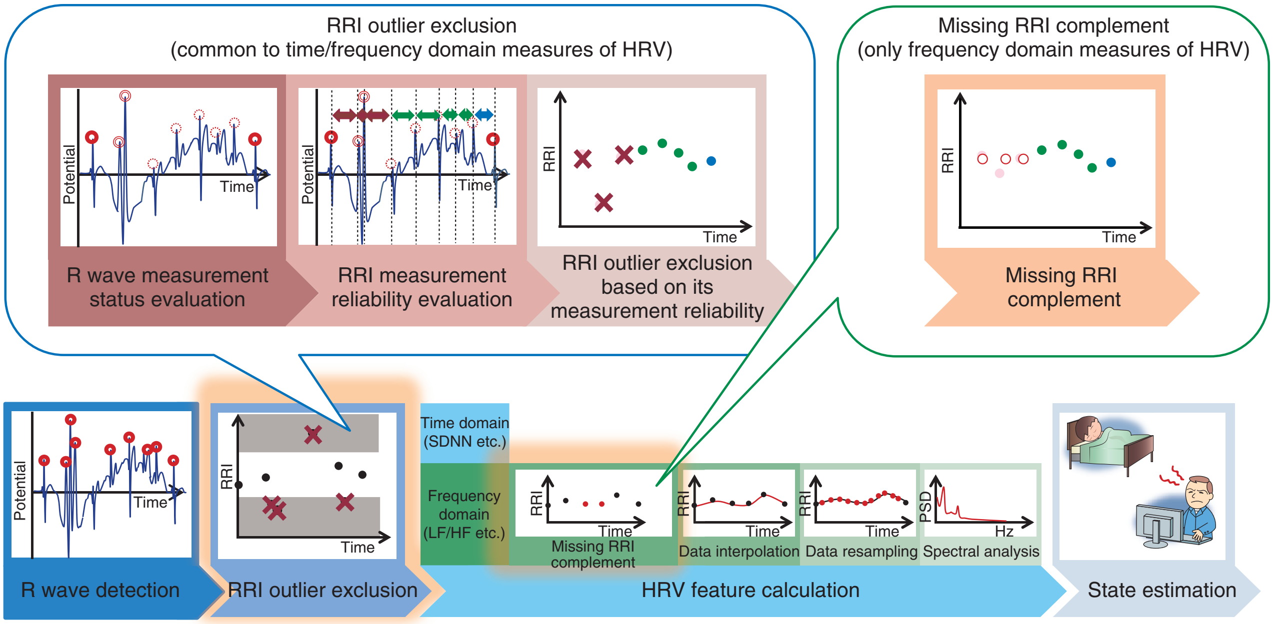


Fig. 6. RRI outlier processing.

reliability. We first evaluate the measurement status of all detected R waves based on their electric potential and then evaluate the measurement reliability of RRIs based on the combined measurement status of two R waves. Thus, we can exclude certain RRIs with low measurement reliability as possible outliers derived from measurement faults [9].

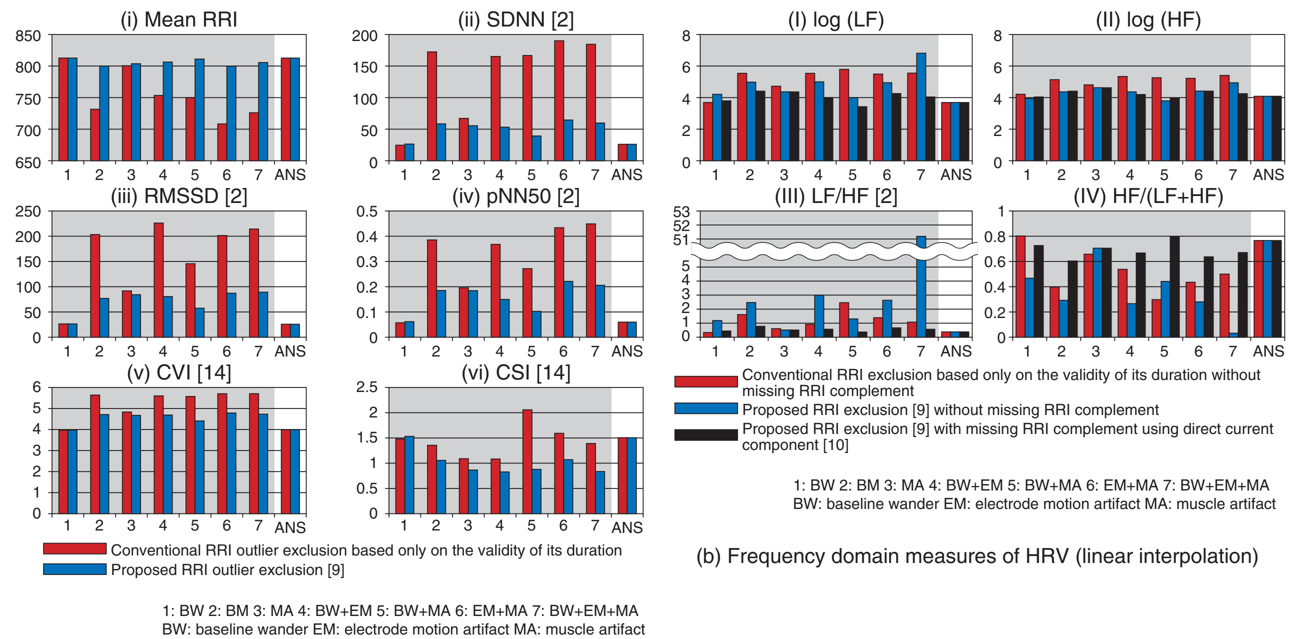
Furthermore, when calculating fHRV, we complement missing RRIs before conducting data resampling in order to properly resample target data by using an interpolation function [10, 11]. The combination of RRI outlier exclusion and missing RRI complementation can limit the aforementioned problems that may occur in conventional data resampling and enable more accurate fHRV calculation.

As a preliminary evaluation before the validation analysis under actual environmental conditions, we evaluated the effectiveness of RRI outlier exclusion and missing RRI complementation by using pseudo ECGs*4 with correct R wave annotation, which assumes inherent measurement faults occurring with shirt-type wearable ECG devices. The experimental results shown in **Fig. 7** indicate that our proposed RRI outlier processing methods were effective for time domain measures of HRV (Fig. 7(a)) as well as fHRV (Fig. 7(b)), and combining both of them was more effective than RRI outlier exclusion alone in fHRV calculation. Furthermore, even more accurate HRV features could be calculated by combining an R wave detection algorithm based on NCWE and RRI outlier processing methods.

5. Summary

In this article, we presented two new methods, ECG waveform analysis based on NCWE and RRI outlier processing, that work together as preprocessing steps in HRV analysis. The combination of both methods enables more accurate calculation of HRV features when using shirt-type wearable ECG devices such as a “hitoe” shirt and may indirectly contribute to expanding the use of such garments for sports or labor health management purposes. We will continue to pursue research and development activities to achieve healthcare support services using shirt-type wearable ECG devices including “hitoe” shirts in collaboration with other NTT Group companies.

*4 Pseudo ECGs: We artificially generated target pseudo ECGs by mixing the MIT-BIH Arrhythmia Database [12], which we used for ECG signals, and the MIT-BIH Noise Stress Test Database [13], which we used for noise or artifacts.



(a) Time domain measures of HRV

CSI: cardiac sympathetic index [14] CVI: cardiac vagal index [14]
pNN50: The proportion derived by dividing NN50* by the total number of all NN intervals [2]
*NN50: The number of interval differences of successive NN intervals greater than 50 ms
RMSSD: The square root of the mean of the sum of the squares of differences between adjacent NN intervals [2]

Fig. 7. Experimental results of RRI outlier processing.

References

[1] H. Inoue (Eds), "Cardiovascular Diseases and Autonomic Nervous Function," Second Edition, Igaku-Shoin, 2001 (in Japanese).

[2] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, "Heart Rate Variability: Standards of Measurement, Physiological Interpretation, and Clinical Use," European Heart Journal, Vol. 17, pp. 354–381, 1996.

[3] T. Takeda, O. Mizuno, and T. Tanaka, "Time-dependent Sleep Stage Transition Model Based on Heart Rate Variability," Proc. of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2015), pp. 2343–2346, Milano, Italy, Aug. 2015.

[4] T. Takeda, A. Niijima, T. Watanabe, K. Yoshida, and O. Mizuno, "Sleep Stage Classification Based on a Relationship between Sleep Stage Transition and Heart Rate Change," IPSJ Transactions on Databases, Vol. 9, No. 4, pp. 6–10, 2016 (in Japanese).

[5] A. Chiba, K. Tsunoda, H. Chigira, T. Ura, O. Mizuno, and T. Tanaka, "Study of Estimating Critical Flicker Frequency from Heart Rate Variability," IEICE Technical Report, Vol. 115, No. 149, pp. 7–12, 2015 (in Japanese).

[6] T. Kondo, Y. Yamato, M. Nakayama, A. Chiba, K. Sakaguchi, T. Nishiguchi, T. Masuda, and T. Yoshida, "Natural Sensing with "hitoe" Functional Material and Initiatives towards Its Applications," NTT Technical Review, Vol. 15, No. 9, 2017.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201709fa3.html

[7] S. Tsukada, N. Kasai, R. Kawano, K. Takagahara, K. Fujii, and K. Sumitomo, "Electrocardiogram Monitoring Simply by Wearing a Shirt—For Medical, Healthcare, Sports, and Entertainment," NTT Technical Review, Vol. 12, No. 4, 2014.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201404fa4.html

[8] S. Shimauchi, K. Eguchi, T. Takeda, and R. Aoki, "An Analysis Method for Wearable Electrocardiogram Measurement Based on Non-Orthogonal Complex Wavelet Expansion," Proc. of the 39th Annual International Conference of IEEE Engineering in Medicine and Biology Society (EMBC 2017), pp. 3973–3976, Jeju, South Korea, July 2017.

[9] K. Eguchi, R. Aoki, K. Yoshida, and T. Yamada, "Reliability Evaluation of R-R Interval Measurement Status for Time Domain Heart Rate Variability Analysis with Wearable ECG Devices," Proc. of EMBC 2017, pp. 1307–1311, Jeju, South Korea, July 2017.

[10] K. Eguchi, R. Aoki, K. Yoshida, and T. Yamada, "Missing R-R Interval Complement Method for Heart Rate Variability Analysis in Frequency Dominant Using Wearable ECG Devices," Proc. of Multimedia, Distributed, Cooperative, and Mobile Symposium (DICOMO 2017), pp. 888–897, Sapporo, Japan, June 2017 (in Japanese).

[11] K. Eguchi, R. Aoki, S. Shimauchi, K. Yoshida, and T. Yamada, "R-R Interval Outlier Processing for Heart Rate Variability Analysis Using Wearable ECG Devices," Advanced Biomedical Engineering, Vol. 7, pp. 28–38, 2018.

[12] Physionet, "The MIT-BIH Arrhythmia Database,"
http://physionet.org/physiobank/database/mitdb/

[13] Physionet, "The MIT-BIH Noise Stress Test Database,"
http://physionet.org/physiobank/database/nstdb/

[14] M. Toichi, T. Sugiura, T. Murai, and A. Sengoku, "A New Method of Assessing Cardiac Autonomic Function and Its Comparison with Spectral Analysis and Coefficient of Variation of R-R Interval," J. Auton. Nerv. Syst., Vol. 62, pp. 79–84, 1997.

Kana Eguchi

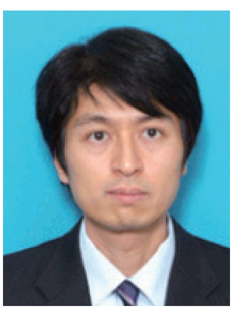
Researcher, Networked Robot and Gadget Project, NTT Service Evolution Laboratories.

She received an M.S. in informatics from Kyoto University in 2012. She joined NTT in 2012 and has been engaged in research on network middleware and medical engineering at NTT Service Evolution Laboratories. Her current research interests include biosignal processing, wearable/ubiquitous computing, and medical engineering. She is a member of the Institute of Electronics, Information, and Communication Engineers (IEICE), the Institute of Electrical and Electronics Engineers (IEEE) Engineering in Medicine and Biology Society, and the Japanese Society for Medical and Biological Engineering.

Ryosuke Aoki

Research Engineer, Networked Robot and Gadget Project, NTT Service Evolution Laboratories.

He received a B.E. in engineering and an M.S. and Ph.D. in information sciences from Tohoku University, Miyagi, in 2005, 2007, and 2014. Since joining NTT in 2007, he has been conducting research in the areas of human computer interaction, medical engineering, motor learning, and service design. He has been a Visiting Assistant Professor at Tokyo Denki University since April 2016.



Suehiro Shimauchi

Senior Research Engineer, Audio, Speech, and Language Media Project, NTT Media Intelligence Laboratories.

He received a B.E., M.E., and Ph.D. from Tokyo Institute of Technology in 1991, 1993, and 2007. Since joining NTT in 1993, he has been researching acoustic signal processing and biosignal processing. Dr. Shimauchi is a member of IEEE, IEICE, and the Acoustical Society of Japan.

Akihiro Chiba

Researcher, Networked Robot & Gadget Project, NTT Service Evolution Laboratories.

He received a B.E. and M.E. in engineering from the University of Electro-Communications, Tokyo, in 2011 and 2013. He joined NTT Service Evolution Laboratories in 2013. His current research interests include sensing of biological information. He received the Society of Instrument and Control Engineers (SICE) Young Author's Award at the 2011 SICE Annual Conference. He is a member of SICE and the Information Processing Society of Japan.

Naoki Asanoma

Research Engineer, Networked Robot & Gadget Project, NTT Service Evolution Laboratories.

He received a master's degree in information and computer science from Waseda University, Tokyo, in 1997. He joined NTT in 1997. He is currently engaged in research and development of healthcare supporting technology.

Advanced Learning Technologies for Deep Learning

Yasutoshi Ida, Sekitoshi Kanai, Yasuhiro Fujiwara, Satoshi Yagi, and Yasuhiro Iida

Abstract

NTT is focusing on artificial intelligence (AI) as a key technology in developing business strategies. The NTT laboratories have been engaged in research and development (R&D) of deep learning algorithms, which play a central role in AI. This article presents an overview of our R&D of novel algorithms that speed up and stabilize deep learning and touches on our collaboration with several partners to verify those algorithms.

Keywords: deep learning, algorithm, optimizer

1. Introduction

NTT places high priority on addressing various social issues and strengthening industrial competitiveness, and the use of artificial intelligence (AI) in our business endeavors in pursuit of these goals is an area of great importance to us. AI includes many technologies such as those used in statistical analysis, machine learning, and deep learning, which are applied to perform classification, regression, and prediction tasks with large amounts of data. Deep learning has been attracting a great deal of interest lately because it has achieved a practical level of accuracy in a variety of tasks. It has already been introduced to improve various business practices and is expected to become a driving force in the creation of new businesses.

Deep learning is a method to extract features from data hierarchically. A layered neural network is usually used as a model. Users select a suitable neural network according to a type of task or data. Convolutional neural networks are suitable for image data, while recurrent neural networks (RNNs) suit time-series data. In a neural network, the system makes predictions by giving weight to input data signals or by applying nonlinear transformation to the signals hierarchically and then having the signals propagate (**Fig. 1**). *Learning* is a process of adjusting weights to reduce the amount of error when using a large training dataset. In this process, the layered structure enables deep learning to achieve high accuracy in a variety of tasks.

However, the layered structure of deep learning models presents some problems such as an increase in learning time and destabilization of learning itself. Therefore, it is important to address this issue when making full use of deep learning techniques. To this end, the NTT laboratories have developed (1) an algorithm that improves learning efficiency and (2) an algorithm that stabilizes learning in an RNN [1, 2]. These algorithms are introduced in the following section.

2. Algorithm to improve learning efficiency

Learning of a layered neural network model is a process of adjusting weights to reduce the amount of error. This adjustment is performed gradually by applying a procedure as shown in **Fig. 2**. First, data are input into the model and a prediction result is obtained (forward propagation). Next, this prediction result is compared with the correct label, and the amount of error is calculated (error computation). The calculated amount of error is propagated to the model (back propagation), and the update direction of weights is calculated. The amount of updating is also

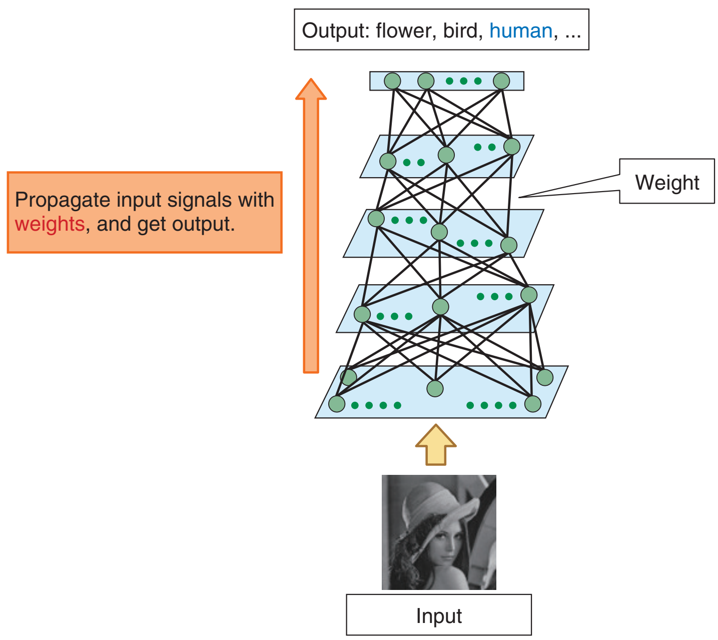


Fig. 1. Layered neural network.

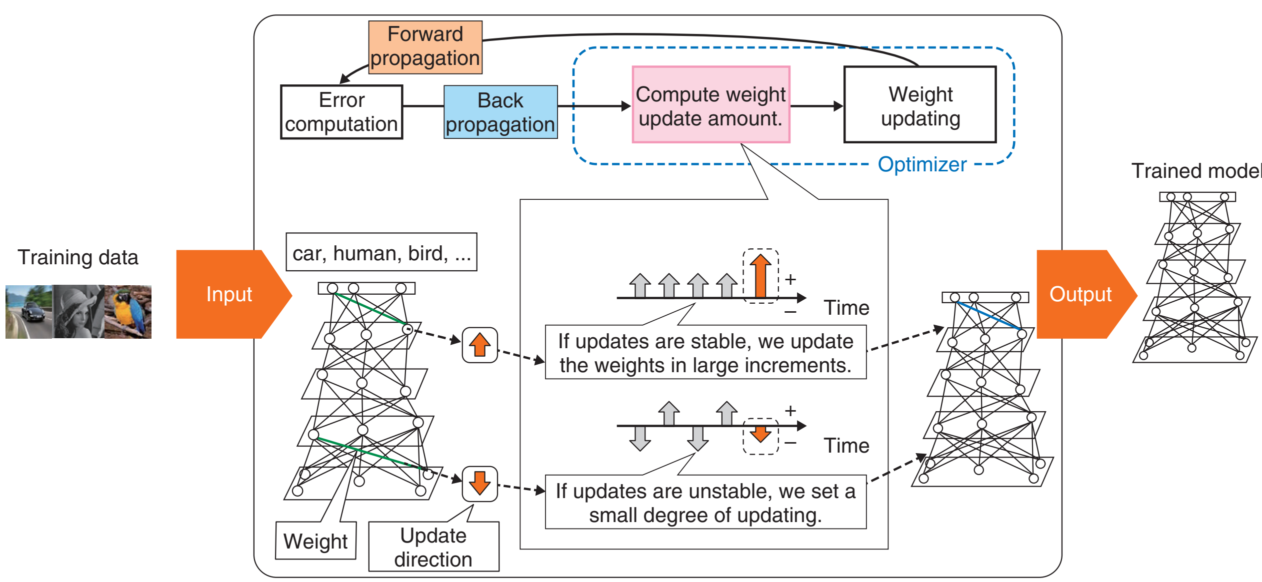


Fig. 2. Learning of a layered neural network model.

calculated (weight update amount computation). Finally, the weights are updated using this update amount (weight update). There are several approaches to calculate the weight update amount, and the learning efficiency varies depending on the approach used. In other words, the amount of error that can be reduced per loop depends on the approach used.

Several approaches adjust the update amount based on information about past update directions. For example, the widely used approaches RMSprop and Adam adjust the update amount based on statistical values calculated from the absolute values of the

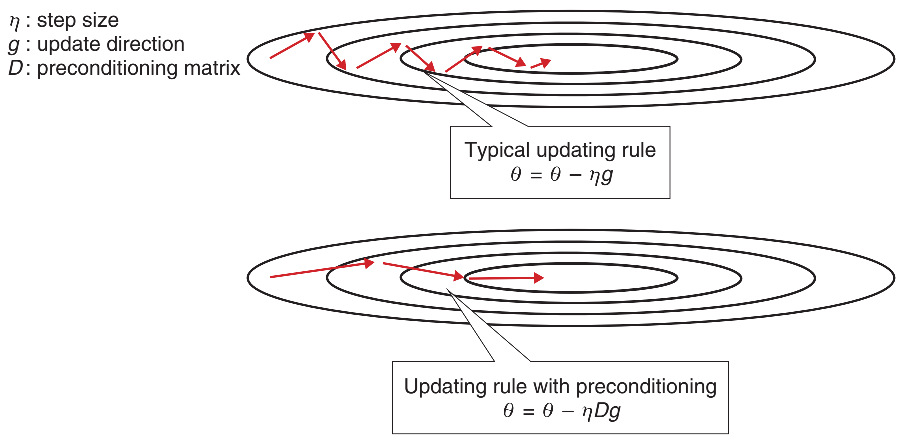


Fig. 3. Behavior of algorithm on loss function with respect to weights θ.

update directions. However, since these approaches use the absolute values of the update directions, they do not generally take variances of those values into consideration. This means that no matter how much variance there is in the update directions, the same update amount is set if the absolute amount is the same.

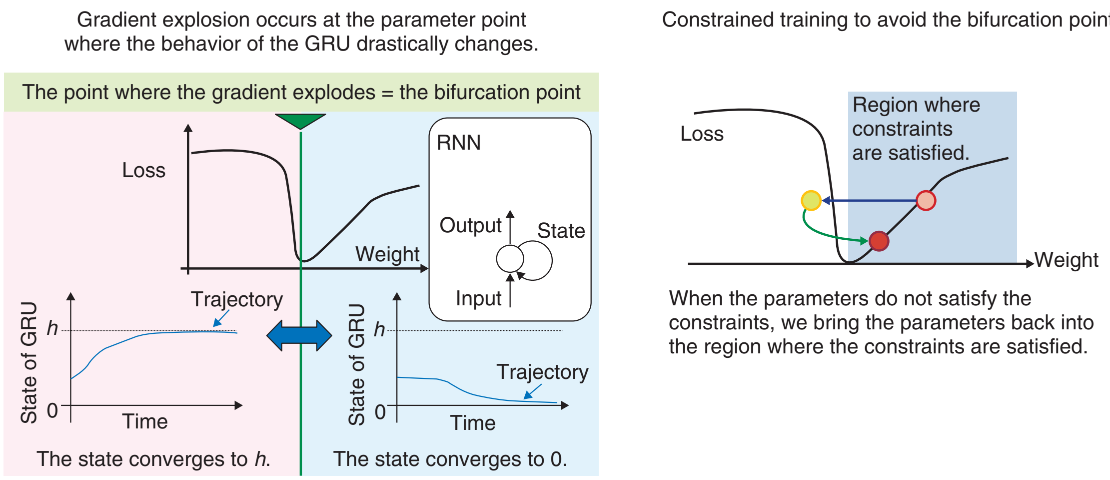
To address this problem, we have developed an algorithm that adjusts the update amount based on the variance in the update directions (Fig. 2). If the update directions vary significantly, it can be intuitively determined that learning is unstable. Therefore, the update amount is reduced to stabilize learning. If the variance of the update directions is small, the update amount is increased. This would increase the variance but would also increase the possibility of learning breaking away from a local solution, enabling it to further reduce the amount of error.

This algorithm is simple and can thus be easily implemented in numerous deep learning frameworks. It also has a theoretically interesting aspect. It can be regarded as a type of optimization with preconditioning. The preconditioning matrix that sets the optimization conditions becomes an approximation of the square root of the diagonal elements of a matrix calculated from gradients, which is called the Fisher information matrix. From the perspective of information geometry, which is a framework for visually understanding information processing problems, this fact suggests that an algorithm that repeats updates can converge faster than existing algorithms (**Fig. 3**).

3. Technique to stabilize training of gated recurrent unit (GRU)

A deep learning model called an RNN is used to handle time-series data such as those used in speech recognition or machine translation. The RNN memorizes information about past data as a state and calculates the output from this state and the current input. For example, consider a task of predicting the next word from a given text. The next word depends on the context of the preceding text. It can be predicted at a high level of accuracy by having the RNN memorize information about the past context as a state.

To process time-series data with a high degree of accuracy, it is necessary to preserve past information for the long term. Therefore, an important indicator of RNN performance is how long the RNN can memorize past information. A model structure called the long short-term memory (LSTM) was proposed in the late 1990s to achieve long-term memorization [3]. The LSTM has a structure called a memory cell that takes in past memories, and a gate structure that forgets unnecessary information so that old but important information can be saved while information that has become unnecessary can be forgotten. The LSTM has already been applied in several machine translation and speech recognition technologies. However, the LSTM has a complicated structure, so the simpler GRU, which combines the input gate and the forget gate of the LSTM into one, was proposed in 2014 [4]. The GRU has a simpler structure and requires less computation; however, it has been demonstrated empirically that the GRU achieves a level of accuracy almost equivalent to that of the



(a) Reveal the bifurcation point

(b) Avoid the bifurcation point

Fig. 4. Technique to stabilize training of GRU.

LSTM.

While the RNN can process time-series data at a high level of accuracy, its learning becomes unstable due to a phenomenon called gradient explosion [5]. This is a phenomenon whereby the gradient, which plays the main role in learning of neural networks, becomes so large as to cause learning to fail. An existing technique to counter this problem is gradient clipping, which clips gradients at a certain threshold value. This technique requires the threshold value to be adjusted through trial and error.

We are studying ways to solve the gradient explosion problem by analyzing the behavior of the GRU. From the perspective of dynamical systems, it can be said that a gradient explosion arises when a bifurcation occurs as a result of a change in parameters of the GRU (**Fig. 4**). A bifurcation is a phenomenon in which the behavior of the GRU changes dramatically when there is a small change in its parameters. NTT's algorithm for stabilizing the training of GRU involves analyzing the behavior of the GRU state to identify the point at which a bifurcation occurs. Furthermore, NTT has proposed a method of learning that efficiently avoids the bifurcation point. Thus, we are studying a method of learning that has a higher degree of accuracy and requires fewer trial-and-error attempts than gradient clipping [2].

4. Collaborations to refine NTT-developed advanced technologies

While devoting our energies to research, we also undertake activities to improve the effectiveness of deep learning technology in collaboration with a broad spectrum of users. For example, we have implemented our efficient learning technology to optimizers of multiple representative deep learning frameworks such as Chainer, TensorFlow, and Caffe, and we are evaluating our technology in collaboration with NTT Group companies. To date, we have examined the efficiency and accuracy of a model for an image recognition service in order to assess the model's feasibility. Through such examinations, we are accumulating know-how regarding the optimum sizes (batch sizes) of the input training dataset based on the data variety and the initial values of the learning rate and weights. We verified our learning stabilization technique for various types of time-series data such as audio data, language data, and sensor data so as to acquire know-how regarding the relationship between our technique when used in combination with optimizers as well as the degrees of stability and accuracy.

In addition, we are studying a wide range of possible approaches to promote collaboration with a broad spectrum of users, for example, providing technology for speeding up or stabilizing deep learning as open source. We are also studying the possibility of providing to users a one-stop service covering a deep learning framework, tuning, and a distributed processing platform. This will be done by building the technologies for speeding up or stabilizing deep learning into an AI processing platform called the corevo Computing Infrastructure (CCI), currently being studied at the NTT laboratories [6]. By building deep learning technology into the CCI, we aim to enable even those users without specialized knowledge to perform advanced data analysis.

References

[1] Y. Ida, Y. Fujiwara, and S. Iwamura, "Adaptive Learning Rate via Covariance Matrix Based Preconditioning for Deep Neural Networks," Proc. of the 26th International Joint Conference on Artificial Intelligence (IJCAI 2017), pp. 1923–1929, Melbourne, Australia, Aug. 2017.

[2] S. Kanai, Y. Fujiwara, and S. Iwamura, "Preventing Gradient Explosions in Gated Recurrent Units," Proc. of the 31st Annual Conference on Neural Information Processing Systems (NIPS 2017), pp. 435–444, California, USA, Dec. 2017.

[3] S. Hochreiter and J. Schmidhuber, "Long Short-term Memory," Neural Computation, Vol. 9, No. 8, pp. 1735–1780, 1997.

[4] K. Cho, B. Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation," Proc. of the 19th Conference on Empirical Methods in Natural Language Processing (EMNLP 2014), pp. 1724–1734, Doha, Qatar, Oct. 2014.

[5] R. Pascanu, T. Mikolov, and Y. Bengio, "On the Difficulty of Training Recurrent Neural Networks," Proc. of the 30th International Conference on Machine Learning (ICML 2013), pp. 1310–1318, Atlanta, USA, June 2013.

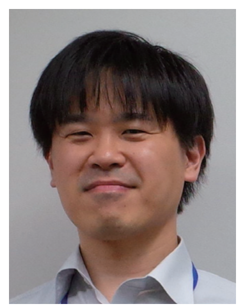
[6] M. Kawashima, "Four Activities to Create Innovative Core Technologies that Support the Development of IoT/AI Services," Business Communication, Vol. 54, No. 12, pp. 8–13, 2017 (in Japanese).

Trademark notes

Yasutoshi Ida

Researcher, Distributed Computing Technology Project, NTT Software Innovation Center.

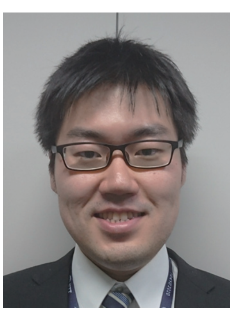
He received a B.E. and M.E. from Waseda University, Tokyo, in 2012 and 2014 and joined NTT in 2014. His research interests lie in the fields of Bayesian modeling and deep learning.



Satoshi Yagi

Senior Research Engineer, Distributed Computing Technology Project, NTT Software Innovation Center.

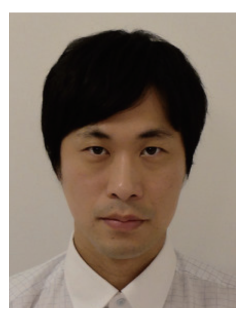
He received a B.E. and M.E. from Waseda University, Tokyo, in 2000 and 2002. He joined NTT Information Sharing Platform Laboratories in 2002. Since then, he has been studying digital identity, data mining, and optimization problems.



Sekitoshi Kanai

Researcher, Distributed Computing Technology Project, NTT Software Innovation Center.

He received a B.E. and M.E. in applied physics and physico-informatics from Keio University, Kanagawa, in 2013 and 2015. He joined NTT in 2015 and has been studying deep learning algorithms.



Yasuhiro Iida

Senior Research Engineer, Distributed Computing Technology Project, NTT Software Innovation Center.

He received a B.S. and M.S. in applied physics engineering from the University of Tokyo in 1998 and 2000. He joined NTT Information Sharing Platform Laboratories in 2000. He is currently researching data mining and data management.

Yasuhiro Fujiwara

Distinguished Technical Member, NTT Software Innovation Center.

He received a B.E. and M.E. from Waseda University, Tokyo, in 2001 and 2003 and a Ph.D. from the University of Tokyo in 2012. He joined NTT Cyber Solutions Laboratories in 2003. His research interests include data mining, databases, and artificial intelligence. He has received several awards including the TAF Telecom System Technology Award from the Telecommunications Advancement Foundation, the IPSJ Nagao Special Researcher Award from the Information Processing Society of Japan (IPSJ), and the Commendation for Science and Technology from the Minister of Education, Culture, Sports, Science and Technology. He is a member of IPSJ, the Institute of Electronics, Information and Communication Engineers, and the Database Society of Japan.

Efforts toward Service Creation with Neural Machine Translation

Sen Yoshida, Masahide Mizushima, and Kimihito Tanaka

Abstract

The accuracy of machine translation has been improving rapidly with the emergence of neural machine translation (NMT) that incorporates deep-learning technology, and it is therefore being utilized in many applications. NTT Media Intelligence Laboratories is working on research and development aimed at creating and enhancing new services using NMT. In this article, we present an overview of NMT and explain some technical points in utilizing it in actual services.

Keywords: machine translation, deep learning, domain adaptation

1. Introduction

Advances in the development and practical application of neural machine translation (NMT), in which deep learning technology is applied, has resulted in substantial improvements in the accuracy of machine translation*1. The outputs of translation services that can be readily utilized from web browsers are considered to be significantly better than they were in years past.

Research on machine translation technology has been underway since around the 1950s. In the early years, translation rules were hand-crafted, but statistical machine translation using machine learning has been under development since the 1990s. This has reduced the cost of building translation systems, resulting in the gradual expansion of machine translation services. Furthermore, the emergence and subsequent improvement in translation accuracy of NMT has led to an increase in the range of its practical use. Consequently, machine translation technology is flourishing.

The NTT Group also provides various types of translation services. In addition to the services offered by NTT DOCOMO via mobile phones and other devices [1], Mirai Translate commenced a machine translation service equipped with an NMT engine [2], which is commercially available through NTT EAST's Hikari Cloud cototoba [3] and NTT Communications' AI Translation Platform COTOHA Translator [4].

Implementing NMT is relatively easy because it has a simpler architecture than conventional machine translation, and various open source software (OSS) libraries for deep learning are available today. In recent years at machine-learning-related academic conferences, authors submitting papers were increasingly being subjected to demands to disclose the source code of their programs. This subsequently led to open-source NMT systems such as OpenNMT [5] becoming easily available to anyone.

2. Overview of NMT

We briefly explain here the mechanism of NMT [6] using examples of English-Japanese translation, in which English text is given as input and is output as text translated into Japanese.

To execute an NMT translator, it is necessary to create a translation model in advance. A translation model in NMT is a neural network that converts English text into Japanese. Like other artificial intelligence (AI) technologies such as image recognition and speech recognition, the neural network of NMT is created by learning from a large number of samples.

*1 This article was translated from Japanese using neural machine translation and post-edited by humans.

In NMT, bitexts, that is, pairs of English and Japanese sentences whose meanings are the same, are used as samples. Below is an example of a bitext.

English: A brown dog is running.
Japanese: 茶色い犬が走っている.
(Chairoi inu ga hashitte iru)

We prepare 0.1–100 million of such English-Japanese bitexts and let the neural network program learn them. With a translation model created in this way, a translator program can translate English texts that are not included in the samples into Japanese.

We explain the structure of the NMT neural network using the case of a sequence-to-sequence model with an attention mechanism [7], which is currently a mainstream scheme. In NMT, we first decompose input text into word sequences*2 as a preprocessing step. Then, each word is mapped onto a vector space of hundreds or thousands of dimensions. Finally, we represent all sorts of information such as the naturalness of the connection between words and the degree of attention given to each word as vectors, and perform translations by doing computations between vectors.

In the process of learning the translation models, the weights in the computation of vectors are adjusted to more appropriate ones. This involves using 0.1–100 million bitexts as learning data. For this reason, an extremely large number of vector operations are performed in the learning of translation models. As with other deep learning technologies, it is therefore common to use a graphics processing unit (GPU)*3, which is hardware that can perform a large number of vector operations at high speed and can be introduced relatively inexpensively. However, even when GPUs are used, the learning of the translation model takes several days to weeks, which is one of the factors in the high cost of developing translation services using NMT.

3. Technical efforts in service creation

At NTT, we are carrying out research and development (R&D) of NMT as part of efforts to create new services. We describe these efforts in this section.

3.1 Adaptation to a limited target domain

A translation service that can be used easily from a web browser does not specifically limit the domain of texts to be translated, and it covers any text from an academic paper on cancer treatment to a menu in a steak house. However, the accuracy of translation is generally higher by limiting the domain of the text to be targeted to a certain extent, for example, by creating and using a dedicated translation model such as for medical papers or for American tourism guides. For this reason, in developing services using NMT, it is vitally important to determine the target domain of a promising translation-related business and to achieve the required translation accuracy by using a customized translation model, which cannot be substituted by a general-purpose translation model.

The simplest and most powerful way to build a translation model that is customized in a targeted domain is to procure a large volume of bitexts in that domain and use them to create a translation model. NTT Media Intelligence Laboratories has begun efforts to improve translation accuracy by collecting bitexts that exist within the NTT Group or by creating them.

3.2 Fine tuning

As explained previously, however, building a translation model with bitexts in the target domain from scratch takes several days or weeks. This makes it too costly to adapt to diverse business needs. Therefore, a technique called *fine tuning* is sometimes used as a means of customizing a translation model at low cost by performing further learning using a relatively small quantity (from several thousand to several hundred thousand) of domain-specific bitexts while using an existing translation model as a basis.

Fine tuning is convenient because it makes it easy to build a custom translation model with only a relatively small number of domain-specific bitexts, although at present, it is more likely to be inferior in terms of translation accuracy compared with the case of building a translation model from scratch. This makes it necessary for a translation business using NMT to use the fine tuning and building-from-scratch in a case-by-case manner by considering the amount of time given and the number of bitexts as well as the scope of the target business domain (**Fig. 1**).

NTT Media Intelligence Laboratories is accumulating know-how on the stabilization of performance in fine tuning, and on the proper use of fine tuning and building-from-scratch.

*2 Word sequence: There are cases in which sentences are segmented into ordinary words and segmented into subwords, which are even finer character strings. Because there is no essential difference in the neural network structure in either case, the segmented character strings are uniformly called “words” in this article.

*3 GPU: Originally an image computing device mounted on a graphics board or the like; however, a technique called general-purpose computing on GPU (GPGPU) that uses GPUs for purposes other than image computation is spreading.

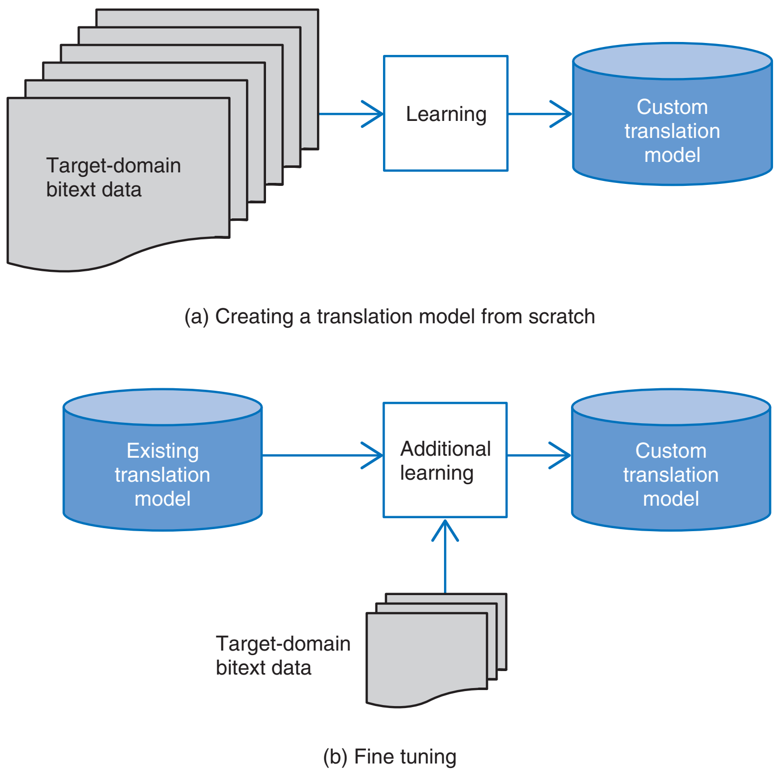


Fig. 1. Method for creating a custom translation model.

3.3 Improving quality of bitexts

Whether we build a translation model from scratch or perform fine tuning, the nature of bitexts used greatly affects the translation accuracy. The nature of bitexts refers to the degree of word variation and the degree of direct translation.

For the degree of word variation, for example, even a sentence representing the same meaning 犬が走っている (A dog is running) can be written in various ways:

・犬が走っています (Inu ga hashitte imasu)
・イヌが走ってる (Inu ga hashitte ru)

It is generally regarded that the fewer word variations there are in bitexts, especially on the target side (Japanese side in English-Japanese translation), the better. This is because when a neural network learns word sequences and the like, it is easier to learn if the sequences are always similar.

Moreover, with regard to the degree of direct translation of bitexts, the readability for humans is emphasized in normal translation (not for machine-translation learning data), so it is often the case that words that are not actually written in English are complemented from the context and incorporated into the Japanese translations. However, the current form of NMT does not handle contextual information, so such complementation can instead become noise in the learning data. Thus, straightforward direct translation is more suitable as bitext data.

NTT Media Intelligence Laboratories has been building guidelines for ensuring good quality of bitext data by suppressing word variations or increasing the degree of direct translation.

3.4 Making named entities variables

Although NMT learns a translation model from a large amount of bitext data, a problem arises in that some expressions that rarely appear in bitext data cannot be learned well. Such problems become prominent especially with numerical values. Numbers exist infinitely, so no matter how large the amount of learning data we collect, we cannot learn everything from the sample.

Similar problems occur even in person names and place names. These phrases indicating proper nouns (e.g., person names and place names), date and time, and other details are referred to as named entities. In machine translation including NMT, it is common to

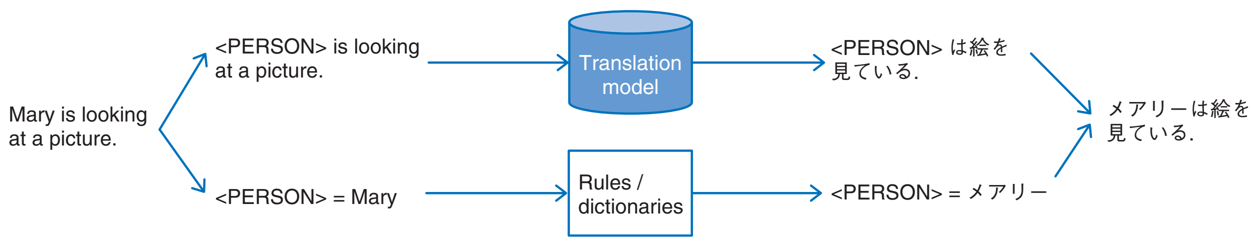


Fig. 2. Converting named entities into variables.

deal with this problem by converting named entities that appear in texts into variables. This is done as follows:

Mary is looking at a picture. →<PERSON> is looking at a picture.

This makes it unnecessary to learn all person names individually in a translation model. However, in a practical translation process, we use a translation model with variables to generate a translation result such as "<PERSON> is looking at a picture," while simultaneously converting "Mary" to "メアリー" and finally returning the variable part to the normal text (**Fig. 2**).

For English-to-Japanese conversion of named entities into variables, for example, a person name and a date, rewriting rules such as regular expression and dictionaries are used. NTT Media Intelligence Laboratories has been working on effective ways of building such rules and dictionaries by utilizing language processing technology such as named entity recognition that we have been developing for many years.

Because treatment of such named entities is not adequately supported by OSS such as OpenNMT, named entity processing is one key to the practical use of NMT. Another important issue is determining how to achieve the functions of a user dictionary, in which users themselves specify phrase translations, in NMT.

4. Future development

NTT Media Intelligence Laboratories will continue to work on R&D aimed at improving the translation accuracy of NMT and adapting it to the business domain efficiently and effectively.

References

[1] Website of translation services by NTT DOCOMO (in Japanese), http://honyaku.idc.nttdocomo.co.jp/
[2] Press release issued by Mirai Translate on December 11, 2017 (in Japanese). https://miraitranslate.com/news/323
[3] Press release issued by NTT EAST on July 3, 2017 (in Japanese). https://www.ntt-east.co.jp/release/detail/20170703_01.html
[4] Press release issued by NTT Communications on January 15, 2018 (in Japanese). http://www.ntt.com/about-us/press-releases/news/article/2018/0115.html
[5] OpenNMT, http://opennmt.net/
[6] Y. Tsuboi, Y. Unno, and J. Suzuki, "Natural Language Processing by Deep Learning," Kodansha, Ltd., 2017 (in Japanese).
[7] D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," Proc. of the 3rd International Conference on Learning Representations (ICLR 2015), San Diego, USA, May 2015.

Trademark notes

Sen Yoshida

Senior Research Engineer, Supervisor, Knowledge Media Project, NTT Media Intelligence Laboratories.

He received a B.E. and M. Info. Sci. from Tohoku University, Miyagi, in 1993 and 1995. He joined NTT Communication Science Laboratories in 1995 and studied multi-agent systems and online social systems. During 2003–2007, he worked at NTT WEST as a member of the IPv6-based commercial public network development team. Since 2007, he has been working on R&D of natural language processing systems, including machine translation, at both NTT Communication Science Laboratories and NTT Media Intelligence Laboratories. He is a member of the Information Processing Society of Japan and the Japanese Society for Artificial Intelligence.

Masahide Mizushima

Senior Research Engineer, Knowledge Media Project, NTT Media Intelligence Laboratories.

He joined NTT in 1992. Since 2012, he has been developing natural language processing systems, including machine translation.

Kimihito Tanaka

Senior Research Engineer, Knowledge Media Project, NTT Media Intelligence Laboratories.

He received a B.E. and M.E. from the University of Tsukuba, Ibaraki, in 1993 and 1995. He joined NTT Human Interface Laboratories in 1995 and studied speech synthesis technologies. During 1999–2009, he worked at NTT EAST as a member of the Customer Premises Equipment Business Division and the R&D Center. Since 2012, he has been working on R&D of artificial intelligence systems, including machine translation, at NTT Media Intelligence Laboratories and R&D Planning Department.

People Flow Prediction Technology for Crowd Navigation

Daisuke Sato, Hisako Shiohara, Masaru Miyamoto, and Naonori Ueda

Abstract

We are investigating the use of incomplete observation data in order to predict the large-scale flow of people for major events such as the Olympic Games and to derive guidance measures in advance to prevent the occurrence of congestion. In this article, we introduce a spatio-temporal variable online prediction algorithm for forecasting congestion risk, and we explain our research and development efforts to utilize the algorithm.

Keywords: spatio-temporal data analysis, congestion risk prediction, artificial intelligence

1. Introduction

Congestion often occurs in places such as tourist regions, commercial areas, and large event spaces where many people gather, and it is conceivable that the congestion will be severe enough in some cases to lead to a major accident. It is therefore very important to provide crowd navigation to help move people along safely and comfortably in such places. This can be done by grasping the whole flow of people in the area, forecasting in advance the occurrence of congestion caused by complex factors, and navigating crowds by taking the appropriate measures. In addition, because the behavior of people changes as they move about, by repeating the cycle of grasping the flow of people that has changed and predicting where congestion will occur next, it is possible to properly navigate the continuously changing people flows (**Fig. 1**).

In this article, we first describe a spatio-temporal variable online prediction algorithm that is applied to predict flows of people in a series of cycles. Then we introduce software that incorporates the prediction technology and explain various approaches to utilizing it.

2. Spatio-temporal variable online prediction

It is important to predict the congestion risk in advance in crowded places such as large events where many people gather. There are several tasks involved in predicting congestion risk. For example, there are tasks to predict in advance the peak congestion time and the size of congestion in places such as entrances that are known to be crowded, as well as tasks to predict the occurrence of unexpected congestion for which we do not know when or where it will happen. We applied our spatio-temporal variable online prediction algorithm to solve the latter prediction problem [1].

An overview of the processing flow of this prediction algorithm is shown in **Fig. 2**. The input of the algorithm is spatio-temporal observation data of the area where the congestion is observed (Fig. 2(a)). For example, this includes position data using GPS (Global Positioning System) and person-count data by position sensors installed at the event site.

Next, spatial data are generated by estimating missing values for each observation time (Fig. 2(b)). With the spatial data of the time series created in this way, we can construct a latent structural model that captures the characteristics of the area common to each time (Fig. 2(c)). This model expresses the latent structure of people flow, and by fitting the model to

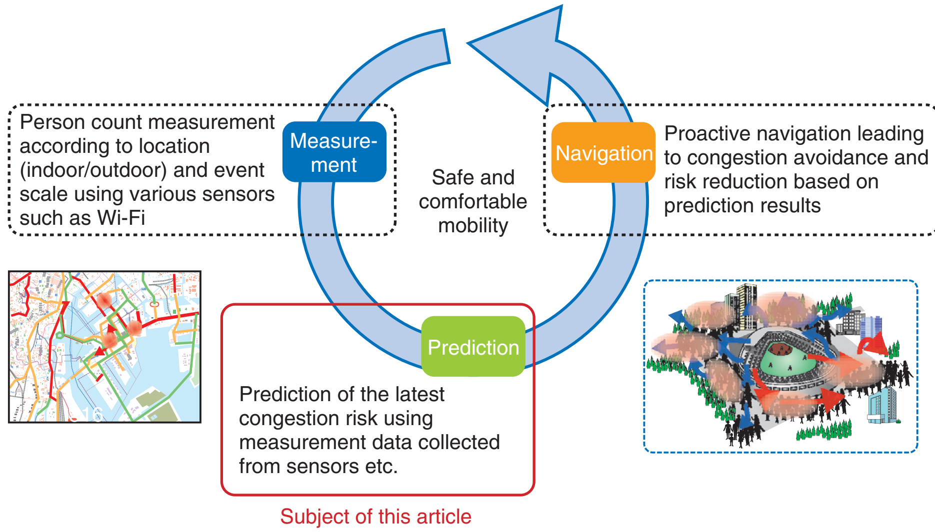


Fig. 1. Cycle to achieve safe and comfortable mobility.

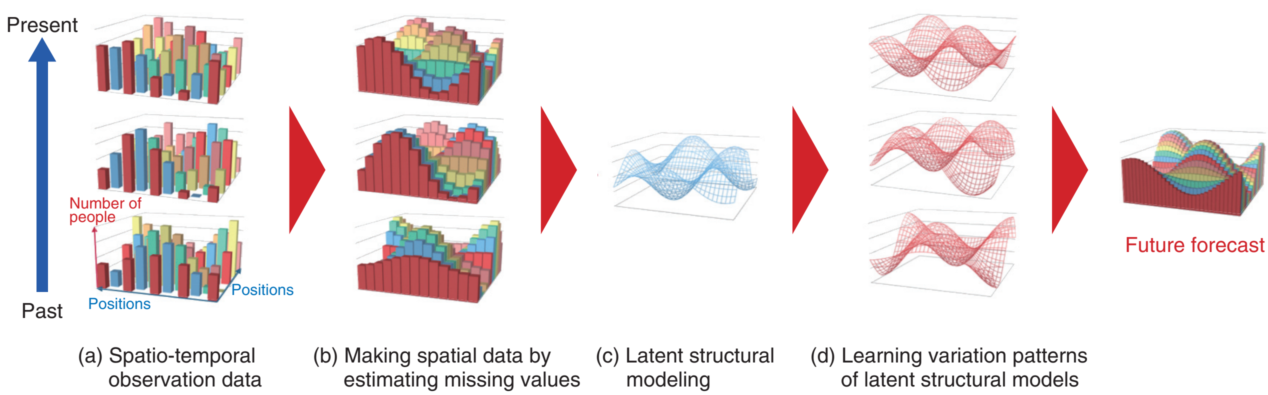


Fig. 2. Processing flow of spatio-temporal variable online prediction.

the input data at each time, the fluctuation in the flow of people can be regarded as a fluctuation of a latent structure. Learning this fluctuation pattern makes it possible to predict the people flow that will occur in the near future (Fig. 2(d)).

The conventional method used for prediction is based on the assumption that data that are spatially and temporally near each other are similar, and this made it difficult to predict sudden fluctuations. However, the spatio-temporal variable online prediction algorithm learns the time-series fluctuation pattern in the latent structural model, and it is therefore effective even when there is large fluctuation.

3. Spatio-temporal people flow data prediction software

The spatio-temporal people flow data prediction software has a function to visualize the current congestion situation from sensor data collected in real time and to predict future congestion situations by using the spatio-temporal variable online prediction algorithm based on accumulated spatio-temporal data.

An overview of the software is shown in **Fig. 3**. The software processes data including general spatio-temporal data, events detected by Bluetooth Low Energy (BLE) beacons, the number of detected

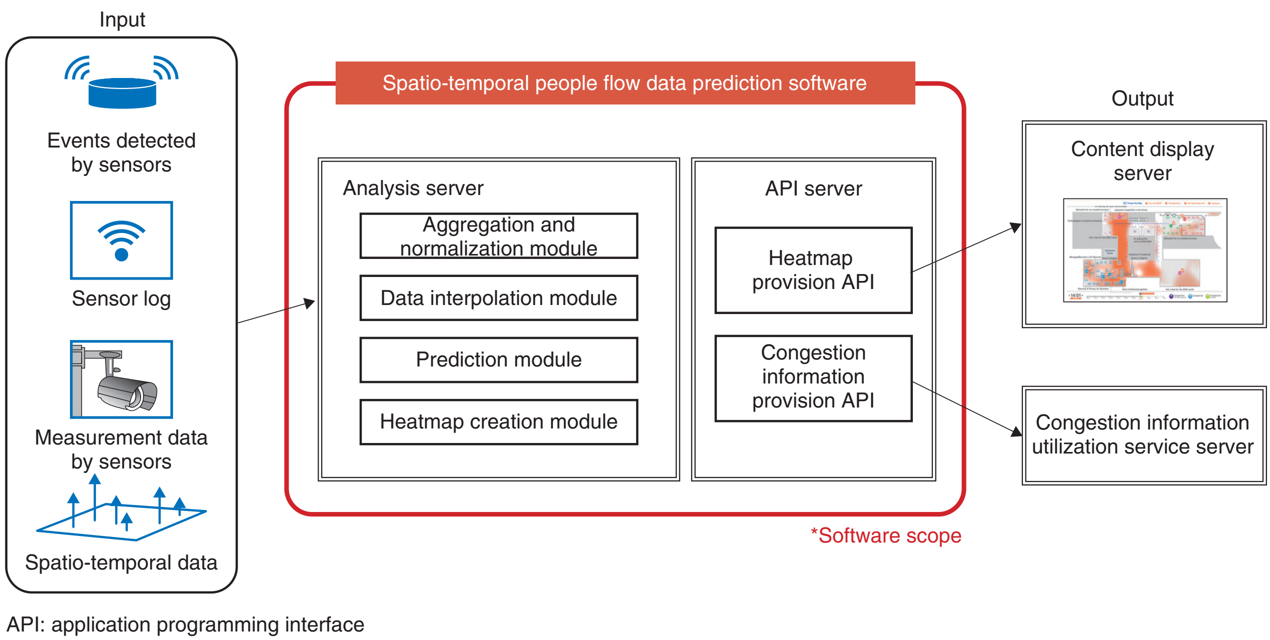


Fig. 3. Overview of spatio-temporal people flow data prediction software.

terminals by Wi-Fi access point logs, and the number of people measured by cameras. The input data are processed by the analysis server, and the results are output through the application programming interface (API) server. There are two kinds of output; the first is current and near future congestion density information, and the other is a heatmap image visualized by expressing congestion information in color.

Sensors for measuring congestion status have different levels of detection accuracy. It is assumed that multiple sensors will be used in combination, depending on the situation and purpose as well as the cost. This software, as a preprocessing step before prediction, has a function to absorb differences between different detection ranges and detection ratios for each sensor and to convert it into congestion information that can be handled uniformly. Thus, data from different types of measurement sensors can be combined and input. This is cost effective and results in higher detection accuracy.

4. Demonstration experiment at large events

We conducted a demonstration experiment using the spatio-temporal people flow data prediction software at two large events where a lot of people gathered (**Fig. 4**).

At Niconico Chokaigi 2016 held in April 2016, we provided services to notify the event operators, facility operators, and visitors of congestion risks by visualizing the congestion status [2, 3]. At this event, we installed about 600 BLE beacons at the venue, provided visitors with an application to upload data that received the radio waves from the beacons, and input the uploaded data to the system.

At NTT R&D Forum 2017 held in February 2017, we set up BLE beacons and Wi-Fi access points in the hall and used data uploaded via the official application as system input [4, 5]. At this event, we provided value in various ways by providing current and predicted congestion information for other services via the API.

At both these events, we used only the latest real-time sensor data acquired on the day to predict the congestion situation 20 minutes ahead. An accuracy evaluation was carried out after the experiment, and it was confirmed that this technique was able to predict large fluctuation of people flow more accurately than the conventional method.

5. Approach to practical application at NTT DOCOMO

NTT DOCOMO is currently investigating services that utilize this technology. The use of population statistics* inferred from DOCOMO's mobile network

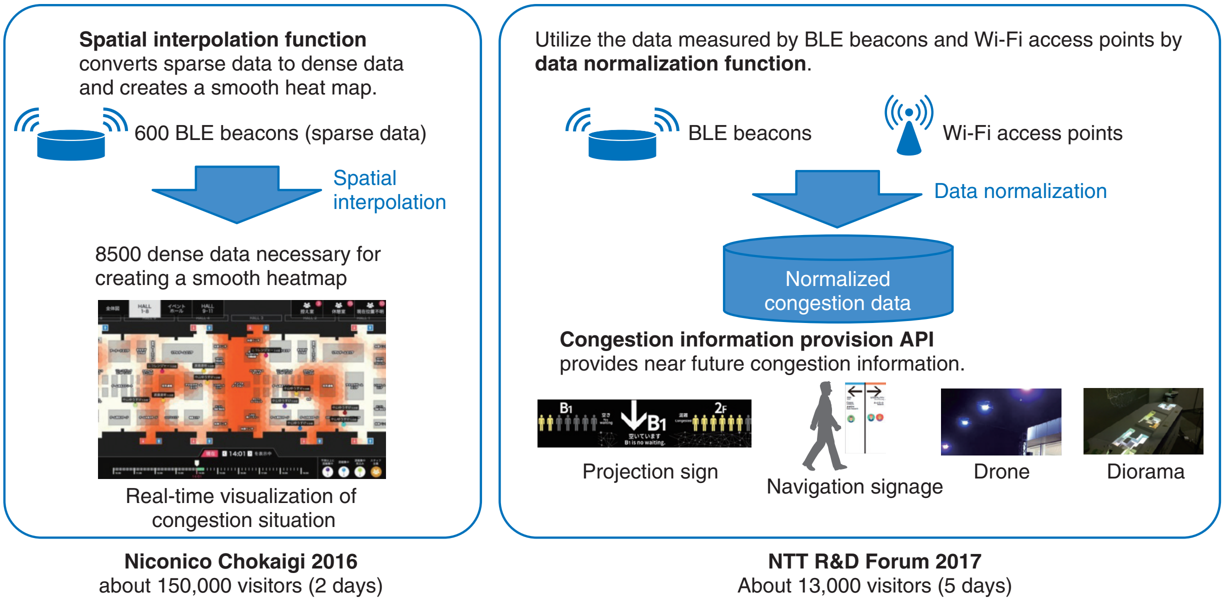


Fig. 4. Demonstration experiment at large events.

together with the spatio-temporal people flow data prediction software makes it possible to predict in 10-minute units the number of people within a 250–500 m square area a few hours ahead [6] (**Fig. 5**).

Utilizing the present and near future numbers of people makes it possible to navigate people to places and routes that are not congested in tourist regions and commercial areas that are crowded. It also makes it possible to efficiently supply the appropriate means of transportation such as buses and taxis when a shortage arises in the supply of transportation. Also, in the event of a disaster or accident, it is expected that rescue teams will be dispatched efficiently according to the scale of the number of victims. At large events such as sports games and fireworks festivals, the nearest stations to the facility are crowded before the start, people concentrate around the facility during the event, and the flow of people spreads to the surrounding area after the event. Consequently, dynamic deployment of security guards using population distribution prediction in the near future can lead to strengthening accident/crime prevention measures.

6. Future outlook

In the future, we will continue to conduct trials and other efforts to make the people flow prediction technology more practical. In addition, we will continue further research and development so that we can provide more new value such as deriving proactive navigation to prevent congestion in advance.

* The population statistics used in this technology consist of statistical information indicating only the number of members of a group for each demographic in a particular time and area, so individual customers cannot be identified. Thus, the behavior of individual customers cannot be determined by others from the population statistics. The population statistics were gathered in compliance with “Mobile Spatial Statistics Guidelines” [7].

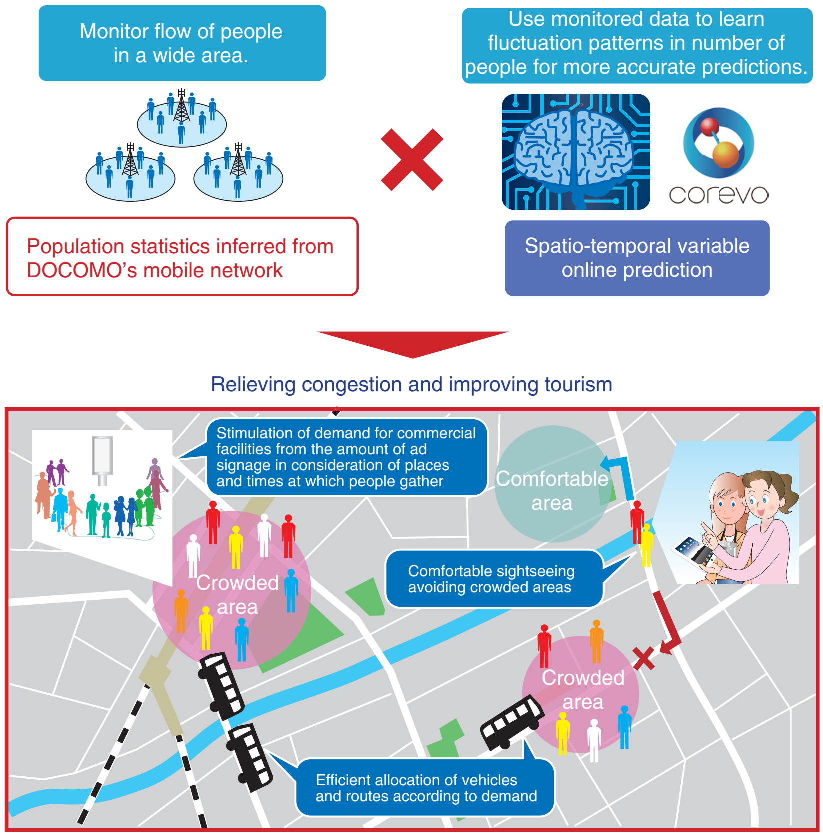


Fig. 5. Practical applications by NTT DOCOMO of proposed technology.

References

[1] N. Ueda, "Proactive People-flow Navigation Based on Spatio-temporal Prediction," Japanese Journal of Applied Statistics, Vol. 45, No. 3, pp. 87–102, 2016 (invited).

[2] NTT press release issued on April 25, 2016 (in Japanese). http://www.ntt.co.jp/news2016/1604/160425a.html

[3] D. Sato, Y. Mihara, Y. Sato, Y. Tanaka, M. Miyamoto, and S. Sakuma, "Visualization Service for Congestion Degree Using BLE Beacons," IPSJ Transactions on Consumer Devices and Systems, Vol. 8, No. 1, pp. 1–10, 2018.

[4] NTT press release issued on February 13, 2017 (in Japanese). http://www.ntt.co.jp/news2017/1702/170213a.html

[5] NTT Service Evolution Laboratories, "R&D Forum Showcase 2017," Business Communication 2017, Vol. 54, No. 4, pp. 22–24 (in Japanese). http://www.bcm.co.jp/ntt-solution-guide/pdf/ntt-2017-04-02.pdf

[6] Press release issued by NTT DOCOMO on September 20, 2017 (in Japanese). https://www.nttdocomo.co.jp/info/news_release/2017/09/20_00.html

[7] NTT DOCOMO, "Mobile Spatial Statistics Guidelines," https://www.nttdocomo.co.jp/english/binary/pdf/service/world/inroaming/inroaming_service/Mobile_Spatial_Statistics_Guidelines.pdf

Trademark notes

Daisuke Sato

Researcher, Proactive Navigation Project, NTT Service Evolution Laboratories.

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

Hisako Shiohara

Senior Research Engineer, Proactive Navigation Project, NTT Service Evolution Laboratories.

She received a B.E. and M.E. in physics from Osaka University in 1990 and 1992. She joined NTT in 1992. She has been developing an advanced system for analyzing people flow data.

Masaru Miyamoto

Senior Research Engineer, Supervisor, Service Innovation Laboratory, NTT Service Evolution Laboratories.

He received a B.E. and M.E. in engineering from Waseda University, Tokyo, in 1995 and 1997. He joined NTT Human Interface Laboratories in 1997. From 2009 to 2014, he was with NTT Research and Development Planning Department. He has been in his current position since 2014. His interests include user experience, management of research and development, and big data.

Naonori Ueda

NTT Fellow, Head of Ueda Research Laboratory and Director of Machine Learning and Data Science Center, NTT Communication Science Laboratories.

He received a B.S., M.S., and Ph.D. in communication engineering from Osaka University in 1982, 1984, and 1992. In 1984, he joined NTT Electrical Communication Laboratories, where he researched image processing, pattern recognition, and computer vision. From 1993 to 1994, he was a visiting scholar at Purdue University, Indiana, USA. In addition to his current positions as head of Ueda Research Laboratory (NTT Fellow) and director of the Machine Learning and Data Science Center, he also serves as deputy director of the RIKEN Center for Advanced Intelligence Project, established in April 2016. He is also a visiting professor at the Graduate School of Informatics, Kyoto University, and at the Graduate School of System Informatics, Kobe University. He is a fellow of the Institute of Electronics, Information and Communication Engineers and a member of IPSJ and the Japan Medical AI Association, and a senior member of the Institute of Electrical and Electronics Engineers.

Cell Encapsulation and 3D Self-assembly Using Multi-layered Polymeric Thin Films

Tetsuhiko Teshima, Hiroshi Nakashima, Yuko Ueno, Satoshi Sasaki, Calum S. Henderson, and Shingo Tsukada

Abstract

Multi-layered thin films are spontaneously folded to form three-dimensional (3D) geometries. In this study, we demonstrate that polymeric thin films are self-folded to encapsulate cells. The films consist of two types of polymers with different mechanical stiffnesses; thereby, the rolled-up 3D tubular architectures with controllable diameters are fabricated based on the strain engineering. A batch release of sacrificial layers forms the multiple cells wrapped in rolled-up films, leading to artificial reconstruction of fiber-shaped cellular 3D constructs with the intrinsic morphologies and functions of living tissues. This system can potentially provide 3D biointerfaces that are necessary for the reconstruction and assembly of functional tissues and implantable tissue grafts.

Keywords: self-folding, polymeric thin film, tissue regeneration

1. Introduction

Three-dimensional (3D) biocompatible structures are strongly required for providing interfaces with cells to achieve cell-based assay [1], tissue engineering [2], and transplantation therapeutics [3]. Above all, the molded structures composed of non-cytotoxic polymer can produce *ex vivo* 3D cellular constructs with intrinsic cell-cell interactions, morphologies, and functions due to high elasticity and biocompatibility [4]. Although a variety of top-down fabrication processes such as laser micromachining or 3D printing techniques have been conventionally proposed, it is technically difficult to construct minute polymeric structures with confined 3D geometries.

To achieve 3D bottom-up fabrication, bimetal film has been utilized for self-assembly of 3D structures with precise geometry [5], which achieves curved bilayer films with heterogeneous mechanical properties. This system enables the assembly of a variety of 3D structures including pyramidal [6], tubular [1], and helical structures [7].

However, this bilayer mechanism is not sufficiently applied to polymeric 3D geometries due to the difficulty of creating polymeric bilayers. A 3D bottom-up fabrication of polymer 3D structures would provide an interface with cells to build 3D cell-laden architectures. Thus, an alternative method to produce a polymeric bilayer with heterogeneous mechanical properties is highly necessary. Cell-friendly polymers are surely applicable for *ex vivo* 3D cellular architectures with cell-cell connections, intrinsic morphologies, and various physiological functions [8].

In this study, we found that multi-layered polymeric films with heterogeneous mechanical properties can form self-folded rolled shapes (micro-rolls) [9]. Integrating the films with the non-cytotoxic batch release of a hydrogel-based sacrificial layer [10] makes it possible to autonomously fold the films into cylindrical shapes based on differential strain gradients

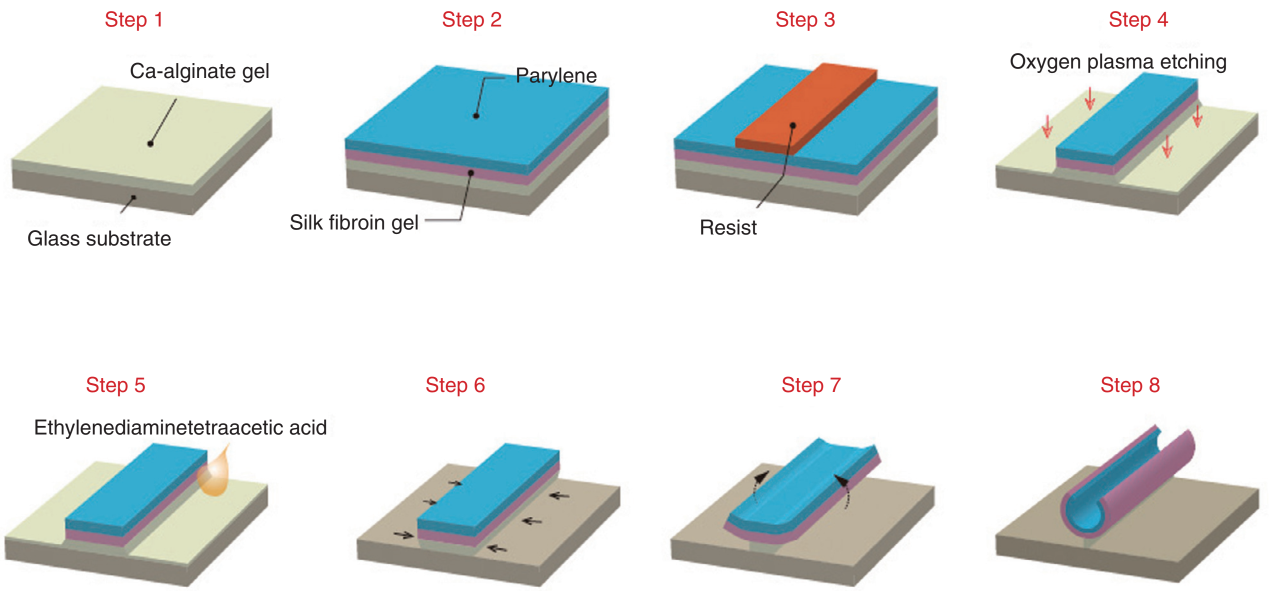


Fig. 1. Self-assembling thin film fabrication process. (Steps 1–4) Three layers of silk fibroin gel, parylene, and Ca-alginate gel are laminated and then processed using lithography to form an arbitrary 2D micro-pattern. (Steps 5–8) The Ca-alginate gel layer is removed with ethylenediaminetetraacetic acid to release the thin film from the substrate.

that depend on the film thickness. Various 3D cell-laden microstructures were formed from two-dimensional (2D) geometrical micro-patterns, enabling the embedded cells to migrate and connect with each other. Within a micro-cavity, the cells tend to form the desired 3D architectures and synchronize their behavior. We also used various cell-lines and primary cultured cells to demonstrate that they can reconstitute the intrinsic cellular morphologies and functions. This system is potentially applicable to realizing 3D biointerfaces for such applications as the reconstruction of functional tissues and implantable tissue grafts.

2. Materials and methods

We fabricated multi-layered films composed of mechanically stable and biocompatible polymers, namely, poly(chloro-*p*-xylylene) (parylene-C) [10, 11] and silk fibroin [12] on an underlying sacrificial layer. We selected alginate hydrogel (calcium alginate (Ca-alginate)) as the sacrificial layer since it is dissolved with chelate agents. Thus, it achieves a spontaneous and non-cytotoxic batch release process with arrayed cell-laden films. As the film components, we used silk fibroin extracted from *Bombyx mori* [13] and parylene-C thanks to their mechanically robust features, high optical transparency, and excellent biocompatibility.

The fabrication process is shown in **Fig. 1** [9]. First, sodium alginate was coated on the glass substrates and then immersed in calcium chloride to hydrogelate alginate with calcium ions (Ca^{2+}) (Step 1). Next, the aqueous silk fibroin was coated and immersed in methanol for gelation in accordance with previously reported protocols [12] and laminated with a parylene layer by chemical vapor deposition (Step 2). Then photolithographical micro-patterning of the photoresist was carried out (Step 3). Thereafter, the trilaminar films consisting of alginate hydrogel, silk fibroin, and parylene were etched with oxygen plasma through a photoresist mask to fabricate a micro-patterned film array (Step 4). When the Ca-alginate layer was dissolved by adding ethylenediaminetetraacetic acid (EDTA), it gradually shrank from the edges and released the remaining bilayers as shown in Fig. 1 (Steps 5–8). SEM (scanning electron microscopy) images show identifiable trilaminar films with highly defined geometries (**Fig. 2**).

3. Results

The self-folding process of the films and the manipulation and reconstruction of cells are explained in

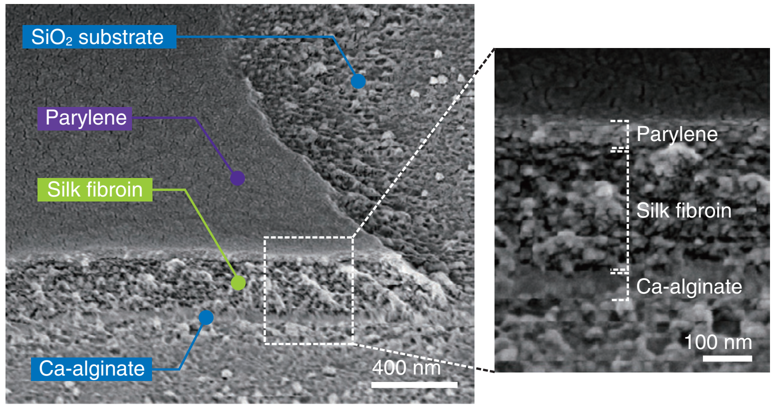


Fig. 2. Low-magnification SEM images of micro-patterned films on glass substrate. The white box indicates enlarged areas. Reprinted from Teshima *et al*. [9]. Copyright 2017.

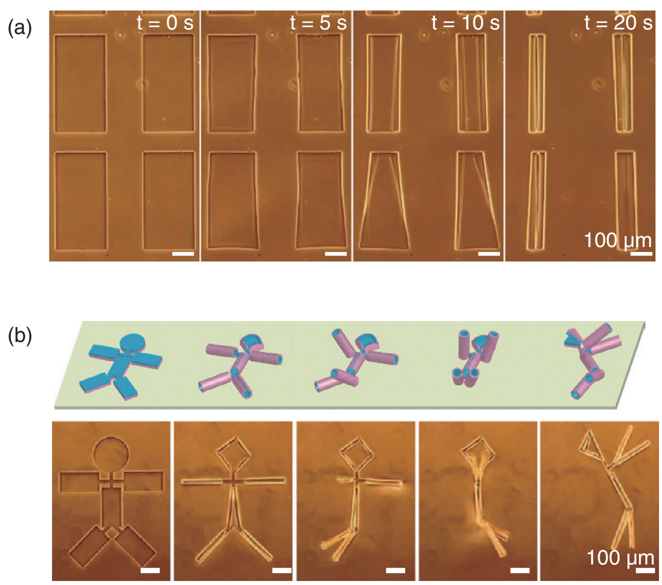


Fig. 3. Using 2D thin films to fabricate arbitrary 3D structures. (a) Fabrication of a cylindrical structure using a rectangular thin film. (b) 3D self-folding with the 3D doll geometry. Reprinted from Teshima *et al*. [9]. Copyright 2017.

this section.

3.1 Self-folding of polymeric bilayer film

The combination of silk fibroin with conformally deposited parylene-C functions as a mechanically heterogeneous bilayer [9]. After the Ca-alginate layer was dissolved by EDTA, the released bilayer films transformed into controlled 3D geometries (**Fig. 3(a)**). This dissolution initiates the formation of tubular structures through self-folding. The curvature radius of the self-folded micro-roll reaches an equilibrium value, which is delayed by more than 10 s after the dissolution of the sacrificial layer. Confocal reconstructed images show that the dimensions of the micro-patterns had a noticeable effect on the diameter of the resulting tubular structures.

In addition to the cylindrical shape, various 3D geometries can be controlled and self-folded depending on the 2D micro-pattern design. For instance, a unique doll-shaped 3D structure was self-folded by combining rectangles with cross-shaped hinges, where the rectangles correspond to the body and

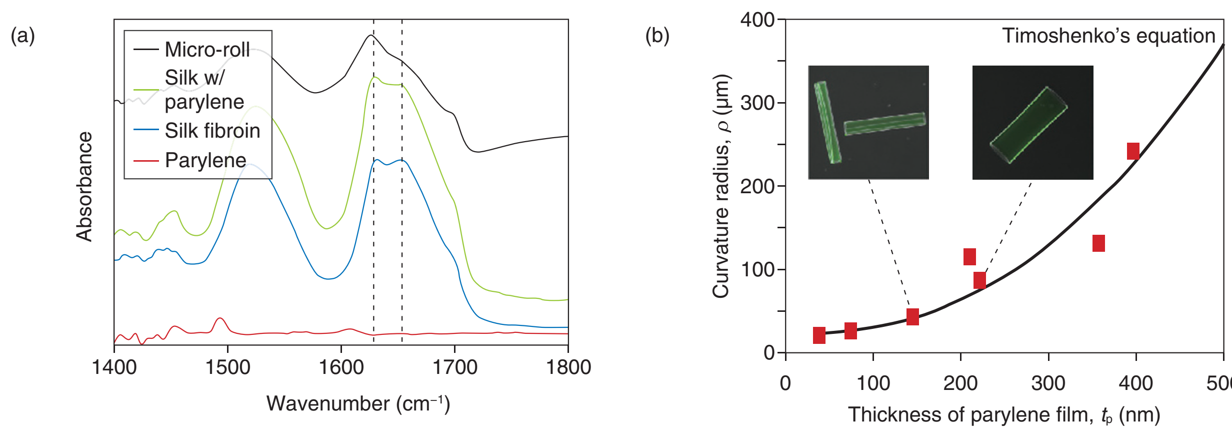


Fig. 4. Characterization of self-folded micro-rolls. (a) FTIR spectra of crystalline silk fibroin in micro-rolls. Resonance absorbances centered at 1650 and 1625 cm^{-1} respectively indicate a random coil conformation in amorphous silk protein and the helical β-sheet conformation of annealed protein. (b) Plot of average curvature radius of free-floating, self-folded 200×400 μm^2 films for various parylene film thicknesses. Reprinted from Teshima *et al.* [9]. Copyright 2017.

limbs and a circle forms the head (**Fig. 3(b)**). The cross portion connecting the arms retained the backbone shape, making it possible to maintain the 3D doll geometry. Hence, by incorporating the branched structure in the pattern interior, it is possible to produce more complex 3D shapes such as dolls, zigzags, and caged skeletal frames.

We used Fourier transform infrared spectroscopy (FTIR) to characterize the molecular conformation of the self-folded micro-rolls [9]. Silk fibroin was methanol-treated to induce β-sheet conformation in its crystalline layer. The spectra ranging from 1450 to 1700 cm^{-1} are assigned to the absorption of the peptide backbones in silk fibroin (**Fig. 4(a)**). All of the films containing the silk fibroin have spectra centered at 1625 cm^{-1}, which is characteristic of the antiparallel β-sheet conformation in crystalline silk. However, the folded micro-rolls exhibit a reduction in absorbance around 1650 cm^{-1}, suggesting a transition from the helical conformation of amorphous silk to a β-sheet-rich conformation. There are no particular differences in the peak wavelengths of the micro-rolls with varying curvature radii. In addition, thanks to its high elastic modulus, the β-sheet-rich matrix was maintained without a phase transition after the self-folding step. Accordingly, the inducement of the β-sheet-rich conformation is an essential process to maintain the 3D shape of the micro-rolls and make them sufficiently durable for long-term incubation.

We applied quantitative predictions of 2D designs to the final 3D geometry to allow us to control the micro-roll architecture [9]. Since the underlying mechanism of self-folding is the stiffness mismatch in the bilayer, the curvature of the micro-rolls ($1/\rho$) was calculated using Timoshenko's bimorph beam theory [14] (**Fig. 4(b)**). This theory predicts the curvature radius (ρ) based on the elastic modulus, layer thickness, and strain. For simplicity, we assumed a constant thickness of silk fibroin. The experimental value ρ is plotted in the equilibrium state as a function of thickness of parylene from 30 to 400 nm. We confirmed that ρ mostly follows the trend predicted by the bimorph beam theory.

3.2 Encapsulation and manipulation of cells

The key challenge with cell-laden micro-rolls is to fabricate tissue-mimicking tubular structures with non-cytotoxicity and to confirm their cellular function. We encapsulated human embryonic kidney (HEK) cells within the micro-rolls. When EDTA was added, the cells on the 2D film surface were encapsulated inside the 3D micro-rolls through the self-folding of the bilayer (**Fig. 5**). Confocal cross-sectional images revealed that the arrayed micro-rolls encapsulated numerous cells at a time, which were separated from the cells outside the patterned area. A live/dead assay revealed that more than 95% of the cells were still alive inside the micro-rolls after exposure to EDTA. This indicates that the self-folding and encapsulation process was sufficiently gentle to avoid cell damage. Since the curvature radius and length of the micro-rolls are controllable, the number of encapsulated

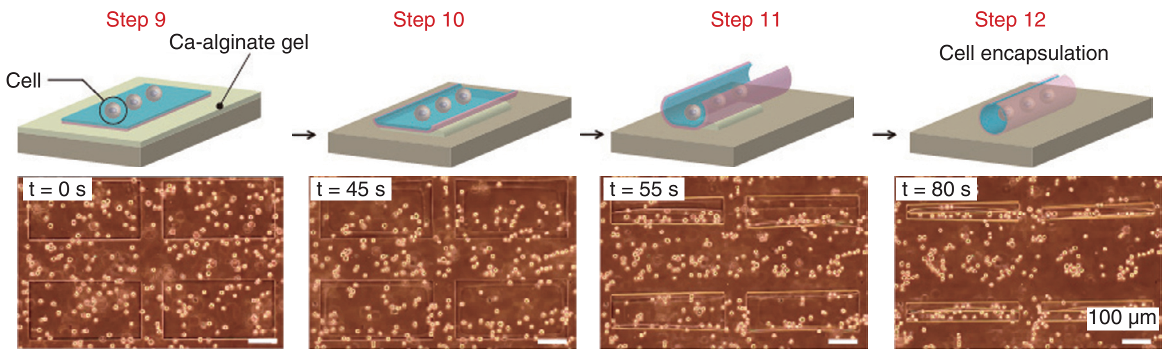


Fig. 5. Encapsulation of cells. As the Ca-alginate gel layer is removed by EDTA, the thin film deforms into a tubular structure. At the same time, cells near the surface of the film are encapsulated inside it. Reprinted from Teshima *et al.* [9]. Copyright 2017.

cells can be easily controlled with the cell density and the volume of the micro-roll. The use of an array of micro-rolls potentially makes it possible to create highly parallel assays that can be used to investigate cellular behaviors including migration, proliferation, and differentiation.

The high biocompatibility and transparency of the micro-rolls enable stable culturing of encapsulated cells. During the incubation, the encapsulated cells migrated and proliferated only inside the micro-rolls without any migration out of the micro-rolls. This leads to the formation of cell-cell contacts along the inner surfaces of the micro-rolls. The micro-rolls with encapsulated cells not only provide the transparency needed for microscope observations, but they also have sufficient mechanical strength to withstand the cell traction force. HEK cells exhibit relatively strong cell-to-cell contact, so the cells in a micro-roll aggregated and filled the micro-roll cavity after one day of culturing. Specifically, the cells proliferated, filled the micro-roll cavity, and then migrated onto the outer silk fibroin surface of the micro-rolls. The shapes of the cell clusters differed depending on the micro-roll volume and the cell-line's adhesive properties.

A pico-liter fluid flow done with a glass capillary enables manipulation of cell-laden micro-rolls [9]. The smooth withdrawal and release of cells from the capillary enables us to relocate cells to the target location (**Fig. 6(a)**). When the micro-cells were being conveyed inside the glass capillary, the micro-rolls stably maintained their cylindrical shapes without buckling or bending. Therefore, individual cells inside the micro-rolls were collected and transported to the targeted position to induce cell-to-cell contact (**Fig. 6(b)**). Note that the cells on micro-rolls were manipulated without stripping them from their growth surfaces, which allowed the adhesive state of the manipulated cells to be preserved. This has the potential to facilitate high-order cellular assembly for potential use in tissue engineering by enabling the manipulation of individual cell types into any desired multi-cell-line combination.

3.3 Reconstruction of functional tissue-like structures

Physiological functions within the cell-laden micro-rolls were investigated using two types of primary cultured cells: cardiomyocytes and hippocampal cells [9]. In contrast to the cell lines, neither of these proliferate spontaneously, and their cell bodies tend to reside in the micro-rolls. After five days of culture, the cardiomyocytes aggregated to form a more than 2-cm-long tubular structure. A time-dependent displacement at corresponding pixel locations was estimated by subtracting the average intensity from the optical images (**Fig. 7**). The intensity corresponding to the displacement shows that all the connecting cells synchronously contracted in the same period at approximately 1 Hz. Time-lapse fluorescence image measurement of fluo-4-labeled cells shows that the oscillation of intracellular Ca^{2+} propagated and was synchronized within the micro-rolls. The microenvironment provided by micro-rolls of any length was able to regulate the cell migration and cell-cell interactions and drive intrinsic morphologies and functions.

Primary hippocampal cells were also encapsulated to reconstitute neural tissues. The encapsulated cells aggregated and subsequently extended their dendrites

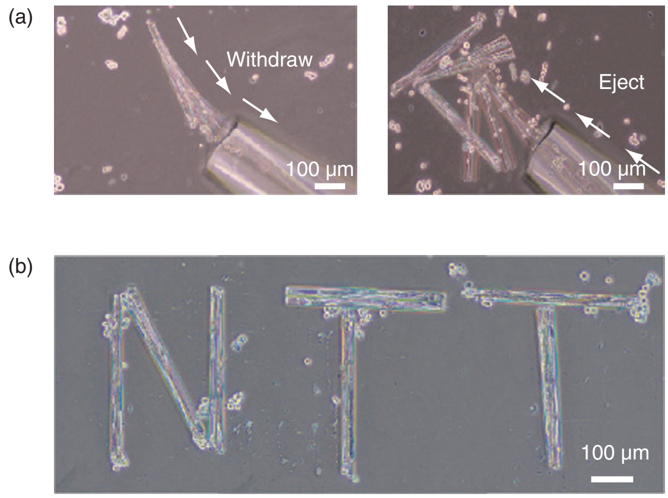


Fig. 6. (a) Manipulation of 400-µm-long cell-laden micro-rolls in culture medium. The micro-rolls can be withdrawn, relocated, and ejected with a picoliter flow using glass capillaries. (b) Micrograph of manipulated multiple cell-laden micro-rolls in culture medium.

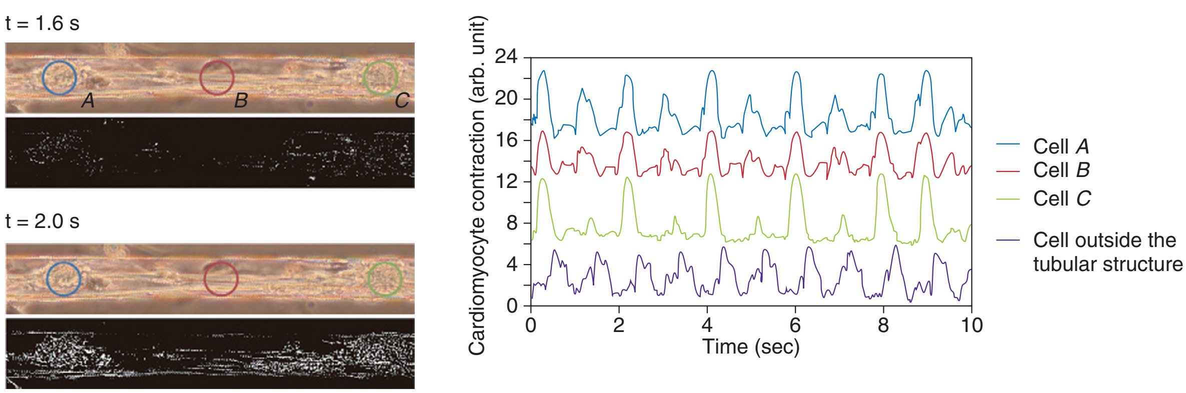


Fig. 7. Reconstruction of minute myocardial tissue structures. (Left) Phase contrast micrographs of cardiomyocytes cultured inside tubular structures, and black and white images showing the displacement of the cells during a heartbeat. (Right) Graph showing the amount of displacement during a heartbeat. In the tubular structure containing the cardiomyocytes, the cells are bound to each other to form a fibrous biological tissue-like structure. The cardiomyocytes beat in unison with a rhythm that is synchronized and constant. Reprinted from Teshima *et al.* [9]. Copyright 2017.

in the micro-roll direction. After three days of culturing, the neurite-mediated cell-cell connection between the paired cell aggregation resulted in the formation of unidirectional neural cell fibers (**Fig. 8(a)**). Somas and their neurites were both located inside the micro-rolls without any protrusions. Interestingly, the neurites were further extended outside the micro-rolls, where they connected with neighboring cells. Immunostaining of neurites and somas showed that both were positionally restricted within the micro-rolls. Some spheroids were located at the edge of the micro-rolls, and the others were evenly distributed inside the micro-rolls. Regardless of their location, the dendrites were extended and evenly distributed in the micro-rolls after long-term incubation.

To further investigate the accessibility of neurites to the outer surface, we examined cellular responses to external stimulation with added glutamic acid. When the glutamic acid receptors in the hippocampal cells were activated, the intracellular Ca^{2+} were detectable by calcium indicators using confocal laser scanning

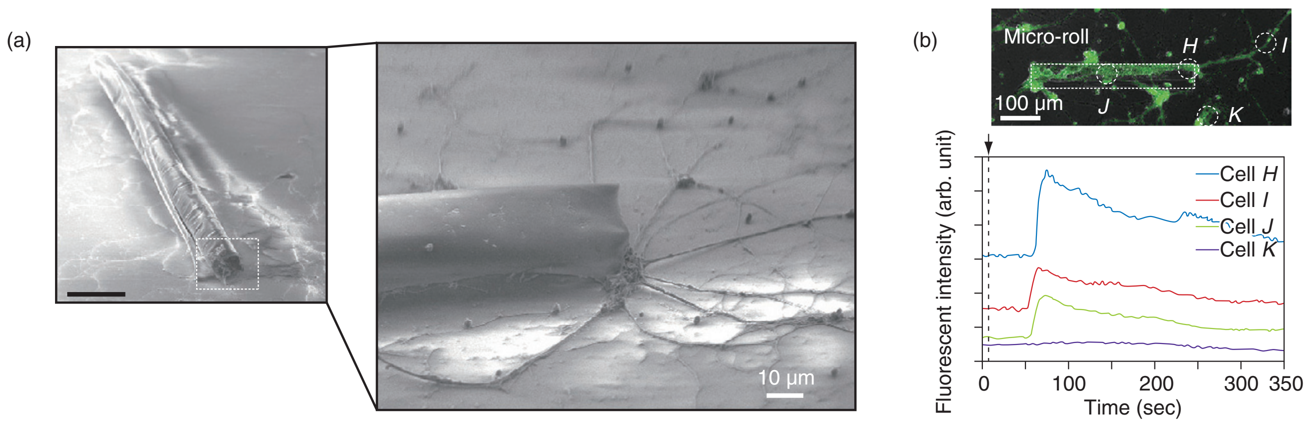


Fig. 8. (a) SEM images of hippocampal cell-laden micro-rolls cultured on poly-D-lysine/laminin-coated dishes. (b) Relative fluorescence intensity profiles of Ca^{2+} oscillations in hippocampal cells labeled with Fluo-4 and stimulated by adding glutamic acid in the segment shown in the confocal fluorescent image. Reprinted from Teshima *et al.* [9]. Copyright 2017.

microscopy (**Fig. 8(b)**). Fluo-4 labeled cells exhibited spontaneous Ca^{2+} oscillations that were synchronized across the cells via synaptic connections (cells *H*, *I*, *J*), whereas the cells that were not connected via neurites showed no oscillation signals (cell *K*). This result indicates that the encapsulated cells come into direct contact and build up a cellular network with the cells outside of the micro-rolls through hollow structures in order to transmit and exchange information with their neighboring cells. After two weeks of culturing, the extended neurites from the manipulated cell-laden micro-rolls were connected and bridged to achieve cell-to-cell contact between neighboring micro-rolls. Daily medium exchanges and cell maintenance ensures the long-term culturing of encapsulated cells without cytotoxicity. Consequently, the relocated cells can recreate intercellular communication with the cells outside the micro-rolls, serving as building units of higher-order tissue assemblies and potential implantable grafts *in vivo*.

4. Summary

In summary, we proposed that a polymeric multilayered film transforms into a programmed 3D configuration that depends on 2D geometry [9]. Conventionally, it is difficult to induce a stress gradient in homogeneous films due to weak adhesion in layered hydrogel film [15]. The micro-rolls proposed in this study are improved by coating the film with parylene using chemical vapor deposition [10, 16], which results in the fabrication of single-cell-sized 3D structures with highly precise geometry in the same manner as bimetal films. The self-folding mechanism is based on the bending of bilayer films; thus, it is potentially applicable to any constituent materials.

The process of fabricating micro-rolls does not require cytotoxic etchant, unlike conventional self-foldable inorganic bimetal thin films. This is attributable to the non-cytotoxic dissolution of the sacrificial Ca-alginate layer, allowing for the batch encapsulation of multiple cells. The biocompatibility of silk fibroin and parylene also ensured the encapsulation, long-term culturing of more than one month, and facile manipulation of encapsulated cells without cytotoxicity. Notably, the higher elasticity of parylene and silk fibroin enables the embedded cell to develop a fiber-shaped construct. As a result, the contraction of the primary myocytes and synaptic connections of highly oriented primary neural cells was achieved. The transparency of the polymers used facilitates the study of any cellular response such as calcium oscillation and the fluorescent observation of an immunostained cytoskeleton. This method could be expanded to many other adherent cell lines for the reconstruction of fiber-shaped functional tissues that mimic muscle fibers, blood vessels, and nerve networks *in vitro*.

5. Future prospects

The system demonstrated here can be used to form scaffolding structures toward regenerative medicine and the behavior analysis of single cells. It also enables encapsulated cells to be freely manipulated in the format of 3D shapes. Importantly, the adherent cells inside the micro-rolls are stably encapsulated and manipulated while retaining their adhesive properties. Although the cell-laden micro-rolls were entirely manually manipulated in this study, our manipulation platform is incorporated with an encapsulated magnet-responsive layer [11, 17] or a microfluidic system [18] for fast, reliable, and automated manipulation. Additionally, the distance-mediated interconnection of elongated neurites between more than two neuron-laden micro-rolls is worthy of a detailed investigation involving optimizing their sizes, location, and patterns from the perspective of building synthetic neural circuits and tissue engineering. This method could be expanded in the future to reconstruct tissues with various shapes *in vitro* that are suitable for exploring single-cell assays, tissue engineering, and implantable *ex vivo* tissue grafts.

References

[1] W. Xi, C. K. Schmidt, S. Sanchez, D. H. Gracias, R. E. Carazo-Salas, S. P. Jackson, and O. G. Schmidt, "Rolled-up Functionalized Nanomembranes as Three-dimensional Cavities for Single Cell Studies," Nano Lett., Vol. 14, No. 8, pp. 4197–4204, 2014.

[2] H. Onoe, T. Okitsu, A. Itou, M. Kato-Negishi, R. Gojo, D. Kiriya, K. Sato, S. Miura, S. Iwanaga, K. Kuribayashi-Shigetomi, Y. T. Matsunaga, Y. Shimoyama, and S. Takeuchi, "Metre-long Cell-laden Microfibres Exhibit Tissue Morphologies and Functions," Nat. Mater., Vol. 12, No. 6, pp. 584–590, 2013.

[3] T. M. Chang, "Therapeutic Applications of Polymeric Artificial Cells," Nat. Rev. Drug Discov., Vol. 4, No. 3, pp. 221–235, 2005.

[4] V. Stroganov, M. Al-Hussein, J. U. Sommer, A. Janke, S. Zakharchenko, and L. Ionov, "Reversible Thermosensitive Biodegradable Polymeric Actuators Based on Confined Crystallization," Nano Lett., Vol. 15, No. 3, pp. 1786–1790, 2015.

[5] O. G. Schmidt and K. Eberl, "Nanotechnology: Thin Solid Films Roll up into Nanotubes," Nature, Vol. 410, No. 6825, p. 168, 2001.

[6] K. Malachowski, M. Jamal, Q. R. Jin, B. Polat, C. J. Morris, and D. H. Gracias, "Self-folding Single Cell Grippers," Nano Lett., Vol. 14, No. 7, pp. 4164–4170, 2014.

[7] Z. L. Wu, M. Moshe, J. Greener, H. Therien-Aubin, Z. H. Nie, E. Sharon, and E. Kumacheva, "Three-dimensional Shape Transformations of Hydrogel Sheets Induced by Small-scale Modulation of Internal Stresses," Nat. Commun., Vol. 4, p. 1586, 2013.

[8] O. Guillame-Gentil, O. Semenov, A. S. Roca, T. Groth, R. Zahn, J. Vörös, and M. Zenobi-Wong, "Engineering the Extracellular Environment: Strategies for Building 2D and 3D Cellular Structures," Adv. Mater., Vol. 22, No. 48, pp. 5443–5462, 2010.

[9] T. F. Teshima, H. Nakashima, Y. Ueno, S. Sasaki, C. S. Henderson, and S. Tsukada, "Cell Assembly in Self-foldable Multi-layered Soft Micro-rolls," Sci. Rep., Vol. 7, p. 17376, 2017.

[10] T. Teshima, H. Onoe, K. Kuribayashi-Shigetomi, H. Aonuma, K. Kamiya, H. Ishihara, H. Kanuka, and S. Takeuchi, "Parylene Mobile Microplates Integrated with an Enzymatic Release for Handling of Single Adherent Cells," Small, Vol. 10, No. 5, pp. 912–921, 2014.

[11] T. Teshima, H. Onoe, H. Aonuma, K. Kuribayashi-Shigetomi, K. Kamiya, T. Tonooka, H. Kanuka, and S. Takeuchi, "Magnetically Responsive Microflaps Reveal Cell Membrane Boundaries from Multiple Angles," Adv. Mater., Vol. 26, No. 18, pp. 2850–2856, 2014.

[12] T. Teshima, H. Nakashima, N. Kasai, S. Sasaki, A. Tanaka, S. Tsukada, and K. Sumitomo, "Mobile Silk Fibroin Electrode for Manipulation and Electrical Stimulation of Adherent Cells," Adv. Funct. Mater., Vol. 26, No. 45, pp. 8185–8193, 2016.

[13] D. N. Rockwood, R. C. Preda, T. Yücel, X. Q. Wang, M. L. Lovett, and D. L. Kaplan, "Materials Fabrication from *Bombyx mori* Silk Fibroin," Nat. Protoc., Vol. 6, No. 10, pp. 1612–1631, 2011.

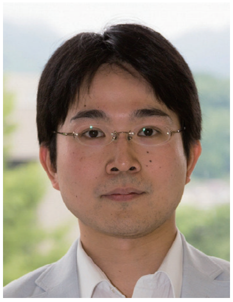
[14] S. Timoshenko, "Analysis of Bi-metal Thermostats," J. Opt. Soc. Am., Vol. 11, No. 3, pp. 233–255, 1925.

[15] M. Jamal, A. M. Zarafshar, and D. H. Gracias, "Differentially Photo-crosslinked Polymers Enable Self-assembling Microfluidics," Nat. Commun., Vol. 2, p. 527, 2011.

[16] T. Teshima, H. Onoe, S. Tottori, H. Aonuma, T. Mizutani, K. Kamiya, H. Ishihara, H. Kanuka, and S. Takeuchi, "High-resolution Vertical Observation of Intracellular Structure Using Magnetically Responsive Microplates," Small, Vol. 12, No. 25, pp. 3366–3373, 2016.

[17] S. Tottori, L. Zhang, F. M. Qiu, K. K. Krawczyk, A. Franco-Obregón, and B. J. Nelson, "Magnetic Helical Micromachines: Fabrication, Controlled Swimming, and Cargo Transport," Adv. Mater., Vol. 24, No. 6, pp. 811–816, 2012.

[18] W. H. Tan and S. Takeuchi, "A Trap-and-release Integrated Microfluidic System for Dynamic Microarray Applications," Proc. Natl. Acad. Sci. USA., Vol. 104, No. 4, pp. 1146–1151, 2007.



Tetsuhiko Teshima

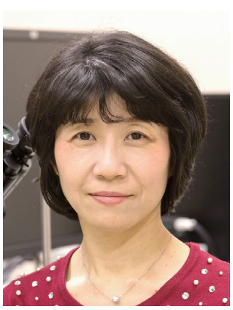
Researcher, NTT Basic Research Laboratories.

He received a B.E., M.E., and Ph.D. from the School of Agriculture, Arts and Science, and Information Science and Technology, University of Tokyo, in 2009, 2011, and 2014. During 2011–2014, he was a research fellow of the Japan Society for the Promotion of Science (JSPS). He joined NTT Basic Research Laboratories in 2014 and is currently working on a silk fibroin-based biocompatible interface with skin and *in vivo* tissue to monitor vital data. He is a member of the Materials Research Society of the USA and the Japan Society of Applied Physics (JSAP).

Hiroshi Nakashima

Senior Research Scientist, Supervisor, NTT Basic Research Laboratories.

He received a B.E., M.E., and Ph.D. in applied chemistry from Waseda University, Tokyo, in 1995, 1997, and 2002. He joined NTT Basic Research Laboratories in 1997. He has since been researching the synthesis and control of optoelectrical properties of (semi)conductive polymers and nano-bio materials. He is a member of the Chemical Society of Japan (and its Colloid and Surface Chemistry Division) and JSAP (and its Molecular Electronics and Bioelectronics Division).



Yuko Ueno

Senior Research Scientist, Supervisor, Distinguished Researcher, NTT Basic Research Laboratories.

She received a B.S., M.S., and Ph.D. in chemistry from the University of Tokyo in 1995, 1997, and 2002. She joined NTT Integrated Information & Energy Systems Laboratories in 1997. She was a post-doctoral fellow at the University of California at Berkeley and at Lawrence Berkeley National Laboratory, Berkeley, CA, USA, from 2004 to 2005. She is a member of the Japan Society of Analytical Chemistry, the Chemical Society of Japan, the Spectroscopical Society of Japan, JSAP, the Japan Society of Vacuum and Surface Science, the Society for Chemistry and Micro-Nano Systems, and the Institute of Electronics, Information and Communication Engineers.

Satoshi Sasaki

Senior Research Scientist, NTT Basic Research Laboratories.

He received a B.S., M.S., and Ph.D. in applied physics from the University of Tokyo in 1988, 1990, and 1993. He joined NTT Basic Research Laboratories in 1993. Since then, he has been studying transport properties of mesoscopic devices, especially quantum dots. He is a member of the Physical Society of Japan and JSAP.

Calum S. Henderson

Internship Student, NTT Basic Research Laboratories.

He is currently studying toward a master's degree in the School of Chemistry, the University of Edinburgh, Scotland. He joined NTT Basic Research Laboratories in 2017 through an industrial placement program. He will join the ISIS Neutron and Muon Source, Oxfordshire, England as a researcher in 2018.

Shingo Tsukada

Senior Distinguished Researcher, NTT Basic Research Laboratories.

He received an M.D. from Toyama Medical and Pharmaceutical University and his medical license in 1990, and a Ph.D. in medicine from University of Tsukuba, Ibaraki, in 2003. He was a visiting researcher at the University of California at San Diego, CA, USA, during 2003–2005. He joined NTT Basic Research Laboratories in 2010. He has been studying the mechanisms and activity control of signal transduction of brain cells. He is a member of the Society for Neuroscience, the Physiological Society of Japan, JSAP, the Japan Neuroscience Society, the Japanese Circulation Society, and the Japanese Orthopaedic Association.

Multi-layer SDN Control Technology

Takuya Tojo, Shingo Okada, Yoshiyuki Hirata, and Seisho Yasukawa

Abstract

To cope with the transition of fixed and mobile networks to fifth-generation services and the further spread of cloud services, NTT Network Technology Laboratories is developing multi-layer software-defined networking (SDN) control technology to achieve integrated control of the IP (Internet protocol) and optical layers, on-demand support, and automatic network operations. In this article, we provide an overview of multi-layer SDN control technology and describe a technical verification test.

Keywords: multi-layer SDN, telemetry, automatic control

1. Introduction

The modern telecom network consists mainly of two types of layers: the optical layer for interconnecting datacenters via optical fiber, and the Internet protocol (IP) layer that performs packet transfer control. The optical layer and IP layer differ in their lead times for facility construction and the technologies that they use, so the conventional approach has been to design each layer independently and optimize the operation and management of each. From here on, however, the migration of the mobile network to fifth-generation (5G) and the further spread of cloud services are expected to create an environment in which diverse services exploiting the high-speed and low-delay characteristics of 5G can be provided, and services can be flexibly rolled out on the cloud as needed by users.

For telecom networks, meanwhile, it is becoming increasingly necessary to provide on-demand communications between users and the cloud and between users themselves. Additionally, a network with high on-demand characteristics features highly dynamic traffic volumes and a tendency for traffic imbalances to occur in specific network intervals. There is therefore a need for a mechanism that can calculate the amount of free circuits and other resources throughout the network and allocate the resources needed, and a mechanism that can eliminate traffic imbalances through traffic routing control.

To cope with this spread of 5G mobile communications and cloud services, we are developing multi-layer software-defined networking (SDN) control technology for performing integrated control of the optical layer and IP layer, which have traditionally been optimized separately, providing on-demand support, and automating network operations.

2. Multi-layer SDN control technology

Multi-layer SDN control technology performs integrated control of the optical layer and IP layer from an SDN controller to provide service-dependent paths such as IP-VPNs (virtual private networks) or Ethernet leased lines in an on-demand manner according to the user's service request (**Fig. 1**). Another aim of this technology is to use each layer's resources to the maximum extent possible by performing packet routing control on the IP layer and optical-wavelength routing control on the optical layer and keeping the network in an optimal state. To this end, NTT Network Technology Laboratories is implementing the functions shown in **Fig. 2** based on the Open Network Operating System (ONOS) [1], an open-source SDN controller, and working to establish multi-layer SDN control technology.

2.1 On-demand virtual path allocation

The allocation of an on-demand virtual path is achieved by implementing (1) a function for configuring

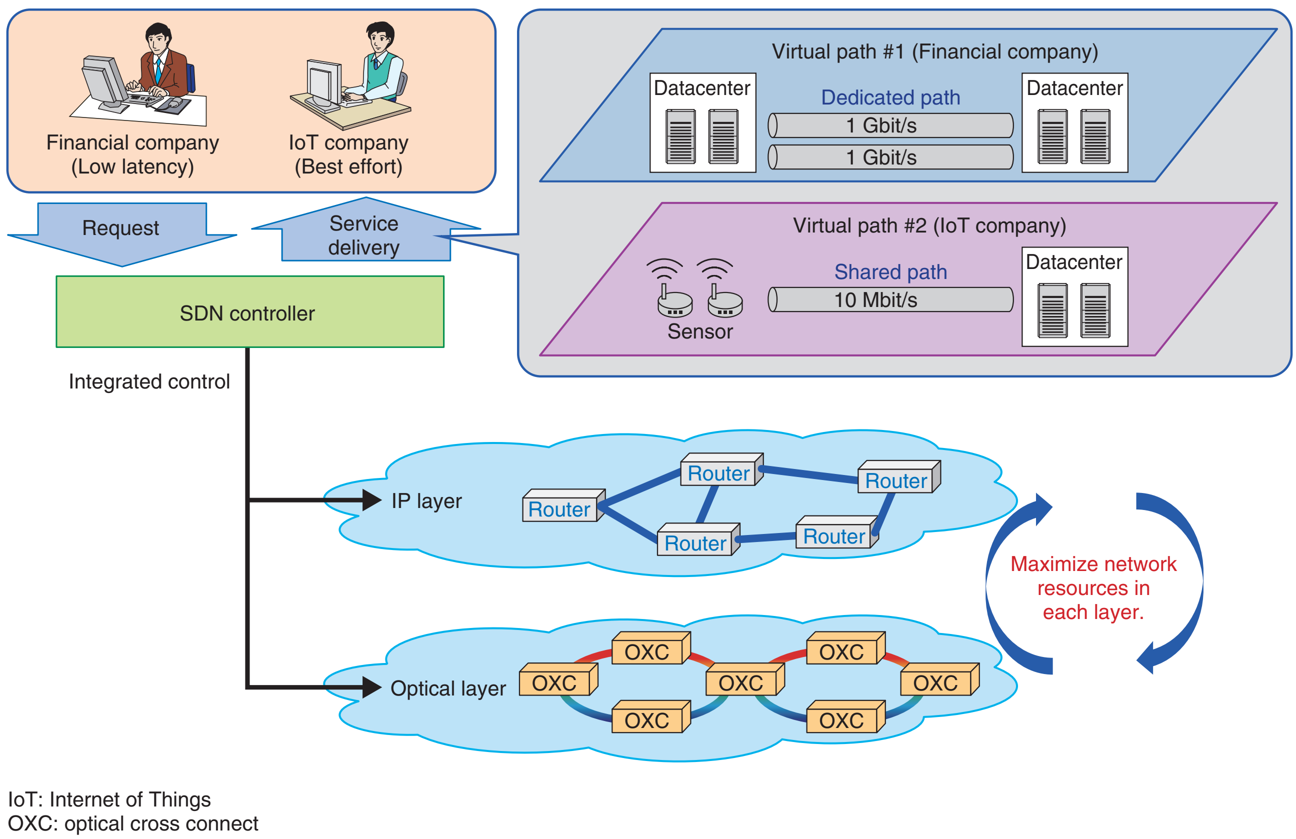


Fig. 1. Integrated control of IP and optical layers.

L3/L2 (layer 3 and layer 2) VPNs on routers, (2) functions for calculating IP routes and wavelength routes, (3) a function for configuring protection and restoration to provide virtual path redundancy, (4) a function for configuring devices, and (5) a device configuration protocol. In addition, on-demand virtual path provision does not simply provide circuits such as IP-VPNs or Ethernet leased lines—it also enables selection of multiple grades of redundancy to deal with faults. In this regard, the conventional approach to configuring redundancy was to adopt a single system for each network, as shown in **Fig. 3**.

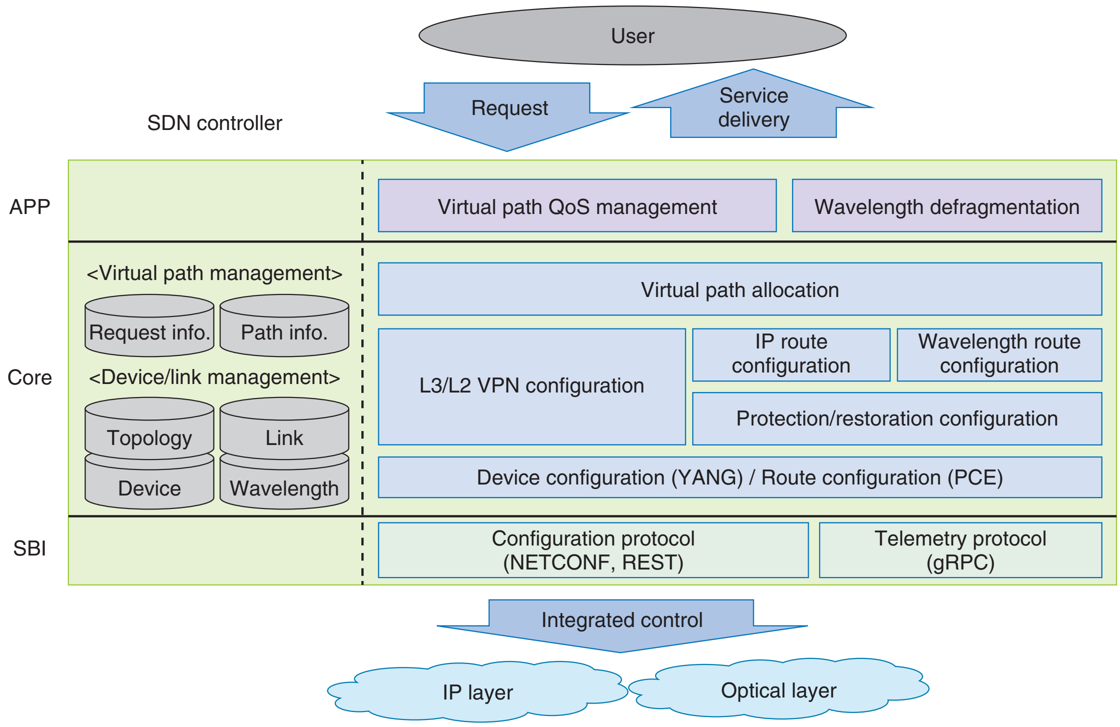
However, network design using multi-layer control technology makes it possible for these two types of redundancy systems—IP protection*1 and optical restoration*2—to coexist on a single physical network and for any redundant configuration to be selected for every virtual path. IP protection features path duplication on both the IP layer and optical layer and achieves high-speed switching, which means high reliability but at a high cost. Optical restoration, meanwhile, provides path duplication on only the optical layer and has relatively slow switching, but it can provide redundancy at a lower cost compared with IP protection. By taking into consideration how redundancy affects cost in this way, the user may select a virtual path with the most appropriate redundant configuration according to service requirements and cost factors.

2.2 Real-time monitoring and control of network conditions

Real-time monitoring and control of network conditions is achieved by (1) functions for virtual-path quality management and routing control, (2) a function for wavelength defragmentation, and (3) a telemetry

*1 IP protection: Technology that prepares links and equipment for an active system and backup system beforehand and that switches immediately to the backup system in the event of a fault in the active system.

*2 Optical restoration: Technology for restoring a system if a fault occurs in links or equipment on the optical layer by calculating a route for detouring around the fault location and changing the transmission routes of optical wavelengths.



APP: application
gRPC: An open source remote procedure call (RPC) system.
L3/L2: layer 3 and layer 2
NETCONF: Network Configuration Protocol
PCE: Path Computation Element
QoS: quality of service
REST: Representational State Transfer
SBI: southbound interface
YANG: Yet Another Next Generation

Fig. 2. Network management functions of SDN controller.

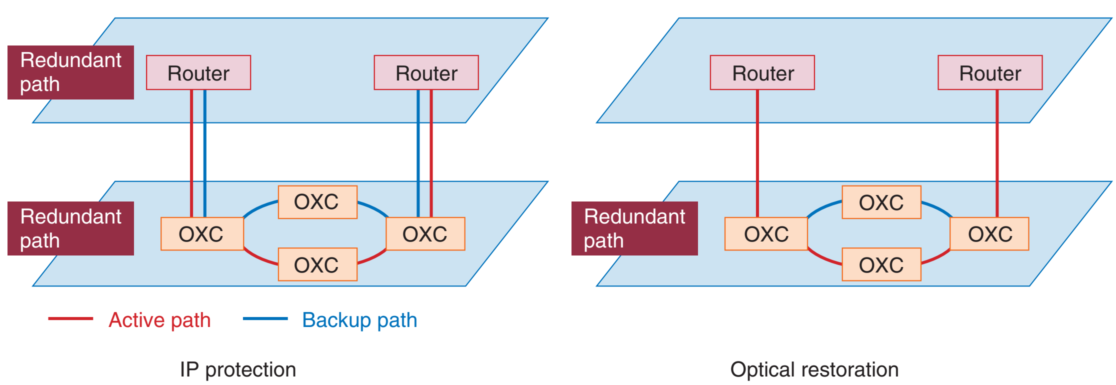


Fig. 3. Difference between IP protection and optical restoration.

protocol. The SDN controller uses a streaming telemetry*3 protocol (gRPC etc.) to receive data from devices on the IP layer and the optical layer and to therefore determine real-time network conditions such as traffic levels, delay times, packet loss ratio, and optical wavelengths in use.

The functions for virtual-path quality management and routing control detect deterioration in network quality (congestion, increase in delay time, etc.) from

*3 Streaming telemetry: A mechanism for continuously obtaining information from equipment for monitoring or other purposes.

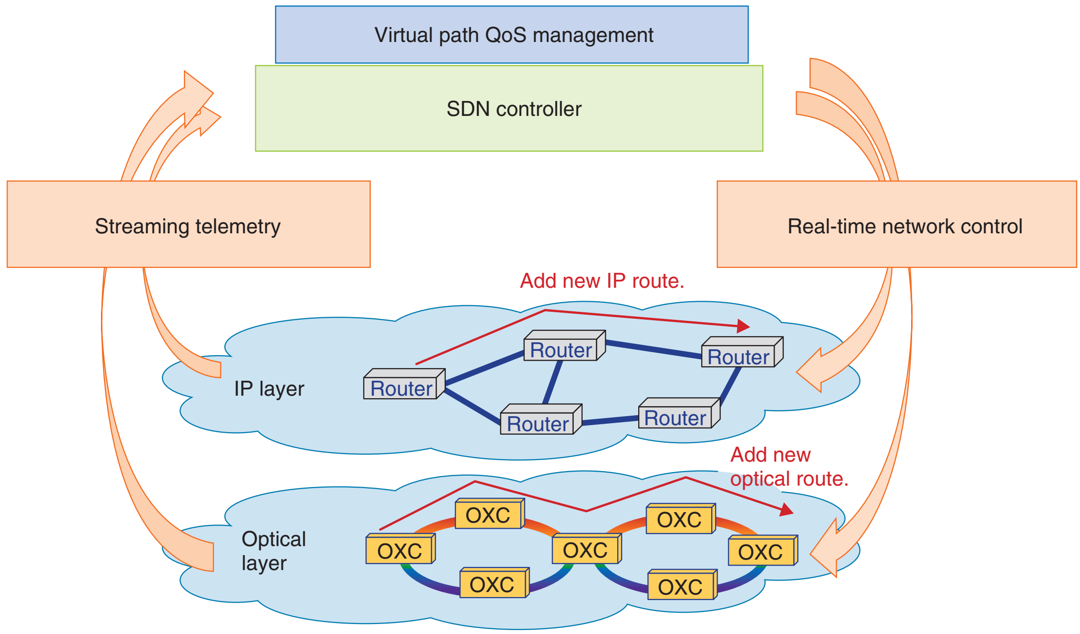


Fig. 4. Real-time network control with streaming telemetry mechanism.

the data obtained and autonomously perform virtual-path routing changes without operator assistance to restore quality (**Fig. 4**). In a network that provides paths on demand, it is assumed that the state of the network can suddenly change, but incorporating such autonomous control is expected to enable flexible response to changes that cannot be handled by manual operations.

Next, the wavelength defragmentation function detects the fragmentation state of optical wavelengths from data on the wavelengths currently in use and changes the wavelength bands in use to other wavelength bands to perform wavelength rearrangement (defragmentation). Changing the optical wavelength bands momentarily disrupts communications, so multi-layer SDN control can be used to set the destination optical wavelength beforehand on the optical layer and perform high-speed switching on the IP layer to minimize this break in communications.

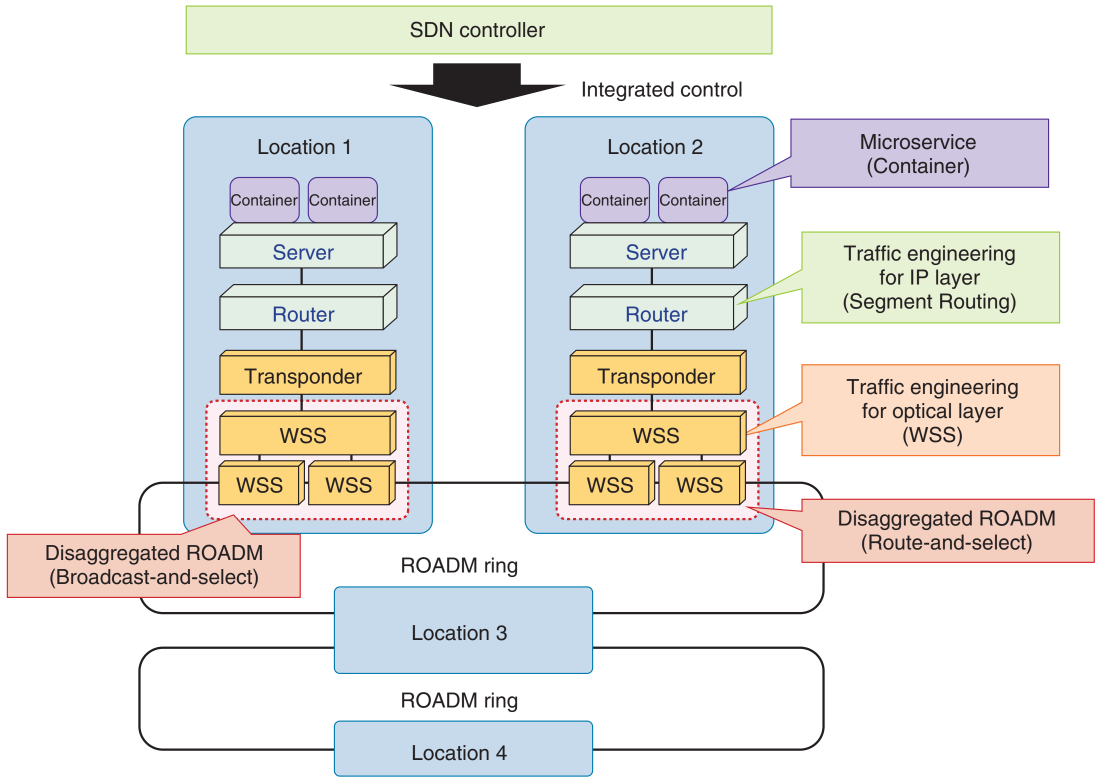
3. Technical verification

We performed a technical verification test to assess the feasibility of multi-layer SDN control. We constructed a network with the configuration shown in **Fig. 5**, and in the test, we allocated virtual paths on demand and monitored and controlled network conditions in real time from the SDN controller. We performed routing control on the IP layer using Segment Routing and achieved IP-VPNs by applying MPLS (Multiprotocol Label Switching) above this routing. We performed routing control on the optical layer by controlling transponders and optical switches separately from the SDN controller and provided Ethernet leased lines by opening up optical wavelengths between target locations. The SDN controller was equipped with a management function for a disaggregation-type reconfigurable optical add/drop multiplexer (ROADM) to combine and operate multiple optical switches as a ROADM, and for this test, it was also equipped with management functions for broadcast-and-select*4 and route-and-select*5 types of ROADM architecture [2].

Redundancy was achieved through IP protection and optical restoration, in the former by performing high-speed switching at the time of a fault using loop-free alternate Segment Routing, and in the latter by controlling optical switches from the SDN controller to make an optical-wavelength detour on detecting a

*4 Broadcast-and-select: A type of routing that broadcasts optical signals to all routes from the transmitting side using a splitter and selects which wavelengths to pass on the receiving side using an optical switch.

*5 Route-and-select: A type of routing that selects a route on the transmitting side using an optical switch and selects which wavelengths to pass on the receiving side using an optical switch.



ROADM: reconfigurable optical add/drop multiplexer
WSS: wavelength selective switch

Fig. 5. Experimental network.

fault on an optical link. With these mechanisms, we confirmed that IP-VPNs or Ethernet leased lines could be provided between target locations from the SDN controller and that either IP protection or optical restoration could be selected to achieve redundancy.

Next, in real-time monitoring and control of the network, we confirmed that traffic levels, delay times, and packet loss ratios could be obtained in real time using the gRPC streaming-telemetry protocol, and that the SDN controller could autonomously change the routing of virtual paths and restore quality when it detected quality deterioration below a certain level. Finally, in wavelength defragmentation, we confirmed that the optical waveform band in use could be changed while minimizing disruption to communications by establishing the destination wavelength beforehand based on the *make-before-break* idea and switching routing to a new wavelength path by Segment Routing.

We also performed a test focused on orchestration. We constructed an environment that provided services by container location 1 and location 2 and implemented a scale-out mechanism for increasing the number of containers at either location to keep up with an increase in the number of service users as a container orchestrator. Additionally, in the same environment, in the event that the number of users increased to a point that the resources needed for providing a service by container at location 1 became insufficient, we also achieved a scale-out mechanism spanning two datacenters by having the SDN controller coordinate with an orchestrator to establish a new virtual path at location 2 so as to provide the service by container at location 2 as well. This mechanism achieves rapid service scale out across datacenters using light containers in response to a sudden increase in demand for a service and establishes new network paths needed by a datacenter from the optical layer to the IP layer on demand. These capabilities make it possible to provide network services that are economical and robust to change.

4. Conclusion

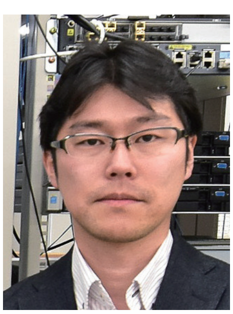
We plan to use the results of the technical verification test that we conducted on multi-layer SDN control technology to continue our research and development efforts in establishing essential infrastructure technologies for the IP and optical layers to support networks in the 5G era.

References

[1] P. Berde, M. Gerola, J. Hart, Y. Higuchi, M. Kobayashi, T. Koide, B. Lantz, B. O'Connor, P. Radoslavov, W. Snow, and G. M. Parulkar, "ONOS: Towards an Open, Distributed SDN OS," Proc. of the ACM SIGCOMM 2014 Workshop on Hot Topics in Software Defined Networking (HotSDN 2014), pp. 1–6, Chicago, USA, Aug. 2014.

[2] B. Collings, "New Devices Enabling Software-defined Optical Networks," IEEE Communications Magazine, Vol. 51, No. 3, pp. 66–71, 2013.

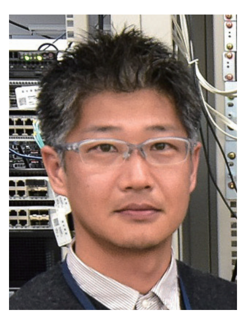
Trademark notes



Takuya Tojo

Senior Research Engineer, Network Architecture Design and Promotion Project, NTT Network Technology Laboratories.

He received a B.E. and M.S. in computers and systems engineering in 2002 and 2004, and a Ph.D. in advanced science and technology in 2007 from Tokyo Denki University. Since joining NTT in 2004, he has worked on a QoS control mechanism of the Next Generation Network (NGN) and future network architecture. His current research includes multi-layer SDN control technology.



Yoshiyuki Hirata

Research Engineer, Network Architecture Design and Promotion Project, NTT Network Technology Laboratories.

He joined NTT in 1998, and has been involved in network operation management. He is currently focusing on multi-layer SDN control for IP and optical layers for automation.

Shingo Okada

Research Engineer, Network Architecture Design and Promotion Project, NTT Network Technology Laboratories.

He joined NTT in 2008 and has been working on solving IPv6 deployment issues. His current research interests include network virtualization technologies such as SDN and NFV (network functions virtualization), especially multi-layer SDN control for the IP layer and optical layer for automation.

Seisho Yasukawa

Group Leader, Senior Research Engineer, NTT Network Technology Laboratories.

He received a B.E. and M.E. in applied physics from the University of Tokyo in 1993 and 1995. Since joining NTT in 1995, he has conducted research and development of asynchronous transfer mode based multimedia switching and NGN architecture. He is also engaged in standardization activities of P2MP (Point-to-Multipoint)-MPLS. His current research involves 5G transport technology.

Standardization by IETF and Discussion in Open Communities on Network Virtualization and Unified Method of Configuration

Hitoshi Irino

Abstract

Network virtualization technologies have progressed with the emergence of software-defined networking, and the standardization of technologies for virtualized networks on autonomous Internet protocol networks is continuing. In addition, progress is being made in the standardization of standard protocols and data models for network devices that enable virtualized networks to be built using overlay networks. This article introduces the progress being made in standardization of network virtualization technologies by the IETF (Internet Engineering Task Force) and presents related discussions underway in open communities.

Keywords: network virtualizations, IETF, open community, SDN

1. Standardization of network virtualization technologies

The implementation of multiple virtual networks on a single network infrastructure using overlay network technologies is becoming more common thanks to software-defined networking (SDN) and network functions virtualization (NFV). This article introduces Virtual eXtensible Local Area Network (VXLAN), Ethernet virtual private network (EVPN), and Segment Routing, which are network virtualization technologies, as well as Network Configuration Protocol (NETCONF) and RESTCONF, which are network management protocols, and the OpenConfig model, a vendor-neutral data model for network management. NTT's activities related to overlay networks are also explained.

1.1 VXLAN

VXLAN is a typical overlay networking technology. It was published as RFC* 7348 by the Internet Engineering Task Force (IETF). VNI (Virtual Network Identifier), a 24-bit field in a VXLAN header, can distinguish around 16 million virtualized networks. VXLAN has already been implemented in various network products. Moreover, Generic Network Virtualization Encapsulation (Geneve) [1] is being discussed for standardization as a new standard overlay protocol in the NVO3 (Network Virtualization Overlay) Working Group (WG) of IETF. A Geneve header can contain not only a 24-bit field but also optional fields (**Fig. 1**).

Internet protocol (IP) multicast is used for transporting BUM (broadcast, unknown unicast, and multicast) frames on VXLAN. However, this method requires an IP multicast network infrastructure such as Address Resolution Protocol (ARP) generated on unicast communication in order to handle broadcast frames. Hence, IP multicast protocols such as PIM (Protocol Independent Multicast) are needed on transport devices.

* RFC: Request for Comments. Documents published by IETF, which include protocols, procedures, programs, and concepts.

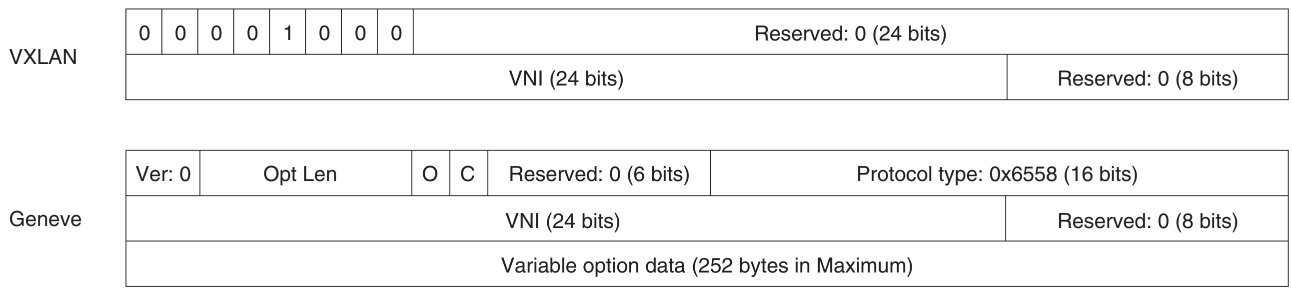


Fig. 1. Comparison of VXLAN header and Geneve header.

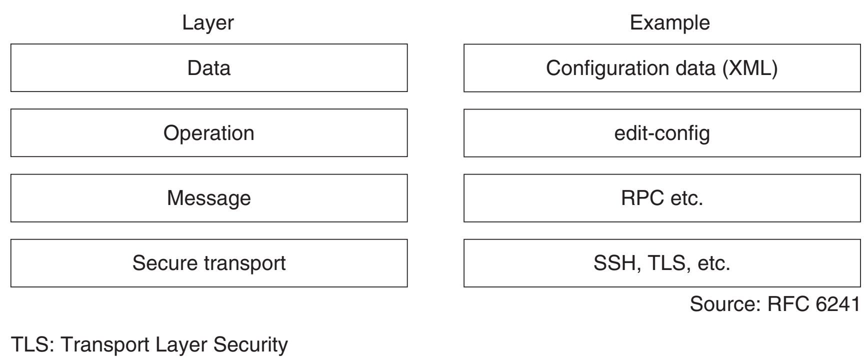


Fig. 2. NETCONF Protocol Layers.

This issue can be resolved by using the Open vSwitch Database (OVSDB) management protocol, which is published as RFC 7047 and developed by VMWare for their SDN solutions. An SDN controller can communicate multiple pairs of IP addresses and MAC (media access control) addresses to software switches (Open vSwitch) and hardware switches that implement OVSDB database schemes by using the OVSDB management protocol instead of using ARP.

1.2 EVPN

In IP networks, autonomous IP routing technologies provide fault tolerance. EVPN is an extension of BGP (Border Gateway Protocol), which brings information corresponding to ARP. EVPN has been published as RFC 7432 for the Multi-Protocol Label Switching (MPLS) data plane and RFC 8365 for VXLAN.

A draft [2] related to EVPN has been proposed that not only enables the sending of information corresponding to ARP but also enables layer 3 virtualizations on VXLAN protocols originally designed for layer 2 virtualization. This proposed technology makes it possible to provide a unified method for layer 2 and layer 3 overlay networks on an IP underlay network. However, concentrative configuration technologies are needed in order to deploy virtualized network technologies in large-scale environments.

1.3 NETCONF

NETCONF 1.1 is a standard protocol for configuring network devices. It was standardized in 2011 as RFC 6241. NETCONF makes it possible to achieve SDN. Basic operations such as get-config, edit-config, copy-config, and delete-config are defined. Their NETCONF operations are commanded using RPCs (remote procedure calls) over transport protocols such as SSH (Secure Shell). Message data in operations are encoded in Extensible Markup Language (XML) (**Fig. 2**). YANG (Yet Another Next Generation) is standardized as a data modeling language that can be mapped to XML. YANG was published as RFC 6020.

1.4 RESTCONF

Many recent web applications have been implemented

Table 1. Comparison between RESTCONF and NETCONF operations.

RESTCONF	NETCONF
OPTIONS	None
HEAD	get-config, get
GET	get-config, get
POST	edit-config (create)
POST	Invoke an RPC operaion
PUT	copy-config
PUT	edit-config (create/replace)
PATCH	edit-config
DELETE	edit-config (delete)

Source: RFC 8040

using Representational State Transfer (REST), which is based on HTTP (Hypertext Transfer Protocol). The same trend is seen in networking. RESTCONF is a protocol that applies REST to NETCONF and has been standardized as RFC 8040 (**Table 1**). RESTCONF can use XML as well as JSON (JavaScript Object Notification), which is generally used for REST to encode message data. The emergence of RESTCONF is therefore very promising, although it does not support full NETCONF functions such as a two-phase transaction.

2. Trends in standardization of data models

The standardization documents concerning NETCONF and RESTCONF described above define only protocols for communication. They do not define the content of messages. YANG, as mentioned above, is the language used for the data model (data structure) in messages. Standardization of YANG-based data models is continuing to progress. These data models can be classified into data models for service descriptions and data models for network devices (**Fig. 3**).

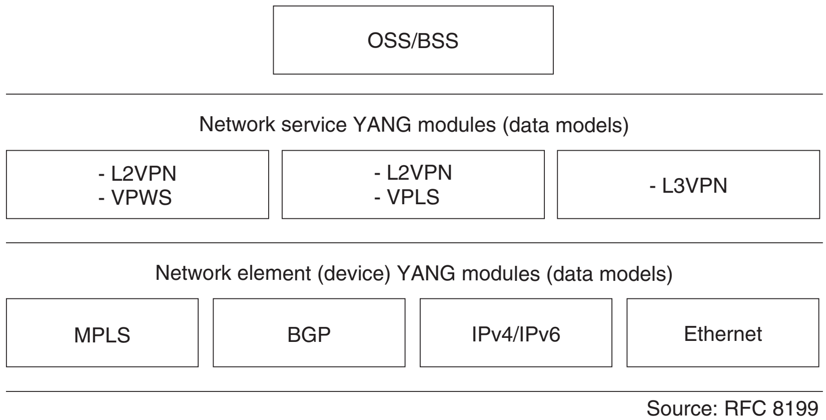
The data models for service description are used for layer 2 and 3 virtual private networks (VPNs) as examples. Layer 3 VPN service data models are already standardized as RFC 8299. Network device data models are used for configuring network interfaces and routing protocols. These data models for configuring network devices have been standardized by IETF and also defined by OpenConfig [3], an organization of service providers such as Google, Microsoft, Facebook, AT&T, BT, and Comcast. Some OpenConfig data models are already implemented in the devices of well-known router vendors. Hence, OpenConfig has the potential to become a common implementation led by service providers for network environments employing the products of multiple vendors. Furthermore, OpenConfig defines not only data models for IP transport devices but also data models for optical devices.

3. Trends in new network virtualization technologies

To date, virtualized networks have been achieved by using MPLS or IP tunneling such as VXLAN. The concept of Segment Routing [4] discussed in the SPRING (Source Packet Routing in Networking) WG in IETF may change the situation. The data plane technology expected to be used for Segment Routing is MPLS (SR-MPLS) or IPv6 (SRv6). Segment Routing is a source routing technology. It can control the forwarding path from the source node by using an embedded node list in a packet header. Therefore, Segment Routing is one of the technologies expected to be implemented to achieve service function chaining. It also enables fast rerouting achieved by TI-LFA (Topology Independent Loop Free Alternative) when routes change in failure cases. Moreover, SRv6 will be able to provide layer 2 and 3 VPNs [5], which have thus far only been achieved on MPLS, on IPv6 data planes by giving new functions to part of the IPv6 address. Hence, SRv6 is expected to be a unified data plane.

4. NTT activities

NTT Network Service Systems Laboratories is developing Multi-Service Fabric (MSF) [6] and publishing it as open source software [7]. MSF is an SDN system that configures and manages multiple vendors' merchant-silicon-based network devices including white-box switches via NETCONF and other protocols. It is used for deploying layer 2 overlay networks based on VXLAN with EVPN and layer 3 overlay networks based on MPLS to a wide area. MSF was used in a collaborative proof of concept (PoC) [8, 9] with an APAC (Asia Pacific) carrier. In addition, it was introduced into a PoC environment by Dimension Data Asia Pacific [10], which is an NTT Group company. MSF uses IP routing technologies including EVPN to achieve fault tolerance and to prevent high loads from concentrating at a particular node. NTT has joined open communities that carriers and OTT (over the top) players lead collaboratively such as ONF (Open Networking Foundation) and TIP



L2VPN: layer 2 virtual private network
OSS/BSS: operations support system/business support system
VPLS: virtual private LAN (local area network) service
VPWS: virtual private wire service

Fig. 3. Classification of data models.

(Telecom Infra Project) to discuss fault tolerant networks achieved by autonomous mechanisms.

References

[1] Geneve, https://datatracker.ietf.org/doc/draft-ietf-nvo3-geneve/
[2] IP Prefix Advertisement in EVPN, https://datatracker.ietf.org/doc/draft-ietf-bess-evpn-prefix-advertisement/
[3] OpenConfig, http://www.openconfig.net/
[4] Segment Routing, https://datatracker.ietf.org/doc/draft-ietf-spring-segment-routing/
[5] SRv6, https://datatracker.ietf.org/doc/draft-filsfils-spring-srv6-network-programming/
[6] 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
[7] Multi-Service Fabric (MSF), https://github.com/multi-service-fabric/msf
[8] NTT press release issued on December 12, 2017. http://www.ntt.co.jp/news2017/1712e/171212a.html
[9] T. Kuwahara, H. Irino, and K. Suzuki, "Global Collaboration Initiatives Revolutionizing Network Research and Development," NTT Technical Review, Vol. 16, No. 6, 2018. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201806fa7.html
[10] A. Masuda, "Creating a New Ecosystem for NFV/SDN Technical and Business Development: the Challenge of NTT Laboratories and Dimension Data APAC," NTT Technical Review, Vol. 15, No. 5, 2017. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201705gls.html

Trademark notes

Hitoshi Irino

Senior Research Engineer, Transport Network Innovation Project, NTT Network Service Systems Laboratories*.

He received a B.A. in environmental information and an M.A. in media and governance (I.T.) from Keio University, Kanagawa, in 2003 and 2005. He joined NTT Network Service Systems Laboratories in 2005. He researched traffic analysis technologies using flow information such as NetFlow/IPFIX (IP Flow Information Export) and proposed drafts about IPFIX in IETF as part of standardization efforts. He has been one of the original architects of NTT's SDN technology for Multi-Service Fabric (MSF) since July 2014. He also initiated a project to construct an original networking OS called Beluganos using open source software and OpenFlow on merchant-silicon-based switches. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE).

*He moved to NTT Communications in July 2018.

Case Study of Problem in Multifunction Telephones Connected to a Business Phone System

Technical Assistance and Support Center, NTT EAST

Abstract

This article describes a non-activation problem that was occurring in customer's multifunction telephones connected to a business telephone system and the investigation that was carried out to determine the cause. This is the forty-seventh article in a series on telecommunication technologies.

Keywords: business phone, multifunction telephone, characteristic impedance

1. Introduction

When a customer added multifunction telephones to NTT's αNXII business phone system, the added terminals failed to activate (called a non-activation problem). There were no past examples of this problem, and since it was possible that it would occur again in the future, we investigated this problem and conducted tests to determine the cause of the fault. In this article, we describe the investigation of the fault, the results of analysis, and countermeasures.

2. Fault overview

The equipment configuration is illustrated in **Fig. 1**. As shown, multifunction telephones are connected to NTT's αNXII business phone system, with 23 phones connected in range A, 2 in range B, and 3 in range C. A star wiring scheme is used to connect all of these multifunction telephones using existing cables manufactured by another company.

The 23 multifunction telephones in range A and the 3 in range C were all determined to be operating normally. However, it was found that the 2 multifunction telephones in range B were displaying the message "One moment please" and were failing to activate.

On-site maintenance personnel replaced the other company's cables that were connecting the non-activated multifunction telephones with NTT cables. As a result, the two terminals were activated, and the fault was resolved.

3. Investigation of cause

Although the telephones had been activated, we carried out an investigation to determine what caused the fault and prevent future occurrences.

3.1 Investigation 1 (evaluation of cable characteristics)

Here, we describe the system components that were investigated and explain the results.

3.1.1 Measurement items

Given that the fault was resolved by on-site maintenance, we focused our attention on the two types of cables used and evaluated their characteristics. The components that were measured are listed in **Table 1** and described below.

Evaluated cables:

A: Other company's cable (110 m)

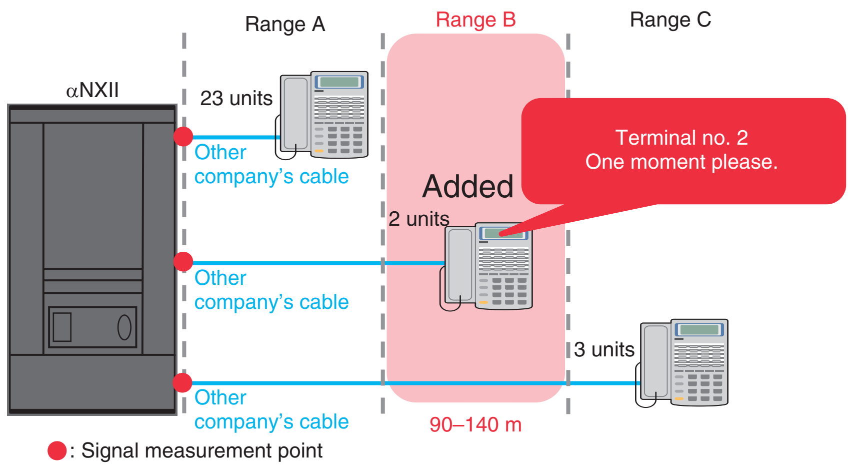


Fig. 1. Equipment configuration (customer's premises).

Table 1. Measurement items.

No major difference

Measurement target	Measurement item	Measuring instrument
Cable only	(1) Cable characteristics	Cable characteristics tester
	(2) Cable loss	Network analyzer
	(3) Crosstalk characteristics	
	(4) Characteristic impedance	Impedance analyzer
Cable + multifunction telephone	(5) Reflection	Network analyzer

B: NTT cable (110 m)

The measurements were conducted using the business phone's signal frequency of 250 kHz.

(1) Cable characteristics (cable only)

(2) Cable loss (cable only)

(3) Crosstalk characteristics (cable only)

(4) Characteristic impedance*1 (cable only)

(5) Reflection*2 (cable + multifunction telephone)

3.1.2 Investigation results

No major difference was found in these two types of cables for items (1) to (3) above.

The characteristic impedance (item (4)) of the cable measured at the business phone's signal frequency of 250 kHz was approximately 85 Ω for cable A, the other company's cable, and 110 Ω for cable B, NTT cable, a difference of about 25 Ω. These measurement results are listed in **Table 2**.

The results of measuring the amount of reflection (item (5)) when connecting a multifunction telephone at 250 kHz were –8.0 dB for cable A and –11.5 dB for cable B, a difference of 3.5 dB. The measurement results are shown in **Fig. 2**.

3.1.3 Discussion

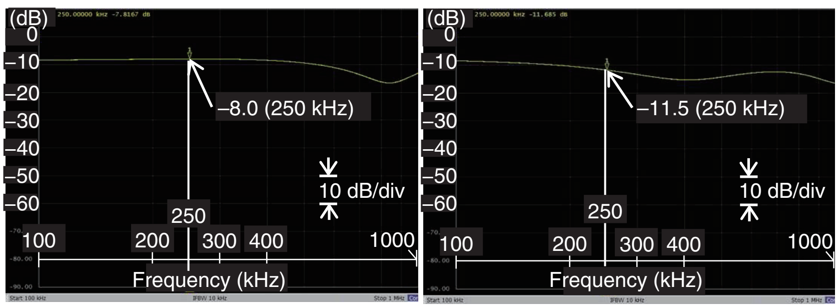
The impedance of the NTT cable was 110 Ω, which satisfies the business-phone operating conditions of 110±20 Ω, but the impedance of the other company's cable was 85 Ω, which does not satisfy the conditions. Furthermore, on examining the reflection when connecting a multifunction telephone, we found that the reflection of the other company's cable was found to be 3.5 dB greater than that of the NTT cable, so it could be considered that this reflection was affecting the signal.

*1 Characteristic impedance: Impedance unique to a transmission path.

*2 Reflection: Ratio of reflected wave to incident wave expressed in dB.

Table 2. Characteristic impedance measurement results.

Frequency (kHz)	Characteristic impedance (Ω)	
	A: Other company's cable	B: NTT cable
1	94.5	114.7
10	94.6	115.5
100	90.9	114.3
250	85.1	110.7



A: Other company's cable B: NTT cable

Fig. 2. Measurement results for reflection.

3.2 Investigation 2 (effect of reflected wave)

3.2.1 Measurement items

This problem occurred in multifunction telephones that were connected with another company's cable and when the telephones were installed within a certain distance from the business phone system. We therefore focused our testing on how the signal waveform differed between the normal-operation time and the non-activated time. We set up equipment at the Technical Assistance and Support Center using the same configuration as that in Fig. 1 and measured the various components at the measurement points indicated by "●" using an oscilloscope.

3.2.2 Investigation results

The results of measuring signal waveforms (for ranges A, B, and C) by oscilloscope are shown in **Fig. 3**. When we examine the portion within the red circle for each range, we can see that the voltage of the waveform was about 0 V for normal operation in ranges A and C and greater than 1 V at the non-activated time in range B.

3.2.3 Discussion

The αNXII business phone uses the Alternate Mark Inversion (AMI) coding scheme. If the voltage at the end of the signal waveform within the red circle is greater than 1 V, the signal cannot be read as normal data (**Fig. 4**). The above results show that this voltage was about 0 V in ranges A and C, indicating normal operation. In range B, however, a voltage of 1.28 V (reflected wave) occurred due to the reflected-wave effect, with the result that the terminal could not recognize the signal normally and could not be activated.

4. Inference of cause of fault

The test results presented above indicate that the characteristic impedance of the other company's cable used for equipment wiring was 85 Ω, which does not satisfy the characteristic impedance of the business phone system used here (110±20 Ω). Consequently, when a multifunction telephone was connected to this type of cable, there was large signal reflection, which affected the signal within a certain range of cable length.

We inferred from these results that the fault was caused by a breakdown in the signal transmitted by the multifunction telephone. The breakdown was due to the impedance mismatch in the other company's cable, which prevented the main equipment from correctly reading the signal. The fault generation mechanism is depicted in **Fig. 5**.

In range A, voltage greater than 1 V at the end of the

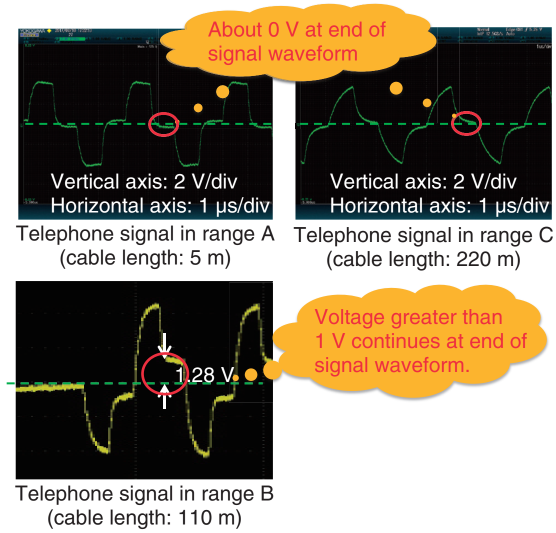


Fig. 3. Results of measuring signal waveforms with oscilloscope.

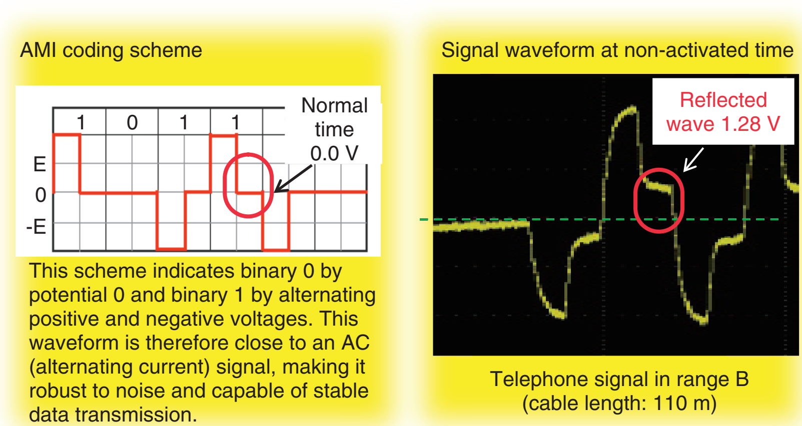


Fig. 4. AMI coding scheme.

signal waveform does not continue, and although a reflected wave exists, the reflection duration is short, so there is little impact on the signal.

Likewise, in range C, voltage greater than 1 V at the end of the signal waveform does not continue. This can be explained by considering that while a reflected wave exists, the cable length is long, which has the effect of attenuating the signal itself as well as the reflected wave. The effect of the reflected wave on the signal is consequently small.

5. Countermeasures

Two possible countermeasures were considered:

a) Replace with an NTT cable

Effect: Achieves impedance matching between the terminal and cable, suppresses reflected waves, and activates the terminal normally.

b) Insert a countermeasure device (attenuator)

Effect: Attenuates the reflected wave and activates the terminal normally.

Although countermeasure a) was implemented effectively in the initial investigation, as described in

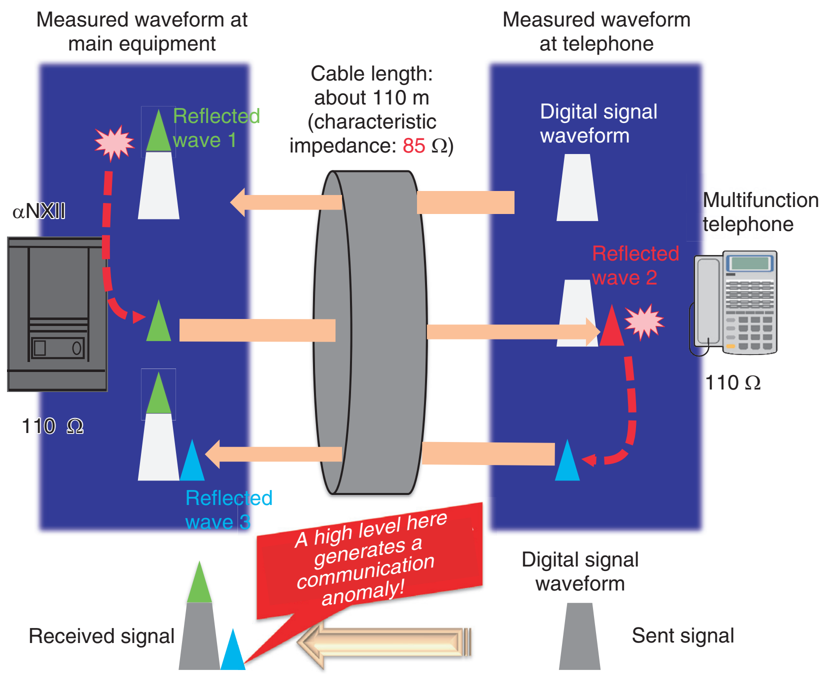


Fig. 5. Fault generation mechanism.

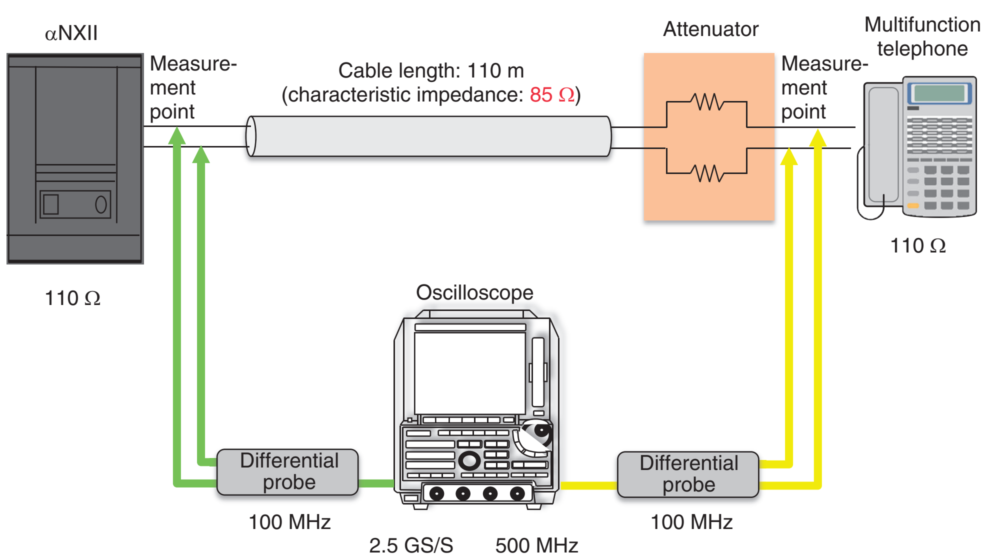
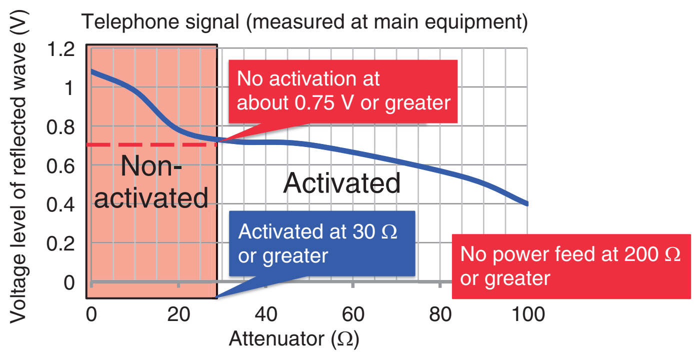


Fig. 6. Test configuration.

section 2, we considered that it may not always be possible to use this countermeasure and thus implemented countermeasure b). We therefore prepared and tested an attenuator at the Technical Assistance and Support Center.

5.1 Test method

We inserted an attenuator (resistance) in a 110-m length of the other company's cable and checked the effect of reducing the voltage level in the reflected wave (**Fig. 6**).



Resistance (Ω)	Reflected voltage level at main equipment (V)	
0	1.08	Non-activated
10	0.98	
20	0.78	
33	0.72	Activated
51	0.70	
85	0.54	
100	0.40	

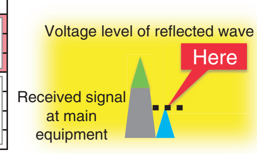


Fig. 7. Checking effect of attenuator.

5.2 Test results

The test results are shown in **Fig. 7**. The following conclusions were made based on these results.

- Inserting a resistance of 30 Ω or greater in the attenuator results in activation.
- No activation occurs if the voltage level of the reflected wave is approximately 0.75 V or greater.
- Inserting a resistance of 200 Ω or greater disables power feeding.

We found based on the above results that setting resistance in the range of 30–100 Ω would work well as a countermeasure to this fault.

6. Conclusion

In this article, we introduced a problem in multifunction telephones connected to NTT's αNXII business phone system handled by the Technical Assistance and Support Center. As shown by this case study, it is important that the cable used between business phone equipment (main equipment and terminal) be impedance-matched with the business phone system. The EMC Engineering Group of the Technical Assistance and Support Center aims to achieve prompt resolution of noise faults related to conduction and radiation and to contribute to the smooth provision of communication services. To this end, it is proactively engaged in technology dissemination activities through technical collaboration, technology development, and technical seminars.

External Awards

ITU-AJ Encouragement Award

Winner: Kei Harada, NTT Network Innovation Laboratories
Date: May 17, 2018
Organization: The ITU Association of Japan (ITU-AJ)

For her more than one year of involvement in oneM2M standardization efforts and related dissemination activities.

Best Magazine Paper Award

Winner: Nahoko Kasai, Takayuki Ogasawara, Hiroshi Nakashima, and Shingo Tsukada, NTT Basic Research Laboratories
Date: May 18, 2018
Organization: The Institute of Electronics, Information and Communication Engineers (IEICE)

For "Development of Functional Textile "hitoe": Wearable Electrodes for Monitoring Human Vital Signals."

A journal article award was given by IEICE to the authors of this paper describing the development history of the functional material "hitoe" and its application examples such as sports monitoring, heart rate/posture/breathing estimation, and worker safety management.

Published as: N. Kasai, T. Ogasawara, H. Nakashima, and S. Tsukada, "Development of Functional Textile "hitoe": Wearable Electrodes for Monitoring Human Vital Signals," IEICE Communications Society Magazine, Vol. 41, pp. 17–23, 2017 (in Japanese).

CHEMINAS Technology Award

Winner: Suzuyo Inoue, Yuzuru Iwasaki, Katsuyoshi Hayashi, and Michiko Seyama, NTT Device Technology Laboratories; Tsutomu Horiuchi, Institute of Technologists
Date: May 21, 2018
Organization: Society for Chemistry and Micro-Nano Systems (CHEMINAS)

For development of a portable surface plasmon resonance sensor and its combination with a microfluidic device for a new biosensing technology.

Contribution Award for Standardization

Winner: Toshimori Honjo, NTT Basic Research Laboratories
Date: May 28, 2018
Organization: Information Technology Standards Commission of Japan (ITSCJ)

For his contribution to standardization activities in ITSCJ, an organization of the Japanese Industrial Standards Committee that is participating in standardization activities corresponding to ISO/IEC (International Organization for Standardization/International Electrotechnical Commission) Joint Technical Committee 1.

ISCS Quantum Devices Award

Winner: Hiroshi Yamaguchi, NTT Basic Research Laboratories
Date: May 29, 2018
Organization: The International Symposium on Compound Semiconductors (ISCS)

For his leading contributions to the development of compound semiconductor opto/electromechanical systems.

The 42nd Laser Society of Japan Encouragement Award

Winner: Hiroki Mashiko, NTT Basic Research Laboratories
Date: May 31, 2018
Organization: The Laser Society of Japan

For research on petahertz electron manipulation with wide-bandgap semiconductors.

We successfully observed electron oscillation with petahertz frequency using semiconductor and insulator solid-state materials characterized by attosecond light source.

Best Paper Award

Winner: Masahiro Kohjima and Tatsushi Matsubayashi, NTT Service Evolution Laboratories; Hiroshi Sawada, NTT Communication Science Laboratories
Date: June 7, 2018
Organization: IEICE

For "Probabilistic Models Based on Non-negative Matrix Factorization for Inconsistent Resolution Dataset Analysis."

Published as: M. Kohjima, T. Matsubayashi, and H. Sawada, "Probabilistic Models Based on Non-negative Matrix Factorization for Inconsistent Resolution Dataset Analysis," IEICE Trans. Inf. & Syst. (Japanese Edition), Vol. J100-D, No. 4, pp. 520–529.

Achievement Award

Winner: Hiroyuki Oto and Yasuyuki Uchiyama, NTT DOCOMO; Kazuaki Obana, NTT Network Innovation Laboratories
Date: June 7, 2018
Organization: IEICE

For commercial deployment of network functions virtualization technology enabling multi-vendor EPC (evolved packet core) software.

Achievement Award

Winner: Tadao Ishibashi, NTT Electronics Techno Corporation; Fumito Nakajima, NTT Device Technology Laboratories; Yoshifumi Muramoto, NTT Device Innovation Center
Date: June 7, 2018
Organization: IEICE

For their pioneering work on ultrahigh-speed, high-output-power, and high-sensitivity photodiodes.

Papers Published in Technical Journals and Conference Proceedings

Local Linear Predictive Coding for High Resolution Time-frequency Analysis

F. Ishiyama

Proc. of 2017 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT 2017), pp. 1–6, Bilbao, Spain, December 2017.

We are developing a method of time-frequency analysis on the basis of linear predictive coding (LPC). In contrast to standard LPC, which assumes an infinite number of samples, our method obtains instantaneous frequencies, instantaneous amplitude vary rates, and instantaneous amplitudes from a small number of samples. The time width for single local analysis can be much narrower than the period of oscillation of a given time series, enabling high resolution time-frequency analysis. We outline our method and use it to analyze electro-magnetic noise in comparison with short time Fourier transform to show its resolution.

Electrical and Kinesthetic Stimulation for Virtual Walking Sensation

H. Kaneko, R. Koide, Y. Ikei, T. Amemiya, K. Hirota, and M. Kitazaki

IEEE Haptics Symposium 2018, San Francisco, USA, March 2018.

The effect of electrical stimulation applied to the lower limb tendons on the sensation of virtual walking was investigated. The virtual walking, here, is a passive experience in which the user sees and acts virtually as some other person did to learn through his/her experience. A display device was developed to provide electrical stimulation to sensory nerves in the Achilles tendon and the tibialis anterior tendon. A kinesthetic stimulus and visual stimulus were simultaneously provided to a seated participant. We measured the sensations of walking and translational motion regarding three factors of electrical, kinesthetic and visual stimuli. The result obtained from eleven participants revealed that the three factors were significant in enhancing the sensation of virtual walking. The electrical stimulation for proprioception seemed effective to compensate the characteristics of passive kinesthetic playback of experiences.

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

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

arXiv: 1805.03385 [quant-ph], May 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.
