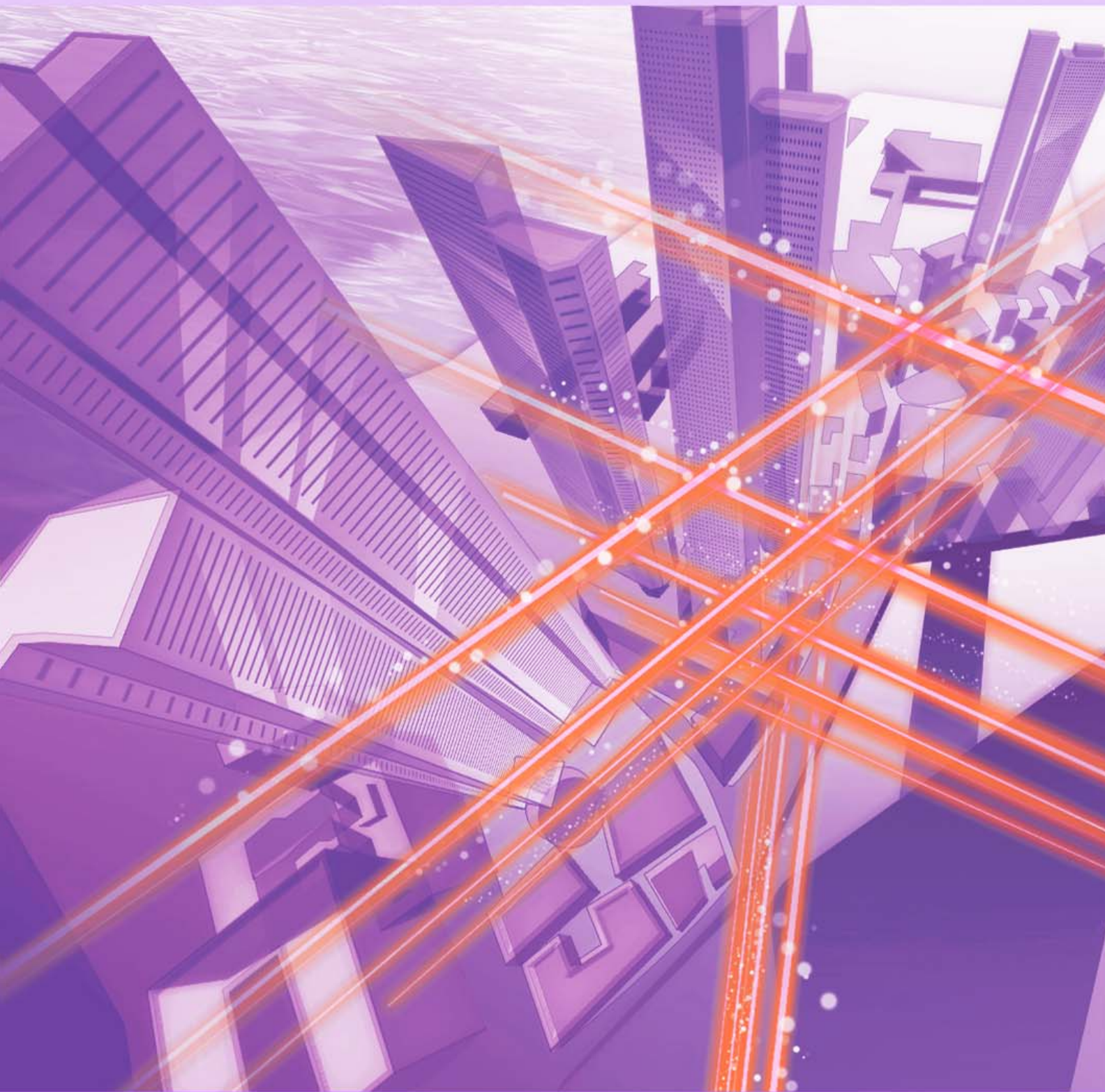
NTT Technical Review

8
2017



August 2017 Vol. 15 No. 8

NTT Technical Review

August 2017 Vol. 15 No. 8

A Scientist who Studies Human Beings from an Artistic and Philosophical Perspective

Junji Watanabe
Distinguished Researcher, NTT Communication Science Laboratories

Overview

The ability to convey human sensory information via vibrations on a smartphone—a device now used by about 65% of the Japanese population—would no doubt lead to a whole new style of communication. Distinguished Researcher Junji Watanabe of NTT Communication Science Laboratories is researching the relationship between the human senses and the environment. We asked him about the essence of his current research and his approach as a researcher.

Keywords: subconscious, haptic information, onomatopoeia

Standing in the middle of science, art, and philosophy

—Dr. Watanabe, please tell us about your recent research activities.

In person-to-person communication, the information exchanged by words is no more than a small part of what each person thinks about in his or her mind. When I encounter another person and engage in a conversation, I give more importance to things that are not expressed in words or things that I am not conscious of, even though I may have subconsciously felt them. This is what I believe to be the meaning of directly encountering and talking with another person, and I am interested in the moment that such things come out.

This kind of relationship between the conscious and subconscious also holds for things that move one's body. In everyday life, when we consciously set out to move our own body, there are actually more complex movements taking place on a subconscious level. Even for movements that we may think to be simple, there are many subconscious movements in the background. With this in mind, I recently broke down movements that are usually performed subconsciously into detailed movements and sensations and created a video of the results like a musical score. This video was exhibited under the title "A Dialogue with the Subconscious, the Most Familiar Other" at the exhibition "ATHLETE" held by 21_21 DESIGN SIGHT in Roppongi, Tokyo from February–June 2017 (**Fig. 1**).

Athletes are known to practice while paying detailed attention to each and every bodily movement and to then train themselves until those movements

Fig. 1. Screenshot of the work “A Dialogue with the Subconscious, the Most Familiar Other” (Junji Watanabe, 2017) showing a video of hand movements (right) and the same movements like a musical score (left).

become subconscious actions. They feel that such a disciplined approach will enable them to excel at the time of a match without having to think about it. Talking about this with the exhibition director led to this work of mine. Thinking that not just athletes but ordinary people too should have access to such a process, I decided to make conscious in an easy-to-understand form those things that people are not usually aware of. Specifically, I took up actions that require some practice in our daily lives such as typing or doing finger motions on a smartphone screen (tapping, swiping, etc.), and I analyzed them from the viewpoints of muscle movement in the bending and extending of fingers and sensations such as fingertip pressure and position. I then represented the results of this analysis like a musical score.

—It’s surprising that presenting such an exhibition would be included in a researcher’s activities; you’re acting more like an artist.

I am interested in human beings and how they are changed through technology. I think that it’s important to understand the essence of human beings from a scientific viewpoint. However, in creating works, my aim is to convey that understanding as a personal experience much as an artist would do, so when viewed from the outside, I may appear to be a mix of a scientist and an artist.

In any field, there is context and format, and to date, I have mainly been involved in writing journal papers, the main format of science. However, I have also participated in exhibitions, one of the main formats in the world of art. In that case, though, it’s not that I would exhibit a “work” within the context of art. Rather, I would focus on making experiential inquiries into the meaning of human beings and technology by questioning the meaning of various things, as in “Can’t this also be interpreted in this way?” or “Is there anything else that has this meaning?” Of course, activities such as these have some relation to art and philosophy.

“Now” is the result of trial and error—retracing your footsteps to find your objective

—What motivated you to use this approach in representing your research?

From 2005 to 2009, I served as a PRESTO (Precursory Research for Embryonic Science and Technology) researcher at the Japan Science & Technology Agency (JST), where I conducted research and exhibited that research under the theme of “Fundamental Science and Technology for New Tactile/Haptic Expressions.” During my time as a PRESTO researcher, I showcased my work every month, if not every week, and I can say that my present way of thinking grew out of a trial and error process at that time.

Something I try to be careful about when demonstrating or exhibiting my research is to refrain as much as possible from limiting attendees in the way they view that research in the sense of “It must be thought of in this way.” Of course, I give some explanations on what I am exhibiting, but regardless of how it is interpreted, that interpretation is what constitutes truth for any one attendee, so I always hope that a dialogue about interpretation occurs as a result. The type of medium used to present one’s research—journal paper or exhibit—depends on the content of that research, but the way of thinking behind that research does not change. I choose an optimal method of output for each case.

In fact, a variety of formats can be used to output the contents of one’s research, and my use of several types of formats is perhaps unique. At a design or art exhibit, I try to present my research in a way that can be experienced without an explanation, and in research demonstrations such as at the NTT R&D Forum, I tend to explain my research in a step-by-step manner. Another approach I might employ is to have recipients experience the results of my research in a sequential manner over a somewhat long period of time, as in a workshop, or I might convey my research

theme through the medium of a magazine. Furthermore, in addition to such momentary experiences, I strive for exhibitions in which recipients can take something back with them—that is, something they can keep in mind long after the exhibit is over.

—You have been involved in diverse research fields. Can you tell us something about your research history up to the present?

I am currently involved in a variety of endeavors including research on the human sensory mechanism, haptic design, and creation of wellbeing guidelines. This may seem curious from the outside, and people may ask: "Why are you involved in so many disconnected things?" But for me, these are all connected.

In the third year of my university studies, I had to choose a field to specialize in. My first choice at that time was virtual reality, my second choice was cognitive science, and my third choice was philosophy. Consequently, I joined the virtual reality laboratory and researched human perception, and it was not long before I developed an interest in the relationship between the human senses and technology.

Moreover, looking back at my childhood, I remember being more interested in principles than material phenomena. For example, I would be more interested in "Why is this the case?" or "How was this derived?" than coming up with conclusions or answers in a test. Studying by rote memorization even in fields such as physics and chemistry was not very appealing to me; I was more interested in the underlying rules. Likewise, with respect to human beings, I was not very interested in superficial speech or behavior. What interested me were the principles of sensation and movement giving rise to such human behavior and the principles of the human mind explaining why it occurred.

I believe that researching the human sensory mechanism is a primary way to approach the principles of the human mind, and that the sense of touch can act on emotions and the subconscious. One example is the "Heartbeat Picnic" workshop that enables participants to experience what it would be like to hold their own hearts in their hands. In this workshop, when you hold a stethoscope against your chest, it causes a cuboid-shaped box to vibrate in sync with your heartbeat, thereby producing a feeling through the sense of touch that you are really alive (**Fig. 2**). I have also conducted a workshop that attempts to visualize the perceptual space of tactile texture through the use of Japanese onomatopoeia (**Fig. 3**). The idea here is to draw a map of tactile sensations that are attuned by using onomatopoeia, which are words that mimic the sound of various objects, or words having a more delicate and intuitive nature rather than simple adjectives such as "rough" and "hard." In addition, I recently began research in the field of wellbeing as the ultimate motivation behind human behavior.

Fig. 2. Feeling your own heartbeat through the vibrations of a box (heartbeat box) in the palm of your hand.

Searching for the essence of all kinds of things and the relationships between them

—What, then, is the common characteristic of these diverse research fields?

I would say that my personal research theme is to examine how the senses of individual human beings create communication and how society came into existence. I also want to search for the essence of these developments and the relationships between them.

For example, onomatopoeia refers to special words that make a connection with the senses by relating sound and meaning. It's easy to relate the 's' sound in a Japanese word like "sara sara" with a smooth sensation. In contrast, the word "zara zara" vocalized by vibrating the vocal chords with the same mouth position as 'z' relates to a rough sensation. This is probably because air vibrates more vigorously at the gums

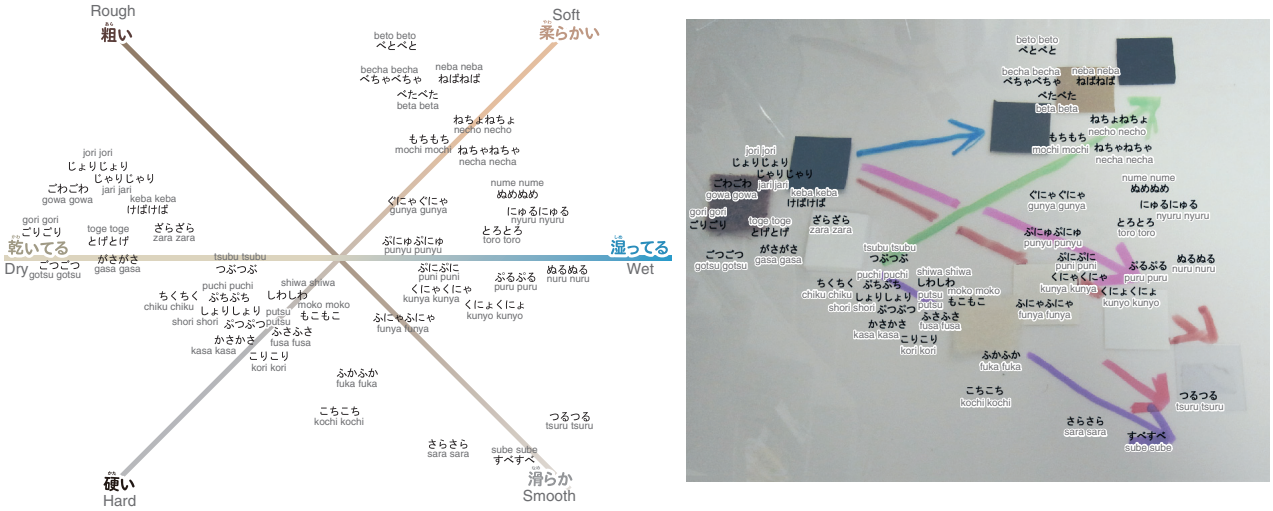


Fig. 3. Two-dimensional distribution map of tactile onomatopoeia arranged according to sensory images having words (left) and an example of arranging tactile materials on that onomatopoeia distribution map with arrows pointing in the direction from unfavorable to favorable materials (right).

when we make 'z' sounds. Through the research on the connection between sensations and words in this way, the "image of a human" in my mind gradually began to utter words.

In addition, while I had been observing humans and researching what they were feeling and thinking at certain moments in time, I am now beginning to view them from a continuous perspective. For example, when a person is motivated to behave in a certain way, how then does that person change the way of dealing with whatever he or she encounters? This constitutes a system loop within the human mind, which is something that I would like to pursue. In this regard, the increasingly popular practice of "mindfulness" can train a person on how to focus one's awareness on the present, that is, on what one is experiencing in the here and now. However, in addition to the immediate present, a human being features an ongoing internal loop that uses present happenings to trigger a look back at the past and maybe a look into the future. I am examining the role of the senses in this loop and searching out what might be possible in this area through technology.

For example, we touch our smartphones to get information. If haptic information should now be given in response to such an action, our feelings may intensify with respect to that information, and our motivation to do something may likewise become stronger. I believe that new technology of this type could be used to influence a person's mental cycle, change a person's behavior, or even improve a person's emotional state.

—The key to finding out what people are feeling or trying to convey seems to be interaction and communication.

We cannot feel the exact same sensations as those of another person, and in terms of communication, it is impossible to "completely" understand what another person is saying. Consequently, since words uttered to someone cannot be fully understood as the speaker intended, the sensation that "something is being understood" can be just an illusion. In other words, when a verbal exchange with someone proceeds without any problems, the sensation that "we are communicating" can likewise be an illusion. Speech consists of a "what" element indicating what is to be conveyed, and a "how" element indicating how that information is to be conveyed. "What" is often expressed by language in a conscious manner, while "how" is strongly related to subconscious movements such as body and hand gestures and facial expressions. Furthermore, communication can be achieved simply on a "how" basis without understanding "what." In fact, there are times in which "how" plays a more important role in communication. We can read a person's intent not only from "what" but also from "how" in a way that enriches communication.

When thinking alone, one only thinks about what can be conveyed through words, but if I can get another person to pick up on what I am subconsciously thinking about, I can uncover what was ambiguous and hidden in whatever I was thinking about. At that time, such an idea is not only mine nor the other person's—rather, it is something that arises between both of us. In communication between two people, each person is thinking about a variety of things above and beyond what each is saying. These thoughts influence the subconscious "how" portion of communication and are somehow conveyed to the other person. In addition, a mutual exchange of "how" tends to influence the substance of "what." Consequently, whatever idea is described in the end is a result born out of a mutual exchange beyond words. For two people who have created such a relationship, both contribute to such an outcome.

Pursuing whether technology can really make people happy

—How far has your research progressed? How will it develop in the future?

In relation to haptic research, I have been working on haptic design principles and wrote the book "Tactility for Generating Information" (**Fig. 4**, recipient of the 69th Mainichi Publishing Culture Award in the Natural Science Category). I am also involved in a project to map onomatopoeia in industries specializing in the use of tactile sensations, based on the onomatopoeia distribution map in Fig. 3 (joint research with Professor Maki Sakamoto of the University of Electro-Communications). Preparing and sharing texture maps in this way can lead to mutual understanding and communication in relation to the sense of touch. Moreover, in 2016, I conducted interviews with lumber-related companies and trading companies specializing in paper and created onomatopoeia distribution maps compiling words that I obtained from those interviews. At present, these are maps corresponding to the products of companies involved in the interviews, but in the future, I think it would be beneficial to create such maps in collaboration with other interested companies and organizations.

I would also like to move forward on the research of haptic communication. We can somehow sense whether being touched by someone else is good and what kind of emotion the other person is experiencing. If this is the case, it should be possible to read human sensory information by combining that information with the sense of touch or smartphone vibrations. At present, there is a lot of research on individual sensations of touch, but in the future, I think the principles of combining tactile materials as in the case of music (combinations of sounds) will become clear.

Fig. 4. "Tactility for Generating Information" (Kagaku-Dojin Publishing, Kyoto, Japan, 2014, in Japanese).

In January 2017, I supervised the Japanese translation of "Positive Computing: Technology for Wellbeing and Human Potential" [1]. This book describes how technology influences wellbeing and how wellbeing can be approached using technology. Today, society is overflowing with information technologies, and I would like to consider the results of accommodating and getting on well with those technologies instead of rejecting them outright and to then ask the question: "Can technology really make people happy?"

—Dr. Watanabe, can you leave us with a message for young up-and-coming researchers?

A sense of balance is vitally important for researchers. You should not research only what interests you. It is also important to choose something that others are not doing, that others are not good at, or that other people may enjoy. In the sense of creating something

of value for society on the whole, it would be better to have someone else do that work if that person is more qualified than you. With this approach, you will be able to demonstrate your own abilities in an optimal way. I believe it is necessary to have both a passion for continuing what you think is important and a calm, objective mind when thinking about value and what it means to society and a particular research field.

Reference

[1] R. A. Calvo and D. Peters, "Positive Computing: Technology for Wellbeing and Human Potential," The MIT Press, Cambridge, MA, USA, 2014.

■ **Interviewee profile**
Junji Watanabe

Senior Scientist (Distinguished Researcher), Human Information Science Laboratory, NTT Communication Science Laboratories.

He received a Ph.D. in information science and technology from the University of Tokyo in 2005. His academic work has been published in scientific journals in the field of neuroscience and interface technologies. He has also presented his work at technology showcases, science museums, and art festivals such as at SIGGRAPH (2006–2009, 2014) and Ars Electronica (2002, 2004, 2007–2017). His research is focused on cognitive science and communication devices with applied perception.

Creation of Artificial Intelligence Services through Open Innovation

Takashi Yagi and Hideaki Ozawa

Abstract

Amid growing expectations of service creation using artificial intelligence (AI), the NTT Group has announced the adoption of corevo® as a unified brand name encompassing the AI technologies born out of its research and development (R&D) efforts and the initiatives applying those technologies. The NTT Group aims to create new value using corevo in collaboration with a variety of partners. This article introduces the directions in R&D to drive the evolution of elemental AI technologies making up corevo and outlines NTT Group initiatives toward the creation of AI services through open innovation.

Keywords: artificial intelligence, corevo, open innovation

1. Introduction

With the aim of accelerating collaboration with a wide variety of partners in the field of artificial intelligence (AI), the NTT Group has unified its AI-related technologies and the initiatives using those technologies under a new brand name called corevo®. This word, meaning collaboration + revolution, evokes the idea of promoting innovation through a collaborative effort with a variety of players [1].

NTT classifies the AI technologies that make up corevo into four types: Agent-AI, Heart-Touching-AI, Ambient-AI, and Network-AI [2]. This classification reflects the type and meaning of the data handled by each AI and the functions and value provided. For example, Agent-AI deals mostly with media data including human speech, language, actions, and facial expressions, and features the ability to converse with humans. Ambient-AI, meanwhile, deals mostly with sensor data obtained from various types of equipment and environments such as automobiles and production facilities and aims to predict and control the near future.

Yet regardless of these four classifications, AI technology today is based for the most part on machine learning typified by *deep learning* and depends on data collection for achieving AI services. With this in mind, we have organized the elemental technologies needed for achieving AI services, that is, AI elemental technologies, data processing technologies, and data collection technologies, into the four layers shown in **Fig. 1** as an AI service architecture.

Here, the main role of data collection technologies making up the lowest layer is to gather data from the outside. This layer can also be treated as one made up of Internet of Things (IoT) technologies that are attracting attention today along with AI. The next higher layer consists of basic data processing technologies such as machine learning for processing the data collected in the lower layer. Continuing upward, the next layer consists of AI elemental technologies that apply those data processing technologies to perform tasks such as identification and inference as needed by services. Finally, the highest layer consists of AI services that are based on the processing results of the lower layers. Organizing the elemental technologies for AI services in this way makes it easy to understand that the four types of AI described above are not substantially different when it comes to configuring services.

In other words, Agent-AI, Heart-Touching-AI, Ambient-AI, and Network-AI are not necessarily independent when thinking about the final service configuration. For example, when conversing with a human by Agent-AI, it is sufficient to think of information obtained from peripheral equipment and the

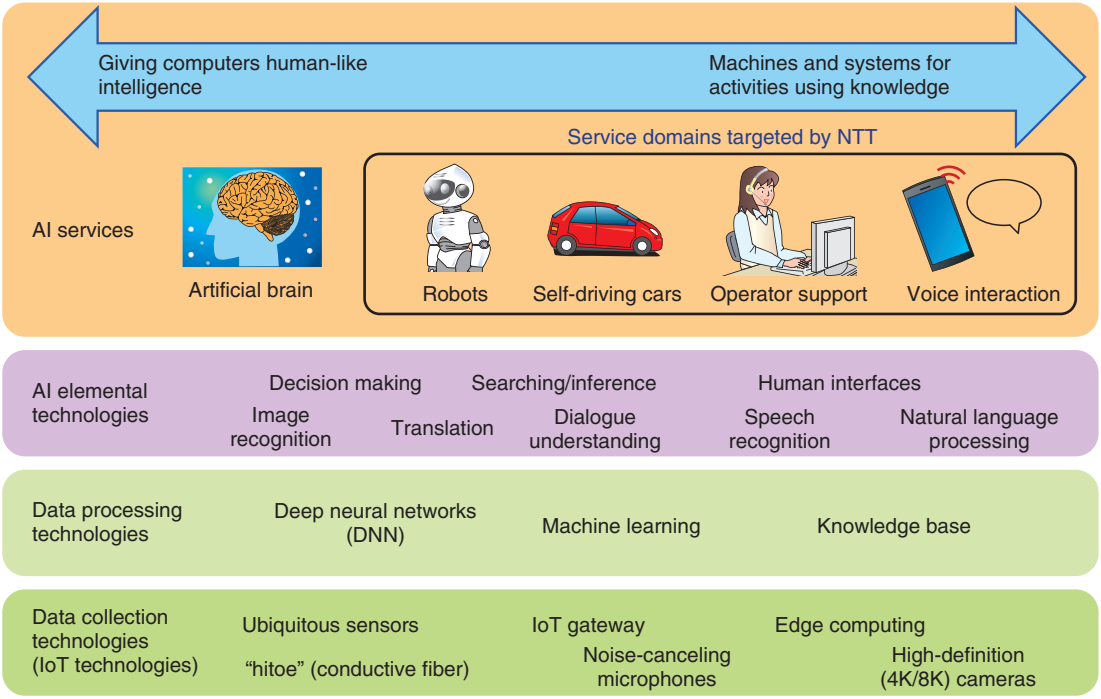


Fig. 1. AI service architecture.

environment as a knowledge source. Information predicted by Ambient-AI, meanwhile, can be used not only to control equipment but also to provide beneficial outcomes by presenting it to humans. That is, when viewed from an Agent-AI perspective, Ambient-AI can be treated as a knowledge source, but when viewed from an Ambient-AI perspective, Agent-AI can be treated as an advanced user interface. In this way, the linking and merging of these four types of AI will become a necessity as the demand for more advanced services grows. NTT is committed to researching and developing ways of linking and merging these four types of AI.

2. Evolution of AI elemental technologies

Each of the layers in the AI service architecture described above is considered to be evolving toward more advanced AI services. For example, we can expect data collection technologies to become more ubiquitous as sensors become smaller and use less power, which should drive progress in data conversion techniques enabling the integrated use of diverse types of data. Similarly, we can expect progress in data processing technologies as in dedicated hardware and distributed processing to enable high-speed processing of even larger volumes of data.

In parallel with these technological advances, the NTT laboratories will promote research and development (R&D) toward the evolution of AI elemental technologies. Although there are many AI elemental technologies in corevo that are already being put to practical use, we do not consider them to be complete technologies. For example, speech recognition technology has already been introduced in voice mining systems for call centers, where it has been shown to be sufficiently accurate for analyzing the trend of a received call. Nevertheless, considering that a call center will automatically perform a knowledge search based on the content of a call, erroneous recognition would tend to lower search accuracy in subsequent steps, so improving recognition accuracy even further is desirable.

Expanding the scope of speech recognition technology as in multilingual support and handling of regional dialects is also desirable. Our aim is to improve the accuracy and utility of these AI elemental technologies by combining the collection of large volumes of data with a high-speed processing platform and promoting model construction and tuning.

The AI elemental technologies making up corevo are wide-ranging, but R&D toward the practical use

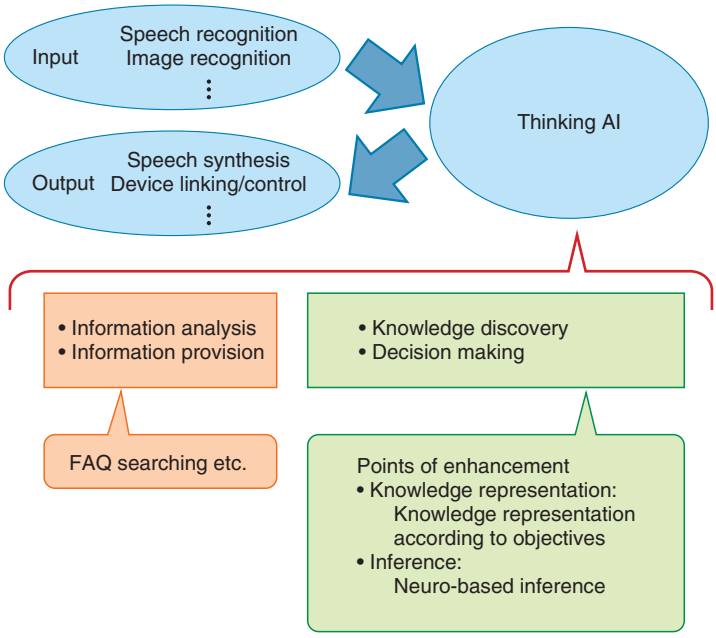


Fig. 2. Enhancement of Agent-AI elemental technologies.

of these technologies is moving ahead, particularly in the area of connecting people and computers. In terms of the four types of AI introduced above, this area would coincide most with Agent-AI. In addition, R&D within Agent-AI itself is focused on technologies such as speech recognition, image recognition, and device linking/control that involve input and output for interacting with the outside world. These are important elemental technologies for achieving AI services that *get close* to people (are considerate of human feelings), but achieving advanced AI services will require AI that understands input information and uses a variety of knowledge sources to derive output information.

We call this *thinking AI*, where *thinking* in this case refers to a processing mechanism achieved by a computer that corresponds to functions similar to those performed by humans. These would include basic processes such as comparing and analyzing the relationships among various items of data and then determining which are the same or different, judging which are good or bad, and classifying, associating, and surveying that data. We consider that thinking AI can be achieved by uncovering patterns of these functions through analytical analysis and machine learning such as deep neural networks for large volumes of data.

In fact, in applications such as frequently asked questions (FAQ) searching based on natural language queries, progress has already been made in using large volumes of data to learn semantic closeness from a linguistic viewpoint so that an explicit question can be understood and an appropriate reply returned [3]. Our plan going forward is to enhance our R&D efforts toward AI that can achieve even higher levels of knowledge discovery and decision making (**Fig. 2**).

3. Service creation through open innovation

The application scope of AI technology is extensive. In the future, it is not inconceivable that AI will be applied to services in all kinds of fields. However, at the current level of technology, the scope of applicability is limited. It is therefore necessary to search for areas in which the use of AI technology is likely to produce a substantial effect while increasing the potential for application by improving the accuracy of the technology. The technologies needed to configure AI services are likewise many and varied, and it is unrealistic to expect the NTT Group to cover all of those technologies. Open innovation with partners that can complement the NTT Group in terms of service ideas, know-how, field (data), and technology is therefore essential for the creation of AI services.

At present, the NTT Group is promoting collaboration

with many partners with the aim of creating services based on the B2B2X (business-to-business-to-X) business model. In the AI area, meanwhile, technology development and business creation are moving forward in the key areas of customer service, facilities maintenance support, healthcare and wellness, and hospitality services (navigation).

3.1 Customer service

The call center business is presently attracting attention as a very promising area for application of AI technology. For a business enterprise, a call center is an important point of contact with its customers, but the load on operators can increase as the quantity and complexity of products and services increase. As a result, operator turnover is high, while much time and effort is needed to train new people, and this presents major problems for call center businesses.

Efforts are underway to solve these problems at the NTT laboratories, and they have developed technologies such as voice mining and FAQ searching to support operators and supervisors. These technologies are already being put to use in actual services at many call centers [4]. Furthermore, as part of this trend, there are high expectations for automatic call reception by AI as a means of making call center operation even more efficient and for saving labor. The Feature Articles in this issue include an introduction to NTT's COTOHA™ communication engine for achieving automatic responses by AI [5].

3.2 Facilities maintenance support

Maintenance and inspection of production facilities/equipment at plants and of infrastructure facilities/equipment have often been dependent on the intuition and experience of veteran technicians. However, the baby boomer generation reached 65 years of age in 2012, and the decrease in the number of veteran technicians is becoming a major issue. As a result, there are also expectations in the facilities maintenance field to use AI technology to make business more efficient and advanced. In response, the NTT laboratories are developing technology for inferring the degree of wear in manholes using image analysis and working on advanced facilities maintenance through the use of three-dimensional point cloud data [6]. These Feature Articles also introduce work on detecting equipment anomalies by expanding NTT-developed acoustic technology to analyze sounds emitted by equipment [7].

3.3 Healthcare and wellness

The population aging rate in Japan (the population ratio of people age 65 or older compared to the total population) exceeded 21% in 2010, marking the arrival of a super-aged society and making healthcare and wellness an important theme. The NTT laboratories are researching the prediction and prevention of lifestyle-related diseases, dementia, and other conditions by reading and interpreting a person's physical and mental state by linking and interacting with medical and healthcare equipment and providing effective feedback. Under the themes of health management support using communication robots and therapeutic-recreation support, we are conducting trials in collaboration with hospital-affiliated universities and nursing institutions to identify issues and test the utility of developed technologies and services [8, 9]. These Feature Articles introduce technology for predicting patient behavior through the use of medical data in collaboration with university hospitals [10].

3.4 Hospitality services (navigation)

Hospitality services, while covering a conceptually broad area, are aimed at guiding the user to the information or location desired in real or virtual space in a safe and comfortable manner. In Japan, it is a theme driven not only by the growing number of foreign visitors to Japan and the need for regional revitalization but also by expectations that hospitality services can help enrich our daily lives.

Against this background, the NTT Group has been conducting trials involving tourism, event guidance, and other scenarios using device integration services [11]. These Feature Articles introduce two key activities in this area. The first involves the provision of information based on an image taken by a smartphone [12]. This makes it easy for the user to access desired information simply by inputting an image in settings where text or speech input is difficult. The second activity involves the inference of a person's destination or potentially dangerous driving from that person's mobility history or automobile driving data [13]. The aim here is to support safe and comfortable mobility.

4. Future prospects

The NTT Group is promoting open innovation in a variety of fields toward the creation of AI services using AI technologies under its corevo brand name. Initiatives for applying these services to actual

business efforts have been launched, but many are still in the trial stage. Looking to the future, we will continue to promote open innovation with a variety of players with the aim of developing AI technologies and creating services that can truly help to solve social problems and strengthen industrial competitiveness.

References

[1] Website of corevo, http://www.ntt.co.jp/activity/en/innovation/corevo/

[2] E. Maeda, "From Information Transmission to Mutual Understanding: Paradigm Shift in the Age of Data," NTT Technical Review, Vol. 14, No. 11, 2016. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201611fa1.html

[3] H. Asano, K. Bessho, K. Sadamitsu, K. Nishida, and K. Saito, "Natural Language Processing Technology for Agent Services," NTT Technical Review, Vol. 15, No. 8, 2017. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201708fa3.html

[4] S. Kawamura, K. Machida, K. Matsui, D. Sakamoto, and M. Ishii, "Utilization of Artificial Intelligence in Call Centers," NTT Technical Review, Vol. 14, No. 5, 2016. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201605fa7.html

[5] N. Furutaka and M. Hamada, "COTOHA™: Artificial Intelligence that Creates the Future by Actualizing Natural Japanese Conversation," NTT Technical Review, Vol. 15, No. 8, 2017. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201708fa4.html

[6] K. Sasaki, "Research and Development of Innovative Operation Technology for Access Network Infrastructure," NTT Technical Review, Vol. 15, No. 4, 2017. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201704fa4.html

[7] H. Uematsu, Y. Koizumi, S. Saito, A. Nakagawa, and N. Harada, "Anomaly Detection Technique in Sound to Detect Faulty Equipment," NTT Technical Review, Vol. 15, No. 8, 2017. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201708fa5.html

[8] NTT press release issued on July 28, 2015 (in Japanese), http://www.ntt.co.jp/news2015/1507/150728a.html

[9] NTT press release issued on November 25, 2015 (in Japanese), http://www.ntt.co.jp/news2015/1511/151125a.html

[10] 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

[11] T. Mochizuki, M. Norota, and K. Suga, "2020 Town—Developing MACHINAKA Service, a Device Integration Service that Utilizes Artificial Intelligence Technology," NTT Technical Review, Vol. 14, No. 12, 2016. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201612fa4.html

[12] S. Ando, I. Igarashi, T. Kinebuchi, T. Nakamura, D. Namikawa, R. Yamashita, Y. Yao, Y. Kusachi, and N. Takei, "Image Recognition Based Digital Watermarking Technology for Item Retrieval in Convenience Stores," NTT Technical Review, Vol. 15, No. 8, 2017. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201708fa2.html

[13] H. Toda, S. Yamamoto, and T. Nishimura, "Spatio-temporal Activity Recognition Technology for Achieving Proactive Navigation," NTT Technical Review, Vol. 15, No. 8, 2017. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201708fa6.html

Takashi Yagi

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

He received a B.E. in electrical engineering and an M.E. in computer science from Keio University, Kanagawa, in 1990 and 1992. He joined NTT Human Interface Laboratories in 1992. His research interests include human-computer interaction, computer-mediated communication, and artificial intelligence. He is a member of the Association for Computing Machinery, the Institute of Electronics, Information and Communication Engineers, the Information Processing Society of Japan, and the Virtual Reality Society of Japan.

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. He joined NTT in 1991 and 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 was involved in developing and managing "goo" internet services including search engine and healthcare services and also worked on establishing new services at NTT Resonant Inc. In 2004, he assumed his current position at NTT Media Intelligence Laboratories. Since May 2013, he has also held the position of president of NTT Resonant Technology, Inc., which develops software for mobile operating systems.

Image Recognition Based Digital Watermarking Technology for Item Retrieval in Convenience Stores

Shingo Ando, Isamu Igarashi, Tetsuya Kinebuchi, Taiji Nakamura, Daichi Namikawa, Ryo Yamashita, Yasuhiro Yao, Yoshinori Kusachi, and Nobukatsu Takei

Abstract

With image recognition based digital watermarking technology, the cameras of mobile devices are used to detect with high accuracy invisible ID (identifier) information embedded in printed matter such as item packaging. This article overviews this technology and a collaborative experiment conducted with the retail group Seven & i Holdings that began in November 2016.

Keywords: digital watermark, angle-free object search, service for inbound passengers

1. Information retrieval by image capture

Recent years have been characterized by the high performance and rapid acceptance of mobile terminals such as smartphones and tablets. Concurrently, various services for mobile terminals have been released and adopted. Among them, the service called mobile visual search is attracting the interest of many users. Mobile visual search refers to services and technologies that recognize objects from images captured by the cameras built into mobile terminals and search or present various bits of information related to the objects. Paintings, buildings, books, and DVDs (digital versatile discs) are typical service targets. Related web pages, the names or locations of the objects, and images of similar objects are the most commonly returned results presented to the user. When users do not know the name of an object or scene and want to learn more about it or want to find out related information such as guidance and personal communication, they can search simply by directing the camera towards the target. This makes mobile visual search much more convenient than word-based search.

Such services are also being developed by NTT, for example, SightX [1] and Hospitality UI/UX (user interface/user experience) [2] and are being actively researched through the cooperation of several laboratories.

This article first overviews mobile visual search and then introduces a digital watermark service, angle-free object search, and image recognition based digital watermarking technology, which is the fusion of both technologies.

2. Mobile visual search

The processing flow of mobile visual search is illustrated in **Fig. 1**. A mobile terminal reads the identifier (ID) code such as a barcode on the target from the query image (which is the image serving as the search key) captured by the camera and sends it to a server, or sends the query image itself to a server. On the server side, the uniform resource locator (URL) of the related web page associated with the received ID is returned to the mobile terminal, or the subject is identified from the received query image by one of the image recognition technologies, and the server

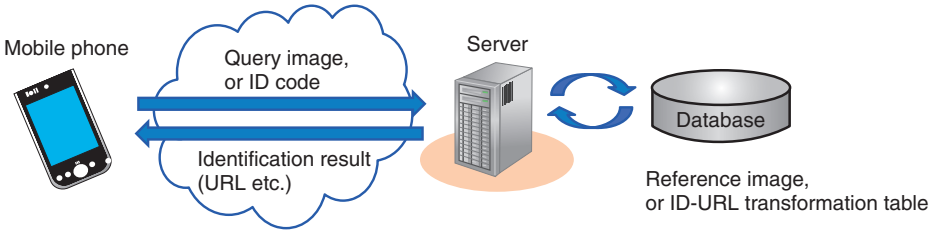


Fig. 1. Mobile visual search.

returns the URLs of related web pages to the mobile terminal.

Various means are available for identifying the target, but there are two broad categories. The first identifies each target by its printed ID code such as a gray scale pattern or barcode or a two-dimensional code or Q-code. The digital watermarking technologies described later also fall into this category. In this case, it is necessary to add unique code patterns to the objects in advance, and the user must find the patterns; the advantage of this is that highly accurate identification is possible.

The other category encompasses image recognition technologies. This is achieved by registering feature values extracted from images of objects in a database in advance and identifying what the object in the query image is by comparing the feature values extracted from the query image with the feature values in the database. The angle-free object search technology described later falls into this category. In this case, there is no need to add anything to the object in advance, but if very similar items exist, it may be difficult to distinguish them. In this way, since the two approaches complement each other, they should be used appropriately according to the purpose or in combination.

3. Digital watermarking technologies

Researchers at NTT Media Intelligence Laboratories have been actively researching digital watermarks for many years, and our proprietary algorithm boasts high reading accuracy and high-speed operation [3–5]. It is assumed to be mainly used for intermedia synchronization. If still images, printed material, or movies contain the watermarks, it is possible to read the watermark ID at high speed just by directing the camera of the mobile terminal towards the target and to access related information.

Our digital watermarking technology [3] first detects the watermark-embedded regions from the image using the quadrilateral fast tracking method called Side Trace Algorithm (STA) [6]. It is assumed that the watermark-embedded region lies within a thin frame with four sides. Next, projective transformation distortion is corrected so that the detected area becomes a square of predetermined size. Then, the digital watermark pattern is extracted from the corrected image, and the watermark ID is read. Since the pre-embedded digital watermark pattern can be extracted by very simple image processing operations, it can be processed very rapidly even by low power devices such as old style mobile terminals. Due to the characteristics of the digital watermark, the appearance of the target printed material is changed, but the degradation in image quality is slight, and it is rare for anyone to notice the watermark in normal use. Furthermore, the projective transformation distortion correction process enables the watermark ID to be read very stably even when captured at an oblique angle.

4. Angle-free object search

Angle-free object search [7] is a technology developed by NTT Media Intelligence Laboratories that can recognize and retrieve three-dimensional (3D) objects with high accuracy and present relevant information no matter which direction the 3D objects are viewed from. With this technology, surrounding buildings, historical sites, signboards, electronic devices, and other objects are accurately recognized through the camera of the mobile terminal, and information such as tourist contents, route guidance information, and operation manuals are presented.

The angle-free object search technology is based on NTT's robust media search (RMS), which is a fast search technology for sound and video [8], and robust object search technology (RMS-object), which has evolved from object identification technologies.

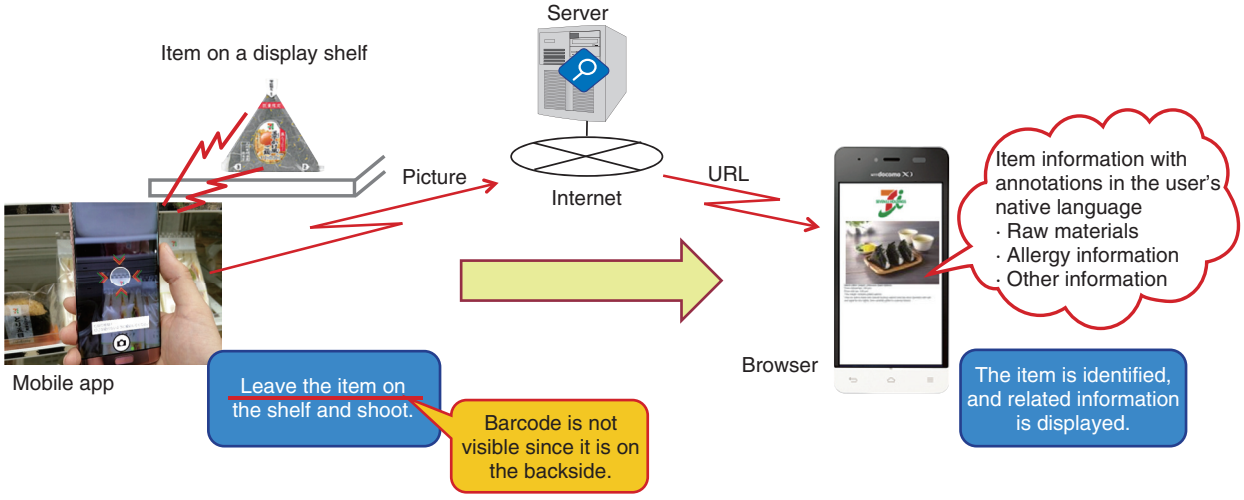


Fig. 2. Item retrievals in convenience store.

First, the relationship between the image features between the input image and the reference image is accurately specified by a unique matching process using constraint conditions on the same solid object derived from projective geometry. This allows the number of reference images prepared in advance to be reduced by about 90% from the conventional technique. The identification accuracy is very high since the importance of image features is statistically estimated based on their frequency of occurrence, and the matching process is carried out taking their importance into account. Furthermore, the image feature database is indexed by hashing to yield short codes by using an original method that considers the distribution in feature space. As a result, the method can locate the image feature group that matches the input image in the image feature database about two times faster than the previous technique.

5. Image recognition based digital watermarking technology

As mentioned above, the digital watermarking technology and angle-free object search technology offer a wide range of use cases. However, several important use cases are not covered well by either technology. For example, it is difficult for foreign tourists to understand the contents or ingredients of certain objects, for example, rice balls sold at convenience stores. Therefore, it would be very useful to have a service that displays raw materials and allergy information in the visitor's native language just by capturing the item with the camera (**Fig. 2**). Barcodes are usually attached to items for inventory control, but the barcode is often in an inconspicuous place such as the backside of the product, so you would have to handle the item to uncover the barcode. This need for frequent handling of food items is troublesome to both the store and the visitor. Therefore, there is a need for a system that recognizes and retrieves items displayed in showcases without having to pick up the items.

Although image recognition technology such as that based on angle-free object search has evolved significantly in recent years, it is still unable to perfectly discriminate very similar products. Therefore, it is considered unsuitable for high-risk information presentation services such as allergy labeling. However, NTT laboratories' digital watermarking technology achieves almost 100% accurate identification, but as described above, it is necessary to add an explicit frame around the watermark-embedded area. It is still considered unsuitable for commercial products, as it changes the visual design of the items.

Therefore, we have developed image recognition based digital watermarking technology that integrates angle-free object search with digital watermarking. In this method, a watermark is embedded within a predetermined pattern such as a private brand logo. The watermark embedded area is extracted from the image using the logo as the clue, and the watermark ID within the area is read.

The processing outline of our proposed technology is shown in **Fig. 3**. Angle-free object search is used

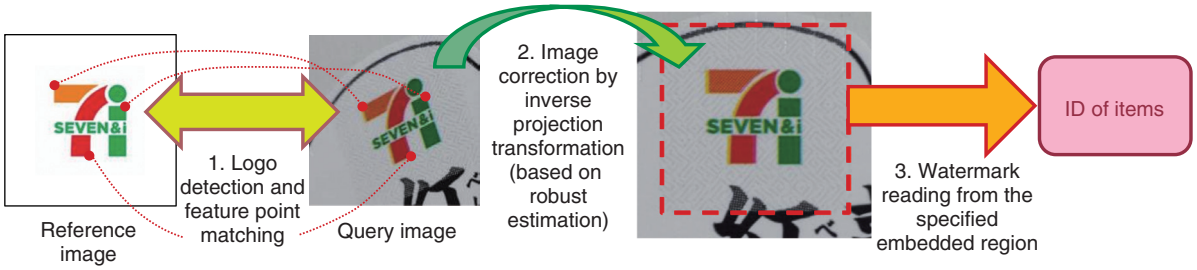


Fig. 3. Overview of the image recognition based digital watermarking technology.

for pattern detection. In addition to detecting patterns, angle-free object search can also identify multiple feature point locations present within the logo with high accuracy. This information can be used to correct projection transformation distortion by applying STA, which detects quadrilateral frames and corrects the distortion of the frames. In particular, by using the robust estimation method called RANSAC (random sample consensus), a projective transformation matrix is automatically derived from the spatial relationship of the feature points, and the projective transformation distortion is then corrected so that the detected area becomes a square of predetermined size. Subsequent processing is the same as that performed in the above-mentioned digital watermarking technology.

The key advantage of this method is that watermarks can be printed without using explicit frames or markers, which makes them visually inconspicuous. Moreover, watermark IDs can be stably read even from oblique angles of up to 45 degrees and when the watermark size is extremely small at about 1 cm square. This makes it suitable for small packages such as rice ball items. These advantages are not offered by any existing technology.

6. Field trial

We conducted experiments of this technology in the laboratory as well as in the field. The experiments and their results are reported in this section.

6.1 Preliminary experiments in laboratory environment

First, the results of a simple preliminary experiment in our laboratory are described. Logo patterns of Seven & i Holdings containing watermark patterns were printed to create the labels often used on food items such as rice ball packaging. We calculated the identification rate based on hundreds of query images taken using commercial smartphone cameras. Query images were obtained by taking images at various angles (up to a 45-degree inclination from the front) while rotating the items themselves by up to 360 degrees from a distance of around 15 cm from the target items. As a result, the correct identification rate was 98.4%, the incorrect identification rate was 0%, and the reject rate (where the application urged the user to re-take the image) was 1.6%. The total processing time on the server was about 1.5 seconds on average per query image. Furthermore, simulations involving about 1 billion artificial samples confirmed that the incorrect identification rate was on the order of approximately 10^{-7}, which is lower than the misreading rate of the barcode standard (less than one in 3 million).

6.2 Field trials at a real store

Next, we explain the results of field trials conducted at an actual 7-Eleven convenience store in Chiyoda-ku in Tokyo; the experiments began in November 2016. In this trial, several subjects shot the logo of the item sample placed in the showcase of the store with the mobile terminal under test, and we verified whether the watermark ID embedded in the logo could be recognized correctly. We ordered the watermark-embedded items from the printer who would normally print the actual item packages. The packing was then used to wrap actual products in the standard way.

We tested three items: a rice ball with red salmon (paper label), a rice ball with chicken and five ingredients (transparent film), and a mixed sandwich (transparent film). We wrote a special application and installed it on the mobile terminal. A target scope was superimposed on the preview screen at the time of image capture, and various instructions were given so that the user could perform image capture reliably

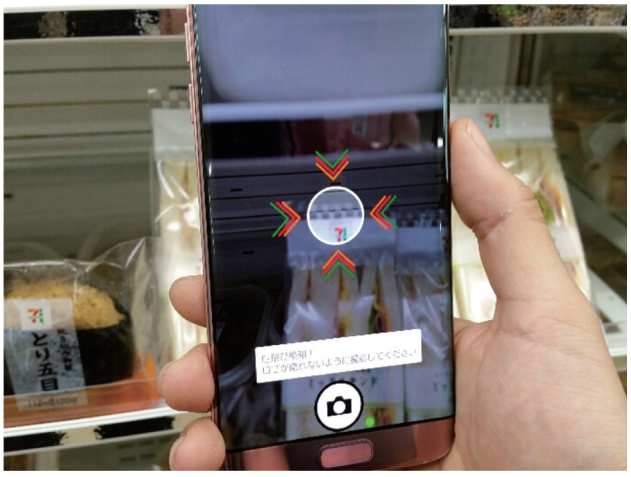


Fig. 4. Image captured by the mobile application in the collaborative experiment.

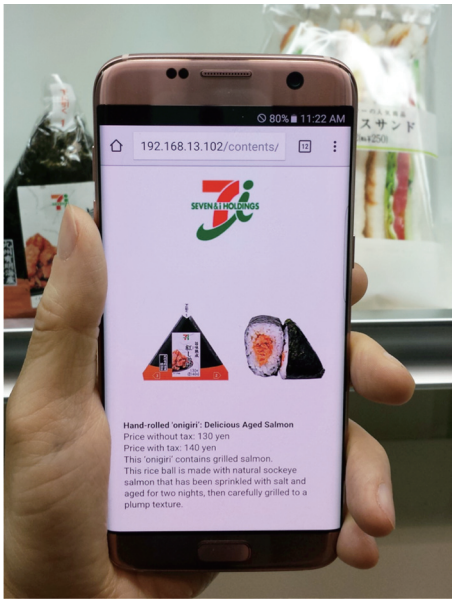


Fig. 5. Screen capture of the item retrieval result.

(**Fig. 4**). In particular, it was important that the application automatically detected *clipped whites* (i.e., overexposed areas, where there is a loss of highlight detail) and *defocus* (an unfocused image), which greatly affect the accuracy of reading the watermark ID, so an application was included that issued a warning to the user if the image was improperly captured. Furthermore, after the watermark ID was read, detailed information on the item was displayed via the web browser on the terminal in the user's native language (automatically selected from the language setting information of the operating system of the mobile terminal) (**Fig. 5**). As a result, we recorded a misreading rate of 0%, which is similar to the results of the laboratory environment, and we were able to achieve high customer satisfaction.

The field trial is being continued as of April 2017, and we plan to conduct usability verification with more customers as subjects.

7. Future development

The image recognition based digital watermarking technology developed by NTT Media Intelligence Laboratories was introduced, and the results of field trials were reported. We are working to solve the problems identified in the field trial and are continuing research and development with the aim of achieving comprehensive service creation.

References

[1] D. Namikawa, H. Minami, H. Kataoka, M. Makiguchi, and M. Shimomura, "SightX: Obtaining Information on a Scene by Pointing a Camera," NTT Technical Review, Vol. 11, No. 9, 2013. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201309fa5.html

[2] Y. Ichikawa, Y. Nakamura, T. Nakamura, H. Tezuka, H. Seshimo, and S. Fukada, "2020 Airport/Station—Hospitality for Foreign Visitors at Airports and Train Stations," NTT Technical Review, Vol. 14, No. 12, 2016. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201612fa2.html

[3] T. Nakamura, A. Katayama, M. Yamamuro, and N. Sonehara, "Fast Watermark Detection Scheme from Analog Image for Camera-equipped Cellular Phone," IEICE Trans. Inf. & Syst. (Japanese Edition), Vol. J87-D-II, No. 12, pp. 2145–2155, 2004.

[4] S. Ando, S. Yamamoto, K. Tsutsuguchi, A. Katayama, and Y. Taniguchi, "Field Test of Mobile Video Watermarking Technology over Terrestrial Digital Broadcasting," ITE Tech. Rep., Vol. 37, No. 38, pp. 57–62, 2013 (in Japanese).

[5] S. Ando, S. Yamamoto, H. Tanaka, K. Tsutsuguchi, A. Katayama, and Y. Taniguchi, "Visual SyncAR: Video Synchronized AR based on Mobile Video Watermark," IIEEJ Trans. Image Electronics and Visual Computing, Vol. 4, No. 2, pp. 114–123, 2016.

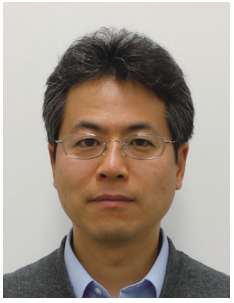
[6] A. Katayama, T. Nakamura, M. Yamamuro, and N. Sonehara, "A New High-speed Corner Detection Method: Side Trace Algorithm (STA)," IEICE Trans. Inf. & Syst. (Japanese Edition), Vol.J88-D-II, No. 6, pp. 1035–1046, 2005.

[7] J. Shimamura, T. Yoshida, and Y. Taniguchi, "View-directional Consistency Constraints for Robust 3D Object Recognition," IIEEJ Trans. Image Electronics and Visual Computing, Vol. 3, No. 2, pp. 164–173, 2015.

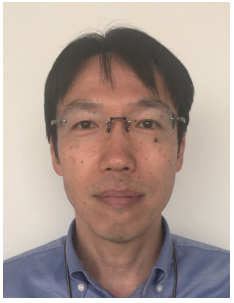
[8] K. Kashino, "Searching and Utilizing Large-scale Media Data," NTT Technical Journal, Vol. 26, No. 4, 2014 (in Japanese).

Shingo Ando
Research Engineer, 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 the Institute of Electronics, Information and Communication Engineers (IEICE) and the Institute of Image Information and Television Engineers (ITE).

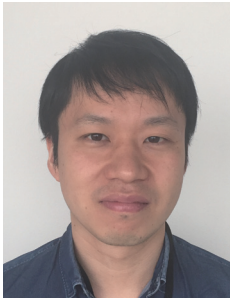
Isamu Igarashi
Researcher, Visual Media Project, NTT Media Intelligence Laboratories.
He received a B.E. and M.E. in electronic engineering from Tohoku University, Miyagi, in 2007 and 2009. He joined NTT in 2009. He is involved in research and practical application development in the fields of image processing and pattern recognition. He is a member of IEICE.



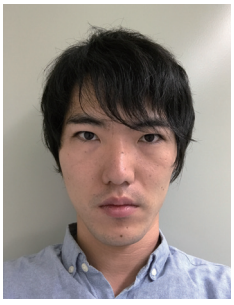
Tetsuya Kinebuchi
Senior Research Engineer, Group Leader of Image Media Processing Group, Visual Media Project, NTT Media Intelligence Laboratories.
He received an M.S. in physics from Tohoku University, Miyagi, in 1997. He joined NTT in 1997. His research interests include pattern recognition and image processing. He is a member of IEICE.



Taiji Nakamura
Senior Research Engineer, 2020 Epoch-making Project, NTT Service Evolution Laboratories.
He received a B.E. in industrial engineering from Tokyo University of Science. He joined NTT DATA Communications Systems (now NTT DATA) in 1991. He has over 20 years' experience in information systems planning and development in the national public sector.



Daichi Namikawa
Researcher, 2020 Epoch-making Project, NTT Service Evolution Laboratories.
He received a B.E. and M.E. from Kagoshima University in 2007 and 2009. He joined NTT in 2009 and has been researching and developing enhanced network services and systems. He is a member of the Information Processing Society of Japan.



Ryo Yamashita
Researcher, 2020 Epoch-making Project, NTT Service Evolution Laboratories.
He received a B.E. and M.E. from Nagoya University in 2010 and 2012. He joined NTT in 2012 and has been engaged in the research and development of human computer interaction interfaces. He is a member of the Association for Computing Machinery.

Yasuhiro Yao
Researcher, Cross Media Project, NTT Media Intelligence Laboratories.
He received a B.S. in pharmaceutical sciences from the University of Tokyo in 2007 and an M.E. in quantum engineering and system sciences from the University of Tokyo in 2010. He joined NTT in 2010. He is involved in research and business development in the fields of image processing and pattern recognition.

Yoshinori Kusachi
Senior Research Engineer, Cross Media Project, NTT Media Intelligence Laboratories.
He received an M.E. from Nara Institute of Science and Technology in 1997 and a Ph.D. in information engineering from the same institute in 2007. He joined NTT in 1997. He has been engaged in research and practical application development in the fields of image processing and pattern recognition.

Nobukatsu Takei
Senior Manager, R&D Produce Group, Research and Development Planning Department, NTT.
He received a Bachelor of Physics from the University of Tokyo in 1989 and joined NTT the same year. He worked on the development of optical transmission systems and network operations at NTT EAST. He has also been involved in general management of NTT R&D laboratories.

Natural Language Processing Technology for Agent Services

Hisako Asano, Katsuji Bessho, Kugatsu Sadamitsu, Kyosuke Nishida, and Kuniko Saito

Abstract

Artificial intelligence agents are beginning to take on the work traditionally performed by people at contact centers and information desks. This article introduces natural language processing technology for achieving agents that can absorb business knowledge, understand customer speech and respond to their requests, engage in casual conversation, and behave like a human being overall.

Keywords: agent, natural language processing, artificial intelligence

1. Introduction

An agent service is a type of service in which an artificial intelligence (AI) agent takes the place of a human being. Businesses such as contact centers and information desks that respond to inquiries based on specialized knowledge have growing needs. The user (customer) of an agent service makes inquiries or requests by talking or chatting by text with an agent. The agent, meanwhile, makes an attempt to understand what the user is saying, and based on a previously learned business flow and business knowledge, provides the information the user desires, or executes a procedure requested by the user after confirming with the user the information needed for that procedure. At this time, a natural response from the agent might include casual conversation in addition to dealing with the current task. Such an exchange between a user and agent is carried out in everyday language used by people, that is, in *natural language*.

We note here that the huge amount of specialized knowledge in business documents such as manuals that must be acquired by business personnel is also written in natural language. An AI agent, however, performs processing using computer-oriented (machine) languages (database query languages, web application programming interfaces (APIs), etc.) and structured data (relational databases, etc.). Therefore, information in user utterances and business documents expressed in natural language must be converted to a machine-readable form.

Agent-AI powered by the NTT Group AI technology called corevo® is technology that supports humans by interpreting the information they generate. Natural language processing technology is one important technology making up Agent-AI [1]. This article introduces natural language processing technology for achieving agents that can respond accurately to specialized inquiries from users while being equipped with a variety of knowledge groups enabling intelligent and friendly conversation on a wide range of topics with users (**Fig. 1**).

2. Natural-language equivalence judgement technology for understanding diverse expressions

An agent possesses knowledge expressed in natural language, and its function is to compare natural-language utterances made by the user with that natural-language knowledge. For example, frequently asked questions (FAQ) is a document or website link consisting of knowledge written in natural language in the form of question (Q) and answer (A) pairs. As such, it can be used to respond to a query from the user by comparing user utterances with FAQ questions and presenting to the user the answer (A) corresponding to the question (Q) matching those utterances.

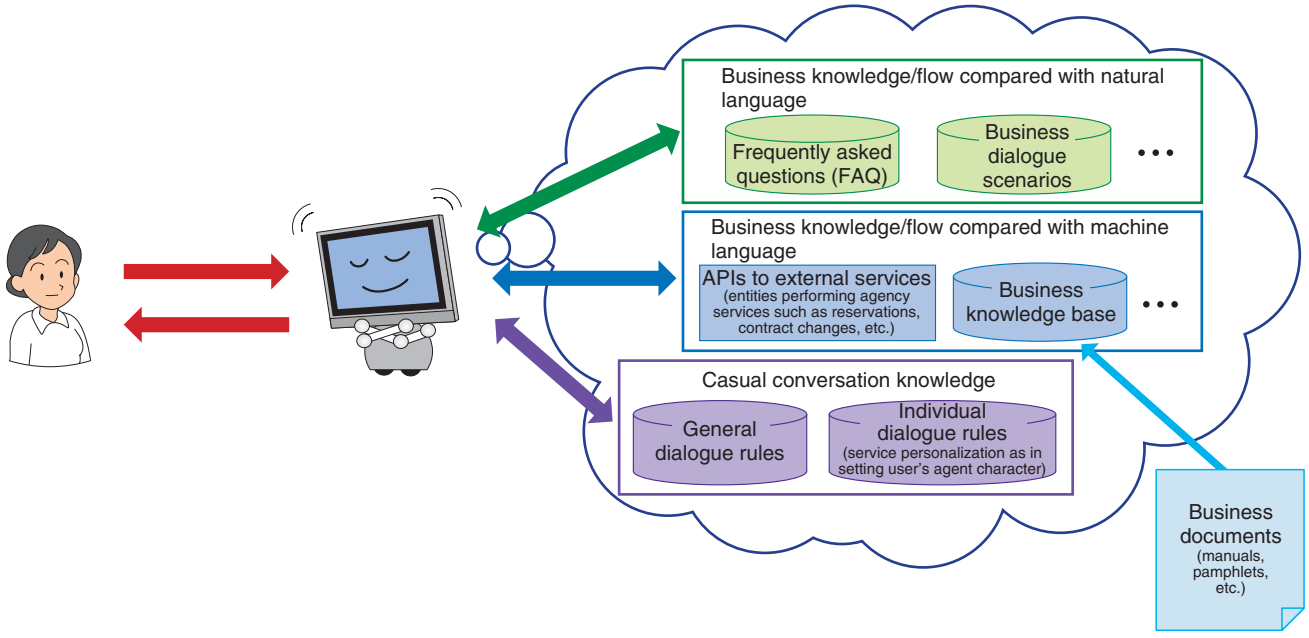


Fig. 1. Agent equipped with diverse knowledge groups.

In addition, each dialogue rule in business dialogue scenarios (a set of dialogue rules based on a business flow) consists of a user utterance (U) and system utterance (S) pair. These rules can be used to compare a user's utterance with stored user utterances (U) and return the system utterance (S) corresponding to the user utterance (U) matching the actual utterance made by the user. However, agent knowledge targeted for comparison rarely matches a user utterance at the character-string level given the diversity of user expressions, so there is a need for technology that can determine equivalence between expressions in terms of intention.

To meet this need, NTT Media Intelligence Laboratories developed natural-language equivalence judgement technology (**Fig. 2**). This technology dynamically constructs quantified semantic concepts of words based on a large-scale Japanese language semantic dictionary and on word occurrence distribution in large volumes of text. It then judges semantic similarity in natural-language statements targeted for comparison. This approach makes it possible to understand diverse expressions characteristic of the Japanese language and to identify with high accuracy agent knowledge that semantically agrees with user utterances.

3. Information extraction technology for providing information and executing external services from user utterances

NTT Media Intelligence Laboratories is also researching and developing technology for extracting information from user utterances for two main purposes. The first is to provide pinpoint information using business knowledge such as fees for specific product options. The second is to perform procedures or obtain information using systems external to the agent as in making/checking reservations or checking inventory.

First, we describe the flow of providing pinpoint information based on business knowledge, referring to **Fig. 3**. In response to a user request for information, the agent provides appropriate information using a knowledge base. For example, given the question "Until when will contract X provide compensation?" from the user, the system interprets this natural-language statement and returns "insurance period" for "contract X" as a reply.

Of course, there are many and varied products other than insurance policies in the world, and with this in mind, NTT Media Intelligence Laboratories has developed technology that enables robust information extraction even for user utterances related to unknown products through the use of general

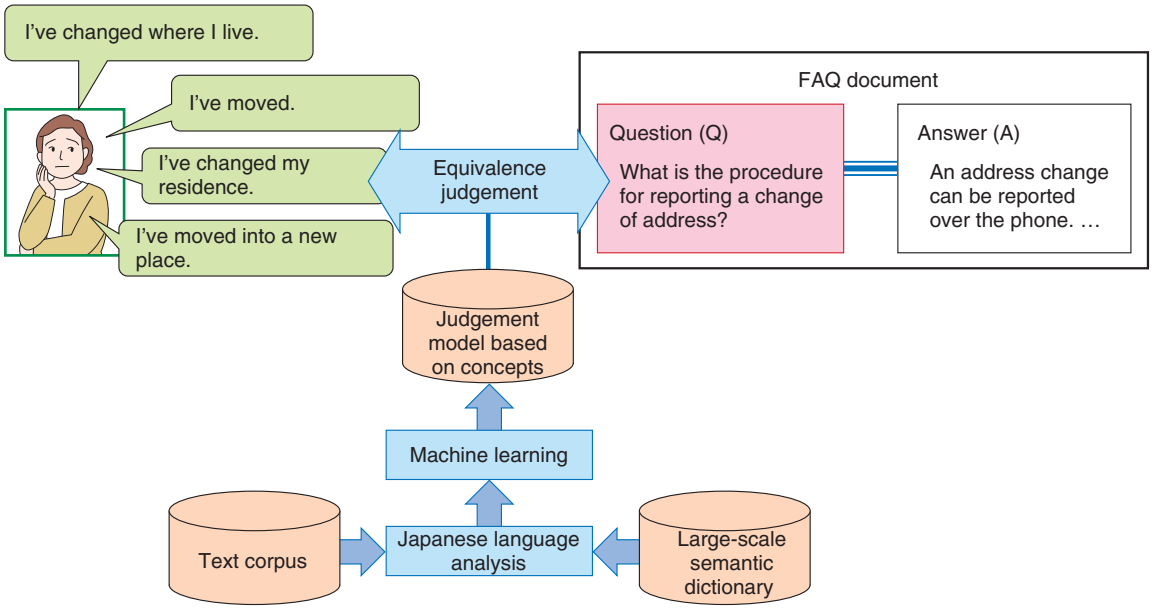


Fig. 2. Natural-language equivalence judgement technology.

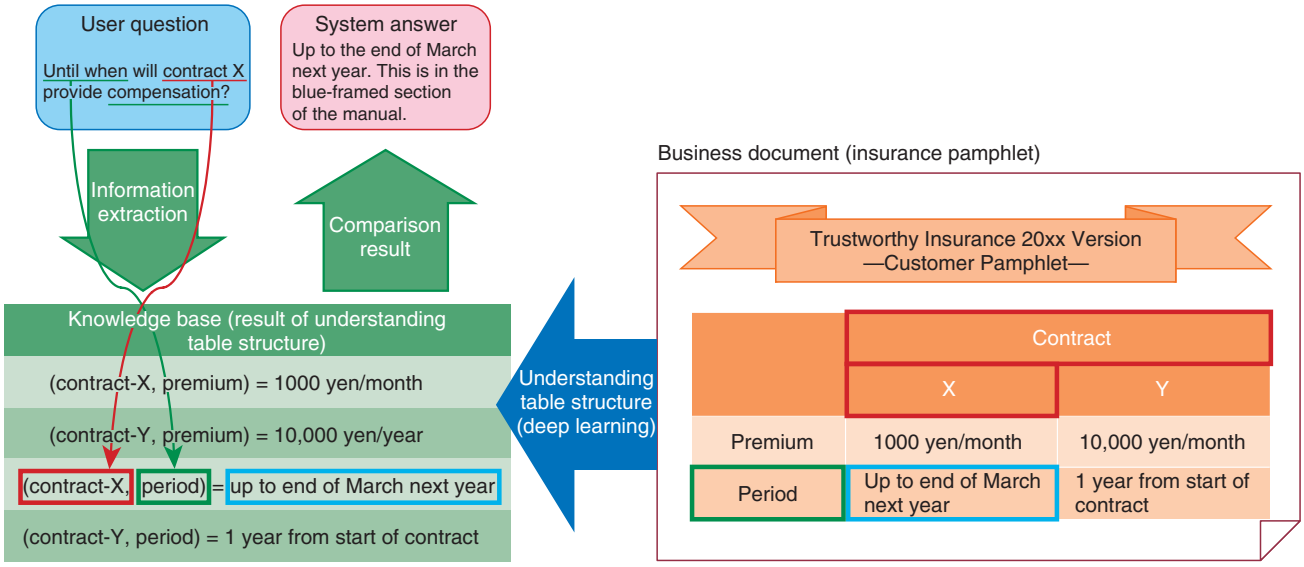


Fig. 3. Provision of pinpoint information based on business knowledge.

language features. This is achieved by automatically inferring the type of question posed. In this example, the technology would infer that the user is asking about "period" when speaking the words "until when" and would use this inference as a general language feature [2].

Next, in the case of external services, the system automatically performs processing to connect to an external API and satisfy the user's request. For example, in response to a user who simply says, "I would like to receive an insurance consultation on Saturday this week," it would be possible to automatically execute an appointment procedure by understanding the intention of specific keywords in the user's

utterance. Here, the system would infer the desired date to be "July 24" (the actual date indicated by "Saturday this week"), the user's objective to be "insurance consultation," and the desired processing to be "appointment."

4. Table classification technology for converting knowledge from human use to AI use

Discovering intelligent information within documents and putting it into an AI-readable form is an important role of natural language processing technology. Knowledge acquired from documents is useful for improving information retrieval and Q&A services, which search for information from a massive amount of data, and for supporting human activities in contact centers and elsewhere.

Documents are filled with information in various formats, and tabular data, in particular, group together important information. However, AI has not been able to make maximum use of such data as knowledge. Tables that we usually unconsciously read fall into a number of types, and we cannot acquire knowledge correctly from the table unless we can distinguish between the tables (Fig. 3). Research to identify different types of tables has been carried out since the first half of the 2000s, but the ability to accurately and exhaustively acquire knowledge from documents has not reached human level.

What kind of ability is necessary to identify table types that AI has not been able to achieve? It is the ability to understand both the meaning of the text written in each cell in the table and the semantically characteristic blocks of the cells. To address these two requirements, NTT Media Intelligence Laboratories proposed hybrid deep learning technology called TabNet that combines a neural network specialized in understanding sequence data as in text and a neural network specialized in discovering features from matrix data as in an image [3]. This technology makes it possible to read tables as humans do and to acquire knowledge from documents with high accuracy.

5. Dialogue rule database enabling replies in casual conversations with humans

Is it enough for agents to play a role only in replying to inquiries or supporting information desks? It has been reported that we humans have a tendency to chat with inanimate objects that have the look and feel of a human being [4]. To achieve agents that are more human in nature, more research is being carried out to find the means of achieving natural replies in casual conversations with humans.

To deal with casual conversation, an agent requires dialogue knowledge for responding appropriately to a wide range of user utterances. Although techniques exist for automatically creating dialogue knowledge from large volumes of text data, manually describing dialogue rules is an effective technique for achieving more accurate and effective replies. NTT Media Intelligence Laboratories has constructed a large-scale dialogue rule database. This database deals with a wide variety of expressions in user utterances by taking into account synonyms like *liquor* and *alcohol* and *drink* and *enjoy* and expressions like *what is* and *I want to know* that express the same intent. Furthermore, by repeatedly improving rules through dialogue experiments, we have succeeded in constructing a database comprising approximately 300,000 rules, which enables natural dialogues with humans with high accuracy.

In actual services, there are ways of creating more attractive agents with high-quality replies. For example, agent characters can be configured as needed, and a separate set of dialogue rules customized to the main task of a service can be prepared and used together with the dialogue rule database.

6. Future development

The technology presented here is gradually being implemented in the COTOHA^{TM} communication engine introduced in the feature article in this issue entitled, "COTOHA^{TM}: Artificial Intelligence that Creates the Future by Actualizing Natural Japanese Conversation" [5]. The plan going forward is to expand the use of this technology in NTT Group agent services.

NTT is committed to researching and developing technologies for enhancing functions that enable AI to learn on its own and for achieving more intelligent and personalized agents.

References

[1] Y. Matsuo, R. Higashinaka, H. Asano, and T. Makino, "Natural Language Processing Supporting Artificial Intelligence," NTT Technical Review, Vol. 14, No. 5, 2016.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201605fa2.html

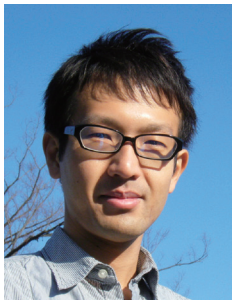
[2] K. Sadamitsu, Y. Homma, R. Higashinaka, and Y. Matsuo, "Zero-shot Learning of Natural Language Understanding for Information Provision in Open Domains," The 30th Annual Conference of the Japanese Society for Artificial Intelligence, 2016 (in Japanese).

[3] K. Nishida, K. Sadamitsu, R. Higashinaka, and Y. Matsuo, "Understanding the Semantic Structures of Tables with a Hybrid Deep Neural

Network Architecture," Proc. of AAAI 2017 (Thirty-First AAAI Conference on Artificial Intelligence 2017), pp. 168–174, San Francisco, CA, USA, Feb. 2017.
[4] B. Reeves and C. Nass, "The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places," CSLI Publications and Cambridge University Press, 1996.
[5] N. Furutaka and M. Hamada, "COTOHA™: Artificial Intelligence that Creates the Future by Actualizing Natural Japanese Conversation," NTT Technical Review, Vol. 15, No. 8, 2017. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201708fa4.html

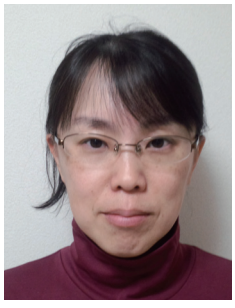
Hisako Asano
Senior Research Engineer, Supervisor, Audio, Speech, and Language Media Laboratory, NTT Media Intelligence Laboratories.
She received a B.E. in information engineering from Yokohama National University, Kanagawa, in 1991. She joined NTT Information Processing Laboratories in 1991. Her research interests include natural language processing, especially morphological analysis, and information extraction. She is a member of the Information Processing Society of Japan (IPSJ) and the Association for Natural Language Processing (NLP).

Katsuji Bessho
Senior Research Engineer, Advanced Speech and Language Technology Group, Audio, Speech, and Language Media Laboratory, NTT Media Intelligence Laboratories.
He received a B.S. and M.S. in mathematics from Osaka University in 1992 and 1994 and a Ph.D. in engineering from Niigata University in 2009. He joined NTT Information and Communication Systems Laboratories in 1994. He is currently engaged in the research and development of natural language processing. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE), IPSJ, and NLP.



Kugatsu Sadamitsu
Researcher, Audio, Speech, and Language Media Project, NTT Media Intelligence Laboratories.
He received a B.E., M.E., and Ph.D. in engineering from the University of Tsukuba, Ibaraki, in 2004, 2006, and 2009. He joined NTT Cyber Space Laboratories (now NTT Media Intelligence Laboratories) in 2009. His current research interests include natural language processing and machine learning. He is a member of NLP.

Kyosuke Nishida
Senior Research Engineer, NTT Media Intelligence Laboratories.
He received his B.E., MIS, and Ph.D. in information science and technology from Hokkaido University in 2004, 2006, and 2008. He joined NTT in 2009. He was a chief engineer at NTT Resonant from 2013 to 2015. Since 2015, he has been a senior research engineer at NTT Media Intelligence Laboratories. His current interests include natural language processing, web mining, and ubiquitous computing. He is a member of IPSJ, IEICE, and the Database Society of Japan.



Kuniko Saito
Senior Research Engineer, NTT Media Intelligence Laboratories.
She received a B.E. and M.E. in chemistry from the University of Tokyo in 1996 and 1998. She joined NTT Information and Communication Systems Laboratories in 1998. Her research focuses on natural language processing and spoken dialogue systems. She is a member of NLP and IPSJ.

COTOHA™: Artificial Intelligence that Creates the Future by Actualizing Natural Japanese Conversation

Natsumi Furutaka and Masahiro Hamada

Abstract

NTT Communications began offering the communication engine "COTOHA™" (COTOHA) in October 2016. This is a service in which interactive conversations are conducted with customers through the use of artificial intelligence (AI) technology. In the hundreds of contacts from customers since this service was launched, we have been able to understand the high expectations customers have of AI, as well as other issues that have come to the surface. Here, we introduce COTOHA and the related business opportunities and market trends we have observed since launching the service.

Keywords: artificial intelligence, natural language processing, communications

1. Introduction

The communication engine "COTOHA™" (COTOHA) is a service that achieves interactive dialogue naturally. It was created by combining Japanese language processing technology developed by NTT Media Intelligence Laboratories (MD Labs) and an inference engine developed by IPsoft[*1]. The concept of COTOHA is a human-like artificial intelligence (AI) that replaces humans in providing certain services (**Fig. 1**). Our aim was to develop an AI that can act and work like a human instead of creating an AI that simply does things humans cannot do. COTOHA can converse naturally and is also able to process simple tasks such as creating orders or making appointments.

Certain problems have become apparent in interactive AI systems developed so far. Most of them are not proficient at understanding the Japanese language, are only able to execute conversations based on the scenario, consume a huge amount of time for scenario creation, and also take a lot of time when they require tuning. However, COTOHA overcomes these problems and achieves more flexible dialogue with efficient functions in construction and operation (**Fig. 2**).

Another important feature of COTOHA is that it can solve problems that other AI systems cannot by automatically escalating the inquiry to a human operator. Therefore, it is possible to improve the outcome of a response, which cannot be achieved by either a human or AI alone, through cooperation between them and by handling what each excels at. Moreover, with COTOHA, it is possible to extend the process range when needed. After the escalation to an operator, COTOHA will accumulate real conversation data between the end user and the operator.

COTOHA can also analyze, classify, and accumulate the customer's dialogue during the sessions, understand the context, and utilize it while having a conversation with a customer in real time. In the future, we will use this powerful Japanese language analyzer to automatically analyze the knowledge

*1 IPsoft: AI technology and information technology automation management service provider.

COTOHA is a communication engine that provides appropriate responses and conducts business processing like a real human by analyzing and understanding the context and meanings of multiple sentences.

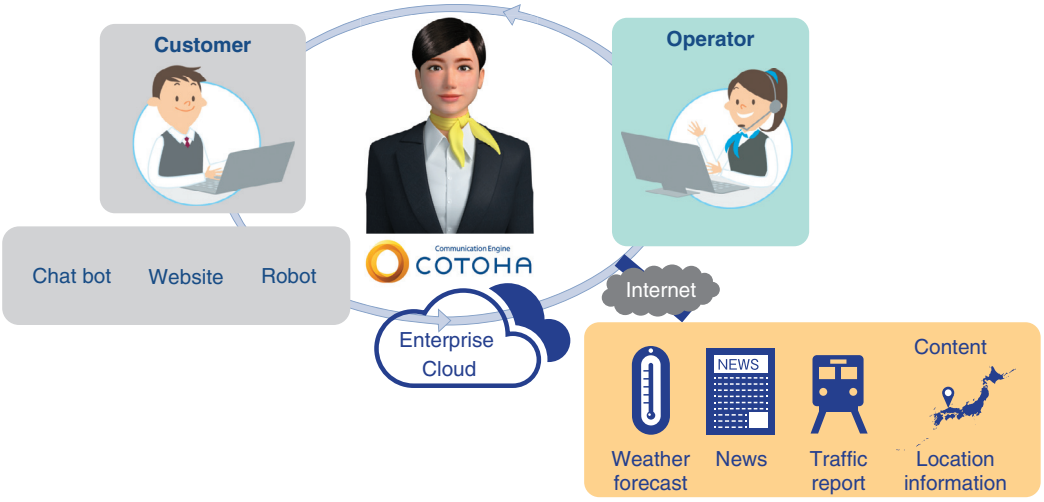


Fig. 1. What is COTOHA?

■ COTOHA's strengths are as follows:

	Previous AI	COTOHA
Accuracy	Understands keywords	1. Good understanding of Japanese • Analyzes the meaning and context • Memorizes customer data
Expression	Expression based on rules	2. Flexible communication • Understands customer's intention and asks automatically what customer doesn't mention • Answers customer's questions during conversation flow
Implementation	Necessary to set all scenarios	3. New setting style; goal matching • Adopts not only ordinary scenario settings but also new settings that don't require complex scenario setting capabilities
Maintenance	Requires a lot of time for tuning	4. Reuse escalation flow • If COTOHA can't answer a customer's question, COTOHA escalates it to an operator and memorizes the dialogue between the customer and operator.

Fig. 2. Strengths of COTOHA.

from manuals and textbooks. We also aim to completely automate the settings. Thus, the feature functions mentioned above are supported by our Japanese language processing technology.

2. Japanese language processing technology that supports COTOHA

In COTOHA, the Japanese language processing technology developed by MD Labs has been applied to understand user speech. We explain two features of the technology here: predicate-argument structure analysis and Japanese language natural sentence identity determination technology.

2.1 Predicate-argument structure analysis

The predicate-argument structure analyzer can identify the semantics of things and actions in a user's speech. For example, "Haha to yakiniku wo tabeta." (I ate Yakiniku with Mom.) and "Sarada to yakiniku wo tabeta." (I ate salad and yakiniku.). In the Japanese language, the subject is often omitted, so those sentences have no subject. However, we need to add a subject when translating them into English. From the first sentence, we know that "Haha" (Mom) and "sarada" (salad) have different roles and functions. The predicate-argument structure analysis technology can analyze the role of "taberu" (eat) as a basis predicate, "Haha" as a joint subject, and "sarada" as an object. With this technology, we can prevent incorrect utterances.

2.2 Japanese language natural sentence identity determination technology

Japanese language natural sentence identity determination technology can determine how similar two sentences are. It is mainly used in two situations in COTOHA. The first is in creating a conversation scenario. COTOHA is able to branch off scenarios according to the user's speech. However, since users tend to speak freely, it is necessary for COTOHA to determine whether or not the estimated user's speech that has been set in advance matches the actual user's speech. The other situation occurs with FAQ (frequently asked questions). By determining the similarity between a user's actual question and an anticipated question, COTOHA can identify the appropriate answer.

Both of these technologies use morphological analysis technology. Open source software has often been used recently to support this technology, but MD Labs has achieved high accuracy analysis by using a dictionary that is constantly updated by experts and a large-scale thesaurus[*2] of the Japanese lexical system that was modified independently by MD Labs.

3. Social issue in Japan and the hope for AI

COTOHA has a high-level ability to understand the Japanese language. It is also *close* to humans in that it is considerate of human feelings. At the same time, it can replace humans when performing various functions. It is therefore expected to be a solution for the declining labor force, which Japan will face in the near future. According to statistical data provided by the Ministry of Internal Affairs and Communications, 40% of Japan's population will be elderly people in 2060—44 years from now—and the working population will halve from its peak to 44.2 million [1]. Furthermore, according to the "New Industrial Structure Vision" issued by the Ministry of Economy, Trade and Industry [2], if this situation is not addressed, it is estimated that by the year 2030 the working population will decrease by 7.35 million people.

In contrast, if the labor force can be shifted to high-value added work by effectively utilizing AI and robots, the number is only expected to decrease by 1.61 million, and Japan's gross domestic product is estimated to increase to 222 trillion yen. Approximately 49% of the Japanese working population will be technically replaced by AI in the next 10–20 years [3]. This number is higher than 47% for the United States and 35% for the UK. The utilization of AI and robotics is expected to be an important factor in fostering the growth and prosperity of Japan by changing the way we work, replacing routine tasks, and increasing productivity.

However, AI is not *God-like* or almighty in its capabilities. In a narrow sense, areas replaced by AI will be different depending on the technology making up that AI. For instance, as introduced in the previous section on technology that supports COTOHA, in order to realize a chat bot, technology and/or functions such as deep learning, a similarity engine, a semantic analyzer, parsing, and various others need to be combined. However, there are only a few products that use all of the mentioned technologies. That is why customers must identify which method or technology they want to use in order to improve the

*2 Thesaurus: A reference work that lists words grouped together according to similarity of meaning (containing synonyms and sometimes antonyms).

efficiency of the current conversation and work process.

In a broad sense, it is impossible to leave the entire process to AI itself. There are still areas that humans are good at but that are difficult for AI, for example, the ability to sympathize with others, judge moral values, and understand irregular or unexpected scenarios. Thus, the task is not to replace everything with AI but identifying which areas need to be handled by AI, and which areas still need to be handled by humans. By creating synergies and approaching this task in a complex manner, we can achieve positive effects such as higher efficiency and improved customer satisfaction.

4. Opportunity for chat bot business

Some leading industry players such as Google, Apple, and Microsoft are also working to develop chat bots and conversational platforms. This has resulted in innovative new services for consumers such as messaging applications and SNS (social networking service) chat platforms, which are spreading as a new medium for information acquisition and communication interfaces with companies and services. The interface becomes a conversation, and everyone can easily use it; consequently, the added value of technologies and systems that were not widely used until now can finally come to light.

Such conversational exchanges can be a major factor in evaluating the value of a service. For instance, Domino's Pizza Japan launched an ordering system on LINE*3 (an SNS) using a chat bot, and they have steadily and significantly increased their sales through LINE, reaching 100 million yen in sales in 4 months and 200 million yen in 7 months [4].

Chat systems have become a third channel after telephone and the Internet, and their operation is becoming increasingly easier to use. This is expected to reduce the load on ordering systems—such as those used by Domino's and other companies—as the acceptance of this new technology increases. It is considered that services such as e-commerce, news searching, and customer support, in which users must be able to handle a certain degree of operability on a website, will soon be converted to chat bot systems in accordance with the necessity of the service.

Communication between chat bots and users can be a new hope for marketing utilization. Since one-way communication has mostly been used in online marketing so far, there have been only a handful of opportunities to get honest responses from users. Chat bots that can conduct conversations with hundreds or even thousands of users at once will be able to compile conversation logs. Although chat bots can only answer questions, the service provider will be able to obtain valuable data such as recordings of live voices that will convey information such as user reactions that service providers have not been able to estimate so far. This is expected to improve the quality of chat bot conversations and to improve the usage in various situations such as determining the potential needs of users and improving services.

Various tools are available to develop chat bots today, but the chat bots typically have limited capabilities. For example, they are confined to a certain fixed conversation, they exit the conversation unnaturally, they are only able to answer general questions, and most of them do not have any developed personality. To increase the utilization of chat bots and the number of users in the future, it is important to have a personalized chat bot that can answer complicated questions while drawing out the user's individual intentions. We will most likely require specialty chat bots for different fields, with another chat bot serving as a *hub* among them. The important elements to achieve this are the abilities to carry out natural conversation and to understand speech. As previously mentioned, we have steadily been improving the Japanese language accuracy and the inference function in chat bots.

To achieve a more sophisticated level of conversation and the necessary information and tools, we will continue to approach AI—which will continue to expand around COTOHA—by focusing on communication.

*3 LINE is a registered trademark of LINE Corporation.

References

[1] Ministry of Internal Affairs and Communications, "White Paper 2016: Information and Communications in Japan," Chapter 1: ICT-based Innovation and Economic Growth, 2016. http://www.soumu.go.jp/johotsusintokei/whitepaper/eng/WP2016/chapter-1.pdf#page=1

[2] Ministry of Economy, Trade, and Industry, "New Industrial Structure Vision," 2016 (in Japanese). http://www.meti.go.jp/committee/sankoushin/shin_sangyoukouzou/pdf/008_05_01.pdf

[3] Press release issued by Nomura Research Institute on December 2, 2015 (in Japanese). https://www.nri.com/jp/news/2015/151202_1.aspx

[4] S. Kinjo, "Chat Bot—Future Society Innovated by the Evolution of AI-based Chat Bots and Robots," Sotechsha, Tokyo, Japan, 2016 (in Japanese). (*Translation of the book title is provided for the readers' convenience.)

Natsumi Furutaka

Project Coordinator, Sales and Promotion, Artificial Intelligence Taskforce, Applications and Content, NTT Communications Corporation.

She received a B.E. in economics from Keio University, Kanagawa, in 2006. She joined NTT Communications in 2006, where she was engaged in network management and developing new businesses. She is currently involved in sales and promotion of COTOHA.

Masahiro Hamada

Engineer, Artificial Intelligence Taskforce, Application and Content, NTT Communications Corporation.

He received a B.E. and M.E. in systems engineering from the University of Electro-Communications, Tokyo, in 2004 and 2007. He joined NTT Advanced Technology Corporation in 2007 and worked on developing the backend of a web search service. He is currently involved in developing a Japanese analysis engine for COTOHA.

Anomaly Detection Technique in Sound to Detect Faulty Equipment

Hisashi Uematsu, Yuma Koizumi, Shoichiro Saito, Akira Nakagawa, and Noboru Harada

Abstract

The Internet of Things has become an increasingly active research field in recent years, and it is useful for collecting information from diverse sensors that can be analyzed to detect the operating status and anomalous behavior of equipment. We introduce here an anomaly detection technique in sound that can be used to detect anomalies in equipment by analyzing sounds picked up by microphones, even in environments where special sensors cannot be installed.

Keywords: anomaly detection in sound, deep learning, noise reduction, acoustic features

1. Introduction

Suppose you use a washing machine every day, but it starts making an unusual rattling noise and then stops working. Or suppose your refrigerator starts making a strange groaning sound and then breaks down after a few weeks. Most of us have probably experienced situations such as these at some point. Maintenance procedures are often triggered when equipment starts making unusual sounds. This applies not only to household appliances but also to commercial devices such as manufacturing equipment and building air conditioning systems. In recent years, services have started to emerge whereby various sensors are used to monitor the functioning of equipment and to detect any anomalies instead of relying on workers to perform this task. This article introduces a system that can automatically determine if equipment is operating normally or has an anomaly based on the sounds it makes.

2. Difficulties in automatically detecting anomalies in sounds

It can be hard to detect anomalies from the sounds made by equipment because it is difficult to collect a large volume of anomalous sounds (i.e., sounds made when equipment is operating abnormally). Remarkable progress has recently been made with machine learning techniques such as deep learning. These techniques make it possible to learn discrimination rules as to whether the operating state is normal or anomalous by training deep neural networks (DNNs) using a huge amount of training sound data obtained from equipment operating in normal and anomalous states. However, the frequency of equipment failure in real environments is very low, and the number of ways in which equipment can fail is also very large. Thus, it is not feasible to collect a sufficient amount of training sound data corresponding to anomalous operating states. It has therefore been difficult to apply these approaches to anomaly detection in sound.

Moreover, if it does become possible to detect anomalous states based on the sounds made by equipment, then this technology is likely to be introduced in factories, where there are many other types of equipment operating alongside the equipment being tested for anomalies. All the sounds emitted by the additional equipment will constitute background noise that interferes with the collection of sounds from the target equipment. In such environments, the background noise can drown out the sounds made by the operation of the target equipment. Therefore, it is essential to reduce this background noise in order to use sound as a means of detecting equipment operating

anomalously.

In this article, we present an overview of how we overcame the two above issues, and we introduce some practical examples of anomaly detection in sound in factories and other noisy environments.

3. Anomaly detection technique in sound using only normal operating sounds as training data

It is difficult to apply conventional machine learning approaches to detect anomalies using sound, as mentioned above. Furthermore, depending on factors such as the environment where the target equipment is installed and its modes of failure, it is possible that a variety of different anomalous sounds may occur even for essentially the same anomalous operating states. In the method involving learning discrimination rules, all of these anomalous sounds have to be collected and learned, making this approach increasingly impractical.

Therefore, we used a method that only requires normal operating sounds instead of having to collect sounds from anomalous states. This method is achieved by calculating the *normalness* of acoustic features. The normalness is determined by measuring the deviation from the normal state, and the state of the machine is identified as having an anomaly when this deviation exceeds a predefined threshold. We have applied this concept to develop an acoustic feature extraction method using DNNs to increase the normalness of normal operating sounds [1, 2].

In this method, discrimination is carried out by only deciding whether or not the acoustic features of the sound to be judged are normal, which is advantageous in that it does not depend on the type of anomalous sound. This method only determines whether or not a sound is the same as a normal sound. It does not detect anomalies in a strict sense but only detects when the sound being generated is not normal. However, if it is possible to detect anomalous sounds, then it should be possible to link with systems to subsequently deal with the anomaly.

4. Noise reduction technique

It is assumed that anomaly detection in sound will be introduced in factories and other such environments. In these environments, there are many other machines running in addition to the equipment to be checked for normal/anomalous operation, so the operating sounds of the target equipment are mixed with noise consisting of the operating sounds of other equipment coming from all directions. An algorithm that detects anomalies by learning only the sounds made during normal operation can be used when the operating sounds of the equipment to be checked have been collected.

In theory, if the waveforms of the sounds generated by all noise sources are subtracted from the sound entering the collecting microphone, then only the operating sound of the target equipment will remain. This can be achieved by obtaining information including the transfer characteristics from the noise sources to the collecting microphone. However, since there are many noise sources, each with transfer characteristics that vary according to diverse factors including the shape of the room and the distance from each noise source to the microphone, it is difficult to accurately calculate the transfer characteristics of each noise source.

Rather than handling many noise sources one by one, we addressed this issue by first creating a model whereby multiple noise sources are treated as a single noise source group. That is, instead of obtaining the transfer characteristics from individual noise sources, we approximated them with a single transfer characteristic. This approximation holds because we can regard each noise source as being much further away from the sound collection microphone than the target equipment. Next, instead of strictly calculating these transfer characteristics, we approximate them in terms of the time delay and transfer gains. In this way, transfer characteristics that would previously have been calculated strictly can be modeled and estimated using only the information needed for noise reduction, thereby facilitating the noise reduction and the detection of anomalous sounds in noisy environments.

5. Experimental detection of anomalous sounds in real environments

We used the method introduced above to conduct an experiment on three machines—an air blower pump, three-dimensional (3D) printer, and water pump—to determine whether or not anomalous sounds could be detected in a real environment.

(1) Air blower pump

An overall view of the air blower pump used in the experiment is shown in **Fig. 1**, together with the microphone mounting position and a close-up view of the pump. The microphone is installed by attaching it to a pole adjacent to the pump. First, we performed preliminary learning of normal sounds by using the

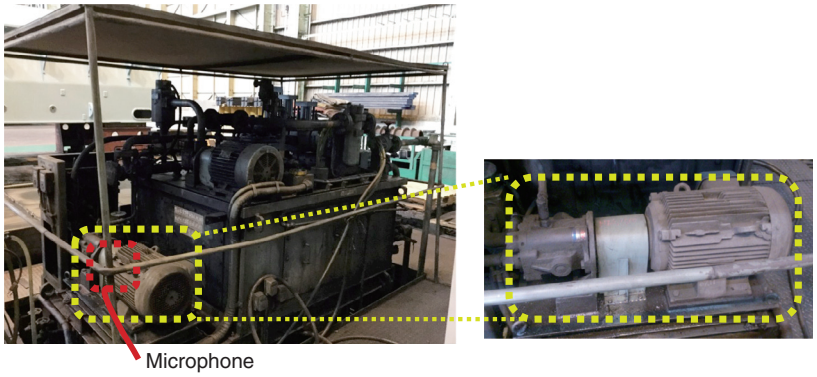


Fig. 1. Air blower pump.

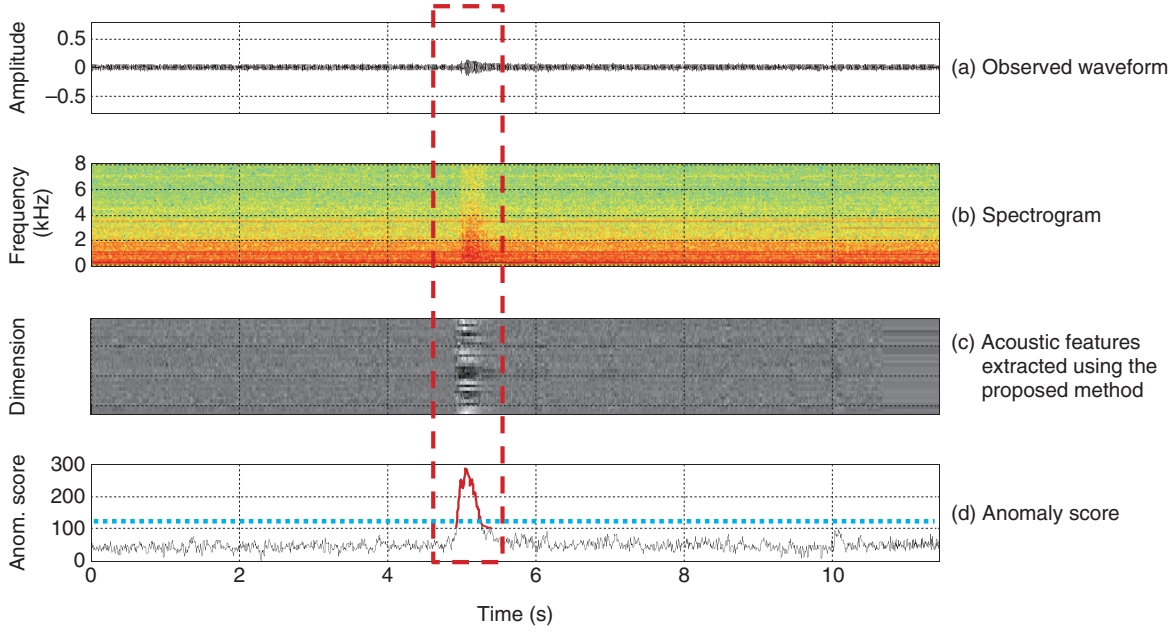
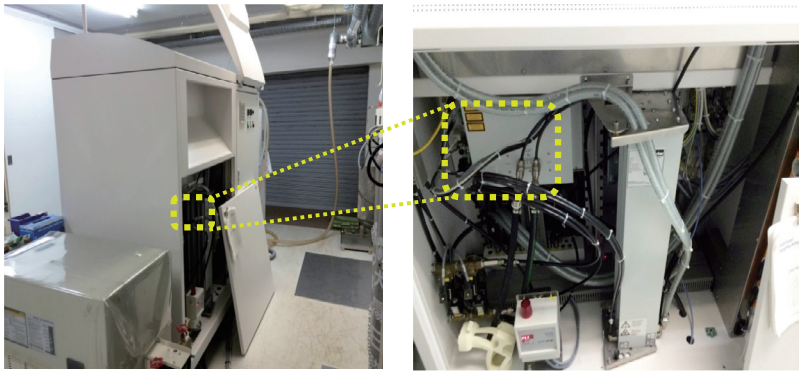


Fig. 2. Air blower pump anomalous sound detection results.

normal operating sound of the air blower pump for 20 minutes.

The results of this experiment are shown in **Fig. 2**. The observed waveform (Fig. 2(a)) and spectrogram (Fig. 2(b)) clearly show that there are different waveforms at around the 5-second point. This is known to be the noise produced when a blockage occurred due to a foreign object stuck in the air blower duct. A characteristic change can also be seen in the acoustic features (Fig. 2(c)) corresponding to the period when this blockage occurred. The anomaly score (Fig. 2(d)) indicates that it was possible to detect the presence of an anomalous sound due to the foreign object blockage.

Although in this experiment we were able to ascertain the cause of the anomalous sound through constant manual observation of the air blower pump's operation, it would be impractical to constantly monitor all pumps in a real environment. Therefore, as in this method, we can expect that if it is possible to detect an anomalous state automatically, then this would yield various benefits in terms of operability



(a) Overall view (b) Microphone arrangement

Fig. 3. 3D printer.

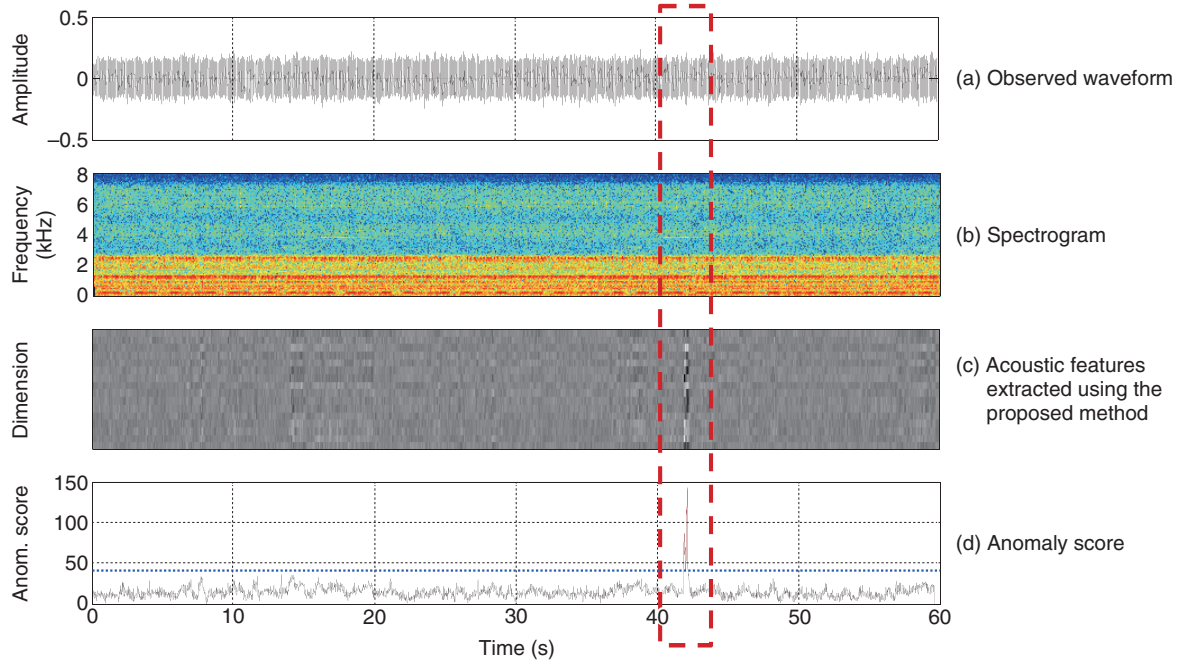


Fig. 4. 3D printer anomalous sound detection results.

and cost.

(2) 3D printer

An optical fabrication 3D printer was used in the experiment (**Fig. 3**). The microphone was placed inside the body of the 3D printer (the part outlined in yellow in Fig. 3) to record the operating sounds. We used 30 minutes of normal operating sounds to train the system with normal sounds.

The experimental results are shown in **Fig. 4**. The results in this experiment differed from those of the air blower pump in that the sound's observed waveform (Fig. 4(a)) and spectrogram (Fig. 4(b)) did not exhibit any changes over the entire displayed period, so it was not possible to distinguish any anomalous states from this information. However, in the acoustic features extracted using the method proposed here (Fig. 4(c)) and the anomaly score calculated based on these acoustic features (Fig. 4(d)), we observed anomaly changes at around 43 seconds (the part outlined in red). This means that a sound that does not

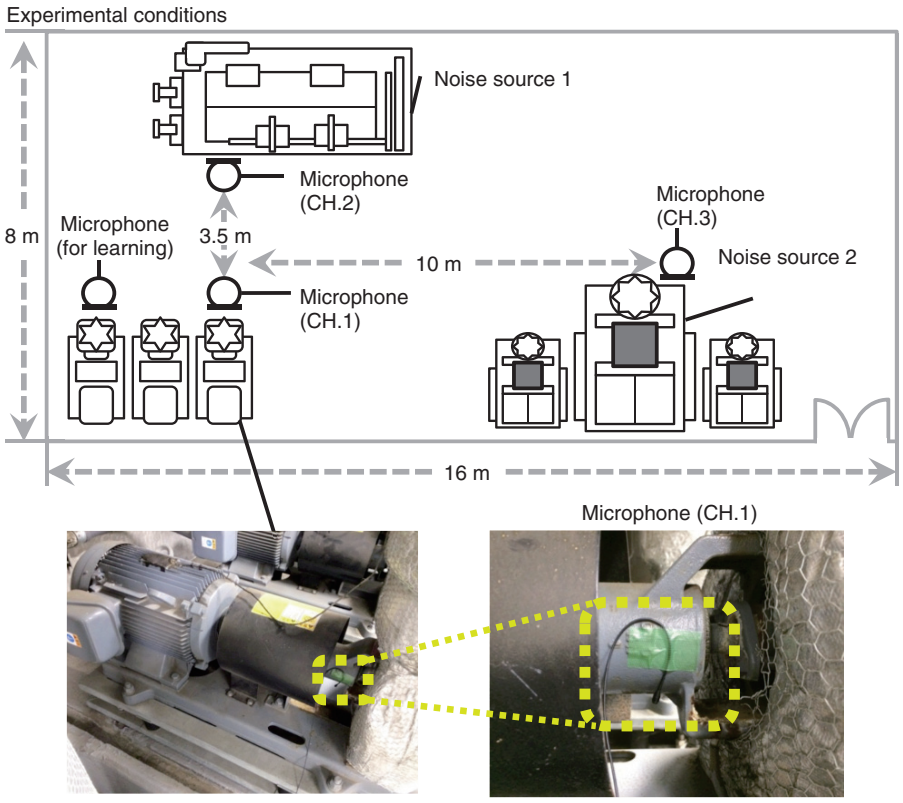


Fig. 5. Water pump and microphone arrangement.

occur during normal operation was observed at around 43 seconds, suggesting that some sort of anomaly had occurred. In fact, the 3D printer stopped unexpectedly about 5 minutes later. Furthermore, from the results of other observations, we found that an unusual action was performed at around 43 seconds whereby the sweeper collided with the model under construction. As in the example of this result, even when the existence of anomalous sound has not been clearly found in the analysis of data such as sound waveforms and spectrograms, this proposed method can be used to clarify the existence of anomalous sounds.

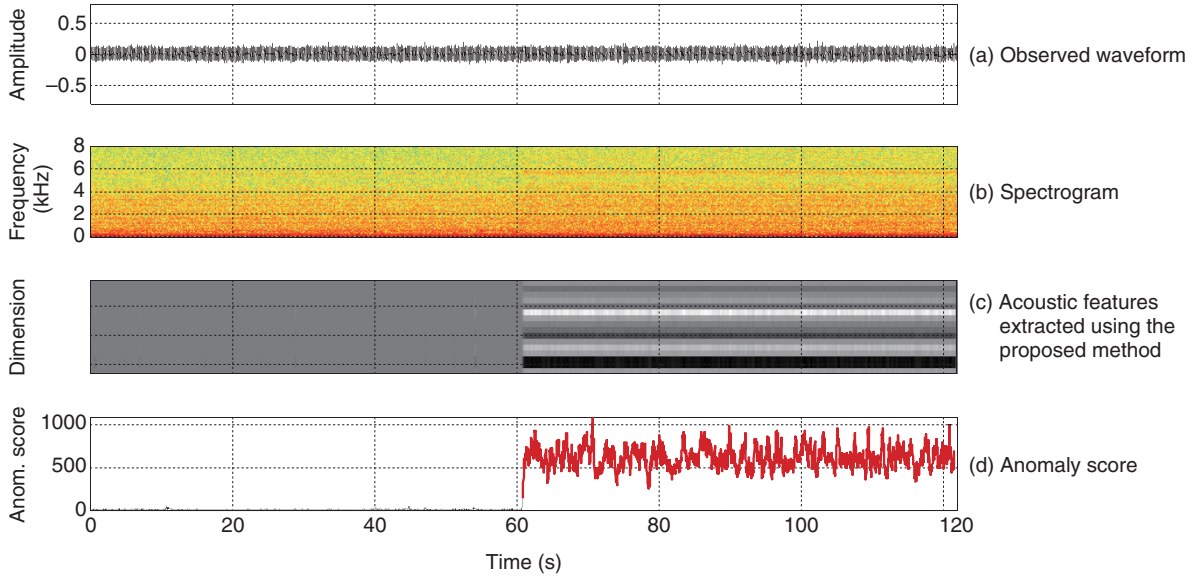
(3) Water pump

Finally, we present an example where this method was applied to a pump used to supply water to a building. In the machine room where this water pump is installed, noise is produced by a variety of other machines in the vicinity, and the operating sound of the target water pump is masked by the noise in this environment. Therefore, in this experiment we arranged the microphone as shown in **Fig. 5**.

In the examples of the air blower pump and 3D printer discussed above, a sudden anomalous sound was produced in the middle of normal operating sounds. With this water pump, however, anomalous sounds were continuously produced due to wearing of the pump's bearings. Thus, for the sound data for learning normalness, we used the sounds of other water pumps of the same type that were installed next to the target pump and that were operating normally.

The experimental results are shown in **Fig. 6**. In this figure, the sounds of the water pump in normal operating states for 60 seconds are shown in the first half, and in the second half these are compared with 60 seconds of sounds of the water pump in an anomalous state. In the observed waveform (Fig. 6(a)) and spectrogram (Fig. 6(b)), no major difference can be seen between the normal operating state (first half) and the anomalous operating state (second half), so from this information alone, it is not possible to distinguish between normal and anomalous states.

In contrast, there is a clear difference between the tendencies of the first half and second half with regard to the acoustic features (Fig. 6(c)) and the anomaly score (Fig. 6(d)), which clearly indicates



The first 60 seconds are the operating sound of the water pump during normal operation, and the last 60 seconds are the operating sound of the water pump during anomalous operation.

Fig. 6. Water pump anomalous sound detection results.

that the operating states are different. As these results show, our technique can be used to detect not only anomalous sounds that occur suddenly, as in the air blower pump and 3D printer, but also anomalous sounds that are produced continuously.

6. Future work

In principle, the anomaly detection technique in sound introduced above can detect anomalies automatically if an equipment produces sounds that appear to be anomalous when heard by naïve listeners in an environment with hindrances such as noise. In the future, we hope to develop algorithms that can even detect anomalous sounds that would be difficult for naïve listeners to detect but could be detected by a well-trained inspection engineer with decades of experience in this field.

References

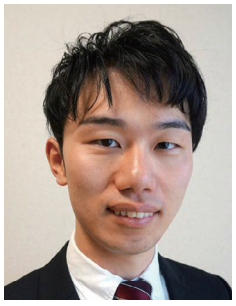
[1] Y. Koizumi, S. Saito, H. Uematsu, and N. Harada, "Optimizing Acoustic Feature Extractor for Anomalous Sound Detection Based on Neyman-Pearson Lemma," to appear in Proc. of the 25th European Signal Processing Conference (EUSIPCO 2017), Kos Island, Greece, Aug./Sept. 2017.

[2] Y. Koizumi, S. Saito, and H. Uematsu, "Anomalous Sound Detection for Machine Operating Sounds using Deep Neural Networks," The 2017 Spring Meeting of the Acoustical Society of Japan, pp. 473–476, Kawasaki, Japan, Mar. 2017.

Hisashi Uematsu

Senior Research Engineer, Cross Modal Computing Project, NTT Media Intelligence Laboratories.

He received a B.E., M.E., and Ph.D. in information science from Tohoku University, Miyagi, in 1991, 1993, and 1996. He joined NTT in 1996 and has been engaged in research on psychoacoustics (human auditory mechanisms) and digital signal processing. He is a member of the Acoustical Society of Japan (ASJ).



Yuma Koizumi

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

He received a B.S. and M.S. from Hosei University, Tokyo, in 2012 and 2014. Since joining NTT in 2014, he has been researching acoustic signal processing and machine learning. He was awarded the IPSJ Yamashita SIG Research Award from the Information Processing Society of Japan (IPSJ) in 2014 and the Awaya Prize from ASJ in 2017. He is a member of the Institute of Electrical and Electronics Engineers (IEEE), ASJ, and the Institute of Electronics, Information and Communication Engineers (IEICE).

Shoichiro Saito

Research Engineer, Audio, Speech, and Language Media Laboratory, 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 IEEE, IEICE, and ASJ.

Akira Nakagawa

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

He received a B.E. and M.E. from Kyushu Institute of Technology, Fukuoka, in 1992 and 1994. He joined NTT in 1994 and since then has been conducting research on acoustic signal processing for echo cancellation. He received a paper award from ASJ in 2002. He is a member of ASJ.

Noboru Harada

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

He received a B.S. and M.S. in computer science and systems engineering from Kyushu Institute of Technology, Fukuoka, in 1995 and 1997, and a Ph.D. in computer science from the University of Tsukuba in 2017. He joined NTT in 1997. His main research areas are lossless audio coding, high-efficiency coding of speech and audio, and their applications. He is an editor of ISO/ IEC 23000-6:2009 Professional Archival Application Format, ISO/IEC 14496-5:2001/ Amd.10:2007 reference software for MPEG-4 ALS, and ITU-T G.711.0, and has contributed to 3GPP EVS standardization. He is a member of IEICE, ASJ, IPSJ, the Audio Engineering Society (AES), and IEEE.

Spatio-temporal Activity Recognition Technology to Achieve Proactive Navigation

Hiroyuki Toda, Shuhei Yamamoto, and Takuya Nishimura

Abstract

We are developing spatio-temporal data analysis technologies to achieve proactive navigation that supports user activities by taking changes in the user's situation and in the surrounding environment into account. We introduce here spatio-temporal activity recognition technologies, which are key elements of spatio-temporal data analysis. The recognition technologies understand the user's activities from spatio-temporal movement data (moving trajectories, acceleration of movement, etc.) recorded by mobile devices.

Keywords: spatial trajectories, spatio-temporal data, data mining

1. Introduction

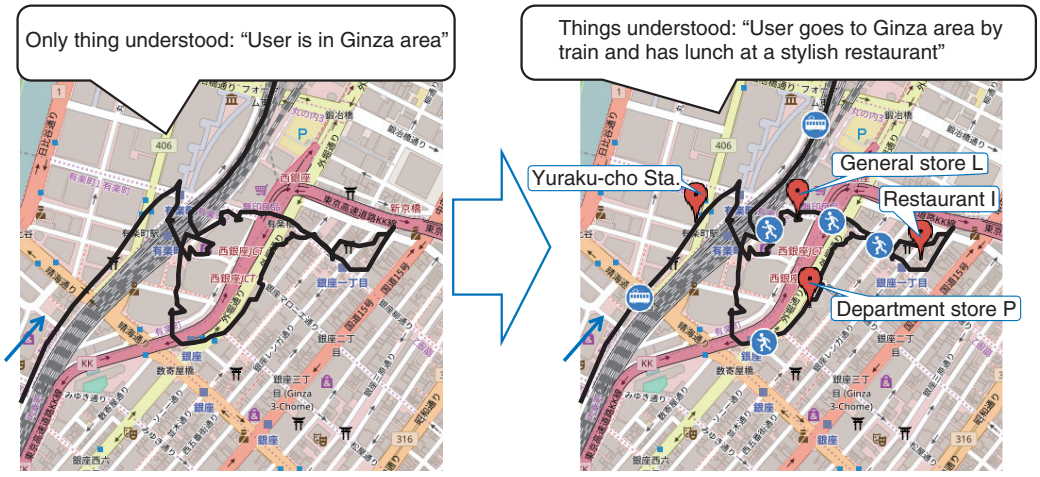
The word *navigation* suggests the idea of a routing assistance service such as a car navigation system that assists users to get from a starting location to a destination. The goal of our proactive navigation includes but is not restricted to routing assistance. For example, it provides users with suggestions for that day's dinner plans on the basis of their past activity histories, and it provides information to alert them of locations where traffic accidents or near misses frequently occur. In summary, the goal of our proactive navigation is to provide useful information for supporting users' spatio-temporal activities on the basis of past and present situations, as well as future situations in which changes may occur with respect to users and their surroundings [1].

2. Spatio-temporal activity recognition technologies

There has been a surge in popularity recently for mobile devices that can record people's activities using sensors. For example, about 1.36 billion smartphones with GPS (global positioning system) sensors were shipped around the world in 2016, a good indication that the market for these devices is continuing to grow [2]. Various other devices that use sensors are also coming into widespread use. These include wearable devices such as smart watches or smart bands that can sense vital data and motion data, and dashboard cameras (dashcams) that can record driving situations. These devices make it possible to record a wide variety of people's activities.

However, what these devices record is basically time-series numerical data. For example, moving trajectories are the sequence of the combination of timestamp, latitude, and longitude. When we draw a moving trajectory based on the latitude and longitude on a map, people who know the area can understand the outline of the moving behavior, but those who do not know the area understand only that the person went to that area (**Fig. 1(a)**).

We have developed spatio-temporal activity recognition technologies to tackle this problem. Using the technologies to analyze moving trajectories enables us to understand not only the areas that the users visited, but also the shops or places they visited [3] and



(a) Moving trajectory

(b) Analysis results in recognizing spatio-temporal activity

Fig. 1. Examples of moving trajectory and spatio-temporal activity recognition results.

the transportation mode they used to get there (**Fig. 1(b)**) [4]. This knowledge makes it possible for us to select suitable information for users.

In the following section we describe two state-of-the-art techniques employed in our spatio-temporal activity recognition technologies. One is a destination prediction technique that can predict the destination of a moving user. The other is a travel condition prediction technique that can detect travel conditions such as slow-moving traffic jams or near miss incidents while driving, through the use of time-series videos and sensor data sequences such as dashcam data.

3. Destination prediction technique

This method consists of two phases: a learning phase and a prediction phase. In the learning phase, a user model is constructed that memorizes a user’s movement trends from the user’s past moving trajectories. In the prediction phase, the destination candidates are predicted using the learned model and the user’s movement data from the original location to the current one (**Fig. 2**) [5].

Two requirements need to be simultaneously satisfied to achieve accurate prediction. One is that the long-term dependency of the movement from an original location to the current one must be taken into account (1). The other is that the data sparsity problem must be factored in (2). However, although the two requirements are related, a simple solution for requirement (1) does not satisfy requirement (2).

To satisfy both requirements, we propose the use of a recurrent neural network (RNN)[*1] for predicting destinations. An RNN is a neural network for modeling sequence data. Specifically, we regard moving trajectories as transitions on a grid space (**Fig. 3(a)**), and the transitions are modeled by the RNN. The RNN model we utilize can use long-term dependency to achieve accurate prediction when the current transitions match the past transition data, and can predict destinations on the basis of only the latest transitions even when the current transitions from the start location to the current one do not match the past transition data. Thus, this method satisfies the two requirements.

However, this RNN based model has a huge computational cost when we use it for prediction. This is because in the simplest terms, it predicts one transition with one calculation. It therefore needs G^M times calculation when the number of grids is G (several hundred or more), and the number of steps to the destination is M (several dozen or more). These are not realistic numbers for calculation purposes. To solve this problem, we use a sampling simulation

*1 RNN: A type of neural network that mainly uses modeling of sequence data.

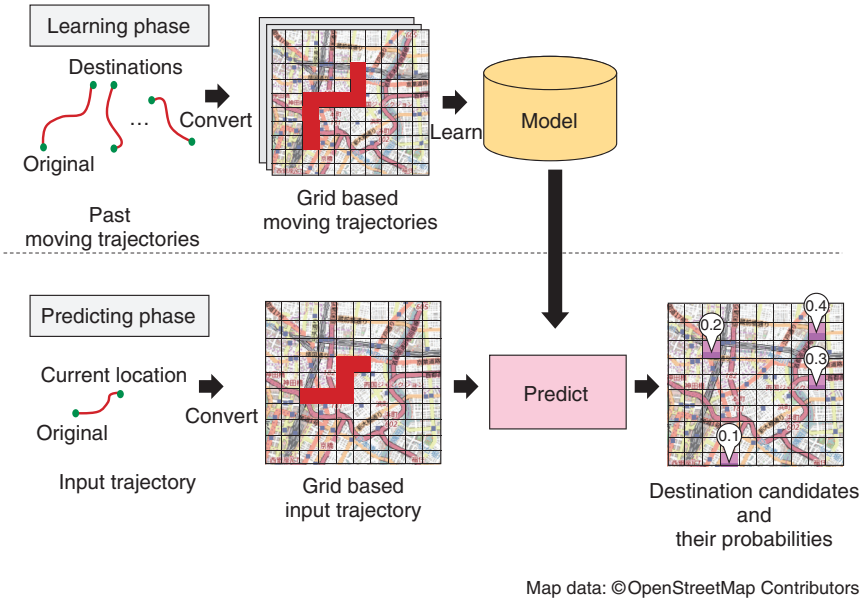


Fig. 2. Outline of destination prediction technique.

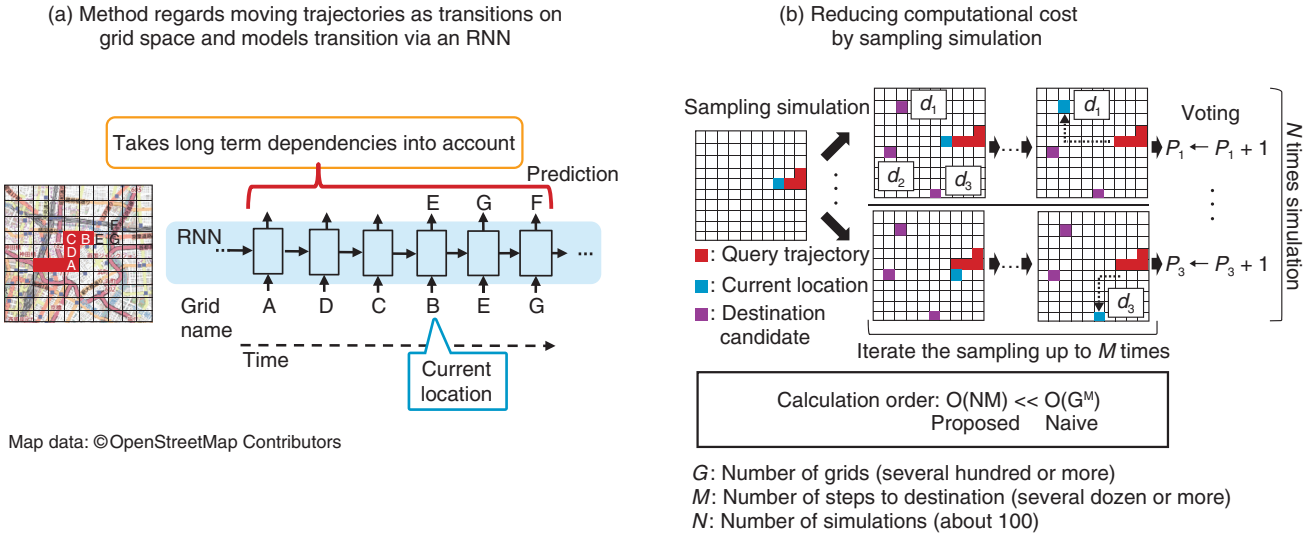


Fig. 3. Technical point of destination prediction technique.

(**Fig. 3(b)**) to reduce the number of calculations needed. The results we obtained confirmed that our method can predict destinations in about one second by using a conventional personal computer server environment.

When this technique is used to analyze the personal trajectories of smartphone users, it predicts the users' future movements and provides the users with useful information about future destination candidates. When it is used in a car navigation system, it provides traffic information about the route to the destination, and information about alternative routes on the basis of the traffic information even if the destination is not set.

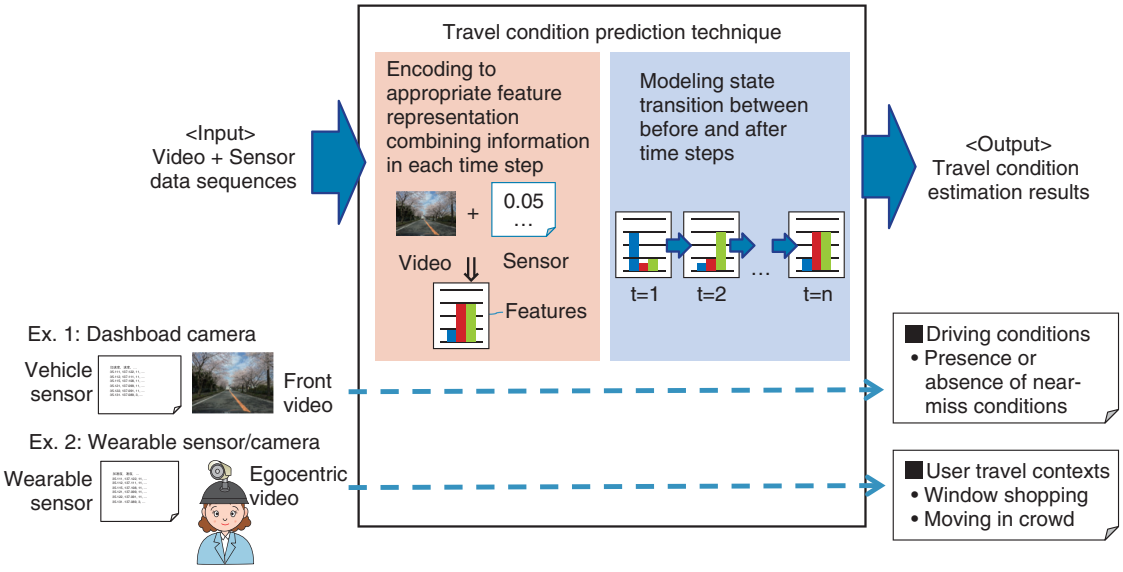


Fig. 4. Overview of travel condition technique.

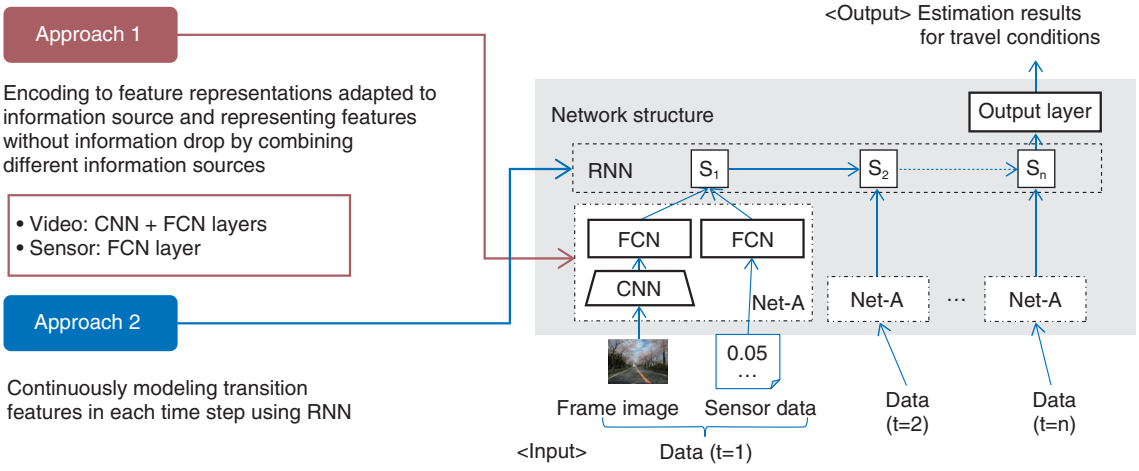


Fig. 5. Neural network structure of travel condition prediction and our approaches.

4. Travel condition prediction technique

Our travel condition prediction technique estimates the moving situations of people from multimodal time-series data recorded by both video images and sensors. For instance, it automatically detects the presence or absence of near-miss scenes from a dashboard video camera in a vehicle and identifies specific user contexts such as walking in a crowd or window shopping from a wearable camera and wearable sensor (**Fig. 4**).

Two difficulties arise when using such data to accurately estimate travel conditions: extracting feature representations for travel conditions from different information sources and modeling state transitions that have time variations.

We solved these problems by using a novel neural network with our travel condition prediction technique. The novel neural network is based on an RNN that can treat state transitions of time-series data. An overview of our proposed model is shown in **Fig. 5**. First, video images and sensor data, which are input

in each time step, are encoded to feature representations by different components of the neural network. A frame image in each time step obtained from video acquires the appropriate feature representation via a fully connected network (FCN) layer and a convolutional neural network (CNN)[*2] layer, which are frequently used in image analysis. Also, sensor data are encoded to suitable feature representations using the FCN layer. Feature representations are extracted by concatenating a feature representation of a frame image and that of sensor data in each time step. Modeling these features using the RNN in correct time sequences makes it possible to take state transitions with time variations into account.

In our experimental evaluation, a near-miss scene was identified by a dashboard video camera, and our travel condition prediction technique detected dangerous driving scenes with higher accuracy than that obtained with three baseline methods: without video images, without sensor data, or without state transitions with time variations, that is, using an RNN. Moreover, we confirmed that the detection accuracy was approximately 90% [6].

NTT Communications and Nippon Car Solutions are considering using our technique as a way to reduce traffic accidents [7]. We are also considering another way to use it, namely by combining location information with actual near-miss cases collected from many vehicle dashboard video cameras. This will enable us to make a *hiyari-hat* (near-miss accidents) map that can pinpoint dangerous places where near-miss situations frequently occur. Various other applications can be considered such as using wearable cameras and sensors to understand specific user contexts and to accurately identify sports scenes from videos recorded within a stadium and sensors fitted for each player.

5. Future work

We plan to collaborate with partner companies to achieve practical use of these technologies. However, we face several challenges in attempting to improve them. For example, the current destination prediction technique cannot predict places that the users have never been to, but we expect it to be able to do so in a practical manner in the future. The current travel condition prediction technique can detect the presence or absence of near misses but cannot accurately understand the types of near misses that have occurred. One future task will be to improve the accuracy and coverage range of the existing technologies.

We should also note the fact that various sensors can record wide-ranging activities through the development of the IoT (Internet of Things) environment. For this environment, we will need to develop our technologies so that they can accurately understand users' spatio-temporal behavior through the use of various sensors and in various situations. To make the proactive navigation truly effective, we must develop our technologies by utilizing understandings of past events to predictions, from known areas to unknown areas, and from simple behavior to diverse behavior.

References

[1] J. Ikedo, T. Horioka, Y. Niikura, Y. Koike, H. Sawada, and Y. Muto, "Proactive Navigation Optimized for Individual Users," NTT Technical Review, Vol. 13, No. 7, 2015.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201507fa4.html

[2] Press release issued by TrendForce on January 25, 2017, "TrendForce Reports Global Smartphone Production Volume Totaled 1.36 Billion Units; Samsung Held On as Leader While OPPO and Vivo Burst into Global Top Five."
http://press.trendforce.com/node/view/2741.html

[3] K. Nishida, H. Toda, T. Kurashima, and Y. Suhara, "Probabilistic Identification of Visited Point-of-interest for Personalized Automatic Check-in," Proc. of UbiComp2014 (the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing), pp. 631–642, Seattle, WA, USA, Sept. 2014.

[4] Y. Endo, H. Toda, K. Nishida, and A. Kawanobe, "Deep Feature Extraction from Trajectories for Transportation Mode Estimation," In Advances in Knowledge Discovery and Data Mining, Proc. of the Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD) 2016, pp. 54–66, Auckland, New Zealand, Apr. 2016.

[5] Y. Endo, K. Nishida, H. Toda, and H. Sawada, "Predicting Destinations from Partial Trajectories Using Recurrent Neural Network," In Advances in Knowledge Discovery and Data Mining, Proc. of PAKDD 2017, pp. 160–172, Jeju, South Korea, May 2017.

[6] K. Ito, "AI Auto-detects Dangerous Driving," Image Laboratory, Vol. 28, No. 2, pp. 45–47, 2017 (in Japanese).

[7] Press release issued by NTT Communications on September 26, 2016 (in Japanese).
http://www.ntt.com/about-us/press-releases/news/article/2016/20160926_2.html

*2 CNN: A type of neural network inspired by the organization of the visual cortex of humans. It is mainly used for image recognition.

Hiroyuki Toda

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

He received a B.E. and M.E. in materials science from Nagoya University in 1997 and 1999, and a Ph.D. in computer science from University of Tsukuba in 2007. He joined NTT in 1999. His current research interests include information retrieval, data mining, and ubiquitous computing. He is a member of the Information Processing Society of Japan (IPSJ), the Institute of Electronics, Information and Communication Engineers (IEICE), the Japanese Society for Artificial Intelligence (JSAI), the Database Society of Japan (DBSJ), and the Association for Computing Machinery (ACM).

Shuhei Yamamoto

Researcher, Proactive Navigation Project, NTT Service Evolution Laboratories.

He received a B.L.S., M.I., and Ph.D. in informatics from University of Tsukuba in 2012, 2014, and 2016. He joined NTT Service Evolution Laboratories in 2016. He is currently studying data mining and analysis of real spatio-temporal data obtained from devices such as dashboard camera sensors and wearable sensors. He is a member of IPSJ and DBSJ.

Takuya Nishimura

Researcher, Proactive Navigation Project, NTT Service Evolution Laboratories.

He received a B.E. and M.E in social informatics from Kyoto University in 2012 and 2014. He joined NTT Service Evolution Laboratories in 2014. He is currently studying data mining from geographic information systems, social media, and human trajectories. He is a member of DBSJ.

Predicting Patients' Treatment Behavior by Medical Data Analysis Using Machine Learning Technique

Hisashi Kurasawa, Akinori Fujino, and Katsuyoshi Hayashi

Abstract

The analysis of medical information including electronic health records and medical image data using artificial intelligence (AI) has been an active area of research in recent years. The objective is to develop a means of supporting decision-making about treatment and medicine prescription by medical doctors. NTT has also been designing an AI for supporting diabetes treatment in collaboration with the University of Tokyo Hospital. We introduce here our machine-learning based technique to predict missed scheduled clinical appointments, which are likely to trigger treatment discontinuation by patients with diabetes.

Keywords: diabetes, treatment discontinuation, machine learning

1. Treatment discontinuation in diabetes treatment

The number of diabetes patients has been gradually increasing throughout the world. There are now over 3 million diabetic patients in Japan and approximately 380 million worldwide. Diabetes progression causes complicating diseases including diabetes retinopathy and nephropathy and results in a reduction in the quality of life and an increase in medical costs. Thus, it is vital for diabetic patients to begin medical treatment early and keep their regular hospital appointments to control their blood glucose level. However, in Japan, about 10% of diabetic patients discontinue treatment and return to the hospital after diabetes progression. This is one of the major problems in diabetic care (**Fig. 1(a)**).

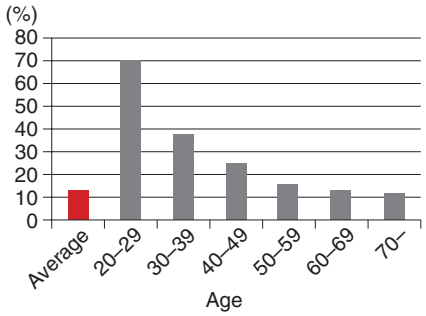
Researchers working to solve this problem have been studying the factors related to treatment discontinuation and have classified diabetic patients into two groups: those continuing and those discontinuing treatment. Multifactorial data were used to identify the differences between the two groups as risk factors by testing their statistical significance. Factors such as being *male* or *holding a job* proved to be high risks for treatment discontinuation [1] (**Fig. 1(b)**).

There are various factors related to treatment discontinuation, and it is therefore quite difficult for diabetes specialists to identify the patients who need a higher level of support. Therefore, NTT has been collaborating with the University of Tokyo Hospital in designing a prediction model that can predict treatment discontinuation at an individual level (**Fig. 2**). We designed the model to have features generated by a machine-learning technique that NTT developed and by applying knowledge of medical data analysis and the clinical experience of staff at the University of Tokyo Hospital [2].

2. Design of prediction model of treatment discontinuation

We focused our efforts on designing a model to predict a missed clinical appointment (MA) that can be a trigger for treatment discontinuation in order to identify the diabetic patients who require stronger

(a) Treatment discontinuation ratio of diabetes patients by age bracket



(b) Examples of patient characteristics related to treatment discontinuation

- Males; people holding jobs
- Young people (under 50 years of age, especially those 20–30 years old)
- Patients whose blood glucose level is well-controlled/poorly controlled
- Patients who have discontinued treatment in the past

Source: Heisei 24-year National Health and Nutrition Survey (in Japanese).
http://www.mhlw.go.jp/bunya/kenkou/eiyou/dl/h24-houkoku-06.pdf

Fig. 1. Statistics related to treatment discontinuation by diabetes patients.

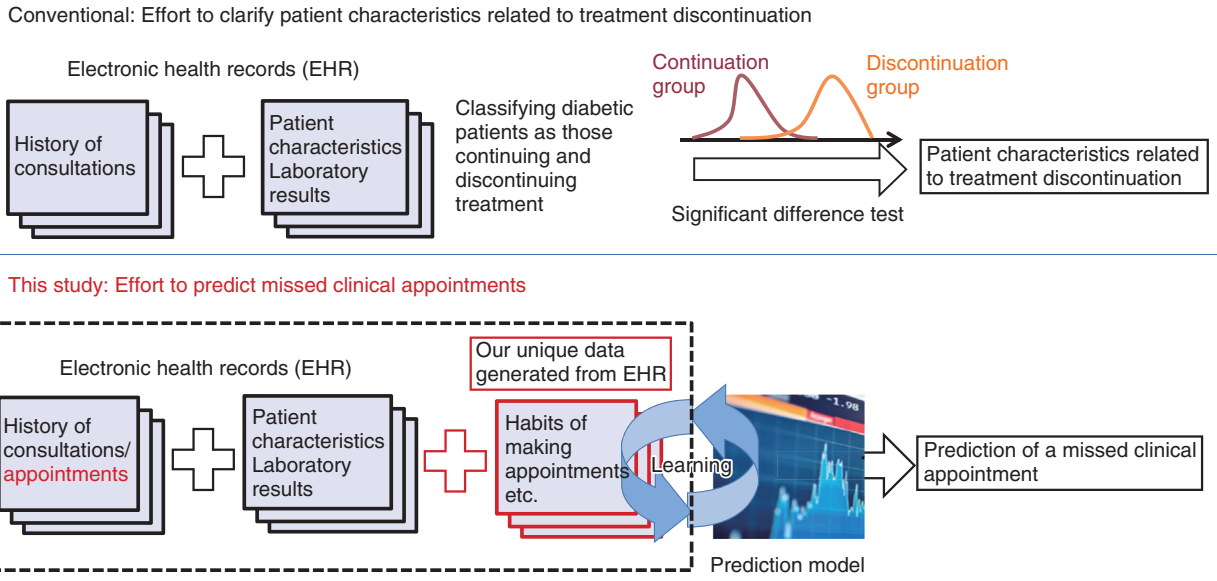


Fig. 2. Prediction model to identify the likelihood of treatment discontinuation by diabetes patients.

support. We designed a logistic regression model that predicts a class y from a feature vector x of a patient's target appointment, with y representing a clinical appointment missed ($y = +1$) or kept ($y = -1$) and x generated from the patient's electronic health record (EHR), representing the time from the initial visit to just before the target appointment. We modeled the probability of an appointment with x attributed to y with a logistic regression,

$$P(y \mid \mathbf{x}; \mathbf{w}) = \frac{1}{1 + \exp(-y\mathbf{w} \cdot \mathbf{x})}.$$

We computed the w estimate by using training data $\{(x1, y1), ..., (xn, yn), ..., (xN, yN)\}$. In the feature design explained below, observed values of each quantitative variable, for example, blood test results, were linearly transformed (normalized) to make a variance of each variable = 1 and assigned to vector x.

When the training data size N is smaller than the dimension of feature vectors, or a training data sampling is biased, maximum-likelihood estimation often overfits a logistic regression model into the training

Table 1. Examples of X1 group.

Category	Features
Consultation	Past/recent consultation of medical departments
Disease	Past/recent diagnosed diseases; recovered from/under treatment
Medicine	Past/recent prescribed medicines, prescription term, number of doses taken per day

Table 2. Examples of X2 group.

Category	Features
Property	Gender, age Distance and time from home to hospital by public transport
Consultation	Frequency of clinic visits Probability of visit on a given day of the week
Appointment	Interval between the date on which a clinical appointment was made and the scheduled appointment date; Day of the week when an appointment was made; Whether the patient kept an appointment on the day it was made or missed the appointment without notice; Weather on the appointment day
Medicine	Length of the prescription term of the latest medicine Total amount of medication a day How many times a day medication is taken Maximum size of prescribed tablets
Laboratory results	HbA1c, HDL-C, LDL-C, TG, and T-Cho

data, leading the model to classify many appointments inaccurately. Thus, we used an L2-norm regularization method to mitigate overfitting and improve the model's generalizability.

3. Feature design

To obtain a prediction model with higher accuracy, we designed features using the knowledge of diabetes specialists, patients' opinions, and the results of our behavior analysis. We designed two types of features: one that was automatically generated from EHR (X1 group, n = 29,025), and another that could be related to treatment discontinuation.

Examples of the X1 group are listed in **Table 1**. These features concern the diagnosis and treatment department that each patient visited before the clinic appointment day, the name of their disease and the prescribed medication, the length of the prescription term (number of weeks, months, etc.), surgical procedures, and exams.

Examples of the X2 group are listed in **Table 2**. We designed these features based on patients' behavior. Thus, the behavior characteristics and environmental factors of each patient weight the MA. For example, we included the day of the week and the weather on the day of a clinical appointment, and the pill size of the prescribed medication. We also referred to interviews with specialists and patients. A diabetes specialist explained that with patients who often forget to take their medication, the specialist monitored compliance by comparing the intervals between scheduled clinical appointments, the length of the prescription term, and the amount of medication the patient still had, so we incorporated the calculation as a check on compliance.

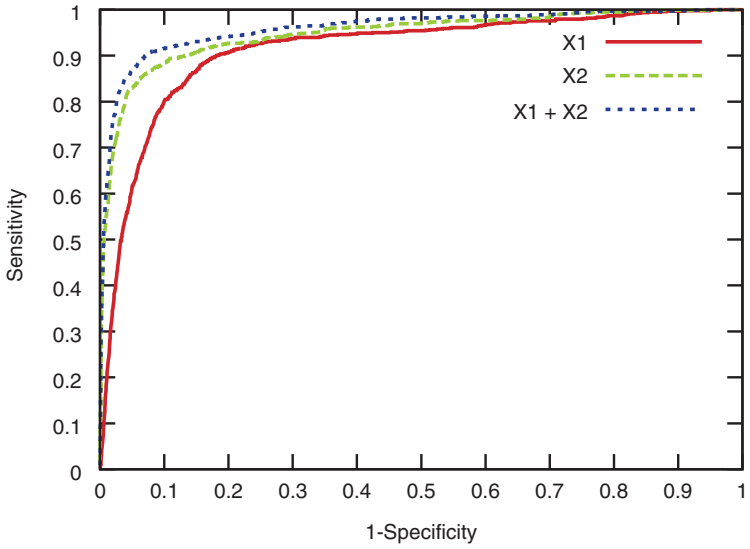
A diabetic patient told us that the accumulated experience with medical treatment in general stimulated his motivation to continue diabetes treatments. Thus, we used the history of clinic visits and the number of diagnosed diseases. We referred to databases maintained by the Japan Meteorological Agency and the Pharmaceuticals and Medical Devices Agency to respectively find weather information and the size of prescribed medication. In addition, we used gender, age, distance from home to hospital, and travel time, which previous research had identified as factors related to treatment discontinuation.

4. Evaluation of prediction performance

All prediction experiments were performed using records from the University of Tokyo Hospital, which included the history of 16,026 clinical appointments

Table 3. Actual results based on patients' behavior in making appointments.

Item description	Number (number of patients)
New appointment registration	32,962 (879)
Appointment schedule change	4,680 (619)
Medical consultation with an appointment (continuation group)	15,104 (869)
Medical consultation without an appointment	228 (116)
Missed appointment but patient had other appointment in the diabetes department on a later day	493 (232)



The larger the area under the ROC curve (AUC value) (i.e., the closer the curve is to the upper left), the higher the prediction performance (0.5 in a completely random case).

Fig. 3. ROC curve measuring prediction performance of logistic regression model with L2-norm regularization by using the X1, X2, and X1 + X2 groups.

scheduled by 879 patients whose initial clinical visit had been made after January 1, 2004, who had diagnostic codes indicative of diabetes, and whose HbA1c had been tested within three months after their initial visit. The records were dated between April 1, 2011 and June 30, 2014. We used the data to predict MAs in appointments. The actual number of missed appointments and kept appointments was 922 and 15,104, as indicated in **Table 3**.

We examined the difference in prediction performance between models by using the X1, X2, and X1 + X2 groups with a receiver operating curve (ROC) (**Fig. 3**). The model compared the area under the curve (AUC) values between them. An AUC value of 0.943 was obtained for the X2 group, leading to a prediction of MAs with higher accuracy than for the X1 group (AUC = 0.912). The AUC for both the X1 and X2 groups (AUC = 0.958) yielded even greater accuracy, as indicated in **Table 4**. An AUC value of more than 0.9 generally means excellent discrimination in such performance evaluations. We can therefore say that our model can predict MAs with high accuracy and help to identify the patients needing

Table 4. AUC value, precision, recall, and F-measure.

Features	AUC	Precision	Recall	F-measure
X1	0.912	0.466	0.438	0.449
X2	0.943	0.750	0.577	0.652
X1 + X2	0.958	0.757	0.659	0.704

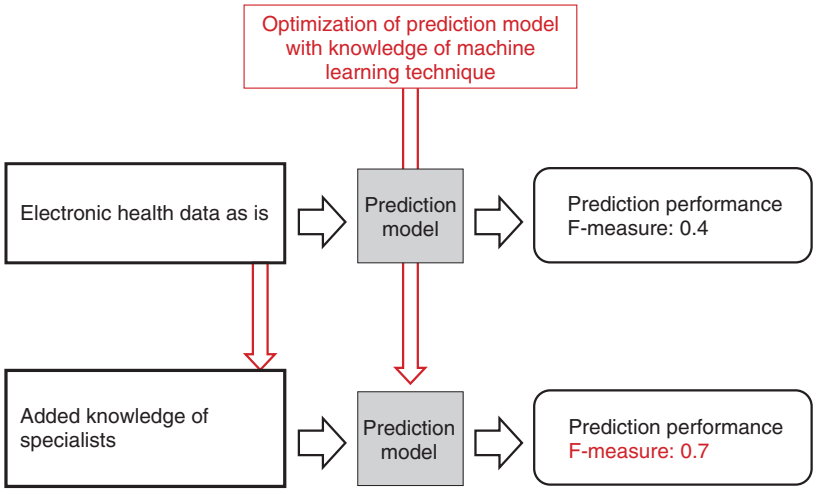


Fig. 4. Relationship between input data and prediction performance.

Table 5. Features contributing to MA prediction.

Feature	Weight
Appointment was made on a Sunday	2.22
Appointment was scheduled for a Friday	1.91
Interval between when a clinical appointment was made and the date for which it was scheduled	–0.912
Graves' disease was diagnosed and recovered from	–1.36
Probability the previous appointment was made on a Monday	0.897
Rilmazafone hydrochloride hydrate recently prescribed	1.29

stronger support to maintain treatment; it also helps to determine when doctors should support them. The precision, recall, and F-measure of the model when using both X1 and X2 groups were 0.757, 0.659, and 0.704, respectively, although the F-measure was 0.449 when using only the X1 group, as indicated in Table 4. We can also say that we succeeded in obtaining a model with high prediction performance by designing features related to treatment behaviors. The relationship between input data and prediction performance is shown in **Fig. 4**.

5. Features contributing to MA prediction

We compared the weights of features used in this study to investigate what the contributive features were for MA prediction. The contributive features are listed in **Table 5**. These features were given the largest absolute weight in our trained model and thus contributed strongly to MA prediction. We found that features related to when and how appointments were made, rather than to patients' clinical condition, influenced the accuracy in predicting MAs. Some contributive features were, for example, "Appointment was made on Sunday," "Appointment was scheduled on Friday," and "Interval between when a clinical appointment was made and the date for which it was scheduled." Of these, the last characteristic amount is a negative value, which means that the shorter the interval between when a clinical appointment was made and the date for which it was scheduled, the more likely a missed clinical appointment will occur. These features found by our model were valued by a diabetes specialist because it is generally difficult for medical doctors to find them.

We believe that our model will be a powerful tool to improve patients' medical condition since it makes it possible to identify patients who need more support, to assess when they should be strongly supported, and to control the strength of intervention with each

patient.

6. Future development

We designed a model that can predict with high accuracy when a diabetic patient is likely to miss a scheduled hospital appointment, possibly leading to discontinuation of treatment. This is the first study utilizing machine learning to design this kind of model.

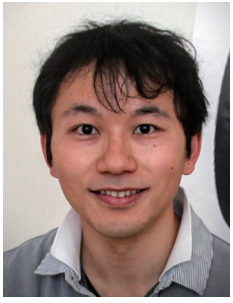
Various techniques that can provide information appropriate to humans have been proposed as recommendation techniques in the machine learning field. We plan to continue studying machine learning in order to develop solutions to prevent treatment discontinuation. One of our collaborators who has been studying telemedicine demonstrated that patients' behavior related to their meals can be improved if they are given feedback on their lifestyle habits based on the data [3]. Thus, we believe that giving patients feedback on the confirmation of appointments will be one of the solutions to motivate patients who may potentially discontinue treatment. The standards for EHR and SS-MIX2 (Standardized Structured Medical Information Exchange) are now being widely implemented, and an environment for handling massive amounts of clinical data is being prepared, so we will focus our efforts on improving our prediction model while utilizing these advances.

References

[1] M. Noda, et al., "Diabetes Consultation Discontinuation Countermeasure Comprehensive Guide," 2014 (in Japanese).

[2] H. Kurasawa, K. Hayashi, A. Fujino, K. Takasugi, T. Haga, K. Waki, T. Noguchi, and K. Ohe, "Machine-learning-based Prediction of a Missed Scheduled Clinical Appointment by Patients with Diabetes," J. Diabetes Sci. Technol., Vol. 10, No. 3, pp. 730–736, 2016.

[3] K. Waki, H. Fujita, Y. Uchimura, K. Omae, E. Aramaki, S. Kato, H. Lee, H. Kobayashi, T. Kadowaki, and K. Ohe, "DialBetics: A Novel Smartphone-based Self-management Support System for Type 2 Diabetes Patients," J. Diabetes Sci. Technol., Vol. 8, No. 2, pp. 209–215, 2014.



Hisashi Kurasawa

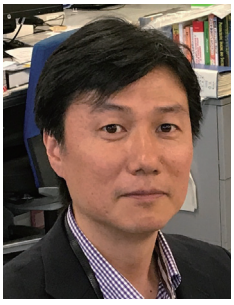
Research Engineer, Ubiquitous Service Systems Laboratory, NTT Network Innovation Laboratories.

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 is studying prediction models of clinical condition, including treatment discontinuation prediction. He is also currently studying biomedical signal analysis at NTT Basic Research 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.

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, NTT Device Technology Laboratories.

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 and began studying electrochemical biosensing devices to detect biomolecules. 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 studying prediction models of clinical conditions, including treatment discontinuation prediction. He is also currently studying surface plasmon resonance-based microfluidic sensing devices to detect biochemical reaction and biomarkers for point-of-care testing at NTT Device Technology Laboratories. He is a member of the Chemical Society of Japan, the Electrochemical Society of Japan, and the Japan Association of Chemical Sensors.

Polarization Switch of Carbon Nanotubes

Ken-ichi Sasaki

Abstract

Polarized light is used in liquid crystal display televisions and other products and is therefore familiar in our daily lives. It is also closely related to optical communications. In this article, we explain a theoretical prediction that a polarizer composed of carbon nanotubes exhibits a novel phenomenon of switching between two polarization components.

Keywords: polarizer, carbon nanotubes, plasmons

1. Introduction

When we write a straight line on a blank sheet and look at it through a transparent mineral called calcite, the single straight line appears as two parallel lines to the eye. This phenomenon is an easy way to understand that light consists of two polarization components. When those two parallel lines are viewed through a polarizer (a linear polarized filter), there is an angle of the polarizer in which only one straight line is visible. If the angle of the polarizer is rotated a further 90°, the straight line that was visible is not visible, and the straight line that was not visible becomes visible. In this way, we can also see that the two components of polarized light are orthogonal, so we call the two polarization components corresponding to the XY orthogonal coordinate axes as X-polarization and Y-polarization components. These two polarized components are closely related to our daily lives, although we do not have much consciousness of them. For example, we can see that with a polarizer the image on a liquid crystal display television is polarized.

Interestingly, polarization is also closely related to advanced technology in modern optical communications. The reason is that if two polarization components are used, it is possible to double the amount of information to be transmitted at the same time in optical communications. Different information containing voice and video data can be transferred to the two polarized components and propagated at the speed of light.

2. Basic science of polarized light

Thought experiments using polarized light have resulted in new knowledge concerning the laws of nature [1]. To cite one example, the well-known proposition in quantum information science that a general quantum-mechanical state cannot be copied was made based on the polarization of light [2]. According to quantum mechanics, light shows not only properties as waves (electromagnetic waves) but also behavior as particles. The light that we can see in our everyday life is propagated in a massive number of point particles called photons. Two components of polarized light correspond to two states that can be taken by one photon, and the state where the two X- and Y-polarization components are superimposed is a general polarization state. We can theoretically prove that this superimposed general polarization state cannot be cloned.

In light of this pioneering achievement, we surmise that the polarization possessed by classical electromagnetic waves is itself a degenerate quantum state of photons, and by applying these corresponding relationships, researchers like us—who are accustomed to classical mechanics—will be able to understand the strange mechanism of quantum mechanics represented in the superposition of states. In fact,

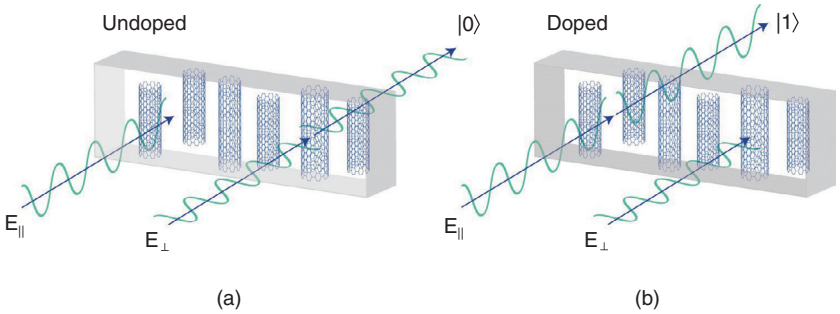


Fig. 1. (a) Undoped carbon nanotubes transmit only perpendicular polarized light. (b) Doped nanotubes transmit only parallel polarization. Since the polarization of the transmitted light rotates 90 degrees by doping, the aligned nanotubes function as a polarization switch.

well-known textbooks on quantum mechanics use polarization to introduce the subject of quantum mechanics [3].

Polarized light therefore has many interesting aspects in a wide range of fields from classical electromagnetism to quantum mechanics and is therefore of great interest to us. Devices (a polarization beam splitter and waveplate) currently exist that can rotate the direction of polarization, but there are no devices that can freely switch the two polarized components. We introduce here a theoretical prediction that such a polarization switch can be done with carbon nanotubes.

3. Polarizer of carbon nanotubes

Carbon nanotubes are cylindrical molecules with diameters as small as 1 nm and lengths up to several micrometers. They consist of carbon atoms and can be thought of as graphene sheets wrapped into cylinders. The diameter of a human hair is about 50 μm and is therefore approximately 50,000 times larger than the diameter of a nanotube. This comparison indicates just how thin carbon nanotubes are. By dispersing a large number of carbon nanotubes on an organic film and then pulling on the film, we can produce a film in which the axial direction of the carbon nanotube is oriented towards the pulling direction.

It has been confirmed that aligned carbon nanotubes function as a polarizer [4, 5]. That is, a carbon nanotube absorbs light whose linear polarization is parallel to the tube's axis, but not when the polarization is perpendicular to it.

Why does a carbon nanotube transmit only perpendicular polarized light? The electric field has a screening effect—called a depolarization effect—that plays a very important role. The depolarization effect occurs when an electric field is applied perpendicularly to the axis. Positive and negative charges are then polarized on the surface of the tube, which creates a new electric field. This newly generated electric field plus the added external electric field becomes the total electric field that the electron actually *sees*. Interestingly, the new electric field is almost the same as the added electric field in magnitude, but the sign is opposite, which cancels them out. As a result of this cancellation, the total electric field almost disappears. Because the absorption of light is proportional to the total electric field, the absorption is suppressed, and a nanotube is almost transparent against a perpendicular polarization.

4. Polarizer of doped carbon nanotubes

Suppose that the polarization component to be transmitted by a nanotube polarizer is X-polarization. To switch it to Y-polarization, we can rotate the nanotube film by 90° in the same manner as with a normal polarizer. We found theoretically that with carbon nanotubes, the direction of the transmitted light could be changed by 90° without any spatial rotation by adjusting the number of electrons in the nanotube through charge doping [6]. That is, a doped carbon nanotube transmits parallel polarized light and absorbs perpendicular polarized light, as illustrated in **Fig. 1**.

The absorption of perpendicular polarized light in a doped nanotube is a *many-body effect* that excites electrons in a collective manner, which is known as plasmon resonance. Doping increases the contribution of the depolarization field. Even if an infinitesimal

external electric field is applied, a finite electric field is generated. The plasmon excitation is essentially different from the single-particle excitation caused by parallel polarized light in an undoped nanotube.

In our study, we intend to provide the opportunity to investigate transitions between two polarized components by means of an electrical method (doping) as well as general-purpose optical measurements (namely, optical absorption). Accordingly, while the study will experimentally investigate the scientific potential of nanotubes through basic research, it will also revitalize applied research on optical devices.

5. Significance of results

To change the polarization direction of light transmitted through a typical polarizer, it is necessary to rotate the lens itself. However, on the basis of our theory of doping dependence, it is anticipated that a nanotube polarizer can reverse the polarization of the transmitted light by 90° without the need for such spatial rotation. The significance of the phenomenon we have discovered can be easily understood from the standpoint of optical transmission. In regard to cutting-edge optical-transmission technology, as represented by digital-coherent devices, the two degrees of freedom of polarization of light are utilized to double the amount of information that can be transmitted. Different kinds of information such as images and sound are transcribed to orthogonally polarized light and transmitted. A polarization switch based on carbon nanotubes can be used to manipulate information within a highly miniaturized structure for optical transmission.

References

[1] B. H. Billings (Ed.), "Selected Papers on Applications of Polarized Light," SPIE: Bellingham, WA, USA, 1992.

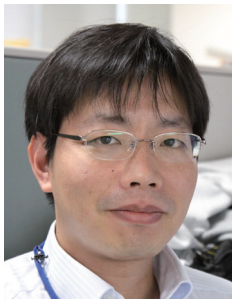
[2] W. K. Wootters and W. H. Zurek, "A Single Quantum Cannot Be Cloned," Nature, Vol. 299, pp. 802–803, 1982.

[3] G. Baym, "Lectures on Quantum Mechanics," W. A. Benjamin, New York, USA, 1969.

[4] S. Shoji, H. Suzuki, R. Proietti Zaccaria, Z. Sekkat, and S. Kawata, "Optical Polarizer Made of Uniaxially Aligned Short Single-wall Carbon Nanotubes Embedded in a Polymer Film," Phys. Rev. B, Vol. 77, 153407, 2008.

[5] B. G. Kang, Y. J. Lim, K.-U. Jeong, K. Lee, Y. H. Le, and S. H. Lee, "A Tunable Carbon Nanotube Polarizer," Nanotechnology, Vol. 21, 405202, 2010.

[6] K. Sasaki, S. Murakami, and H. Yamamoto, "Theory of Intraband Plasmons in Doped Carbon Nanotubes: Rolled Surface-plasmons of Graphene," Appl. Phys. Lett., Vol. 108, No. 16, 163109, 2016.



Ken-ichi Sasaki

Researcher, Low-Dimensional Nanomaterials Research Group, Materials Science Laboratories, NTT Basic Research Laboratories.

He received an M.S. and Ph.D. in physics from Tohoku University, Miyagi, in 1999 and 2003. He joined NTT Basic Research Laboratories in 2010 and has been studying theoretical aspects of graphene and carbon nanotubes. He is a member of the Physical Society of Japan and the American Physical Society.

Trends in Web-based Signage Standardization

Kiyoshi Tanaka, Mushin Nakamura, Kenya Suzuki, and Kei Takegami

Abstract

Web-based signage is expected to be the next generation of digital signage for presenting information and advertisements on displays. Its key feature is the ability to provide rich and compelling services at low cost using web technology. This article introduces trends in the standardization of web-based signage both in Japan and in international standardization bodies such as the W3C (World Wide Web Consortium) and ITU-T (International Telecommunication Union - Telecommunication Standardization Sector).

Keywords: web-based signage, digital signage, standardization, W3C, ITU-T, DSC

1. Introduction

Digital signage shows great promise as an information service for foreign visitors to Japan including at international sports events and as a social infrastructure for delivering information at the time of a natural disaster. The market for digital signage is expected to expand rapidly as the price of displays drops.

The Ministry of Internal Affairs and Communications (MIC) recently held the Conference on the Promotion of Introduction of ICT in the Entire Society in View of 2020. The Digital Signage Working Group was organized at the conference and is now studying the functions needed for implementing information and communication technology (ICT) throughout society via digital signage with a view to successfully staging the events in 2020 and achieving sustainable growth in Japan beyond 2020. It seeks to implement methods to achieve dissemination of disaster-related information, provide information via smartphones based on individual user attributes, and provide multilingual information via signage all with the aim of expanding digital signage functions and ensuring their interoperability [1].

Web-based signage is attracting attention as a means of achieving such a digital signage system. It is a form of digital signage using web technology. As long as an HTML5[*1]-compatible web browser is installed on the signage terminal, signage functions can be provided via a web application called a signage player. In other words, web-based signage can be configured with a general-purpose terminal without regard to the type of terminal or operating system being used, thereby lowering the cost of providing services (**Fig. 1**). It also enables content creation through web application programming interfaces (APIs) and abundant cloud assets such as open data. To achieve interoperability among digital signage systems with different specifications, it must be possible to deliver general content through a standard connection interface. An effective solution to this problem is a system configuration centered on web-based signage that applies widely used web technology.

NTT Service Evolution Laboratories has been researching and developing web-based signage. It aims to turn the results of its research and development (R&D) efforts into useful products and contribute to both domestic and international standardization activities to propagate web-based signage throughout society.

*1 HTML5: Hypertext Markup Language, fifth revision.

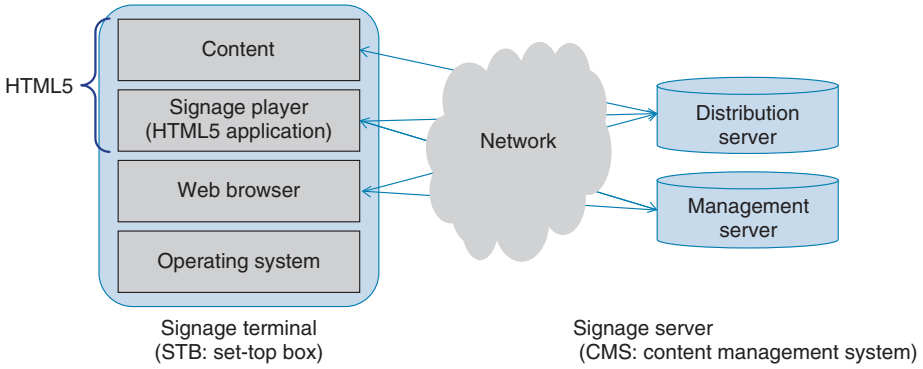


Fig. 1. Web-based signage.

2. Standardization in Japan

The motivating force behind the standardization of digital signage in Japan is the Digital Signage Consortium (DSC), whose core members include NTT Group companies. The DSC serves as a forum for all stakeholders in Japan's digital signage industry. It consists of 112 corporate members (as of April 4, 2017) including manufacturers, content providers, location owners, system integrators, and carriers. Its activities include conducting surveys and doing research in addition to carrying out its standardization efforts. The DSC has so far released "Digital Signage Standard System Guidelines" in 2008 and "Digital Signage Operation Guidelines at Times of Disasters and Emergencies" (Version 1: 2013; Version 2: 2014), which have contributed to international standards through an upstream[*2] process, becoming the basis for ITU-T Recommendations H.780 [2] and H.785.0 [3] published by the International Telecommunication Union - Telecommunication Standardization Sector (ITU-T) [4].

As part of this flow, DSC released "Digital Signage Standard System Interoperability Guidelines" in April 2016 with the aim of providing advanced digital signage functions such as dissemination of disaster information and information provision based on individual user attributes. The goal here is to promote the spread of digital signage with a view to 2020, and envisioning use by the increasing number of foreign visitors to Japan and by people with disabilities.

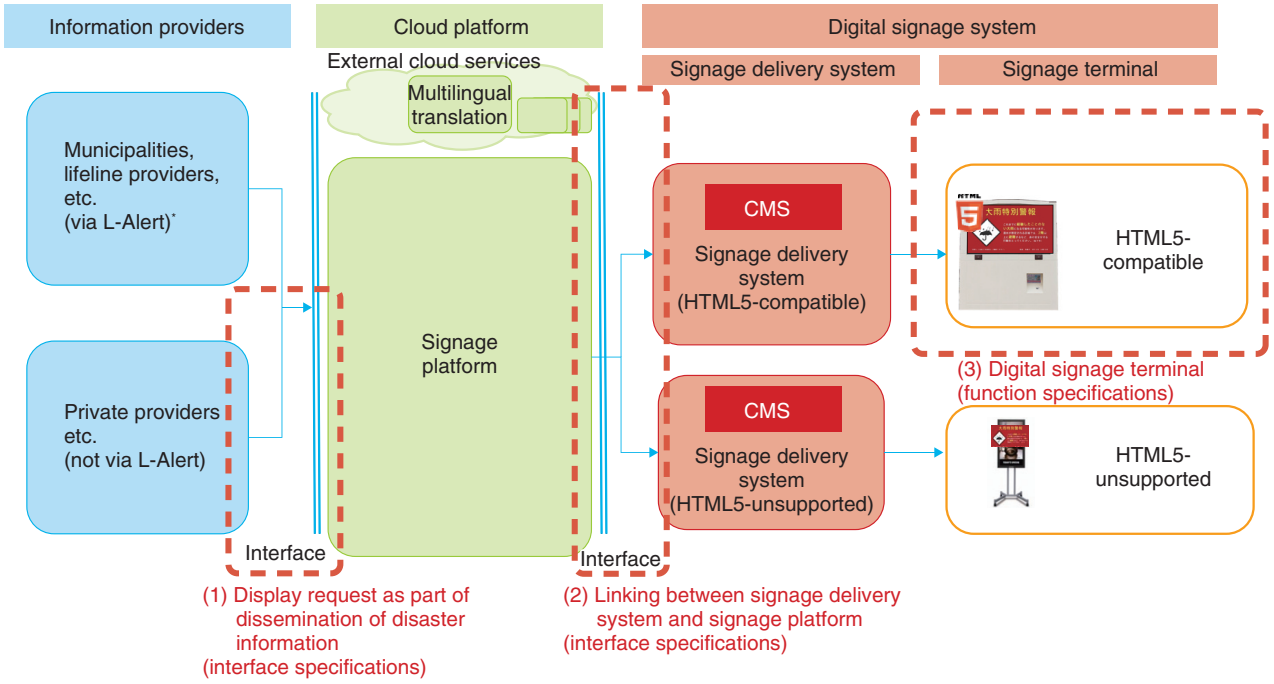
These guidelines present a system configuration that enables the provision of diverse services such as dissemination of urgent information including disaster information on signage, the linking of signage and smartphones to display signage information on smartphones, and the optimal delivery of information through multilingual support that switches the display based on personal attributes (preferred language) (**Fig. 2**). In short, a system configuration that includes a signage platform as a linking infrastructure plus the interface specifications for that platform are being turned into guidelines to achieve interoperability between digital signage systems having different specifications. These guidelines include terminal specifications for achieving functions using HTML5-compatible web-based signage.

3. Standardization at W3C

The World Wide Web Consortium (W3C) is an international, membership-based industry-academia consortium that aims to standardize and promote web technologies [5]. Standardization activities related to web-based signage began with studies on use cases and requirements in the Web-based Signage Business Group (BG) established in response to a proposal from Japan in 2012. The results of these studies were released as a BG report on use cases and requirements [6]. A web-based signage terminal features content presentation and terminal control through a web application called a signage player that runs on a web browser. The basic functions of this signage player can be achieved using existing HTML5 technology. Methods for implementing these functions have been released in the form of profiles (core profile [7], basic media profile [8], and storage profile [9]).

However, from the viewpoint of digital signage operation, APIs for controlling terminal hardware

*2 Upstream: In standardization activities, a proposal for making a domestic standard into an international standard.



* L-Alert: A disaster information sharing system in Japan.
Source: DSC "Digital Signage Standard System Interoperability Guidelines"

Fig. 2. System configuration for achieving envisioned services.

from the browser had been insufficient, and it was proposed that a Working Group (WG) for promoting the standardization of these APIs be established. A basic consensus on establishing this WG was reached in a BG meeting held at the W3C Technical Plenary/Advisory Committee Meetings Week 2015 (W3C TPAC 2015). A charter outlining the activities and mission of this WG was then prepared, and discussions commenced. Then, in 2016, these API discussions continued as face-to-face meetings at the September TPAC and became even more intense at face-to-face BG meetings held in Seoul in November, where the WG charter was enhanced and the preparation of documents related to web-based signage architecture began. At present, APIs targeted for standardization are those related to control of terminal power from the browser and to the acquisition of information necessary for terminal and content control (**Table 1**). The plan is to pursue standardization in earnest once the establishment of this WG is approved in W3C.

4. Standardization at ITU-T

As an agency of the United Nations, ITU-T produces ITU Recommendations that become international standards centered on telecommunications. The area of digital signage is addressed in Question 14 (Q14) by Study Group 16 (SG16), the group in charge of multimedia matters. Japan has been active in making contributions and proposals pertaining to Q14.

Recent activities include contributing an upstream standard from the Digital Signage Standard System Interoperability Guidelines produced by DSC and in addition, making proposals for establishing new work items relating to the topics. In this regard, the "Requirements of interoperable information services in public places" (H.DS-PISR) and "Framework for interactive service" (H.DS-FIS) were established as work items in June 2015 and September 2016, respectively, based on proposals from Japan. In addition, the draft technical paper "Technical Paper on Digital signage: Web-based digital signage" (HSTP.DS-WDS) for clarifying the framework and architecture of web-based signage is being studied as guidelines for configuring digital signage using web technology. The preparation of this paper is progressing based on a collaborative relationship with W3C.

Table 1. API proposals targeted by WG for recommendations.

API		Function	
Power status management		Automatic power saving of the terminal device	
	Power status control	Changing the power status mode	• Switching to stand-by mode • Display on/off
	Power status scheduling	Scheduled resume	• Resuming from stand-by
Contextual information		Information retrieval for player enrichment	
	Signage operational information	Acquiring unique information about the device	• Unique terminal information (serial number) • Manufacturer information (manufacturer name, brand name) • Terminal type
	Signage functional information	Retrieving functional information on the device	• Logical resolution • Display size

5. Verification trial

In line with the standardization trends described above, a verification trial using digital signage conforming to the Digital Signage Standard System Interoperability Guidelines was conducted from January to March 2017 in the Takeshiba redevelopment district in Minato ward, Tokyo [10]. This trial was conducted by the Contents innovation Program (CiP) together with Albero Grande (a joint venture of Tokyu Land Corporation and Kajima Corporation for developing Takeshiba district), NTT Communications, NTT IT (now NTT TechnoCross), and NTT as part of the MIC fiscal year 2016 project "Survey Contracts in Relation to Regional Trials toward IoT Hospitable Environments."

In the trial, research achievements of NTT Service Evolution Laboratories were used in the signage platform and web-based signage delivery system. The trial involved the construction of a signage platform capable of disseminating content to multiple signage systems, automatic acquisition of disaster-related municipal information and web-based information, conversion of that information to signage-compatible content, dissemination of that content to the signage delivery systems of various providers, and provision of a multilingual information service during a disaster. In addition, the web-based signage delivery system made it possible to display priority information rather than other information and to automatically display information on a smartphone in the user's default language.

6. Future outlook

In the near future, standards for web-based signage will become recommendations. The Digital Signage Standard System Interoperability Guidelines already include web-based signage and dissemination, which are also promoted by NTT. These guidelines were upgraded to Version 2 (2017) and reflect the results of the verification trial to further the standardization of web-based signage in Japan. In addition, the contents of these guidelines are expected to become an international standard in 2018, which means a standard that can be used both in Japan and throughout the world and that can contribute to the expansion of web-based signage business.

References

[1] Website of MIC, Conference on the Promotion of Introduction of ICT in the Entire Society in View of 2020 (in Japanese), http://www.soumu.go.jp/main_sosiki/kenkyu/2020_ict_kondankai/index.html

[2] Recommendation ITU-T H.780: "Digital Signage: Service Requirements and IPTV-based Architecture," Geneva, Switzerland, June 2012.

[3] Recommendation ITU-T H.785.0: "Digital Signage: Requirements of Disaster Information Services," Geneva, Switzerland, Oct. 2014.

[4] K. Tanaka, S. Matsuo, H. Ohmuro, and S. Sakaya, "ITU-T SG16 Sapporo Meeting Report," NTT Technical Review, Vol. 13, No. 2, 2015. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201502gls.html

[5] K. Tanaka, S. Fujimura, S. Fukatsu, T. Tokunaga, F. Hamamura, S. Homma, R. Aoki, and T. Matsumoto, "Standardization Activities at W3C TPAC 2015," NTT Technical Review, Vol. 14, No. 5, 2016. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201605gls.html

[6] Web-based Signage Business Group, "Web-based Signage Use Cases and Requirements," Nov. 2015. https://www.w3.org/2016/websigns/ucr/

[7] Web-based Signage Business Group, "Architecture and Requirements for Web-based Signage Player - Core Profile," Nov. 2013. https://www.w3.org/2016/websigns/core/

[8] Web-based Signage Business Group, "Architecture and Requirements for Web-based Signage Player - Basic Media Profile," Nov. 2015. https://www.w3.org/2016/websigns/basic-media/

[9] Web-based Signage Business Group, "Architecture and Requirements for Web-based Signage Player - Storage Profile," Nov. 2015. https://www.w3.org/2016/websigns/storage/

[10] Press release issued by NTT on January 16, 2017 (in Japanese).
http://www.ntt.co.jp/news2017/1701/170116a.html

Kiyoshi Tanaka
Senior Research Engineer, Supervisor, 2020 Epoch-making Project, NTT Service Evolution Laboratories.
He received a B.E., M.E., and Ph.D. in communication engineering from Osaka University in 1992, 1994, and 2005. He joined NTT in 1994 and since then has been researching video-on-demand systems and metadata-related interactive video systems and services, especially those related to IPTV and digital signage services. His current interests include standardization of digital signage. He is one of the co-chairs of the W3C Web-based Signage Business Group. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE), and the Human Interface Society in Japan.

Mushin Nakamura
Senior Research Engineer, 2020 Epoch-making Project, NTT Service Evolution Laboratories.
He received a B.S. in mathematics and an M.E. in computer science from the University of Tokyo in 2001 and 2003. He joined NTT in 2003. He has been engaged in research and development of NGN (Next Generation Network) appliances and web-based signage systems.

Kenya Suzuki
Senior Research Engineer, Supervisor, 2020 Epoch-making Project, NTT Service Evolution Laboratories.
He received a B.E. in electro-communications from the University of Electro-Communications, Tokyo, in 1993 and an M.E. in electronic engineering from the University of Tokyo in 1995. He joined NTT Human Interface Laboratories in 1995. He has been researching human-computer interaction and multimedia applications including web-based signage system development. He is a member of IEICE and the Physical Society of Japan.

Kei Takegami
Senior Manager, R&D Produce Group, NTT Research and Development Planning Department.
He received an M.E. in engineering from Osaka University in 1997 and an MBA from Waseda University, Tokyo, in 2007. He joined NTT in 1997, where he was initially involved in selling communication and networking solutions to large corporations. During 2000–2006, he developed marketing strategies by analyzing information technology market trends at Info-Com Research Inc. From 2006 to 2009, he took part in launching three new businesses at NTT WEST. After that, he was a manager in the corporate planning department and the business development department at a subsidiary of NTT WEST. He is currently responsible for new business development in the media industry.

Metallic Cable Fault Location Search Technology

Abstract

This article describes faults that occur in the core wire of metallic cables and the methods used to locate them. This is the forty-first article in a series on telecommunication technologies. This contribution is from the Access Engineering Group, Technical Assistance and Support Center, Maintenance and Service Operations Department, Network Business Headquarters, NTT EAST.

Keywords: metallic cable, open-circuit fault, insulation fault

1. Introduction

Massive amounts of metallic cables were installed outside NTT offices during the era when the number of fixed-line telephone subscribers was increasing. In recent years, however, faults in the core wire of these cables have begun to appear as a result of aging. The current approach at maintenance hubs is to conduct repairs whenever such a fault is discovered. However, although metallic cable facilities are expected to gradually decrease in number with the shift toward an optical-fiber cable network, it cannot be assumed that all metallic cables will soon be gone. To prevent circuit problems, it is important to quickly estimate the location of faults in metallic cables and deal promptly with those faults.

2. Metallic cable faults

Faults in metallic cables include cross (short-circuit) faults, insulation faults, and ground faults due to factors such as humidity or cable submersion in water (**Photo 1**). Such faults can cause circuit malfunctions such as noise and interrupted communications (**Fig. 1**).

The conventional response to the occurrence of such a fault has been to search for the fault using a labor-intensive technique by which maintenance personnel open up a cable-connection point in the field, determine whether the fault lies in the interval above that point (facility-center side) or below that point (customer side) using a device such as an insulation-resistance meter, and repeat that process until the fault location is isolated.

3. Present methods of estimating fault location

A typical method of evaluating the soundness of a metallic cable is to measure characteristics such as insulation resistance, electrostatic capacitance, loop resistance, and extraneous voltage/current using testers and testing tools. However, it is not standard to estimate a fault location using a measurement device designed for that purpose in the field.

Measurement devices for estimating fault locations in metallic cables have actually existed for some time. However, obtaining estimations with such a device requires a sufficient understanding of measurement methods, knowledge of the state of the core wire being measured (open-circuit, short-circuit, or ground fault), and selection of an appropriate measurement method depending on the state of the core wire, line configuration, and other factors. These requirements have made fault-location operations difficult in practice.

4. Fault-location searching methods

We introduce here two key methods for estimating fault locations in metallic cables using commercially available devices.

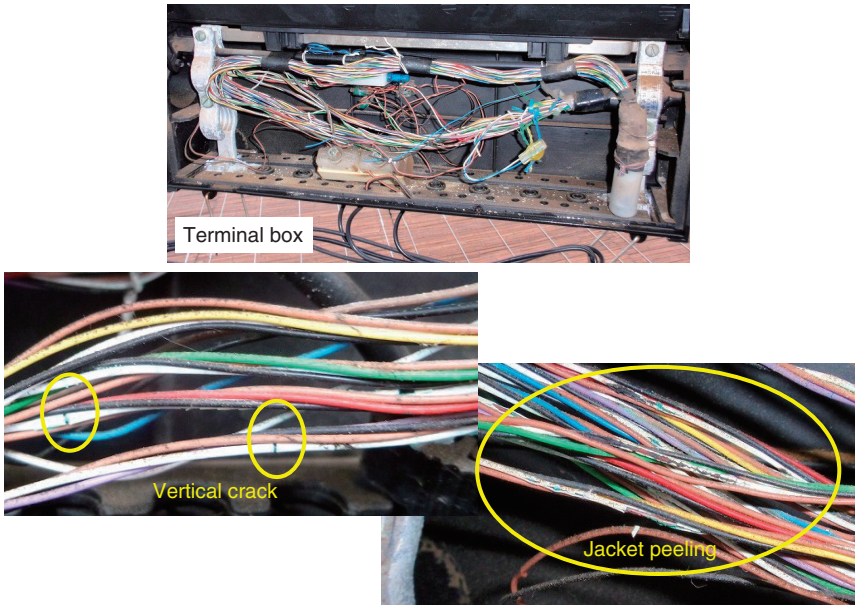


Photo 1. Types of core-wire faults in metallic cables.

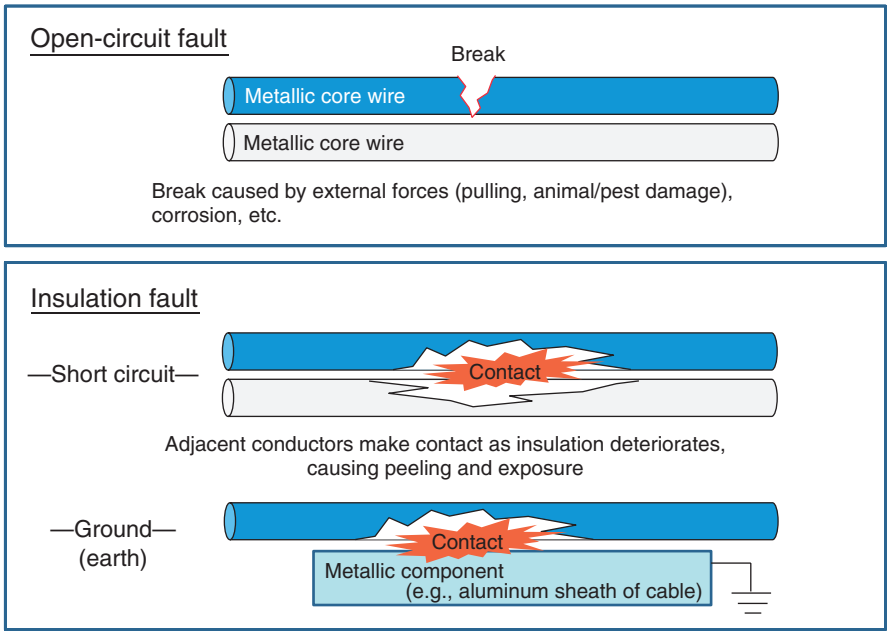


Fig. 1. Mechanisms of core-wire faults.

(1) Pulse radar method

This is an effective measurement method for searching out fault locations for open-circuit and short-circuit faults. A pulse voltage is applied to the core wire, the round-trip propagation time of the pulse voltage reflecting off the abnormal point is graphed, and the distance to the abnormal point is estimated. Specifically, this method derives the distance to the fault location from the product of the round-trip propagation time and the propagation

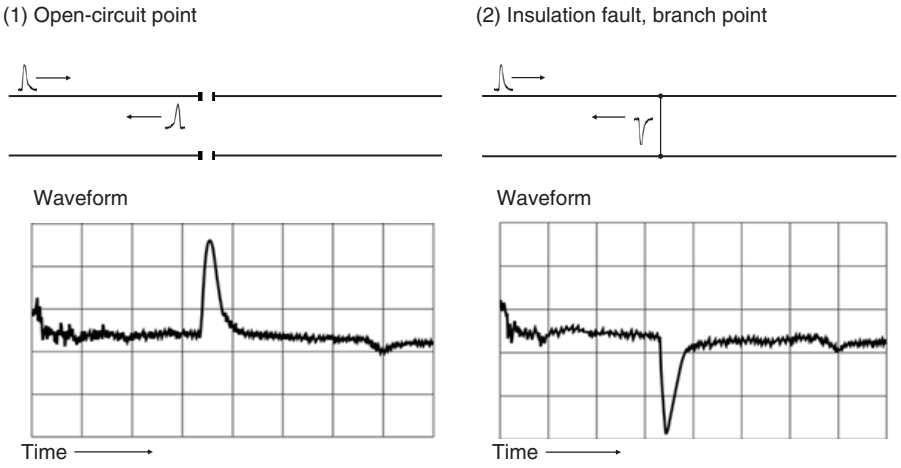


Fig. 2. Waveforms in pulse radar method.

speed of the pulse voltage, the latter of which depends on the type of cable. For example, propagation speed is 96 m per microsecond in the case of colored code polyethylene (CCP) conductors with a diameter of 0.4 mm.

A property of a pulse signal is that it will be reflected when encountering an impedance mismatch in the cable. The point where such an impedance mismatch occurs may indicate an open-circuit point, insulation-fault point, or branch point in the cable. An open-circuit point corresponds to a state of high impedance and is therefore displayed as an upward waveform. An insulation-fault point or branch point, meanwhile, corresponds to a state of low impedance and is therefore displayed as a downward waveform (**Fig. 2**).

With this method, it may be difficult to read an open-circuit waveform depending on the diameter of the core wire and the configuration of the target line. The comparison method has been found to be effective in such cases. This method simultaneously connects a faulty line and a sound line under identical conditions to a measurement device and displays the difference between the obtained pulse waveforms to clarify the fault location (**Photo 2**).

A weak point of the pulse radar method is that reflection from each end point (open end) of each branch in a branched configuration is the same as that from an open-circuit point. This may complicate the pulse waveform graph, preventing the fault location from being displayed (inferred). Therefore, the pulse radar method is an effective means of searching for the location of an open-circuit or short-circuit fault in a line with no branches.

(2) Murray loop method

This is an effective measurement method for finding the location of a short-circuit fault or ground fault. The principle of the Wheatstone bridge is applied to estimate the location of a fault point on the basis of line resistance values (**Fig. 3**). The length of the measured line must be input to the measurement device beforehand so that resistance can be read as distance and so the end point of the measured line can be strapped (looped) to form a measurement circuit.

While measurement in the case of a branched line is relatively difficult for the pulse radar method described above, the Murray loop method can be used to estimate the fault location even on a line with branches. For example, when measurements are conducted when the end point of the main line is strapped, and given that the fault point lies on the branch line, the location (distance) at which the branch on the main line begins would be measured as the fault point. However, if that location were assumed to be sound, the line interval branching from that position would become suspect. In this way, the Murray loop method is an effective method for finding fault locations regardless of branching in the case of short-circuit or ground fault. The pulse radar method and Murray loop method are compared in **Table 1**.

5. Example of searching for fault location

We present here an example of locating a fault in a metallic cable (**Fig. 4**). First, it is important to confirm the present state of the core wire to be searched,

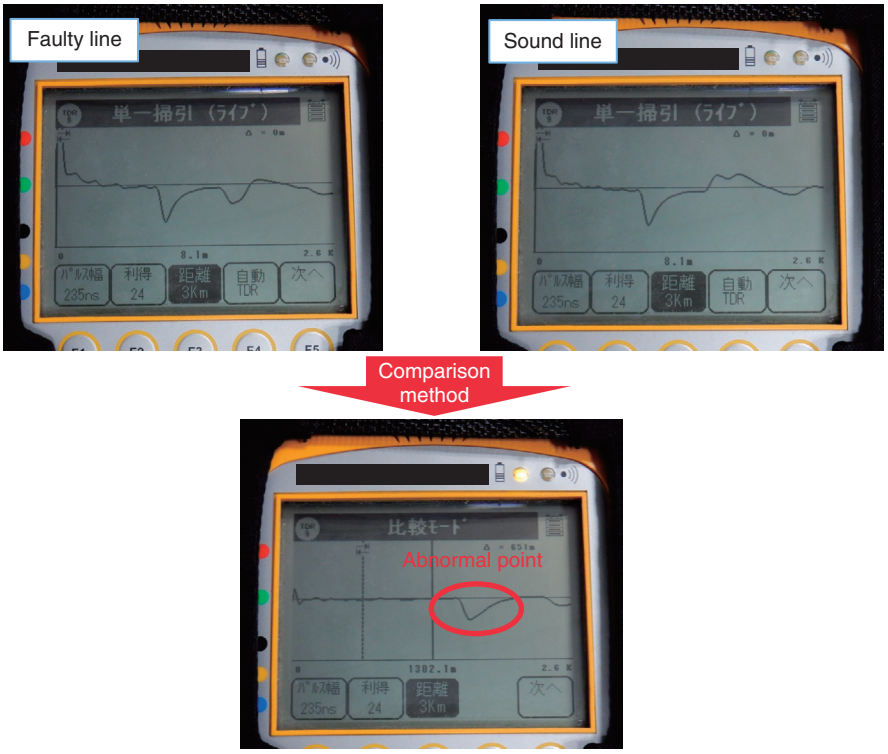


Photo 2. Comparison method.

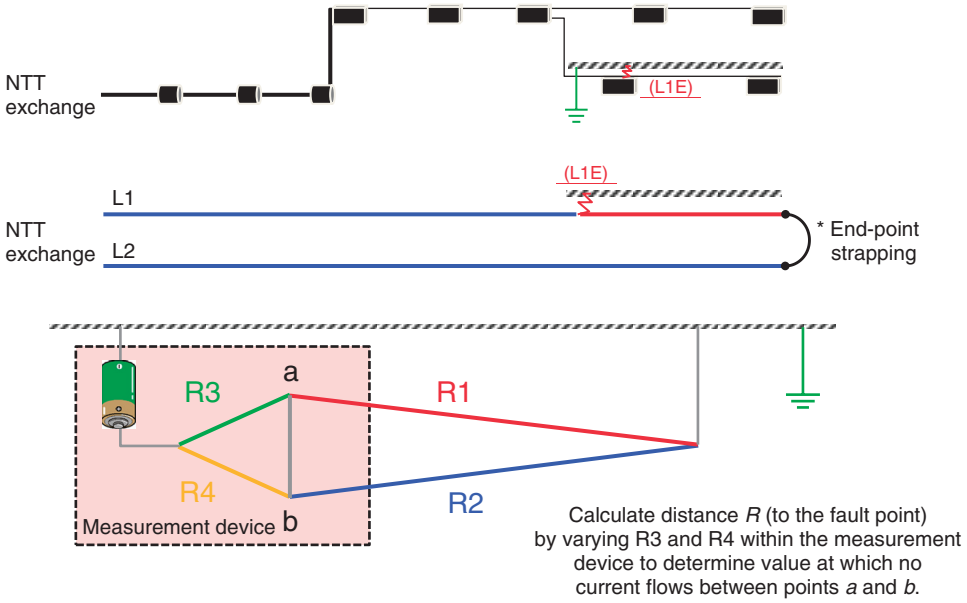


Fig. 3. Concept of Murray loop method.

Table 1. Comparison of measurement methods.

Method	Line format	State of measured core wire	Measurement notes
Pulse radar method (TDR) • Estimates suspected location from waveform obtained by measuring reflection of pulse signal	Multi: No Non-multi: Yes	Open circuit Insulation fault (SC)* * Apply RFL depending on line conditions	• Requires waveform analysis
Murray loop method (RFL) • Estimates suspected location by converting obtained resistance value to line length	Multi: Yes Non-multi: Yes	Ground fault (L1/L2 E, L1L2E) Insulation fault (SC)	• Requires looping at end of measured wire • Requires sound wire for comparison measurement

TDR: time domain reflectometry
RFL: resistance fault location
SC: short circuit

(1) Determine state of core wire (insulation fault, ground fault) at NTT exchange.
(2) Apply strap (1) to end of measured line.
(3) Conduct measurements using Murray loop method from exchange. ⇒ distance to abnormality = 2500 m (point (a))
(4) Check point (a). If abnormal, repair. If not abnormal, the branch line from that point is suspect.
(5) Apply strap (2) to end of this branch line.
(6) Conduct measurements using Murray loop method from exchange. ⇒ distance to abnormality = 2700 m (point (b))

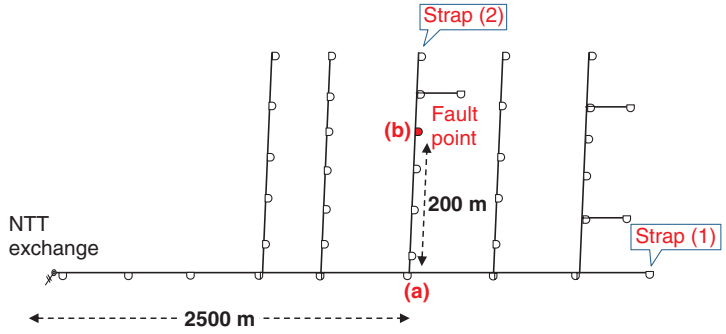


Fig. 4. Example of searching for fault location in aerial line.

that is, whether the fault is of the open-circuit, short-circuit, or ground type, so that the most appropriate fault location search method can be selected. If the core wire in question is in a state other than an open-circuit, short-circuit, or ground fault, care should be taken since such a state is not suited to fault location searching using a measurement device.

Next, the fault is evaluated to determine whether it occurred in the underground interval or the aerial interval. If it can be determined to lie on the underground interval, the position is identified and the search is completed. However, if it is judged to lie on the aerial interval and branch lines exist, the procedure is to obtain measurements using the Murray loop method after strapping any one of the branch lines on the main line to determine a fault point. Then, if that position after checking turns out to be sound, the branch line connected to that point is suspect. If that branch line itself has no branching, the pulse radar method can be used to obtain measurements from the start point of that branch. On the other hand, if the branch line itself includes branching, its end point can be looped and the Murray loop method used to search again for the fault location. The fault location can be found in this way.

In an aerial interval, metallic cables are said to be easily affected by external factors such as wind, rain, and ultraviolet rays that can give rise to many faults. Since it is normal for the aerial interval of cabling to have branch lines, an appropriate approach would be to use the Murray loop method to perform rough interval identification and narrow down the interval, and to use the pulse radar method to isolate the fault position. In other words, it is more effective to use multiple methods in a composite manner instead of being tied to a single method.

6. Conclusion

This article introduced techniques for quickly estimating fault locations in metallic cable with the aim of preventing circuit failures.

While leveraging the knowledge and expertise it has gained in the maintenance and operation of diverse facilities, the Technical Assistance and Support Center will continue to provide technology consultations, reply to support requests, and promote technology dissemination for all concerned parties in order to strengthen the reliability of access network facilities.

External Awards

OFT Young Researcher's Award
Winner: Atsushi Nakamura, NTT Access Network Service Systems Laboratories
Date: February 17, 2017
Organization: The Institute of Electronics, Information and Communication Engineers (IEICE) Technical Committee on Optical Fiber Technologies (OFT)

For "Mode Field Diameter of LP11 Mode for Estimating Coupling Efficiencies at a Splice and Its Measurement Technique."
Published as: A. Nakamura, K. Okamoto, Y. Koshikiya, and T. Manabe, "Mode Field Diameter of LP11 Mode for Estimating Coupling Efficiencies at a Splice and Its Measurement Technique," IEICE Tech. Rep., Vol. 116, No. 198, OFT2016-12, pp. 5–10, Aug. 2016.

TELECOM System Technology Award
Winner: Shohei Kamamura, Daisaku Shimazaki, Yoshihiko Uematsu, NTT Network Service Systems Laboratories; Kouichi Genda, Nihon University; Koji Sasayama, National Institute of Informatics
Date: March 24, 2017
Organization: The Telecommunications Advancement Foundation

For "Multi-staged Network Restoration from Massive Failures Considering Transition Risks."
Published as: S. Kamamura, D. Shimazaki, Y. Uematsu, K. Genda, and K. Sasayama, "Multi-staged Network Restoration from Massive Failures Considering Transition Risks," 2014 IEEE International Conference on Communications, Sydney, Australia, June 2014.

Young Researcher's Award
Winner: Hiroshi Yamamoto, NTT Network Service Systems Laboratories
Date: March 24, 2017
Organization: IEICE

For "Availability Improvement Schemes for Multi-carrier Optical Transmission Systems."
Published as: H. Yamamoto, K. Kitamura, M. Yokota, H. Date, S. Kamamura, H. Maeda, and Y. Uematsu, "Availability Improvement Schemes for Multi-carrier Optical Transmission Systems," Proc. of the 2016 IEICE Society Conference, BS-5-3, Hokkaido, Japan, Sept. 2016.

Young Researcher's Award
Winner: Daisuke Goto, NTT Access Network Service Systems Laboratories
Date: March 24, 2017
Organization: IEICE

For "Validation of the Demodulation of the Multi-satellite/Multi-beam Systems in Time-frequency Asynchronous MIMO Channels."
Published as: D. Goto, F. Yamashita, and K. Kobayashi, "Validation of the Demodulation of the Multi-satellite/Multi-beam Systems in Time-frequency Asynchronous MIMO Channels," Proc. of the 2016 IEICE General Conference, B-3-11, Fukuoka, Japan, Mar. 2016.

Young Researcher's Award
Winner: Hirotaka Ujikawa, NTT Access Network Service Systems Laboratories
Date: March 24, 2017
Organization: IEICE

For "Skipping Power-on Self Test for Energy Efficient Optical Access Equipment" and "Evaluation of Constrained Performance of Optical Access Equipment with Limited Functionality."
Published as: H. Ujikawa, T. Harada, M. Yoshino, K. Suzuki, and A. Otaka, "Skipping Power-on Self Test for Energy Efficient Optical Access Equipment," Proc. of the 2016 IEICE General Conference, B-8-77, Fukuoka, Japan, Mar. 2016.
H. Ujikawa, T. Harada, M. Yoshino, K. Suzuki, and A. Otaka, "Evaluation of Constrained Performance of Optical Access Equipment with Limited Functionality," Proc. of the 2016 IEICE Society Conference, BS-8-30, Hokkaido, Japan, Sept. 2016.

Special Technical Award in Smart Radio
Winner: Doohwan Lee, Hirofumi Sasaki, Hiroyuki Fukumoto, Ken Hiraga, Tadao Nakagawa, and Hiroyuki Shiba, NTT Network Innovation Laboratories
Date: May 25, 2017
Organization: IEICE Technical Committee on Smart Radio

For "An Experimental Study of the Orbital Angular Momentum (OAM) Multiplexing."
Published as: D. Lee, H. Sasaki, H. Fukumoto, K. Hiraga, T. Nakagawa, and H. Shiba, "An Experimental Study of the Orbital Angular Momentum (OAM) Multiplexing," IEICE Tech. Rep., Vol. 116, No. 276, SR2016-56, pp. 7–8, Oct. 2016.

Best Paper
Winner: Hirofumi Sasaki, Doohwan Lee, Hiroyuki Fukumoto, Tadao Nakagawa, and Hiroyuki Shiba, NTT Network Innovation Laboratories
Date: May 25, 2017
Organization: IEICE Technical Committee on Smart Radio

For "Beam Propagation Control Method for OAM Multiplexing Using Gaussian Beam."
Published as: H. Sasaki, D. Lee, H. Fukumoto, T. Nakagawa, and H. Shiba, "Beam Propagation Control Method for OAM Multiplexing Using Gaussian Beam," IEICE Tech. Rep., Vol. 116, No. 276, SR2016-73, pp. 105–110, Oct. 2016.

Honorable Mention Award
Winner: Maya Okawa, Hideaki Kim, and Hiroyuki Toda, NTT Service Evolution Laboratories
Date: May 29, 2017
Organization: The 18th IEEE International Conference on Mobile Data Management (IEEE MDM 2017) General Co-Chairs

For "Online Traffic Flow Prediction Using Convolved Bilinear Poisson Regression."
Published as: M. Okawa, H. Kim, and H. Toda, "Online Traffic Flow Prediction Using Convolved Bilinear Poisson Regression," IEEE MDM 2017, Daejeon, South Korea, May/June 2017.

Best Paper Award
Winner: Yo Yamaguchi and Takana Kaho, NTT Network Innovation Laboratories; Ichihiko Toyoda, Saga University; Kazuhiro Uehara and Tadao Nakagawa, NTT Network Innovation Laboratories; and Kiyomichi Araki, Tokyo Institute of Technology
Date: June 1, 2017
Organization: IEICE

For "A Balanced Mixer Using Asymmetric Stacked Lines for a Quasi-millimeter Wave Band."
Published as: Y. Yamaguchi, T. Kaho, I. Toyoda, K. Uehara, T. Nakagawa, and K. Araki, "A Balanced Mixer Using Asymmetric Stacked Lines for a Quasi-millimeter Wave Band," IEICE Trans. Commun. (Japanese Edition), Vol.J99-C, No. 5, pp. 266–275, May 2016.

Papers Published in Technical Journals and Conference Proceedings

Presenting Changes in Acoustic Features Synchronously to Respiration Alters the Affective Evaluation of Sound

T. G. Sato, J. Watanabe, and T. Moriya

International Journal of Psychophysiology, Vol. 110, pp. 179–186, August 2016.

Synchronization of respiration to cyclic auditory stimuli is a well-observed phenomenon and known to have an effect on affective evaluation of the presented sound. However, no studies have separated the effect of the change in respiratory movement itself and that when there is synchrony between respiration and sound. In this study, we used a system that can change the acoustic features synchronously with the respiration phase and directly investigated the effect the synchrony has on affective ratings without changing respiratory movements. An acoustic stimulation was presented where the sound intensity (SI) or fundamental frequency (F0) was modulated in response to the participant's respiration phase. Affective evaluations of the acoustic stimuli were made by using the Self-Assessment Manikin (SAM). The experiments compared synchronous and asynchronous conditions.

In the synchronous condition, SI (or F0) was increased with inhalation (decreased with exhalation) or decreased with inhalation (increased with exhalation). In the asynchronous condition, a sound identical to that presented in the synchronous condition was replayed. The participants evaluated sounds that were acoustically the same but where the temporal relationship differed between respiration and the acoustic features. In our results, significantly higher arousal ratings were observed when the change in SI and respiration (inhalation or exhalation) was synchronous and when the increase in F0 and inhalation was synchronous. This suggests that the synchronous phenomenon between respiration and auditory stimuli can play a critical role in affective evaluation.

Long-haul Dense Space Division Multiplexed Transmission over Low-crosstalk Heterogeneous 32-core Transmission Line Using Partial Recirculating Loop System

T. Mizuno, K. Shibahara, F. Ye, Y. Sasaki, Y. Amma, K. Takenaga, Y. Jung, K. Pulverer, H. Ono, Y. Abe, M. Yamada, K. Saitoh, S. Matsuo, K. Aikawa, M. Bohn, D. J. Richardson, Y. Miyamoto, and T. Morioka

Journal of Lightwave Technology, Vol. 35, No. 3, pp. 488–498, February 2017.

In this paper, we present long-haul 32-core dense space-division-multiplexed (DSDM) unidirectional transmission over a single-mode multicore transmission line. We developed a low-crosstalk heterogeneous 32-core fiber with a square lattice arrangement, and a novel partial recirculating loop system. The span crosstalk of the 51.4-km 32-core transmission line was less than –34.5 dB. This allowed the transmission of polarization-division-multiplexed 16 quadrature amplitude modulation (PDM-16QAM) signals through all 32 cores over a long distance exceeding 1000 km. We demonstrate 32-core DSDM 20 wavelength-division-multiplexed PDM-16QAM transmission over 1644.8 km with a high aggregate spectral efficiency of 201.46 b/s/Hz. Additionally, we examine the effect of crosstalk on the transmission performance of each core, and show that the Q-penalty has strong correlation with intercore crosstalk.

Outlier Management for Robust Visual SLAM in Dynamic Environments with Easy Map and Camera Pose Initialization

J. Shimamura, K. Sudo, M. Morimoto, T. Osawa, and Y. Taniguchi

IIEEJ Transactions on Image Electronics and Visual Computing, Vol. 5, No. 1, pp. 20–33, June 2017.

We present a robust monocular visual simultaneous localization and mapping (Visual SLAM) method, an algorithm capable of estimating robust camera poses in dynamic environments and reducing the user operation load for Visual SLAM initialization. In dynamic environments, standard monocular Visual SLAM using 3D/2D matched points between a recovered 3D map and feature points on a current frame tends to fail because the 3D map is distorted by moving objects with duration. To address this issue, we classify outliers from a robust estimator in the camera pose estimation into feature points on moving objects and mismatched points on occlusions, specular reflections, textureless regions, and so on. To achieve this, we first construct an angle histogram based on outlier flows that are vectors between reprojected points and matched points at a current frame, then approximate the obtained angle histogram using a mixture of Gaussian functions. Finally, we estimate the parameters for Gaussian mixtures by using an expectation maximization algorithm. We also introduce weighted tentative initial values of a 3D map and a camera pose to reduce the user operation load. Experimental results demonstrate that our system can work robustly in highly dynamic environments and initialize by itself without user assistance.

Heart Rate Measurement Based on Event Timing Coding Observed by Video Camera

T. G. Sato, Y. Shiraki, and T. Moriya

IEICE Transactions on Communications, Vol. E100-B, No. 6, pp. 926–931, June 2017.

The purpose of this study was to examine an efficient interval encoding method with a slow-frame-rate image sensor, and show that the encoding can work to capture heart rates from multiple persons. Visible light communication (VLC) with an image sensor is a powerful method for obtaining data from sensors distributed in the field with their positional information. However, the capturing speed of the camera is usually not fast enough to transfer interval information like the heart rate. To overcome this problem, we have developed an event timing (ET) encoding method. In ET encoding, sensor units detect the occurrence of heart beat event and send their timing through a sequence of flashing lights. The first flash signal provides the rough timing and subsequent signals give the precise timing. Our theoretical analysis shows that in most cases the ET encoding method performs better than simple encoding methods. Heart rate transfer from multiple persons was examined as an example of the method's capabilities. In the experimental setup, the developed system successfully monitored heart rates from several participants.

Vibration on the Soles of the Feet Evoking a Sensation of Walking Expands Peripersonal Space

T. Amemiya, Y. Ikei, K. Hirota, and M. Kitazaki

Proc. of 2017 IEEE World Haptics Conference, pp. 234–239, Munich, Germany, June 2017.

Processing of audio-tactile multisensory stimuli presented within the space immediately surrounding our body, i.e., peripersonal space (PPS), is known to be facilitated and the boundaries of the PPS extended by body action such as walking. However, it is unclear whether the boundaries change when a sensation of walking is induced with no physical body motion. Here, we presented several vibration patterns on the soles of the feet of seated participants to evoke a sensation of walking and examined the change in reaction times to detect a vibrotactile stimulus on the chest while listening to a looming sound approaching the body, which was taken as a behavioral proxy for the PPS boundary. Results revealed that a cyclic vibration consisting of lowpass-filtered walking sounds presented at the soles that clearly evoked a sensation of walking decreased the reaction times, indicating that the PPS boundary was expanded forward by inducing a sensation of walking.

100-Gb/s Transmission over a 2520-km Integrated MCF System Using Cladding-pumped Amplifiers

C. Castro, S. Jain, E. De Man, Y. Jung, J. Hayes, S. Calabrò, K. Pulverer, M. Bohn, S. Alam, D. J. Richardson, K. Takenaga, T. Mizuno, Y. Miyamoto, T. Morioka, and W. Rosenkranz

IEEE Photonics Technology Letters, Vol. 29, No. 14, pp. 1187–1190, July 2017.

A 10.5-Tb/s optical transmission (15 × 100 Gb/s QPSK channels per core) over 2520 km of multicore fiber is achieved using an integrated multicore transmission link consisting of directly spliced multicore components, such as fan-in/fan-out fiber couplers, a 60-km trench-assisted seven-core hexagonal fiber and cladding-pumped erbium-ytterbium-doped fiber amplifiers.