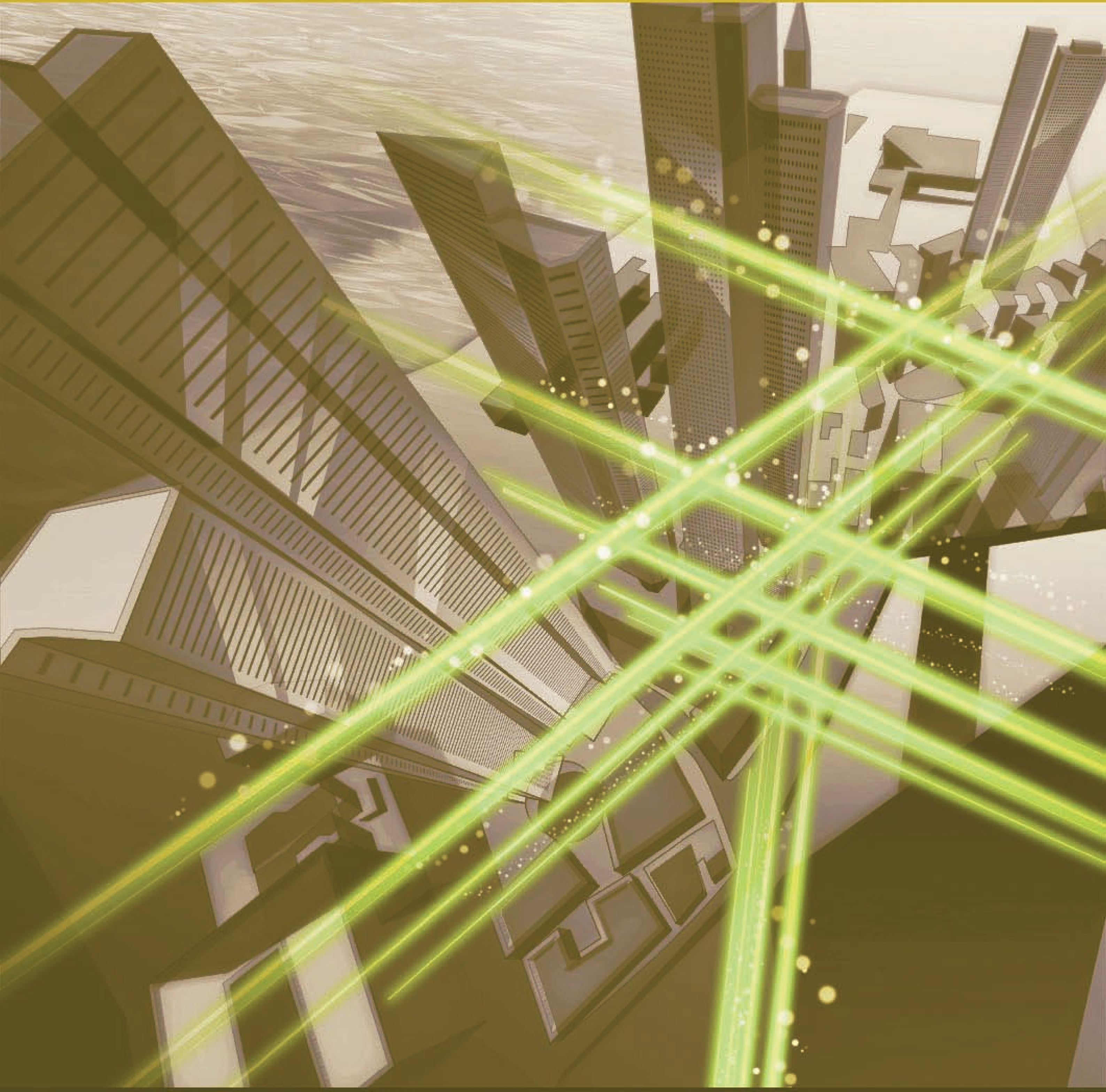
NTT Technical Review

7
2015



July 2015 Vol. 13 No. 7

NTT Technical Review

July 2015 Vol. 13 No. 7

Feature Articles: The Challenge of Creating Epoch-making Services that Impress Users

Regular Articles

Global Standardization Activities

External Awards/Papers Published in Technical Journals and Conference Proceedings

Creating Epoch-making Services that Excite Users

Hirohisa Jozawa, Shuichi Nishioka, and Kota Hidaka

Abstract

When the telephone first made it possible for people to chat with friends hundreds of miles away, and when the Internet made it possible to obtain thousands of books' worth of information without having to visit a library, people were excited by these technological advances. NTT Service Evolution Laboratories has set itself the challenge of developing technologies for the creation of new services that will generate as much excitement as the phone and Internet did. This article presents an overview of our work.

Keywords: highly realistic, agents, big data

1. Services based on four viewpoints

NTT Service Evolution Laboratories (EV Labs) has set itself the goal of creating exciting new services by 2020. With the aim of contributing to the success of the Olympic and Paralympic Games and other events to be held in Japan during that year, we are promoting research and development (R&D) of services that excite users from the following four viewpoints (see **Fig. 1**):

(1) Shared excitement: Sharing our excitement with people from all over Japan and from the rest of the world during the 2020 international festival

(2) Supporting growth: Addressing questions about what information and communication technology (ICT) has to offer for the growth of society

(3) Optimal navigation: Understanding the current location and situation, discovering the causes of any problems, and dealing with them effectively

(4) Hospitality by ICT: Providing the best possible hospitality with the latest ICT, to maintain a culture where all people can enjoy the convenience of ICT equally

1.1 Shared excitement

During international soccer matches, many users support their teams at public viewings in places with large screens such as in sports bars and in open areas outside train stations. Users gathering in the same place can be drawn into the ups and downs of the match, and when their team wins, they experience the same excitement as people who are actually at the stadium.

We expect that public viewings will be introduced for a more diverse range of sports in the future and will happen in more places. At EV Labs, we are conducting R&D aimed at providing more realistic match viewing, forming large screens more easily, distributing content simultaneously all over the world, and appropriately sharing content. Further details can be found in the article "Developing Technologies for Services that Deliver the Excitement of Games Worldwide" [1].

1.2 Supporting growth

If you think back to the time when you first learned to ride a bicycle, or when you first managed to swim 25 meters, you may recall that these achievements did not happen in isolation but occurred with the help of others such as family, friends, and teachers. Support is considered to include not only teaching ways of realizing goals, but also providing verbal encouragement (e.g., a mixture of praise and constructive criticism).

To offer support, it is first necessary to know about

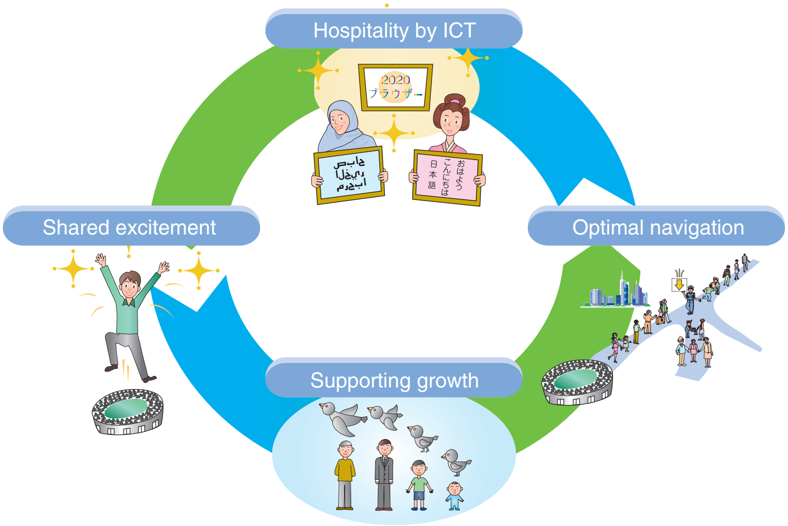


Fig. 1.   Creating services that excite users.

the user's situation. Once that has been done, assistance for growth is provided by interacting with the user by giving a pep talk in order to decide how support should be provided. Since each user has different needs, the needs must be personalized. At EV Labs, entities called *personal agents* are being developed to provide personalized growth support. To ensure the convenience of users, the data and know-how accumulated through interactions with users should be stored on the network side (e.g., in the cloud), and should be accessible at any time as the users grow. This also means that the system should not be reliant on any particular type of network. Further details can be found in the article "Personal Agents to Support Personal Growth" [2].

1.3   Optimal navigation

If we could create a system that would make it possible for people to arrive at their destination without getting lost or stuck in traffic, it would be very convenient for users. One might even say that an optimal navigation system could help users to discover new things and find things they are interested in by making stopovers to fill up spare time instead of always taking the shortest path to the destination.

To realize such a system, we first estimate where the user is, and how the user is moving. Then we predict the motion as a crowd targeting a huge number of users.

Optimal navigation is implemented by analyzing individuals and groups. There are many different usage scenarios, but we are considering a framework that allows navigation in disaster situations. Further details can be found in the article "Proactive Navigation Optimized for Individual Users" [3].

1.4   Hospitality by ICT

People traveling overseas can feel anxious about many different things such as potential difficulties with the local language and food. Many foreigners will visit Japan in 2020, and we are investigating ways to provide useful services for them, such as providing information on food that is translated into their native language. Further details can be found in the article "Creating *Omotenashi* Services for Visitors and Spectators in 2020" [4].

We are also studying the form in which information can be suitably provided not only to foreigners but also to people of any nationality, gender, or age. For example, it will be necessary to consider language notations that require vertical writing or are dependent on the language characteristics of specific countries. Other considerations include providing *ruby* characters (pronunciation guides) for users who are unable to read *kanji*, and large characters for users who are unable to read small print. Hospitality will be provided by building on previous work relating to universal web design initiatives and behavior observations. Further details can be found in the article "Towards the Creation of Attractive Services Based on an Understanding of Users" [5].

2. Future development

Some aspects of the initiatives discussed in these Feature Articles—such as their commercialization and applications—are being studied together with partner companies. This makes it essential to facilitate co-innovation through collaboration. Further details can be found in the article "Efforts toward Co-Innovation Promotion" [6]. We will continue to promote research so that the services created through the efforts of EV Labs will one day be described as epoch-making.

References

[1] A. Akutsu, K. Hidaka, M. Inoue, N. Ito, T. Yamaguchi, S. Fujimura, and A. Nakadaira, "Developing Technologies for Services that Deliver the Excitement of Games Worldwide," NTT Technical Review, Vol. 13, No. 7, 2015.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201507fa2.html

[2] Y. Tomita, T. Tanaka, T. Yamada, and K. Kitamura, "Personal Agents to Support Personal Growth," NTT Technical Review, Vol. 13, No. 7, 2015.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201507fa3.html

[3] 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

[4] N. Kanamaru, H. Tezuka, A. Fukayama, Y. Nakamura, H. Yamaguchi, and M. Motegi, "Creating *Omotenashi* Services for Visitors and Spectators in 2020," NTT Technical Review, Vol. 13, No. 7, 2015.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201507fa6.html

[5] Y. Asano, T. Ohno, S. Miyahara, M. Wanabe, M. Nakashige, A. Kimura, and K. Wago, "Towards the Creation of Attractive Services Based on an Understanding of Users," NTT Technical Review, Vol. 13, No. 7, 2015.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201507fa5.html

[6] K. Shindo, T. Sugiyama, S. Azuma, and K. Kurokawa, "Efforts toward Co-Innovation Promotion," NTT Technical Review, Vol. 13, No. 7, 2015.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201507fa7.html

Hirohisa Jozawa

Executive Research Engineer, Director of Research, NTT Service Evolution Laboratories.

He received the B.E. and M.E. in electrical engineering from Waseda University, Tokyo, in 1987 and 1989, respectively. He joined NTT in 1989. He has engaged in R&D of video coding and standardization of MPEG-4. His current research interests are big data analysis and its applications.

Shuichi Nishioka

Senior Research Engineer, Supervisor, NTT Service Evolution Laboratories.

He received the B.E. in electrical and computer engineering and the Dr.Eng. in information media and environment sciences from Yokohama National University, Kanagawa, in 1995 and 2005, respectively. Since joining NTT in 1995, he has been engaged in research on database management systems, copyright management systems, XML processing systems, and information retrieval systems. He is a member of the Information Processing Society of Japan and the Database Society of Japan.

Kota Hidaka

Senior Research Engineer, Planning Section (double as Natural Communication Project), NTT Service Evolution Laboratories.

He received the M.E. in applied physics from Kyushu University, Fukuoka, in 1998, and the Ph.D. in media and governance from Keio University, Kanagawa, in 2009. He joined NTT in 1998 and studied video skimming methods based on emphasized speech. From 2009 to 2013, he worked at NTT EAST and engaged in commercial developments of Home Gate Way (HGW), Home-ICT based on HGW platform, and smart mater system. He has been studying the immersive telepresence technology called *Kirari!* since 2014.

Developing Technologies for Services that Deliver the Excitement of Games Worldwide

Akihito Akutsu, Kota Hidaka, Masayuki Inoue, Naoki Ito, Tetsuya Yamaguchi, Shigeru Fujimura, and Atsushi Nakadaira

Abstract

NTT Service Evolution Laboratories has floated the concept of an ultra-high-presence service that will enable viewers to experience the feeling of actually being in a sporting venue or the like, wherever they are, and is therefore developing an immersive telepresence technology called *Kirari!*, towards its implementation. This article introduces the supporting technologies used with Kirari!, which involve ultra-high-presence media generation, spatial and environmental information synchronized delivery, and content license management.

Keywords: ultra-high-presence services, immersive telepresence, media synchronization technology

1. Introduction

NTT Service Evolution Laboratories floated the concept of delivering complete game spaces in real time—first within Japan and then to the world—with introduction intended for 2020, and is working on the immersive telepresence technology called *Kirari!* together with ultra-high-presence media generation technology, spatial and environmental information synchronized delivery technology, and content license management technology to support Kirari!. Our aim is to establish these technologies and implement a world where viewers can experience the feeling and excitement of being at a sporting venue, wherever they are.

2. Techniques behind ultra-high-presence services

2.1 Kirari! immersive telepresence technology

We have been working on the research and development (R&D) of the immersive telepresence technology called Kirari!, which is a technique for directly transmitting not just the images and sounds of players but also the environment in which the players exist. It also links real objects that exist in the environment, such as lighting and robots, at the transmission destination and reproduces them together with the sound by projection mapping. Kirari! is a combination of the next-generation video compression technology H.265/HEVC (High Efficiency Video Coding) [1], the lossless speech coding technology MPEG-4 ALS (Moving Picture Experts Group-4 Audio Lossless Coding) [2] that can compress audio information without distortion, and the highly realistic media synchronization technology Advanced MPEG Media Transport (MMT) that we have recently started researching and developing. The objective with Advanced MMT is to make it seem as if a game is being played in front of the viewers' eyes, in a form that far exceeds the reality enjoyed from a flat image. It does this by using synchronized reproduction that transmits a life-size image of the subject, for example, a player on a team, while transmitting all

Fig. 1. The world of immersive telepresence technology *Kirari!*.

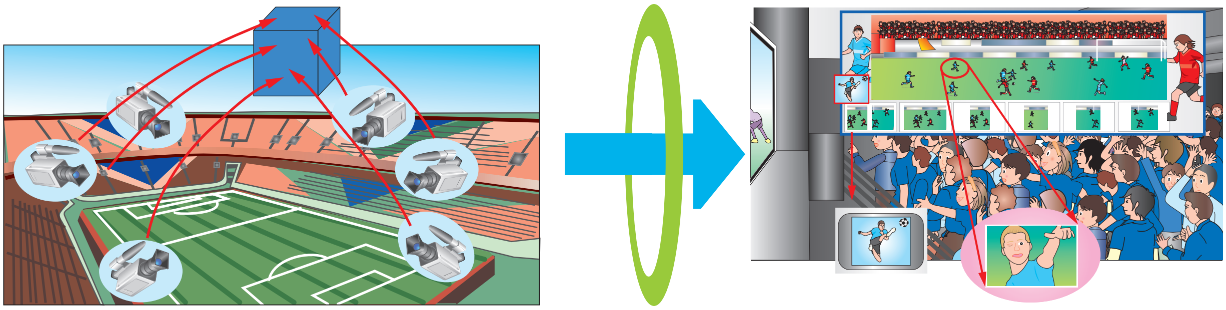


Fig. 2. Images of the highly realistic service suited to the reproduction environment.

information other than the subject to fit into the dimensions of the transmission destination, such as a television (TV) screen in a bar or restaurant, even when the transmission destination facilities have different dimensions, lighting, and acoustics.

It is predicted that in 2020, many viewers will be able to share the excitement of sports from sites around the world on TV or by public viewing, due to the spread of 4K and 8K broadcasting, although there are limits to the realism that can be portrayed on a flat screen. That is why our intention is to apply Kirari! technology to provide viewers at multiple points around the world with the physical sensations of being at games, as if they were at the sporting venue at the same time (**Fig. 1**).

2.2 Ultra-high-presence media generation technology

Ultra-high-definition video is not limited to 4K video. We intend to establish media processing technologies that can transmit ultra-high-definition video that surpasses even 8K video. More specifically, we aim to implement ultra-high-presence media generation technology that combines and integrates high-definition video and sound data that has been captured at a number of locations by a number of cameras, for reproduction using multiple display devices. This technology adjusts flexibly to the reproduction environment; it enables us to display content that is tailored to the user on a portable terminal such as a smartphone, or to display ultra-high-definition video that has been put together on a large screen at a public viewing venue, for example. It also reproduces a high level of realism that has not been seen before, by combining different kinds of media data such as video and sound (**Fig. 2**).

We previously implemented trials of 4K video transmission over a commercial network, taking

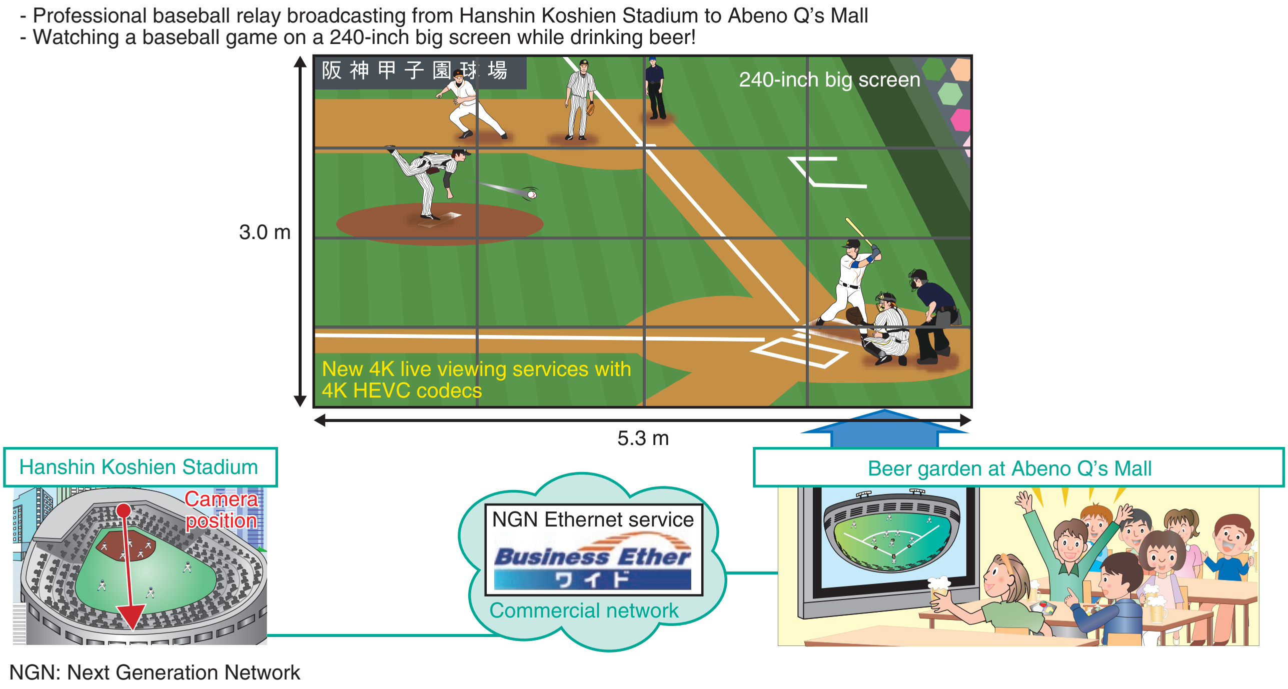


Fig. 3. Trial conducted at Hanshin Koshien Stadium.

advantage of the HEVC encoding technologies that NTT has developed. In a trial of 4K live viewing carried out at Hanshin Koshien Stadium [3], video of a professional baseball game captured by 4K cameras at the stadium was compressed by an H.265/HEVC encoder, and 4K video was transmitted in real time to a live-viewing venue through the Business Ether Wide business-oriented network [4] of NTT WEST (**Fig. 3**). In a trial carried out with public viewing at the All-Japan Phone-Answering Competition [5], we demonstrated 4K video delivery to multiple sites by best-effort IPv6 multicast delivery, using FLET'S Cast service for content providers who have the FLET'S Hikari Next service users. In a trial of live viewing at the 2014 Saga International Balloon Fiesta [6], we demonstrated 4K video delivery to multiple points from video captured by 4K cameras at the Saga Balloon Fiesta site, using the best-effort service FLET'S Hikari Next Hayabusa [7].

Through these trials, we were able to confirm that the HEVC encoding technology and commercial network were effective as a means of transmitting high-definition 4K video to remote sites. We plan to continue working on the R&D of ultra-high-presence media generation technologies intended for the transmission of ultra-high-definition video of 8K and higher.

2.3 Spatial and environmental information synchronized delivery technology

To achieve natural communication services, it is necessary to concentrate video, sound, and other spatial and environmental information, synchronize it, and deliver it efficiently with little lag. For example, when presenting ultra-high-resolution video that has been extracted and concentrated from various spaces and environments and that is to be shared between a number of playback devices, or when carrying out switching control of environmental information such as lighting and temperature in addition to video and sound to match timing, it is necessary to present various different bits of information in an integrated form to the user rather than individually for each playback device, which makes it important to have a mechanism for synchronized delivery and presentation. To exchange spatial and environmental information bidirectionally in the future, information will need to be transferred with little lag, so techniques for reducing the processing time for information transfer and the buffering time at the receiving terminals will be important. Images of a system that delivers synchronized ultra-high-resolution video, sound, and

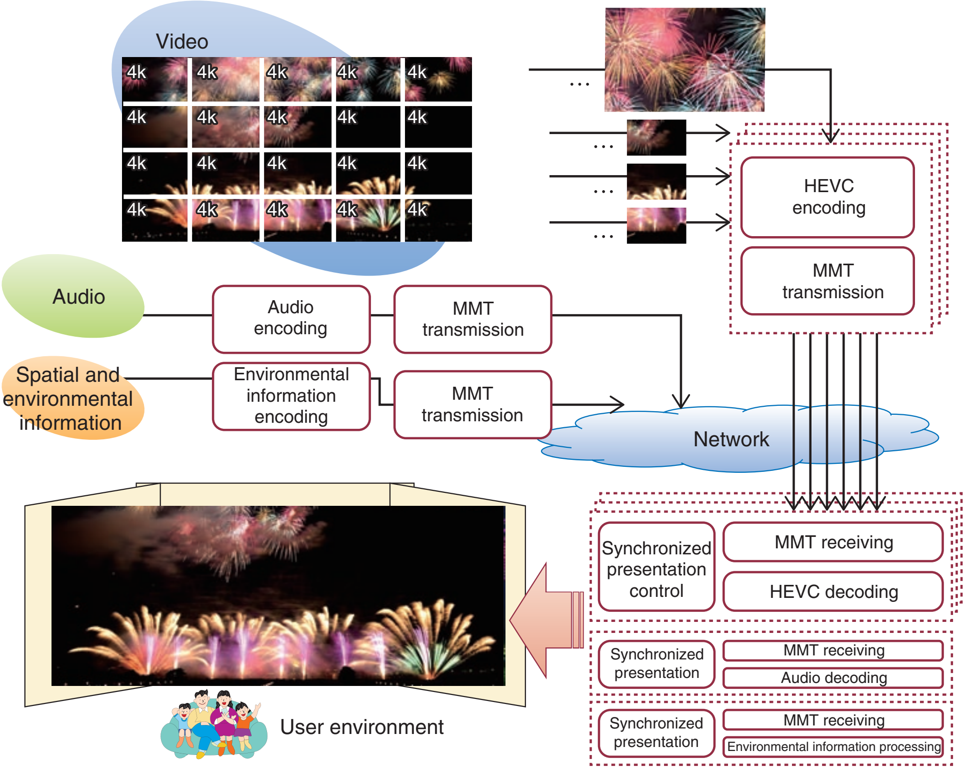


Fig. 4. Images of system that reproduces space and environment in the users' environment.

environmental information to reproduce the space and environment in the user's environment are shown in **Fig. 4**.

NTT Service Evolution Laboratories is first implementing synchronized delivery and presentation that focus on high-definition movies and sound, and is also working on the R&D of a multi-content synchronized delivery and presentation system, as part of efforts to establish spatial and environmental information synchronized delivery technology. By adopting the MMT method that is specified as the transmission format in ISO/IEC*1 23008-1 and ARIB*2 STD-B60, this multi-content synchronized delivery and presentation system delivers presentation timestamp information (UTC: Coordinated Universal Time) together with the video and sound content, in order to implement synchronized presentation at the receiving terminal.

It is also possible to freely integrate and present a number of different sources of video and sound content at the receiving terminal to suit the user's intentions, by transferring and receiving MPEG Composition Information (CI) that consists of XML (extensible markup language) information for presentation control as defined by ISO/IEC 23008-11. The multi-content synchronized delivery and presentation system implements a content synchronized delivery function and a presentation control information delivery function in the configuration shown in **Fig. 5**, making it possible to deliver a number of high-definition video images and partial images generated by extracting them from high-definition video, and to present these images in synchronization to a number of terminals to suit the user's intentions. This means that users can freely enjoy the content they want to play, without any constraints on the size of the display or terminal, or of processing power.

2.4 Content license management technology

In ultra-high realistic services, because smartphones capable of shooting 4K video are becoming available in the market, it is assumed that utilization of high-quality user-generated content such as photos or videos will increase. To ensure that this content can

---

*1 ISO/IEC: International Organization for Standardization, International Electrotechnical Commission

*2 ARIB: Association of Radio Industries and Businesses

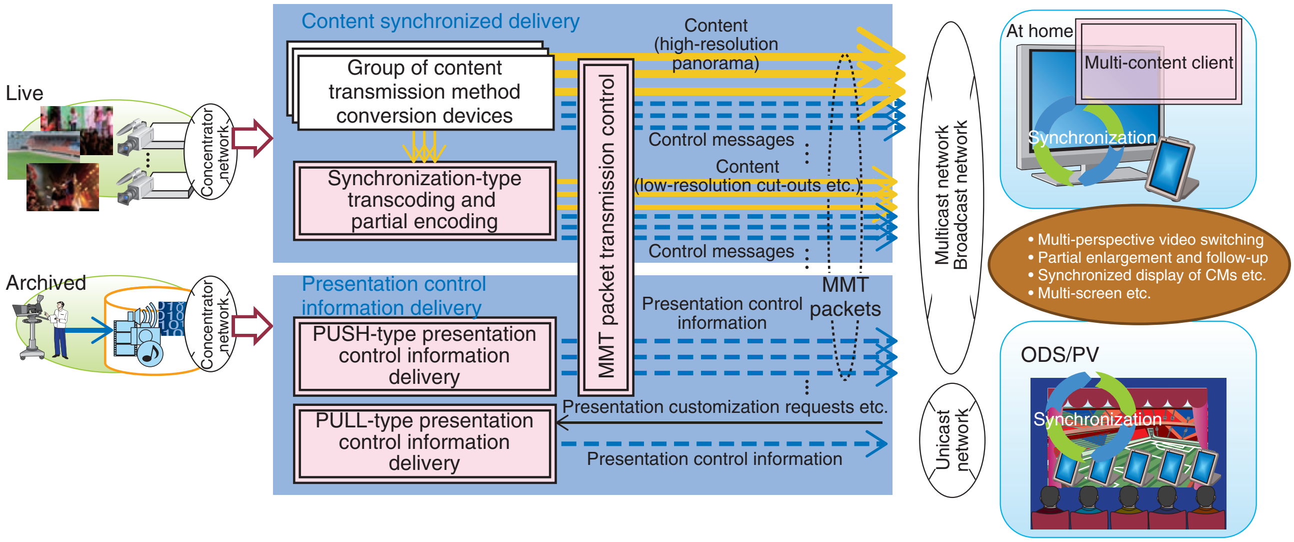


Fig. 5. Multi-content synchronized delivery and presentation system.

be utilized securely, a new highly convenient license management technology is needed; therefore, we are currently researching this kind of technology to ensure that such content is never used in unintended ways, as high-quality digital content can easily be digitally copied.

It is generally assumed that centralized control is needed in conventional license management technology. However, there have been several problems with this style. For example, if the security of the centralized management is breached, the whole system will entirely collapse, and it is difficult to provide flexible management with licenses set by the content creators themselves. In consideration of these problems, NTT is developing completely new license management technology through collaborative research with Muroran Institute of Technology, and our proposed technology can ensure sufficient reliability without centralized control by gathering small amounts of computing resources from each of the participants in the network.

Based on a so-called blockchain[*3] that records the entire history related to license information and that is shared by all network participants, our technology achieves highly reliable license management by combining blockchain and cryptographic technologies such as electronic signatures created by a public-key encryption method. In this case, the data structure of a blockchain is literally like chains; that is, each piece of data is recorded in a form that links the data before and after it.

The license issuing mechanism using a blockchain is shown in **Fig. 6**.

(1) First of all, the content creator uses a private key that is held only by the content owner, and broadcasts license information that records who has been given the license, with an electronic signature attached, to the entire network. At this point, it is guaranteed by the electronic signature that no one other than the content owner can issue a license. Accessibility and durability are ensured by recording the broadcast license information in a blockchain.

(2) When license information is added to the blockchain—in other words, when a new block is added to the blockchain—the computationally expensive problem, which involves searching for additional information (nonce) to ensure that the leading n digits are zero within the hash value calculation relating to the added block, is calculated competitively by all the participants. To make it difficult to fake the blockchain, a computationally expensive problem is set deliberately.

(3) When someone eventually finds the appropriate nonce, a new block containing license information is added to the blockchain, and the license information is approved with its

---

[*3] Blockchain: Used in applications such as digital (virtual) currency, a blockchain is a public ledger in which the exchanges of data are recorded in a chain-like form. The blockchain is shared with network users participating in the service.

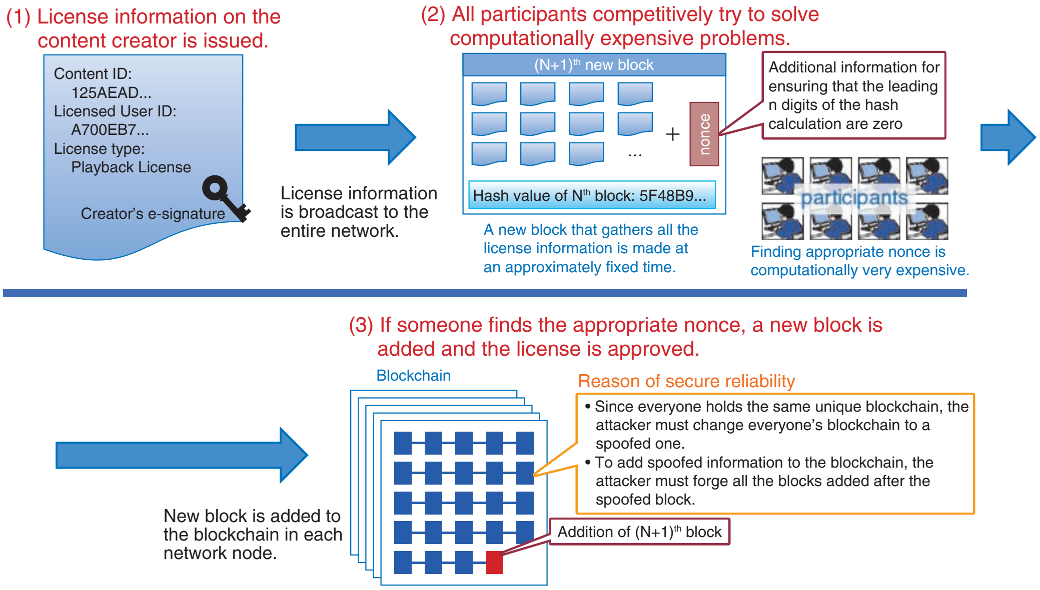


Fig. 6. Blockchain mechanism.

reliability being ensured by the efforts of all the participants.

To be accepted as a standard, it is not only important for this new content license management technology to be secure and convenient, but efforts must also be made to achieve increased use of it through community activities. In the future, we intend to continue working on this approach with the cooperation of many stakeholders such as content creators and manufacturers.

3. Future expansion

At NTT, we are involved in R&D that will enable people around the world to physically experience ultra-high-presence technology that allows them to feel like they were spectators at sports and live events, through collaboration with various partners up to 2020. Kirari! enables spectators at remote sites to observe games as they happen, by transmitting and reproducing entire game spaces at many distant sports arenas and live venues, so they can share physical sensations such as those conveying speed and strength. We also plan to investigate applications to events that have been difficult for many people to appreciate in the past because of their remoteness, such as local festivals that serve as examples of our intangible cultural heritage.

References

[1] Feature Articles: "Video Technology for 4K/8K Services with Ultra-high Sense of Presence," NTT Technical Review, Vol. 12, No. 5, 2014.
https://www.ntt-review.jp/archive/2014/201405.html

[2] Selected Papers: "Research Activities in Laboratories of New Fellows Part1," NTT Technical Review, Vol. 5, No. 12, 2007.
https://www.ntt-review.jp/archive/2007/200712.html

[3] NTT WEST press release issued on June 27, 2014 (in Japanese).
https://www.ntt-west.co.jp/news/1406/140627a.html

[4] NTT WEST, Overview of Business Ether Wide solution (in Japanese).
https://www.ntt-west.co.jp/solution/solution/category/wide/

[5] NTT WEST press release issued on November 10, 2014 (in Japanese).
https://www.ntt-west.co.jp/news/1411/141110a.html

[6] NTT WEST press release issued on October 24, 2014 (in Japanese).
http://www.ntt-west.co.jp/saga/s_release/2014/20141024.pdf

[7] NTT WEST, FLET'S Hikari.
http://flets-w.com/english/flets_hikari/next/

Akihito Akutsu

Executive Research Engineer, Project Director, Natural Communication Project, NTT Service Evolution Laboratories.

He received the M.E. in engineering from the Center for Environmental Remote Sensing, Chiba University. In 2001, he received the Ph.D. in natural science & technology from Kanazawa University. Since joining NTT in 1990, he has been engaged in R&D of video indexing technology based on image/video processing, and man-machine interface architecture design. From 2003–2006, he worked at NTT EAST, and engaged in managing a joint venture between NTT EAST and Japanese broadcasters. In 2008, he was appointed as Director of NTT, Cyber Solutions Laboratories. He was engaged in the R&D project of the broadband and broadcast services. In October 2013, he was appointed as Executive Producer of 4K/8K HEVC in NTT Media Intelligence Laboratories.

He received the Young Engineer Award and Best Paper Award from the Institute of Electronics, Information and Communication Engineers (IEICE) in 1993 and 2000. He is a member of IEICE.

Kota Hidaka

Senior Research Engineer, Planning Section (double as Natural Communication Project), NTT Service Evolution Laboratories.

He received the M.E. in applied physics from Kyushu University, Fukuoka, in 1998, and the Ph.D. in media and governance from Keio University, Kanagawa, in 2009. He joined NTT in 1998 and studied video skimming method based on emphasized speech. From 2009 to 2013, he worked at NTT EAST and engaged in commercial developments of Home Gate Way (HGW), Home-ICT based on HGW platform, and smart mater system. He has been studying the immersive telepresence technology called Kirari! since 2014.

Masayuki Inoue

Senior Research Engineer, Natural Communication Project, NTT Service Evolution Laboratories.

He received the B.E. and M.E. in electrical engineering from Tokyo University of Science in 1994 and 1996, respectively, and the Ph.D. in engineering from Tokyo Institute of Technology in 2008. He joined NTT in 1996 and has been engaged in research on 3D cyberspace and video communication systems. He is currently working on the business development of H.265/HEVC 4K LSI. His research interests also include user interaction and video system evaluation. Dr. Inoue is a member of IEICE and the Institute of Image Information and Television Engineers (ITE).

Naoki Ito

Senior Research Engineer, Natural Communication Project, NTT Service Evolution Laboratories.

He received the M.E. from Toyohashi University of Technology, Aichi, in 2001. He joined NTT in 2001 and engaged in research on character recognition. He moved to NTT EAST in 2004 and engaged in the development of security systems. He moved to NTT Cyber Space Laboratories (now, NTT Media Intelligence Laboratories) in 2008. Since 2014, he has been working on the development of high-definition visual communication service systems. He is a member of IEICE.

Tetsuya Yamaguchi

Senior Research Engineer, NTT Service Evolution Laboratories.

He received his B.E., M.E., and Ph.D. in information engineering from Osaka University in 1997, 1999, and 2008, respectively. Since joining NTT in 1999, he has been engaged in R&D of content distribution and navigation systems. His current interests are advanced media transport and ultra-realistic communications.

Shigeru Fujimura

Research Engineer, Natural Communication Project, NTT Service Evolution Laboratories.

He received the M.S. in information science and technology from the University of Tokyo in 2005 and joined NTT the same year. Since then, he has been engaged in research on web mining and web engineering, especially on effective implementation methods for web applications.

Atsushi Nakadaira

Senior Research Engineer, Natural Communication Project, NTT Service Evolution Laboratories.

He received the B.S. in physics from Kyoto University in 1992, the M.S. in physics from the University of Tokyo in 1994, and the Ph.D. in applied physics from the University of Tokyo in 2001. He joined NTT in 1994 and has been studying III-nitride semiconductors, display devices of polymer distributed liquid crystals, interactive systems of 3D displays, and visual communication systems. He is a member of the Japan Society of Applied Physics and ITE.

Personal Agents to Support Personal Growth

Yasuhiro Tomita, Tomohiro Tanaka, Tomohiro Yamada, and Kazuo Kitamura

Abstract

We at NTT Service Evolution Laboratories have been focusing on implementing a personal agent that will co-exist with us in the real world, empathize with us, encourage us to stay active, always be at our side to support our personal growth, and even grow on its own. In this article, we introduce sensing technology for understanding people and their environment; multimodal interaction technology for supporting people and helping them to grow; and virtualization technology for implementing device features that are used to continuously support people in any place or at any time.

Keywords: vital signs sensing, multimodal interaction, M2M

1. Introduction

Over the course of our lifetimes, we can acquire new skills, maintain our current skills even as we age, and be energized and motivated to action. We use the term *personal agent* to refer to an entity that understands our personal situations and helps us to grow. Although a personal agent is a lifelong assistant, we want it to push us to improve rather than become a source of our complete dependence. Our research and development efforts at NTT Service Evolution Laboratories have been focused on implementing just such a personal agent.

If we can build a personal agent like the one we just described, it will allow us to provide agent services that support and push people to improve over the course of their lifetimes, as shown in **Fig. 1**.

In order to do this, we need to understand the environment around the person being supported (using sensing technology) and to continuously support that person in his or her activities over a lifetime (using multimodal interaction technology). These, in turn, require a mechanism for taking in differences between the robots and other devices that serve as the physical manifestation of the personal agent, effectively accommodating and controlling a vast number of sensors, and providing an abstraction layer for applications that use these devices (device feature virtualization).

2. Sensing technology for inferring internal human state

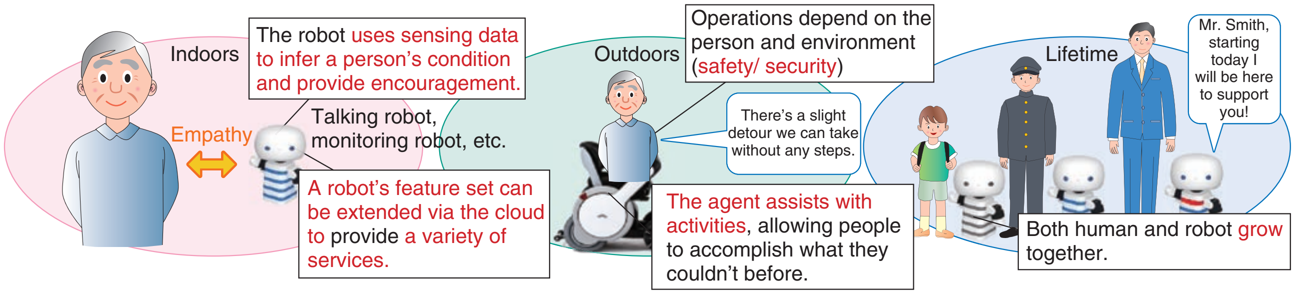
Pulse rates and electrocardiograms are useful biological data for understanding a person's psychological and physiological state. This kind of data is generally collected by means of sensors attached to the fingertips, arms, ears, and other parts of the body. An alternative method of naturally and unobtrusively collecting data is to use *hitoe*™,[*1] a smart fabric jointly developed by NTT and Toray Industries, Inc. By simply donning a hitoe shirt, the wearer can provide acceleration data and information on the electrical activity of his or her heart [1].

We have also developed a more conventional pulse wave sensor surface for people who would find it inconvenient to wear or attach sensors in advance. Unlike existing pulse wave sensors that take measurements at a single point, our sensor has been augmented with optical technology so that it can take measurements anywhere on a surface [2]. These expanded

---

[*1] *hitoe*™: NTT and Toray Industries, Inc. jointly developed the hitoe fabric using a conductive polymer coating (PEDOT-PSS) on a fibrous surface.

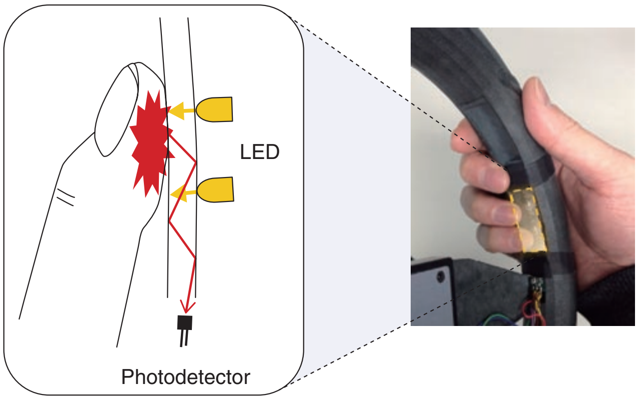
Services are implemented through a personal agent that will always stay close by you like a friend, allowing you to relax and rely on it anytime, anywhere—even as you age, circumstances change, and the personal agent itself takes different forms.



* The robot was created in collaboration with AOI Pro. Inc.

Fig. 1. Sample agent services.

Existing pulse wave sensors have been optically augmented to implement a wide sensor region.



LED: light-emitting diode

Fig. 2. Pulse wave sensing surface.

capabilities allow it to take measurements even when it is only lightly touched. For example, we could combine this sensor with a mouse or steering wheel to respectively measure a person's pulse wave data during ordinary computer use or while driving, as shown in **Fig. 2**.

However, because hitoe and surface-based pulse wave sensors take measurements during everyday activities, movement and other factors introduce noise into the data and thus make it more difficult to analyze than data measured by stationary, dedicated devices. To grapple with this problem, we are researching technologies that will remove noise from the measured data and very precisely quantify features such as heartbeats; we are also researching technologies that will estimate mental fatigue, changes in cognitive performance, sleep quality, balance and posture while walking, movement, and other properties from the sensor data and its characteristics.

We can also expect to roll out sports and healthcare solutions by employing these technologies for fatigue risk management, productivity improvements, sleep treatments, and activity tracking.

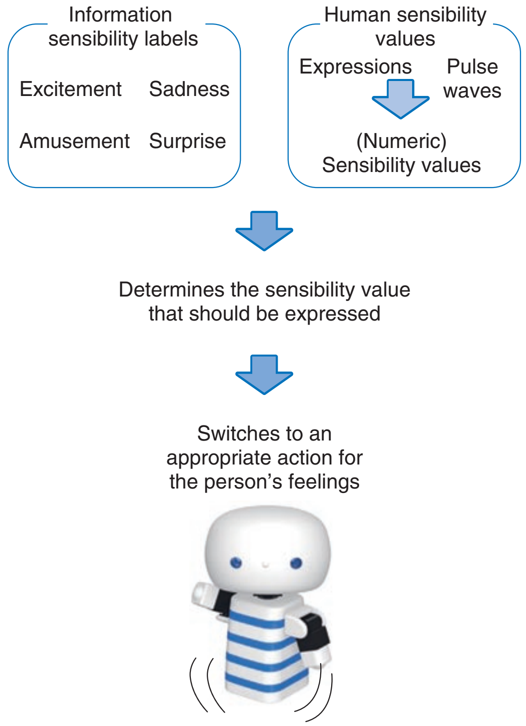
3. Multimodal interaction technology for helping people grow

In order for a personal agent to provide active encouragement that is based on personal and surrounding circumstances, it is important for the agent to use expressions and give feedback in ways that are appropriate to the time and place as well as each individual person.

Multimodal interaction utilizes a combination of sensing, voice recognition, speech synthesis, and interactive dialogue to understand human expressions and conditions; it then verbally or physically responds in kind through a robot or some other gadget, thus presenting information and content to the user. This makes it possible to put the user at ease, evoke emotions, and assist the user in his or her personal growth.

Here, we must also pause to consider how a person's age, generation, location, and situation will affect the type of robot or electronic gadget that he or she will come into contact with. Even if the type of robot or gadget changes, its control procedures and methods for interacting with humans will be implemented as a data model that can be personalized and stored for each user; this allows any robot or gadget to appropriately and seamlessly reassure, empathize with, and otherwise support the personal growth of the user.

One specific use case that we are considering is an intelligent robot[*2] that provides active encouragement to patients in their own homes as well as in nursing homes. As shown in **Fig. 3**, the robot would first use biological sensors to come to an understanding of the patient's physical condition; then, on the basis of that condition, it could use speech, gestures, and other forms of emotional expression to encourage conversation and keep dementia in check [3].



* The robot was created in collaboration with AOI Pro. Inc.

Fig. 3. Varying emotions caused by robot actions.

We are also considering ways to support people who are disabled or have physical impairments, in anticipation of futuristic wheelchairs and the spread of other personal mobility technologies. We are striving to allow people to enjoy going outside—even if they cannot walk about freely—with the help of a personal mobility vehicle, which will have its own personality and thus provide an easy, interactive way for anyone to learn how to drive it as well as providing directions to a destination. In order to collect subjective information that is difficult to interpret through sensors (such as physical sensations and anecdotes), we are considering methods for prompting the user to recall information during the course of natural conversation. By accumulating and extracting this information along with map data, we will be able to provide an optimal navigation experience for everyone, as shown in **Fig. 4**.

We are also exploring multimodal interaction technologies based on neuroscience and psychology in the field of sports and child development. Specifically, we anticipate a need for assistance with skill acquisition in sports and are therefore researching the possibility of a *coaching suit* that will detect muscle movements via biological sensors and then use force feedback, visual feedback, and audio feedback to teach each individual wearer appropriate form. After internalizing proper muscle and joint movements through sports practice done while wearing the coaching suit, athletes should be able to reproduce proper form even without the suit on.

4. Device feature virtualization to support personal agent functionality

To implement the aforementioned technologies, it will be important to efficiently manage robots and other gadgets (or a wider range of devices) along with the data they collect. It will also be important to make it easy for device functionality to be used over a network (via the cloud).

The basic architecture for implementing all of this

[*2] Intelligent robot: A robot that does not simply try to ingratiate itself with a patient; instead, it actively makes suggestions while taking individual circumstances into account.

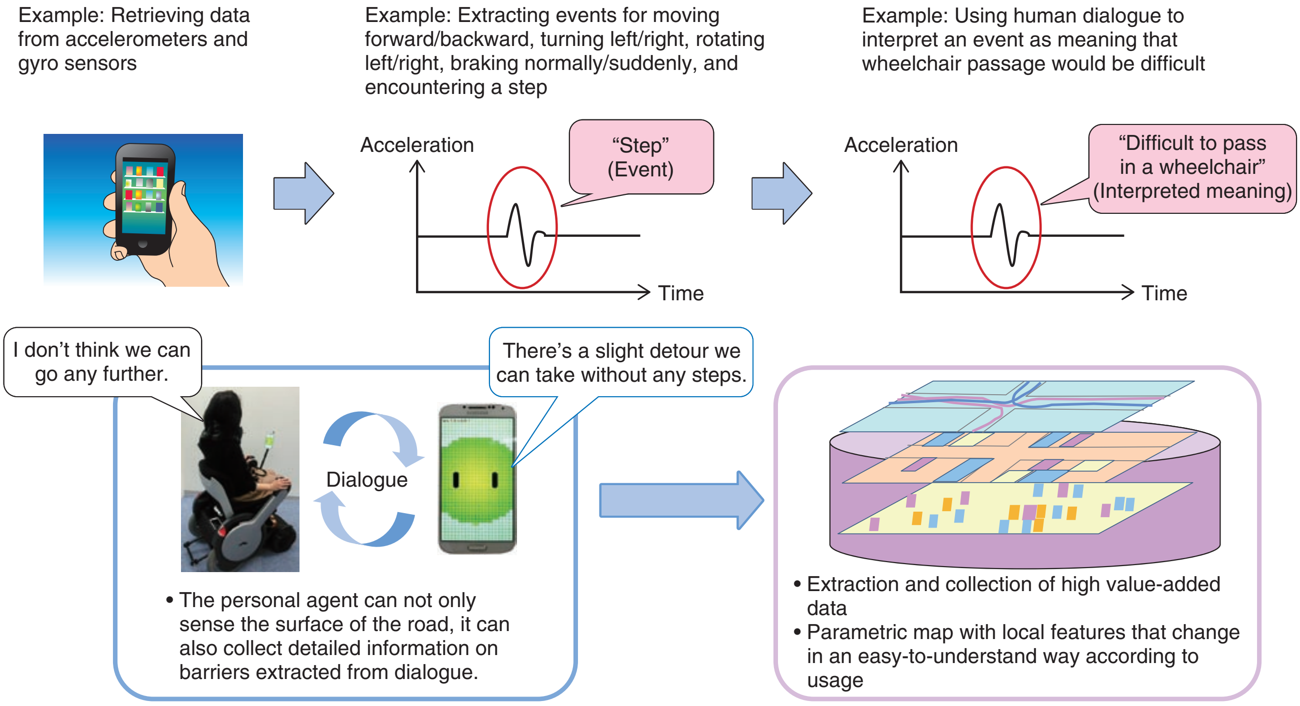


Fig. 4. Collecting subjective information through dialogue.

must make it easy to control each device and make use of data without being aware of the differing protocols for each individual device and industry domain. In many cases, practical data applications and device management policies have common requirements. Furthermore, services that are provided and implemented with a common architecture can be used in cross-sector applications and are thus desirable in terms of efficiency during system development.

As we examined this architecture, we also paid attention to trends related to the international standardization of machine-to-machine (M2M)[*3] technologies because the goal of international standardization is in line with our own efforts: to establish specifications for service platforms and system architectures applicable in many different industries [4].

oneM2M[*4] is an international standards organization for M2M technologies that is jointly run by telecommunications standards bodies in Japan, the United States, China, Korea, and Europe. The first version of an international standard—with prescriptions for architecture implementations—was officially released in January 2015. We have been referring to this standardized M2M architecture as we explore basic architectures for virtualizing device features. The standardized M2M architecture and an example application of device feature virtualization are shown in **Fig. 5**.

In device virtualization, the international M2M architecture standard is used for reference in defining physical device features as a logical device in a datacenter via an M2M-GW (gateway)[*5], which can then be controlled by applications via an application programming interface (API)[*6]. To get the most use out of this architecture, we are moving forward with a oneM2M standard proposal that will give common device model specifications for a variety of different industries.

In the future, we will augment devices with wearable sensors, robots, and other technology; support a wide variety of M2M area networks[*7]; and explore even more opportunities that will allow us to apply

---

[*3] M2M: A system that connects machines via a communication network so that they can exchange data autonomously.

[*4] oneM2M: An organization that was established in 2012 by the major telecommunications standards bodies in Asia, the United States, and Europe to undertake global M2M standardization efforts.

[*5] M2M-GW: A device that terminates connections from core/access networks and relays data to area networks.

[*6] API: An interface that provides functionality to client programs.

[*7] M2M area network: A short-range network that connects sensors and other M2M devices with M2M-GWs.

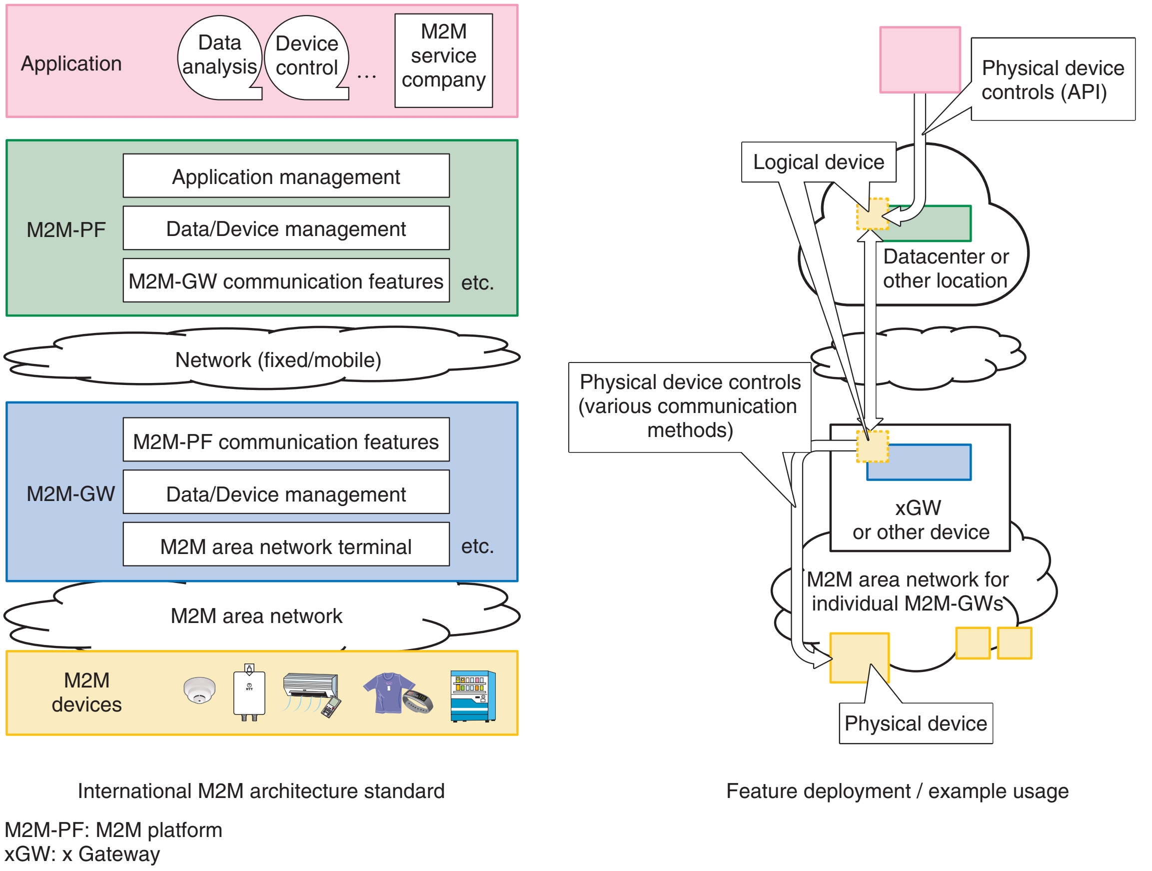


Fig. 5. Standard M2M architecture and sample application of device feature virtualization.

this technology to a wide range of services.

5. Future work

Personal agents will coexist with us in the real world—empathizing with us, encouraging us to act, growing together with us, and pushing us to take small steps in areas where we may have given up before. In this way, we believe that personal agents will enrich our lives.

In order to make personal agents a reality, we need a diverse collection of technologies: video processing to act as an agent's eyes; speech and language processing to act as an agent's mouth; artificial intelligence and big data analysis to act as an agent's brain and memories; and a robot to act as an agent's body. In addition to making use of the technologies being researched at NTT, we would like to partner with a large number of companies. Through co-innovation, we hope to create a personal agent that will help people grow.

References

[1] T. Ogasawara, K. Ono, N. Matsuura, M. Yamaguichi, J. Watanabe, and S. Tsukada, "Development of Applications for a Wearable Electrode Embedded in Inner Shirt," NTT Technical Review, Vol. 13, No. 1, 2015.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201501fa2.html

[2] H. Chigira, A. Maeda, M. Kobayashi, K. Fujiwara, T. Hiraoka, A. Tanaka, and T. Tanaka, "A Study on Heart Rate Monitoring in Daily Life by Using a Surface-type Sensor," SICE Journal of Control, Measurement, and System Integration, Vol. 8, No. 1, pp. 74–78, 2015.

[3] T. Matsumoto, S. Seko, R. Aoki, A. Miyata, T. Watanabe, and T. Yamada, "Affective Agents for Enhancing Emotional Experience," Proc. of the Second International Conference on Human-Agent Interaction (HAI 2014), pp. 169–172, Tsukuba City, Ibaraki, Japan, Oct. 2014.

[4] A. Tsutsui and Y. Goto, "Standardization Activities of oneM2M," NTT Technical Review, Vol. 12, No. 8, 2014.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201408gls.html

Yasuhiro Tomita

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

He received the B.S. and M.S. in physics from Tokyo Institute of Technology in 1987 and 1989, respectively. He joined NTT in 1989 and engaged in research on the development of visual communication systems, the planning of corporate network services and cloud services, and the business development of security and video communication technology. He has been in charge of the research project on robotics and communication gadgets since 2014. He is a senior member of the Institute of Electronics, Information and Communication Engineers (IEICE).

Tomohiro Tanaka

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

He received the B.S. and M.S. in mechanical engineering from Osaka University in 1987 and 1989, respectively. He joined NTT in 1989.

Tomohiro Yamada

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

He received the M.S. in electronics engineering from Niigata University in 1992 and the MBA from the University of Birmingham, UK, in 2007. He joined NTT in 1992, where he has been engaged in research on content distribution systems and technologies to support human-robot communication. He is the W3C (World Wide Web Consortium) Advisory Committee Representative of NTT and also a member of the Association for Computing Machinery (ACM), the Institute of Electrical and Electronics Engineers (IEEE), and the Information Processing Society of Japan (IPSJ).

Kazuo Kitamura

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

He received the B.S. and M.S. in materials science from Waseda University, Tokyo, in 1990 and 1992, respectively. He joined NTT in 1992.

Proactive Navigation Optimized for Individual Users

Jotaro Ikedo, Tsutomu Horioka, Yasuhiro Niikura, Yoshimasa Koike, Hiroshi Sawada, and Yuichi Muto

Abstract

In this article, we first give an overview of the proactive navigation project that NTT Service Evolution Laboratories is striving to accomplish. Then we provide a simple explanation of the technologies currently being researched and developed to implement the project: information navigation, big data analysis, and resilient communication. Finally, we give an overview of experiments currently being run in Fukuoka City along with their results.

Keywords: proactive, navigation, big data

1. Introduction

Over one million tourists visit Japan from abroad every month—a number that is projected to increase as we approach the year 2020. In fact, even domestic Japanese tourism is expected to grow in the coming years. Here at NTT Service Evolution Laboratories, we are doing our utmost to research and develop technologies—primarily for navigation—so that we can provide useful services to all of these visitors with a humble spirit of hospitality.

Rather than simply trying to provide navigation instructions from point A to point B, we are considering ways to provide information that could influence a visitor's activities (for example, suggestions for that day's dinner plans or tourist destinations for the next day). Our ideal proactive navigation system would be able to carefully select the most useful information for a customer on the basis of his or her tastes, past activity, and plans for the day—as well as the weather—to make suggestions before the customer even has to ask for them.

2. Technological components of proactive navigation

Every customer naturally has different needs and preferences. To implement proactive navigation, we need to understand or observe customers to intuit their wishes before we can determine what to suggest and then present it so it is easy to understand. We will probably also need to understand customers' circumstances and predict how those circumstances will change so that we can suggest the most appropriate activities for them. Furthermore, we believe that more than a few visitors from abroad will still have vivid memories of the 2011 Great East Japan Earthquake. Consequently, we feel that one of the tenets of hospitality involves providing appropriate information to visitors to ease any anxiety in the event of an emergency.

We plan to address these requirements by using three technologies to implement proactive navigation: information navigation, big data analysis, and resilient communication, as shown in **Fig. 1**.

2.1 Information navigation

To provide useful information to the customer, we first need to know where the customer currently is as well as whether the customer is stationary or in transit, and, if so, what mode of transport he or she is using. We have developed two technologies to obtain this information: location estimation and transport mode estimation, as shown in **Fig. 2**.

(1) Location estimation

Although the Global Positioning System (GPS)

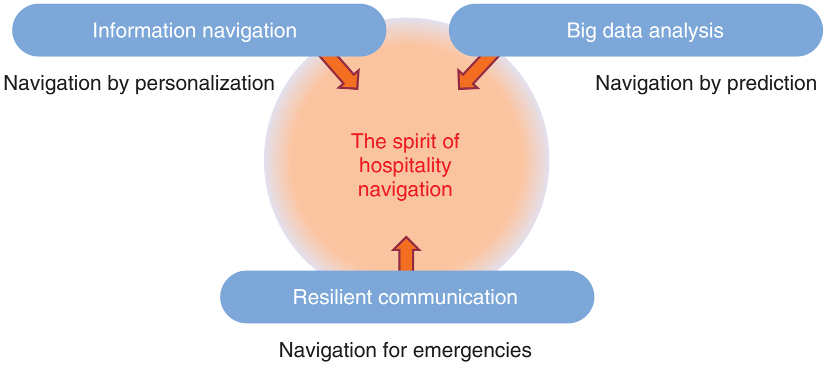
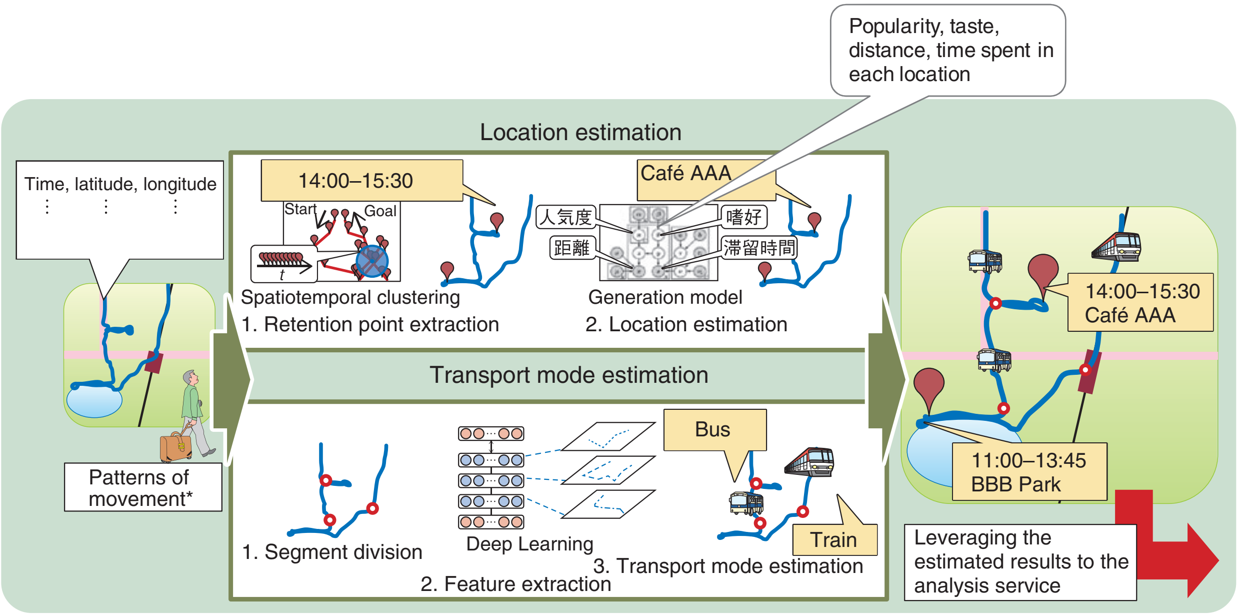


Fig. 1. Three technologies that constitute proactive navigation.



*Patterns of movement are obtained by GPS on cellular phones and smartphones.

Fig. 2. Location estimation and transport mode estimation.

allows us to very precisely pinpoint a smartphone's current location, it still has a margin of error of 5–30 m. This has made it difficult to pinpoint a customer's actual location while he/she is underground or in densely packed urban areas and buildings. Our location estimation technology looks at the probabilities associated with GPS measurements scattered in the temporal and spatial directions to accurately extract accumulated locations; then, using characteristics associated with various places such as stores or restaurants (for example, that a person is likely to stay in the same location for a period of 15–90 minutes to have a meal), the technology is able to improve the accuracy of a customer's estimated location. If we are also able to measure customer locations over a longer period of time, we can expect to further improve the accuracy of our estimates by applying individual customer trends, such as stores they visited in the past.

(2) Transport mode estimation

Using GPS tracking, we can determine a customer's traveling speed and thus to some extent guess that customer's mode of transport. However, pedestrians sometimes move at the same speed as cars stuck in traffic jams in urban areas, making such guesswork more difficult. Furthermore, we would often like to distinguish between similar modes of transport such as buses and taxis. In situations like these, simple rules on the basis of whether a customer's speed has exceeded a certain threshold are insufficient to make a decision. Our transport mode estimation technology therefore predicts patterns of movement using machine learning with the same deep neural networks as pattern recognition algorithms. This has allowed us to very precisely estimate modes of transport.

With these technologies, we can provide tourist information tailored to each individual's situation. For example, we can suggest nearby tourist spots to customers when they are likely to be finishing a meal at a restaurant, or send information on upcoming points of interest along a train route that customers are riding on.

2.2 Big data analysis

At popular tourist spots and stadiums filled with excited fans, there is a danger that a crush of crowds moving in the same direction could lead to serious accidents. Our big data analysis technology examines measurements on where people congregate in order to predict what events will transpire and when. For example, by taking measurements of crowd density distributions and extrapolating up to approximately an hour into the future, we can anticipate mass confusion before it occurs and thus suggest the appropriate actions to prevent it. However, by intervening in this way, we will also cause our original predictions—which did not account for such intervention—to become inaccurate. As a result, to keep up with actual conditions, it will be necessary to continually iterate between taking measurements, making predictions, and suggesting actions in real time. We are currently pursuing research and development (R&D) efforts to effectively accomplish this processing loop in real time.

Our R&D efforts also encompass the technological building blocks for making suggestions that meet each individual customer's needs by intimately understanding the customer's point of view and even considering factors that he or she has not consciously noticed. So far we have established technologies for recommending routes to tourists (e.g., from point A to B to C); technologies for determining whether a visitor's trip is routine or unusual on the basis of an analysis of the trip's time of day, day of the week, and location; and technologies for analyzing multidimensional composite data to determine when and where it would be easy for someone with particular attributes to easily visit. These technologies allow us to provide fine-grained navigation appropriate for individual customer circumstances.

2.3 Resilient communication

Because Japan is a country where natural disasters occur frequently, we feel that our navigation system must convey a sense of security that it will get a customer to safety whenever and wherever disaster strikes.

It is now common for many ICT (information and communication technology) services to assume that client devices are always connected to the cloud. However, power and Internet connections can be cut off in the event of a disaster, preventing people from using cloud-based services. In situations such as these, our resilient communication technology allows end-user devices to communicate with each other using only Wi-Fi and an HTML5* -compliant browser—even without a dedicated application—as shown in **Fig. 3**. Furthermore, our implementation will allow devices to be carried between disconnected Wi-Fi service areas to relay information back and forth and make continued communication possible.

Next we plan to target facilities of local municipalities that are expected to be used as evacuation centers in times of emergency. Our goal is to use only the equipment available at these evacuation centers to implement a communication system that provides the essential function of checking on the safety of evacuees as well as a system that can be used to build infrastructure for supplying information.

In 2015, we are planning to conduct a trial of application services that can achieve the minimum level of individual safety reporting when evacuation centers lose power and network connections. Following the trials, we will work to address any issues that arose.

3. Examples of current work

Here, we introduce a regional revitalization trial that has been ongoing in Fukuoka City since October 2014, involving tourism services for visitors from abroad. The trial's primary participants are the JTB Group, a leading travel agency in Japan, and the NTT Group. During the trial, we provided both foreign and domestic visitors to Fukuoka City with free tourist guide applications and services that can be used by connecting to the Internet via Wi-Fi. Then, with the visitors' mutual consent, we collected information

---

* HTML5: hypertext markup language version 5

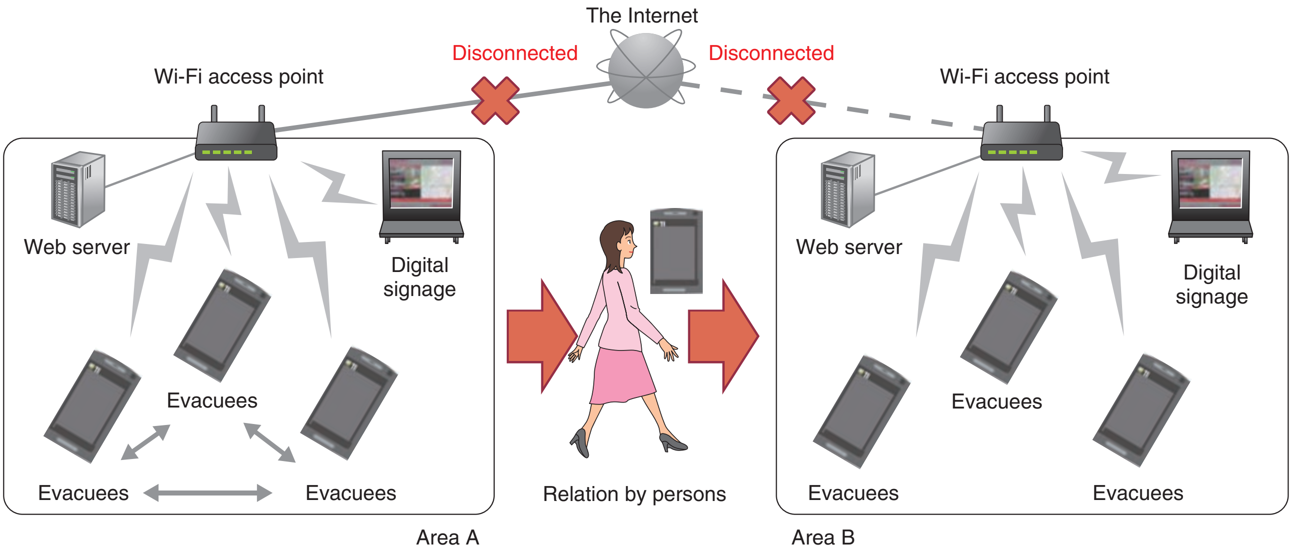
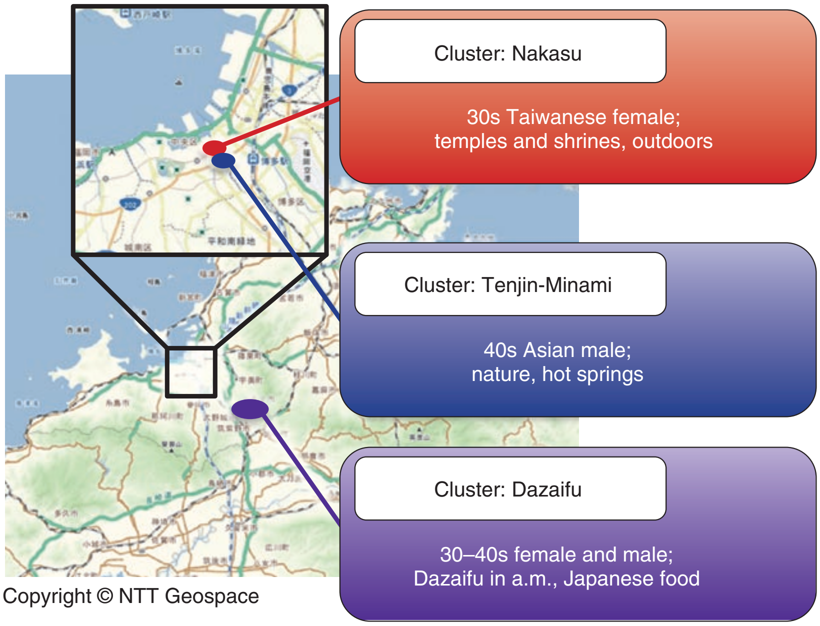


Fig. 3. Resilient communication.



Copyright © NTT Geospace

Fig. 4. Sample results of analyzing various visitor activities.

related to their usage history and applied multidimensional composite data analysis techniques to it (**Fig. 4**). This information allowed us to find groups (clusters) of people visiting typical tourist sites as well as differing characteristics between groups in the northern and southern parts of the downtown area. By identifying groups with different characteristics, we can provide information tailored to individual customers on the basis of the characteristics of the group that the customer is presumed to be traveling with.

In the future, we will strive to make specific predictions of when and where events will happen on the basis of our analysis of the data to provide even more useful information to customers.

4. Future plans

Here at NTT Service Evolution Laboratories, we do not believe that the technologies we have presented in this article are yet sufficient to achieve our vision of hospitality; there are still areas that need further attention. We will endeavor to promptly establish R&D projects that address these areas—or collaborate with people both inside and outside our company—and thus realize the spirit of hospitality through proactive navigation.

Jotaro Ikedo
Project Director, Proactive Navigation Project, NTT Service Evolution Laboratories.

He received the B.E. in electronic engineering and the M.E. in electrical engineering from Kogakuin University, Tokyo, in 1989 and 1991, respectively. Since joining NTT in 1991, he has been engaged in R&D of low-bit-rate speech coding and wireless transmission, and in developing VoIP/videoconference systems. He has recently been managing an R&D project focusing on big data analysis methods and geographic information systems. He is a member of the Institute of Electrical and Electronics Engineers (IEEE), the Institute of Electronics, Information and Communication Engineers (IEICE), and the Acoustical Society of Japan (ASJ).

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

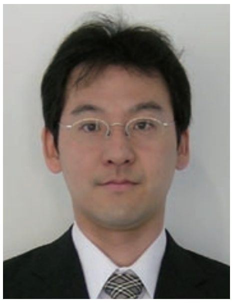
He received the M.I. in information systems from Japan Advanced Institute of Science and Technology, Ishikawa, in 1995. Since joining NTT Human Interface Laboratories, he has been engaged in developing a secure content distribution system. During 2004–2005, he was a visiting scholar at Stanford University, USA, where he studied peer-to-peer content distribution methods from a juristic perspective. He is currently in charge of developing an analytical system for big data.

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

He received the B.E. in electronics and electrical engineering from Keio University, Tokyo, in 1994. He joined NTT Human Interface Laboratory in 1994 and studied video and image processing. He worked at NTT Communications from 1999 to 2004, where he developed the Media Asset Management system for TV broadcasters. During 2004–2011, he worked at NTT EAST producing customer premises equipment. He is currently studying a home and office network service that handles many kinds of customer premises equipment.

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

He received the M.E. in material chemistry from Tohoku University, Miyagi, in 1989. He joined NTT Information Processing Laboratory in 1989 and studied pattern recognition and computer aided instruction. During 1999–2005 he worked at NTT Resonant, where he developed a web-based learning system. He also worked at NTT Advanced Technology, where he built a support system for a virtual private server service. He is currently studying geographic information systems.



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

He received the B.E., M.E., and Ph.D. in information science from Kyoto University in 1991, 1993, and 2001, respectively. He joined NTT in 1993. His research interests include statistical signal processing, audio source separation, array signal processing, machine learning, latent variable models, graph-based data structures, and computer architecture. From 2006 to 2009, he served as an associate editor of the IEEE Transactions on Audio, Speech & Language Processing. He is an associate member of the Audio and Acoustic Signal Processing Technical Committee of the IEEE Signal Processing Society. He received the Best Paper Award of the IEEE Circuit and System Society in 2000, the SPIE ICA Unsupervised Learning Pioneer Award in 2013, and the Best Paper Award of the IEEE Signal Processing Society in 2015. He is a member of IEEE, IEICE, and ASJ.

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

He received the B.E. and M.S. in information sciences from Tohoku University, Miyagi, in 1998 and 2000, respectively. He joined NTT Service Integration Laboratories as an engineer in 2000 and engaged in development activities in the Communication Traffic & Service Quality Project. From 2004 to 2009, he worked at the R&D center of NTT WEST and then went back to NTT Service Integration Laboratories. He is now with NTT Service Evolution Laboratories and is involved in promoting and producing NTT R&D products.

Towards the Creation of Attractive Services Based on an Understanding of Users

Yoko Asano, Takehiko Ohno, Shinji Miyahara, Masahiro Watanabe, Mutsuhiro Nakashige, Atsunobu Kimura, and Kae Wago

Abstract

To create information and communication technology services that are fun for users, it is necessary to ensure that such services match user needs and characteristics, and to do this, we need to fully understand users. We introduce here the processes involved in service creation and the design area that NTT Service Evolution Laboratories is focusing on. We also show a specific example of what we do to understand the user and visualize a service image.

Keywords: service design, human-centered design, design thinking

1. Introduction

1.1 Human-centered design

The NTT Group is continually striving to create attractive information and communication technology (ICT) services for users. An attractive service is a service that users will want to use and will continue to enjoy using. The methodology of creating such services is called *service design* [1], and it has attracted a lot of attention in leading companies. NTT Service Evolution Laboratories has been conducting research on the process and specific methods of service creation. We have also been working on the development of support tools and related guidelines, as we believe that service design will become more and more important in the future.

1.2 Process of service design

The process and the target area of service design are illustrated in **Fig. 1**. The process of human-centered service design begins by acquiring a deep understanding of different kinds of people. It is especially important to estimate the needs of users who are expected to use the service. The needs of users can be clarified by observing and finding out about their lifestyle, values, hobbies, and attitudes toward ICT services, among other things. For some services, it is also necessary to understand the relevant environmental and social background.

Obtaining such information makes it possible to create a service concept. In this phase, we use various techniques such as holding workshops to generate excellent ideas. In this approach, we focus on obtaining data in order to understand people and society and use the data to generate a number of ideas. We then gradually narrow down the ideas.

We believe that the user is best understood by the user him/herself. Therefore, the approach may involve a user participatory design approach such as obtaining user input from the early phases of service design. We continue to gradually embody the concept born in this phase as a service image, and we use a technique to actually draw the concept as a story, for example, as a three-frame comic.

By interviewing target users, we are able to review the service concept to make sure that the concept

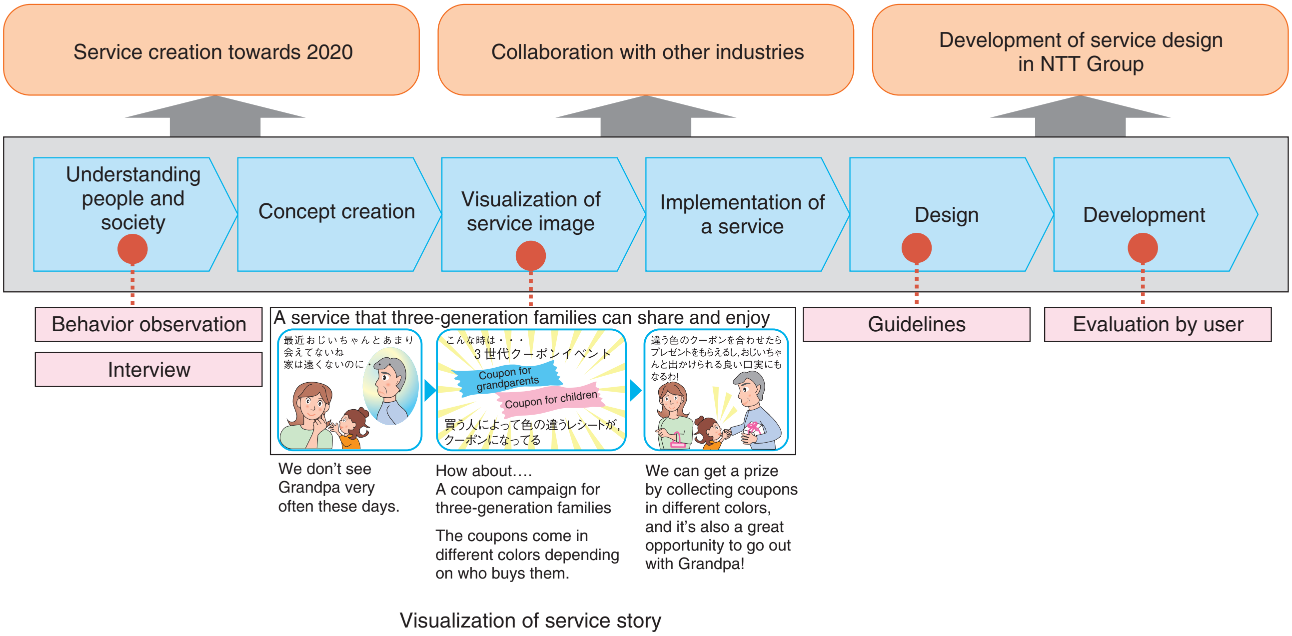


Fig. 1. Process and main target area of human-centered service design.

matches the needs of users. If the service concept does not seem to match user needs, we can modify it. It often happens that service developers determine that an idea does not match the needs of the user very well. As a result, the user may not want to use the service. To prevent this, it is necessary to interview potential users at an early stage so as to align the service idea to the user needs. In this way, the idea can gradually be refined while repeatedly obtaining user feedback, and an appealing service can be created that users will want to use.

1.3 Designing specific services

After the user needs are clarified, we start designing a specific service. In this phase, we also consider the design of the screen transitions and the user interface, as they will greatly affect the ease of use. To streamline this process, we assembled guidelines on user characteristics that summarize the points necessary to create easy-to-use services.

When developing a service, we conduct user evaluations as often as possible to see if the users are able to use the service as we intended [2]. If some aspect of the service seems difficult for the user to use, we modify the user interface. Fixes carried out in a more advanced phase of service development will increase costs. Therefore, when considering the screen specifications, it is important to evaluate them in the early stages as much as possible by using a technique such as prototyping on paper.

NTT Service Evolution Laboratories has so far focused on consumers as the target area of service design. In the future, we will further develop the service process in order to create hospitality services looking towards 2020 and to create the B2B2X (business-to-business-to-X) service through collaboration with other industries. We will accumulate information on the characteristics of target users and will continue to work on creating guidelines focusing on user characteristics. In addition, we plan to establish a system to strengthen the service design process for the entire NTT Group.

2. Research methods to understand the user

In human-centered service design, it is important to understand the characteristics of the user. To understand such characteristics, we typically conduct behavioral observations and user interviews.

2.1 Behavior observation

Behavior observation is an ethnographic technique of studying a society or culture. In this technique, we visit places where the target user is active and observe the user behavior. We also surmise the reasons for the behavior of the user. As a result, we derive the user's needs and values, and also the challenges the user may potentially face. Here, we introduce an example of behavior observation targeting foreign tourists.

In this observation, we focused on the behavior when foreigners were in contact with people and things, and we observed characteristic actions such as engaging in dialog with local people and fellow travelers. In this observation, we also focused on actions that appeared to be unrelated, as it is important to observe the behavior without forming any preconceived ideas. We recorded anything we noticed in the observations in a notebook along with the time and the surrounding environment. Here, the user actions and the viewer's interpretations are both recorded.

As a result of observing foreigners, we found that they exhibited some interesting behavior before and after purchasing activities. For example, in a Tsukiji sushi shop, foreign tourists enjoyed not only eating sushi but also watching the chef prepare the sushi. In addition, foreigners also enjoyed conversing with the clerk and the surrounding customers. We also observed a case in which a chef who did not speak any foreign language displayed the ingredients on a tablet in order to give detailed information on the sushi. In this observation, we found that foreign tourists are often interested in things that do not typically interest Japanese people and that they find pleasure in casually communicating with Japanese people.

2.2 Interviews

There are a variety of approaches when conducting interviews. In our approach, we interview users about their past experience in order to extract their needs and challenges. To know the users deeply, however, we do not just listen to their answers to our questions that we prepared in advance. Rather, we ask the reasons for the action the user took or how the user felt in a certain situation. We also ask them whether or not they had experienced another situation where they felt the same way. In addition, questions that can only be answered Yes or No are useless. It is necessary to ask open questions in order to obtain more diverse answers. For example, when listening to a user's story about going to a tavern, we should not just ask whether the tavern was fun. It is better to ask the user to explain their feelings by instructing them, "Please tell me anything you felt when you went to the tavern." This will lead to a more relevant and substantial interview.

Here, we introduce examples of interviews with foreigners living in Japan and foreigners traveling in Japan, which we conducted in order to understand more about the foreigners' lives. We interviewed foreigners living in Japan about the following items:

- Opportunities they are interested in experiencing in Japan
- Changes in their impressions of Japan
- Experience when coming to Japan for the first time
- Current life

These interviews revealed useful information on the daily lives of foreigners and how they feel. In contrast, we mainly interviewed travelers about the following:

- Behavior while traveling
- Background of actions
- Habits

The results revealed information about how travelers feel and act. For example, one traveler remarked, "In convenience stores in Japan, the packaging of goods is not the same as in my home country. Therefore, I could not buy an item I was looking at because I did not know the contents." As a result, we found that foreigners find some things inconvenient that are a matter of course for Japanese people. We gather the data from these behavioral observations and interviews and organize the data into interesting points and anticipated needs. The results enable us to learn more about the true nature and behavior of foreigners without stereotyping them according to a previous image. For example, in one behavioral observation, we observed that foreigners do not use their smartphones outdoors as often as Japanese people do. Therefore, we conducted an interview and found that foreigners generally use their smartphones only at free Wi-Fi spots such as guesthouses. Because the cost of mobile services in Japan is high, the foreigners do not have a mobile service contract. It is therefore possible that services combined with free Wi-Fi would strongly appeal to foreigners. In this way, the resulting data can be used to construct an image of the service user. We can also use the data when conceiving an idea for a service.

3. Service design of Video Nabitto

We introduce here Video Nabitto as a specific example of service design. This service was initiated as a trial by NTT WEST in October 2014 [3].

Video Nabitto was developed with the aim of improving the convenience of the Hikari BOX+ set-top box provided by NTT WEST. This service enables users to search for videos via audio input on a smartphone or tablet, to display the search results on

TV combined with a smartphone results
in a richer communication tool.

From personal smartphone to the family smartphones

Fig. 2.   Service concept.

television (TV) with the Hikari BOX+, and to view the videos on the big screen. This service was achieved by combining NTT's voice interaction platform technology that can narrow down search criteria by vocal input and multi-screen technology to display video on the screen of different devices such as smartphones, tablets, and TVs. In addition, this service enables users to reduce the number of operations for video searching and to find the videos they want to watch right away so that they can enjoy YouTube videos on a large-screen TV.

We contributed to the concept creation, visualization of the idea, and the prototype evaluation. We first interviewed multiple involved parties that were close to the target user image and investigated their backgrounds and needs. We conducted a mini-workshop based on the survey results and created three service ideas. Then we interviewed users to confirm their needs and ways of utilizing such a service. Finally, we evaluated the service idea, brushed it up based on the evaluation results, and created a single concept (**Fig. 2**). To further crystallize the idea, we created service usage scenarios and a draft of a user interface (**Fig. 3**). We also asked potential users to evaluate a developed prototype, and then we improved the prototype by identifying what areas were difficult to use. The prototype was used to implement an actual trial. By repeating the evaluation from the initial steps of the service design, we created a service that matches the needs of users who want to easily watch video content on TV. Such a process is often time-consuming and expensive, but this time it was possible to carry out a minimal investigation in a short time and to prevent the development time from being significantly extended.

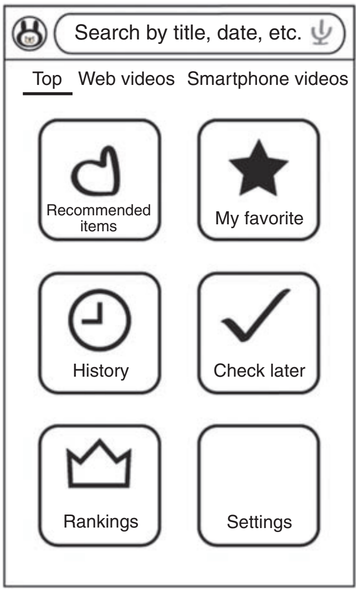


Fig. 3.   Proposed user interface.

4. Universal web design initiatives

By furthering our understanding of users, we will be able to provide information in accordance with user characteristics. We have been working thus far to conduct research and create guidelines on universal web design that is intended to convey easy-to-understand information to a variety of people, including the elderly and people with disabilities.

Web content is used by different kinds of users in different environments. There is not only one form of

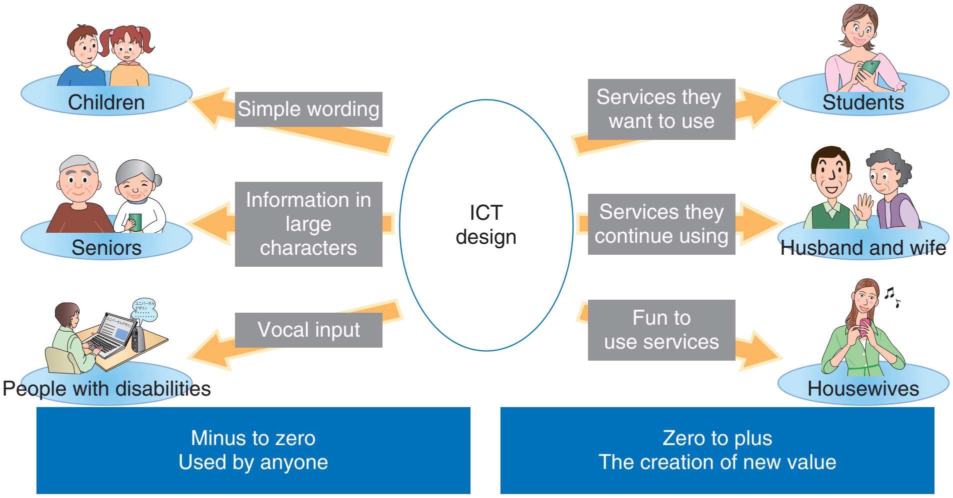


Fig. 4. ICT service design tailored to the user.

information (**Fig. 4**). In universal web design, the content is displayed using simple wording for children and displayed in large characters for seniors. In addition, text-to-speech software, which represents information by sound, can be used for disabled users. By devising the web content based on such design criteria, it can be changed in a variety of ways to suit specific users.

In addition to web content, ICT services also have a feature enabling information to be presented and changed flexibly. Changing the representation according to the user makes it possible to develop more attractive and easy-to-use services.

Before creating guidelines, we collected knowledge on user characteristics by obtaining ratings of various services. User characteristics include not only physical characteristics of the user, but also their lifestyle, values, hobbies, and attitudes about ICT services. We then created the guidelines based on the obtained characteristics. These guidelines can then be used by people who design services.

It is important that anyone can easily create services for people who have difficulties in using information technology such as the elderly and people with disabilities. The guidelines will make it possible to create services for such users. In addition, we are also working on applying the resulting knowledge to the development of hospitality services towards 2020.

5. Future work

NTT Service Evolution Laboratories will continue to study the design of ICT services that users will want to use and will continue to use. By leveraging people-friendly design, we can develop easy to understand and attractive services for users and can contribute to enhancing people's lives and work.

References

[1] M. Stickdorn and J. Schneider, "This is Service Design Thinking: Basics, Tools, Cases," Wiley, 2012.

[2] M. Nakatani, M. Yasuoka, and T. Ohno, "Service Design Which is Built upon User Understanding—How to Utilize User Data for Designing Service—," Proc. of HCG Symposium 2013, Matsuyama, Japan, December 2013.

[3] Website of NTT WEST, FLET'S, Video Nabitto (in Japanese), http://flets-w.com/douga-navitsan/

Yoko Asano

Executive Research Engineer, Project Director, Universal UX Design Project, NTT Service Evolution Laboratories.

She received the B.E. in administration engineering from Keio University, Kanagawa, in 1988 and joined NTT Human Interface Laboratories the same year. She moved to NTT Cyber Solutions Laboratories (now NTT Service Evolution Laboratories) in 1999. Since then, she has been conducting research on human interfaces. She is a member of the Human Interface Society (HIS), the Japan Ergonomics Society, and the Institute of Electronics, Information and Communication Engineers (IEICE).

Takehiko Ohno

Senior Research Engineer, Supervisor, Universal UX Design Project, NTT Service Evolution Laboratories.

He received the B.Sc. and M.Sc. from Tokyo Institute of Technology in 1992 and 1994, respectively. He joined NTT Basic Research Laboratories in 1994 and studied cognitive science and human-computer interaction. He has been researching human-computer interaction, human-centered system design, user experience design, usability, gaze tracking technology and its applications, cognitive modeling, information appliances, and computer-mediated communication. He is a member of the Association for Computing Machinery, the Information Processing Society of Japan, the Japanese Cognitive Science Society, and IEICE.

Shinji Miyahara

Senior Research Engineer, Universal UX Design Project, NTT Service Evolution Laboratories.

He received the M.E. in information system engineering from Osaka University in 2000. Since joining NTT in 2000, he has been working on human interfaces.

Masahiro Watanabe

Senior Research Engineer, Universal UX Design Project, NTT Service Evolution Laboratories.

He received the B.E. and M.E. in mechanical engineering and the Dr.E. in electronics, information, and communication engineering from Waseda University, Tokyo, in 1991, 1993, and 2003, respectively. He joined NTT Basic Research Laboratories in 1993 and studied the human auditory system. Since moving to NTT Cyber Solutions Laboratories (now NTT Service Evolution Laboratories) in 1999, he has been studying human-computer interaction, especially universal design and user-centered design.

Mutsuhiro Nakashige

Senior Research Engineer, Universal UX Design Project, NTT Service Evolution Laboratories.

He received the Dr.E. in advanced interdisciplinary studies from the University of Tokyo in 2005. Since joining NTT in 2005, he has been studying human-computer interaction. He received the Excellence Interactive Presentation Award by the Human Communication Group Symposium in 2013. He is a member of HIS and the Japanese Society of Child Science.

Atsunobu Kimura

Senior Research Engineer, Universal UX Design Project, NTT Service Evolution Laboratories.

He received the M.E. and Dr.E. in information processing from Nara Institute of Science and Technology in 2003 and 2007, respectively. Since joining NTT in 2003, he has been working on human interfaces.

Kae Wago

Researcher, Universal UX Design Project, NTT Service Evolution Laboratories.

She received the B.A. in agricultural science from Kyoto University in 2009. Since joining NTT WEST in 2009, she has been working on human interfaces. She received the Excellence Interactive Presentation Award by the Human Communication Group Symposium in 2015.

Creating *Omotenashi* Services for Visitors and Spectators in 2020

Naoyoshi Kanamaru, Hirohisa Tezuka,
Atsushi Fukayama, Yukihiro Nakamura,
Hitoshi Yamaguchi, and Manabu Motegi

Abstract

NTT Service Evolution Laboratories is working to develop services for people from both within and outside the country who will visit various places in Japan. This includes the spectators and organizers who will be in Tokyo for the 2020 Olympic and Paralympic Games. This article describes four conceptual *Omotenashi* services that will be available on smartphones and tablets to provide people with spectator and transportation assistance.

Keywords: service concepts, PoC, epoch-making

1. Introduction

What we aim to do is to provide new ICT (information and communication technology) services to give positive and memorable experiences and impressions to people from all over the globe, including the spectators and organizers of the 2020 Tokyo Olympic and Paralympic Games. The four conceptual services described here employ media technologies to deploy new services to assist not only those who will be attending the Games but also those who organized them.

2. Travel support service through multilingual translation and location identification

The first service is intended to enable visitors to Japan to use transportation facilities smoothly and with a minimum of problems. It will provide collected and selected information on locations and other factors that will help people reach their destination by making suitable transportation connections or taking detours to avoid traffic congestion or blockages due to accidents.

The service will also use image recognition and location prediction functions to collect and provide information that visitors will need. The service's translation function will enable visitors to receive the information on their smart devices in English, Chinese, and Korean (**Fig. 1**) [1].

The service will enable visitors to use their smart devices to take photos of information displayed on signboards and other places, after which the device will show them the right way to go in order to reach their destination. They will also be able to get pictorial information about special requirements such as those for disabled persons or people traveling with children. Urgent messages such as wrong-way warnings and emergency alerts will be issued as needed.

3. Personalized navigation service through object recognition

The second service will allow visitors to use their smart devices to recognize objects around them to guide them. NTT's proprietary recognition technology rapidly and robustly extracts 3D (three-dimensional) objects, after which the service tailors the object information to individual users on the basis of factors such as their age, gender, language, location, and travel history, as well as on the actions and preferences of similar users. This enables visitors to obtain personalized, interesting, and useful tourist information on what they are watching or looking at (**Fig. 2**) [1].

The service will provide visitors with personalized

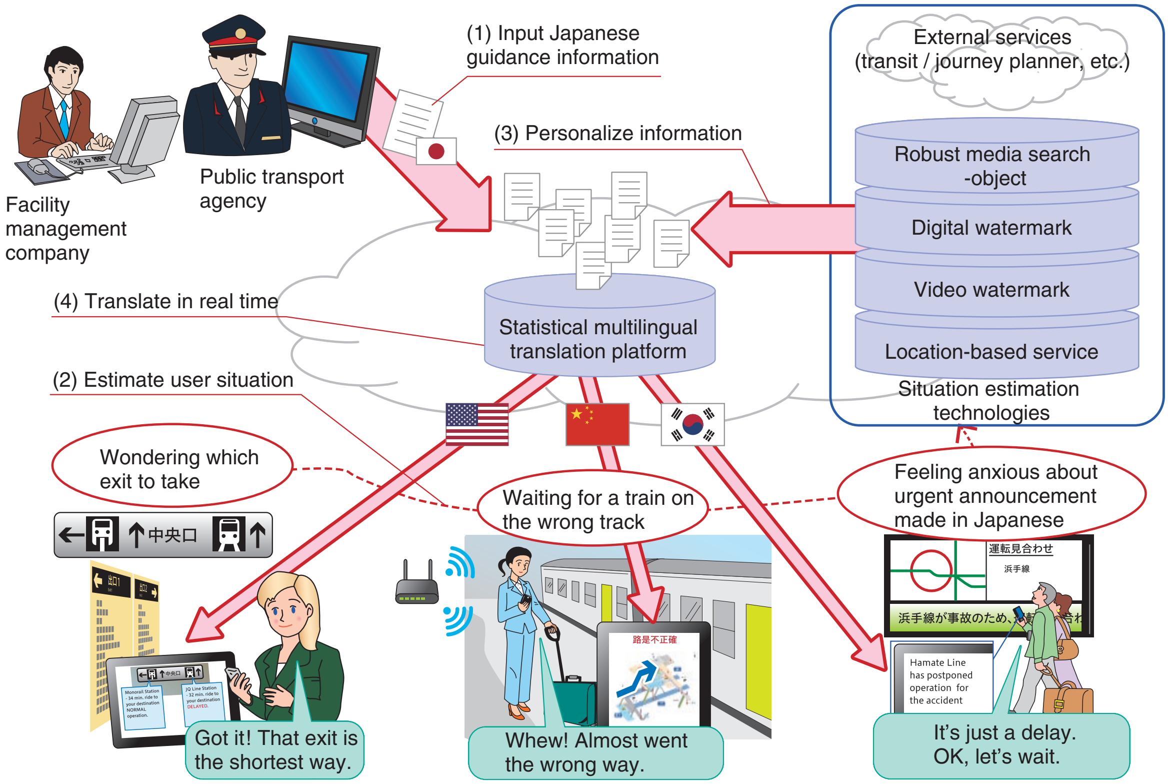


Fig. 1. Travel support service through multilingual translation and location identification.

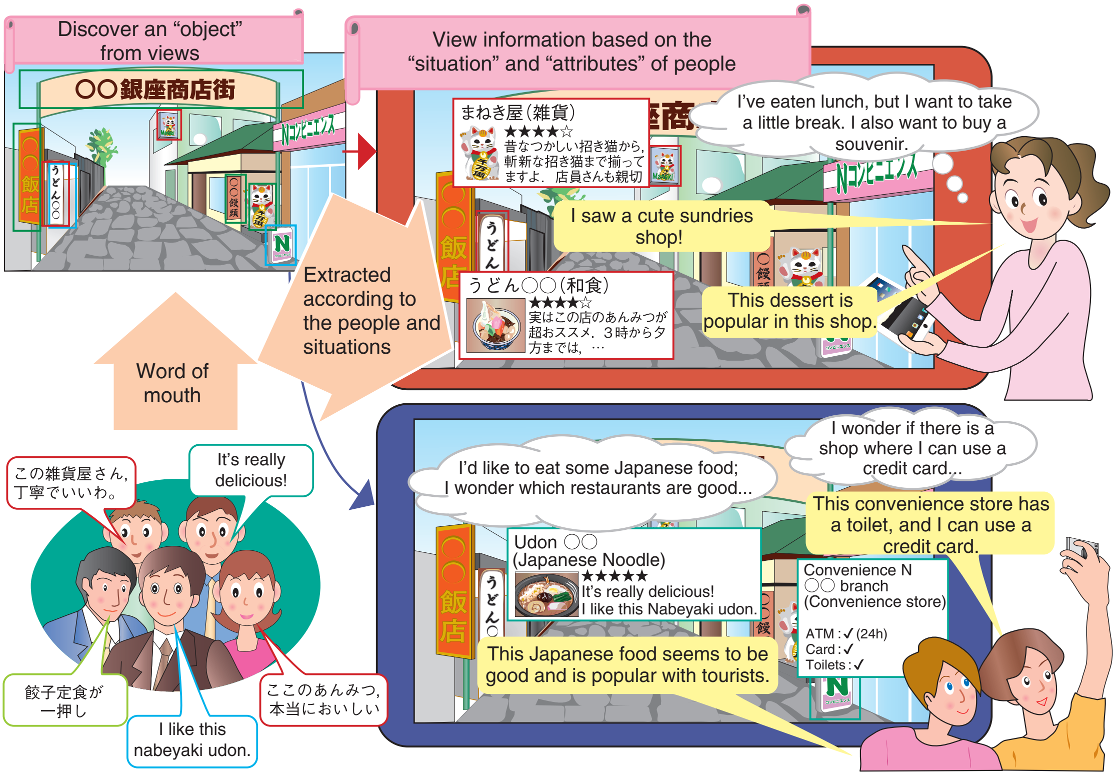


Fig. 2. Personalized navigation service through object recognition.

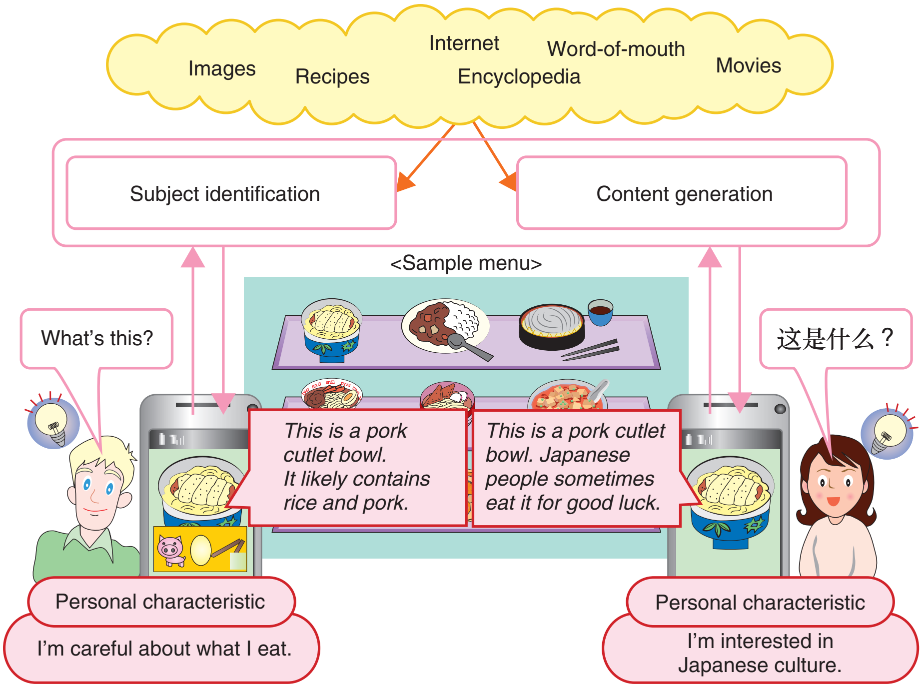


Fig. 3. Personalized information service through subject identification.

guide information on tourist spots on the basis of, for example, their cultural backgrounds. It will also provide them with personalized instructions on dealing with heavy crowds at stadiums or auditoriums on the basis of their preferences, and personalized information to navigate them to booths at exhibitions and to stores in shopping streets or malls.

4. Personalized information service through subject identification

The third service will instantly provide visitors with information on Japanese food and other goods to meet their needs and preferences. This service combines extended SightX [2] technology with the subject identification function and will enable visitors to more fully enjoy their stay in Japan.

Once a picture of various foods and goods is taken by a visitor's smart device, it can be easily identified as subjects in an image database that can be put together with minimal effort. The result the database provides will be composed from related information on the Internet that is personalized according to the visitors' preferences, cultural backgrounds, and other factors. If the information is available only in Japanese, it will be translated into the visitor's language (**Fig. 3**) [1].

The service provides explanations about food in restaurants and goods in stores in the visitor's language. It will also give further related information including the history, cultural backgrounds, and habits of these items if visitors desire. It also shows ingredients and allergens in foods for those concerned about health, dietary, and regional restrictions, and gives opinions on the dishes and any trending information about them.

5. Personalized sports spectating through information on selected athletes

The fourth service will provide timely information about a selected athlete on a smart device in a selected language. The information includes images automatically gathered from the Internet and social media by using CGM (consumer generated media) Automatic Zapping and organized according to scenes. Organizers of sports events can archive and edit the uploaded content and the images in it to use them as further business resources even after the sporting event has ended (**Fig. 4**). The service will enable spectators along a marathon race course, for example, to watch video images of a selected runner from the viewpoints of other spectators at different locations along the course.

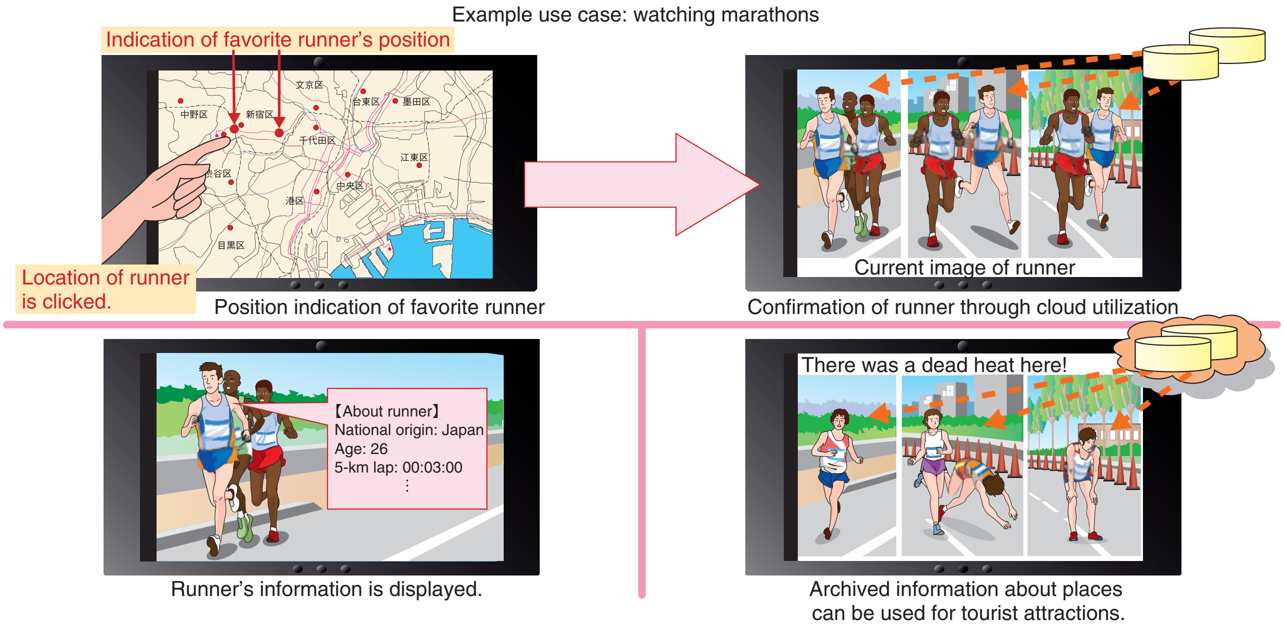


Fig. 4. Personalized sports spectating through information on selected athletes.

6. Future work

This article described four conceptual services targeted for use in 2020. For the next step, we hope to work with partners in the transportation and tourism industries to achieve co-innovation, that is, to implement real services with partners and users through service trials executed for proof of concept (PoC) purposes. Development and trial environments (XFARM [2], GEMnet [3]) will be prepared and improved to carry out the trials. Our objective will be to provide advanced and impressive services by executing the quick create-test-demolish cycle for proposing service concepts and their PoC.

References

[1] Feature Articles: Media System Technology for Creating an Appealing User Experience, NTT Technical Review, Vol. 13, No. 4, 2015. https://www.ntt-review.jp/archive/2015/201504.html

[2] Feature Articles: Activities Toward Service Incubation to Speed Up R&D of Service Systems, NTT Technical Review, Vol. 11, No. 9, 2013. https://www.ntt-review.jp/archive/2013/201309.html

[3] M. Shimomura, A. Masuda, H. Uose, N. Inoue, and N. Nakayama, "GEMnet2 R&D Testbed Network," NTT Technical Review, Vol. 11, No. 1, 2013. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201301ra4.html

Naoyoshi Kanamaru

Senior Research Engineer, Supervisor, Project Director, 2020 Epoch-making Project, NTT Service Evolution Laboratories.

He received the B.E., M.E., and D.E. in engineering from Tohoku University, Miyagi, in 1987, 1989, and 1992, respectively. He joined NTT in 1992. His research interests include human-computer interaction. He is a member of the Robotics Society of Japan (RSJ) and the Japanese Association for an Inclusive Society.

Hirohisa Tezuka

Senior Research Engineer, Supervisor, 2020 Epoch-making Project, NTT Service Evolution Laboratories.

He received the B.E. and M.E. in engineering from Kyoto University in 1991 and 1993, respectively. He joined NTT in 1993. His research interests include system management, networked robotics, and lifelog systems. He is a member of the Society of Instrument and Control Engineers (SICE).

Atsushi Fukayama

Senior Research Engineer, 2020 Epoch-making Project, NTT Service Evolution Laboratories.

He received the B.S. and M.S. in precision engineering from Kyoto University in 1997 and 1999. He joined NTT in 1999 and has been engaged in basic research on image recognition, human-computer interaction, and application service development utilizing technologies such as network storage and augmented reality. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE).

Yukihiro Nakamura

Senior Research Engineer, 2020 Epoch-making Project, NTT Service Evolution Laboratories.

He received the B.E. and M.E. in engineering from the University of Tsukuba, Ibaraki, in 1992 and 1994, respectively, and the Ph.D. in functional control systems from Shibaura Institute of Technology, Tokyo, in 2014. He joined NTT in 1994. From 2001 to 2005, he was with NTT Communications Corporation. From 2011 to 2014, he was with NTT Advanced Technology Corporation. In 2014, he joined NTT Service Evolution Laboratories. His research interests include network robot platforms and human-robot interaction. He is a member of the Japan Society of Mechanical Engineers (JSME), RSJ, and SICE.

Hitoshi Yamaguchi

Senior Research Engineer, 2020 Epoch-making Project, NTT Service Evolution Laboratories.

He received the B.S. in physics from Keio University, Kanagawa, in 1997 and the M.S. in applied physics from Tokyo Institute of Technology in 1999. He joined NTT Network Service Systems Laboratories in 1999 and has been engaged in research on network and communication services. He is a member of IEICE and the Japanese Society for Artificial Intelligence (JSAI).

Manabu Motegi

Senior Research Engineer, 2020 Epoch-making Project, NTT Service Evolution Laboratories.

He received the M.E. and Ph.D. in mechano-micro engineering from Tokyo Institute of Technology in 1995 and 2008, respectively. He joined NTT in 1995. From 2003 to 2006, he was with NTT-IT Corporation. In 2006, he joined NTT laboratories. His research interests include multi-robot collaboration, human-machine collaboration, and lifelog systems. He received the Encouragement Award from RSJ in 1998. He is a member of IEICE and RSJ.

Efforts toward Co-Innovation Promotion

Kazuhiko Shindo, Takeshi Sugiyama, Shozo Azuma, and Kiyoshi Kurokawa

Abstract

NTT is now shifting its business model from the conventional B2C (business-to-consumer) model to the B2B2C (business-to-business-to-consumer) model in order to create novel services in alliance with various external players. This article introduces the Co-Innovation Promotion Project that was launched at NTT Service Evolution Laboratories to promote partnerships with them.

Keywords: co-innovation, B2B2C, technical marketing

1. Introduction

In 2014, the NTT Group announced the launch of the Hikari Collaboration Model, which involves wholesaling optical-fiber access services—as a worldwide first—from NTT EAST and NTT WEST. This model is aimed at activating the ICT (information and communication technology) market in which novel value creation is achieved. Consequently, NTT is now shifting its business model from conventional B2C (business-to-consumer) to B2B2C (business-to-business-to-consumer). NTT laboratories, as part of the NTT Group, are now required to contribute to value creation for various external players and also for the NTT Group itself by developing new business domains.

NTT laboratories launched the Co-Innovation Promotion Project (EIP: Service Evolution Lab. Co-Innovation Promotion Project) in NTT Service Evolution Laboratories in December 2014 to achieve the contributions described above. EIP is more than a project, however; it is an organization of people working to achieve various outcomes. The mission of EIP is to draft cross-organizational research and development (R&D) strategies for service creation, to enhance promotional activities, and to promote industry-academia alliances directed toward the new services expected to be launched around the year 2020 with the Hikari Collaboration Model and the Co-Innovation Promotion Project.

2. Conventional technical marketing activities

The Business Promotion SE (System Engineer) Project, the former name of EIP, was focused on technical marketing and business promotion of R&D products as a one-stop representative organization of NTT laboratories in collaboration with developing companies and producers[*1]. The actual activities are described below:

- General consulting: Presenting R&D products and promoting collaboration between the representatives of each business company.
- Solution Business Produce: Promoting the *trinity*: laboratories that possess R&D products, developing companies that build commercial products, and business companies that provide services using the products.
- Industry-academia alliance promotion: Promoting collaborations with academic parties, for example, collaborative research projects.

NTT laboratories have various R&D themes, which for the most part have been put into practical use as components of services from business companies, as

---

[*1] Producers: Representatives who are responsible for creating new services by connecting the wide range of technological products from NTT laboratories with those of partners, and in so doing, forming alliances with various firms or parties.

components for internal use by the business companies, and as components of products sold by developing companies. However, quite a few of these companies have failed in these activities, whereas many of them might have been able to put the research to commercial use by collaborating with an external partner, which may have enabled the companies to extract unexpected value from the products and make full use of sales channels and proper commercialization processes that the partner companies possess. Recent changes in the circumstances around the NTT Group suggest that EIP needs to discover prospective alliance partner firms to create new services, to promote academic alliances in inexperienced fields, and to collaborate with venture firms, in addition to working on improving conventional practices.

3. Evolution of activities leading to Co-Innovation

In order to achieve Co-Innovation, it is important for NTT laboratories that EIP remains a reliable representative that various players can rely on. The players may be firms, universities, venture companies, or other organizations. They may also be NTT researchers. EIP thus assists the NTT Group with *encounter* and *awareness* services. The five characteristic services of EIP are *planning*, *implementation*, and *development* in addition to the abovementioned *encounter* and *awareness* (**Fig. 1**). The following sections describe three of these services: encounter, to receive acknowledgement from external partners and to build relationships with them; awareness, to provide opportunities to create novel ideas via technical surveys and discoveries; and planning, to promote business based on these ideas.

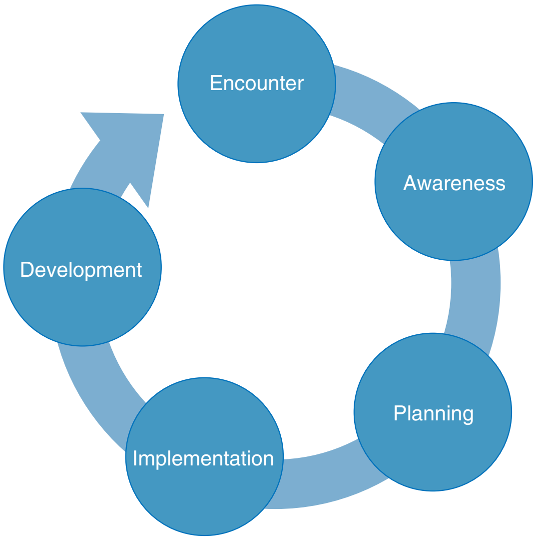


Fig. 1. Five services of EIP.

3.1 Encounter: evolved activity of promotion

EIP plans and holds internal technical exhibitions of the NTT Group that present opportunities to communicate the interests of related organizations. The latest event conveyed a favorable impression of several focused exhibitions that were designed to show application services, using the business model canvas (BMC[*2]). It was an experimental event designed to go beyond a mere introduction of technologies. The event will be expanded to a forum for making business proposals by presenting usage scenes of R&D products complemented by external technology. EIP employs the BMC to make technical proposal documents to increase the impact of technical marketing.

EIP also plans and conducts cooperative observation and discussion meetings to identify opportunities for business creation. For example, EIP held a meeting to discuss business in agriculture and to introduce related products. The discussion resulted in actual business proposals. EIP plans to hold such meetings in fields other than agriculture to promote cross-company relationships for organizations engaged in the same fields.

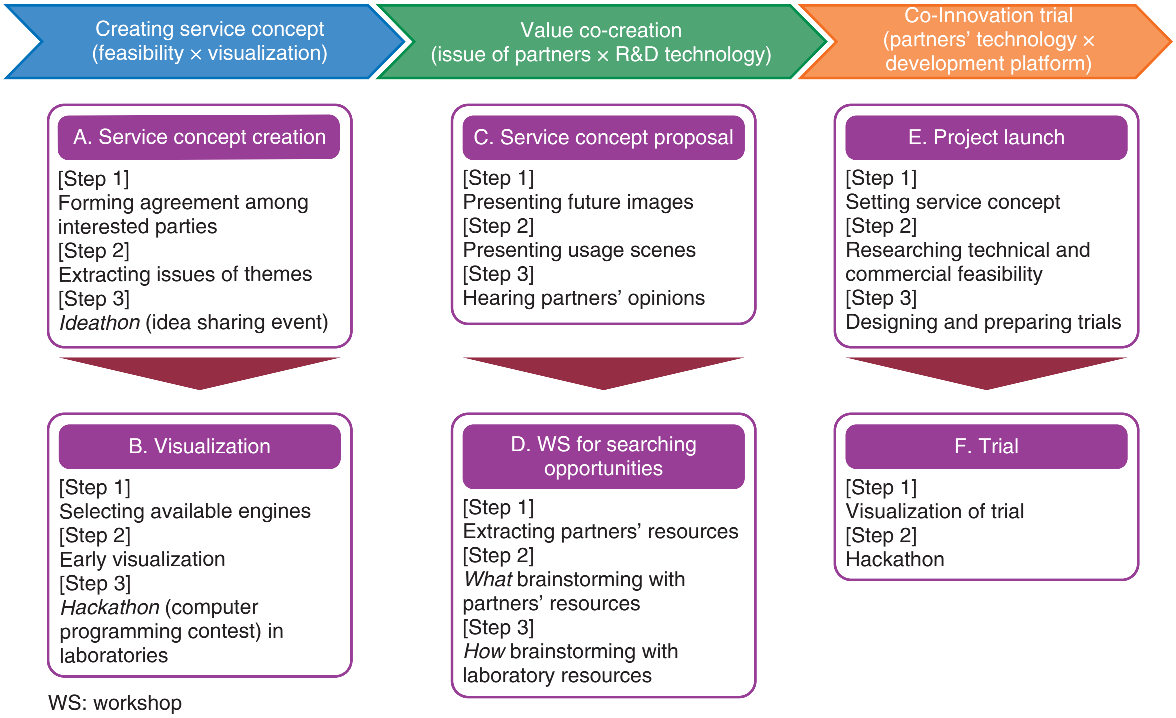
3.2 Awareness: novel activities for Co-Innovation and evolution of academic alliances

EIP is working on a value co-creation project to generate ideas for new services with prospective partner firms. EIP members present comprehensive information on NTT R&D products to the firms and work on creating business concepts based on the service resources of the firms that accommodate NTT R&D products (**Fig. 2**).

In the value co-creation project, a video is shown that presents a future vision using NTT R&D products, and the viewers are given opportunities to experience them. EIP members thus build typical user models that suggest usage scenes and introduce related technologies. NTT R&D products are no longer based only on a technical perspective but also on a user perspective. The prospective partner firms can imagine service scenes from their own viewpoints

---

[*2] BMC: A general term for a set of tools used to build business models; it consists of nine blocks: customer segments, customer relationships, channels, value proposals, main actions, resources, partners, money flow, and systematic cost evaluation.



WS: workshop

Fig. 2. Flow of a value co-creation project.

and those of their customers. The firms can also gain an understanding of the expected future vision by experiencing the latest NTT R&D products. The members thereafter hold brainstorming sessions to create service ideas based on the above resources (referred to as *What* brainstorming), and events to devise implementation processes by utilizing NTT R&D products and complementary resources of business companies (referred to as *How* brainstorming).

EIP induces the parties concerned to find partner firms at an early stage and to plan new services.

Meanwhile, conventional academic alliances have focused on cooperative research to complement internal resources by matching themes. The extended focus is to find desired alliances with unfamiliar academic organizations and various types of firms including ventures derived from universities. A greater selection of R&D products can thus be put into practical use.

3.3 Planning: evolution of commercialization

It is essential to analyze the businesses of partner firms to identify business issues and to implement strategic technical marketing of related R&D products and to promote the partnerships. EIP members create account plans to concentrate on the tasks they are focusing on. They administer and analyze their tasks according to the phase of their activity in order to clarify purposes, visualize the activities, and improve their activities. When an analysis phase is added and a clear perspective of the next actions is gained, EIP members can expect their efforts to result in commercialization of a substantial part of their work (**Fig. 3**). Several R&D products thus have been put into commercial use. To effectively promote alliances with partner firms, an information system to timely share information on key persons and their activities is needed. EIP members carry out effective technical marketing using sales force automation to share the negotiation history of projects, the status of projects, and other important information on their activities.

EIP's contributions are to attain wider awareness of R&D products, to activate R&D and commercialization, and to serve as a one-stop representative for various players including NTT laboratories themselves through the encounter, awareness, and planning services described above.

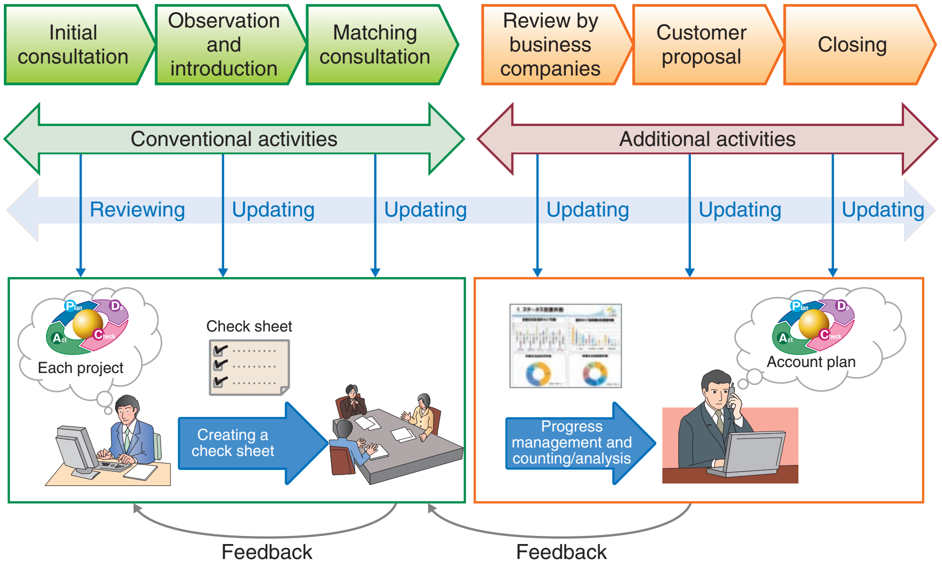


Fig. 3. Evolution of commercializing activities.

4. Future tasks

EIP will conduct R&D activities concerning future business of the NTT Group through Co-Innovation, targeting commercialization around the year 2020.

Kazuhiko Shindo

Senior Research Engineer, Supervisor, Co-Innovation Promotion Project, NTT Service Evolution Laboratories.

He received the M.S. and Ph.D. in science from Tokyo Institute of Technology in 1991 and 2012. He joined NTT in 1991 and researched carbon materials, thermocell technology, and hydrogen storage systems. He was with NTT R&D in Okinawa from 1997 to 2000 and the NTT R&D Strategy Department from 2003 to 2006. He joined the Co-Innovation Project in December 2014.

Shozo Azuma

Senior Research Engineer, Supervisor, Co-Innovation Promotion Project, NTT Service Evolution Laboratories.

He received the B.E. in information science engineering from Osaka University in 1994 and the M.E. in information science engineering from Nara Institute of Science and Technology in 1996. He joined NTT Human Interface Laboratories in 1996. He has been engaged in R&D of natural language processing, geographical information systems, agent systems, and multimedia processing.

Takeshi Sugiyama

Senior Research Engineer, Co-Innovation Promotion Project, NTT Service Evolution Laboratories.

He received the M.S. in mathematics education from Kobe University in 1998. He joined NTT in 1998, and began researching collaboration and distance learning methods via the Internet or video conference systems. He was with NTT Communications Corporation (NTT Com) from 2003 to 2014 as a system engineer, where he worked on integrating a video conference system for enterprises. He has been with NTT Service Evolution Laboratories since 2014.

Kiyoshi Kurokawa

Executive Research Engineer, Project Director, Co-Innovation Promotion Project, Service Evolution Laboratories.

He received the B.E. in computer science and the M.E. in electrical engineering from Kyushu Institute of Technology, Fukuoka, in 1988 and 1990, respectively. He joined NTT in 1990. He has been engaged in research on information resource management, data visualization, and connected homes.

Predicting *Who Will Be the Next Speaker and When* in Multi-party Meetings

Ryo Ishii, Kazuhiro Otsuka, Shiro Kumano, and Junji Yamato

Abstract

An understanding of the mechanisms involved in face-to-face communication will contribute to designing advanced video conferencing and dialogue systems. Turn-taking, the situation where the speaker changes, is especially important in multi-party meetings. For smooth turn-taking, the participants need to predict who will start speaking next and to consider a strategy for achieving good timing to speak next. Our aim is to clarify the kinds of behavior that contribute to smooth turn-taking and to develop a model for predicting the next speaker and the start time of the next speaker's utterance in multi-party meetings. We focus on gaze behavior and respiration near the end of the current speaker's utterance. We empirically demonstrate that gaze behavior and respiration have a relation to the next speaker and the start timing of the next utterance in multi-party meetings. A prediction model based on the results reveals that gaze behavior and respiration contribute to predicting the next speaker and the timing of the next utterance.

Keywords: turn-taking, gaze, respiration

1. Introduction

Face-to-face communication is one of the most basic forms of communication in daily life, and group meetings conducted using this kind of communication are effective for conveying information, understanding others' intentions, and making decisions. To design better communication systems that can enhance our communication beyond conversation *in loco*[*] and to develop social agents/robots that interact naturally with human conversations, it is critical to fully understand the mechanism of human communication. Therefore, ways to automatically analyze multi-party meetings have been actively researched in recent years [1, 2].

Turn-taking, the situation where the speaker changes, is especially important. The participants need to predict the end of the speaker's utterance and who will start speaking next and to consider a strategy for good timing with respect to who will speak next in multi-party meetings. If a model can predict the next speaker and the timing that the next speaker's utterance will start, the model will lay the foundation for the development of natural conversational systems in which conversational agents/robots speak with natural timing and of teleconference systems that avoid utterance collisions with time delays by apprising participants of who will speak next.

The goal of our research is to demonstrate the mechanism of turn-taking, namely what kind of information contributes to determining the next speaker and the timing of the next utterance, and to construct a prediction model that can predict who speaks next and when. To predict the next speaker and the timing of the next utterance, we developed a prediction model that has a three-step processing

---

[*] *in loco*: A Latin term meaning "in the proper place."

sequence: (I) prediction of turn-taking occurrence, (II) prediction of the next speaker in turn-taking, and (III) prediction of the timing of the next utterance. A flowchart of the model is shown in **Fig. 1**.

We focus on gaze behavior and respiration as information related to turn-taking. Gaze behavior is known to be an important cue for smooth turn-taking [3–5]. For example, the next speaker looks away when he/she starts to speak after having made eye contact with the current speaker in two-person meetings. However, previous research has only roughly demonstrated gaze behavior tendencies; it has not quantitatively revealed the relationship between gaze behavior and the next speaker and the timing of the next utterance. It is known that utterances and respiration are closely related. In order to speak, we must breathe out, and we need to take breaths to continue speaking for a long time. When starting an utterance, the next speaker inhales deeply. Moreover, a person's attitude about an utterance is frequently represented figuratively as *breathing*. For example, keeping as low a profile as possible so as not to be yielded the turn is often referred to metaphorically as *holding one's breath* or *saving one's breath*.

Thus, we focused on the detailed transitions of gaze behavior and respiration, which have not been investigated in multi-party meetings analysis. We previously examined the relationship between the transitions of gaze behavior and respiration and the next speaker and next-utterance timing and revealed that transitions in gaze behavior and respiration are useful for predicting them in multi-party meetings [6–8]. In this article, we describe how we analyzed gaze behavior and respiration in multi-party meetings to construct our prediction models.

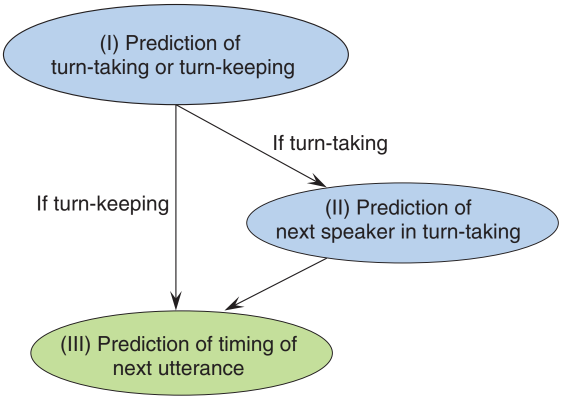


Fig. 1. Process flowchart of prediction model of next speaker and timing of next utterance.

2. Corpus of multi-party meetings

To analyze gaze and respiration, we collected a corpus of conversations in multi-party meetings. We recorded four natural (i.e., unrehearsed) meetings, such as the kind that would be held daily, conducted by four groups consisting of four different people (16 people in total) for about 12 minutes [8]. In each meeting, all four participants were in their 20s and 30s, and this was the first time they had met. They faced each other and sat down. They argued and gave opinions in response to highly divisive questions such as "Is marriage the same as love?" and were instructed to draw a conclusion within 12 minutes.

We recorded the participants' voices with a pin microphone attached to their clothing at chest level and made a video recording of the entire scene and took bust shots (head and shoulders) of each participant (recorded at 30 Hz). We recorded each participant's respiration information using a NeXus-10 MARK II biofeedback system. The respiration sensor of the system records the depth of breathing with a belt wrapped around the participant's body and outputs a value of the degree of breathing depth (called the RSP value hereafter) at 128 Hz. A high RSP value means that the person keeps taking air into the lungs. In contrast, a low RSP value means the absence of air in the lungs. We collected data during four meetings that were each about 12 minutes long (a total of about 50 minutes), and from the recorded data, we built a multimodal corpus consisting of the following verbal and nonverbal behaviors and biological information:

- Utterance: For the utterance unit, we used the inter-pausal unit (IPU) [9]. The utterance interval was extracted manually from the speech wave. The portion of an utterance followed by 200 ms of silence was used as the unit of one utterance. An utterance unit that continued by the same person was considered to be one utterance turn. Pairs of IPUs that adjoined in time, and groups of IPUs at the time of turn-keeping and turn-taking were created. There were 906 groups of IPUs created at the time of turn-keeping and 148 at the time of turn-taking.
- Gaze object: The gaze object was annotated manually using the video showing the upper body of the participants as well as overhead views from the videos by a skilled annotator. The objects of gaze were the four participants (hereafter denoted as P1, P2, P3, and P4) and other

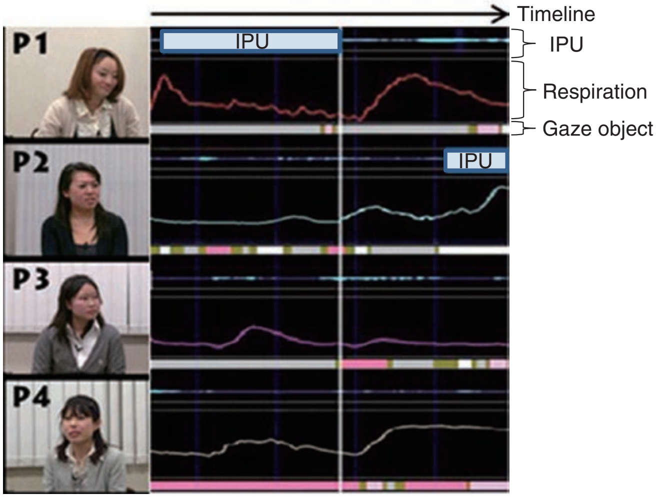


Fig. 2. Corpus data of multi-party meetings.

objects such as the floor or ceiling.

- Respiration: The RSP value measured by the Nexus MARK II has a different magnitude for each participant. To correct for these individual differences, the RSP value of each participant was normalized by the mean value μ and standard deviation δ of each participant. Specifically, the RSP value of each participant was normalized on the basis of the values of $\mu + \delta$ and $\mu - \delta$ for each participant. This enabled us to treat each participant's RSP value data for the analysis on the same scale.

All the above-mentioned data were integrated at 30 Hz for visual display using the NTT Multimodal Meeting Viewer (NTT MM-Viewer) that we developed [10] (See **Fig. 2**).

3. Prediction of next speaker and next-utterance timing based on gaze behavior

3.1 Analysis of gaze behavior

Gaze behavior is known to be an important cue for smooth turn-taking. For example, Kendon [4] demonstrated that the next speaker looks away when he/she starts to speak after having made eye contact with the current speaker in two-person meetings. Thus, it is assumed that these temporal transitions of participants' gaze behavior are important for the prediction of turn-taking situations. We therefore decided to focus on the gaze transition patterns near the end of utterances and to express them as an n-gram, which we defined as a sequence of gaze direction shifts. To generate a gaze transition pattern, we focused on the object a participant gazed at (*gazed object* hereafter) that occurs for 1200 ms during the period 1000 ms before the utterance and 200 ms after it; the candidate gazed objects were first classified as *speaker*, *listener*, or *others* (non-person objects) and labeled. At this time, the existence of mutual gaze was taken into consideration from the knowledge [4–6] that a mutual gaze takes place in two-person dialogs at the time of turn-taking. We used the following five gaze labels for the classification:

- S: Listener looks at a speaker without mutual gaze (speaker does not look at the listener.).
- S_M: Listener looks at a speaker with mutual gaze (speaker looks at the listener.).
- L_1-L_3: Speaker or listener looks at another listener without mutual gaze. L_1, L_2, and L_3 indicate different listeners.
- L_{1M}-L_{3M}: Speaker or listener looks at another listener with mutual gaze. L_{1M}, L_{2M}, and L_{3M} indicate different listeners.
- X: Speaker or listener looks at a non-person object such as the floor or ceiling.

The construction of a gaze transition pattern is shown in **Fig. 3**: P1 finishes speaking, and then P2 starts to speak. P1 gazes at P2 after he/she gazes at others during the interval of analysis. When P1 looks at P2, P2 looks at P1; namely, there is mutual gaze. Therefore, P1's gaze transition pattern is X-L_{1M}. P2 looks at P4 after making eye contact with P1. Therefore, P2's gaze transition pattern is S_M-L_1. P3 looks at others after looking at P1. P3's gaze transition

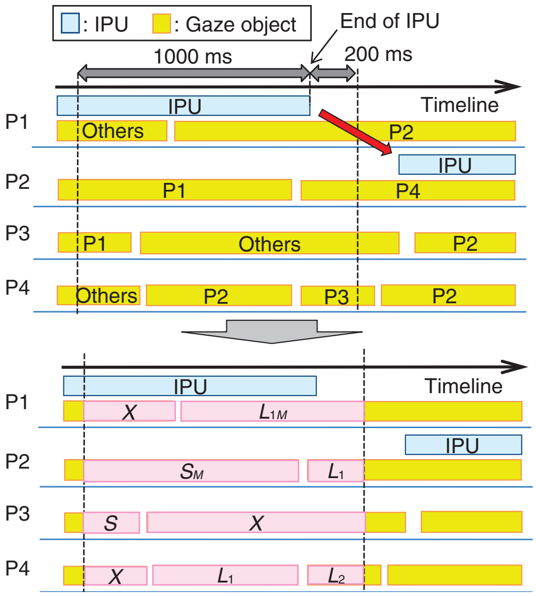


Fig. 3. Sample of gaze transition pattern generation.

pattern is therefore S-X. P4 looks at P2 and P3 after looking at others. Thus, P4's gaze transition pattern is X-L_1-L_2.

If the next speaker and the next-utterance timing differ depending on the gaze transition pattern, the gaze transition pattern may be useful for predicting them in multi-party meetings. To explore the relationship between gaze transition patterns and the next speaker and next-utterance timing, we analyzed the gaze transition pattern when turn-keeping and turn-taking occur and whether the next speaker in turn-taking and the next-utterance timing depend on the gaze transition pattern. After finishing the three analyses, we built a prediction model using gaze transition patterns.

3.2 Analysis of gaze transition pattern and turn-keeping/turn-taking

First, we analyzed how much the change in the gaze transition pattern of the speaker and listeners would differ quantitatively by turn-taking and turn-keeping. In this article, we introduce the results of analyzing the speaker's gaze transition pattern. The frequency of appearance of a speaker's gaze transition pattern under turn-taking and turn-keeping conditions using 1054 data sets is shown in **Fig. 4**. The results indicated that there were 17 gaze transition patterns. The *Others* class includes six patterns, each of which occurred in less than 1% of the data because the amount of data was small. The results of a chi-squared test showed that the frequent appearance of a gaze transition pattern differed significantly between the conditions at the time of turn-taking and turn-keeping ($\chi^2 = 87.03$, $df = 11$, $p < .01$). Next, we conducted a residual analysis to verify which gaze transition pattern differed between turn-keeping and turn-taking. The results are also shown in Fig. 4, from which we understand the following:

- A speaker's gaze transition pattern has a significantly high frequency of turn-keeping at the time of X and X-L_{1M}-X. That is, when a speaker does not look at a listener at all, or a mutual gaze with a listener is started and a gaze is stopped immediately, the frequency of turn-keeping is higher than turn-taking.
- A speaker's gaze transition pattern has a significantly high frequency of turn-taking at the time of L_{1M}, L_1, X-L_1, and L_1-X. That is, when a speaker continues to look at a listener (in spite of the presence or absence of mutual gaze), starts to gaze at a listener (not a mutual gaze), or stops looking at a listener (not a mutual gaze), the frequency of turn-taking is high.

We found that the frequency of the different gaze transition patterns for a speaker differed in turn-keeping and turn-taking. Similarly, we found that the frequency of the different gaze transition patterns for a listener differed in turn-keeping and turn-taking. Therefore, these results suggest that gaze transition patterns of the speaker and listeners are valuable information for predicting turn-keeping and turn-taking.

3.3 Analysis of gaze transition pattern and next speaker in turn-taking

Next, we analyzed the frequency of each listener's becoming the next speaker according to the speaker's and listeners' gaze transition patterns. We present here the results of analyzing the listeners' gaze transition patterns. The results of totaling who becomes the next speaker for every listener's gaze transition pattern with 148 turn-taking data sets are shown in **Fig. 5**, where the yellow bars represent the frequency that the listener herself who exhibited a gaze transition pattern (behavior) becomes the next speaker. Moreover, in a gaze transition pattern that has L_1-L_3 or L_{1M}-L_{3M}, listeners are classified into two categories: a listener who is gazed at and a listener who is not gazed at by the listener who exhibited a gaze transition pattern.

For example, in the listener's gaze transition pattern L_{1M}, the frequency of the listener exhibiting that gaze

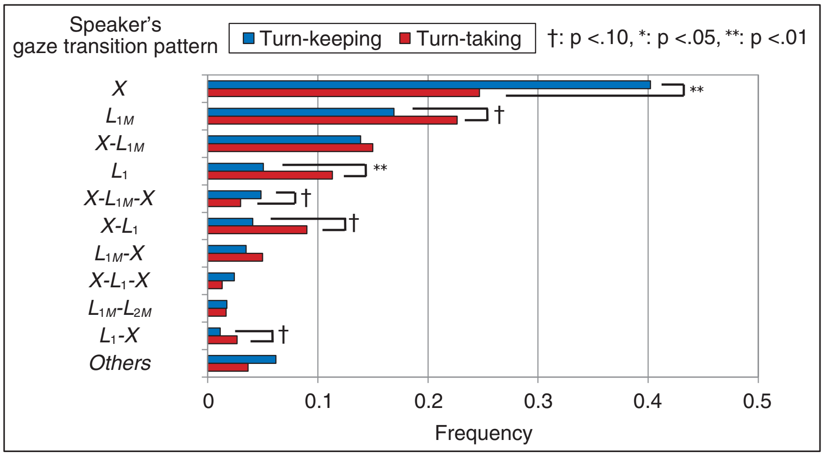


Fig. 4. Relationship between speaker's gaze transition pattern and turn-keeping/turn-taking.

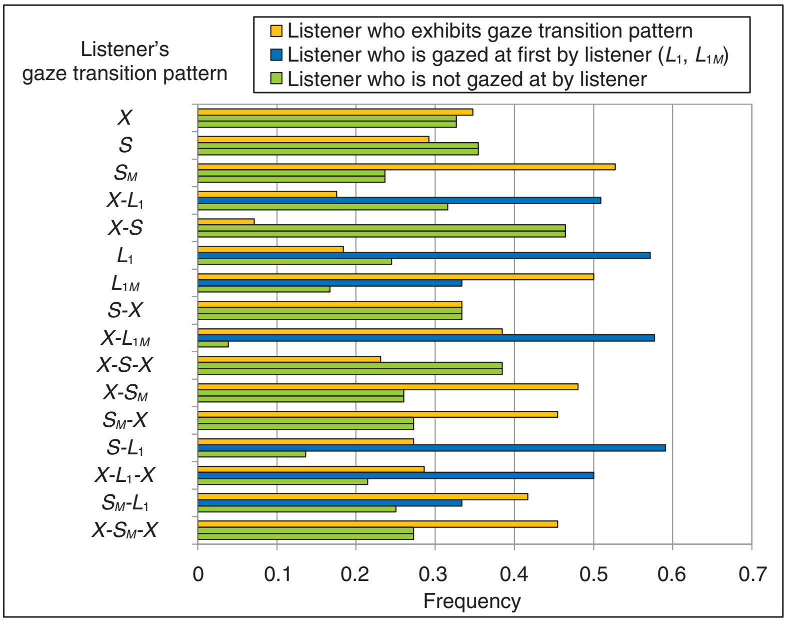


Fig. 5. Relationships between listener's gaze transition pattern and next speaker in turn-taking situation.

transition pattern who becomes the next speaker is 0.50, the frequency that the listener who is gazed at first by the listener (exhibiting the gaze pattern) is 0.33, and the frequency that another listener who is not gazed at by the listener (exhibiting the gaze pattern) becomes the next speaker is 0.17. The following becomes clear when the relationship between a listener's gaze transition pattern and the next speaker is seen in detail.

- When the listener's gaze transition pattern includes S_M, for example, S_M, X-S_M, S_M-X, S_M-L_1, or X-S_M-X, i.e., a listener makes eye contact with the speaker, the frequency that the listener becomes the next speaker is highest.
- When the listener's gaze transition pattern is X-L_1, L_1, X-L_{1M}, S-L_1, or X-L_1-X, namely, when a listener starts to look at another listener, keeps on looking at another listener without mutual gaze, looks from the speaker to another listener (without mutual gaze), or stops looking at the speaker

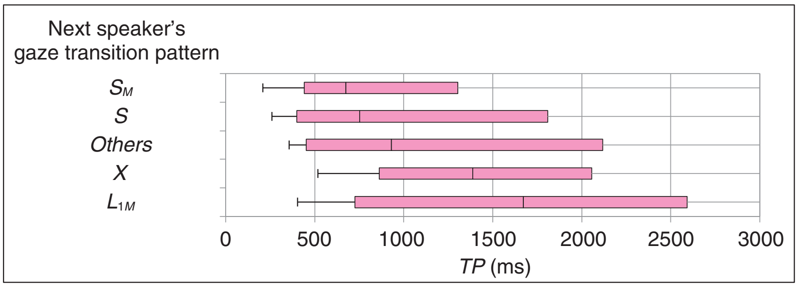


Fig. 6. Relationship between next speaker's gaze transition pattern and *TP* in turn-taking.

immediately after starting to look at the speaker (without mutual gaze), the frequency that the listener falling into pattern L_1 or L_{1M} becomes the next speaker is highest. Conversely, when the listener's gaze transition pattern is L_{1M}, i.e., a listener continues to carry out a mutual gaze with another listener, the frequency that the listener herself becomes the next speaker is highest.

We found that the frequency of each listener's becoming the next speaker in turn-taking differs depending on the listeners' gaze transition patterns. Similarly, we found that the frequency of each listener's becoming the next speaker in turn-taking differs depending on the speaker's gaze transition pattern. These results suggest that gaze transition patterns of the speaker and listeners are valuable information for predicting the next speaker in turn-taking situations.

3.4 Analysis of gaze transition pattern and next-utterance timing

An early study on this topic [4] showed that a listener's response is delayed if a speaker does not look at the listener; consequently, we think that gaze behavior is useful for predicting the timing of the next utterance. We quantitatively analyzed the correlation between the timing of the next utterance and the gaze transition pattern of the speaker and listeners. We defined timing interval *TP* as the interval between the end of the speaker's IPU and the start of the next speaker's IPU.

We analyzed the *TP* for each gaze transition pattern of the speaker and listeners in turn-keeping and of the speaker, listeners, and next speaker in turn-taking.

In this article, we introduce the results for only the next speaker's gaze transition pattern in turn-taking. Box plots of *TP* obtained for each next speaker's gaze transition pattern using 148 data sets are shown in **Fig. 6**. The *Others* class includes 38 patterns, each of which occurred in less than 5% of the data because the number of data was small. A one-way ANOVA (analysis of variance) showed that there is a significant difference in *TP* depending on the speaker's gaze transition patterns (F(4,315) = 2.05, p <.10). Here, S_M and S, which indicate that the next speaker continues to look at the current speaker, have the shortest median values, 675 and 754 ms. In contrast, L_{1M}, which means the next speaker continues to look at the listener with mutual gaze, has the longest median value, 1673 ms. That is, when the next speaker continues to look at the current speaker, the timing of the next speaker's utterance starts early. When the previously reported gaze behaviors mentioned above [4] occur, the next speaker starts to speak quickly. In contrast, when the next speaker does not look at the current speaker, turn-taking is not smooth, and the timing is late.

We found that the next-utterance timing differs depending on the next speaker's gaze transition pattern in turn-taking. Similarly, we found that the frequency of the next-utterance timing differs depending on the speaker's and listeners' gaze transition patterns in turn-keeping and turn-taking. These results suggest that gaze transition patterns of the speaker and listeners in turn-keeping and of the speaker, listeners, and next speaker in turn-taking influence the next-utterance timing situations. Therefore, it would be useful to use the gaze transition patterns to predict the next-utterance timing.

3.5 Prediction model using gaze transition pattern

The analysis results described in the previous subsections indicate that gaze transition patterns provide useful indicators of turn-keeping and turn-taking, the next speaker in turn-taking, and the timing of the next utterance in multi-party meetings. On the basis of these results, we constructed a prediction model that features three processing steps using gaze transition patterns. The model is based on a support vector machine (SVM), in which the method is SMO (sequential minimal optimization) [11]. We also evaluated the accuracy of the model, along with the effectiveness of the gaze transition patterns. The settings of the SVM are the radial basis function (RBF) kernel.

For the turn-keeping/turn-taking prediction model, the data used in the SVM consist of the turn states of turn-taking and turn-keeping as a class and the gaze transition patterns of the speaker and three listeners as features. In this phase of the study, we employed leave-one-out with 296 data sets: 148 data sets obtained by sampling 906 data items in turn-keeping to remove the bias of the number of data, and 148 data sets in turn-taking, four-fold cross validation. We collected data from four groups; we obtained training data from three of them and test data from the remaining one. Then we calculated the average prediction accuracy. The results of the evaluation indicated an accuracy rate of 65.0% in turn-keeping and 68.2% in turn-taking. This suggests that the gaze transition patterns are useful for predicting turn-taking and turn-keeping.

For the prediction model of the next speaker in turn-taking, the data used in the SVM consist of the listener who will be the next speaker as a class and the gaze transition patterns of the speaker and three listeners as features. In this phase of the study, we employed leave-one-out with 148 data sets of four dialogs, four-fold cross validation. The results showed an average prediction accuracy rate of 61.0%. This suggests that the gaze transition patterns contribute to predicting the next speaker in turn-taking.

For the prediction model of the next-utterance timing, the data used in the SVR (SVM for regression) contain the start timing as a class and the gaze transition patterns of the speaker and three listeners in turn-keeping and of the speaker, two listeners, and next speaker in turn-taking as features. We examined two models for turn-keeping and turn-taking situations. In this phase of the study, we employed leave-one-out with 906 data sets in turn-keeping and 148 data sets in turn-taking of four dialogs, four-fold cross validation. We calculated the error in the results predicted from the actual utterance start time was calculated. As a result, the average errors were 324 ms in turn-keeping and 1281 ms in turn-taking. In a similar manner, we examined a base model that outputs the average value of interval *TP*. The average error for the base model were 469 ms in turn-keeping and 1590 ms in turn-taking, which was higher than that for our prediction model. This suggests that the gaze transition patterns contribute to predicting the timing of the next speaker's first utterance in multi-party meetings.

4. Prediction of next speaker based on respiration

Utterances and respiration are known to be closely related. In order to speak, we must breathe out, and we need to take breaths to continue speaking for a long time. When starting an utterance, the next speaker inhales deeply. Moreover, a person's attitude about an utterance is frequently represented figuratively as *breathing*. For example, as mentioned previously, when someone tries to keep as low a profile as possible so as not to be yielded the turn, it is often referred to metaphorically as *holding one's breath* or *saving one's breath*. As a first attempt to deal with respiration, we conducted a fundamental study of the relationships between respiration and the next speaker in multi-party meetings [8]. After that, we devised a prediction model of the next speaker using respiration.

We analyzed how the respiration of the speaker and listeners in turn-keeping and of the speaker, listeners, and the next speaker quantitatively differs in turn-taking. We considered that if the analysis revealed differences in the speaker's respiration between turn-taking and turn-keeping or differences in respiration between the next speaker in turn-taking and the listener in turn-taking and turn-keeping, respiration could be used as a useful indicator to predict the next speaker.

We assumed that the speaker takes a breath quickly right after the utterance to continue to speak in turn-keeping. In contrast, we assumed that the speaker doesn't take a breath quickly right after the utterance in turn-taking. Therefore, we focused on the speaker's inhalation right after the end of IPU for the analysis of speaker's inhalation. We extracted the speaker's inhalation phase just after the end of the IPU and used the following inhalation-phase parameters for the speaker in order to compare differences in inhalation in detail (see **Fig. 7**).

- *MIN*: RSP value at the start of the inhalation phase, i.e., the minimum RSP value of the inhalation phase.
- *MAX*: RSP value at the end of the inhalation phase, i.e., the maximum RSP value of the inhalation phase.
- *AMP*: Amplitude of RSP value of inhalation

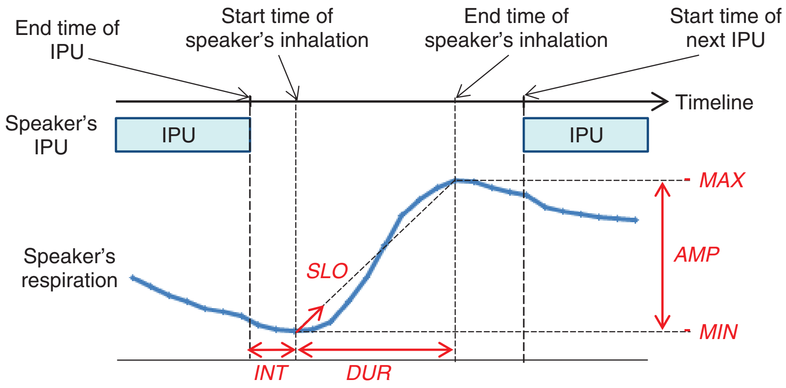


Fig. 7. Analytical parameters of inhalation right after end of IPU.

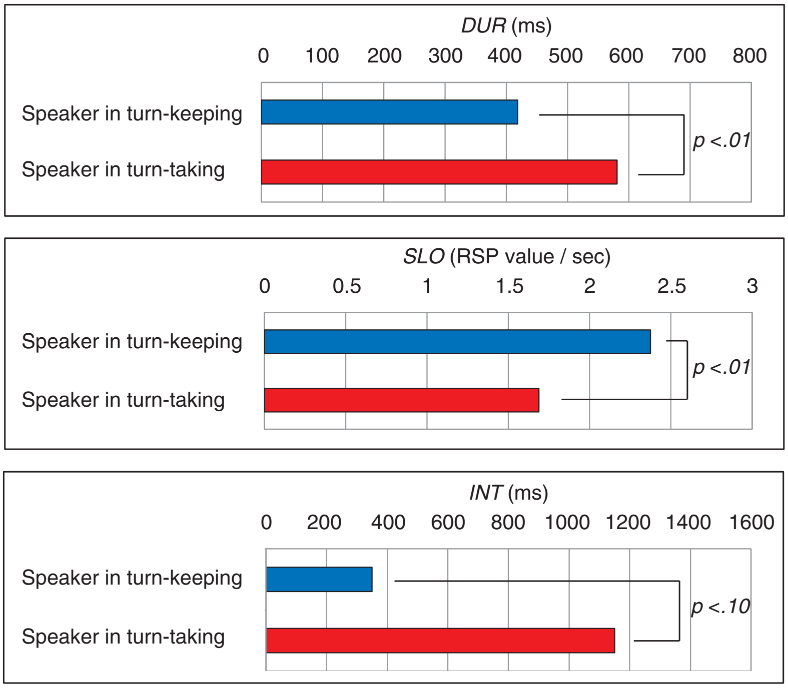


Fig. 8. *DUR*, *SLO*, and *INT* of speaker's inhalation phase right after end of IPU in turn-keeping and turn-taking.

phase.

- *DUR*: Duration of inhalation phase.
- *SLO*: Mean value of slope of RSP value per second during inhalation phase.
- *INT*: Interval between end time of speaker's IPU and start time of inhalation.

We analyzed these parameters in turn-taking and turn-keeping.

We calculated the mean value of the six parameters of the speaker in turn-taking and turn-keeping. Then, we calculated the mean value for all participants. We used a paired t-test to statistically verify whether the each parameter in turn-taking was significantly different from the same value in turn-keeping. The results suggested that there are significant differences in only *DUR*, *SLO*, and *INT* between turn-keeping and turn-taking ($t(30) = 3.08$, $p < .01$ for *DUR*; $t(30) = 2.96$, $p < .01$ for *SLO*; $t(30) = 2.04$, $p < .10$ for *INT*). These average values of *DUR*, *SLO*, and *INT* are shown in **Fig. 8**. These results reveal that the speaker

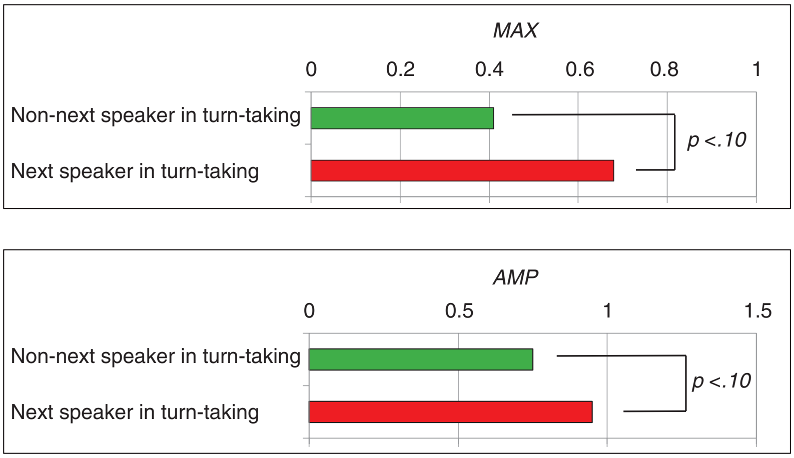


Fig. 9. *MAX* and *AMP* of inhalation phase of listeners who will not become next speaker (i.e., non-next speaker) and listener who will become next speaker (i.e., next speaker) right after end of IPU in turn-taking.

inhales more rapidly and quickly right after the end of a unit of utterance in turn-keeping.

We assumed that the next speaker takes a big breath right before starting to speak in turn-taking. In contrast, the listeners don't take a big breath. Therefore, we focused on the listeners' inhalation right before the start of next speaker's IPU for the analysis of listeners' respiration in turn-taking. We extracted the inhalation phase of the listeners. Then, we calculated the mean value of the six parameters for the listeners who will not become the next speaker (i.e., *non-next speakers*) and the listeners who will become the next speaker (i.e., *next speakers*) in turn-taking. We used a paired t-test to statistically verify whether the each parameter of the inhalation phase of the next speaker was significantly different from that of non-next speakers in turn-taking. The results suggested that there are significant differences in only *MAX*, and *AMP* between the next speaker and non-next speaker (t(30) = 1.98, p <.10 for *MAX*, t(30) = 2.03, p <.10 for *AMP*). The average values of *MAX* and *AMP* are shown in **Fig. 9**. These results reveal that the listener who will be the next speaker takes a bigger breath before speaking than listeners who will not become the next speaker in turn-taking.

The analysis results suggest that *DUR*, *SLO*, and *INT* of the speaker's inhalation right after the IPU can potentially be used to predict whether turn-keeping or turn-taking will occur, and the *MAX* and AMP of the inhalation phase of listeners can potentially be used to predict the next speaker in turn-taking.

To investigate the effectiveness of *DUR*, *SLO*, and *INT* of the speaker's inhalation right after the IPU to predict whether turn-keeping or turn-taking will occur, we constructed a prediction model based on an SVM and evaluated its performance. The SVM settings are RBF kernels. The data used in the SVM consist of the turn states of turn-taking and turn-keeping as a class and *DUR*, *SLO*, and *INT* of the speaker's inhalation as features. We employed the four-fold cross validation with 296 data sets, which includes 148 sets of data obtained by sampling 906 data items in turn-keeping in order to remove the bias of the number of data and 148 data sets in turn-taking. The accuracy rate of the prediction model was 78.7%. This suggests that parameters *DUR*, *SLO*, and *INT* of the speaker's inhalation right after the IPU contribute to predicting whether turn-keeping or turn-taking will occur.

Next, to investigate the effectiveness of *MAX* and *AMP* of the three listeners' inhalation before the next utterance in order to predict the next speaker in turn-taking, we constructed a prediction model based on the SVM as previously explained and evaluated its performance. The data used in the SVM consist of the next speaker as a class and the parameters *MAX* and *AMP* of each of the three listeners' inhalation as features. We employed four-fold cross validation with the 148 turn-taking data samples. The accuracy rate of the prediction model was 40.8%. The chance level was 33.3% because there were three next-speaker candidates in turn-taking. This suggests that the *MAX* and *AMP* parameters of the listeners' inhalation contribute to predicting the next speaker in turn-taking.

Therefore, we found that the participants' respiration is useful for predicting the next speaker in multi-party meetings.

5. Conclusion

We have developed a model for predicting the next speaker and the timing of the next speaker's utterance in multi-party meetings. As an initial attempt, we demonstrated how effective gaze behavior and respiration are in predicting the next speaker and next-utterance timing. We found from the results of analyzing gaze behavior that the next speaker and the timing of the next utterance differ depending on the gaze transition patterns of participants. The results of respiration analysis revealed that a speaker inhales more rapidly and quickly right after the end of a unit of utterance in turn-keeping than in turn-taking. The next speaker takes a bigger breath in preparing for speaking than listeners do who will not become the next speaker in turn-taking. We also constructed prediction models to evaluate how effective gaze behavior and respiration are in predicting the next speaker and the next-utterance timing. The results suggest that gaze behavior is useful for predicting the next speaker and the utterance timing, and respiration is also useful for predicting the next speaker. In the future, we plan to explore a high-performance prediction model using multimodal processing with the goal of achieving highly accurate prediction.

References

[1] K. Otsuka, "Conversation Scene Analysis," IEEE Signal Processing Magazine, Vol. 28, No. 4, pp. 127–131, 2011.

[2] D. Gatica-Perez, "Analyzing Group Interactions in Conversations: a Review," Proc. of 2006 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), pp. 41–46, Heidelberg, Germany, Sept. 2006.

[3] S. Duncan, "Some Signals and Rules for Taking Speaking Turns in Conversations," Journal of Personality and Social Psychology, Vol. 23, No. 2, pp. 283–292, 1972.

[4] A. Kendon, "Some Functions of Gaze Direction in Social Interaction," Acta Psychologica, Vol. 26, pp. 22–63, 1967.

[5] H. Sacks, E. A. Schegloff, and G. Jefferson, "A Simplest Systematics for the Organization of Turn Taking for Conversation," Language, Vol. 50, No. 4, pp. 696–735, 1974.

[6] R. Ishii, K. Otsuka, S. Kumano, M. Matsuda, and J. Yamato, "Predicting Next Speaker and Timing from Gaze Transition Patterns in Multi-Party Meetings," Proc. of the 15th ACM International Conference on Multimodal Interaction (ICMI), pp. 79–86, Sydney, Australia, Dec. 2013.

[7] R. Ishii, K. Otsuka, S. Kumano, and J. Yamato, "Analysis and Modeling of Next Speaking Start Timing based on Gaze Behavior in Multi-party Meetings," Proc. of the 39th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp. 694–698, Florence, Italy, May 2014.

[8] R. Ishii, K. Otsuka, S. Kumano, and J. Yamato, "Analysis of Respiration for Prediction of "Who Will be Next Speaker and When" in Multi-party Meetings," Proc. of the 16th ACM International Conference on Multimodal Interaction (ICMI), pp. 18–25, Istanbul, Turkey, Nov. 2014.

[9] H. Koiso, Y. Horiuchi, S. Tutiya, A. Ichikawa, and Y. Den, "An Analysis of Turn-taking and Backchannels Based on Prosodic and Syntactic Features in Japanese Map Task Dialogs," Language and Speech, Vol. 41, pp. 295–321, 1998.

[10] K. Otsuka, S. Araki, D. Mikami, K. Ishizuka, M. Fujimoto, and J. Yamato, "Realtime Meeting Analysis and 3D Meeting Viewer Based on Omnidirectional Multimodal Sensors," Proc. of the 11th ACM International Conference on Multimodal Interaction (ICMI), pp. 219–220, Cambridge, USA, Nov. 2009.

[11] S. S. Keerthi, S. K. Shevade, C. Bhattacharyya, and K. R. K. Murthy, "Improvements to Platt's SMO Algorithm for SVM Classifier Design," Neural Computation, Vol. 13, No. 3, pp. 637–649, 2001.

Ryo Ishii

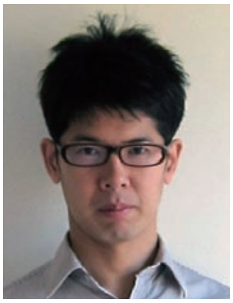
Research Engineer, Sensory Resonance Research Group, Human Information Science Laboratory, NTT Communication Science Laboratories.

He received the B.S. and M.S. in computer and information sciences from Tokyo University of Agriculture and Technology in 2006 and 2008, respectively, and the Ph.D. in informatics from Kyoto University in 2013. He joined NTT Cyber Space Laboratories (now, NTT Media Intelligence Laboratories) in 2008. He moved to NTT Communication Science Laboratories in 2012. He was also an invited researcher at Seikei University, Tokyo, from 2011 to 2013. His current research interests include communication science, multimodal interactions, and human-computer interaction. He received the FY2014 IEICE HCG Research Award and the Association for Computing Machinery (ACM) Int. Conf. on Multimodal Interaction 2014 Outstanding Paper Award. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) and JSAI (The Japanese Society of Artificial Intelligence).

Kazuhiro Otsuka

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

He received the B.E. and M.E. in electrical and computer engineering from Yokohama National University, Kanagawa, in 1993 and 1995, respectively, and the Ph.D. in information science from Nagoya University, Aichi, in 2007. He joined NTT Human Interface Laboratories in 1995. He moved to NTT Communication Science Laboratories in 2001. In 2010, he was a distinguished invited researcher at Idiap Research Institute, Switzerland. His current research interests include communication science, multimodal interactions, and computer vision. He has received several awards including the Information Processing Society of Japan (IPSJ) National Convention Best Paper Award in 1998, the IAPR International Conference on Image Analysis and Processing Best Paper Award in 1999, the ACM Int. Conf. on Multimodal Interfaces 2007 Outstanding Paper Award, the Meeting on Image Recognition and Understanding (MIRU) 2009 Excellent Paper Award, the IEICE Best Paper Award 2010, the IEICE KIYASU-Zen'iti Award 2010, the MIRU2011 Interactive Session Award, the METI of Japan, Innovative Technologies Special Award (Digital Content Expo 2012), the ACM Int. Conf. on Multimodal Interaction Outstanding Paper Awards (2012 and 2014), and the IEICE Human Communication Award 2014. He is a member of the Institute of Electrical and Electronics Engineers (IEEE), IEICE, and IPSJ.



Shiro Kumano

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

He received the Ph.D. in information science and technology from the University of Tokyo in 2009 and joined NTT Cyber Space Laboratories (now, NTT Media Intelligence Laboratories) the same year. His research interests include computer vision and affective computing. He received the ACCV 2007 Honorable Mention Award. He has served as an organizing committee member of the IAPR International Conference on Machine Vision Applications. He is a member of IEEE, IEICE, and IPSJ.

Junji Yamato

Executive Manager, Media Information Laboratory, NTT Communication Science Laboratories.

He received the B.E., M.E., and Ph.D. from the University of Tokyo in 1988, 1990, and 2000, respectively, and the S.M. in electrical engineering and computer science from Massachusetts Institute of Technology, USA, in 1998. His areas of expertise are computer vision, pattern recognition, human-robot interaction, and multi-party conversation analysis. He is a visiting professor at Hokkaido University and Tokyo Denki University and a lecturer at Waseda University. He is a senior member of IEEE and ACM.

Taking the English Exam for the "Can a Robot Get into the University of Tokyo?" Project

Ryuichiro Higashinaka, Hiroaki Sugiyama, Hideki Isozaki, Genichiro Kikui, Kohji Dohsaka, Hirotoshi Taira, and Yasuhiro Minami

Abstract

NTT and its research partners are participating in the "Can a robot get into the University of Tokyo?" project run by the National Institute of Informatics, which involves tackling English exams. The artificial intelligence system we developed took a mock test in 2014 and achieved a better-than-human-average score for the first time. This was a notable achievement since English exams require English knowledge and also common sense knowledge that humans take for granted but that computers do not necessarily possess. In this article, we describe how our artificial intelligence system takes on English exams.

Keywords: artificial intelligence, natural language processing, Todai Robot Project

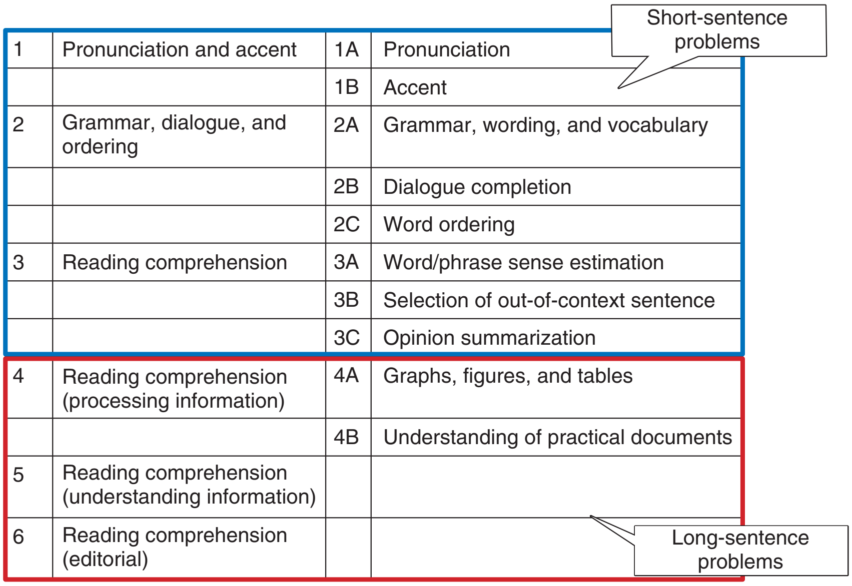
1. Introduction

The National Institute of Informatics (NII) has been promoting the Todai Robot Project (Can a robot get into the University of Tokyo?) [1]. By developing *artificial intelligence* that can pass the entrance exam of the University of Tokyo (Todai), the members of the project aim to clarify the limitations of artificial intelligence and the boundaries between humans and computers. The goal of the project is to achieve high marks on the National Center Test for University Admissions (NCTUA) by 2016 and to be admitted into the University of Tokyo by 2021.

NTT, together with its research partners, Okayama Prefectural University, Akita Prefectural University, Osaka Institute of Technology, and the University of Electro-Communications, joined the project to form an English team in 2014 and achieved a score of 95 out of 200 points) on a mock Center Test (Yozemi mock Center Test by the educational foundation Takamiya Gakuen) in the same year; the score we achieved with our computer program was above the national average of 93.1. This is a big jump from the previous year, because the English score in 2013 was 52, which was a chance-level score considering that all problems are multiple choice, mostly consisting of four choices.

It may be a surprise to readers to learn that English is regarded as one of the most difficult subjects for computers compared with other subjects. This is because problems in English cannot be solved just by having English knowledge such as of vocabulary, grammar, and idioms. Rather, these problems require *common sense knowledge*. For example, there are problems in which it is necessary to estimate human emotion and the causal relationships of events. At NTT, we have been tackling problems that require common sense reasoning through the research and development of dialogue and machine translation systems. We successfully exploited our expertise in these fields to achieve a high score on the mock Center Test.

To pass the entrance exam of the University of Tokyo, it is necessary to achieve high marks on both



1	Pronunciation and accent	1A	Pronunciation
		1B	Accent
2	Grammar, dialogue, and ordering	2A	Grammar, wording, and vocabulary
		2B	Dialogue completion
		2C	Word ordering
3	Reading comprehension	3A	Word/phrase sense estimation
		3B	Selection of out-of-context sentence
		3C	Opinion summarization
4	Reading comprehension (processing information)	4A	Graphs, figures, and tables
		4B	Understanding of practical documents
5	Reading comprehension (understanding information)		
6	Reading comprehension (editorial)		

Short-sentence problems

Long-sentence problems

Fig. 1. Basic structure of English exam of NCTUA.

the NCTUA and second-stage exams. However, since the second-stage exams contain difficult tasks involving writing, we are currently working on the NCTUA, where there are only multiple-choice problems. The typical structure of the English exam of the NCTUA is shown in **Fig. 1**. There are usually six categories of problems; the first three are short-sentence problems (problems related to words and short phrases), and the latter three are long-sentence problems (problems related to long documents that may contain illustrations, graphs, and tables).

In this article, we describe how *Torobo-kun* (*Torobo* stands for Todai Robot, and *kun* is an honorific title in Japanese normally used for boys; the members of the project affectionately refer to the exam-solving software this way) solves questions on the English exam, putting an emphasis on our approach on short-sentence problems. Unfortunately, the current level of artificial intelligence has difficulty solving long-sentence problems, and we have not yet established a concrete method for solving them. We discuss the reason for this difficulty in the last section. Moreover, Torobo-kun is currently a type of computer software. Therefore, it does not visually see and read questions on paper, nor does it fill in answer sheets using pencils. It obtains exams in electronic form and outputs choices that it thinks are correct.

2. Pronunciation and accent (word stress) problems

Problems regarding pronunciation and accent (word stress) are not that difficult for computers because they can consult electronic dictionaries.

With pronunciation, the task is to select a word that has a letter that is pronounced differently from the same letters in the other choices. An example of a pronunciation problem is shown in **Fig. 2**. To solve this problem, the program first consults an electronic dictionary. The dictionary here is one developed for speech recognition and synthesis research*; thus, it has pronunciation and accent information for words. Then, to obtain the pronunciation for a target letter (in this case, “I”), we calculate the most likely alignments between letters and pronunciations within the dictionary by using statistical methods. This tells us which letter corresponds to what pronunciation. Finally, since we know that the pronunciation for “I” in “ignorant” is different from the “I”s in the other choices, we can obtain the correct answer.

With accent, the task is to select, out of four choices, a word that has the most accented syllables in a different position compared with other choices. An example of an accent problem is shown in **Fig. 3**. The same dictionary is used to solve this type of problem.

---

* We used the Carnegie Mellon University Pronouncing Dictionary.

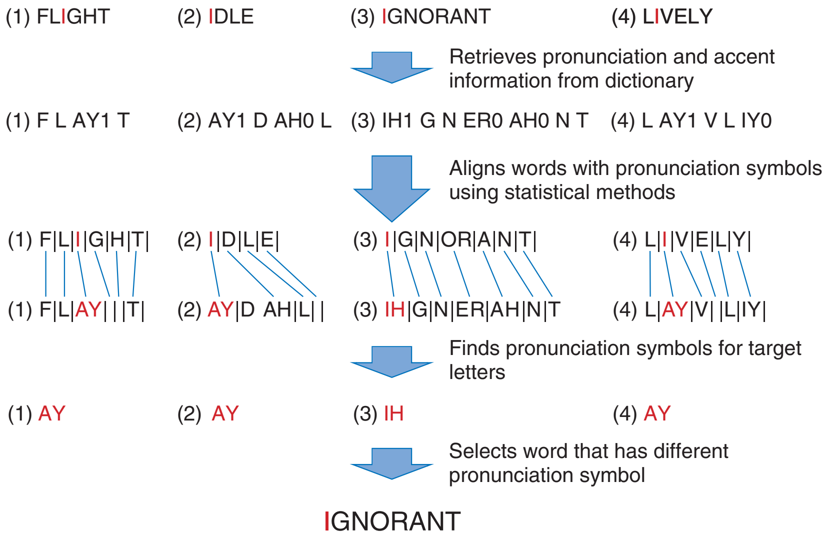


Fig. 2. Pronunciation problem.

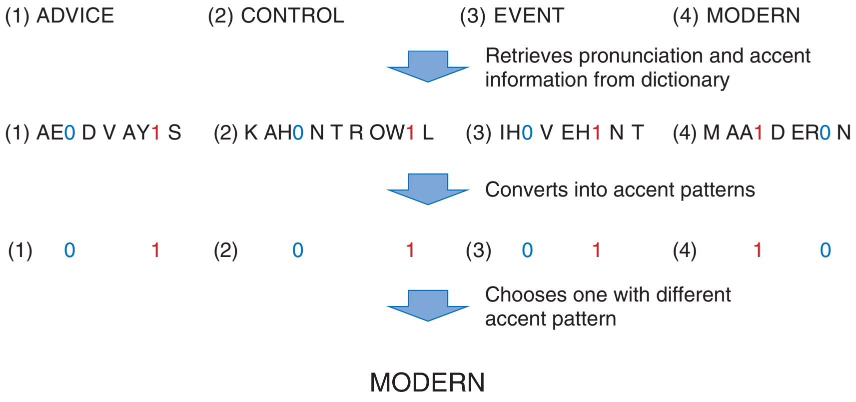


Fig. 3. Accent problem.

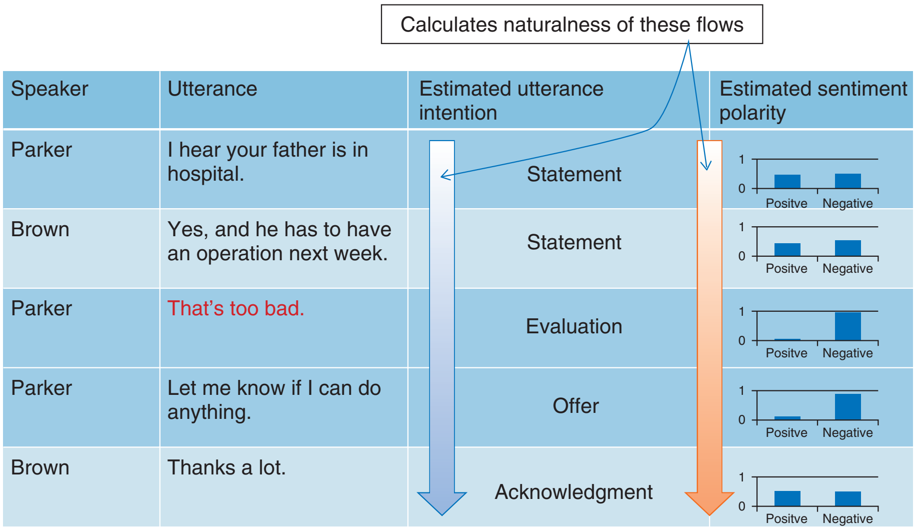
In the figure, alphabetic characters (letters) denote pronunciation symbols, and the numbers 0 and 1 respectively indicate whether a vowel is a non-accented or an accented one. From these 0s and 1s, we can find the accented vowels in the words and immediately determine that *modern* is the odd one out.

3. Grammar, wording, and vocabulary problem

The task here is to select one of the four choices that best fits in the blank in a sentence. Below is an example:

I had a severe toothache, so I made [A] with the dentist.
(1) a promise, (2) a reservation, (3) an appointment, (4) an arrangement.

In this example, the answer that best fits [A] is (3), an appointment. To solve this type of problem, we turn to a technology called statistical language models, which have been widely used in machine translation to produce fluent translations. Language models are created by statistically processing a large number of documents; they contain the probability information about how one expression follows another. Given a question, our program first creates sentences by



Calculates naturalness of these flows

Speaker	Utterance	Estimated utterance intention	Estimated sentiment polarity
Parker	I hear your father is in hospital.	Statement	
Brown	Yes, and he has to have an operation next week.	Statement	
Parker	That's too bad.	Evaluation	
Parker	Let me know if I can do anything.	Offer	
Brown	Thanks a lot.	Acknowledgment	

Fig. 4. Flow of utterance intentions and sentiment polarities.

filling in the blanks with each of the choices and calculates the fluency of each sentence by using the language models. Then, we select the choice that realizes the most fluent (probable) sentence.

4. Dialogue completion problem and opinion summarization problem

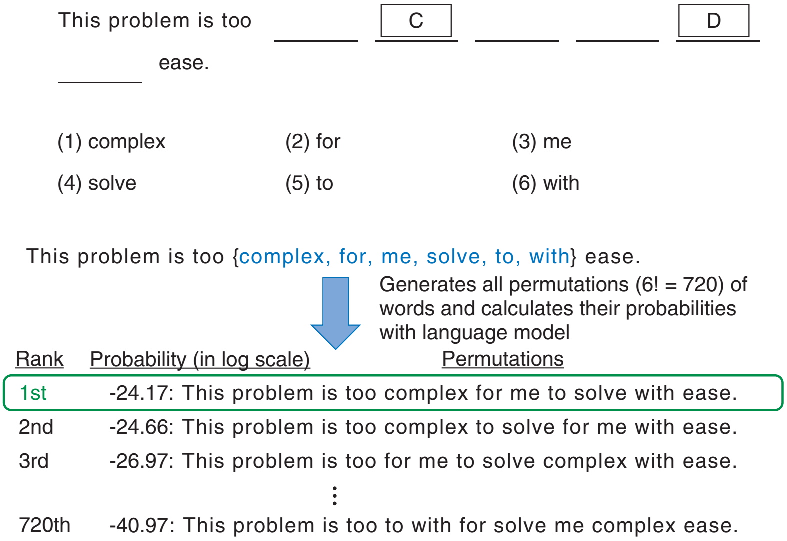
In the dialogue completion problem, it is necessary to select one utterance from four choices to fill in the blank in a given dialogue. In the example below, the utterance that best fits [B] is choice (4).

Parker: I hear your father is in hospital.
Brown: Yes, and he has to have an operation next week.
Parker: [B] Let me know if I can do anything.
Brown: Thanks a lot.
(1) Exactly, yes., (2) No problem., (3) That's a relief., (4) That's too bad.

To solve this type of problem, we first fill in the blank by using each of the choices and, for each case, estimate the naturalness of the conversation; the choice that achieves the highest degree of naturalness is selected as the answer. We estimate the naturalness of a conversation from two viewpoints; one is the flow of utterance intentions (also known as dialogue acts or illocutionary acts, e.g., statement, evaluation, question, etc.) and the flow of sentiment polarities, e.g., positive and negative.

The flow of utterance intentions and sentiment polarities when the blank has been filled in with (4) is shown in **Fig. 4**. Here, the utterance intentions and sentiment polarities are those that have been automatically estimated from dialogue data using statistical methods. The conversation begins with a *statement* about Brown's father being in hospital followed by another *statement* about the fact that he has to have an operation. This is followed by an *evaluation* (evaluative response), *offer*, and *acknowledgment*. The flow of these utterance intentions seems as natural as that in human conversation. As for sentiment polarities, the polarities regarding the admission to hospital and the operation match that of (4); that is, they all have negative polarities. Therefore, this flow of sentiment is also as natural as that in human conversation. On the basis of the naturalness of these two flows, Torobo-kun can select (4) as the most appropriate answer.

The task with opinion summarization problems is similar to that of the dialogue completion problem. The task in this case is to select an utterance that best summarizes a speaker's opinion in a multi-party discussion. Since a discussion is a kind of conversation, we use the same technique as just described; that is, we select an utterance that creates the most natural flow. In this type of problem, the choices are usually of the same utterance intention. Therefore, we only leverage the flow of sentiment polarities.



This problem is too ______ [C] ______ ______ [D] ______ ease.

(1) complex (2) for (3) me

(4) solve (5) to (6) with

This problem is too {complex, for, me, solve, to, with} ease.

Rank	Probability (in log scale)	Permutations
1st	-24.17	This problem is too complex for me to solve with ease.
2nd	-24.66	This problem is too complex to solve for me with ease.
3rd	-26.97	This problem is too for me to solve complex with ease.
⋮		
720th	-40.97	This problem is too to with for solve me complex ease.

Fig. 5. Solving the word ordering problem.

5. Word ordering problem

The task in these problems is to fill in blanks in a sentence by correctly ordering given words in order to create a valid grammatical sentence. An example is shown in **Fig. 5**. In the example, the correct ordering for the given words is “complex for me to solve with” (the answer for this problem consists of the words that correspond to blanks [C] and [D]). For this type of problem, our program creates all possible permutations of word orderings (in the example, 720 permutations are created) and chooses the most appropriate one. To calculate the appropriateness, we turn to the statistical language models we used for the grammar, wording, and vocabulary problems.

6. Word/phrase sense estimation problem

The task in this case is to estimate the meaning of an unknown word/phrase (the unknown word/phrase may not be unknown to the computer program because it has access to dictionaries; here, *unknown* means *unknown or unfamiliar to examinees*) and to select a choice whose meaning is the most similar to the given word or phrase. Below is an example:

George: I must get this paper finished by next Monday.
Paul: So you can’t go to dinner with me this weekend?
George: No, but I’ll take a rain check.
Paul: Sure, how about next weekend.
Unknown word (phrase): *take a rain check*
(1) accept your offer later, (2) change my mind, (3) go with you, weather permitting, (4) refuse the invitation.

In this problem, *take a rain check* is the unknown phrase, and we need to select the choice that has the meaning closest to the phrase. In this example, because *take a rain check* means to request a postponement of an offer, the correct answer is (1). For this type of problem, we utilize a technique called word2vec [2] and an idiom dictionary. Word2vec enables us to calculate the similarity between words on the basis of their usages in large text data. The idea used here is called distributional similarity; that is, the more similar the contexts of words, the closer their meanings are.

Using word2vec, we calculate the similarities between an unknown word/phrase and the choices and select the one that has the highest similarity score. Note that when the unknown word/phrase is listed in the idiom dictionary, that word/phrase is replaced by its gloss (definition statement) before calculating the similarity. This heightens the accuracy in calculating similarities.

7. Long-sentence problems and future directions

With long-sentence problems, we apply the method used for textual entailment. In textual entailment, two documents are given, and the task is to recognize whether the content of one document is contained in the other. In long-sentence problems, there are many questions that ask whether the content of a choice is contained in a given document. Therefore, the technology for textual entailment can be directly applied. However, this approach achieved very poor results in the mock Center Test. In particular, it was difficult to absorb the differences in expressions with the same meaning and to find reasonable referents for referring expressions. It is therefore necessary to further improve the accuracy of textual entailment. In addition, it is necessary to read the minds of characters in stories, understand illustrations, graphs, and tables, and grasp the logical and rhetorical structures of documents. As sentences get longer and longer, the need to *read between the lines* becomes more important. Between the lines, there is common sense, and to solve long-sentence problems, it is necessary to tackle the problem of common sense head on. In the future, we will focus on solving long-sentence problems and creating programs that can achieve high marks on both the NTCUA and the second-stage exams of the University of Tokyo.

Acknowledgment

We thank the educational foundation Takamiya Gakuen for providing us with the mock Center Test data.

References

[1] N. H. Arai and T. Matsuzaki, "The Impact of A.I. on Education—Can a Robot Get into the University of Tokyo?", Proc. of the 22nd International Conference on Computers in Education (ICCE 2014), pp. 1034–1042, Nara, Japan, Nov./Dec. 2014.

[2] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed Representations of Words and Phrases and their Compositionality," Proc. of the 27th Annual Conference on Neural Information Processing Systems (NIPS 2013), pp. 3111–3119, Stateline, NV, USA, Dec. 2013.

Ryuichiro Higashinaka

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

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

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

Hiroaki Sugiyama

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

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

Hideki Isozaki

Professor, Department of Systems Engineering, Okayama Prefectural University.

He received the B.E., M.E., and Ph.D. in engineering from the University of Tokyo in 1983, 1986, and 1997, respectively. He joined NTT Basic Research Laboratories in 1986 and studied artificial intelligence including logical inference. From 1990 to 1991, he was a guest researcher at Stanford University, USA. In 1995, he began studying natural language processing and soon after joined NTT Communication Science Laboratories. He is currently studying machine translation and scientific literature question answering including image processing. He received the 2004 IPSJ Best Paper Award. He joined Okayama Prefectural University in 2011. He is a member of the Association of Computational Linguistics, NLP, IPSJ, JSAI, and IEICE.

Genichiro Kikui

Professor, Department of Computer Science and Systems Engineering, Okayama Prefectural University.

He received the B.S. and M.E. in electrical engineering and the Ph.D. in informatics from Kyoto University in 1984, 1986, and 2007, respectively. He joined NTT in 1986. He was a visiting researcher at the German Research Centre for Artificial Intelligence (DFKI) in 1993 and the Center for the Study of Language and Information (CSLI) at Stanford University, USA, in 1997. He worked for ATR Spoken Language Processing Labs from 2001 to 2005 as the head of Department 2, responsible for integrating speech and natural language processing. He has been a professor with Okayama Prefectural University since 2011. He has worked on written and spoken natural language processing including text/speech translation, automatic text revision, and cross-language information retrieval. He is currently interested in information extraction and its applications. He received the 1997 IPSJ Best Author Award, the 2005 IEICE Information and Systems Society Best Paper Award, and the 22nd Telecom Technology Award. He is a member of IEICE, IPSJ, JSAI, and NLP.

Kohji Dohsaka

Professor, Department of Electronics and Information Systems, Akita Prefectural University.

He received the B.S. and M.E. in information and computer science from Osaka University in 1984 and 1986, respectively, and the Ph.D. in information science from Japan Advanced Institute of Science and Technology, Ishikawa, in 2004. He joined NTT in 1986. He was a visiting researcher at the Massachusetts Institute of Technology (MIT) in Cambridge, MA, USA, in 2002 and an invited professor at Osaka University from 2008 to 2011. He has been a professor with Akita Prefectural University since 2012. He has worked on dialogue systems, human-robot interaction, and natural language processing; he is now interested in communication robots for human support. He received the 1996 IPSJ Best Paper Award and the 2010 and 2011 JSAI Annual Conference Award. He is a member of IEICE, IPSJ, JSAI, and NLP.

Hirotoshi Taira

Associate Professor, Faculty of Information Science and Technology, Osaka Institute of Technology.

He received the B.S. in chemistry and the M.S. in solid-state chemistry from the University of Tokyo and the Ph.D. in information science from Nara Institute of Science and Technology in 1994, 1996, and 2002, respectively. He joined NTT Communication Science Laboratories in 1996 and researched text categorization with machine learning and bioinformatics. He was with the Business Intelligence Deployment Center, Research and Development Headquarters, NTT DATA, where he studied and developed business intelligence techniques from 2005 to 2007. He was a research scientist in the Natural Language Group, Innovative Communication Laboratory, NTT Communication Science Laboratories, where he researched Japanese predicate argument structure analysis and recognition of textual entailment from 2007 to 2014. He joined Osaka Institute of Technology in 2014. His research interests include natural language processing, machine learning, and artificial intelligence. He is a member of IPSJ, JSAI, and NLP.

Yasuhiro Minami

Professor, Department of Information Systems Fundamentals, Graduate School of Information Systems, the University of Electro-Communications.

He received the M.E. in electrical engineering and the Ph.D. in electrical engineering from Keio University, Kanagawa, in 1988 and 1991, respectively. He joined NTT in 1991. He was a visiting researcher at MIT, USA, from 1999 to 2000 before rejoining NTT. He has been a professor with the University of Electro-Communications in Tokyo since 2014. His interest is modeling for speech recognition. He received the 2006 IPSJ Best Paper Award, the 2005 IEICE Information and Systems Society Best Paper Award, and the 2005 and 2008 Telecom System Technology Award. He is a member of IEICE, the Acoustical Society of Japan, and the Institute of Electrical and Electronics Engineers.

Cross-connect System with Packet Transport Technology

Masaya Ogawa, Hidenori Iwashita, Takashi Takashima, Toshiya Matsuda, Kenichi Higuchi, and Katsutoshi Koda

Abstract

The public switched telephone network (PSTN) and existing dedicated line services have been carried over transmission lines using older data link systems based on time-division multiplexing technology. However, as Internet protocol traffic has increased and traffic to the PSTN and existing dedicated services has decreased, there has been a move to use data link systems based on packet transport technology; this in turn has allowed simpler networks to be built. We have developed a packet-based cross-connect system that can efficiently promote this transition by allowing individual low-speed (1.5 Mbit/s) paths to be split up and reassembled.

Keywords: SDH, MPLS-TP, cross-connect system

1. Introduction

1.1 Development goals

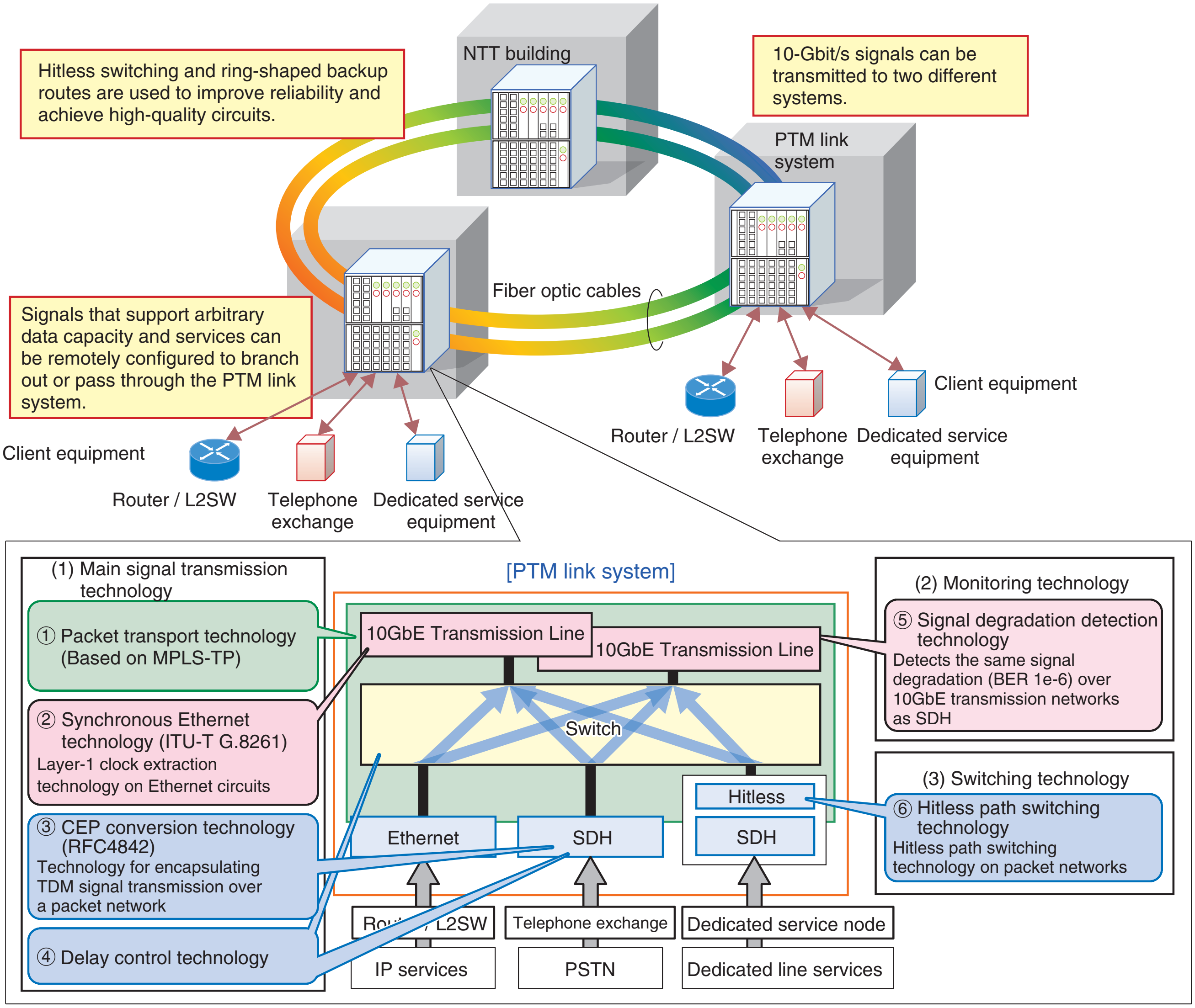
The old data link system and existing dedicated line system were adopted more than 20 years ago, and they have begun to show signs of age. Consequently, replacement parts need to be secured, the cost of maintenance is rising, and there is a shrinking number of technicians familiar with the equipment. NTT Network Service Systems Laboratories is working to address these issues by developing technologies for migrating away from the old data link system and the existing dedicated line system. However, the increase in Internet protocol (IP)-related traffic together with the reduction in traffic to the public switched telephone network (PSTN) and existing dedicated line services means that any future network needs to be able to:

- Handle increased IP traffic with expanded IP-related service areas
- Offer flexible, granular bandwidth settings to efficiently accommodate IP-related services alongside PSTN and existing dedicated line services
- Provide maintenance functionality such as alert monitoring and path settings equivalent to conventional synchronous digital hierarchy (SDH).

1.2 Technologies applied

To satisfy these conditions, NTT Network Service Systems Laboratories has developed a Packet Transport Multiplexer (PTM) link system based on this packet transport technology. An overview of the PTM link system is shown in **Fig. 1**. The following signal transmission technologies are used in the PTM system: packet transport technology based on Multiprotocol Label Switching - Transport Profile (MPLS-TP), Synchronous Ethernet technology that can extract layer-1 clocks over the Ethernet physical layer, CEP (circuit emulation over packet) conversion technology that encapsulates time-division multiplexing (TDM) signal transmission over a packet network, and delay control technology that can forward data over packet networks with the same delay as conventional SDH.

For monitoring, the PTM link system uses signal degradation detection technology that can detect the same amount of signal degradation (bit error rate: 10^{-6}) as SDH over the Ethernet physical layer. For switching, the PTM link system uses hitless path switching over packet networks. The PTM link system is an infrastructure system that will form the



BER: bit error rate
ITU-T: International Telecommunication Union - Telecommunication Standardization Sector
L2SW: layer 2 switch

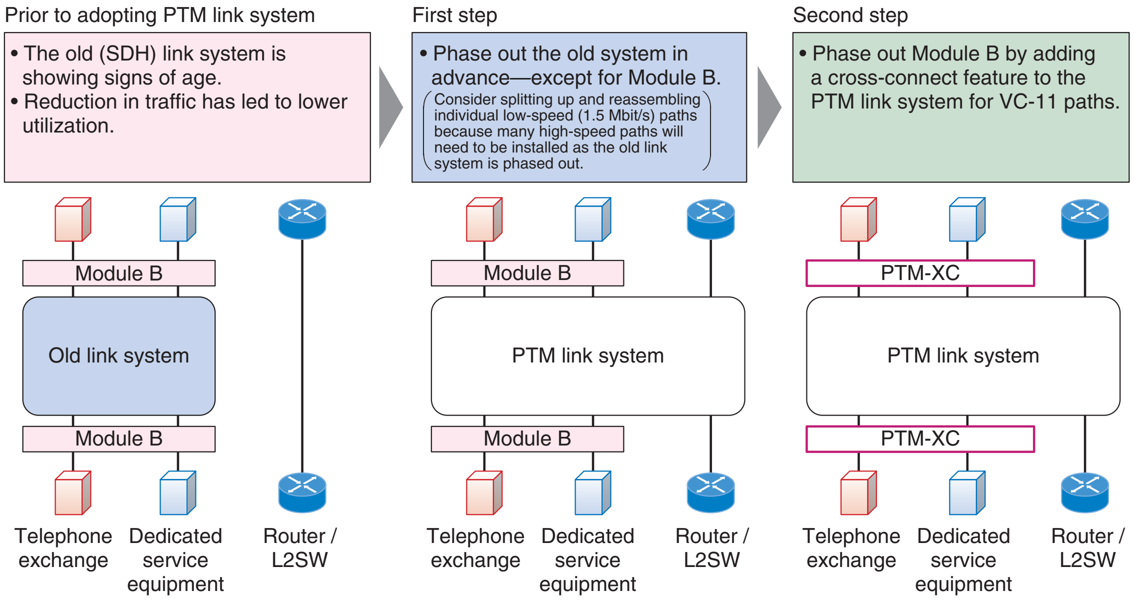
Fig. 1. Overview of PTM link system.

basis of NTT's telecommunications network with the same quality of data transmission as synchronous transport modules (STMs), which can provide TDM services for PSTN and dedicated lines over packet networks. The PTM link system integrates both TDM and packet technology; it can multiplex and demultiplex paths of arbitrary capacity at any datacenter, allowing the network to dynamically adjust in response to demand and the type of service provided.

1.3 Migration process

Development at NTT has proceeded in anticipation of a two-step migration process. The previous-generation link system will be migrated first (as shown in **Fig. 2**), followed by the equipment referred to as Module B. The Module B cross-connect equipment has STM-0/STM-1 interfaces (52M/156M IF) for connecting VC-11 (virtual container 11: 1.5 Mbit/s) paths between any STM-0/STM-1 interfaces.

Subscriber nodes and relay nodes for PSTN and dedicated services will be connected to each other via



PTM-XC: packet transport multiplexer cross-connect

Fig. 2. Sample application of link system to accommodate existing PSTN and dedicated line services.

VC-3 (52 Mbit/s) paths in the future. As a result, we can bundle several low-speed VC-11 and VC-2 (6.3 Mbit/s) path segments with low utilization in the VC-3 path to increase path utilization and enable more efficient migrations. This approach will also be used for connections between businesses and at other times when it is difficult to transition to VC-3 paths. To improve path utilization, we developed a PTM-XC (PTM cross-connect) system with cross-connect features for VC-11 and VC-2 paths. An overview of the PTM-XC is shown in **Fig. 3**.

By basing PTM-XC on a PTM link system with packet switches, we were able to use common parts and minimize the scope of development.

2. Issues of migrating away from existing equipment (Module B)

The equipment for Module B was adopted over 20 years ago, so we need to consider the following issues in the PTM-XC migration.

2.1 High-efficiency, high-density utilization

The utilization of existing equipment has dropped with the decrease in traffic to the PSTN and existing dedicated line services. For this reason—and because many more VC-3 paths will need to be installed when the old link system is phased out—we aim to achieve higher utilization by splitting up and reassembling individual VC-11/VC-2 paths. Sample connections from a single consolidated building to two subscriber buildings using telephone exchange and dedicated service equipment are shown in **Fig. 4**. Without the cross-connect equipment, VC-11 paths (1.5 Mbit/s) cannot be multiplexed, and thus, four under-utilized VC-3 paths are required. By introducing cross-connect equipment, we can multiplex VC-11 paths and thus establish connections via two VC-3 paths with high utilization. Furthermore, ten Module B units are necessary to accommodate the maximum number of (VC-11) paths—4032—when utilization is high. These Module B units occupy a large area once they are installed. We wanted PTM-XC to accommodate the same number of paths as Module B in a denser package to reduce this installation size.

2.2 Reduced delay

PTM-XC must reduce the delay for the existing dedicated line services that it accommodates at least as much as Module B does. In general, the conversion of VC-11/VC-2 signals to MPLS-TP packets—and vice versa—incurs considerable delay that we want to minimize.

2.3 Reduced power consumption

Module B can comprise up to ten units, so its

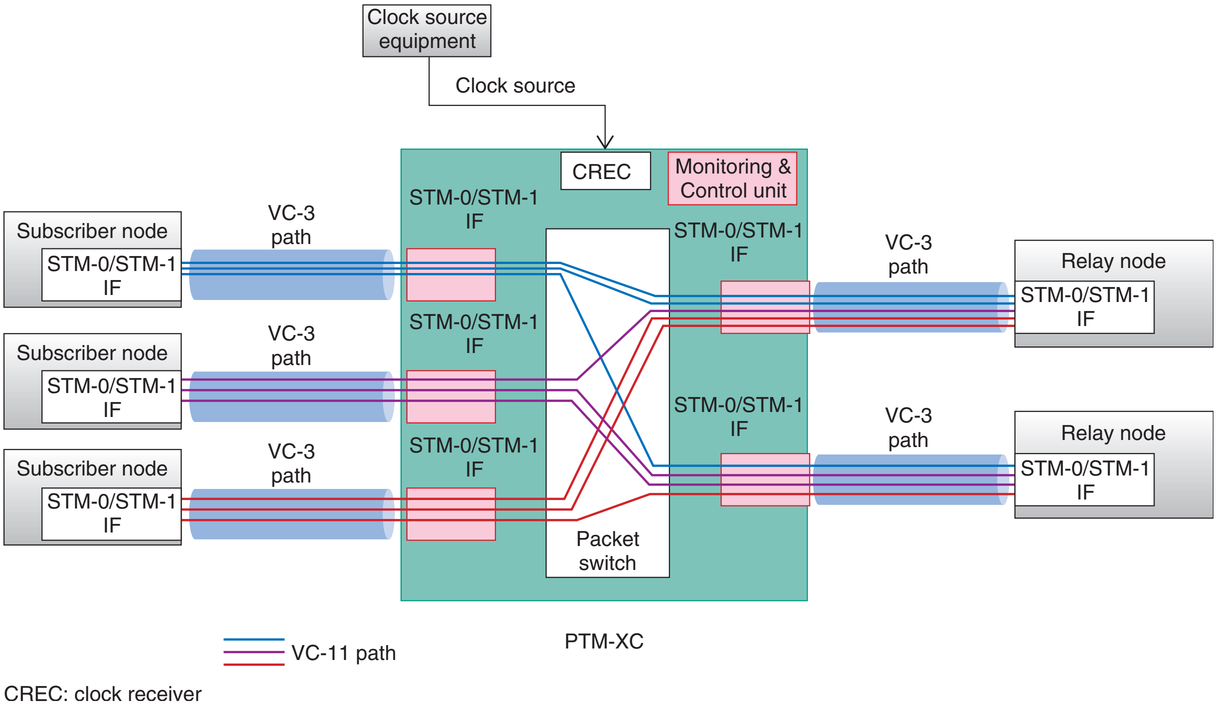


Fig. 3. Overview of PTM-XC.

energy consumption is significant and its operational costs are increasing. The demand for energy conservation has increased in recent years, and therefore, we want PTM-XC to use less power than its predecessors.

2.4 Improved maintainability

Cross-connect equipment must be able to accommodate up to 4032 (VC-11) paths; it also comes equipped with a large number of packages and optical modules, which we want to make easier to register.

In addition, there are numerous interfaces that connect the Module B cabinets to each other. As a result, it can be difficult to pinpoint malfunctions in the interfaces between sending and receiving units. We want PTM-XC systems to be easier to maintain, allowing technicians to quickly find and repair any malfunctioning parts.

Furthermore, semiconductor devices have become more integrated and intricate in response to recent demands for equipment that is smaller and consumes less power. This has made it possible for soft errors (such as bit errors in memory attributed to cosmic rays) to occur and cause malfunctions. We also want to address these soft errors.

3. Technical solutions to address these issues

3.1 High-efficiency path utilization

PTM-XC's cross-connect features use MPLS-TP paths (label switched paths) to forward information between packages, placing information about the destination in internal headers. Internally, the equipment uses MPLS-TP to encapsulate the two-way connection between the STM-0/STM-1 interface and the switch when forwarding data from the input port (A) to the output port (Z). By multiplexing several VC-11/VC-2 paths going in the same direction, we were able to keep the overhead for packetizing data in check. Also, we were able to use a single cabinet to accommodate the same number of circuits that would have required ten cabinets with existing equipment, as shown in **Fig. 5**.

3.2 Low-delay technology

To suppress the delays associated with packetizing and depacketizing data when paths are respectively multiplexed and demultiplexed, TU (tributary unit)-11/TU-2 frames are split when they are packetized. This has resulted in processing delays equivalent to or less than Module B.

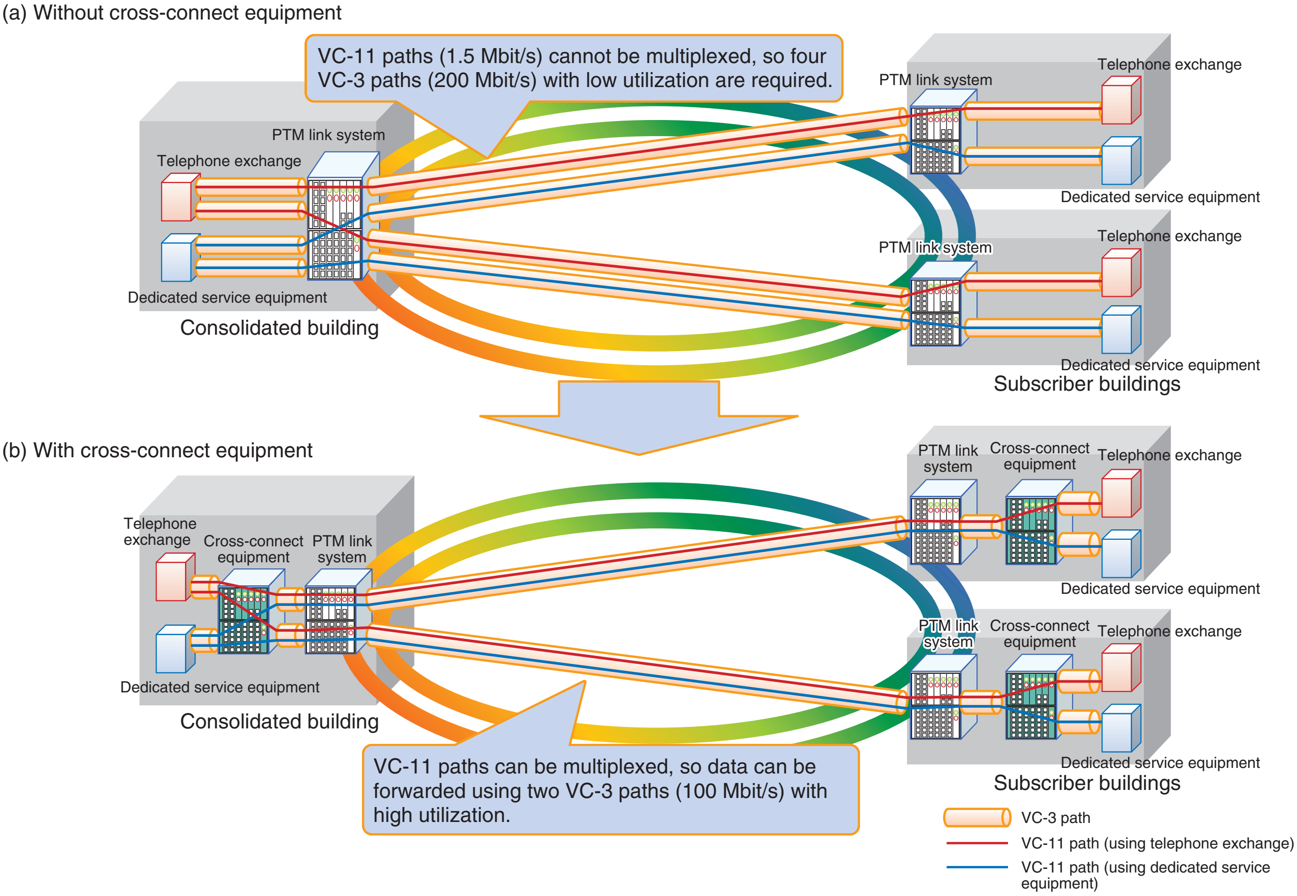
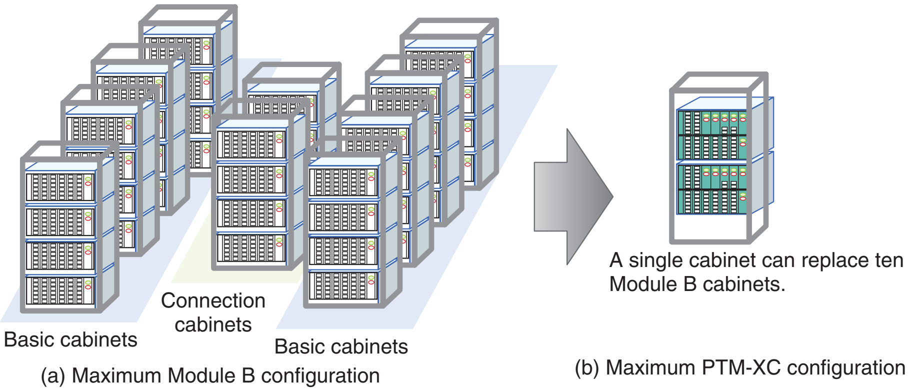


Fig. 4. Increasing path utilization with cross-connect equipment.



(a) Maximum Module B configuration

(b) Maximum PTM-XC configuration

Fig. 5. High-density PTM-XC implementation.

3.3 Reduced power consumption

We have managed to reduce the number of circuits from ten cabinets' worth with Module B, the existing equipment, to only a single cabinet's worth. This has allowed us to reduce the maximum power consumption levels to approximately 5% of Module B's.

3.4 Improved maintainability and operability

PTM-XC automatically registers monitoring controllers, switches, STM-0/STM-1 interfaces, and other packages in every slot when unit types are configured during the construction of new equipment. Multiplex section protection is automatically configured for all identical ports in adjacent slots when the STM-0/STM-1 interfaces are registered, reducing the steps necessary to open paths.

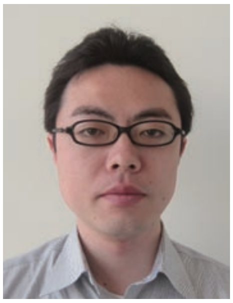
PTM-XC also does not require any complicated connections between cabinet interfaces because it only uses a single cabinet in its maximum configuration; instead, it simply uses STM-0/STM-1 interfaces and switches. As a result, PTM-XC is easier to troubleshoot than Module B, which could involve problems spanning multiple cabinets. PTM-XC furthermore provides the same path testing features as Module B, making it easy to pinpoint where errors have occurred.

In addition to everything mentioned above, PTM-XC takes measures to automatically correct soft errors that can occur in both main signal memory and monitoring controller memory. PTM-XC also provides performance-related information by counting the number of soft errors that occur and sending notifications when a given threshold of errors is exceeded.

When a package needs to be replaced, a remote technician can make that package's light-emitting diodes alternately blink red and green to point it out to an onsite technician.

4. Summary

In this article, we introduced PTM-XC equipment that was developed by NTT Network Service Systems Laboratories and is capable of splitting and reassembling individual low-speed (1.5 Mbit/s) paths. PTM-XC can make the process of migrating away from existing Module B equipment efficient and cost-effective. We are aiming to push forward on migrating to a link system with this equipment and support simple network structures in the future.



Masaya Ogawa

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

He received the B.S. and M.S. in electronics from Keio University, Kanagawa, in 2003 and 2005, respectively. He joined NTT EAST in 2005. He was involved in maintenance for transport networks at East Network Operation Center at NTT EAST. He has been in his current position since July 2009 and is engaged in developing the PTM link system and PTM-XC. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) and the Japan Society of Applied Physics (JSAP).

Hidenori Iwashita

Researcher, Transport Network Innovation Project, NTT Network Service Systems Laboratories.

He received the B.S. and M.S. in nuclear engineering from Hokkaido University in 2006 and 2008, respectively. Since 2008, he has been a researcher at NTT Network Service Systems Laboratories. He is engaged in research and development (R&D) of the PTM link system and PTM-XC. He is a member of IEICE.

Takashi Takashima

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

He received the B.S. and M.S. in applied physics from Tohoku University, Miyagi, in 1997 and 1999, respectively. He joined NTT in 1999 and has been mainly involved in planning transport networks at NTT EAST. He has been in his current position since July 2012 and is working on developing the PTM link system and PTM-XC.

Toshiya Matsuda

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

He received the B.S. and M.S. in electrical information and communication engineering from Waseda university, Tokyo, in 1990 and 1992, respectively. In 1992, he joined the NTT Transmission Systems Laboratories and began researching long-haul large-capacity transmission systems. His current research interests include metro area network systems. He is a member of IEICE.

Kenichi Higuchi

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

He received the B.S and M.S. from Tokyo University of Science in 1989 and 1991, respectively. He joined NTT in 1991. He has been engaged in research on digital switching systems and photonic network systems. He is currently with NTT Network Service Systems Laboratories, where he is involved in R&D of the PTM link system and PTM-XC. He is a member of IEICE.

Katsutoshi Koda

Director, Transport Network Innovation Project, NTT Network Service Systems Laboratories.

He received the B.S. and M.S. in mechanical sciences from Tokyo University of Science in 1987 and 1989, respectively. He joined NTT in 1989 and worked on the development of transport systems at the Network System Development Center. During 1999–2000, he was engaged in human resource management at NTT EAST. He worked in the global procurement office at NTT from 2001 to 2004 and then spent three years at NTT EAST before joining NTT Network Service Systems Laboratories. He has been in his current position since July 2014. He is managing the development project for PTM-XC as well as R&D for future optical networking and transport network management technology.

Recent Activities of ITU-R Study Group 5

Shinya Otsuki and Takumi Togi

Abstract

Radiocommunication systems such as mobile communication systems and fixed wireless communication systems play important roles in telecommunication networks. Mobile communication systems have become widespread, as represented by smartphones, and fixed wireless communication systems provide flexible and resilient communication in times of disasters. The International Telecommunication Union - Radiocommunication Sector (ITU-R) is one of the organizations responsible for developing standards for these important systems. Towards the end of 2015, the ITU-R will hold two very important meetings: the Radiocommunication Assembly (RA) and the World Radiocommunication Conference (WRC). This article introduces recent activities and contributions from NTT and NTT DOCOMO to ITU-R Study Group 5, which deals with terrestrial radiocommunication systems.

Keywords: ITU-R, Study Group 5, terrestrial radiocommunication service

1. Introduction

The International Telecommunication Union - Radiocommunication Sector (ITU-R) is the ITU sector that develops regulations and standards for radiocommunication systems. The structure of ITU-R is depicted in **Fig. 1**.

The Study Groups (SGs) work on developing technical, operational, and procedural bases for efficient use of the radio spectrum and the geostationary-satellite orbit and develop ITU-R Recommendations and Reports. They also carry out studies and provide solutions called *Methods* that are required in discussions at the World Radiocommunication Conference (WRC) on revising the Radio Regulations (RR)*1.

The structure and scope of the six SGs in the ITU-R [1] are listed in **Table 1**. The scope of SG 5 is terrestrial services, and Dr. Akira Hashimoto of NTT DOCOMO has been the chairman of this SG for the last two study periods (2007–2012 and 2012–2015).

As listed in **Table 2**, SG 5 has four Working Parties (WPs); the participants in these WPs meet periodically to discuss ITU-R Recommendations. NTT members regularly attend WP 5A and WP 5C meetings, which deal respectively with land mobile services and fixed services, and NTT DOCOMO members regularly attend WP 5D meetings, which deal with International Mobile Telecommunications (IMT) systems*2.

The ITU-R may also establish Task Groups (TGs) to deal with a specified theme for a short period, or Joint Task Groups (JTGs) to deal with a specific theme involving more than one SG. From 2012 through 2014, JTG 4-5-6-7 was established to discuss topics on WRC-15 agenda item 1.1, additional frequency bands for IMT. This article also introduces the activities of JTG 4-5-6-7.

2. Recent SG and WP activities

2.1 WP 5A

The scope of WP 5A is mainly land mobile radiocommunication systems excluding IMT. This WP

---

*1 Radio Regulations: The RR provide international rules and regulations for spectrum allocation to radio services, use of satellite orbits, and administrative and operational procedures for radio stations, all of which are needed for the use of radio waves.

*2 IMT systems: Includes 3rd generation mobile communication systems and beyond such as IMT-2000 and IMT-Advanced systems.

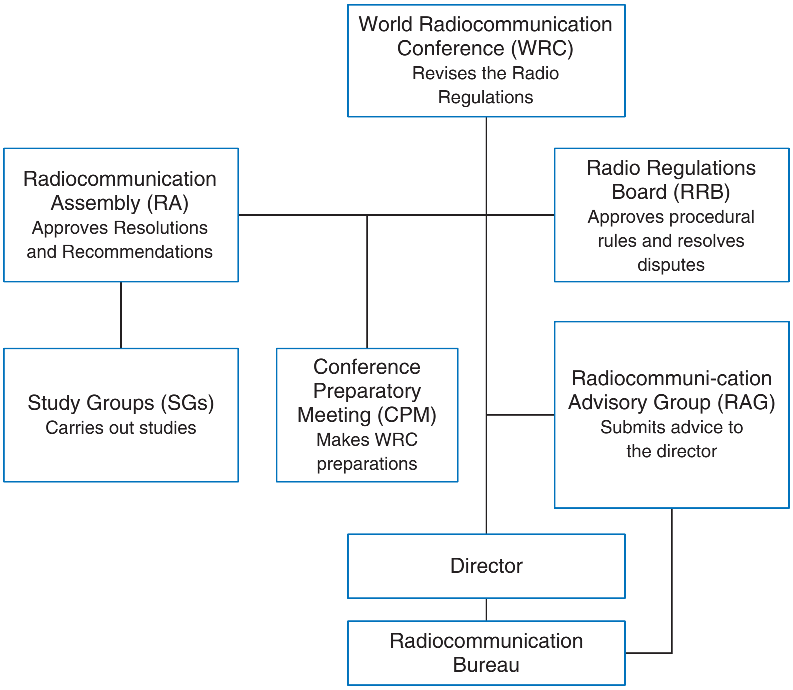


Fig. 1. Structure of ITU-R.

Table 1. Structure of ITU-R SGs.

	Scope of work
SG 1: Spectrum management	Spectrum management principles and techniques, general principles of sharing and spectrum monitoring
SG 3: Radiowave propagation	Propagation of radio waves in ionized and non-ionized media, and the characteristics of radio noise
SG 4: Satellite services	Systems and networks for fixed-satellite, mobile-satellite, broadcasting-satellite, and radiodetermination-satellite services
SG 5: Terrestrial services	Systems and networks for fixed, mobile, radiodetermination, amateur, and amateur-satellite services
SG 6: Broadcasting service	Radiocommunication broadcasting including vision, sound, multimedia, and data services principally intended for delivery to the general public
SG 7: Science services	Systems for space operation, space research, earth exploration and meteorology; systems for remote sensing including passive and active sensing systems, radio astronomy, and standard frequency and time signals

*Note that there is no SG 2.

carries out studies on wireless access systems (WAS) including wireless local area networks (WLANs) and public protection and disaster relief (PPDR) systems. It also studies new technologies for land mobile radiocommunication systems and develops ITU-R Recommendations and Reports on these topics.

At the October/November 2014 WP 5A meeting, work was completed on an ITU-R Report on cognitive radio systems (CRS)[*3]. These systems are

---

[*3] Cognitive radio system: A radio system employing technology that allows the system to obtain knowledge of its operational and geographical environment, to dynamically and autonomously adjust its operational parameters and protocols according to its obtained knowledge, and to learn from the results obtained.

Table 2. Structure of WPs in SG 5.

	Scope of Work
WP 5A	Land mobile service above 30 MHz (excluding IMT); wireless access in fixed services; amateur and amateur-satellite services
WP 5B	Maritime mobile service including Global Maritime Distress and Safety System (GMDSS); aeronautical mobile service and radiodetermination service
WP 5C	Fixed wireless systems; HF and other systems below 30 MHz in fixed and land mobile services
WP 5D	IMT systems

HF: high frequency

expected to improve efficient use of the spectrum allocated to land mobile communication systems.

This WP also carries out standardization activities on intelligent transport systems (ITS). One of the agenda items to be discussed at WRC-15 that concerns ITS is "Allocation of the band 77.5–78 GHz to the radiolocation service to support automotive short-range high-resolution radar operations," which is intended to achieve high resolution automotive radar for more precise collision avoidance by using the entire 77–81 GHz frequency range. In cooperation with WP 5B, which deals with radar, WP 5A has completed sharing and compatibility studies with other existing radiocommunication systems and provided solutions to achieve co-existence with these systems. These activities are in line with Japan's position of supporting new allocations for the 77.5–78 GHz frequency band.

2.2 WP 5B

The scope of WP 5B is the maritime mobile service, aeronautical mobile service, and radiodetermination service (e.g., radar), and this WP deals with a lot of topics discussed at WRCs. Although telecom operators have made few contributions to this WP, the NTT Group addresses these topics via liaison statements from other WPs that reflect the Group's opinions.

2.3 WP 5C

The scope of WP 5C is mainly fixed wireless systems (FWSs). This WP is developing ITU-R Recommendations and Reports on transport/trunking wireless network systems, fixed wireless access (FWA) systems, mobile backhaul (MBH) for land mobile radiocommunication, and systems for temporary use in disaster relief.

At the October/November 2014 WP 5C meeting, work was finalized on a new ITU-R Report on fixed service use and future trends [2]. The details of this ITU-R Report are described later in this section.

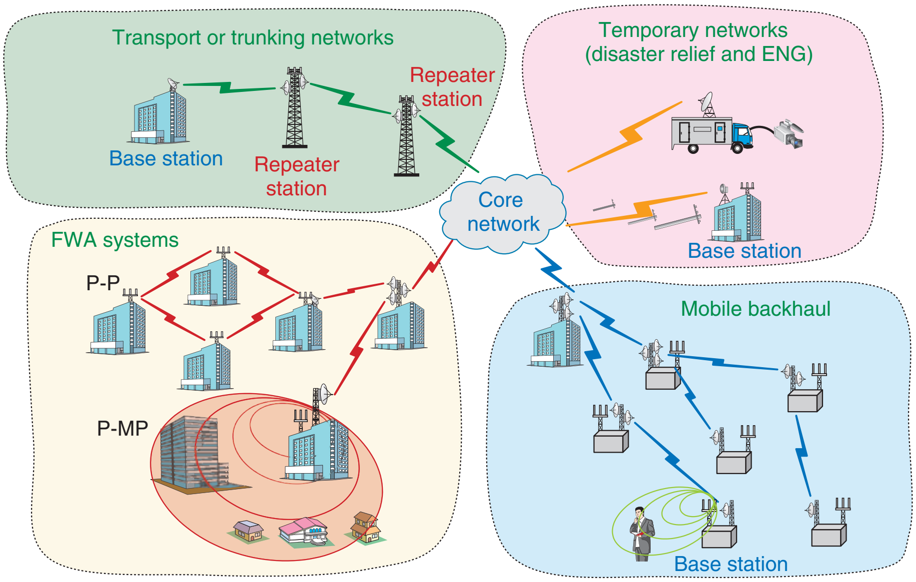
One study currently being carried out in WP 5C is aimed at developing a new ITU-R Report on technical characteristics for MBH such as frequency bands, network topologies, and required capacity. The WP has been developing this report through liaison statements in cooperation with WP 5D and ITU-T SG15, which deal respectively with IMT and fiber-optic transport networks.

In this section, we summarize the abstract of the new ITU-R Report on the aforementioned fixed service use and future trends. FWSs are used in telecommunication networks in various situations. FWSs are used for transport/trunking networks, MBH networks, FWA systems, and temporary networks, as shown in **Fig. 2**. The demand for MBH capacity in such networks is increasing with the explosive increase in traffic in mobile communication systems due to the introduction of smartphones.

To fulfill this demand, studies on FWSs using millimeter-wave bands (above 30 GHz) are being carried out, since a broader bandwidth is available in these bands than in the frequency bands below 30 GHz. Report ITU-R F.2323, "Fixed service use and future trends," is aimed at providing guidance on the future development of fixed services by taking into account the evolution of current use and technology development, application trends for FWSs, and future requirements for FWSs. This Report covers the following areas: (1) FWS use in telecommunication networks, (2) FWS band usage, (3) FWS technologies and trends, and (4) spectrum aspects and requirements. It was NTT that led the way in developing the report and that made a lot of contributions to it. These contributions included applications making use of FWA technologies as follows: (1) an example of an MBH network to provide Internet access service for passengers in trains on the Tsukuba Express Line in Japan, (2) home network systems in the 60-GHz frequency band as an extension of FWA systems, and (3) gigabit millimeter-wave links in the 120-GHz frequency band. The above contributions were reflected in this ITU-R Report.

2.4 WP 5D

The scope of WP 5D is IMT, and this WP deals with IMT radio interfaces, IMT frequency arrangements (frequency use within the ranges identified for IMT), sharing and compatibility studies between IMT and other radiocommunication systems, and the



ENG: electronic news gathering
P-P: point to point
P-MP: point to multipoint

Fig. 2. Applications of FWS.

development process for and concepts of future IMT (fifth generation (5G) of mobile communications). This WP develops ITU-R Resolutions, ITU-R Recommendations, and ITU-R Reports on these topics. The WP has been the recipient of many NTT DOCOMO contributions including studies on 5G. The following is a review of the studies currently being carried out on 5G, which is expected to be introduced around the year 2020.

(1) Current studies toward 5G

Studies on IMT-2000[*4] (3G) and IMT-Advanced[*5] (4G) radio interfaces have been completed by WP 5D (and by former WP 8F, the predecessor of WP 5D), which has also adopted numerous technical specifications proposed by external standards developing organizations. The WP will complete the development of specifications for IMT-2020[*6] radio interfaces by the year 2020. In addition, it is actively discussing ITU-R Resolutions for the future development process of IMT, including IMT-2020, and an ITU-R Recommendation (Recommendation on Vision) on use cases, applications, requirements, and technology trends of IMT-2020. Studies on IMT-2000 and IMT-Advanced were conducted following the same process. However, for IMT-2020 the duration from initial studies to deployment is projected to be reduced to 8 years (as opposed to 15 years for IMT-2000), as shown in **Fig. 3**.

(2) Recommendation on 5G Vision

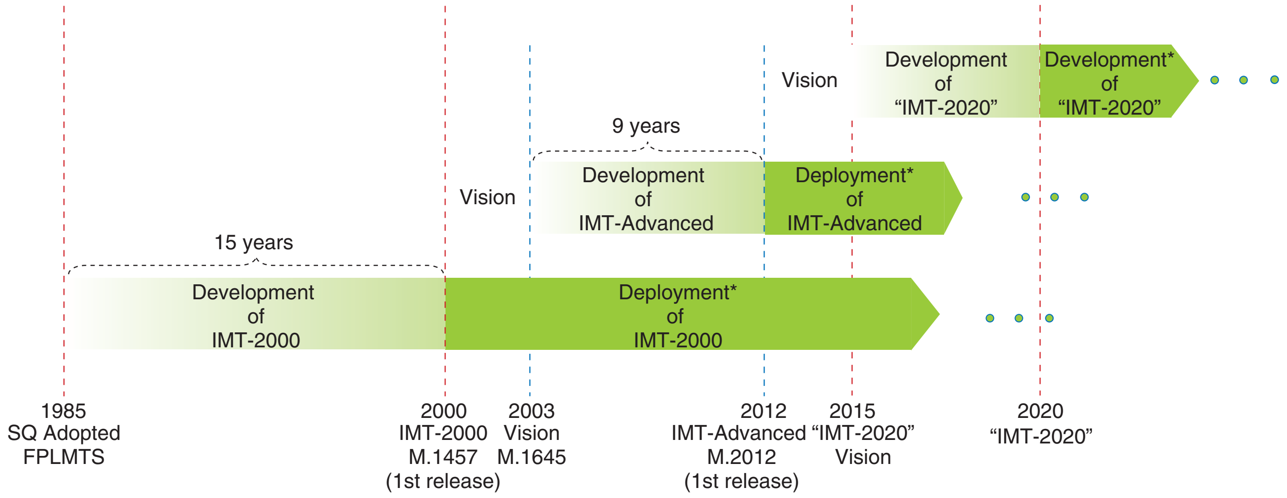
Three main use cases are assumed for IMT-2020 in the Recommendation on Vision. One is for further enhancement of mobile broadband devices such as smartphones (Enhanced Mobile Broadband). Another is for communications involving a massive number of mobile terminals in the era of Internet of Things (IoT) and Internet of Everything (IoE) (massive machine-type communications). The other is for communications required for high reliability such as vehicle to vehicle (V2V) communication and public safety communication systems (ultra-reliable and low latency communications). The Recommendation on Vision contains considerations for eight parameters for the basic requirements, while Recommendation ITU-RM.1645 [3] on the Vision of IMT-2000 and

---

[*4] IMT-2000: W-CDMA (Wideband Code Division Multiple Access), HSPA (High-Speed Packet Access), and LTE (Long Term Evolution) are included as radio interfaces.

[*5] IMT-Advanced: LTE-Advanced is included as a radio interface.

[*6] IMT-2020: The term “3G” is used in ITU and in related organizations all over the world. However, the term “4G” is not used in ITU since there are certain ambiguities in how it was generated, likewise for the term “5G.” The term “IMT-2020” is provisionally used in this article, but “IMT-2020 Connect” is also a candidate term. The formal name will be approved within ITU in October 2015.



*Deployment timing may vary across countries.

FPLMTS: Future Public Land Mobile Telecommunication Systems
SQ: Study Question

Fig. 3. Overview of timeline for IMT development and deployment (derived from the Recommendation on Vision).

Table 3. Comparison of system requirements for IMT-2020 and IMT-Advanced.

	Rec. ITU-R M.1645 (IMT-2000/IMT-Advanced)	Rep. ITU-R M.2134 (IMT-Advanced)	Rec. on Vision (IMT-2020)
Year of approval	2003	2008	2015
Mobile terminal mobility	250 km/h	350 km/h	500 km/h
Peak transmission rate	0.1–1 Gbit/s	1 Gbit/s	20 Gbit/s
User throughput	–	10 Mbit/s	0.1–1 Gbit/s
Mobile terminal density	–	10^5/km^2	10^6/km^2
Radio link latency	–	10 ms	1 ms
Power efficiency (/bit)	–	–	100 times that of IMT-Advanced
Spectrum use efficiency	–	–	2–5 times that of IMT-Advanced
Area traffic capacity	–	0.1 Mbit/s/m^2	10 Mbit/s/m^2

IMT-Advanced specified only mobility and peak data rate*7. This shows that the uses of IMT are expected to become wider and more varied with the further development of IMT technologies (**Table 3**). It should be noted that the values shown in the table are tentative and will be discussed and finalized through discussions after completion of the Recommendation on Vision.

2.5 JTG 4-5-6-7

At the previous WRC (WRC-12), an agenda item on additional spectrum allocations and identifications for IMT was adopted and specified as Agenda Item 1.1 for WRC-15. Since this item covers a variety of candidate frequency ranges and involves a lot of existing services in those ranges, JTG 4-5-6-7 was established as the group responsible for it. In contrast, WP 5D studied the spectrum requirements for IMT and the parameters for sharing and compatibility studies between IMT and other radiocommunication systems. JTG 4-5-6-7 took the study results of WP 5D into account and carried out sharing and compatibility studies between IMT and other radiocommunication systems. It also compiled candidate frequency bands to be identified for IMT and provided draft

---

*7 Requirements were updated in 2008 in Report ITU-R M.2134 after Recommendation ITU-R M.1645 was developed.

texts for the revision of RR in the case of an IMT identification. As a result of these activities, the group was able to prepare draft texts for the Conference Preparatory Meeting (CPM).

(1) Spectrum requirements for IMT

It has been agreed that spectrum requirements for IMT in 2020 will be 1340 MHz and 1960 MHz for lower and higher user density settings, respectively. These values are based on a number of contributions NTT DOCOMO members made to WP 5D; NTT DOCOMO also took the lead in developing methodologies and tools for estimating spectrum requirements. These results were approved as Report ITU-R M.2290 [4] at the December 2013 SG 5 meeting.

(2) Candidate frequency bands for IMT

On the basis of contributions made from a number of countries, agreement has been reached on the following candidate frequency bands for IMT:

470–694/698 MHz, 1350–1400 MHz, 1427–1452 MHz, 1452–1492 MHz, 1492–1518 MHz, 1518–1525 MHz, 1695–1710 MHz, 2700–2900 MHz, 3300–3400 MHz, 3400–3600 MHz, 3600–3700 MHz, 3700–3800 MHz, 3800–4200 MHz, 4400–4500 MHz, 4500–4800 MHz, 4800–4990 MHz, and 5925–6425 MHz.

It is expected that at WRC-15, some frequency bands will be identified for IMT from the above list, and it should be noted that the C-band (3400–4200 MHz and 4400–4900 MHz) and the L-band (1427.9–1462.9 MHz and 1475.9–1510.9 MHz), which are supported by Japan, are included in the candidate frequency bands.

(3) Sharing and compatibility studies between IMT and other radiocommunication systems

A number of sharing and compatibility studies have been carried out by JTG 4-5-6-7 on IMT and other radiocommunication systems that use the candidate frequency bands. From the results, the group has identified sharing conditions such as separation distance, transmission power limit, interference mitigation techniques, and feasibility of sharing. The parameters (power, antenna height, and densities of base stations and mobile terminals) the group used in the studies were discussed in WP 5D on the basis of contributions made mainly by NTT DOCOMO and were approved as Report ITU-R M.2292 [5] at the December 2013 SG 5 meeting.

For the C-band, the main subjects have involved sharing issues between IMT and fixed-satellite systems. A number of contributions have been made by NTT DOCOMO regarding sharing studies and interference mitigation techniques such as the use of IMT base stations placed at low height with low transmission power and small cells. These contributions have helped to show how the interference into fixed-satellite systems from IMT systems can be mitigated.

For the L-band, one of the serious issues is unwanted emissions from IMT into Earth Exploration Satellite Service (EESS) using the 1400–1427 MHz band, which is adjacent to the L-band. Japan contributed the measured values of unwanted emissions from IMT base stations and mobile terminals already in operation in Japan. These values show that it is possible to deploy IMT without harmful interference to EESS operated in the adjacent frequency bands, and these results are properly reflected in the relevant documents.

3. Activities toward the 2015 meetings

ITU-R Resolutions on the name and development process of the 5th generation of IMT and the Recommendation on Vision will be adopted at the July 2015 SG 5 meeting. It is expected that the ITU-R Resolutions will be approved at RA-15 to be held in October 2015.

The results of considerations compiled by JTGs regarding Agenda Item 1.1 were finalized at the 2nd session of CPM (CPM15-2) as the CPM Report[*8]. Using this Report as a basis, WRC-15 will make decisions on the identification of frequency bands and associated conditions for IMT.

4. Future subjects

The ITU-R will soon hold the two most important meetings of 2015, RA-15 and WRC-15. During the latter meeting, discussions will be held on the agenda item most important to Japan, which regards additional IMT allocations and identifications. There will also be discussions on other agenda items for new allocations of the frequency bands that the NTT Group wireless systems use. It may be required that these frequency bands be shared with other wireless systems internationally.

Agenda items for the future WRC (expected to be held in 2019) will also be discussed at WRC-15. Additionally, NTT DOCOMO will propose a new agenda item regarding additional IMT allocations and identifications beyond the year 2020 under

---

[*8] CPM Report: CPM Reports submitted to the WRC contain technical studies, solutions based on the studies, and examples of RR revisions. The reports are finalized during the CPM, which is held half a year before the WRC.

instructions from the Ministry of Internal Affairs and Communications. Since demand for spectrum is growing year by year, the future WRC is expected to see a variety of discussions on much wider frequency ranges.

This year and in years to come, NTT and NTT DOCOMO will continue to actively participate in ITU-R, always taking into account how best use of the frequencies can be made by the NTT Group.

References

[1] Structure of Radiocommunication Study Groups, ITU-R. http://www.itu.int/pub/R-RES-R.4

[2] Report ITU-R F.2323, http://www.itu.int/pub/R-REP-F.2323

[3] Recommendation ITU-R M.1645, http://www.itu.int/rec/R-REC-M.1645

[4] Report ITU-R M.2290, http://www.itu.int/pub/R-REP-M.2290

[5] Report ITU-R M.2292, http://www.itu.int/pub/R-REP-M.2292

Shinya Otsuki

Research Engineer, Wireless Access Systems Project, NTT Access Network Service Systems Laboratories.

He received the B.E., M.E., and Ph.D. from Osaka University in 1993, 1995, and 1997, respectively. He joined NTT in 1997 and studied wireless access systems, wireless local area network systems, and wireless systems for Internet services in trains. During 2008–2011, he worked on studies and international standardization efforts in evolved packet core and services using IP (Internet protocol) multimedia subsystems at NTT Service Integration Laboratories. Since 2011, he has been with NTT Access Service Systems Laboratories, where he is engaged in studies on wireless local area networks and international standardization efforts. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) and the Institute of Electrical and Electronics Engineers (IEEE).

Takumi Togi

Manager, Wireless Standardization Department, NTT DOCOMO.

He received the B.E. in advanced chemistry and the M.E. in materials engineering from Hokkaido University in 1998 and 2000, respectively. He joined NTT DOCOMO in 2000. He is currently responsible for spectrum planning and coordination issues for IMT systems and ITU-R activities as well.

External Awards

MTRAJ Award

Winner: Akira Ito, NTT Access Network Service Systems Laboratories; Hiroyuki Saito, Tokyo Denki University; Takanobu Suzuki, Toyo University
Date: April 20, 2015
Organization: Material Testing Research Association of Japan (MTRAJ)

For "Quantification of Seismic Performance Deterioration Caused by Corrosion on a Telecommunication Conduit."
Published as: A. Ito, H. Saito, and T. Suzuki, "Quantification of Seismic Performance Deterioration Caused by Corrosion on a Telecommunication Conduit," Journal of Material Testing Research Association of Japan, Vol. 59, No. 4, pp. 188–193, Oct. 2014.

2014 LOIS Research Award

Winner: Shigeki Takeuchi and Goro Inomae, NTT Service Evolution Laboratories; Manabu Okamoto, NTT Media Intelligence Laboratories; and Hiroyuki Tate, NTT Service Evolution Laboratories
Date: April 22, 2015
Organization: Institute of Electronics, Information and Communication Engineers (IEICE), Information and Systems Society, Technical Committee on Life Intelligence and Office Information Systems (LOIS)

For "The Tourist Information System for the Route Bus with Smartphone—the Evaluation Result Report on the Bus Line—."

Buses for tourists are used in tourist resorts. We are developing a system that provides passengers with route sightseeing information so that buses do not need a guide on board or special digital signage equipment. In this paper, a tourist information system is proposed that is integrated with the location of a bus and shows users necessary details from a large amount of sightseeing information. An evaluation using an experimental system is also described.
Published as: S. Takeuchi, G. Inomae, M. Okamoto, and H. Tate, "The Tourist Information System for the Route Bus with Smartphone—the Evaluation Result Report on the Bus Line—," IEICE Tech. Rep., Vol. 114, No. 150, LOIS2014-15, pp. 23–28, Jul. 2014.

45th Senken Gousen Prize New Frontier Award

Winner: Nippon Telegraph and Telephone Corporation and Toray Industries, Inc.
Date: April 24, 2015
Organization: Senken Shimbunsha

For the development of conductive fabric called *hitoe* that enables continuous measurement of the biological signals of the person wearing it.

IEICE Communications Society Best Tutorial Paper Award (Japanese edition)

Winner: Yusuke Asai, NTT Network Innovation Laboratories; Koichi Ishihara, Tomoki Murakami, Riichi Kudo, Takeo Ichikawa, Yasushi Takatori, and Masato Mizoguchi, NTT Access Network Service Systems Laboratories
Date: May 15, 2015
Organization: IEICE, Communications Society

For "Overview of Very High Throughput Wireless LAN Standard IEEE 802.11ac and Experimental Evaluation of Multiuser-MIMO Transmission."
Published as: Y. Asai, K. Ishihara, T. Murakami, R. Kudo, T. Ichikawa, Y. Takatori, and M. Mizoguchi, "Overview of Very High Throughput Wireless LAN Standard IEEE 802.11ac and Experimental Evaluation of Multiuser-MIMO Transmission," The Transactions of the Institute of Electronics, Information and Communication Engineers. B, Vol. J97-B, No. 1, pp. 1–18, Jan. 2014.

ITU-AJ International Corporation Award

Winner: Atsushi Takahara and Toshikazu Sakano*, NTT Network Innovation Laboratories

* Currently, he is with Advanced Telecommunications Research Institute International (ATR).

Date: May 15, 2015
Organization: The ITU Association of Japan (ITU-AJ)

For the development of the information and communication technology (ICT) architecture called the movable and deployable ICT resource unit (MDRU) that enables instant recovery of telecommunications services in damaged areas. In addition, they conducted a feasibility study on the effectiveness of MDRUs in disaster-stricken areas in the Philippines under an International Telecommunication Union (ITU) project. They also actively worked on standardization efforts within ITU in order to promote broad use of MDRUs.

ISCIE Encouragement Prize

Winner: Masaya Murata, NTT Communication Science Laboratories
Date: May 21, 2015
Organization: Institute of Systems, Control and Information Engineers (ISCIE)

For "Gaussian Unscented Filter."
Published as: M. Murata, H. Nagano, and K. Kashino, "Gaussian Unscented Filter," Proc. of the 46th ISCIE International Symposium on Stochastic Systems Theory and Its Applications (SSS'14), Kyoto, Japan, Nov. 2014.

Papers Published in Technical Journals and Conference Proceedings

A Spectrum- and Energy-efficient Scheme for Improving the Utilization of MDRU-based Disaster Resilient Networks

T. Ngo, H. Nishiyama, N. Kato, T. Sakano, and A. Takahara

IEEE Transactions on Vehicular Technology, Vol. 63, No. 5, pp. 2027–2037, June 2014.

The lack of spectrum and energy resources is a major problem in movable and deployable resource unit (MDRU)-based networks, where the throughput requirement is very high. In this paper, we propose a scheme to improve the utilization of both spectrum and energy resources to increase performance in the gateway portion of the MDRU-based network.

Attenuators Enable Inversely Proportional Setting Transmission Power and Carrier Sense Threshold in WLANs

D. Okuhara, F. Shiotani, K. Yamamoto, T. Nishio, M. Morikura, R. Kudo, and K. Ishihara

Proc. of the 25th International Symposium on Personal Indoor and Mobile Radio Communications (IEEE PIMRC 2014), pp. 986–990, Washington DC, USA, September 2014.

The insertion of attenuators between transceivers and antennas is proposed in order to improve spatial channel reuse in carrier sense multiple access with collision avoidance based wireless local area networks (WLANs). Using attenuators enables the access points or stations to reduce the power levels of transmission signals and received signals, which results in an increase in the carrier sense threshold. Thus, the attenuation value control enables inversely proportional setting of the transmission power and carrier sense threshold, which is known to provide a novel solution for the unfairness problem caused by variable transmission power or variable carrier sense threshold. We first derive a simple and effective condition such that the aggregate spectral efficiency is higher when attenuators are used than they are when using some approximations. Through a numerical evaluation as well as a testbed for using attenuators, the condition is shown to be valid despite using the approximations, and throughput improvement is verified. The result from the testbed also disclosed a new unfairness problem due to individual differences in transmission power or carrier sense threshold.

AP Cooperative Diversity in Wireless Network Using Interference-aware Channel Segregation Based Dynamic Channel Assignment

M. T. H. Sirait, Y. Matsumura, K. Temma, K. Ishihara, B. A. H. S. Abeysekera, T. Kumagai, and F. Adachi

Proc. of IEEE PIMRC 2014, pp. 1185–1189, Washington DC, USA, September 2014.

Recently, we proposed interference-aware channel segregation based dynamic channel assignment (IACS-DCA) which forms a channel reuse pattern with low co-channel interference (CCI) in a distributed manner. The transmission performance of a wireless station (STA) located far from access points (APs) degrades due to path loss and shadowing loss. AP cooperative diversity is a well-known technique to improve the transmission performance. Since IACS-DCA operates so as to assign different channels to different APs located nearby each other, it is not an easy task to introduce AP cooperative diversity to a wireless network using IACS-DCA. In this paper, we propose an AP grouping and channel selection method for AP cooperative diversity in a wireless network using IACS-DCA. By computer simulation, we show that the proposed AP cooperative diversity can improve the transmission performance.

Broadband Spectrum Sensing Platform Based on Received Waveform Cross-correlation Using Distributed Sensors

T. Yamada, D. Lee, H. Shiba, Y. Yamaguchi, T. Kaho, T. Nakagawa, and K. Uehara

Proc. of SmartCom (Singapore - Japan International Workshop on Smart Wireless Communications) 2014, Singapore, October 2014.

The rapid increase in mobile traffic due to the spread of smartphones is expected to cause a frequency shortage. The frequency range from 400 MHz to 6 GHz has been allocated to various wireless systems in Japan. Therefore, the use of white space technology is expected, and many studies on it have been reported. We present a broadband spectrum sensing platform for a white space system that covers a frequency range from 400 MHz to 6 GHz. In our platform, we elaborated and implemented a cooperative sensing method that exploits cross-correlations of received signals among distributed sensors to estimate the received signal strength. This method enables us to detect signals under the noise floor (e.g., –114 dBm/6 MHz according to Federal Communications Commission rules), to identify transmitters, and to recognize the border of radio coverage. We developed a prototype of our platform and successfully verified its performance.

An Advanced Wi-Fi Data Service Platform Coupled with a Cellular Network for Future Wireless Access

R. Kudo, Y. Takatori, B. A. H. S. Abeysekera, Y. Inoue, A. Murase, A. Yamada, H. Yasuda, and Y. Okumura

IEEE Communications Magazine, Vol. 53, No. 11, pp. 46–53, November 2014.

Wireless LAN (local area network) devices are now everywhere because of the rapid spread of smart wireless devices. Demand for far richer content services is also driving the expansion of mobile traffic. Converging the cellular network with Wi-Fi is a reasonable way to support the increasing mobile traffic because most mobile user terminals already have Wi-Fi interfaces. Creating more opportunities for Wi-Fi use will require further enhancement of system capacity and manageability, especially in the high-density Wi-Fi network. This is because the chronic depletion of system resources is becoming a significant problem in the Wi-Fi network given the increases in Wi-Fi density and traffic. This article introduces a Wi-Fi data service platform coupled with cellular networks, which strengthens the synergy of two networks. Enhanced monitoring and performance prediction are essential to provide a high-grade user experience in high-density Wi-Fi environments.

Cooperative Inter-cell Interference Mitigation Scheme with Downlink MU-MIMO Beamforming for Dense Wireless LAN Environment

K. Ishihara, T. Murakami, Y. Asai, Y. Takatori, and M. Mizoguchi

Wireless Personal Communications, DOI: 10.1007/s11277-014-2220-2, December 2014.

A cooperative inter-cell interference (ICI) mitigation scheme with transmit beamforming for dense wireless LAN systems is proposed. The proposed scheme applies transmit beamforming used for downlink multi-user multiple-input and multiple-output (MU-MIMO) in order to mitigate the effect of ICI and selectively determines whether an access point (AP) performs null beamforming for each station (STA) in overlapping basic service sets (OBSSs) according to the ICI power. Null beamforming is used to suppress ICI if its power exceeds a threshold; otherwise, it is not carried out and the transmit antenna is used to obtain a diversity gain for STAs associated with the AP. Computer simulations confirmed that the achievable rate obtained with the proposed scheme is superior to that obtained with either time resource sharing or conventional ICI mitigation in an OBSS environment.

A Novel Application of Massive MIMO: Massive Antenna Systems for Wireless Entrance (MAS-WE)

K. Maruta, A. Ohta, S. Kurosaki, T. Arai, and M. Iizuka

Proc. of ICNC (International Conference on Computing, Networking and Communications) 2015, Anaheim, USA, February 2015.

This paper proposes a practical application of Massive MIMO technology, Massive Antenna Systems for Wireless Entrance (MAS-WE), and its related inter-user interference (IUI) cancellation and scheduling techniques. MAS-WE, on which the entrance base station (EBS) employs a large number of antennas, can effectively provide high capacity wireless entrance links to a large number of Wi-Fi access points (APs) distributed in a wide coverage area. The proposed techniques have been ultimately simplified to have less impact on their practical implementation in order to spatially multiplex more than 16 signal streams with around 100 antenna elements on the EBS side. The SIR (signal-to-interference ratio) performance was evaluated by system level simulation considering imperfect channel state information (CSI), and the results showed that the proposed MAS-WE with simplified techniques can achieve high spectral efficiency with high level space division multiplexing.

Experimental Validation of Digital Pre-distortion Technique for Dual-band Dual-signal Amplification by Single Feedback Architecture Employing Dual-band Mixer

I. Ando, G. K. Tran, K. Araki, T. Yamada, T. Kaho, Y. Yamaguchi, and T. Nakagawa

IEICE Transactions on Electronics, Vol. E98-C, No. 3, pp. 242–251, March 2015.

In this paper, we propose and experimentally validate a dual-band digital predistortion (DPD) model that takes account of the intermodulation and harmonic distortion produced when the center frequencies of input bands have a harmonic relationship. We also describe and experimentally validate our proposed novel dual-band power amplifier (PA) linearization architecture consisting of a single feedback loop employing a dual-band mixer. Experiment results show that the DPD linearization the proposed model provides can compensate for intermodulation and harmonic distortion in a way that the conventional two-dimensional (2-D) DPD approach cannot. The proposed feedback architecture should make it possible to simplify analog-to-digital converter (ADC) design and eliminate the time lag between different feedback paths.

Different Roles of the COMT and HTR2A Genotypes in Working Memory Subprocesses

H. M. Kondo, M. Nomura, and M. Kashino

PLOS ONE, Vol. 10, No. 5, e012651, May 2015.

The present study used an imaging genetics approach to examine the interaction between neurochemical functions and working memory performance. We focused on functional polymorphisms of the catechol-O-methyltransferase (COMT) Val158Met and serotonin 2A receptor (HTR2A) -1438G/A genes, and devised a delayed recognition task to isolate the encoding, retention, and retrieval processes of visual information. The COMT genotypes affected recognition accuracy, whereas the HTR2A genotypes were associated with response times.

Dense Space Division Multiplexed Transmission Technology

T. Mizuno, H. Takara, A. Sano, and Y. Miyamoto

CLEO (Conference on Lasers and Electro-Optics) 2015, SW4M. 1, San Jose, USA, May 2015.

We review recent progress in ultra-high capacity transmission based on dense space division multiplexing (DSDM) for future scalable optical transport networks and present the latest multi-core multi-mode fiber, spatial multi/demultiplexers, and MIMO signal processing technique.