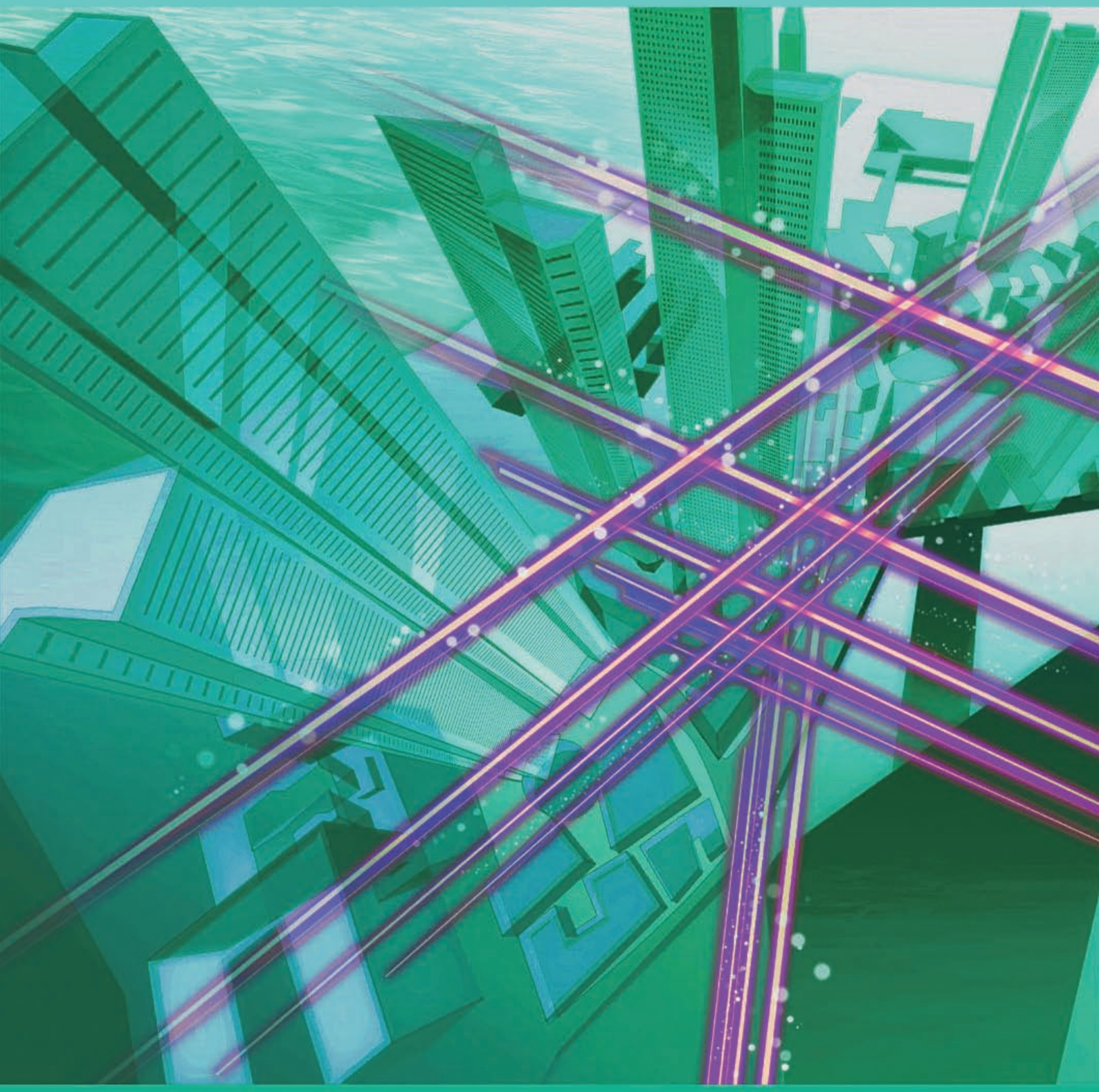
NTT Technical Review

4
2015



April 2015 Vol. 13 No. 4

NTT Technical Review

April 2015 Vol. 13 No. 4

Value Creation by Collaboration

Hiroo Unoura
Representative Director and President, Chief Executive Officer, NTT

Overview

This article introduces NTT Group initiatives towards 2020 centered on global cloud business and network services. The content of this article is based on the keynote lecture presented by NTT Representative Director, President and Chief Executive Officer Hiroo Unoura at NTT R&D Forum 2015 held February 19–20, 2015.

Keywords: cloud, collaboration, cybersecurity

1. 30 years since NTT privatization

The year 2015 marks 30 years of NTT being a private corporation. NTT's consolidated revenues, which were a little over 5 trillion yen in the first year of its privatization, now exceed 11 trillion yen. In this regard, I'd like to acknowledge the efforts made by each company in the NTT Group and our customers' gracious support over the years. During this time, our business structure has changed considerably. I think that we have become a completely different group in the last 15 years compared to the first 15 years of this period.

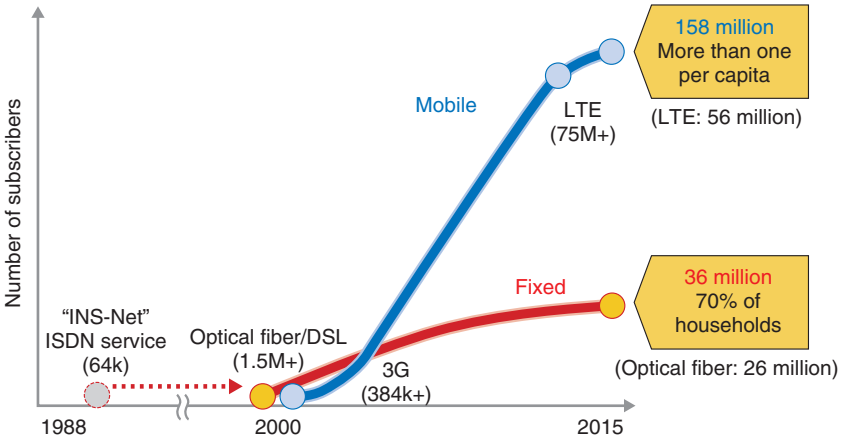
It has also been 15 years since NTT was reorganized under a holding company system. In other words, this year marks a milestone for NTT as it reaches 15 years since its transition to a holding company system and 30 years since its transformation into a private company. Furthermore, in January of this year, NTT became the first Gold Partner of the Tokyo 2020 Olympic and Paralympic Games. I believe that this will provide us with a great opportunity to take on a variety of challenging issues with a renewed sense of purpose.

One major change over these last 30 years has been the penetration of broadband access. The number of mobile broadband subscribers for mobile phones and smartphones now stands at about 158 million, which equates to more than one broadband-capable handset per capita in Japan. The number of fixed broadband subscribers, meanwhile, is now about 36 million, which means that broadband has spread to about 70% of all households in Japan, with about 26 million of these subscribers using optical fiber (**Fig. 1**).

This evolution of broadband services corresponds to an improvement in transmission speed. In mobile broadband, the transmission speed has jumped by about 600 times from the 384 kbit/s of 3G services to the 225 Mbit/s maximum bit rate of Long-Term Evolution (LTE)-Advanced services (launched in March 2015) provided by NTT DOCOMO. In fixed broadband, the transmission speed has increased by about 16,000 times from the 64 kbit/s of ISDN (Integrated Services for Digital Network) to the 1 Gbit/s of optical services that are now being provided. To be sure, these amazing improvements in transmission speed have been instrumental in bringing about a variety of changes in our business.

2. Evolution towards the cloud

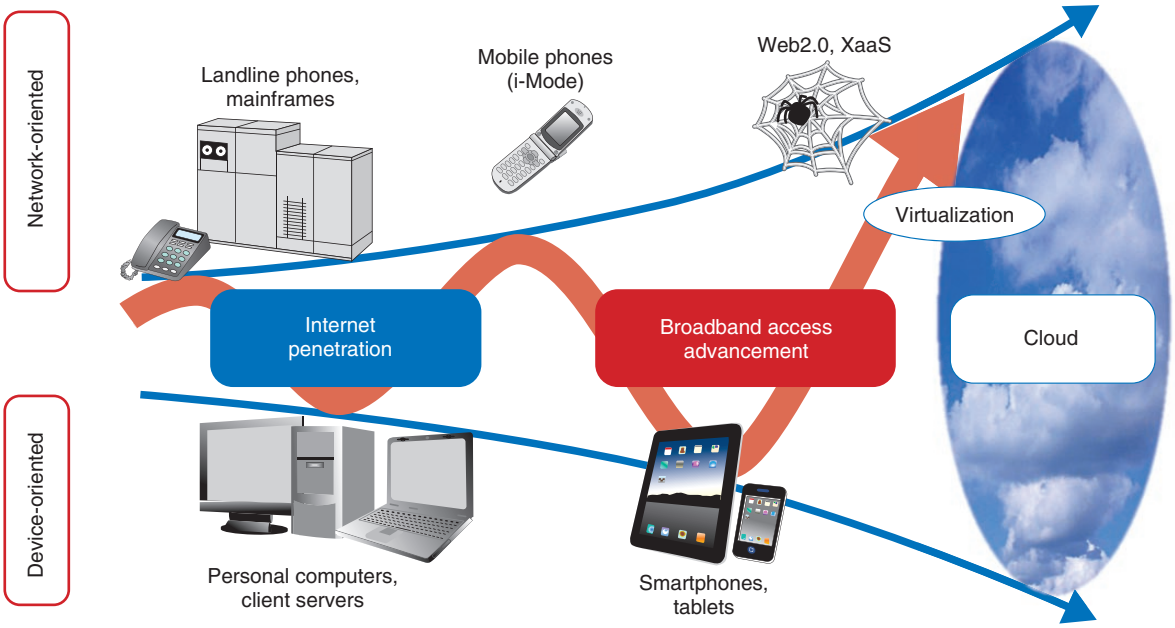
The evolution of broadband services has had profound effects on the user's information and communication technology (ICT) environment. As you probably know, the greatest of these is evolution towards the cloud (**Fig. 2**). After the appearance of mainframes,



*Mobile: 3rd generation mobile phone, LTE, BWA (broadband wireless access) Fixed: Optical fiber, DSL, CATV (cable television)
Source: "Number of Telecommunications Service Subscriptions and Market Share (As of the end of September 2014)," Ministry of Internal Affairs and Communications.

DSL: digital subscriber line
ISDN: Integrated Services for Digital Network
LTE: Long-Term Evolution

Fig. 1. Penetration of broadband access.



XaaS: X as a service

Fig. 2. Evolution towards the cloud.

advances in personal computers and terminals brought about a shift in the ICT system configuration from a centralized type to a distributed type, but with the emergence of mobile services, the ICT system reverted once again to a centralized configuration. In this way, the development of communications has involved passing through repeated cycles of centralized and distributed operations. However, I believe

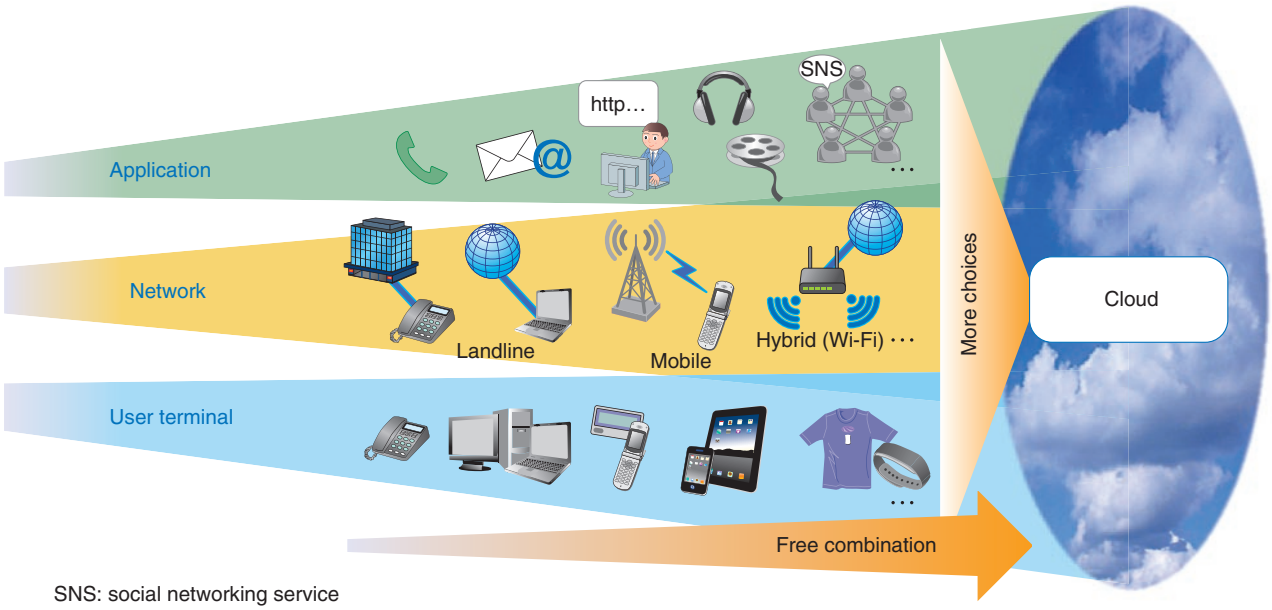


Fig. 3. The collaboration era.

that the present cloud could mark an end to these cycles and represent the final format for communications.

2.1 Towards the collaboration era

How, then, have cloud services affected the business of the NTT Group? For users, the cloud era has expanded the range of carriers and vendors to choose from and has made it easy to combine them as desired. It also means that users can immediately switch to a newly available service and do so at very low cost.

In the cloud era, how should competitiveness be defined? Since users now have a greater number of choices, we could say that being competitive means the ability to be selected. However, being chosen once does not exclude the possibility that a user will soon switch to another provider. For this reason, I would define a competitive edge as having the ability to be "continually selected." I would like to transform the NTT business model in order to acquire this ability to be continually selected by customers.

The cloud era corresponds to a business environment in which many and varied players can participate. There is no doubt that a telecommunications carrier is simply one of them. I believe, however, that the NTT Group should become one of these players "with value," which we define as a *value partner*.

Becoming a value partner means, of course, that the NTT Group will have to compete with a variety of players, but it also means that we will have to incorporate collaboration into our business model (**Fig. 3**).

2.2 Towards the Next Stage 2.0

In November 2012, NTT announced its Medium-Term Management Strategy called "Towards the Next Stage." The idea at that time was that the NTT Group needed to transform itself beyond its traditional role of being a provider into being a value partner who could help transform the user's lifestyle and the corporate customer's business model. To this end, two main objectives were established: make global cloud services the cornerstone of NTT's business, and thoroughly enhance the competiveness of NTT's network services within Japan.

We are upgrading this Medium-Term Management Strategy to "Towards the Next Stage 2.0" with the aim of clarifying the path to recovery of profitability and growth. Specifically, we are slightly modifying our objectives in two ways: to accelerate the speed of generating profits from global cloud services and to enhance the profitability of network services.

We can also consider expansion of the B2B2X (business-to-business-to-X) business to be a pillar of this updated strategy. In fact, we are treating our Hikari Collaboration Model, which involves wholesaling of optical-fiber access services, as a manifestation of our new business model. Moreover, to solidify our

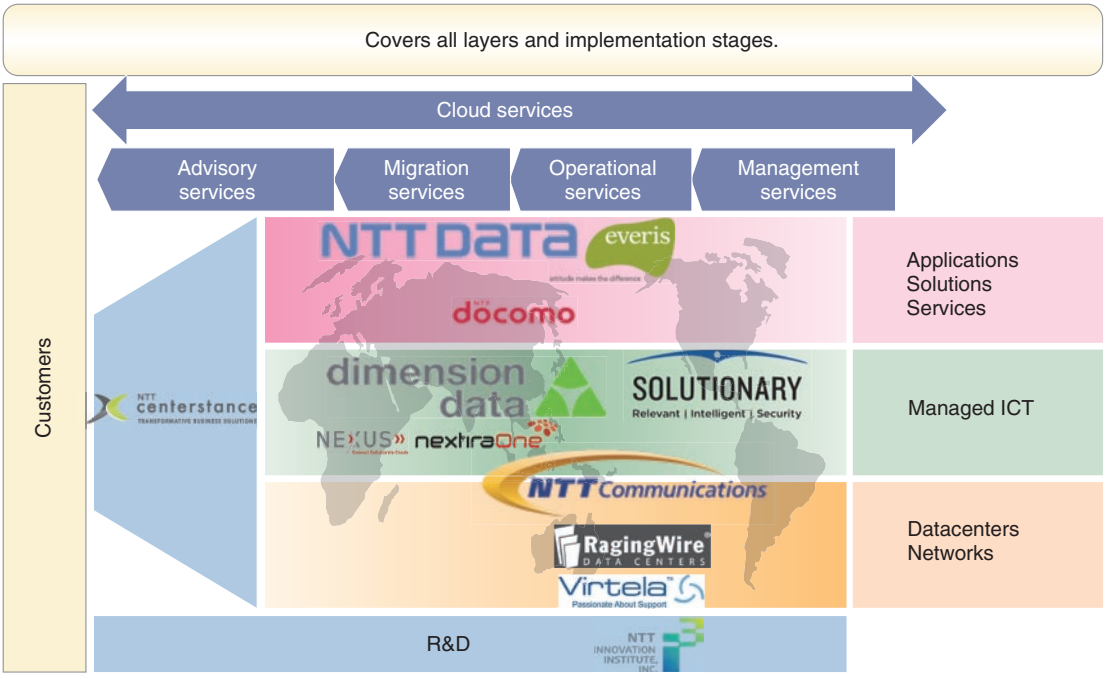


Fig. 4. Formation of NTT Group global business.

role as a value partner, I would like to see the entire NTT Group make significant contributions to themes and initiatives that are currently being undertaken by the Japanese government such as regional revitalization. The 'X' in B2B2X may be 'C' for consumer, 'B' for business, or even 'G' for "government." With this new type of business as a pillar of the NTT Group, I plan to announce a new Medium-Term Management Strategy in May 2015.

3. Global cloud business

Our current Medium-Term Management Strategy aims for global business sales of US$20 billion in fiscal year 2016, ending March 31, 2017, with more than 50% of corporate sales being overseas sales. The fiscal year 2014 goal is $15 billion, which I'm sure we will achieve. The entire NTT Group is advancing steadily toward this goal of $20 billion.

3.1 Global cloud business promotion system

The formation of NTT Group global business covering all layers and stages of cloud services has been completed (**Fig. 4**). I can say that there are no other corporate groups in the world with such a system, which has been highly evaluated as a very unique approach even by parties outside NTT.

Let me give a few examples here. As of the end of September 2014, the NTT Group operated 239 datacenters in 19 countries around the world with a total floor space of about 1 million square meters. This amount of floor space is currently ranked as No. 1 in the world. Of course, our efforts are not limited to datacenters; we are also working to build up our services and technologies. In January 2014, NTT DATA acquired the Spanish company everis, while NTT Communications acquired the U.S. companies RagingWire and Virtela. In addition, Dimension Data is expanding its footprint and service expertise through its acquisitions of NextiraOne and Nexus IS.

3.2 Promotion of cross-selling

A major strength of the NTT Group is its ability to provide one-stop cloud services and to form tie-ups with a variety of partners. Additionally, to further expand the scope of its business, we are intensifying our promotion of cross-selling within the group. Last year, we brought together for the first time NTT Group companies that had achieved results in cross-selling and held a cross-selling award ceremony. Three awards were given for cross-selling achievements in Europe, Asia, and North America. I would like to continue these efforts into the future.

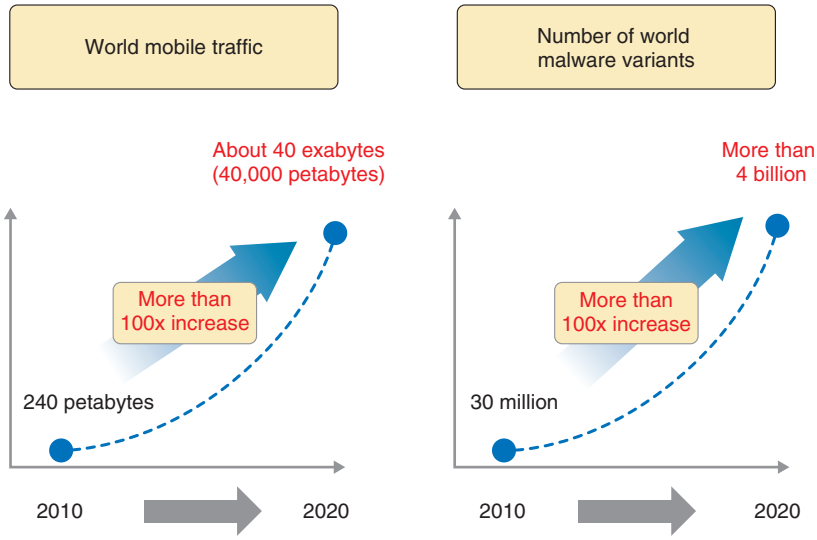


Fig. 5. Explosive increase in data traffic and growing security risks.

4. Network services

4.1 Becoming more competitive through innovation

I will now introduce our activities in network services.

Our first undertaking here is a new billing plan, which was announced by NTT DOCOMO in April 2014. In the past, a user who purchased a mobile phone would receive a discount in monthly usage fees under the condition that the phone would be used for two years. However, amid the situation in which major carriers were selling the same kinds of terminals, a cash-back incentive war arose in the industry, which began in March of last year.

Wishing to extricate itself from such an unsophisticated approach to attracting customers, NTT DOCOMO decided to provide its loyal, long-term customers with more cordial and attentive service and to migrate to a new billing scheme that would enable individual households to enjoy a rich broadband environment.

Our second undertaking in network services is the full-scale wholesaling of optical-fiber access services under the Hikari Collaboration Model of NTT EAST and NTT WEST, announced in May 2014.

This initiative is part of a theme that I have been involved with for many years. Although it began with NTT DOCOMO's offering of simple package discounts to promote the combined use of mobile and optical broadband services, our aspirations did not stop there. Today, our fundamental objective is to further expand the value of Hikari services by encouraging the participation of diverse market players and to create new B2B2C business using NTT's extensive optical fiber infrastructure.

5. Tokyo 2020 Olympic and Paralympic Games

In January 2015, NTT concluded a contract with the Tokyo Organizing Committee of the Olympic and Paralympic Games to be the first Gold Partner of the Tokyo 2020 Olympic and Paralympic Games. On entering this contract, I made the following three promises:

(1) To provide thorough and reliable operations as a provider of network systems for the Games taking into account the possibility of huge volumes of traffic and cyber-attacks
(2) To create new services that express a hospitable and welcoming spirit in collaboration with members of industry
(3) To contribute as a partner to making this great undertaking of the Tokyo 2020 Olympic and Paralympic Games a legacy to be left for future generations in Japan.

5.1 Explosive increase in data traffic and growing security risks

Worldwide mobile traffic, which stood at 240 petabytes in 2010, is expected to increase to 40 exabytes by 2020 (**Fig. 5**). Additionally, the number of malware programs, the detection of which is a specialty

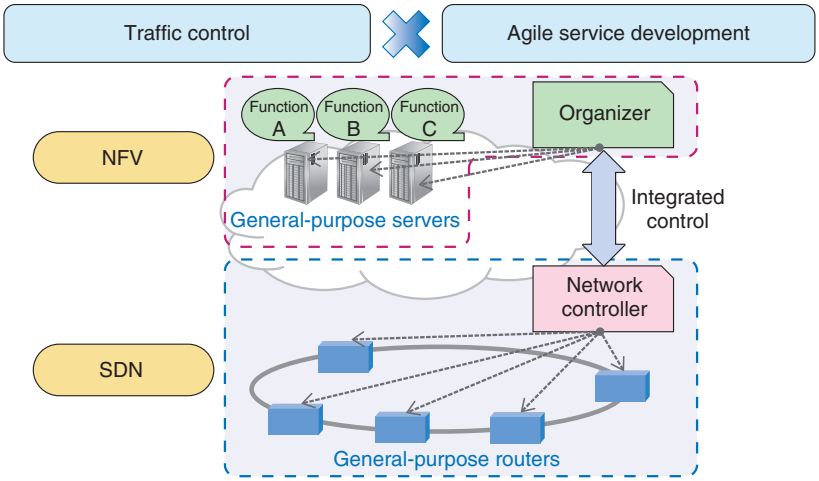


Fig. 6. Sophisticated network.

of NTT laboratories, is expected to increase by more than 100 times from 2010 to 2020. According to the National Institute of Information and Communications Technology (NICT), there were 25.6 billion cyber-attacks directed at the Japanese government and Japanese companies in 2014. We can expect the volume of cyber-attacks to intensify during the period that Tokyo hosts the 2020 Olympic and Paralympic Games.

5.2 Building a sophisticated network

Building a sophisticated network and performing security tasks will, of course, be important from a technology viewpoint, as in network functions virtualization (NFV) and software-defined networking (SDN). Yet it will also be important from the viewpoint of performing thorough and reliable network operations to establish usage rules together with our partners (**Fig. 6**). This incorporation of rules into the network usage method will be taken up in future activities.

5.3 Cybersecurity management

Our plan is to create a system that can provide a certain degree of training in cybersecurity management not by 2020, but by 2018 (**Fig. 7**). Then, we would like to implement countermeasures while performing simulations over a two-year period. The time that we have left to accomplish this is limited, so I would like to pursue this initiative in collaboration with a variety of partners.

5.4 Development of cybersecurity personnel

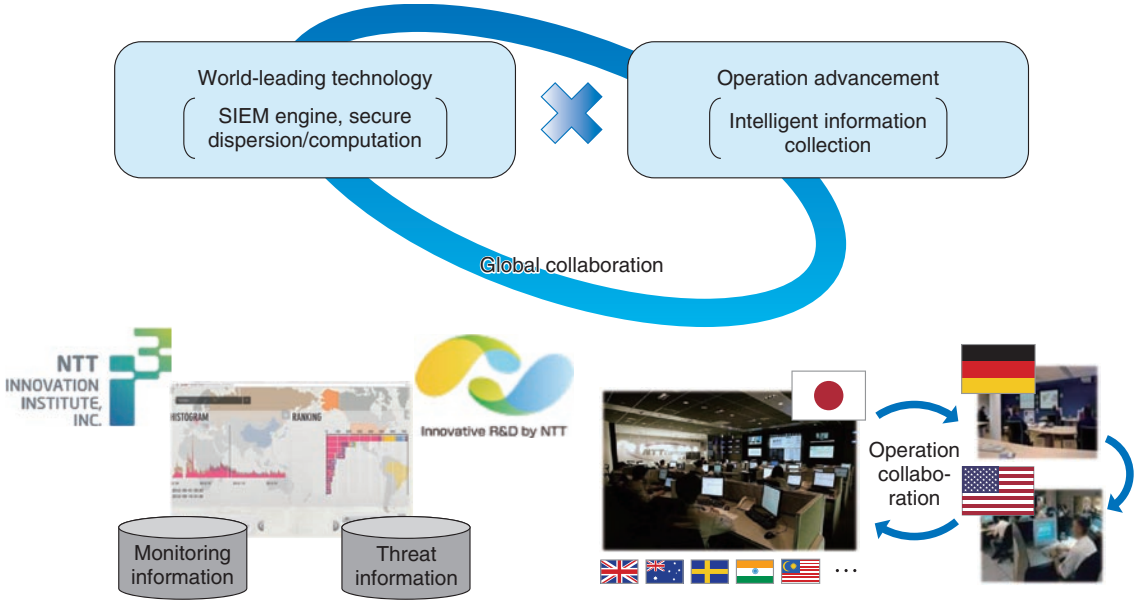
The development of cybersecurity personnel is a matter of urgency. At present, NTT Group companies have a cybersecurity staff of about 2500, from novice to advanced members. This number will be increased by four times by 2020. As for cybersecurity training on the outside, we are working on setting up endowed courses at universities and establishing collaborative efforts between government, industry, and academia. Looking forward, I believe that cybersecurity personnel should take on the role of cross-company specialists.

5.5 New endeavors

NTT announced a service called “t-Room” some years back [1]. The t-Room service presents images of location A at location B and vice versa to make it feel as if all users were gathered together in the same room. I wondered if the technology used by t-Room could somehow be applied to the 2020 Olympic and Paralympic Games. For example, if we attached a camera to a wheelchair of an athlete playing in a wheelchair basketball game and applied the t-Room concept, viewers could experience for themselves the remarkable athletic ability of these amazing athletes. I think this is a real possibility.

6. Toward regional revitalization

The NTT Group is carrying out activities toward regional revitalization in recognition of the fact that this is a major theme in Abenomics, the economic policies of the current government (**Fig. 8**). Today,



SIEM: security information and event management

Fig. 7. Integrated security.

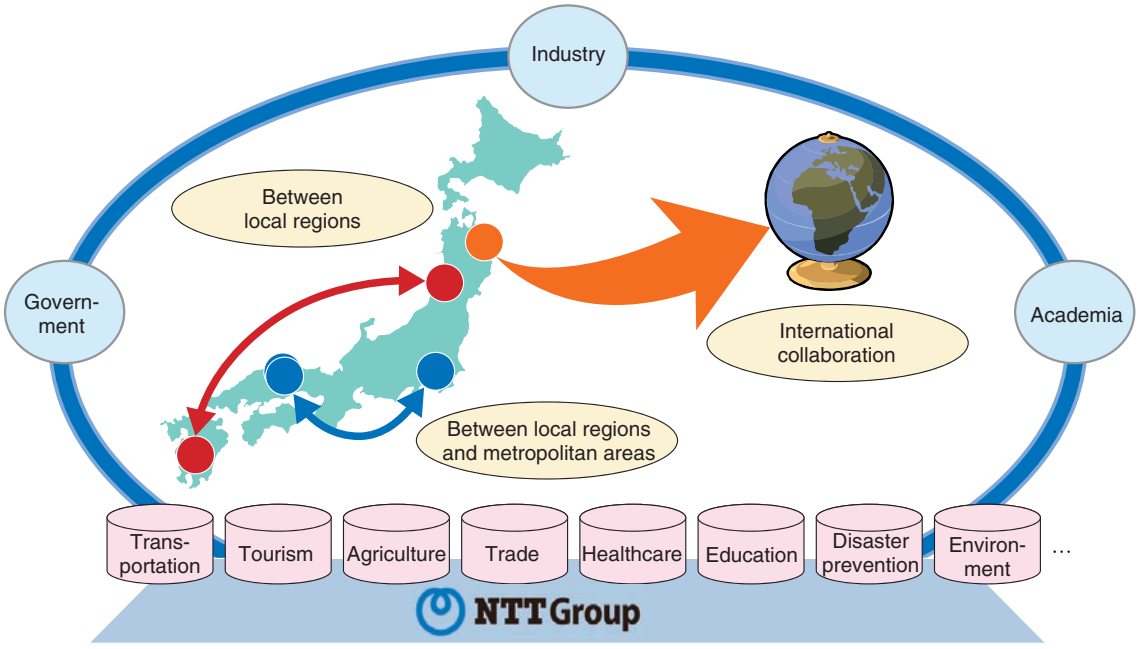
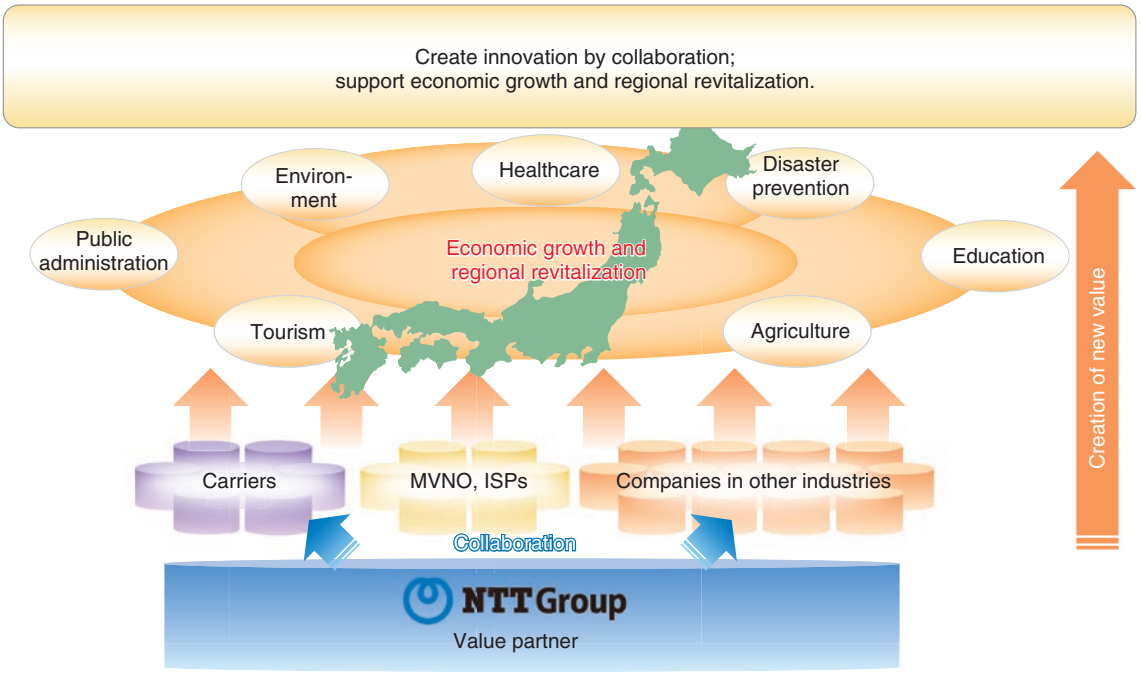


Fig. 8. Collaboration for regional revitalization.

Japan is faced with a variety of unique problems such as the aging society and the declining birthrate, and these problems are multiplied in local regions. How to resolve these problems is an issue that must be addressed to reinvigorate the Japanese economy. The main players in dealing with this issue are the



ISP: Internet service provider
MVNO: mobile virtual network operator

Fig. 9. NTT Group's approach.

specific regions themselves, but I believe that inter-regional collaboration is also an important aspect of regional revitalization. By this, I don't mean collaboration just between close regional districts; it can include, for example, collaboration between Kyushu and Hokkaido, between less populated regions and metropolitan areas, and even between local regions and international regions or parties.

Moreover, in terms of industry-related collaboration, new tie-ups should be formed with a variety of industries in addition to stimulating local industries. Needless to say, collaboration between industry, academia, and government is essential to regional revitalization. I think that having each region take an original approach is a pillar of the Japanese government's regional revitalization plan this time around. Make no doubt about it, though, the appropriate application of ICT is essential to executing this plan.

7. Future activities of the NTT Group

The NTT Group is not the main player at center stage. Rather, it aims to be a value partner in a supporting role, that is, a group that can provide ample assistance and support in economic growth and regional revitalization (**Fig. 9**). In addition to NTT EAST and NTT WEST, NTT DOCOMO, NTT DATA, and NTT Communications, the NTT Group also includes such leading companies as NTT FACILITIES and NTT Finance. I am proud of this organization that has the power to play a major role in Japan's recovery. In this regard, I am fully committed to creating a new legacy of the NTT Group for future generations.

Reference

[1] K. Hirata et al., "The t-Room—Toward the Future Phone," NTT Technical Review, Vol. 4, No. 12, pp. 26–33, Dec. 2006.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr200612026.pdf

Co-Innovation Leading to the Future

Hiromichi Shinohara
Representative Director and Senior Executive Vice President,
Chief Technology Officer and Senior Vice President of Research and Development Planning, NTT

Overview

This article introduces NTT's research and development (R&D) activities undertaken in collaboration with its partners to apply information and communication technology to deal with issues arising from changes in Japan's social environment. It is based on the keynote speech delivered by Hiromichi Shinohara, Representative Director and Senior Executive Vice President, Chief Technology Officer and Senior Vice President of Research and Development Planning of NTT, at the NTT R&D Forum 2015 held on February 19 and 20.

Keywords: R&D, collaboration, open innovation

1. Target directions of NTT's R&D

Japan is confronting a variety of social issues such as an aging population, declining birthrate, ever-increasing medical costs, and the deterioration of social infrastructures. It must also strengthen its industrial competitiveness, revitalize regional economies, and prepare for the big event in 2020. In addition, network traffic will continue to grow, and security threats will increase not only in number but also in the scope of their impact.

The NTT Group is striving to address these issues by exploiting information and communication technology (ICT), which is its forte, in collaboration with partners. The role of NTT's research and development (R&D) is to create technologies that will help to solve these wide-ranging issues. NTT has targeted five major projects that will be built on top of its basic technologies. These are big data, security, clouds, networks, and media/knowledge processing. The activities associated with these projects should not be confined to individual laboratories but rather should span all the laboratories (**Fig. 1**).

1.1 Big data: Predictive

NTT's R&D focuses on predictive big data. This is not simply a vague prediction of what will happen in the future. The aim is to predict not just what will happen, but when and where something will happen (**Fig. 2**). The key technology for this is spatio-temporal multidimensional data analysis. A feature of this technology is that it models time and space on multi-dimensional axes and uses this model to predict the time of occurrence. Since one of the challenges in handling big data is how to protect personal information, we have devised a way to analyze aggregate data without first decomposing the data. By combining the spatio-temporal multidimensional data analysis with *Jubatus*, which is a real-time machine learning engine, and *Grapon*, a kind of graph analysis technology, we are making it possible to analyze flows of people, things, and information in order to predict what will happen and when and where it will happen. For example, we plan to apply this capability to

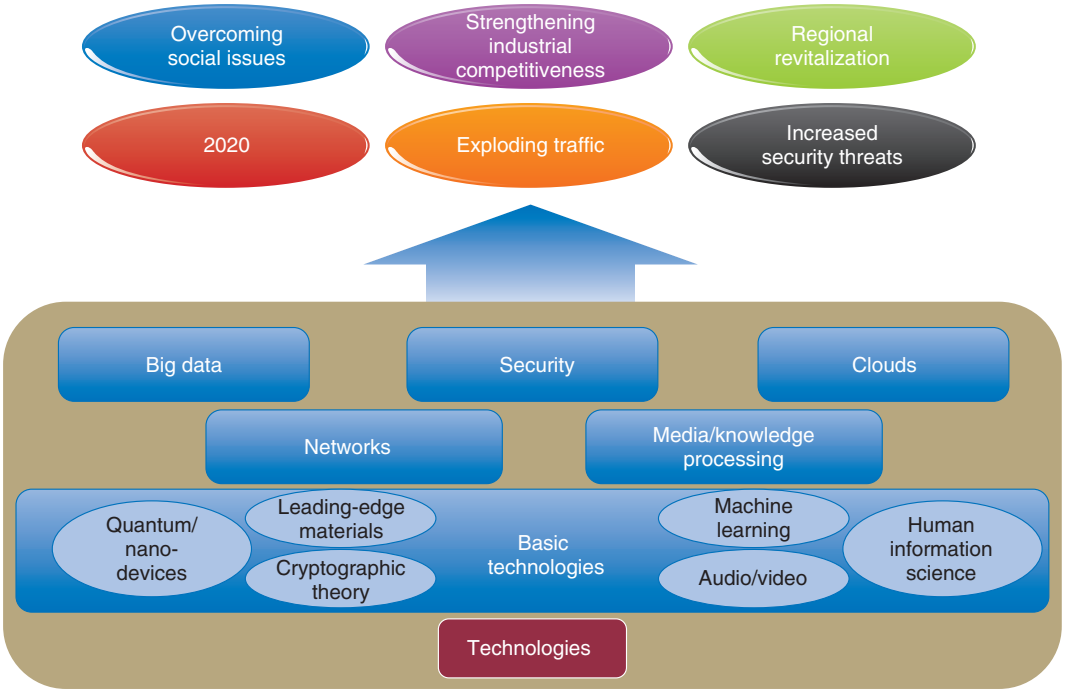


Fig. 1. Roles of R&D.

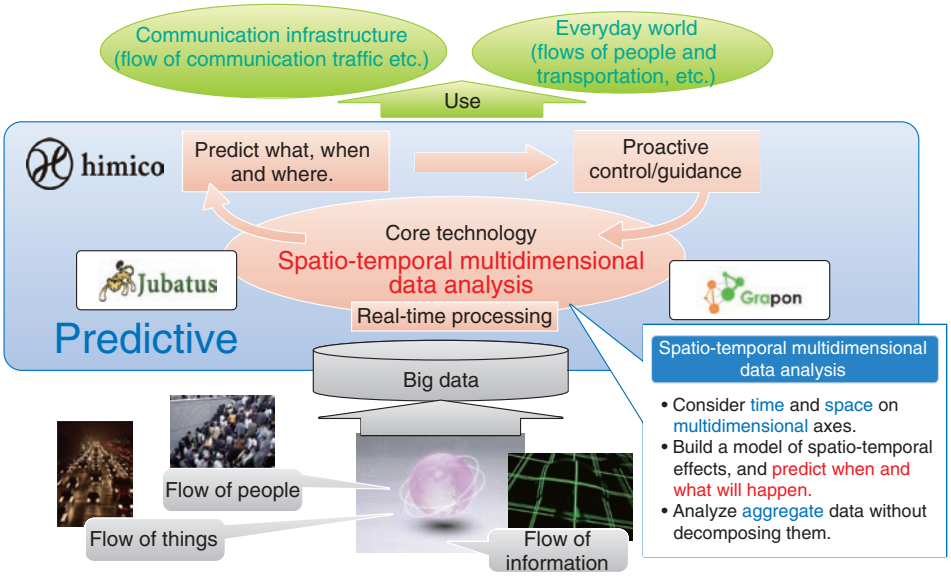


Fig. 2. Big data: Predictive.

control the flow of people or vehicular traffic in society, and to control communication traffic and predict failures in the communication infrastructure.

What motivated us to work on this issue was the Great East Japan Earthquake that struck in 2011. In a bid to escape from the tsunami, many people leapt into their cars, and most of them headed in the same direction. This caused traffic congestion and resulted

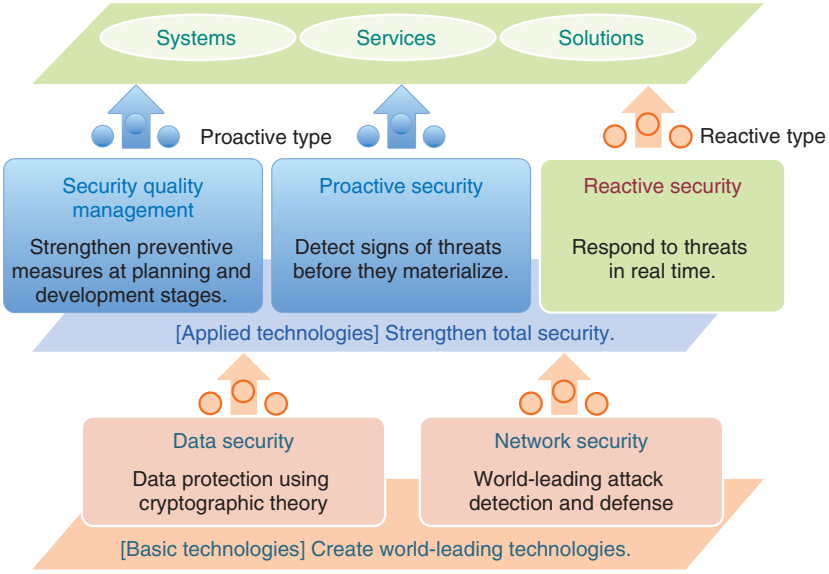


Fig. 3. Security by design.

in the deaths of many people. What struck home then was the need to guide people in a totally, rather than individually, optimized way. We will explore the application of our analysis technologies in this area.

1.2 Security: By design

We have focused on four security areas: data security with an emphasis on cryptographic theory, network security centering on malware analysis, proactive security to detect signs of threats before they materialize, and reactive security to respond to threats in real time. In addition to continuing these efforts, we will start to focus on security quality management. The intent is to incorporate preventive measures at the planning and development stages of networks and systems (**Fig. 3**).

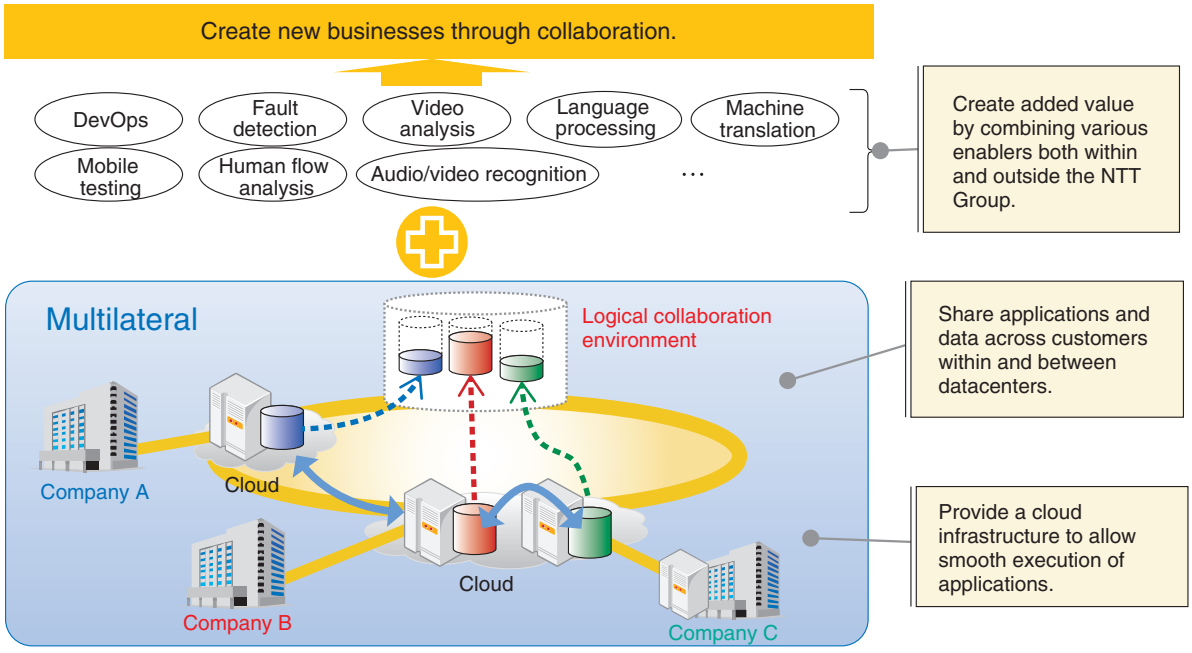
Let us consider as an example the case of applying security quality management to network design. The security task begins with identifying what to protect in the entire network. Then we analyze threats, assess risks, study countermeasures, and incorporate the results into the network and system design. Incorporating security measures at the design stage has been shown to reduce security costs (and risks) to one-hundredth that if measures were to be taken during the operation stage.

1.3 Cloud: Multilateral

The NTT Group considers global clouds one of the pillars of its growth and has been developing cloud technology that enables users to access clouds safely at an affordable price, and to migrate their tasks from existing computers on their own premises to clouds and to do so flexibly and rapidly in a way suited to their requirements. Going forward, we will develop new technology that enables users to engage in multilateral collaboration (**Fig. 4**). If enterprises are to create a wide range of marketable services and products in the years to come, it is extremely important for them to develop businesses that involve collaboration or integration of multiple industries instead of confining their initiatives to their own industry. NTT's R&D will enable a diverse range of enterprises to create new products and services by analyzing data in collaboration with them rather than these enterprises simply entrusting their data to some providers. We intend to achieve this by developing a mechanism that enables multiple customers to share applications and data within a datacenter or across datacenters of the NTT Group, and by combining the big data analysis with video development technologies owned by the NTT Group or external enablers.

1.4 Network: Dependable and elastic

It is about time that we conceive a new network. There are two reasons for this, one of which is technical. With the dramatic increase in the performance of off-the-shelf servers, it is now possible to shift from using only carrier-specific devices to using inexpensive commodity devices when building networks and



DevOps: A coined term that combines development and operations.

Fig. 4. Cloud: Multilateral.

systems. The second reason is related to the carrier's business. We need more dependable and elastic networks and systems. Having witnessed a dramatic change in the telecommunications business environment over the last decade, we consider that our networks and systems should be flexible enough to be able to adapt to a changing business environment down the road. In addition, if we are to collaborate with service providers as the *value partner* that customers continue to select, we need to respond rapidly, flexibly, and economically to their diverse range of requirements. We have named our activities to create such a network *NetroSphere*, a word combining "Netro," which means learning from the history of the Internet and telecommunications to develop the future network, and "Sphere," which is derived from "atmosphere." We plan to create a space that can encompass the needs of customers and adapt to changes in our business environment by integrating our accumulated network technologies and expertise and using commonly available parts.

NetroSphere consists of three elements: the transport, the network functions for controlling the transport, and the operations of the entire network. Its key concept is *separation* and *combination* (**Fig. 5**). Although virtualization technology has strengthened separation of hardware and software, the focus of NTT's research is on separation at a smaller component level.

In the transport element, we will clearly separate optical and electrical parts to achieve all-optical transport. By minimizing electrical processing, we will reduce the number of devices used in networks, which in turn will reduce the number of devices subject to maintenance. In addition, we will implement network functionality in software. This will dramatically enhance our ability to use off-the-shelf hardware. We will advance to this stage rapidly by collaborating with other global carriers instead of attempting to undertake this project on our own.

It is also necessary to break network functions up into applications that can be treated as components and to provide capabilities common to multiple applications as a platform. The idea is to simplify applications. This will facilitate the introduction or modification of services. In addition, by making the platform a distributed one, we can achieve high reliability without incurring additional costs.

Our plan for operations is to provide an API (application programming interface) that has the capacity to operate the aforementioned components effectively and to respond rapidly to diverse requests from a variety of players. In addition, we will develop a proactive maintenance approach by which machine

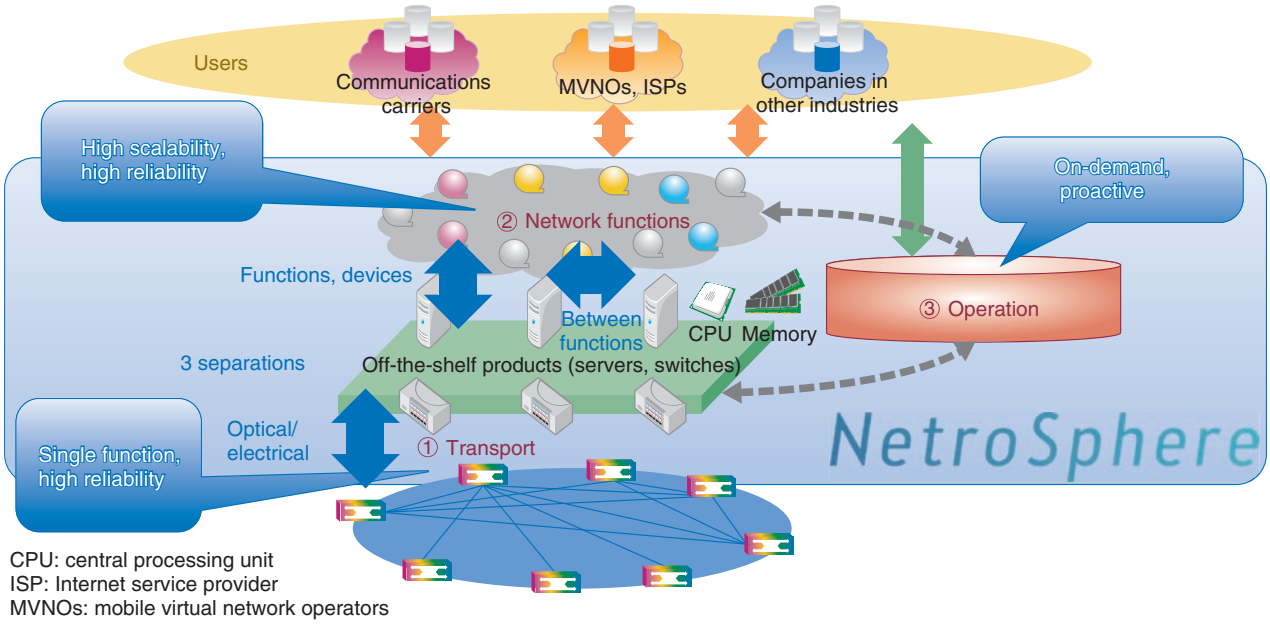


Fig. 5. NetroSphere activities.

learning is used to detect signs of faults and quality degradation before they materialize.

1.5 Media and knowledge processing: Semantic and real

Our overall approach to communication is to consider the meaning of the particular communication from the communicator's perspective, and to provide a high sensation of reality that far surpasses the experience of conventional viewing and listening. This will require video processing, speech and acoustic processing, and language processing. It is important to convey a sensation of reality in a way that matches human sensitivity instead of simply aiming to reproduce reality (**Fig. 6**). This means that we need to delve deeply into the very core of communication on the basis of communication science, taking into account how humans see and hear the things around them.

Let me introduce some of the highlights of our research outcomes in the area of communication science. The first is what we call *HenGenTo* (transforming light) (**Fig. 7**). Just by projecting a special light, you can make a still picture appear animated. Our understanding is that when someone perceives an object, the brain takes in information about the object's shape, color, and movement separately, and then combines these separate pieces of information. When a light that is related to movement based on this principle is projected onto an object, the still object looks as if it were animated. Another example is a robot that NTT used in its project called *Can a robot enter the University of Tokyo?* In this project, we gave the robot the task of sitting the English test, a challenge that completely stumps most computers. In a mock exam given in September 2014, our robot raised its deviation value from 41.0 to 50.5, exceeding the average human score in the exam for the first time.

We consider that R&D in the area of physics and its application to devices are also important for the advancement of ICT. One of our main accomplishments in this area last year was an all-optical RAM (random access memory) chip containing more than 100 bits. We will continue to pour our efforts into the pursuit of leading-edge R&D.

2. R&D activities looking toward 2020

As we look towards the big event of the 2020 Olympic and Paralympic Games, NTT must play three key R&D roles. These are to continue to provide stable and high-quality network services even as we face an explosive increase in network traffic, to take reliable network security measures, and to provide hospitality with an emotional appeal that brings with it new

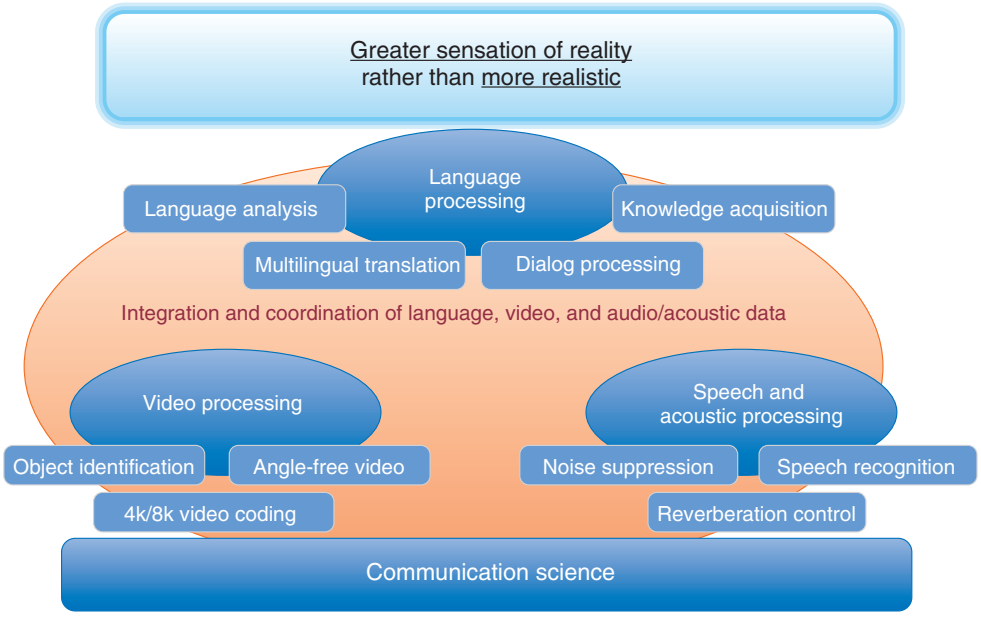


Fig. 6. Media/knowledge processing: Semantic and real.

Fig. 7. Light projection technology: *HenGenTo*.

experiences.

2.1 Resolving issues through open innovation

Whether we are talking about 2020 or not, NTT cannot resolve issues by itself. It is essential to undertake innovation with a variety of partners in the areas of both business and R&D.

For the last two years, we have been promoting open innovation, which can take two forms: working with homogeneous partners and working with heterogeneous partners (**Fig. 8**). The objective of the former is to dramatically raise performance in a specific area or to expedite development. The objective of the latter is to create new value that can be achieved only when we work with customers or partners in fields completely different from our own. We plan to increase collaboration with heterogeneous partners in our open innovation activities.

2.2 Specific activities

Some of our specific activities toward 2020 are introduced below.

(1) Providing stable and high-quality network services

Network traffic is not only increasing at a prodigious rate but also changing in ways we have never

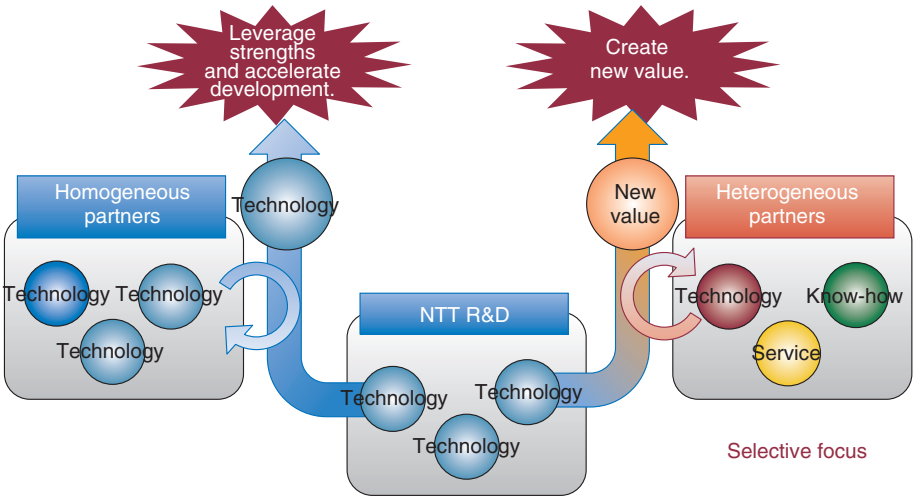


Fig. 8. Objectives of open innovation.

experienced before. Traffic can concentrate in a specific location in the network—just as foot traffic can in certain areas—or it can continue to move, as happens with the front-runners in a marathon race. NTT can contribute to managing such traffic by providing technology that makes it possible to handle high volumes of traffic efficiently with a small amount of equipment, which eliminates the need for a massive investment in lots of equipment. A case in point is proactive network control technology. Instead of observing a change in traffic linearly, this technology generates geographically distributed traffic forecasts by analyzing user behavior and traffic-generating events and controls the network accordingly. This network control makes it possible to handle high traffic volumes with minimal equipment (**Fig. 9**).

An approach from a different perspective is control of the traffic itself. This approach has been adopted in collaboration with Dwango, a video delivery service provider. The objective is to enhance customers' quality of experience by varying the bit rate when sending video signals. A high bit rate is used in a good communication environment, but it is lowered to a bit rate that is just good enough to avoid video interruption in an adverse communication environment. We conducted a small-scale evaluation experiment and found that interruptions to viewing were reduced dramatically from 33% to 2%, and the volume of transmitted data decreased by more than 10%.

(2) Network security measures

Cyber-attacks are expected to continue to increase in number. In view of the fact that all industries are connected to the Internet, this is not simply a problem in the field of communications; rather, it concerns all social infrastructures including those for finance, electric power, and transportation. As a communications carrier, NTT will of course strengthen the resilience of its networks to such attacks but will also need to intensify collaboration with other industries.

Let me introduce the concept of resilient security, which is a representative technology designed to thwart cyber-attacks. This technology automatically detects attacks based on communication patterns collected from the network and semi-automatically implements recovery actions appropriate for the specific type of attack (**Fig. 10**). This is effective for slow DoS (denial of service) and reflection DDoS (distributed DoS) attacks that have become problematic recently. We are refining the technologies required to counter these threats by monitoring and analyzing the latest attack patterns.

We have also taken the initiative in cultivating cybersecurity experts. Considering the wide range of cybersecurity-related tasks undertaken by NTT Group companies, we have classified human resources into three categories: those capable of undertaking cybersecurity management and providing consultation, those capable of undertaking cybersecurity operations, and those capable of developing systems with due attention to cybersecurity. Each category is divided into three levels of proficiency.

We also support cybersecurity education at universities. As the first initiative, we have provided funding

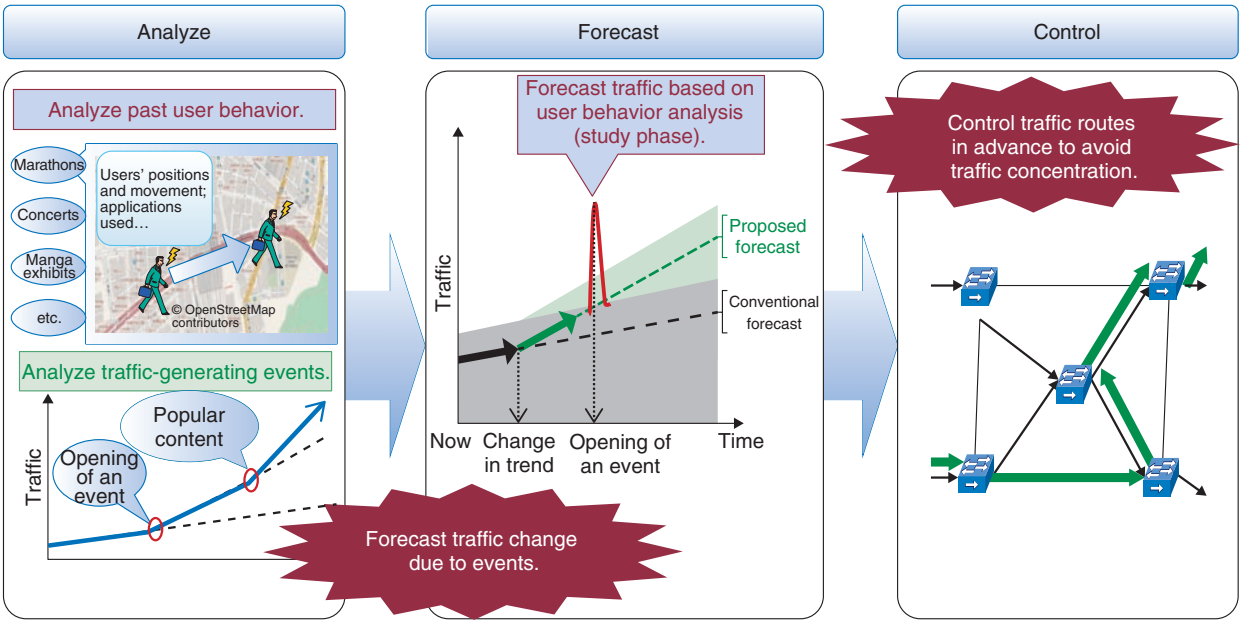
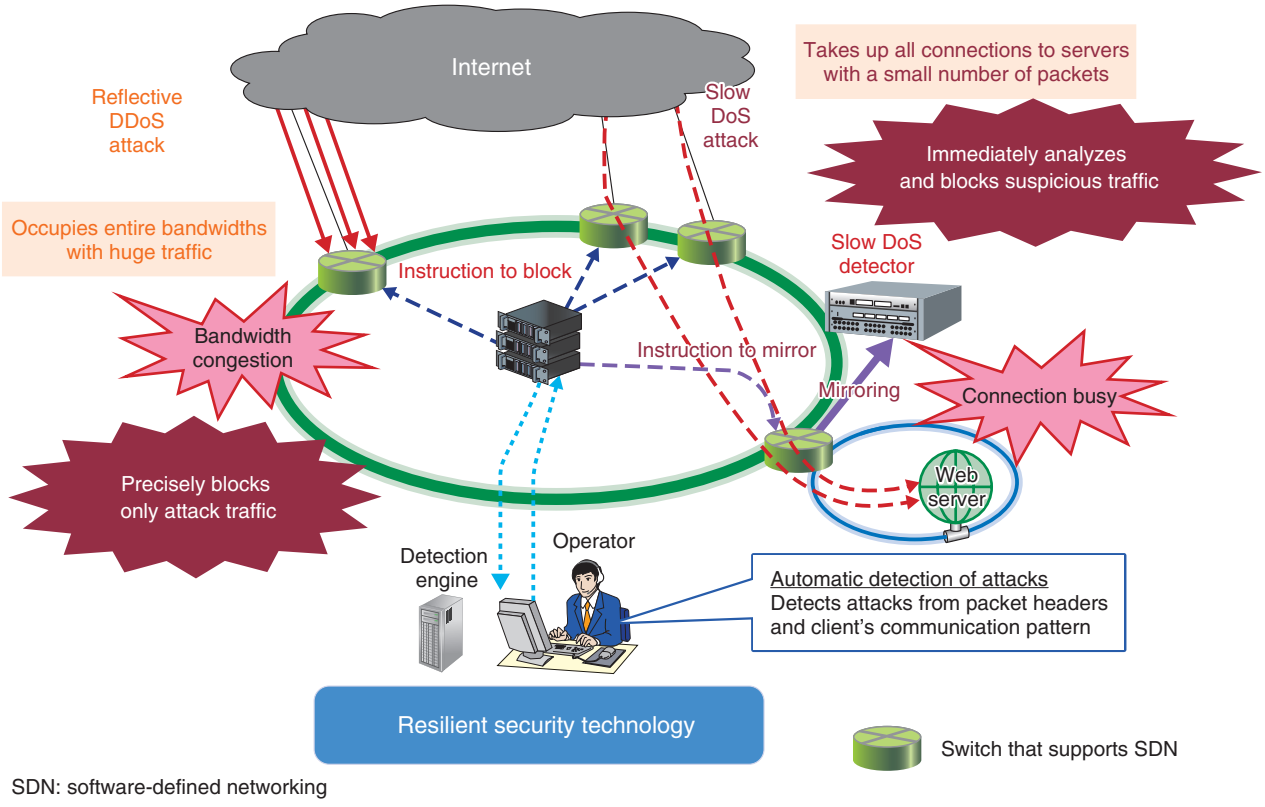


Fig. 9. Proactive network control technology.



SDN: software-defined networking

Fig. 10. Resilient security technology.

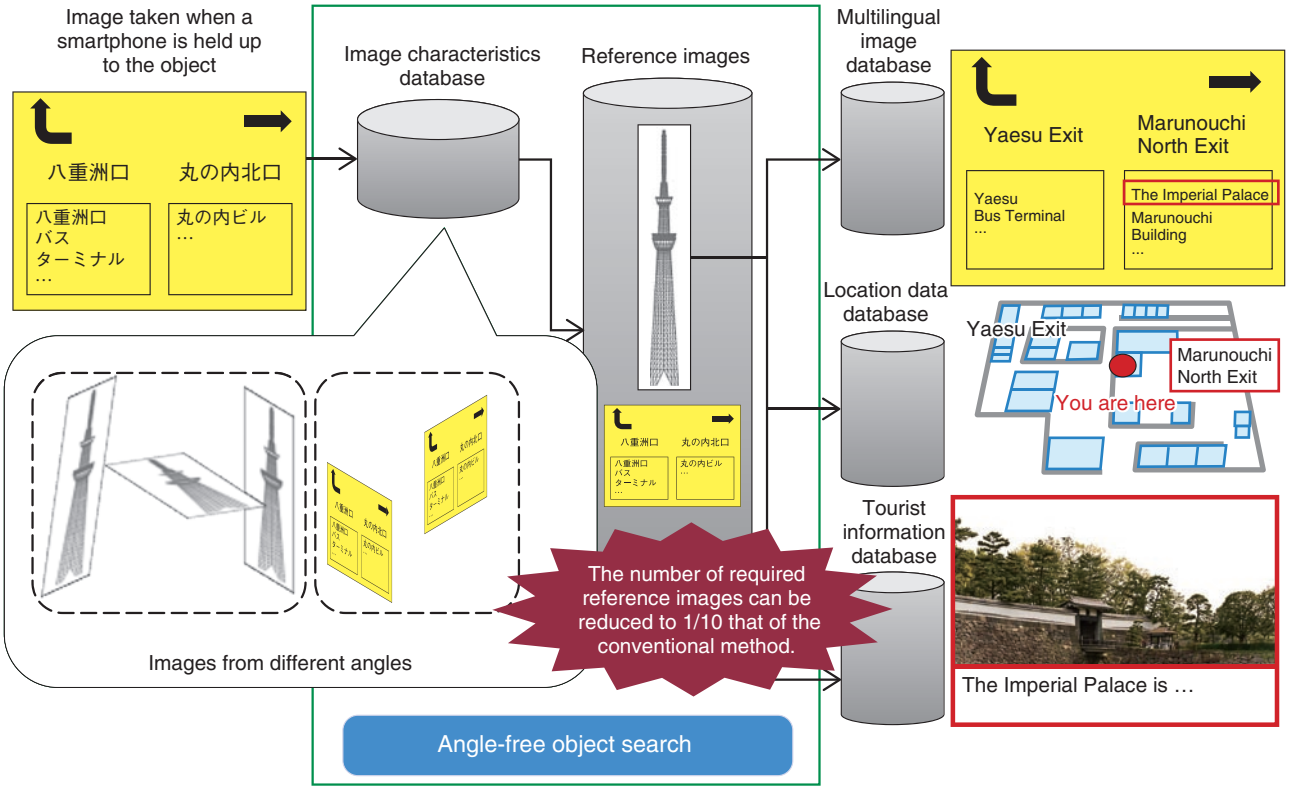


Fig. 11. Angle-free object detection.

for a class designed for graduate and undergraduate students at Waseda University. NTT's malware analysts will give lectures on basic technologies and oversee practical exercises. An online course, *gacco*, which is operated by NTT DOCOMO, will offer the *Ultra-Basic Course on Information Security* starting in May 2015.

We intend to undertake these kinds of initiatives in different industries. We will work with other industries to provide multifaceted education that incorporates knowledge specific to each industry.

(3) Providing new experiences and creating a welcoming and hospitable atmosphere

We will work on creating a hospitable atmosphere by providing new and meaningful experiences that reflect a welcoming attitude to not only spectators and foreign tourists but also all people throughout Japan and around the world. In line with the strength of its technical development capabilities, the R&D arm of NTT will collaborate with a wide range of partners to convert technologies into business opportunities.

Let me introduce angle-free object search as an example of research in the field of video search, (**Fig. 11**). This technology is extremely robust against the angle from which a 3D (three-dimensional) object is photographed and can associate a captured image with reference images even if the number of relevant reference images is small. A potential application of this technology is searching for information displayed on signs. A reference image is linked in advance to a multilingual image (English, Chinese, etc.), information about the location of the sign, and information about the content of the sign. When a visitor from overseas directs the camera of his/her smartphone toward a sign at a station, the information on the sign is displayed in multiple languages. This is not a translation. Rather, the sign is linked to information that was prepared in advance. This arrangement makes it possible to provide information that is taken for granted by Japanese people but is invaluable for foreign visitors. We can create a truly hospitable atmosphere for guests from abroad by providing such easily overlooked information. This technology is undergoing a feasibility test near Tokyo Station. In particular, because it is very easy to get lost in

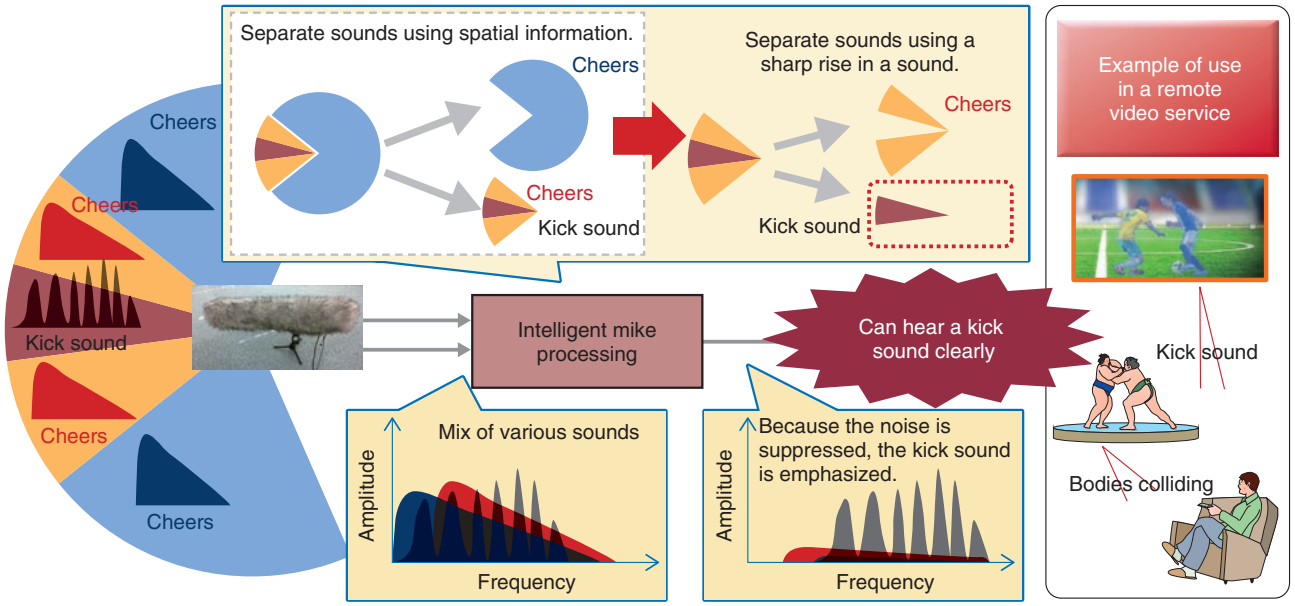


Fig. 12. Target microphone.

underground areas, it will be useful if visitors can ascertain their current location just by directing the camera of their smartphones toward a sign or a pattern on the ceiling.

In the field of high-definition video delivery, NTT collaborated with NHK to hold 8K super-high-definition public viewings of the 2014 FIFA World Cup in Brazil. This project used MPEG (Moving Picture Experts Group) media transport (MMT), which can deliver a number of HEVC (high-efficiency video coding) videos that are fully synchronized with each other. This enables you to view two fully synchronized videos. For example, while you are watching a soccer game on a public screen, you can use your tablet as a second screen and view the game from a different angle.

In the field of acoustics, we developed noise cancelation technology, which can extract a specific voice even in extremely noisy environments where the noise level exceeds 100 dB. This technology uses two general microphones. It extracts a specific voice based on differences in the voice level between the sounds from the two microphones. We have confirmed that the technology can reduce the noise level by a factor of 10,000. Even under an elevated railroad track, one of the noisiest places conceivable, speech recognition experiments using this technology recorded a success rate in excess of 90%. We have been collaborating since last year with Mitsubishi Heavy Industries to test the effectiveness of this technology in a factory. We have confirmed that the speech recognition works well, and conversations can be conducted normally even in this noisy environment.

Another achievement that evolved from noise cancellation technology is target microphone technology, which extracts a target sound clearly using a sharp rise in the sound as a clue in addition to determining the spatial noise distribution for each direction (**Fig. 12**). This technology will make it possible, for example, to extract the sound of a soccer player kicking the ball or the sound of bodies colliding in a sumo bout, thereby providing sound with a high sensation of reality to viewers.

Technology for interactive viewing of omnidirectional audio and video has been developed in collaboration with Dwango. It combines the above acoustic technologies with a 360-degree camera to enable users to experience the sensation of being on the field with the players. As users move their viewing direction, they can hear a stereo sound that is in sync with their movement.

Immersive telepresence is another challenging research project (**Fig. 13**). Information that represents the space and environment of a particular location is sent to a remote location. This information consists of object information, layer information (which represents the depth of an object), illumination information,

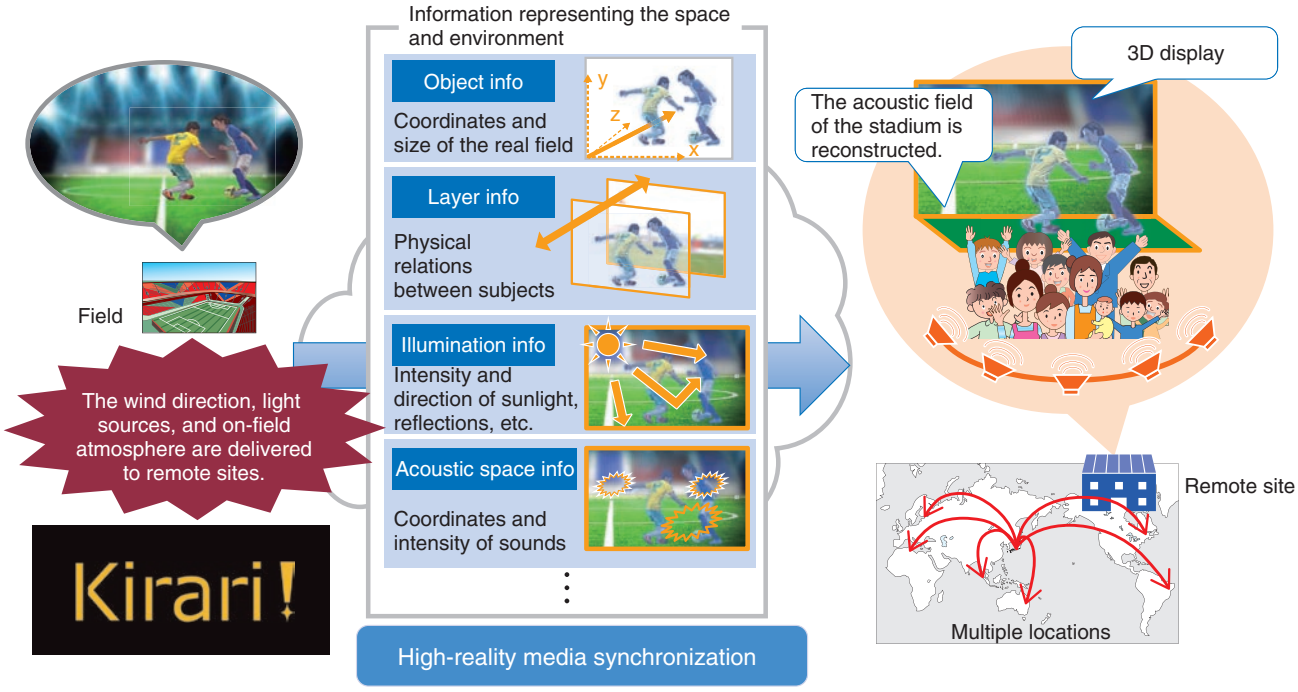


Fig. 13. Immersive telepresence.

and acoustic space information. The space and environment of the original location is reconstructed at the remote location from these pieces of information, enabling users there to feel as if they were at the actual location. We have named this project "Kirari!" and we plan to implement it for public viewing.

In the field of materials, a number of partners have collaborated to create a B2B2X (business-to-business-to-X) service. NTT and Toray worked together to develop a sensing fabric called *hitoe*. Goldwin produced a shirt that incorporates hitoe. Panasonic developed a receiver/transmitter that picks up bioelectric signals from hitoe and transmits them. NTT DOCOMO collaborated with Runtastic to produce an app that uses these bioelectric signals. While hitoe is now used to measure the user's heart rate and check the workload of the current sports activity, it can also be used to measure other signals such as myoelectric and electrocardiographic signals. We consider that by teaming up with other partners, we can expand the applications of hitoe to a wide range of fields such as athlete training and health management. In these areas, hitoe can be used to check muscle condition and even mental states and eventually provide more generalized medical support.

In the field of human information science, we aim to apply bioinformation measurement technology in sports training. This technology measures myoelectric states and represents the time pattern of an exercise by means of synthetic sounds. When it is applied to the training of a baseball pitcher, for example, it is possible to perceive intuitively through sound whether the pitcher bends his/her arm sufficiently or is overly cautious in throwing the ball. This *audibilization* is expected to motivate and raise the awareness of the player.

3. Areas commanding increased attention

In the age of the telephone, the vast majority of people were able to use their phones without problems. With the advent of the Internet age, it became apparent that there were many people who did not have the slightest idea about how to utilize it. The age of the smartphone is now upon us, and there are people who find the latest communication tools too difficult to use or who express anxiety because they need to disclose personal information in various situations when using social networking services. We feel that the gulf between people who can use these forms of communication and those who cannot is gradually widening and that we need to take people to

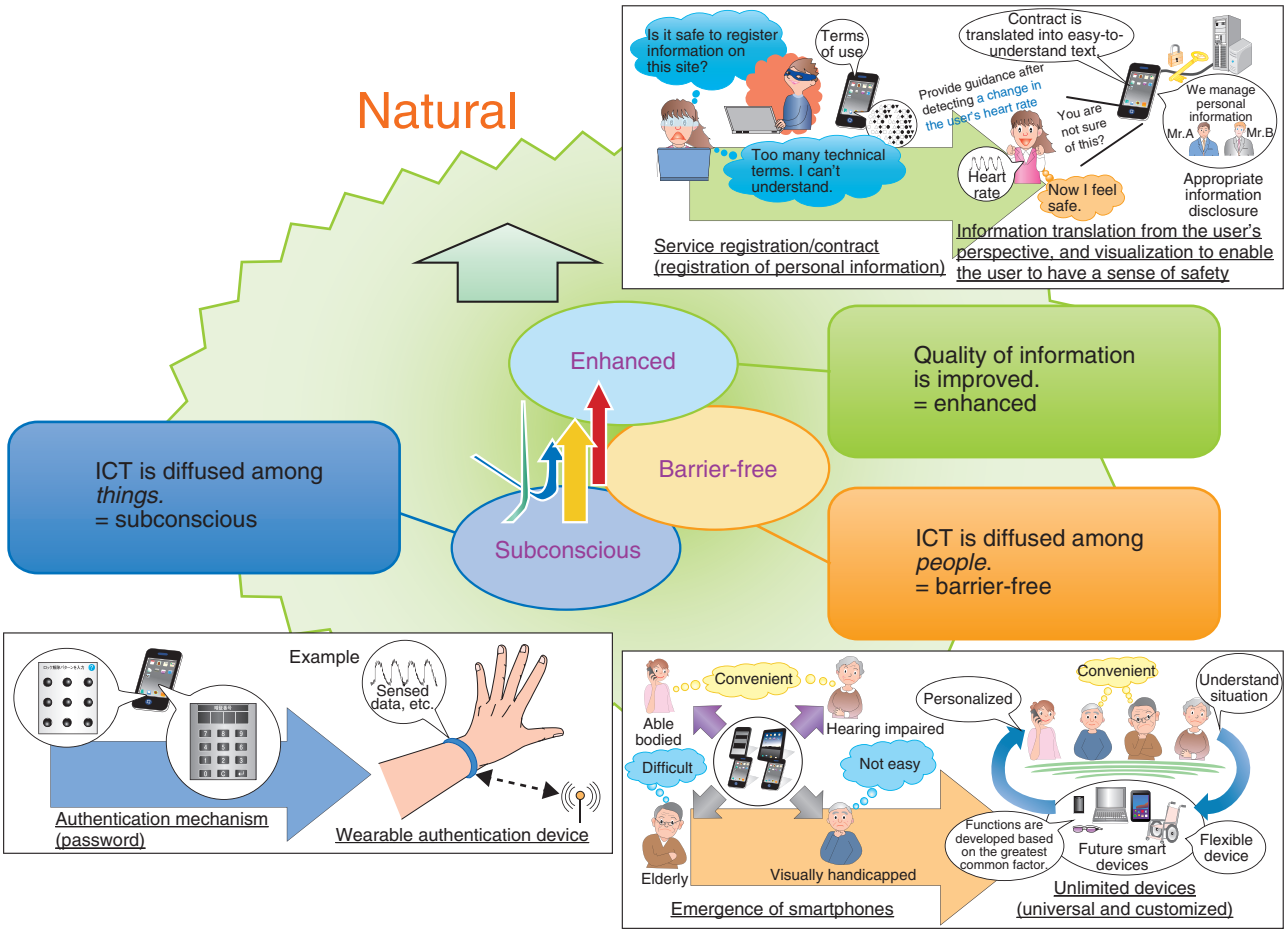


Fig. 14. Approach to a *natural* style.

a safe and secure world in which anyone can enjoy new experiences naturally. We believe that we need to strengthen our pursuit of a *natural* style to resolve the anxiety and the growing gap referred to above. We will adopt three approaches for this purpose: *subconscious*, in which ICT is diffused among things, *barrier-free*, in which ICT is diffused among people, and *enhanced*, in which the quality of information is improved (**Fig. 14**).

(1) Subconscious

Password-based authentication creates a tradeoff between complexity and security. We consider that a potential way of relieving the user of this complexity is a wearable authentication device.

(2) Barrier-free

Although current smartphones are very convenient high-functionality terminals for those who are highly ICT literate, they need to be improved before they can be used universally by anyone. We are aiming to develop a barrier-free terminal that makes it possible for anyone, regardless of personal limitations or disabilities, to use a smartphone effortlessly by taking their individual characteristics into account and drawing upon their inherent abilities.

(3) Enhanced

Suppose you want to register or conclude a contract on a website. When you are requested to input your personal information, you may feel it is unsafe to do so and decide not to register. To solve this problem, it is necessary to enhance the quality of information so that ICT services come closer to and stand on the side of the user.

4. Looking ahead

Communications carriers such as NTT originally worked, for a prolonged period, on developing telecommunications that connected people with people.

Later, they added information processing that connects people with information. Currently, they have gone beyond these two goals and seek to attain ○○×ICT, a concept that involves enhancing the capabilities of various industries through ICT. In the world of ○○×ICT, it is important to create an environment in which anyone can access various services, not just those who want to. From this perspective, we will further strengthen our activities to pursue a natural style.

Creating an Appealing User Experience by Applying Media System Technology

Satoshi Takahashi, Yushi Aono, Shiro Ozawa, Hidenori Okuda, Atsushi Sagata, and Ryuichi Tanida

Abstract

This article introduces NTT efforts aimed at creating appealing user experiences by applying the wide variety of media system technology under development at NTT, including technologies related to speech, language, audio, still images, and video. The work described here focuses on two areas: a personal agent that is intimately close to the user and provides personalized services that stimulate human knowledge and behavior, and high-sense-of-presence media services that enable the user to enjoy extremely natural viewing experiences.

Keywords: personal agent, high sense of presence, media system processing technology

1. Creation of personalized services that stimulate human knowledge and behavior

1.1 Virtual agents

In recent years, virtual agents that can satisfy the various needs of individual users have been attracting attention. In particular, virtual agents that respond to voice input and can retrieve weather reports and answer simple questions have been implemented in smartphones and other personal devices, so such agents are now more readily available. However, these kinds of services are positioned as an input option for web search functions, and they simply present the results of searches performed using the given keywords. This form of use involves a single question and a single response, and is therefore limited to executing just one function of the original roles of the virtual agent. Currently, such virtual agents are mostly confined to virtual worlds such as the web, and the information they can retrieve is limited to that domain. They cannot use information from the user's real-world situation to interact with and influence the user. We consider this to be a major obstacle to be overcome in developing future virtual agents.

1.2 NTT's concept of a virtual agent

In view of the situation described above, NTT has shed the idea of a virtual agent that is confined to a virtual world and interacts with users in the form of a single response to a single question in favor of a personal agent that understands the user intimately and exists together with the user in the real world (**Fig. 1**). NTT is now moving forward with research and development (R&D) to realize such an agent. The NTT concept of a personal agent involves three important elements. One is that the agent can understand the user's situation and intentions in the context of the real world through technology that senses and processes various kinds of media. Another is that it can actively influence the user based on its understanding of the user's situation and intentions. The third element is that it grows together with the user by understanding the user's situation and influencing the user accordingly. We believe that implementing these elements requires technology for understanding real-world situations and organizing and structuring that information (real-world structuring technology), and

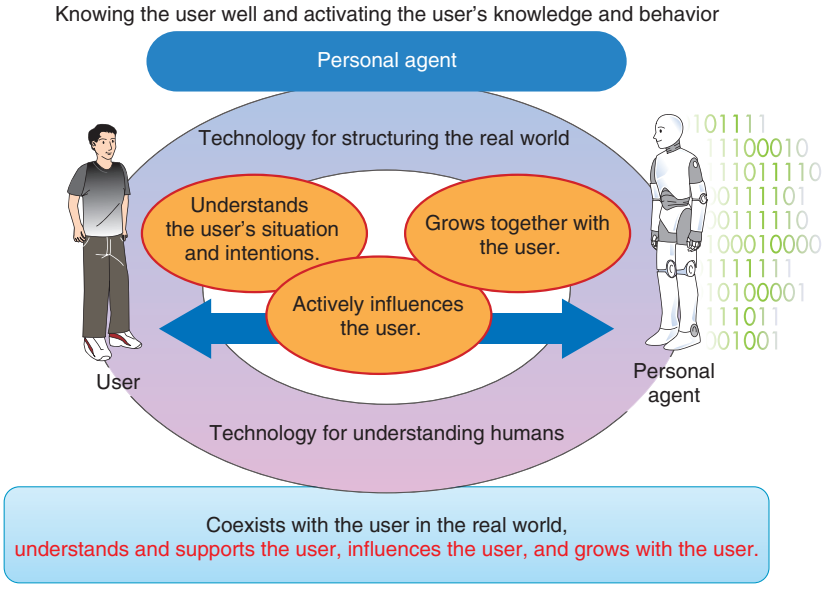


Fig. 1. Important elements of a personal agent.

technology for understanding both the explicit and latent intentions of the user (technology for understanding humans). We are developing various types of media system technology to support the required technology.

1.3 Evolution of personal agents

NTT has set two milestones for developing the personal agent and is moving forward with a policy for achieving the ultimate goal of this technology (**Fig. 2**). The first step is to create a user profile by collecting information on the user's interests and preferences from dialogs with the user and to function as a kind of servant or butler by providing appropriate support based on the profile. The second step is to sense the user's present situation, the ambient mood, and the user's expressions and state, and adaptively support the user accordingly, like a friend. The ultimate goal is to go beyond simply presenting a short-term optimum solution. Rather, the personal agent will anticipate future situations and influence the user with care and understanding on that basis, like a family member.

For example, consider a user who is trying to lose weight. Rather than recommending a nearby fast food restaurant, the agent would respond with proper concern for the user's goals and recommend a restaurant that has a health-oriented menu, even if it involves a circuitous route. We believe the personal agent towards which NTT is working will be capable of providing a new and appealing user experience unlike any virtual agent that now exists.

These Feature Articles describe technology that supports the NTT concept of a personal agent and presents specific examples. The article, "Media Processing Technology for Achieving Hospitality while on the Go," describes a service that guides the user around a city, statistical machine translation technology for presenting guide information, and robust media search technology for recognizing objects in an image [1]. "Media Processing Technology for Achieving Hospitality in Information Search" describes a service for assisting users in their daily activities and subject identification technology for searching the Internet for information related to an image captured by a camera, natural language processing technology for understanding the user's intention and responding in a natural way, and user-designed speech synthesis technology for generating synthesized speech for various speakers and speaking styles [2]. "Media Processing Technology for Business Task Support" introduces technology that holds promise for applications extending beyond these service scenarios to business scenarios [3].

2. Appealing sense-of-presence media services

2.1 High-definition video

In Japan, digital broadcasting via communication satellite began in 1996. That was followed by digital

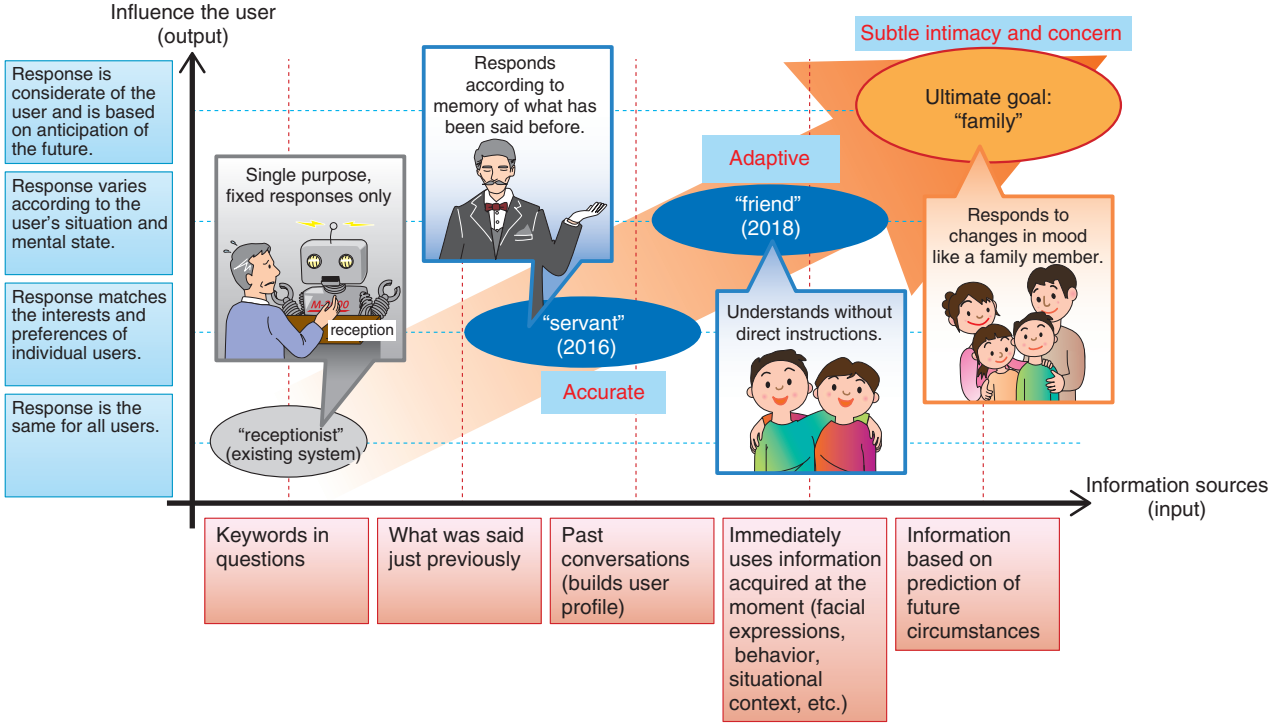


Fig. 2. Evolution of the personal agent.

broadcasting via broadcast satellite in 2000 and by terrestrial digital broadcasting in 2003. The current high-definition television (HDTV) video format[*1] provides a remarkable improvement in image quality compared with analog broadcasting and is now used for almost all programs that are broadcast.

The next generation of high-definition video media is said to be the 4K and 8K formats,[*2] which provide an overwhelmingly superior feeling of detail and representation of reality compared to HDTV and can be used to provide services that create a high sense of presence. The 4K video format was first introduced in movie theaters in 2007, and since then, the number of screens has been increasing. As a result, various types of practical 4K equipment have been developed, including projectors and cameras for professional use. Furthermore, the Next Generation Television and Broadcasting Promotion Forum (NexTV-F) began conducting test broadcasts in the 4K format in June 2014. Consumer-use 4K-resolution displays are also appearing on the market, and home use of 4K TV is becoming more popular as well.

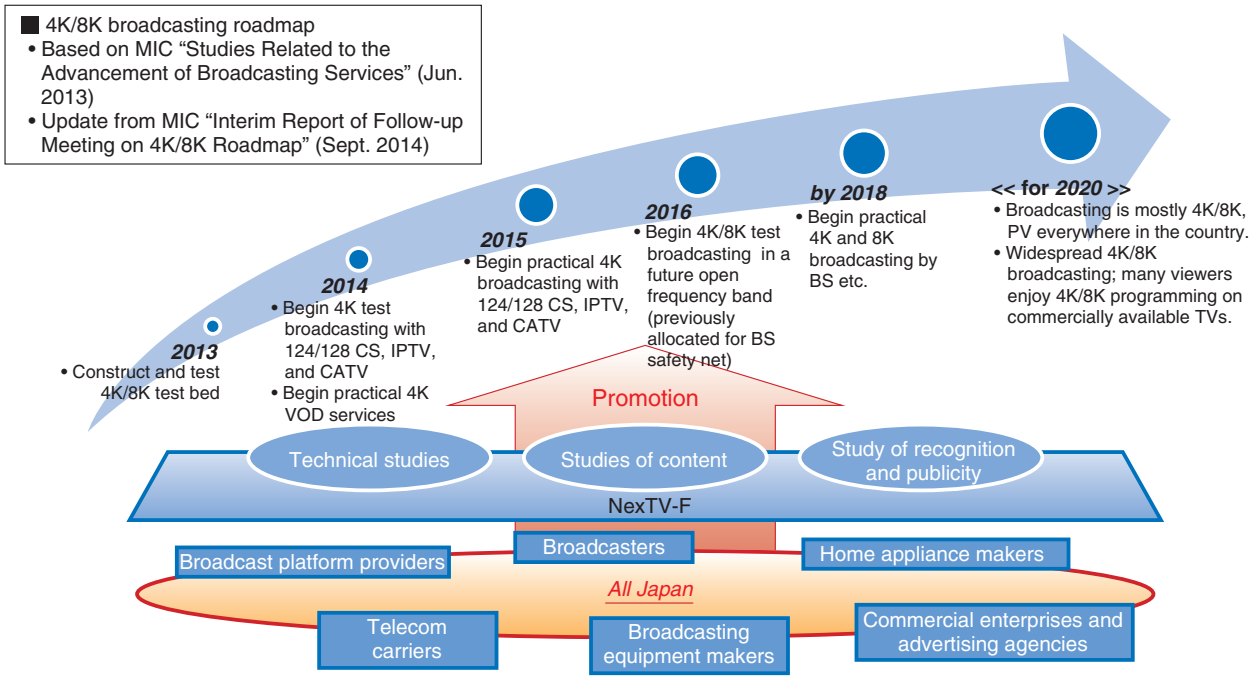
2.2 Trends in Japan and NTT related to high-definition video

Countries around the world are putting more effort into achieving 4K and 8K broadcasting, and a world-leading roadmap for commercialization of these formats has been formulated by a study group of Japan's Ministry of Internal Affairs and Communications. Furthermore, an interim report from a follow-up meeting recommended acceleration of the roadmap to promote 4K and 8K broadcasting (**Fig. 3**) [4].

For the implementation of 4K and 8K telecom and broadcasting services, broadcasters, telecom carriers, and equipment manufacturers have established the NexTV-F as an organization for cooperation. NTT is a proponent of the organization and is therefore collaborating with various enterprises to push forward with the implementation of the world's most advanced 4K and 8K services.

*1 HDTV: 1920 × 1080 pixels, also referred to as 'Hi-Vision' in Japan.

*2 4K/8K: A high-definition video format that has twice the horizontal resolution and four times the vertical resolution of HDTV. 4K and 8K together are also referred to as ultra-high definition (UHD).



BS: broadcasting satellite
CATV: cable television
CS: communication satellite
IPTV: Internet protocol television
MIC: Ministry of Internal Affairs and Communications
PV: public viewing
VOD: video on demand

Fig. 3. 4K/8K broadcasting roadmap for Japan.

2.3 Toward implementing high-sense-of-presence media services

The NTT vision for future high-sense-of-presence media services is to realize rich life environments by providing user experiences that combine high-definition video, high-definition audio, and high sense of presence.

The subjects for ongoing R&D for telecom and broadcasting services that use high-definition media include HEVC (High Efficiency Video Coding) encoding technology, MMT (MPEG Media Transport) transmission technology, FireFort®-LDGM FEC (Low-Density Generator Matrix Forward Error Correction) technology, and other elemental technologies that are essential to service implementation [5–7].

However, implementing high-sense-of-presence media services that go beyond high definition requires more than simply improving resolution, compression quality, and transmission quality. What is needed is innovative R&D that can produce technology for reproducing the sensation of being in a certain place or the feeling of being able to understand even more than one could understand by being in that place.

The article "Audio-visual Technology for Enhancing Sense of Presence in Watching Sports Events" describes five areas of innovative technology that NTT is working on to implement high-sense-of-presence media services [8]:

- Interactive distribution technology for omni-directional video that enables users to view any region within a 360° video image as they choose
- Lossless audio encoding technology for compressing high-resolution, high-quality audio
- Distribution and encoding for arbitrary point-of-view video for composing video from points of view where cameras cannot be placed, such as the line of sight of the goalkeeper or the ball
- A zoom microphone system for extracting voice signals from a remote source to give the user a sensation of being on the playing field

- Reverberation removal and control technology for separating a music signal into direct sound and sound that arises indirectly from reflection from walls, etc., to reproduce sound that creates a sense of presence

3. Future development

We believe that the media system technology NTT is developing will spread throughout the world and bring about new, appealing user experiences via a variety of services. For the personal agent, we will continue investigating forms of service and technology that make it possible to respond to user needs, and we will develop services in collaboration with partners in various fields. For the high-sense-of-presence media services, too, we will continue to push for the implementation of 4K/8K services and promote innovative media technology for the evolution from *high definition to high sense of presence* in cooperation with broadcasters, video distributors, and other partners.

References

[1] M. Horii, K. Arai, M. Nagata, K. Kashino, K. Hiramatsu, A. Fukayama, and H. Yamaguchi, "Media Processing Technology for Achieving Hospitality while on the Go," NTT Technical Review, Vol. 13, No. 4, 2015.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201504fa4.html

[2] K. Sadamitsu, J. Shimamura, G. Irie, S. Tarashima, T. Yoshida, R. Higashinaka, H. Nishikawa, N. Miyazaki, Y. Ijima, and Y. Nakamura, "Media Processing Technology for Achieving Hospitality in Information Search," NTT Technical Review, Vol. 13, No. 4, 2015.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201504fa5.html

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

[4] The Ministry of Internal Affairs and Communications, "Announcement of Interim Report of Follow-up Meeting on 4K/8K Roadmap," press release issued on September 2014 (in Japanese).
http://www.soumu.go.jp/menu_news/s-news/01ryutsu11_02000039.html

[5] T. Onishi, T. Sano, K. Yokohashi, J. Su, M. Ikeda, A. Sagata, H. Iwasaki, and A. Shimizu, "HEVC Hardware Encoder Technology," NTT Technical Review, Vol. 12 No. 5, 2014.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201405fa3.html

[6] K. Iso, M. Kitahara, R. Tanida, T. Mitasaki, N. Ono, and A. Shimizu, "World's Highest-performance HEVC Software Coding Engine," NTT Technical Review, Vol. 12, No. 5, 2014.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201405fa4.html

[7] T. Nakachi, T. Yamaguchi, Y. Tonomura and T. Fujii: "Next-generation Media Transport MMT for 4K/8K Video Transmission," NTT Technical Review, Vol. 12, No. 5, 2014.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201405fa6.html

[8] D. Mikami, Y. Kunita, Y. Kamamoto, S. Shimizu, K. Niwa, and K. Kinoshita, "Audio-visual Technology for Enhancing Sense of Presence in Watching Sports Events," NTT Technical Review, Vol. 13, No. 4, 2015.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201504fa7.html

Satoshi Takahashi
Executive Manager, Executive Research Engineer, Supervisor, Audio, Speech and Language Media Project, NTT Media Intelligence Laboratories.
He received the B.E., M.E., and Ph.D. in information science from Waseda University, Tokyo, in 1987, 1989, and 2002, respectively. Since joining NTT in 1989, he has been engaged in research on speech recognition, spoken dialog systems, and pattern recognition. He is a member of the Acoustical Society of Japan and the Institute of Electronics, Information and Communication Engineers (IEICE).

Yushi Aono
Senior Research Engineer, Supervisor, Producer, Media Component & System Produce, Promotion Project 1, NTT Media Intelligence Laboratories.
He received the B.E. in biotechnology and the M.E. and Ph.D. in control engineering from Osaka University in 1994, 1996, and 1999, respectively. Since joining NTT in 1999, he has been engaged in research on speech synthesis, speech recognition, and spoken dialog systems. He is a member of IEICE and the Information Processing Society of Japan.

Shiro Ozawa
Senior Research Engineer, Producer, Media Component & System Produce, Promotion Project 1, NTT Media Intelligence Laboratories.
He received the B.E. and M.E. from Tokyo University of Mercantile Marine (now, Tokyo University of Marine Science and Technology) in 1997 and 1999. After joining NTT as a researcher, he worked for NTT Cyber Space Laboratories. In 2005, he was transferred to NTT Comware Corporation as an engineer. He returned to NTT Cyber Space Laboratories in 2010 as a research engineer. His fields of interest are computer vision, video communication, user interfaces, and 3D displays. He received the Young Author's Award from ISPRS (International Society for Photogrammetry and Remote Sensing) in 1998 and the Encouragement Award from the Institute of Image Electronics Engineers of Japan (IIEEJ) in 2004. He is a member of the Association for Computing Machinery, the Society for Information Display, IEICE, the Institute of Image Information and Television Engineers (ITE), and IIEEJ.

Hidenori Okuda
Director of Research, Executive Research Engineer, Visual Media Project, NTT Media Intelligence Laboratories.
He received the B.E. and M.E. in precision machinery engineering from the University of Tokyo in 1986 and 1988, respectively, and the M.S. in computer science from Stanford University, CA, USA, in 1994. In 1988, he joined NTT Human Interface Laboratories and was mainly involved in R&D of visual communication systems. From 1997 to 2003, he was engaged in multimedia business development in the business sections of NTT and NTT-ME Corporation. He is currently interested in creating ultra-realistic remote viewing systems. He is a member of IEICE and ITE.

Atsushi Sagata
Senior Research Engineer, Supervisor, Visual Media Project, NTT Media Intelligence Laboratories.
He received the B.E. in electronic engineering from the University of Tokyo in 1994. He joined NTT in 1994 and has been engaged in R&D of video coding systems at various NTT laboratories. During 2001–2005, he was engaged in the commercial development of a video transmission system for terrestrial digital TV broadcasting at NTT Communications. He then worked on the development of an audio-visual transcoder system for terrestrial digital TV broadcasting IP retransmission services at NTT Cyber Space Laboratories until 2011. His current research interests include digital image processing, real-time hardware/software UHD-video coding systems, and UHD-video transmission systems. He is a member of IEICE.

Ryuichi Tanida
Research Engineer, Visual Media Coding Group, Visual Media Project, NTT Media Intelligence Laboratories.
He received the B.E. and M.E. in electrical engineering from Kyoto University in 2001 and 2003, respectively. He joined NTT to study video coding in order to develop a real-time H.264 software encoder application. He moved to NTT Advanced Technology Corporation in 2009 to manage various software development projects related to video compression, distribution, and TV phones. He has been at NTT Media Intelligence Laboratories since 2012 as a research engineer in the visual media coding group. His current research focuses on the development of HEVC software encoders. He is a member of IEICE.

Media Processing Technology for Achieving Hospitality while on the Go

Motoyuki Horii, Kazuhiro Arai, Masaaki Nagata, Kunio Kashino, Kaoru Hiramatsu, Atsushi Fukayama, and Hitoshi Yamaguchi

Abstract

This article introduces a service concept of *hospitality on the go*, in which users are guided while they move around a local area, as well as statistical machine translation technology and robust media search technology for supporting such services.

Keywords: statistical machine translation, robust media search, hospitality

1. Introduction

With the year 2020 in mind, NTT's goal is to implement a navigation service that can provide foreign visitors to Japan, who are moving about a local area, with detailed guidance according to the user's attributes and situation. This concept is described in detail below.

1.1 Service for guiding people to their destination in unfamiliar places

In recent years, the translation of Japanese guidance information into other languages on information displays in public transportation facilities such as train stations has been progressing. Nevertheless, information that can change at any time due to delays or accidents cannot be translated in advance. Moreover, visitors from other countries do not have an intuitive understanding of the local geography, so simply translating the names of places and exits that appear on signs does not help them decide which way to go.

To achieve a more effective navigation aid, detailed information presented in Japanese is translated in real time by using multilingual statistical translation technology, and appropriate guidance information is selected based on an estimation of the user's situation (**Fig. 1**). Robust media search (RMS) technology, which can recognize objects that the user sees around them, and various other types of recognition technology are used to estimate the user's situation. For example, when a train station employee inputs emergency information in Japanese, that information is immediately translated and sent to the smartphones of foreign visitors who are in that station, and digital signage or message boards can be translated and displayed on smartphones held up to the boards by users. It would also be possible to display navigation instructions to destinations in various languages by interworking with smartphones.

1.2 Tourist navigation service based on "What I can see now"

At tourist sites and places being visited for the first time, the surroundings as seen by the user are captured with a smartphone or head-mounted device, and that information is used to provide location-based guidance. Video of scenery can include various photographic angles and objects in the environment. RMS-object is a specialized type of object recognition technology for recognizing photographic subjects. RMS-object can be used to discover multiple objects from different viewing angles and environments with higher accuracy (**Fig. 2**). Combining RMS-object with technology capable of estimating the user's situation makes it possible to provide

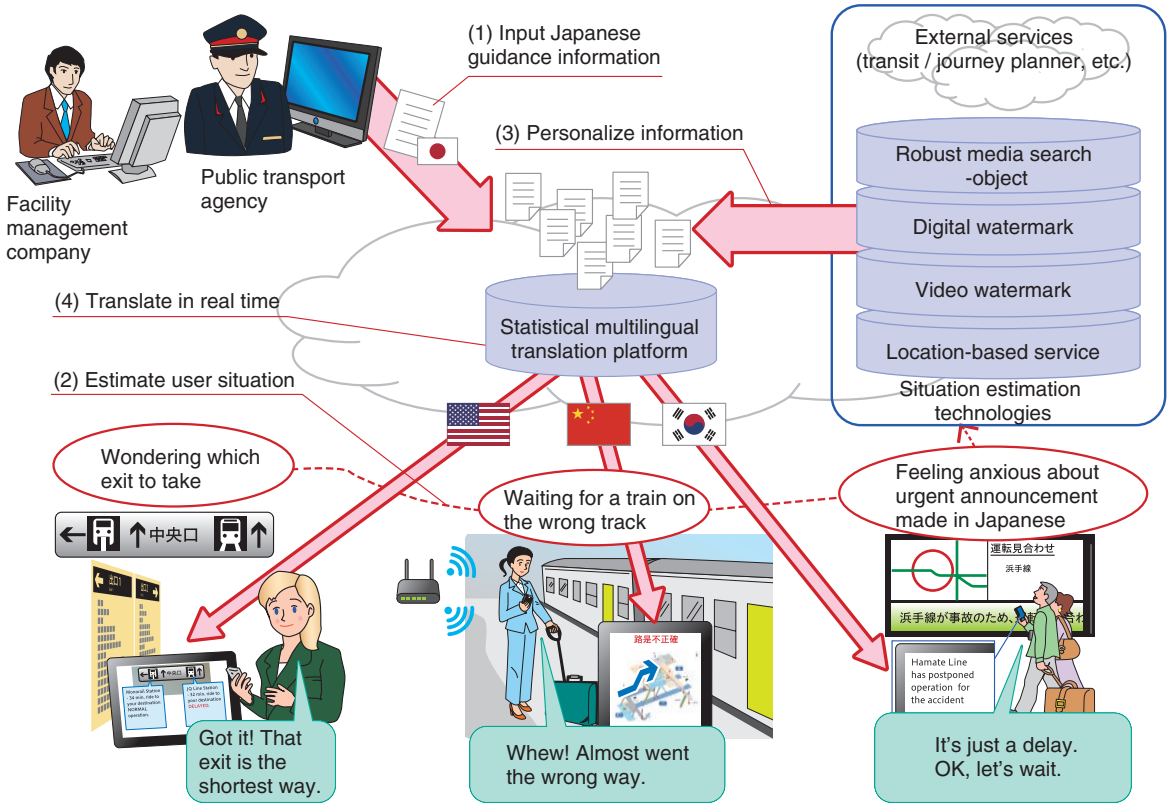


Fig. 1. Hospitable navigation (public transportation).

guidance according to the user's location at the time by selecting and displaying what is appropriate for the user's attributes and situation from the information available on the discovered objects.

2. Media processing technology to support *hospitality on the go*

NTT is moving forward with research and development (R&D) of statistical machine translation and RMS technology to be applied in implementing hospitality on the go services.

2.1 Statistical machine translation

As use of the Internet increases and the reach of globalization spreads further, the need for language translation done by computers, known as machine translation, is also increasing. Work on machine translation to eliminate the barrier of language, including work on a national level, is accelerating as we look toward 2020, and expectations are high.

R&D on machine translation has a long history, and many machine translation systems have already been developed. Nevertheless, existing systems have not really met worldwide needs and expectations, so an innovative advance in technology is needed.

Conventional machine translation systems that use a rule-based translation approach require years of work by many experts to manually produce translation rules and bilingual dictionaries for translation of a new language. Such systems have already reached the limit of accuracy of manual work, and thus, in recent years, a different approach called statistical machine translation has become mainstream. In this approach, a statistical model that is equivalent to translation rules and bilingual dictionaries is learned automatically from large-scale bilingual data on the order of several million sentence pairs.

The outline of statistical machine translation is illustrated in **Fig. 3**. It achieved at an early stage a practical level of accuracy for language pairs that have very similar word orders such as English and French. For languages that have greatly different word orders such as English and Japanese, however, it was not able to outperform the conventional rule-based translation.

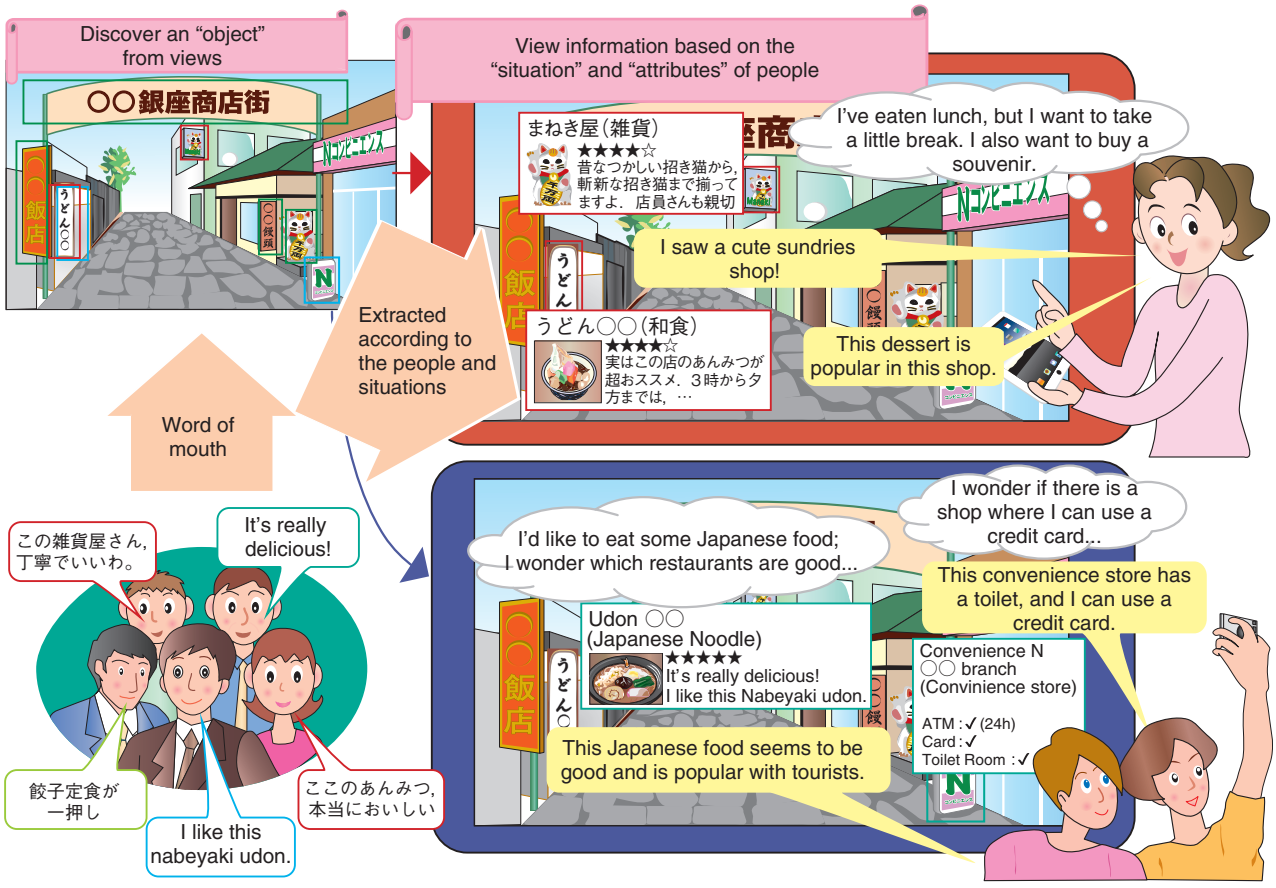


Fig. 2. Hospitable navigation (tourist site).

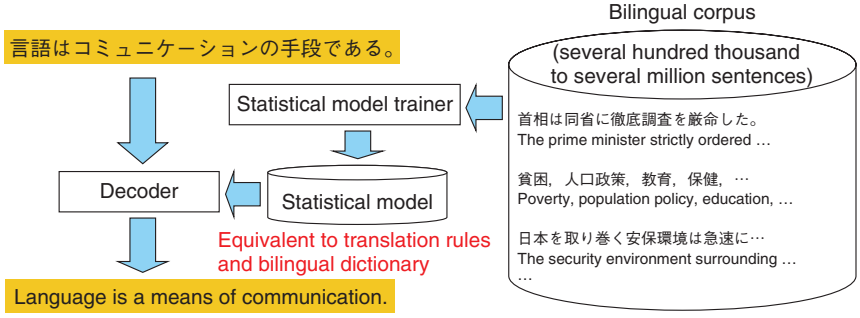


Fig. 3. Outline of statistical machine translation.

NTT has devised a method in which statistical machine translation is applied after reordering English words into Japanese word order. The reordering of English words is based on the single Japanese linguistic property of head finality [1]. For the first time in English-to-Japanese translation, we achieved a result in which statistical machine translation outperformed rule-based translation in accuracy [2].

The concept of word reordering based on the Japanese head-final property is illustrated in **Fig. 4**. The word that determines the grammatical role of a phrase in a sentence is called the *head*; or, as is learned in

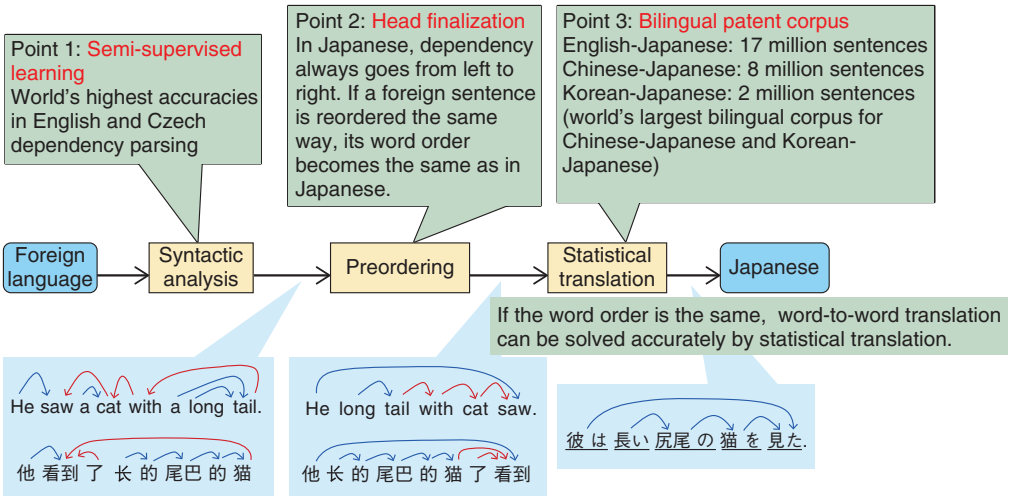


Fig. 4. Preordering based on Japanese head-final property.

Japanese elementary school, the word that is modified by others is the head. In Japanese, the dependency always goes from left to right, which is to say that the modified word is always placed at the sentence-end side. The term that describes this relationship is called *head finality*. Therefore, if we reorder the words of the translation source language (English or Chinese) so that its dependency always goes from left to right, the resulting word order becomes the same as the Japanese word order. If the word order is the same, highly accurate translation is possible through literal word-by-word translation.

Translating from Japanese to foreign languages (English or Chinese), on the other hand, is difficult because we must select the dependency relations in the input Japanese sentence that should be reversed from right to left based on the word order of the target language.

NTT has devised a translation method for changing the word order of Japanese sentences to that of the target language based on the predicate-argument structure of Japanese [3]. The predicate-argument structure is the grammatical relationship between a verb and nouns, namely, which noun is the subject of the verb and which noun is the object of the verb. As is learned in English classes in Japanese middle schools, English has an SVO (subject, verb, object) word order, while Japanese has an SOV (subject, object, verb) order.

Therefore, we first identify the predicate-argument structure of the Japanese sentence and then change the order of the *bunsetsu* phrases (roughly corresponding to noun phrases, prepositional phrases, verb phrases, etc. in English) to convert the Japanese SOV order to the SVO order of English. Because English and Japanese have an opposite word order within a bunsetsu phrase, the next step is to reorder the words in the Japanese bunsetsu phrases to match the English order (e.g., 東京で→in Tokyo). This approach reduces the number of word ordering errors in statistical translation from Japanese to English by about 30% relative to the conventional method.

The Multilingual Statistical Translation Platform (PF) was developed with the machine translation technology described above. The platform currently handles translation from English, Chinese, and Korean to Japanese and from Japanese to those languages. In addition to the main translation function, the platform provides a user dictionary function, an unknown language detection function, and other functions that are needed for business applications. Functions for user convenience, such as support for creating statistical models, which is difficult for ordinary users, are also implemented.

The quality of statistical machine translation depends on the amount of data used to train the statistical model. We achieved high-quality in translating patents by using this platform with large datasets of corresponding sentence pairs that we created from patent documents for English and Japanese (about 17 million sentence pairs), Chinese and Japanese (about 8 million sentence pairs), and Korean and Japanese (about 2 million sentence pairs). Replacing the data used when training the statistical model makes it

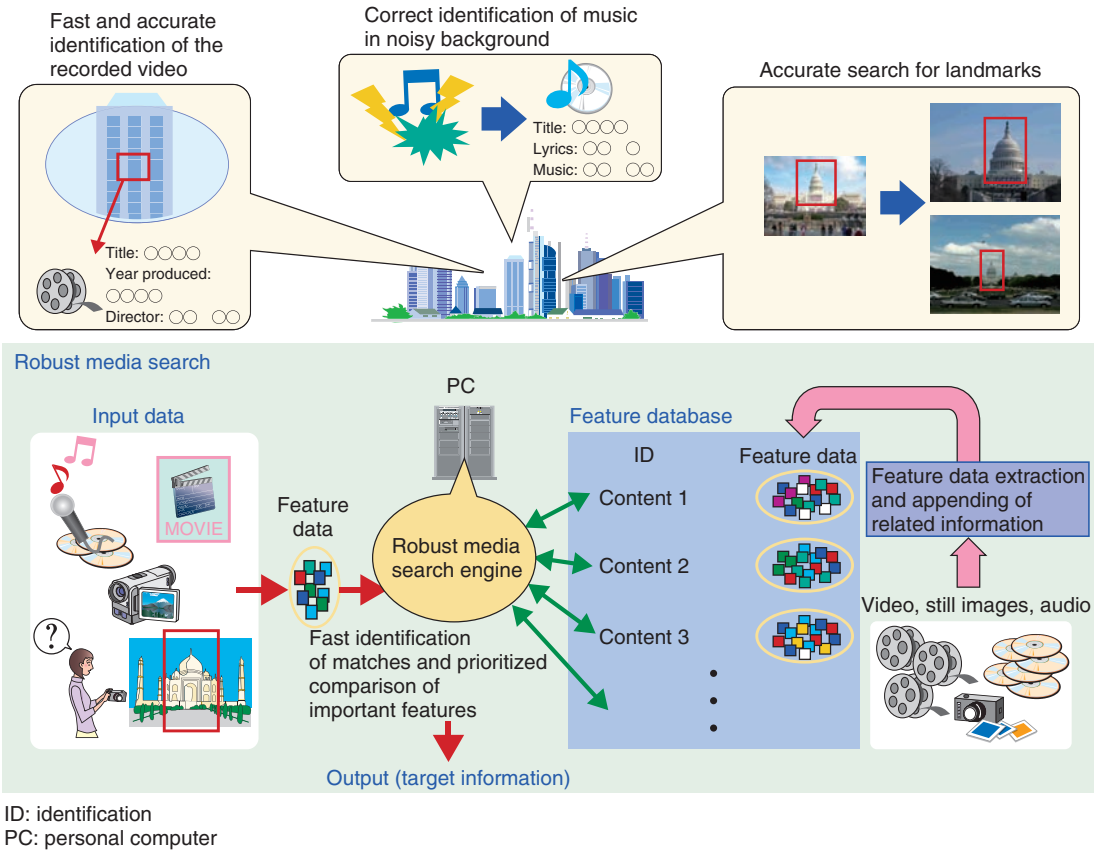


Fig. 5. Robust media search technology.

possible to automatically construct high-quality machine translation systems for particular domains other than patents.

The development of innovative machine translation technology and its implementation in a system as described above has laid the foundation for meeting worldwide needs and expectations. In the future, we will aim for even higher accuracy and a wider range of target domains to achieve machine translation that truly removes the barrier of language.

2.2 RMS

RMS is technology for using small parts of video or still-image signals from a camera or sound signals from a microphone as keys to search a large database that contains videos, music, and still images of landmarks (**Fig. 5**) [4, 5].

This kind of matching-based media search has been the subject of R&D by NTT laboratories for over 20 years, and the results have served as the core of various services, including *net monitoring* for investigating the use of video on the Internet, *music use listing* for automatically creating lists of music used in broadcast programming, and *second screen* for displaying network content related to broadcast programming on smartphones by capturing audio or video in the programming.

RMS is robust against ambient noise or obstacles, distortion in video, and interruptions in audio, and it can also search huge amounts of media data instantly. For example, it is possible to identify the name of a song that can be heard amidst street noise. For video, it is possible to quickly and accurately identify objects that are partially hidden and cannot be seen in their entirety. Because RMS uses video and sound rather than textual information, it is possible to identify what is seen or heard when the names of those things are not known by the user or when text input is difficult. For hospitality while on the go, this function can be used to recognize objects that can be seen in the surroundings and to display appropriate information according to the user's attributes and situation.

We are currently working on increasing the speed and accuracy as well as the ease of use of RMS to enable fast and accurate searches at the moment and on the spot. In the future, we will continue to study actual use environments and to do basic research on media search technology.

3. Future development

The idea of hospitality on the go places importance on understanding the user's situation and intentions, which is one of the main elements of our vision of a personal agent. People who visit Japan find it difficult to gather information in public places because of language differences and indecipherable signs and information displays. Our approach to overcoming this problem is to provide that information in different forms that suit users and to display it in useful ways.

Our goal for the future is to implement services to provide a better user experience by going beyond the translation and video search technology described in this article through interworking with technology related to geographic data and other important technical elements.

References

[1] H. Isozaki, K. Sudoh, H. Tsukada and K. Duh, "HPSG-based Preprocessing for English-to-Japanese Translation," ACM TALIP. Vol. 11, No. 3, Sept. 2012.

[2] I. Goto, B. Lu, K. P. Chow, E. Sumita, and B. K. Tsou, "Overview of the Patent Machine Translation Task at the NTCIR-9 Workshop," Proc. of NTCIR-9, pp. 559–578, Tokyo, Japan, Dec. 2011.

[3] S. Hoshino, Y. Miyao, K. Sudoh, and M. Nagata, "Two-state Preordering for Japanese-to-English Statistical Machine Translation," Proc. of IJCNLP 2013, pp. 1062–1066, Nagoya, Japan, Oct. 2013.

[4] K. Kashino, R. Mukai, K. Otsuka, H. Nagano, T. Izumitani, A. Kimura, T. Kurozumi, and J. Yamato, "Fast Media Search," NTT Technical Journal, Vol. 19, No. 6, pp. 29–32, 2007 (in Japanese).

[5] K. Kashino, "Search and Use of Huge Amounts of Media Data—towards Eliminating the Bottleneck of the Big Media Era," NTT Technical Journal, Vol. 26, No. 4, pp. 31–34, 2014 (in Japanese).

Motoyuki Horii
Senior Research Engineer, Supervisor, Project Leader of Natural Language Processing Systems Development Project, Promotion Project 1, NTT Media Intelligence Laboratories.

He received the B.S. and M.S. in computer science from Keio University, Kanagawa, in 1986 and 1988, respectively. Since joining NTT in 1988, he has mainly been engaged in R&D of natural language processing, agent communication, IPTV systems, and multimedia broadcasting systems. He is currently developing multilingual statistical machine translation systems and introducing them to NTT Group companies.

Kazuhiro Arai
Senior Research Engineer, Supervisor, Promotion Project 1, NTT Media Intelligence Laboratories.

He received the B.E. and M.E. in electrical engineering from Kansai University, Osaka, in 1987 and 1989, respectively, and the Ph.D. in information engineering from Osaka University in 1992. He joined NTT Human Interface Laboratories in 1992 and studied dialogue controls on spoken dialogue systems. During 1996–1997, he was a visiting researcher at AT&T Laboratories in Florham Park, NJ, USA. In 1999, He moved to NTT Communications and joined a commercial system development project. He developed the Voice Portal service system and softphone for voice-over-IP services. During 2009–2014, he promoted the Voice Mining platform in NTT Media Intelligence Laboratories. He is currently promoting the Multilingual Machine Translation platform.

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

He received the B.E., M.E., and Ph.D. in information science from Kyoto University in 1985, 1987, and 1999. He joined NTT in 1987. He was with Advanced Telecommunications Research Institute International (ATR), Interpreting Telephony Research Laboratories, Kyoto, from 1989 to 1993. He was a visiting researcher at AT&T Laboratories, NJ, USA, from 1999 to 2000. His research interests include natural language processing, especially morphological analysis, named entity recognition, parsing, and machine translation. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE), the Information Processing Society of Japan (IPSJ), the Japanese Society for Artificial Intelligence (JSAI), the Association for Natural Language Processing (ANLP), and the Association for Computational Linguistics (ACL).

Kunio Kashino
Senior Research Scientist, Supervisor, Distinguished Researcher, and Leader of Recognition Research Group, Media Information Laboratory, NTT Communication Science Laboratories.

He received the Ph.D. from the University of Tokyo for his pioneering work on music scene analysis in 1995. Since joining NTT in 1995, he has been working on audio and video analysis, search, retrieval, and recognition algorithms and their implementation. He has received several awards including the Maejima Award in 2010, the Young Scientists' Prize for Commendation for Science and Technology from the Ministry of Education, Culture, Sports, Science and Technology in 2007, and the IEEE (Institute of Electrical and Electronics Engineers) Transactions on Multimedia Paper Award in 2004. He is a senior member of IEEE. He is also a Visiting Professor at the National Institute of Informatics, Tokyo.

Kaoru Hiramatsu
Senior Research Scientist, Supervisor, Media Recognition Research Group NTT Communication Science Laboratories.

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

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 laboratories 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 IEICE.

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 JSAI.

Media Processing Technology for Achieving Hospitality in Information Search

Kugatsu Sadamitsu, Jun Shimamura, Go Irie, Shuhei Tarashima, Taiga Yoshida, Ryuichiro Higashinaka, Hitoshi Nishikawa, Noboru Miyazaki, Yusuke Ijima, and Yukihiro Nakamura

Abstract

This article introduces services for achieving *hospitality* in information search activities. These services are designed to assist individual users in their surroundings during day-to-day activities. We also introduce subject identification technology based on images, natural language processing technology for understanding people and responding naturally, and speech synthesis technology capable of generating synthesized speech with a wide variety of speakers and speaking styles, all of which support these services.

Keywords: subject identification, natural language processing, speech synthesis

1. *Hospitality* in information search

At NTT, our aim is to implement services that provide support that is detailed and appropriate to the user context and that is intended for individual users in various everyday activity scenarios. We introduce specific examples of this below.

1.1 Service providing users with information about unknown entities

This service presents users with information about unknown items that they are curious about, such as unfamiliar cuisine or folk art, tailored to those users (**Fig. 1**). This service is based on subject recognition technology developed by NTT laboratories. It rapidly identifies items using the cameras in smartphones and tablets that we use every day. The service uses goods-related information available on the Internet and combines it with individual characteristics of each user's culture and tastes, to provide users with content that is suited to them and in their own language. We can anticipate various services such as those that present information about Japanese cuisine or folk art to foreigners visiting Japan in their native languages, those that display cooking ingredients to people who are careful about what they eat because of their cultural background or food allergies, and those that provide shoppers with word-of-mouth information relating to products.

1.2 Agent service with soft toys that understand user's intention

People enjoy spending time with family and friends, and there are often occasions during activities such as traveling and watching sports where someone wants to pull a smartphone out and check information. However, this immediately results in a sense of being isolated from the circle of family and friends, and the person is enclosed in his own world.

What if it were possible to comprehend a user's intention in the middle of a casual conversation between users and to convey the information that is

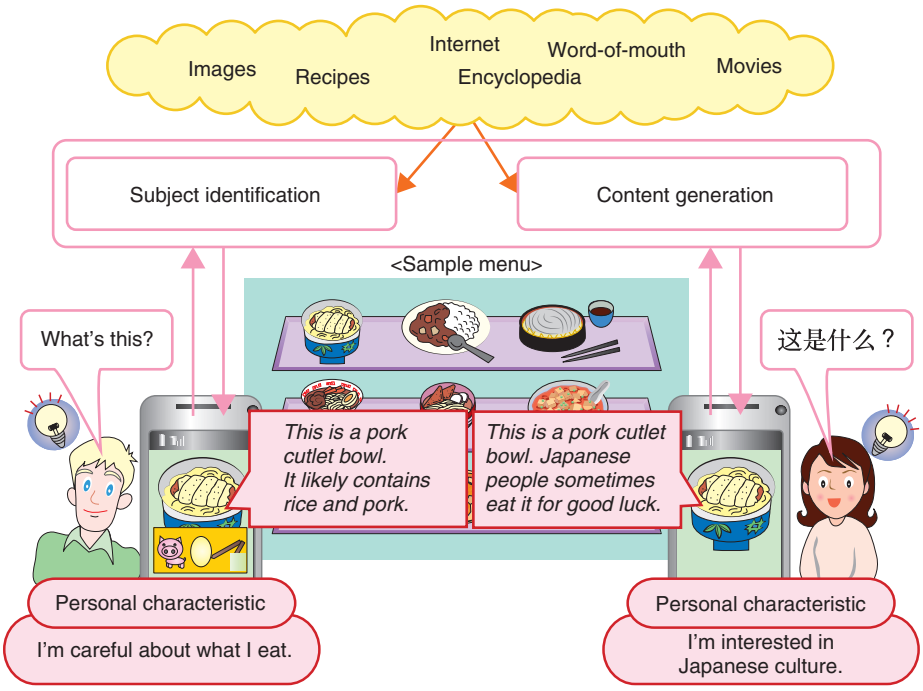


Fig. 1. Example of providing users with relevant information about unknown items.

required in a timely manner? It would seem as if a new member had joined that circle to provide the information to the user, without breaking up the circle of people or isolating any of the existing members.

The agents we have developed have a material form (such as a soft toy or puppet) and coexist physically within a circle of people, as represented by the teddy bear shown in **Fig. 2**. After comprehending the person's intent, the agent then generates utterances with the appropriate content and volume. They can behave naturally within the circle of people by speaking with a wide variety of synthesized speech. In the near future, we could have a world in which such agents are present within various circles of people.

2. Media processing technology that supports hospitality in information search

At NTT, we are working on the research and development (R&D) of subject identification technology based on images, natural language processing technology for understanding people and responding naturally, and speech synthesis technology capable of generating synthesized speech with a wide variety of speakers and speaking styles in order to implement services that add hospitality to information search.

2.1 Subject identification technology based on images

To get close to the user and provide information that depends on the user's situation, it is necessary to have a computer that can identify and comprehend the world and objects in the same way a person can. We introduce here our subject identification technology that identifies a photographed subject from an image captured by a camera that corresponds to a person's eyes. When subjects are identified from images, it is necessary to prepare reference images relating to those subjects beforehand. To cope with differences in photographic, environmental, and illumination conditions, however, it is usually necessary to prepare a large number of reference images for one subject, and such preparation requires a great deal of work. Here, we introduce our technologies for identifying subjects; one is capable of identifying three-dimensional (3D) subjects, and the other uses a cloud data application for subject identification. These technologies are expected to greatly reduce the work of preparing images in advance.

(1) 3D subject identification technology [1]

This technology makes it possible to identify even a 3D subject from a small number of reference images

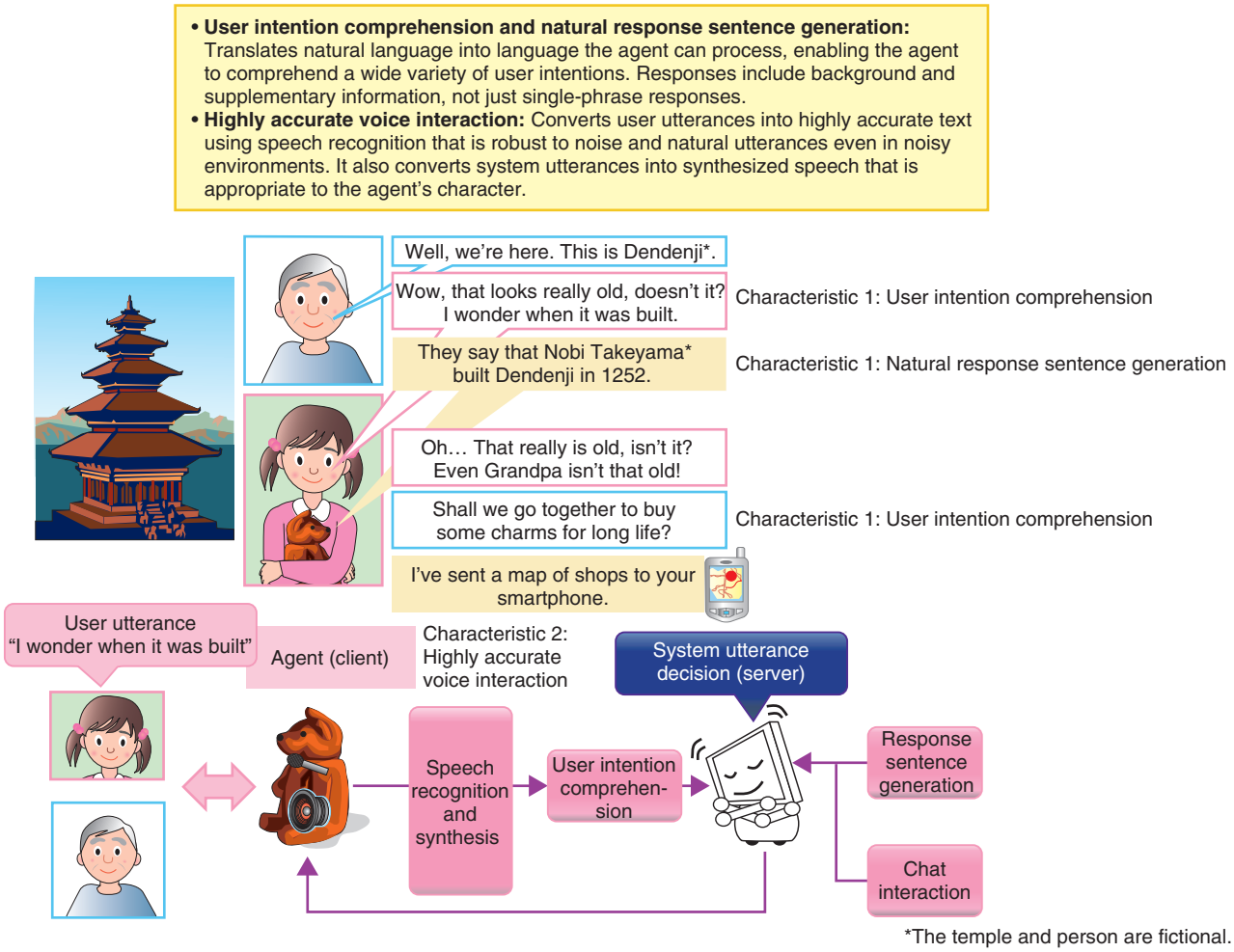


Fig. 2. Agent service using soft toys that understand user's intention.

with a high level of accuracy (**Fig. 3**). In contrast to flat objects such as books or compact discs, the appearance of a 3D object in an image will change with the direction from which it was photographed. Thus, in the past it was necessary to prepare a large number of reference images beforehand. This technology enables more robust identification of 3D objects by automatically estimating the relative direction of shooting with respect to reference images, even from an image that was captured in such a way that the user could not see its front surface. From the viewpoint of service business operators, the number of images to be prepared beforehand can be significantly reduced since images need only be registered from a few directions. In addition, this technology makes it possible to identify a number of objects highly accurately, even if they are in an untidy environment or are partially concealed, by viewing them from feature points that satisfy constraint conditions on 3D objects derived from projective geometry.

(2) Subject identification technology using cloud data application

The optimal reference images are those in which only the subject is captured accurately and there is no unnecessary background. Our cloud data application subject identification technology (**Fig. 4**) is intended to achieve this through the use of our unique subject region extraction technology [2]. This technology makes it possible to identify and extract just the region in which the subject exists from a number of images that show the subject to be identified (**Fig. 5**). Application of this technology will make it possible to create reference images that show only their

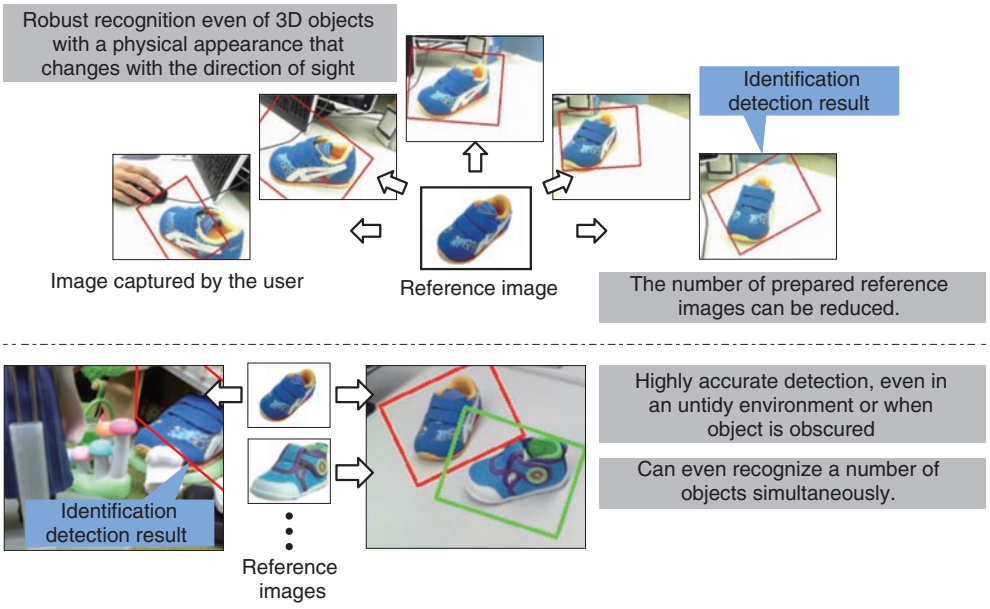


Fig. 3. Features of 3D subject identification technology.

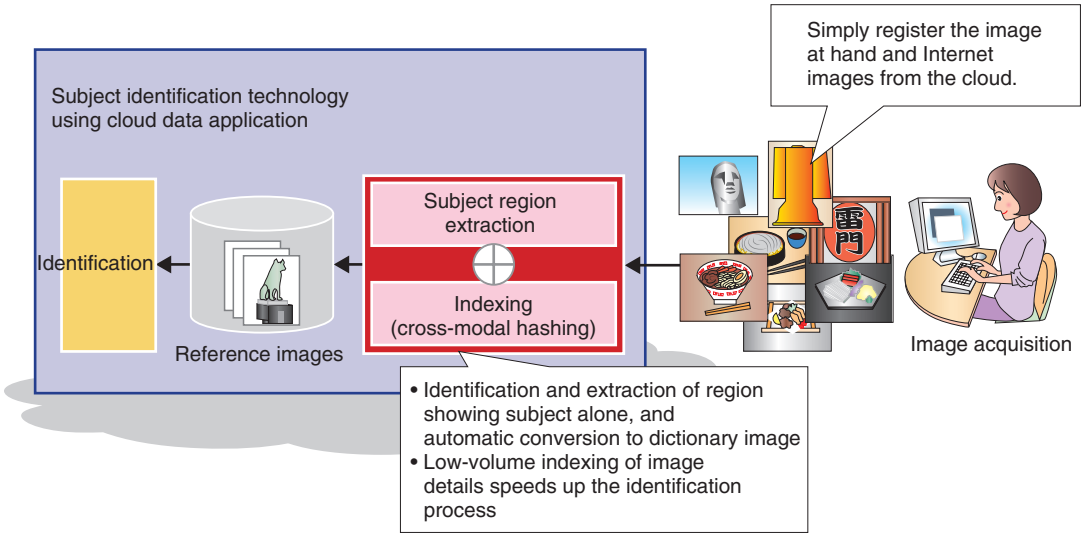


Fig. 4. Cloud data application type of subject identification.

subjects by simply registering the photograph at hand together with images acquired from the Internet. Not only does this lower the barrier to introducing and using the subject identification technology, but it is also expected to make it possible to increase the number of categories of subjects that can be identified by making it easy to prepare a huge number of reference images.

If it becomes possible to handle a large number of reference images, a rapid identification method that facilitates this will be essential. We are working on the R&D of unique indexing technology called *cross-modal hashing* [3]. This technology involves converting the content of each reference image into a very short code (hash) that is stored, making it possible to execute rapid identification processing by using the

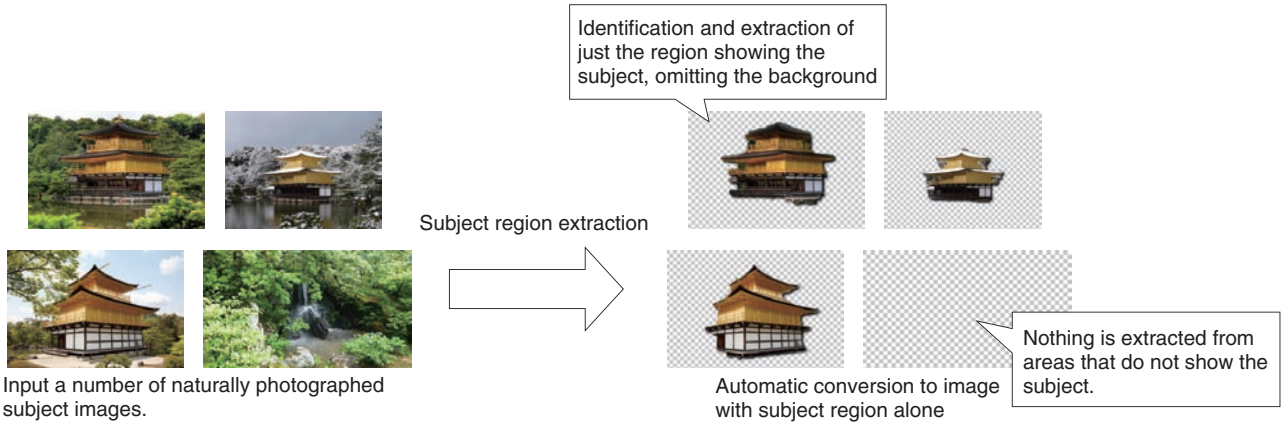


Fig. 5. Subject region extraction technology.

indexes, even with respect to a huge number of reference images. For example, a processing time of approximately 7 seconds was necessary in the past to identify subjects from a collection of 1 million reference images, but our technology has reduced that to less than half a second.

In the future, we will work on improving the technology to identify a large number of subjects rapidly and highly accurately from images, and we will continue to promote the implementation of user-friendly services that will enable smart assistance according to the user situation.

2.2 Natural language processing technology for understanding people and responding naturally

For an agent to behave naturally, it must have physical functions such as actions, but it is also important to have natural language processing technology in order to generate the content that the agent speaks. To that end, we introduce (1) user utterance intention comprehension technology, (2) automatic summary generation technology for generating natural descriptive sentences, and (3) interactive chat technology for putting together the entire interaction in a natural manner.

(1) Utterance intention comprehension technology

An overall image of the agent system is shown in **Fig. 6**. In this case, the description deals with the example of a scenario in which the agent plays the part of a sightseeing guide for a grandfather and grandchild.

The first thing that the agent should do is comprehend the intentions behind the contents of user utterances. Since a computer cannot comprehend human language, we are developing technology that roughly translates human language into a computer language (such as a database query language) [4]. For example, if the sentence "I wonder when it was built" and meta-information such as the current location can be translated into computer language as "s = Dendenji, p = year of construction, o = ?", it is possible to respond with "Dendenji was built in 1252."

(2) Automatic summary generation technology

However, simply returning single- phrase responses as described above does not make for a natural agent. We are therefore conducting research into improving the system's naturalness and intelligence by summarizing descriptive sentences as appropriate and adding them to the responses. Our approach is to take text information relating to a certain subject (Dendenji) that already exists on the web, convert it into colloquial form as a description, and provide information to be uttered by the agent. During this process, our automatic summary technology [5] plays a large role. When converting text on the web into a suitable form as a description, we use the automatic summary technology to generate a concise text passage with the redundant parts of the original text removed, and provide a natural description by converting it into colloquial form. This makes it possible to implement an inexpensive agent that provides natural descriptions. An example of the process when text on the web is converted into more colloquial-seeming speech is shown on the right side of Fig. 6. For example, the text passage: "According to the temple's history, in

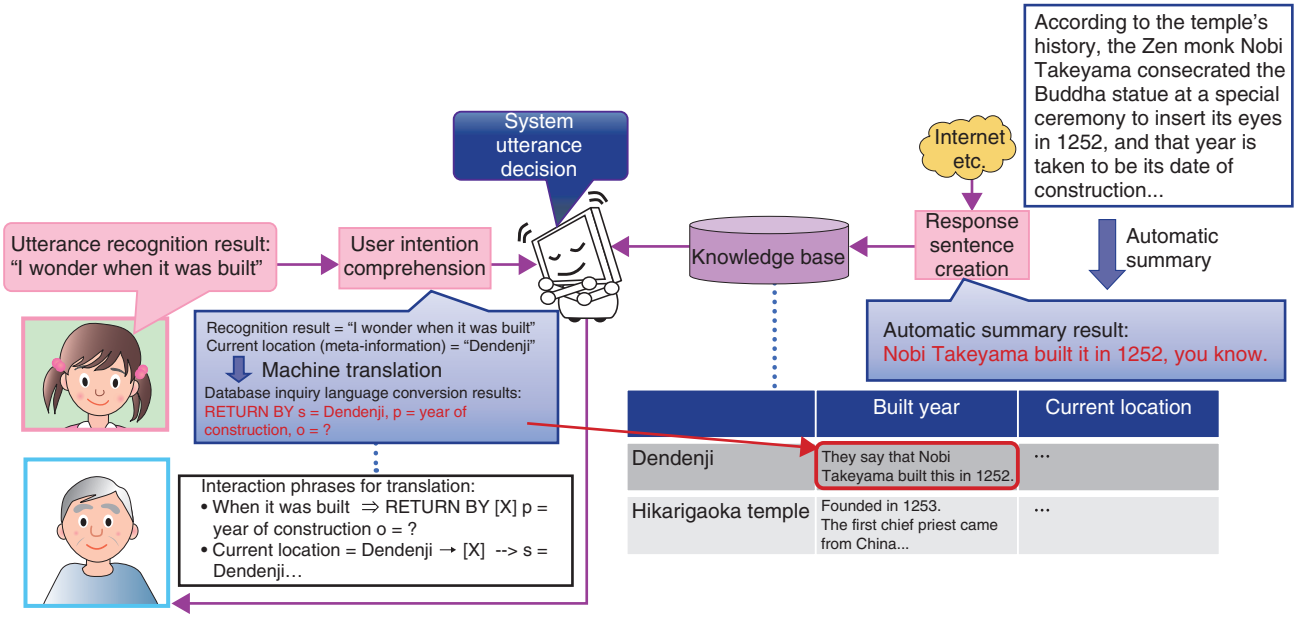


Fig. 6. Depiction of agent system.

1252 the Zen monk Nobi Takeyama* consecrated the Buddha statue at a special ceremony to insert its eyes, and that year is taken to be its date of construction..." can be used as an utterance by converting it into the brief, colloquial expression: "They say it was built by Nobi Takeyama in 1252."

(3) Interactive chat technology

It is possible to improve the quality of conversation by providing worthwhile information within the conversation, but to improve the naturalness of the conversation overall, it is necessary to improve the coverage of topics. Therefore, what we need is a chat function. A chat function is difficult to write in algorithmic form, though, and up until now, this kind of function has been implemented using hand-made rules. However, methods that rely on rules are expensive, and their coverage of topics is low. That is why we have constructed an interaction system that can chat automatically on a wide range of topics by turning text data on the Internet into interaction knowledge using language processing technology [6, 7] (**Fig. 7**). The system generates utterances from the predicate argument structure data (structure based on sentences formed of subjects, objects, etc.), based on the current topic and the utterance intentions of the system, and responds with a wide range of topics by selecting utterance sentences from an utterance database. It is also possible to make the phraseology suit the system characteristics by using the sentence-end expression conversion function.

Thus, natural language processing technology is essential for implementing an advanced interaction agent. In the future, we will continue with our R&D aimed at attaining a deeper understanding of language and on generating language for different objectives.

2.3 Speech synthesis technology capable of generating synthesized speech of a wide variety of speakers and speaking styles

Up until now, the main uses for speech synthesis technology were in information provision services such as those for verbally confirming safety information and those providing automated voice guidance systems in call centers. However, usage purposes have recently expanded to include applications such as voice interaction agents due to the recent spread of high-performance mobile terminals. Since the objective of information provision services is to convey information accurately by voice, it was sufficient to generate a synthesized speech with a single speaker. In contrast, with voice interaction agents, the requirement is now for synthesized speech with a wider

* This temple and the monk are fictional.

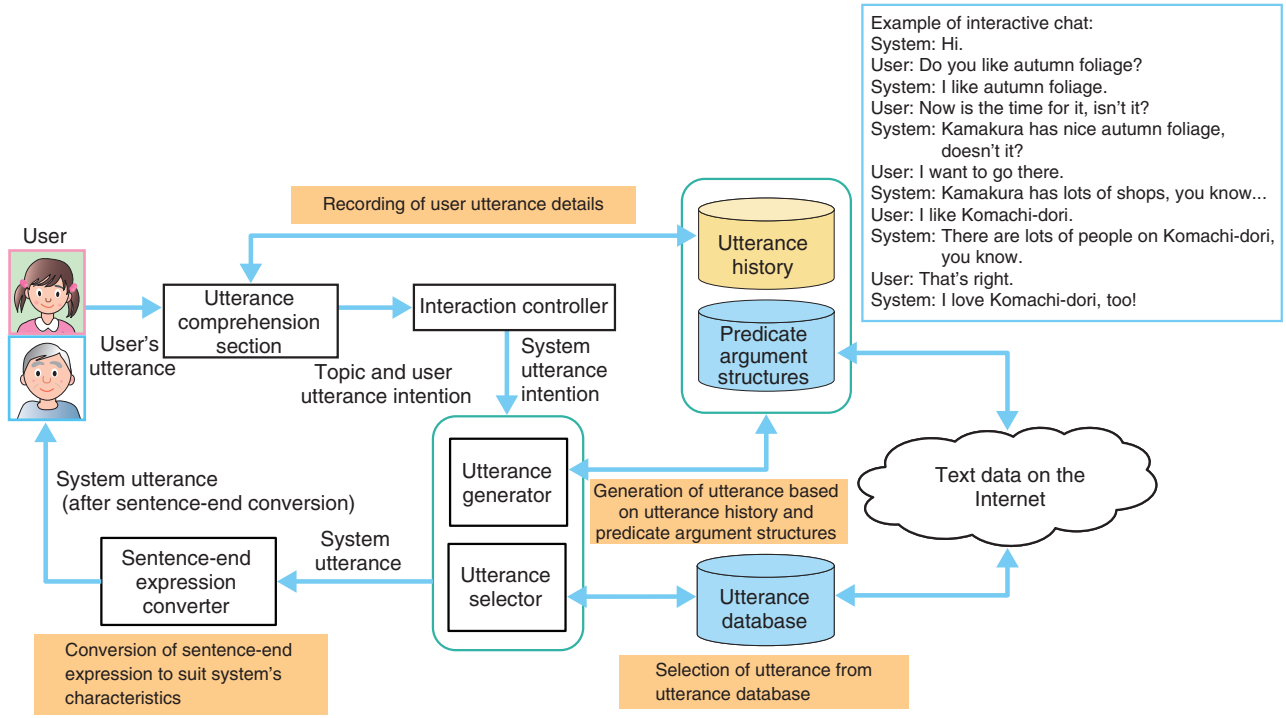


Fig. 7. Processing flow of interactive chat technology.

variety of speakers and speaking styles than in the past, for example, synthesized speech of various speakers that correspond to agents' characters, and synthesized speech with speaking styles that correspond to scenes that induce emotional expressions. We introduce here the novel speech synthesis technology that was developed by NTT Media Intelligence Laboratories for just such usage scenes, which enables the generation of synthesized speech for a wide variety of speakers and speaking styles.

An overview of the speech synthesis system is shown in **Fig. 8**. This technology consists of a training component that trains a model from an arbitrary speaker's speech and retains the characteristics of that speaker's speech, and a speech synthesis component that generates synthesized speech using that trained model.

In the training component, speech data uttered by a specified speaker are recorded. Then a model that has the speaker characteristics (voice quality and speaking style) of the target speaker is trained from the recorded speech of that speaker. The trained model consists of two parts: a speaker model that holds information on the voice quality of that speaker, and a style model that holds information on the speaking style of the speaker, for example, voice pitch and speaking rate.

In the speech synthesis component, a segment of synthesized speech that has the voice quality and speaking style of the speaker is generated from the trained speaker model and style model of the speaker. This technology also makes it possible to impart speaking styles to synthesized speech from within the style model that was trained beforehand, for example, a butler-style speaking style or a dramatic-reading-style speaking style. This means that when the speech synthesis is processed, it is possible to generate synthesized speech that has been given a specific speaking style while still having the voice quality of that speaker (such as speech with a dramatic-reading-style of speaking but with the voice quality of Ms. A).

One feature of this technology is that a shorter time for recording the speech of a speaker is necessary when implementing the speech synthesis of an arbitrary speaker. In order to generate synthesized speech having sufficient quality (reproducibility of speaker characteristics and naturalness of the synthesized speech) using conventional speech synthesis technology, a huge amount of speech data uttered by the

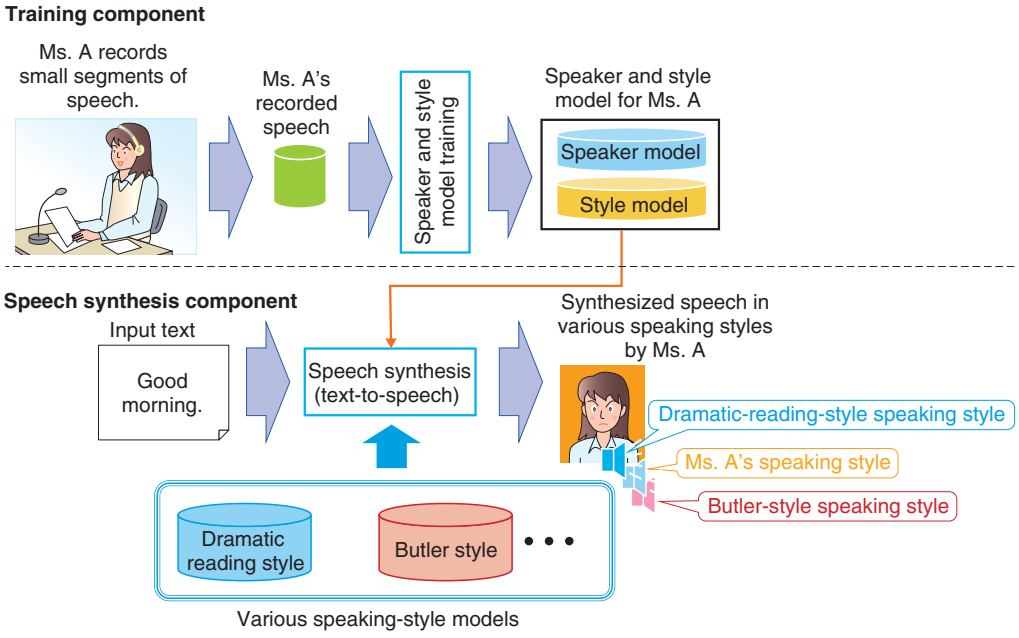


Fig. 8. Overview of speech synthesis system.

speaker are required. Since an extended speech recording taking between several hours to a dozen hours is necessary for collecting such speech data, it is difficult to create synthesized speech of various speakers because of the cost. This technology implements the simple creation of synthesized speech of various speakers corresponding to agents' characters, by greatly reducing the necessary speech recording time to between several dozen minutes and a couple of hours.

Our future task is to improve the basic performance such as the naturalness of the synthesized speech and the reproducibility of speaker characteristics. We will also investigate the speech synthesis required for the voice interaction interface, for example, the generation of synthesized speech with appropriate speaking styles for various usage scenes.

3. Future plans

Following our theme of hospitality in information search, we are focusing on issues that users encounter in their everyday lives and are working to involve agents in this process. We believe that further technical developments in the image recognition, language processing, and speech synthesis technologies introduced in this article will help achieve this.

In the future, we will pursue research on a new style of providing information naturally by using an approach that will help us understand potential preferences and cultural backgrounds, and by the entry of agents into user circles, with the aim of implementing services that will develop new relationships between users and agents that have not been seen before.

References

[1] J. Shimamura, T. Yoshida, and Y. Taniguchi, "Geometric Verification Method to Handle 3D Viewpoint Changes," MIRU2014 (the 17th Meeting on Image Recognition and Understanding), OS3-4, Okayama, Japan, Jul. 2014.
[2] S. Tarashima, G. Irie, H. Arai, and Y. Taniguchi, "Fast Web Image Object Cosegmentation with Region Matching," Proc. of Media Computing Conference 2014 (the 42nd Annual Conference of the Institute of Image Electronics Engineers of Japan), R4-2, Tokyo, Japan, Jun. 2014.
[3] G. Irie, H. Arai, and Y. Taniguchi, "Hashing with Locally Linear Projections," IEICE Transactions on Information and Systems (Japanese Edition), Vol. J97-D, No. 12, pp. 1785–1796, 2014.
[4] R. Higashinaka, K. Sadamitsu, W. Uchida, and T. Yoshimura, "Question Answering Technology in Shabette-Concier," NTT Technical Journal, Vol. 25, No. 2, pp. 56–59, 2013 (in Japanese).
[5] H. Nishikawa, K. Arita, K. Takana, T. Hirao, T. Makino, and Y. Matsuo, "Learning to Generate Coherent Summary with Discriminative Hidden Semi-Markov Model," Proc. of COLING 2014 (the 25th International Conference on Computational Linguistics), pp. 1648–1659, Dublin, Ireland, Aug. 2014.
[6] R. Higashinaka, "Towards an Open-domain Dialogue System," Proc. of the 70th SIG-SLUD (Special Interest Group on Spoken Language Understanding and Dialogue Processing) of the Japanese Society for Artificial Intelligence, Vol. 70, pp. 65–70, 2014.
[7] R. Higashinaka, K. Imamura, T. Meguro, C. Miyazaki, N. Kobayashi,

H. Sugiyama, T. Hirano, T. Makino, and Y. Matsuo, “Towards an Open-domain Conversational System Fully Based on Natural Language Processing,” Proc. of COLING 2014, pp. 928–939, Dublin, Ireland, Aug. 2014.

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

Jun Shimamura
Senior Research Engineer, Visual Media Project, NTT Media Intelligence Laboratories.
He received the B.E. in engineering science from Osaka University in 1998 and the M.E. and Ph.D. from Nara Institute of Science and Technology in 2000 and 2006, respectively. He joined NTT Cyber Space Laboratories (now, NTT Media Intelligence Laboratories) in 2000. From 2006 to 2010, he worked at NTT Communications. He received the 2005 TAF (Telecommunication Advancement Foundation) TELECOM System Technology Award. He is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) and the Institute of Image Information and Television Engineers (ITE).

Go Irie
Research Engineer, Visual Media Project, NTT Media Intelligence Laboratories.
He received the B.E. and M.E. in system design engineering from Keio University, Kanagawa, in 2004 and 2006, respectively, and the Ph.D. in information science and technology from the University of Tokyo in 2011. He joined NTT Cyber Solutions Laboratories in 2006. He is a member of IEICE and ITE.

Shuhei Tarashima
Researcher, Visual Media Project, NTT Media Intelligence Laboratories.
He received the B.E. and M.E. in engineering from the University of Tokyo in 2009 and 2011. He joined NTT Cyber Solutions Laboratories in 2011. His research interests include machine learning and pattern recognition in computer vision, for example, object classification, detection, and segmentation.

Taiga Yoshida
Researcher , Visual Media Project, NTT Media Intelligence Laboratories.
He received the B.E. and M.E. in informatics from Kyoto University in 2007 and 2009, respectively. He joined NTT Cyber Solutions Laboratories in 2009 and studied video recommendation functions based on metadata and audio-visual features. He is currently studying object recognition. He is a member of IEICE.

Ryuichiro Higashinaka
Senior Research Scientist, Audio, Speech, and Language Media Project, NTT Media Intelligence Laboratories.
He received the B.A. in environmental information, the Master of Media and Governance, and the 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 IEICE, the Japanese Society for Artificial Intelligence (JSAI), IPSJ, and NLP.

Hitoshi Nishikawa
Research Engineer, Audio, Speech, and Language Media Project, NTT Media Intelligence Laboratories.
He received the B.A. in policy management and the Master of Media and Governance from Keio University, Kanagawa, in 2006 and 2008, respectively, and the Ph.D. in engineering from Nara Institute of Science and Technology in 2013. He joined NTT in 2008. His research interests lie in the area of natural language processing, especially the study of automatic summarization. He received the Best Paper Award (first place) and Annual Meeting Excellent Paper Award from NLP in 2013 and 2014, respectively. He is a member of the Association for Computational Linguistics (ACL), NLP, JSAI and IPSJ.

Noboru Miyazaki
Senior Research Engineer, Audio, Speech, and Language Media Project, NTT Media Intelligence Laboratories.
He received the B.A. and M.E. from Tokyo Institute of Technology in 1995 and 1997, respectively. He joined the NTT Basic Research Laboratories in 1997. From 2004 to 2007, he was with NTT Cyber Space Laboratories (now, NTT Media Intelligence Laboratories). From 2008 to 2012, he was with NTT-IT Corporation. His current research interests include speech synthesis and spoken dialogue systems. He is a member of IEICE, the Acoustical Society of Japan (ASJ), and JSAI.

Yusuke Ijima

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

He received the B.E. in electric and electronics engineering from the National Institution for Academic Degrees and University Evaluation upon graduation from Yatsushiro National College of Technology, Kumamoto, in 2007, and the M.E. in information processing from Tokyo Institute of Technology in 2009. He joined NTT Cyber Space Laboratories (now, NTT Media Intelligence Laboratories) in 2009, where he engaged in R&D of speech synthesis. He is a member of ASJ and 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. 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 Robotics Society of Japan (RSJ), the Japan Society of Mechanical Engineers (JSME), and the Society of Instrument and Control Engineers (SICE).

Media Processing Technology for Business Task Support

Takanobu Oba, Kazunori Kobayashi, Hisashi Uematsu, Taichi Asami, Kenta Niwa, Noriyoshi Kamado, Tomoko Kawase, and Takaaki Hori

Abstract

This article introduces two aspects of work toward implementation of services to support business tasks, specifically the use of speech recognition in very noisy environments such as factories and construction sites, and technology for recording minutes of meetings. The most recent audio and speech processing technology that is applied in these services is also described.

Keywords: intelligent microphone, meeting minutes support, speech recognition

1. Introduction

The progress achieved in information and communication technology (ICT) has increased the efficiency of various tasks. For example, conversion of paper documents to electronic form allows information to be accessed and managed over a network and simplifies search and display of statistical information. A disadvantage, however, is that it is sometimes difficult to enter information using a keyboard in environments where there are no desks such as outdoors and in situations where both of a person's hands are occupied by a task such as driving a vehicle or operating machinery. These problems have created a barrier to the use of ICT. Another problem is the recording of meeting minutes and similar situations in which people exchange information verbally. The process involves manually entering information into electronic documents by relying on memory or written notes, which is time-consuming and may produce incomplete results.

2. Speech recognition and noise reduction technology

We describe here two examples of the use of voice recognition in noisy environments such as factories or construction sites and the use of speech recognition in meetings to support the production of minutes.

2.1 Intelligent microphone capable of speech recognition in noisy environments

The intelligent microphone consists of multiple microphones and is based on acoustic signal processing technology that segregates the target speech from other sounds such as other voices or background noise. The intelligent microphone makes it possible to pick up the speaker's voice clearly even in very noisy environments (**Fig. 1**).

This microphone enables high-quality telephone calls (clear hands-free calling) (**Fig. 2**) and highly accurate voice recognition, even in noisy environments such as in factories or construction work sites or in an automobile traveling on a highway.

The acoustic signal processing involves the use of spatial power distribution, frequency characteristics, and temporal fluctuation characteristics to estimate the spectral filter in order to segregate the target speech from other noise (**Fig. 3**). Adapting the signal processing to microphone observations makes it possible to reduce the ambient noise power by a factor of 1/10,000 with less degradation of the target voice signal. This performance enables accurate speech

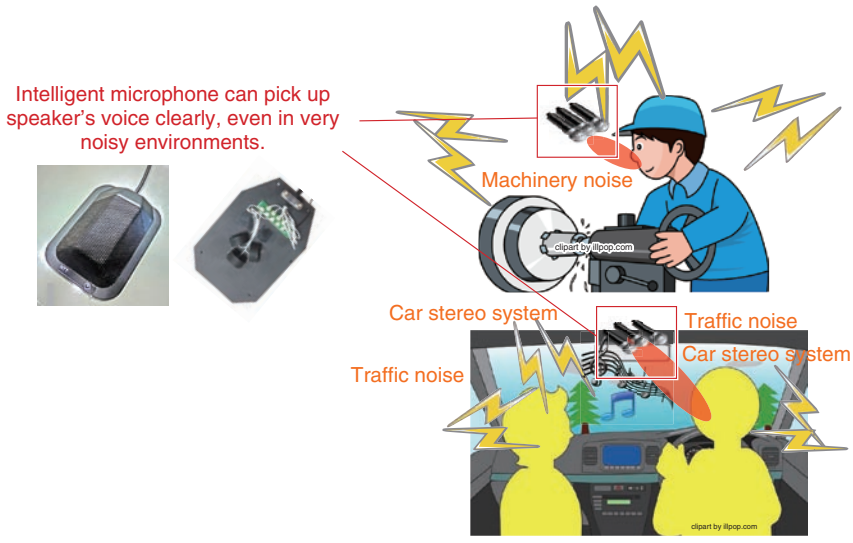


Fig. 1. Concept of the intelligent microphone.

Headset-type intelligent microphone

Fig. 2. Photo and use scenario of intelligent microphone.

recognition and high quality telephone calls even in extreme 100-dB-level noise environments. The steps involved in acoustic signal processing are as follows:

(1) Hybrid of beamforming and spectral filtering

With some microphones, sharp directivity cannot be formed. Thus, enhancement of the target speech cannot be achieved by simply applying beamforming. Adapting a spectral filter to the beamforming output makes it possible to suppress noise efficiently.

(2) Estimation of spatial interference noise power distribution

The desired speech source cannot be enhanced by simply applying conventional beamforming. Therefore, multiple beams are formed to estimate the spatial interference noise power distribution. If the estimated power of the desired sound and interference noise differ, the noise output power will be efficiently reduced.

(3) Reduction of diffuse noise using temporal fluctuation

The observed signals include diffuse noise such as that from an air conditioner. The power spectrum of diffuse noise changes over time. By utilizing these characteristics, we can accurately estimate the frequency spectrum of diffuse noise.

This newly developed intelligent microphone can be implemented for various terminals and contribute to speech services in noisy environments.

2.2. Support for producing meeting minutes

Meetings occur frequently in the business world. What was talked about in detailed and complex discussions is often forgotten a few days later, though, so the content is recorded in the form of minutes. That, however, is not a simple task and requires considerable time. Taking careful notes during the meeting for later preparation of minutes makes it difficult to

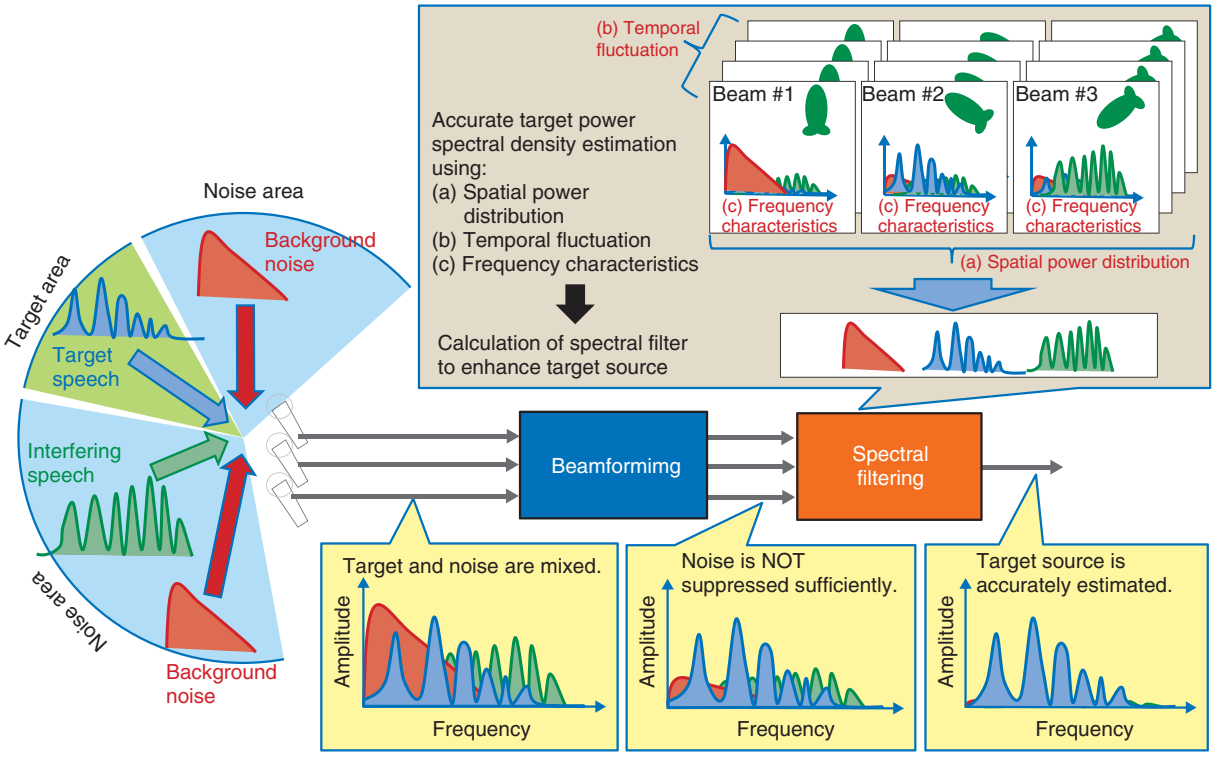


Fig. 3. Acoustic signal processing with the intelligent microphone.

follow the discussion, and priority on note-taking interferes with participation in the meeting.

NTT is developing a service for efficient production of minutes for National Diet sessions and the Assembly meetings of local governments. As a replacement for stenographers, the system supports the creation of minutes by using speech recognition for speech-to-text conversion. The next development target in this project is a system for recording everything that is spoken in business conferences and meetings as text. Having the content in text form would make it possible to rapidly search for particular parts of particular meetings and to retrieve the results. Because the voice recordings also remain, it is also possible to listen to any part of the meeting again.

In this work, real-time speech recognition during meetings is important. This has multiple advantages, one of which is that important comments can be tagged at the meeting (**Fig. 4**). As soon as a speaker utters a comment, the content of the utterance is recognized, transcribed, and displayed on the personal computer. Any participants looking at the display can tag an important utterance simply by clicking it, which is depicted as a green star in Fig. 4. The text consisting of only the tagged comments would appear as simple meeting minutes. Another advantage is that it can be used as a tool for promoting conversation, thus extending use beyond the framework as a simple meeting minutes creation support system. The system in current development has a tag cloud function, which displays words that have been used in the meeting with high frequency (**Fig. 5**). When participants can view the keywords during the discussion, they can keep and organize a view of what has been said as long as the discussion continues, which is likely to stimulate ideas.

Real-time speech recognition also makes it immediately clear which participants speak more and which speak less. Also, speakers may create a negative impression by speaking too rapidly, a manner that we should correct. It is difficult to realize how we are speaking while concentrating on the meeting, but this system may help us by bringing the problem to our attention.

Performing real-time speech recognition in this way can be expected to increase the productivity of meetings. The reason many people think that long

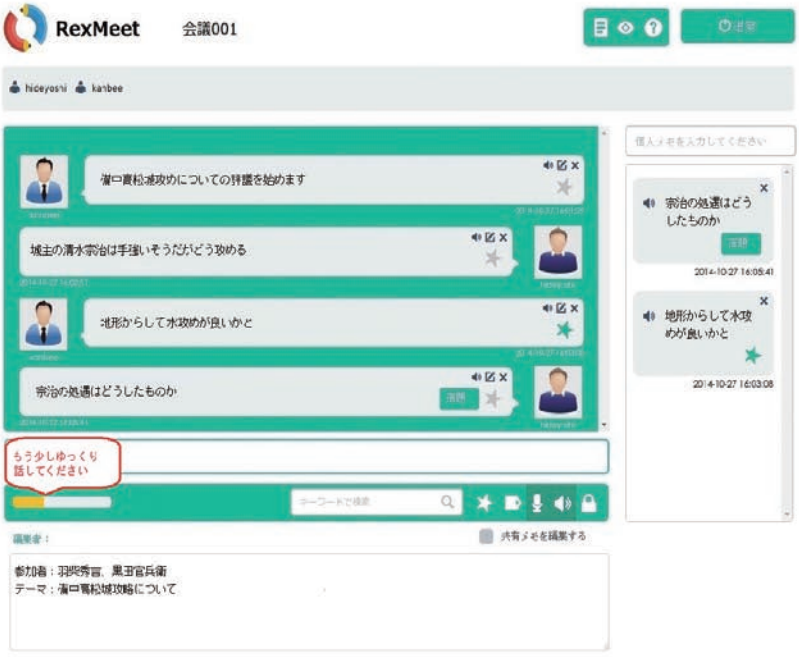


Fig. 4. Screen display of meeting minutes support system.

Fig. 5. Tag cloud.

meetings and conferences are not interesting is probably that daily meetings and other such meetings are low in productivity. We believe that speech recognition can contribute to improving this situation. Although minutes production support is one current target for this technology, we are moving forward with research and development (R&D) aimed at providing support for the meeting process itself.

3. Voice activity detection (VAD)

Voice activity detection (VAD) is a basic and essential

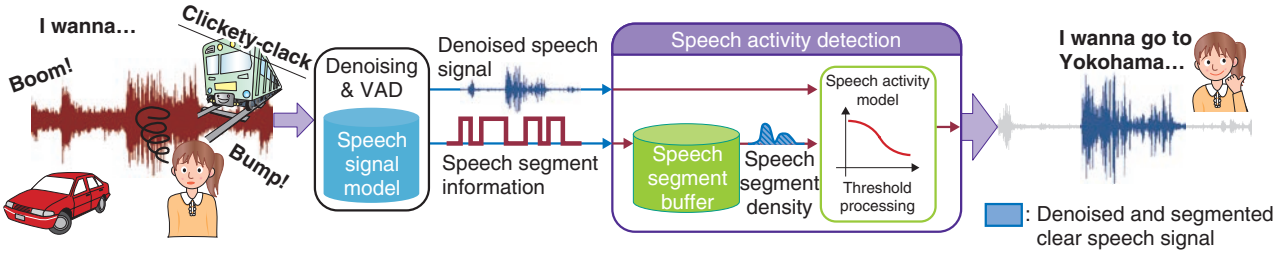


Fig. 6. VAD taking voice activity density into account.

function for many voice applications that involve speech recognition. VAD technology identifies the parts of a microphone signal that actually contain speech by monitoring the microphone signal input and detecting the points where speech begins and ends. Errors in detecting those points are a major problem that greatly reduces the utility of a speech interface for application programs. Detection errors can cause the application to react before the user finishes speaking or result in the application failing to react even after the user finishes speaking, causing a delay.

One difficult problem in VAD is determining whether a short interval is the end of an utterance or part of the utterance. NTT has addressed this problem by developing a technique to identify the temporal density of speech signals that takes the density of speech segments into account (**Fig. 6**). Speech varies in a characteristic way as an utterance progresses. By processing that variation according to our original speech model, we succeeded in extracting the utterance segments of a user with high accuracy. Furthermore, by combining VAD technology previously developed by NTT with technology for simultaneously executing noise suppression, we were able to reduce VAD errors by at least 15% compared to conventional technology, even in a high-noise environment.

4. VoiceRex NX2014 real-time DNN speech recognition engine

The term *deep learning* is currently attracting a lot of interest in media processing fields such as speech recognition and image recognition. Deep learning is mainly pattern processing that uses deep neural networks (DNNs), where *deep* refers to many layers in a neural network. The use of DNNs has remarkably improved the accuracy of speech and image recognition.

This has surprised many researchers because it had been strictly shown mathematically that the representational power of a neural network of three or more layers cannot be increased by adding layers. Nevertheless, when it comes to the question of whether or not the representative capability can be controlled by using an engineering approach, increasing the number of layers while restricting the number of nodes per layer (as in DNNs) is completely different from the conventional approach of increasing the number of nodes while keeping the number of layers at three.

DNNs are used in speech recognition as an acoustic model for determining what phoneme (“ah,” “ee,” etc.) is represented by a particular part of the speech input (**Fig. 7**). Conventional techniques have not been able to achieve a practical level of recognition accuracy when applied to speech recognition in conversations. With the emergence of DNN, however it has become possible to achieve highly accurate conversational speech recognition.

NTT developed the VoiceRex NX2014 speech recognition engine, introducing DNNs at an early stage. One feature of the VoiceRex series is high-speed recognition processing. However, the high computational cost of DNNs resulted in slower processing. That problem spurred the development of various techniques for increasing speed, and the result was a speech recognition engine that is capable of processing speed at the same high level as the original system. From the standpoint of speech recognition service providers, this technology makes it possible to provide services with a speech recognition function that is more accurate yet just as fast as was possible previously. From the user’s viewpoint, the speech recognition engine returns highly accurate results immediately after the user speaks.

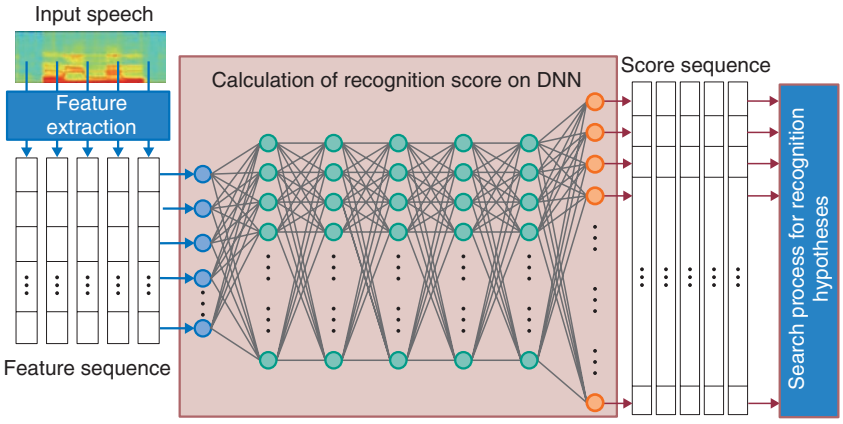
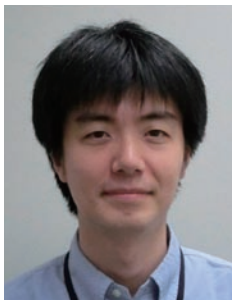


Fig. 7. DNN speech recognition mechanism.

5. Future development

We have described our work on speech processing as a core technology for supporting tasks in business situations. Language processing and image processing also play important roles in supporting business tasks. By gaining a deeper understanding of human language and of the objects that people see and the situations that exist around them, we can achieve more intelligent support in a wider variety of scenarios through the use of ICT equipment. The NTT laboratories are engaged in R&D of various types of media including speech, language, and images. Our objective in the future is to provide more advanced support for work tasks and for people in general through an organic integration of media.



Takanobu Oba
Research Engineer, Audio Speech and Language Media Project, NTT Media Intelligence Laboratories.
He received the B.E., M.E., and Ph.D. in engineering from Tohoku University, Miyagi, in 2002, 2004, and 2011, respectively. In 2004, he joined NTT, where he engaged in research on spoken language processing at NTT Communication Science Laboratories. He has been with NTT Media Intelligence Laboratories since 2012. He received the 25th Awaya Kiyoshi Science Promotion Award from the Acoustical Society of Japan (ASJ) in 2008. He is a member of ASJ, the Institute of Electrical and Electronics Engineers (IEEE), and the Institute of Electronics, Information and Communication Engineers (IEICE).

Kazunori Kobayashi
Senior Research Engineer, Audio Speech and Language Media Project, NTT Media Intelligence Laboratories.
He received the B.E., M.E., and Ph.D. in electrical and electronic system engineering from Nagaoka University of Technology, Niigata, in 1997, 1999, and 2003, respectively. Since joining NTT in 1999, he has been engaged in research on microphone arrays, acoustic echo cancellers, and hands-free systems. He is a member of ASJ and IEICE.

Hisashi Uematsu
Senior Research Engineer, Audio Speech and Language Media Project, NTT Media Intelligence Laboratories.
He received the B.E., M.E., and Ph.D. in information science from Tohoku University, Miyagi, in 1991, 1993, and 1996. He joined NTT in 1996 and has been engaged in research on psychoacoustics (human auditory mechanisms) and digital signal processing. He is a member of ASJ.

Taichi Asami
Research Engineer, Audio Speech and Language Media Project, NTT Media Intelligence Laboratories.
He received the B.E. and M.E. in computer science from Tokyo Institute of Technology in 2004 and 2006, respectively. He joined NTT Cyber Space Laboratories (now NTT Media Intelligence Laboratories) in 2006 and studied speech recognition, spoken language processing, and speech mining. He received the Awaya Prize Young Researcher Award in 2012 and the Sato Prize Paper Award in 2014 from ASJ. He is a member of ASJ, IEEE, IEICE, and the International Speech Communication Association (ISCA).

Kenta Niwa
Research Engineer, Audio Speech and Language Media Project, NTT Media Intelligence Laboratories.
He received the B.E., M.E., and Ph.D. in information science from Nagoya University, Aichi, in 2006, 2008, and 2014, respectively. Since joining NTT in 2008, he has been engaged in research on microphone array signal processing. He was awarded the Awaya Prize by ASJ in 2010. He is a member of IEEE, ASJ, and IEICE.

Noriyoshi Kamado
Research Engineer. Audio Speech and Language Media Project, NTT Media Intelligence Laboratories.
He received the B.E. from Nagaoka National College of Technology, Niigata, in 2007, the M.E. in electrical and electronic systems engineering from Nagaoka University of Technology in 2009, and the Ph.D. from Nara Institute of Science and Technology in 2012. He joined NTT Media Intelligence Laboratories in 2012, where he has been working on speech signal processing technologies for a speech recognition system. He is a member of ASJ, IEEE, and the Audio Engineering Society (AES).

Tomoko Kawase
Researcher, NTT Media Intelligence Laboratories.
She received the B.E. and M.E. in system design engineering from Keio University, Kanagawa, in 2011 and 2013, respectively. Since joining NTT in 2013, she has been working on signal processing technologies for a speech recognition system. She is a member of ASJ.

Takaaki Hori
Senior Research Scientist, Signal Processing Group, Media Information Laboratory, NTT Communication Science Laboratories.
He received the B.E. and M.E. in electrical and information engineering from Yamagata University in 1994 and 1996, respectively, and the Ph.D. in system and information engineering from Yamagata University in 1999. Since 1999, he has been engaged in research on spoken language processing at NTT Cyber Space Laboratories (now NTT Media Intelligence Laboratories). He was a visiting scientist at the Massachusetts Institute of Technology, USA, from 2006 to 2007. He is currently a senior research scientist at NTT Communication Science Laboratories. He received the 22nd Awaya Prize Young Researcher Award from ASJ in 2005, the 24th TELECOM System Technology Award from the Telecommunications Advancement Foundation in 2009, the Kiyasu Special Industrial Achievement Award from the Information Processing Society of Japan in 2012, and the 58th Maejima Hisoka Award from the Tsushinbunka Association in 2013. He is a senior member of ASJ, IEEE, and IEICE.

Audio-visual Technology for Enhancing Sense of Presence in Watching Sports Events

Dan Mikami, Yutaka Kunita, Yutaka Kamamoto, Shinya Shimizu, Kenta Niwa, and Keisuke Kinoshita

Abstract

The number of people who experience sporting events remotely by viewing via television or the Internet far outnumber those who see the event in person. In recent years, there has been a diversification in individual preferences and viewing styles, and viewers want to enjoy the event in the way they prefer. This article introduces NTT's efforts concerning video and audio technology that enables high-sense-of-presence viewing. This work involves achieving both a high-presence remote viewing experience, which closely reproduces the actual on-site experience, and an ultrahigh sense of presence, which provides an experience that exceeds the on-site experience.

Keywords: High sense of presence, sports viewing, video and audio

1. Introduction

The dictionary defines *sense of presence* as, "the feeling that one is actually in a distant place." However, is that all that is needed to enjoy high-sense-of-presence sporting events? Sense of presence has two main aspects. One is the feeling that you are actually at that location, which we call *high sense of presence*. The other is the feeling that you see or know more than you would if you were actually at the location, which we call *ultrahigh sense of presence*. Sports events require both of these aspects. Many people want to have an experience in their own homes that is like being in the stands or even on the field of the event. There are also many viewers that want an even higher sense of presence that includes video that cannot be seen from the stands or conventional television (TV) such as video from the athlete's point of view, and sounds that cannot usually be heard such as talking among the players, which can provide an even richer experience than is available from the spectator's seat. This article describes elemental technology that NTT is working on to implement ultrahigh–sense-of-presence viewing.

2. Interactive distribution technology for omnidirectional video

In recent years, inexpensive head-mounted displays (HMDs) and cameras that can capture images that have a field of view of close to 360° have appeared on the market. Such devices have stimulated wider active interest in virtual reality viewing, which has previously been limited to some specialists and enthusiasts. NTT Media Intelligence Laboratories has been moving forward with research on interactive panorama distribution technology [1] that enables users to view in any direction they prefer. A specific application of this technology is interactive distribution technology for omnidirectional video viewing, which separates omnidirectional (360-degree) video into a number of regions and feeds the data to high-quality encoders. Then, high-quality video is selectively distributed according to the direction in which the user is looking (**Fig. 1**). Being able to view high-quality video for only the direction of viewing makes

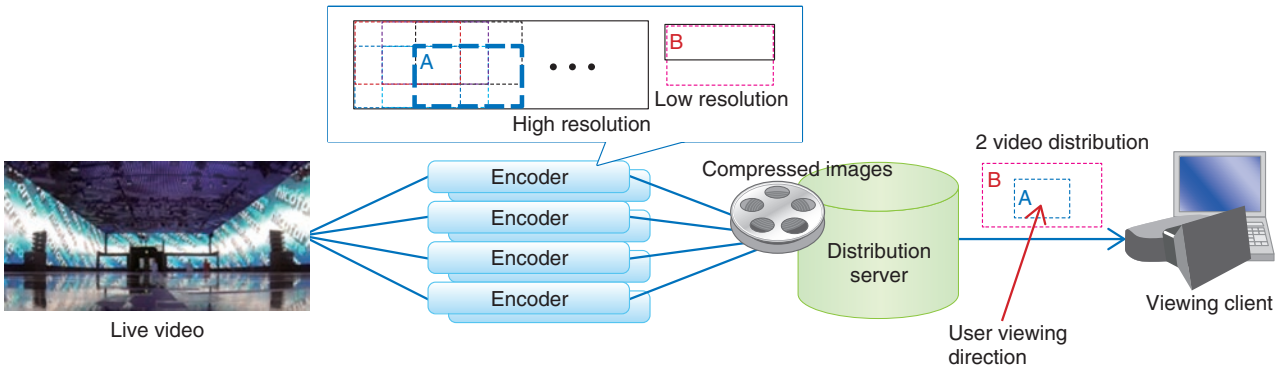


Fig. 1. Overall configuration of interactive distribution system for omnidirectional video.

it possible to deliver the video with less bandwidth than would be needed to deliver the omnidirectional video at high quality.

When selective distribution technology that is based on interactive panorama distribution technology is applied in omnidirectional distribution, the user's viewing experience varies greatly in terms of the features listed below.

(1) Wide field of view: Video that covers the entire field of view can give the user the sensation of being in a space (immersion). The visual field of humans has higher spatial resolution closer to the center and low spatial resolution at the periphery. That feature can be used to provide a high sense of presence with a limited bandwidth by transmitting and displaying video with high resolution only in the central area of the field of view.

(2) Head tracking: The HMD has sensors for acceleration and position that can be used to detect movement of the user's head. This makes it possible to present video to both eyes according to head motion, creating the feeling of looking around in a space. In contrast to viewing with a tablet or other conventional means, the user does not need to consciously select what part to view, so the viewing experience is more intuitive.

The space that can be experienced in current games and attractions at amusement parks is mostly produced by computer graphics. The technology we describe here, however, does not use computer graphics yet has been confirmed to deliver a sense of presence using live video from music performances and other events. We expect that applying this technology to the viewing of sports events can provide users with the exciting atmosphere of being in any spectator's seat in the stadium and convey the effect of being on the playing field.

3. Lossless audio encoding

Audio compression technology such as MP3 (MPEG*-1/2 Audio Layer 3) and AAC (Advanced Audio Coding), which are used in portable audio players and digital broadcasting, is widely used to provide audio at reasonable quality under the constraints of transmission bandwidth and memory capacity. To achieve a high sense of presence, however, transmission of sound with fidelity to the original source is needed. NTT has been participating in the standardization of MPEG–4 ALS (Audio Lossless Coding) and working to expand the use of lossless audio encoding [2].

Lossless coding makes it possible to completely reproduce the original sound waveform, even with compression, so audio data can be transmitted with no degradation of sound quality and with efficient use of network resources. A video and lossless audio system that we developed jointly with NTT Network Innovation Laboratories was used in trials of high-sense-of-presence live audio distribution conducted by NTT WEST and others. In those trials, users experienced a much stronger sense of being part of the scene than with conventional distribution methods, including joining in naturally with spectator applause and cheering [3, 4]. This increase in sound quality has also influenced the broadcasting of 4K and 8K video. In a survey of opinions on ultrahigh-definition TV conducted by the Ministry of Internal Affairs and Communications (MIC) of Japan in the spring of

* Moving Picture Experts Group

2014, nearly half of the responses concerned higher sound quality, and many of those were requests for use of lossless audio coding [5]. As a result, the MIC issued a Ministerial Ordinance for the capability of using MPEG-4 ALS in 4K/8K broadcasting, which was standardized by ARIB (Association of Radio Industries and Businesses) as ARIB STD-B32 in the summer of 2014.

We can thus see that there is a demand for higher sound quality to increase the sense of presence. In response to this demand, we have also moved forward with implementation of lossless audio coding for tablet terminals and set-top boxes and conducted verification testing for efficient use of the radio frequency band. In the future, wider use of lossless audio coding can be expected to improve the sense of presence for TV broadcasting and content distribution. Compression by lossless audio coding will also enable efficient transmission of the audio data acquired by the zoom microphone technology that is described later in this article, meaning that we are approaching the day when control of reverberation according to the listening environment will allow users to enjoy high-sense-of-presence content.

4. Distribution and encoding for arbitrary viewpoint video

Arbitrary viewpoint video allows the viewing of cuts from any position or orientation, regardless of the position or orientation of the camera that captured the scene. This technology is intended to provide a sense-of-presence video experience that is not possible with conventional video technology. That higher sense of presence is achieved by providing video from locations where ordinary cameras cannot be placed, such as the line of sight of players or the ball itself in a soccer match.

Arbitrary viewpoint video is created by using multiple-viewpoint video images taken simultaneously in different locations and orientations in a scene. The number of cameras necessary for taking the videos depends on the degrees of freedom of the viewpoints and the quality of the video to be created, but generally, many cameras are needed. However, video photography using many cameras, and the storage and transmission of the large volume of resulting video data are difficult. One method of creating arbitrary viewpoint video with less video data is to use depth mapping, which represents the distance of objects from the camera. We describe here our work on arbitrary viewpoint video using depth mapping together with video taken from multiple viewpoints.

Progress in sensor technology in recent years has made it possible to obtain depth maps directly by using depth cameras or rangefinders. However, the depth maps obtained in this way have low spatial resolution and contain a lot of noise. Therefore, the quality of arbitrary viewpoint video created with this technology is not high. To solve this problem, we have developed noise reduction processing and depth map up-sampling processing that uses the correlation between video and depth maps and the consistency of depth maps between viewpoints. Furthermore, we achieved real-time composition of arbitrary viewpoint video from the multi-viewpoint video and the depth maps obtained from the sensors by implementing the processing with a GPU (graphics processing unit).

Multiple viewpoint video and depth maps are a compact representation of arbitrary viewpoint video, but the amount of data is still huge compared to ordinary video. Therefore, efficient compression technology is essential for actual distribution of arbitrary viewpoint video. We previously developed a number of techniques for encoding arbitrary viewpoint video, including the use of viewpoint synthesis prediction and palette-based prediction. Viewpoint synthesis prediction is a technique applied in the synthesis of arbitrary viewpoint video to achieve efficient prediction between points of view by synthesizing predicted images using viewpoint video and depth maps that have already been encoded. Palette-based prediction is a method for generating predicted images by using the depth map feature, the value of which varies greatly between objects in the scene but varies little within a single object. In addition to achieving highly accurate prediction, this technique can also prevent degradation of performance in the synthesis of arbitrary viewpoint video by depth map encoding. These techniques that we have developed have been adopted in the 3D-HEVC (High Efficiency Video Coding) standard, which is an extension of the most recent HEVC international standard for video encoding [6].

In addition to the technology that we have described so far, implementation of arbitrary viewpoint video requires a lot of technology for elements ranging from imaging to the display and the user interface. With the current arbitrary viewpoint video synthesis technology using depth maps, the degree of freedom in moving the viewpoint and the quality of the synthesized image are limited. In the future, we plan to continue developing technology for realizing arbitrary viewpoint video that provides a video experience

Fig. 2. Zoom microphone system.

for larger scene spaces such as moving down into the playing field at sporting events.

5. Zoom microphone system

To improve the experience of sports events viewed via broadcasting or via telecommunications methods such as the Internet, we are developing technology for generating video that gives the viewer the feeling of being on the playing field. The zoom microphone system makes it possible to pick up distant sound sources clearly. It would be an elemental audio technology that is required for producing such video.

Our research started with a simple question: If a camera can zoom in on a target object, why can't we *zoom in* on distant sound sources in order to pick them up clearly? If distant sound sources were clearly recorded, it would be possible to provide audio that gives the viewer an impression of being on the playing field in the future.

The zoom microphone system consists of the two technologies shown in **Fig. 2**.

(1) Microphone array design for segregating sound sources from each other

We previously proposed a basic principle concerning how spatial signals should be captured with multiple microphones used to separate sound sources [7]. We defined the mutual information between sound sources and multiple microphones based on the information theory. To maximize that information, we placed microphones at optimum positions in front of parabolic reflectors. The implemented system shown in Fig. 2 includes 96 microphones and 12 parabolic reflectors.

(2) Noise suppression processing with less output degradation

We developed a signal processing algorithm for segregating sound sources arriving from target beamspace from other noise. With our algorithm, the output noise level would be reduced while maintaining the output signal quality. By utilizing phase/amplitude differences between microphones, a spectral filter that reduces the noise output power by as much as a factor of 1/10,000 can be generated [8]. So far, we have established principles for clearly picking up sound sources and have confirmed accurate estimation of sound sources at arbitrary locations 20 m away.

In the future, we will investigate the performance on an actual field and make technical improvements to reduce the number of microphones needed.

6. Reverberation removal and control

Hearing the cheers of the audience is an important element of sense of presence in viewing sporting events. Although being surrounded by cheering can greatly add to the sense of presence, suppression of that cheering may allow viewing that is more analytical. NTT has been working on reverberation removal and control technology, which plays an important role in controlling the sense of presence. The major application of this technology is in sound recording at concerts, so we describe it here in that context.

Recordings of dynamic and memorable performances and music from the past are available all around us in the form of compact discs (CDs), records, and other formats. We might wonder whether a sense of presence such as that obtained when hearing the music in the acoustic field of the site where it was recorded could be restored if those recordings could be played back in stereo, but that is not

necessarily the case. The reason is that it is difficult to reproduce the acoustic environment that existed at the time the performance was recorded when the recording is played back.

When we listen to music from a seat in a concert hall, two types of sound arrive at our ears from the stage. One is direct sound, which comes straight to our ears from the stage in front of us, and the other is reverberation, which comes to our ears indirectly as reflections from the walls and ceiling in four directions. Recordings on media such as CDs generally record a mixture of direct sound and reverberation that is picked up at a location near the audience seats. Thus, ordinary stereo recordings cannot reproduce the original acoustic environment that existed at the time of the recording.

We have developed the world's first technology for separating direct sound from the reverberation component in an audio signal, a technique we call *reverberation control*. We can enhance the sense of presence by applying this technique to separate the music signal into direct sound and reverberation components, and then create an acoustic environment that is similar to the acoustic environment at the time of recording by playing the direct sound component through the front speakers of a surround-sound system and playing the reverberation component through the front and rear/surround speakers [9]. Application of this technique to the past work of famous international artists and to consumer audio products has been well received thus far. In the future, we will continue our basic research with the objectives of applying it to broadcasting and achieving more accurate reverberation control processing.

7. Future development

We have described here some elemental technologies for achieving a high sense of presence in the viewing of sports events. Achieving high-sense-of-presence viewing of sports is a complex task that involves a variety of audio and video elements, including recording, encoding, distribution, processing, and the viewing system. In the future, we hope to continue creating viewing experiences that provide an even higher sense of presence in remote locations and at meeting sites by further integrating elements from the wide range of research done by NTT laboratories.

References

[1] Y. Tanaka and D. Ochi, "Interactive Distribution Technologies for 4K Live Video," NTT Technical Review, Vol. 12, No. 5, 2014. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201405fa5.html

[2] T. Moriya, N. Harada, and Y. Kamamoto, "Scope of Research on High-quality Audio Signal Processing and Coding,"
Y. Kamamoto, T. Moriya, N. Harada, and C. Kos, "Enhancement of MPEG-4 ALS Lossless Audio Coding,"
N. Harada, T. Moriya, and Y. Kamamoto, "MPEG-4 ALS: Performance, Applications, and Related Standardization Activities,"
in Selected Papers: Research Activities in Laboratories of New Fellows Part I, NTT Technical Review, Vol. 5, No. 12, 2007. https://www.ntt-review.jp/archive/2007/200712.html

[3] H. Yamane, A. Yamashita, K. Kamatani, M. Morisaki, T. Mitsunari, and A. Omoto, "High-presence Audio Live Distribution Trial," NTT Technical Review, Vol. 9, No. 10, 2011. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201110fa4.html

[4] Y. Kamamoto, N. Harada, T. Moriya, S. Kim, T. Yamaguchi, M. Ogawara, and T. Fujii, "Multichannel Audio Transmission over IP Network by MPEG-4 ALS and Audio Rate Oriented Adaptive Bit-rate Video Codec," NTT Technical Review, Vol. 11, No. 7, 2013. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201307ra1.html

[5] MIC, Results of Appeal for Opinions on Draft Report of Broadcasting System Subcommittee (Technical requirements for ultra-high-definition television broadcasting system), Mar. 2014 (in Japanese). http://www.soumu.go.jp/main_content/000283104.pdf

[6] S. Shimizu, "Recent Progress of 3D Video Coding Standardization in JCT-3V," Journal of the Institute of Image Information and Television Engineers, Vol. 67, No. 7, pp. 557–561, 2013.

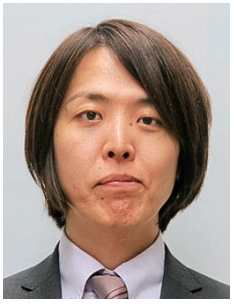
[7] K. Niwa, Y. Hioka, K. Furuya, and Y. Haneda, "Diffused Sensing for Sharp Directive Beamforming," IEEE Transactions on Audio, Speech and Language Processing, Vol. 21, No. 11, pp. 2346–2355, 2013.

[8] K. Niwa, Y. Hioka, and K. Kobayashi, "Post-filter Design for Speech Enhancement in Various Noisy Environments," Proc. of IWAENC 2014 (14th International Workshop on Acoustic Signal Enhancement), pp. 35–39, Juan-les-Pins, France, Sept. 2014.

[9] K. Kinoshita, "Enhancing Speech Quality and Music Experience with Reverberation Control Technology," NTT Technical Review, Vol. 12, No. 11, 2014. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201411fa3_s.html

Dan Mikami
Senior Research Engineer, Visual Media Project, NTT Media Intelligence Laboratories.

He received the B.E and M.E in engineering from Keio University, Kanagawa, in 2000 and 2002, respectively, and the Ph.D. in engineering from University of Tsukuba, Ibaraki, in 2012. He joined NTT in 2002. His current research activities are mainly focused on multimedia content handling. He was awarded the Meeting on Image Recognition and Understanding 2009 Excellent Paper Award, the IEICE Best Paper Award 2010, the Institute of Electronics, Information and Communication Engineers (IEICE) KIYASU-Zen'iti Award 2010, and the IPSJ SIG (Special Interest Group)-CDS (Consumer Devices & Systems) Excellent Paper Award 2013. He is a member of IEICE, the Information Processing Society of Japan (IPSJ), and the Institute of Electrical and Electronics Engineers (IEEE).



Yutaka Kunita
Senior Research Engineer, Visual Media Project, NTT Media Intelligence Laboratories.

He received the B.E., M.E., and Ph.D. in mathematical engineering and information physics from the University of Tokyo in 1996, 1998, and 2001, respectively. He joined NTT Cyber Space Laboratories in 2001 (now, NTT Media Intelligence Laboratories) and has been studying high-reality visual communication. He is a member of the Association for Computing Machinery, IEICE, and the Virtual Reality Society of Japan.

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

He received the B.S. in applied physics and physico-informatics from Keio University, Kanagawa, in 2003 and the M.S. and Ph.D. in information physics and computing from the University of Tokyo in 2005 and 2012, respectively. Since joining NTT Communication Science Laboratories in 2005, he has been studying signal processing and information theory, particularly lossless coding of time-domain signals. He was also with NTT Network Innovation Laboratories, where he developed the audio-visual codec for ODS (other digital stuff / online digital sources) from 2009 to 2011. He has contributed to the standardization of coding schemes for MPEG-4 ALS, ITU-T (International Telecommunication Union, Telecommunication Standardization Sector) Recommendation G.711.0, and 3GPP (3rd Generation Partnership Project) Enhanced Voice Services (EVS). He received the Telecom System Student Award from the Telecommunications Advancement Foundation (TAF) in 2006, the IPSJ Best Paper Award from IPSJ in 2006, the Telecom System Encouragement Award from TAF in 2007, and the Awaya Prize Young Researcher's Award from the Acoustical Society of Japan (ASJ) in 2011. He is a member of IPSJ, ASJ, IEICE, and IEEE.

Shinya Shimizu
Research Engineer, Visual Media Project, NTT Media Intelligence Laboratories.

He received the B.Eng. and M.Info. in social informatics from Kyoto University and the Ph.D. in electrical engineering from Nagoya University, Aichi, in 2002, 2004, and 2012, respectively. He joined NTT in 2004 and has been involved in research and development of 3D video coding algorithms and standardization in MPEG and ITU-T. His research interests include signal processing for free viewpoint video, light fields, computer vision, and computational photography. He is a member of IEICE.

Kenta Niwa
Research Engineer, Visual Media Project, NTT Media Intelligence Laboratories.

He received the B.E., M.E., and Ph.D. in information science from Nagoya University, Aichi, in 2006, 2008, and 2014, respectively. Since joining NTT in 2008, he has been engaged in research on microphone array signal processing. He was awarded the Awaya Prize from ASJ in 2010. He is a member of IEEE, ASJ, and IEICE.

Keisuke Kinoshita
Senior Research Engineer, Signal Processing Research Group, Media Information Laboratory, NTT Communication Science Laboratories.

He received the M.Eng. and Ph.D. from Sophia University, Tokyo, in 2003 and 2010, respectively. Since joining NTT in 2003, he has been researching various types of speech, audio, and music signal processing, including speech enhancement such as 1ch/multi-channel blind dereverberation, noise reduction, distributed microphone array processing, and robust speech recognition. He received the 2006 IEICE Paper Award, the 2009 ASJ Outstanding Technical Development Prize, the 2011 ASJ Awaya Prize, and the 2012 Japan Audio Society Award.

R!SE Initiative Activities of the United Nations

Osamu Takino, Yasuo Ishigure, and Yuji Maeda

Abstract

Natural disasters result in serious human and economic damage and are a serious threat to continuing global growth. To accelerate disaster prevention and disaster risk reduction and to make knowledge dissemination more widespread, the United Nations has initiated cooperative projects between the public and private sectors. NTT is also involved with this initiative. The ongoing activities of the initiative are described here.

Keywords: R!SE initiative, disaster risk reduction, UNISDR,

1. Overview of United Nations' efforts in disaster risk reduction

In recent years, large-scale natural disasters have occurred worldwide due in part to the effect of climate change and geological phenomena, and they have wrought serious human and economic damage (**Table 1**).

The United Nations International Strategy for Disaster Reduction (UNISDR) was established in 2000 to assist countries and communities to become more resilient against disasters by implementing disaster risk reduction strategies, and its efforts are continuing. UNISDR sponsored the World Conference on Disaster Risk Reduction (WCDRR) held in January 2005 in Kobe, Japan. The *Hyogo Framework for Action (HFA) 2005–2015: Building the Resilience of Nations and Communities against Disasters* was adopted there. To make HFA possible, UNISDR has been supporting the building of cooperative relationships between member countries and advising and aiding in the planning of disaster risk reduction strategies for each country.

Conventionally, disaster prevention planning and execution was largely led by the public sector of a country and local government, and the effects of their actions were limited to the respective local communities. However, globalization means that businesses are connected through many types of worldwide supply chains. As a result, a distant natural disaster can result in unexpected economic damage to businesses [1]. For example, when the Thai flood crisis occurred in 2011, many manufacturers far from Thailand suffered because of the unavailability of components manufactured in Thailand.

This article introduces the R!SE (Disaster Risk-Sensitive Investment) initiative, a strategy launched by UNISDR from a global perspective to reduce the risks of natural disasters. It is a cooperative organization between the public and private sectors. We also report on NTT's contributions to the R!SE initiative.

2. Background of R!SE initiative

UNISDR launched the R!SE initiative based on the two following objectives [2].

(1) A natural disaster has the potential to economically damage many far away stakeholders because of the structure of worldwide supply chains. Therefore, such disasters are not exclusively regional problems. However, it may be possible to reduce the damage from natural disasters if we treat them as a common risk, and if the private sector in a foreign country and the public sector of a local community aggressively invest in disaster prevention.

(2) It is important to share the large amount of

Table 1. Large-scale disasters and resulting damage.

Date of occurrence	Countries most affected (Main area)	Disaster name	Human toll	Economic toll
December 2004	Indonesia (Sumatra offshore)	2004 Indian Ocean earthquake and tsunami	Death toll: 230,000 (estimated) Total victims: 2 million*1	US$6.8 billion*1
August 2005	United States (Southeastern area)	Hurricane Katrina	Death toll: 1336 (as of Dec. 20, 2005)*2	US$96 billion*2
March 2011	Japan (Tohoku area)	The Great East Japan Earthquake and Tsunami	Death toll: 15,890 Missing: 2590 (police report on Feb. 10, 2015)*3	16.9 trillion yen*4 (approx. US$210 billion at Jun. 2011 exchange rate)
July 2011	Thailand (Mekong and Chao Phraya river basins)	2011 Thailand floods	Death toll: 800 (estimated)*5	1.44 trillion Thai baht*5 (approx. US$45.7 billion as of Dec . 2011)
November 2013	Philippines (Central region)	Typhoon Haiyan (Typhoon Yolanda)	Death toll: 6201 Missing: 1785 Total victims: 16 million*6	39.8 billion pesos*6 (approx. US$2.86 billion at 2013 exchange rate)

Sources:
*1 Cabinet Office, Government of Japan, "White Paper on Disaster 2006," Jun. 2006 (in Japanese). http://www.bousai.go.jp/kaigirep/hakusho/h18/BOUSAI_2006/html/honmon/hm01040103.htm
*2 A. Kato, "Physical Damage Caused by Hurricane Katrina," Disaster Prevention Research Group, National Research Institute for Earth Science and Disaster Prevention, Japan, Mar. 2006. http://dil-opac.bosai.go.jp/publication/nied_natural_disaster/pdf/41/41-03.pdf
*3 National Police Agency of Japan, Emergency Disaster Countermeasure Headquarters, "Damage Situation and Police Countermeasures associated with 2011 Tohoku District," Feb. 2015. http://www.npa.go.jp/archive/keibi/biki/higaijokyo_e.pdf
*4 Cabinet Office, Government of Japan, "Damage Estimation of the Great East Japan Earthquake," Jun. 2011 (in Japanese). http://www.bousai.go.jp/2011daishinsai/pdf/110624-1kisya.pdf
*5 Ministry of Foreign Affairs of Japan, "Japan's ODA White Paper 2012," Mar. 2013 . http://www.mofa.go.jp/policy/oda/white/2012/html/honbun/b2/s2_2.html
*6 Ministry of Land, Infrastructure, Transport and Tourism, "Damage of Typhoon Haiyan in Philippines," Mar. 2014 (in Japanese). http://www.mlit.go.jp/river/shinngikai_blog/shaseishin/kasenbunkakai/shouiinkai/r-jigyouhyouka/dai04kai/siryou6.pdf

disaster prevention knowledge gained in the private sector with many other enterprises and with countries throughout the world. As experience and knowledge of disaster prevention investment is increasingly shared, the spread of disaster prevention and disaster risk reduction worldwide may accelerate.

Therefore, the R!SE initiative was begun as a way for the public and private sectors to cooperate. Public and private sector organizations such as insurance companies, investment-related organizations, and educational institutions work together to accelerate the investment in disaster prevention measures [2] (**Fig. 1**).

Fig. 1. Basic alliance of R!SE initiative.

Activities to increase the disaster resilience of the global economy are underway and are aimed at incorporating disaster prevention measures into a regional development investment strategy in order to maintain continuous global development. The activities are also aimed at turning communities into disaster-resilient communities, which will attract investment in disaster prevention and disaster risk reduction.

A launch event of the R!SE initiative was held by UNISDR in May 2014 [3]. NTT—the only Japanese company—received a request from the UN to participate in the R!SE initiative launch ceremony, and Yasuyoshi Katayama, Senior Executive Vice

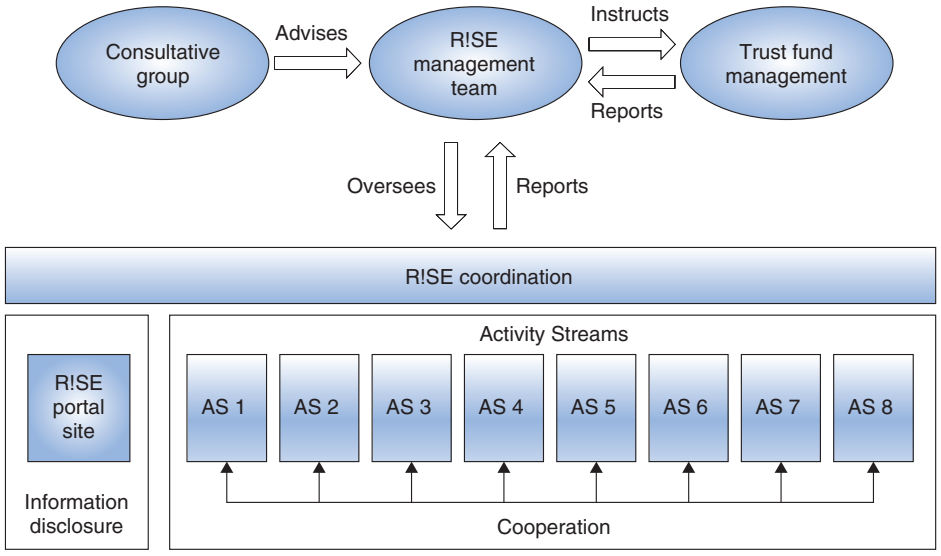


Fig. 2. Organization of R!SE initiative.

Table 2. AS structure of R!SE initiative.

AS	Activity	AS leader
AS 1	DRM strategies	PwC
AS 2	Risk metrics	EIU
AS 3	DRM industry standards	PwC
AS 4	DRM higher education	FIU
AS 5	Responsible investing	PRI
AS 6	Resilient cities	AECOM
AS 7	Insuring resilience	Willis
AS 8		UN

DRM: disaster risk management

President of NTT (at that time) attended the ceremony and approved the main ideas of UNISDR regarding disaster risk reduction. He also said he would like to utilize NTT's experience in dealing with the aftermath of the Great East Japan Earthquake in future R!SE initiative activities [4].

3. Organization and activities of the R!SE initiative

3.1 Organization

The concept of the R!SE initiative is based on disaster risk-sensitive investment [2]. This means that a safer world is possible with widespread investment activities that are intended to have a disaster prevention effect. The initiative concurrently promotes multilateral activities such as the development of a common index to evaluate business corporations and the disaster risk reduction solutions adopted by them, and a common index to estimate the current state of disaster prevention measures in a city.

UNISDR determines the entire directionality of the R!SE initiative by communicating with the R!SE management team at the UN Headquarters. Eight Activity Streams (ASs) have been created. Through development of a common index, higher educational programs and cooperation efforts between the public and private sectors have begun. The R!SE organization [2] and AS structure [2] are presented in **Fig. 2** and **Table 2**.

3.2 Eight Activity Streams

(1) AS 1: Disaster risk management (DRM) strategies

This AS is led by PricewaterhouseCoopers (PwC). Through this activity, standard tools to evaluate the DRM status of a company will be developed in order to gather knowledge on disaster prevention while considering the intellectual property and reputation of the donating company, then to share this knowledge over the portal site. Many businesses will be able to refer to good practices of disaster risk reduction and to advance their own DRM.

(2) AS 2: Risk metrics

This activity is aimed at developing a method for assuming the degree of risk to business in the event of a natural disaster and to predict the effect such a disaster would have on the national economy and the gross domestic product (GDP). This AS is led by The Economist Intelligence Unit (EIU) and is aimed at developing a risk model for every national disaster from their experience in developing a risk rating method for emerging companies. The developed method will make it easier to calculate what effect the occurrence of a disaster would have on business and on a country's GDP.

(3) AS 3: DRM industry standards

This AS is led by PwC. It will help the respective industrial sectors to increase their disaster risk resilience, and create an industrial standard to enable cooperation for quick recovery from natural disasters.

(4) AS 4: DRM higher education

This AS is aimed at advancing education programs that promote disaster risk reduction efforts. This AS is led by Florida International University (FIU), and is intended to improve the curricula of business risk management and the development of education programs for academic institutions, training centers, and business associations for small- and medium-sized businesses.

(5) AS 5: Responsible investing

This AS is aimed at assisting owners and investment managers to consider disaster risk in their investment portfolios. It is led by Principles for Responsible Investment (PRI) [5], and will help guide global investment behavior to consider disaster risk reduction.

(6) AS 6: Resilient cities

This AS is led by AECOM to increase the disaster resilience of cities. This AS will involve inspecting and improving the evaluation tool called Disaster Resilience Scorecard for Cities, which is designed for use by local governments to evaluate their disaster resilience situation. This tool includes many factors such as the degree of cooperation between the public sector, communities, and companies in the area. It is aimed at ensuring that the current state of disaster resilience of a city can be understood. This tool will help local governments consider a detailed strategy for disaster risk reduction investment through the results of self-evaluation.

(7) AS 7: Insuring resilience

This AS reconsiders disaster insurance from the viewpoint of disaster risk reduction. This AS is led by Willis and is aimed at providing insurance, especially in developing countries, in order to promote sustainable development.

(8) AS 8: DRM in United Nations

This AS is led by the UN and is aimed at increasing the disaster resilience of the UN against disasters and climate change, and supporting UN staff to promote education on disaster risk resilience. This AS is also aimed at providing education and training to UN staff while advancing cooperation between UN organizations and cooperation among the R!SE initiative members.

4. Mid- and long-term targets of the R!SE initiative and NTT's future activities

Each AS activity of the R!SE initiative has set a target for 2020, as indicated in **Table 3**. The next milestone will be to report their progress at WCDRR, which will be held in Sendai, Japan, in March 2015.

NTT is planning to contribute to the R!SE initiative through cooperation with each participating company of AS 1, 3, and 6, using its knowledge and experience in the area of disaster reduction. In AS 1 (DRM strategies), NTT is planning to share a case of disaster risk management through disaster prevention and disaster risk reduction examples that NTT has obtained. In AS 3 (DRM industry standards), NTT is planning to contribute to establish standards that should be decided and adopted by businesses to increase disaster prevention and disaster resilience in a particular area through its experience as a communication carrier and an ICT (information and communication

Table 3. Achievement target of each AS by 2020.

AS	Achievement targets by 2020
AS 1	AS 1 will have helped 100 global businesses make decisions on disaster prevention investment.
AS 2	Revised metrics will have been developed to quantify disaster risk for economic policy and corporate planning and will have been tested in at least 20 country contexts.*
AS 3	AS 3 will have helped 10 industry sectors adopt DRM standards and certification.
AS 4	In three global regions, at least 10 leading business schools will offer new and/or enhanced DRM courses or modules in their MBA-type programs, and at least 10 training or business education providers will offer cutting-edge DRM content, methodologies, and ready-to-use tools and/or software to business leaders in large, medium, and small enterprises.*
AS 5	One thousand asset owners and investment managers will support principles for disaster risk-sensitive investments as part of their ongoing commitment to responsible investment.*
AS 6	AS 6 will have helped in evaluating 50 cities worldwide using the Disaster Resilience Scorecard for Cities.
AS 7	At least 50 businesses will have benefited from access to improved insurance capacity, and 200 (re-) insurers will have revised their pricing of risk based on up-to-date risk data.*
AS 8	New risk information, risk management tools, and good practices from the private sector will have been shared, on demand, with organizations of the UN system.*

*Obtained from reference [2]

technology) partner. In AS 6 (Resilient cities), NTT is planning to use its knowledge in activities regarding the Disaster Resilience Score Card for Cities.

NTT will take part in disaster prevention in the R!SE initiative and standardization activities for disaster reduction in Japan and throughout the globe.

References

[1] Global Assessment Report on Disaster Risk Reduction 2013, http://www.preventionweb.net/english/hyogo/gar/2013/en/home/download.html

[2] The R!SE Initiative program summary, http://www.preventionweb.net/rise/sites/default/files/R!SE%20Program%20Summary%20V2.pdf

[3] Official Launch Event of the R!SE Initiative, http://www.preventionweb.net/rise/news/official-launch-event-rse-initiative

[4] NTT Public Relations Twitter feed on May 21, 2014 (in Japanese). https://twitter.com/nttpr/status/469263220788985856

[5] Financial Services Agency, "Overview of International Corporate Governance Network and Principle for Responsible Investment," Nov. 27, 2013 (in Japanese). http://www.fsa.go.jp/singi/stewardship/siryou/20131127/07.pdf

Osamu Takino
Senior Research Engineer, Security Risk Management Project, Emergency Management Operations Group, NTT Secure Platform Laboratories.

He received the B.E. and M.E. in computer control mechanical engineering from Osaka University in 1991 and 1993, respectively. He joined NTT Software Laboratories in 1993 and studied workflow management systems, office systems, location management systems, and smartcards. He also participated in a global multimedia experiment project in East Asia. He served as a Technical Committee member of the Global Platform and as a member of the team compiling IC card standards in the Digital Apartment project for the Chinese Ministry of Construction. After a transfer to NTT Communications, he worked in the areas of security solutions, application service development, and voice service development. He is currently working on a project to develop ICT systems that support disaster prediction. He received the Scholarship Encouragement Award from the Institute of Electronics, Information and Communication Engineers (IEICE) in 1996. He is a member of IEICE. He has been certified by $(ISC)^2$ as a Certified Information Security Systems Professional.

Yasuo Ishigure
Senior Research Engineer, Security Risk Management Project, Emergency Management Operations Group, NTT Secure Platform Laboratories.

He received the B.E., M.E., and Ph.D. in electronic and information engineering from National Toyohashi University of Technology, Aichi, in 1995, 1997, and 2000, respectively. He joined NTT in 2000 and studied ultrahigh-realistic communication technologies including super-high-resolution (4K) and 3D image display systems and their human factors. During 2005–2010, he was Deputy Manager of the Life Science Business Promotion Office, Business Innovation Sector, and then Deputy Senior Analyst at the Research Institute for System Science, Research and Development Headquarters, at NTT DATA. Since 2010, he has been studying technologies related to e-Health services including electronic health records (EHRs) and personal health records (PHRs) at NTT Secure Platform Laboratories.

Yuji Maeda
Vice President, Senior Manager, Security Risk Management Project, NTT Secure Platform Laboratories.

He received the Ph.D. in systems information science from Future University Hakodate, Hokkaido, in 2013. He joined NTT Telecommunication Networks Laboratories in 1991. He is currently supervising a development project on medical and healthcare information systems and disaster-resilient information systems. He received the Scholarship Encouragement Award from IEICE in 1998. He is a senior member of IEICE, and a member of IEEE (Institute of Electrical and Electronics Engineers) and the Japanese Telemedicine and Telecare Association.

Inspection Method for Near-ground Corrosion in Steel Poles for More Efficient Facility Management

Abstract

This article introduces a method of inspecting corrosion that occurs near the ground in steel poles to achieve more efficient facility management. This is the twenty-eighth of a bimonthly series on the theme of practical field information on telecommunication technologies. This month's contribution is from the Materials Engineering Group, Technical Assistance and Support Center, Maintenance and Service Operations Department, Network Business Headquarters, NTT EAST.

Keywords: steel pole, corrosion, inspection method

1. Introduction

Corrosion that occurs on a steel pole (SP) near the ground can have a major effect on its structural strength, so SPs must be appropriately maintained and managed through facility inspections. In the past, such inspections were performed visually to evaluate the extent of corrosion near the ground and to determine whether the facility in question was failing. However, the number of deficient facilities that have been in operation for more than 30 years and that will have to be replaced is increasing, so there is a need to rethink the inspection method so that a priority can be assigned to the replacement of SPs. To this end, the Technical Assistance and Support Center has investigated a method of performing inspections to evaluate SP strength. This inspection method makes it possible to easily and quantitatively measure the degree to which an SP member is degrading due to corrosion so that facilities that require urgent handling can be identified and targeted for replacement. This article introduces the method studied here as a new and efficient approach to inspecting existing SPs.

2. Corrosion near the ground in SPs and conventional inspection method

NTT EAST and NTT WEST manage about 4 million SPs in total, and a large number of these have been in operation for many years since their initial installation. Since deterioration owing to long-term use can be visibly apparent, the approach to managing the maintenance of SPs has been to inspect their state of deterioration and to eventually replace them, starting with those with noticeable deterioration.

The main cause of deterioration in communication facilities made of steel is corrosion, which is greatly influenced by the presence of water and oxygen. An SP is an outdoor facility embedded in the ground, so corrosion can easily progress on it in the area near the ground (the boundary between the above-ground and below-ground portions of the SP), where oxygen in the air is present and rain water collects. Checking the state of corrosion near the ground is therefore an important inspection task.

In the conventional inspection method for this type of corrosion in SPs, the inspector visually examines the extent, or range, of corrosion or checks for holes caused by corrosion by referring to sample photographs. The inspector will then place a priority level on problem resolution (such as replacement) according to the range of corrosion (**Fig. 1**).

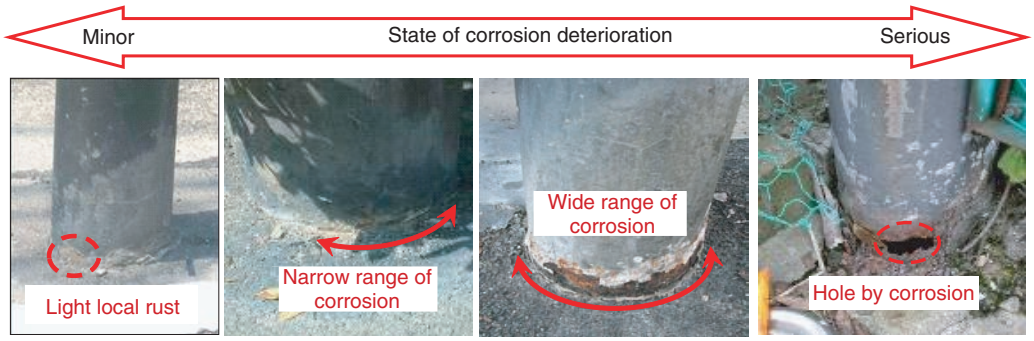


Fig. 1. Examples of corrosion near the ground in SPs.

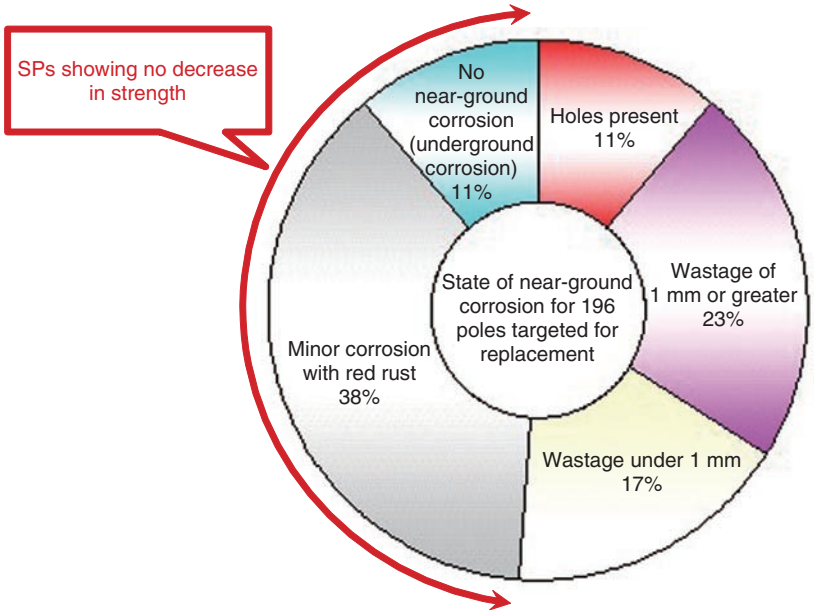


Fig. 2. Breakdown of corrosion conditions in SPs targeted for replacement.

3. Understanding the actual condition of SPs judged deficient

Inspection using the method described in section 2 had become well established to the point that it was possible to determine the corrosion conditions at the base of SPs, but at the same time, a new problem arose. Specifically, in deciding whether replacement was necessary based on inspection results, a massive number of facilities became targeted for replacement. However, mounting a prompt response was difficult because of cost and labor considerations.

In light of this situation, the cooperation of NTT branch offices and subcontractors in various areas was enlisted to survey the actual condition of deficient SPs (about 200) for which replacement was deemed necessary. This survey involved analyzing the corrosion conditions based on the progress of wastage (reduction in thickness) from corrosion, which affects the SP strength. The results of this survey revealed that almost half of the poles analyzed were affected by minor corrosion without wastage (**Fig. 2**). This type of corrosion does not contribute to a reduction in pole strength, so these SPs were classified as facilities not in need of prompt replacement. Thus, when the conventional inspection method was used, SPs showing no decrease in strength had been judged and managed as if they were equally deficient

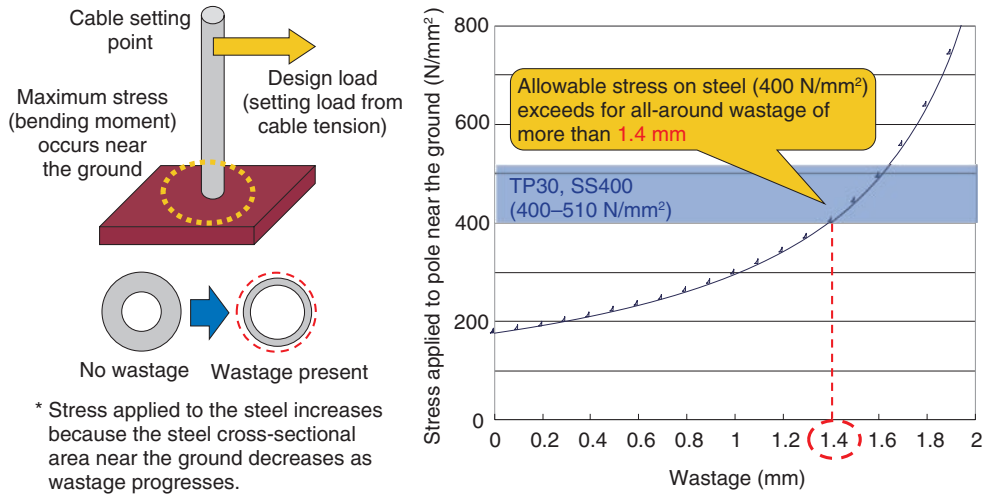


Fig. 3. Stress applied to SP near the ground with corrosion wastage.

to SPs that did show a decrease in strength.

At the Technical Assistance and Support Center, we used these findings as a basis for studying inspection methods that would assess the state of an SP according to the decrease in strength associated with corrosion wastage instead of making visual judgments based on the range of corrosion near the ground. A scheme for setting priorities in replacing SPs was also studied.

4. Study of inspection methods for quantifying near-ground corrosion wastage

The strength of an SP depends on the thickness of its steel material. In addition, the stress applied to the SP increases as the cross-sectional area of the SP decreases with the progression of corrosion wastage (**Fig. 3**). We therefore considered the need for an inspection/management method based on the thickness of the steel material that could be applied to identify those SPs whose safety could not be ensured up to the next inspection. This could be accomplished by setting the amount of wastage at which the stress applied to the near-ground part of the SP when the SP design load is acting on the overhead position of the pole exceeds 400 N/mm^2—the allowable stress on steel—as a threshold for pole replacement, and by estimating the speed of steel corrosion. Information on the steel thickness of existing SPs near the ground level is needed for this, so we studied alternative methods to visual inspection that could be used to obtain this information. In revising the inspection method in this way, we considered the importance of ensuring measurement accuracy and keeping inspection tasks simple while minimizing the costs incurred by such a revision. We tested the following two methods using commercial products.

(1) Measurement of corrosion wastage using a depth gauge
- This method involves measuring the depth at a near-ground location where the thickness has decreased because of corrosion, using as a reference an original (healthy) location having no corrosion on the surface of the SP (**Fig. 4**).

(2) Measurement of remaining thickness using an ultrasonic thickness gauge
- This method involves measuring the remaining thickness at a near-ground location where thickness has decreased based on the relationship between the propagation speed of an ultrasonic pulse and its transmit/receive time.

Test results revealed problems with measurement accuracy in both methods. When measurements are obtained using the depth gauge, the reference surface on the device needs to be set firmly at a sound location on the SP, but this was found to be difficult if that location had a spherical surface. By contrast, measurements obtained using the ultrasonic thickness gauge are based on the time taken for a pulse emitted from the transmit/receive sensor through one end of the steel material (incident surface) to be reflected back to the sensor from the other side of the steel material (reflecting surface). However, the reflected pulse will be difficult to measure if the reflecting

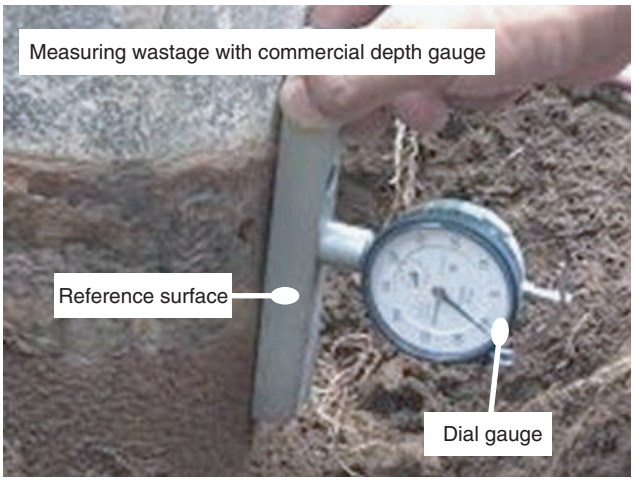


Fig. 4. Measurement of wastage using commercial depth gauge.

Table 1. Comparison of near-ground wastage measurement methods.

	Ultrasonic thickness gauge	Commercial depth gauge
Accuracy	Poor (error greater than ± 0.1 mm)	
Reasons for low accuracy (problem)	Difficulty in using ultrasonic measurement on samples with spherical or uneven surface	Difficulty in affixing reference surface of a commercial depth gauge to spherical surface of an SP
Difficulty in resolving problem	High (expensive ultrasonic measuring equipment is needed)	Low (can be improved using a simple positioning tool)

surface is spherical, and an ultrasonic pulse will be difficult to insert if the incident surface is uneven. In short, problems existed with both of these measurement methods, but considering the ease (cost) of making improvements, we selected the measurement method using a depth gauge (**Table 1**).

5. Development of SP near-ground wastage measurement tool

To improve the accuracy in measuring corrosion wastage using a depth gauge, a function is needed for reliably positioning the depth gauge at a sound location on the SP where the material has no corrosion (expected reference surface) and also at the location of corrosion wastage. We therefore investigated a method for attaching a depth gauge mount to the SP, and after fabricating such a mount, we constructed a prototype version of an SP near-ground wastage measurement tool consisting of a depth gauge and gauge mount (**Fig. 5**).

Fig. 5. SP near-ground wastage measurement tool.

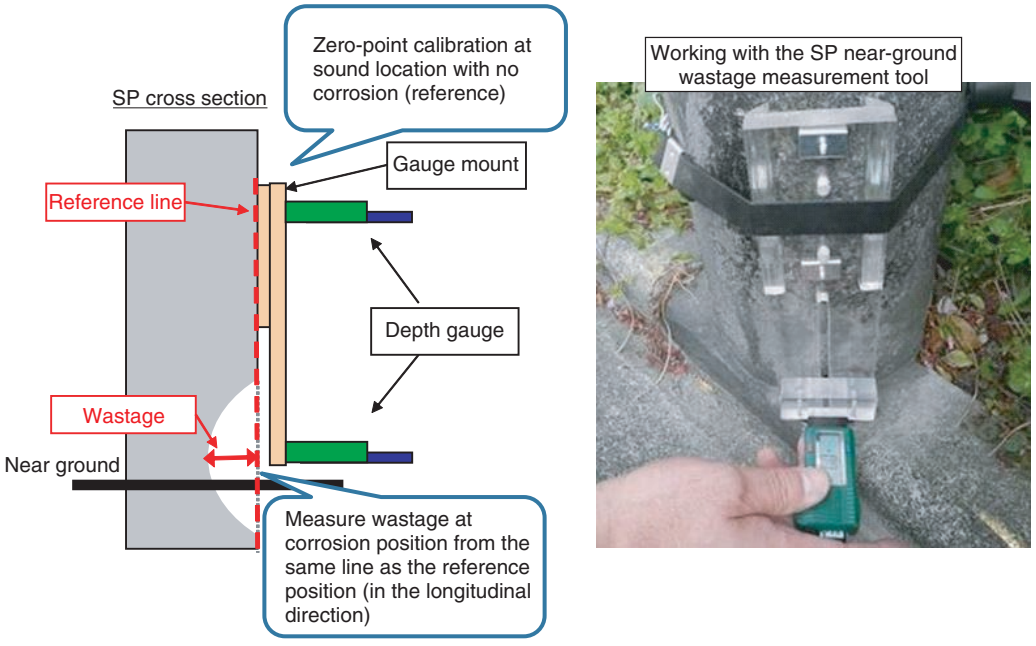


Fig. 6. Measurement of near-ground wastage using depth gauge and gauge mount.

Table 2. Comparison of conventional and new inspection methods.

	Inspection method	Features
Conventional	Visual determination of range of corrosion	• Deals equally with poles having a high risk and low risk of collapsing • Visual evaluations result in variations in inspection data.
New	Use of a tool to measure wastage	• Separates poles at high risk of collapsing from those at low risk • Quantitative evaluations using a tool result in consistency of inspection data.

This tool uses the position of the gauge mount surface as a reference to measure the distance from the pole surface at a sound location to the pole surface at a location affected by near-ground wastage, as shown in **Fig. 6**. This distance is taken to be wastage (reduction in thickness). Since an SP has a tapered structure in which its diameter decreases in the upward direction (towards the top), we gave the gauge mount a shape such that the near-ground wastage location and the sound location to be measured can be positioned along the same straight line in the SP's longitudinal direction. Furthermore, to affix the tool to the SP, we use both magnets and a fixing band. We also selected a gauge with a digital display to make measurement results easier to read in order to ease the workload on inspection personnel as much as possible. Differences between the conventional inspection method and the new inspection method are summarized in **Table 2**. A major difference here is that in the conventional method, the range of corrosion is evaluated by visual means, while in the new method, a measuring device is used to assess the reduction in thickness of the SP caused by corrosion (wastage).

6. Effect and acceptance of new inspection method

In the conventional method of assessing the range of corrosion by visual inspection, SPs at a high risk of collapsing were grouped with those not at risk, which made it difficult to identify poles in urgent need of replacement. The results of a sample survey of exist-

ing SPs using the SP near-ground wastage measurement tool revealed that nearly half of the faulty SPs that had been evaluated as needing replacement by the conventional visual inspection method did not actually need to be replaced in the period up to the next inspection.

Furthermore, inspections carried out using the SP near-ground wastage measurement tool enable us to finely classify SPs from those with a high risk of deteriorating and collapsing to those with no risk at all, which makes for efficient replacement and management of SPs. This new method has already been adopted as a standard inspection method, and the plan is to promote its use even further.

External Awards

Young Researcher's Award
Winner: Hisashi Kurasawa, NTT Network Innovation Laboratories
Date: February 17, 2015
Organization: The Institute of Electronics, Information and Communication Engineers (IEICE)

For "Evaluation of the Estimation for Missing Sensor Values Using a Field Trial."
Published as: H. Kurasawa, Y. Fujii, H. Sato, A. Yamamoto, H. Kawasaki, M. Nakamura, M. Kamiya, A. Tsutsui, and T. Miyazaki, "Evaluation of the Estimation for Missing Sensor Values Using a Field Trial," Proc. of the 2014 IEICE General Conference, B-18-48, Niigata, Japan, Mar. 2014 (in Japanese).

IEICE Information Network Research Award
Winner: Takeru Inoue, Toru Mano, Kimihiro Mizutani, NTT Network Innovation Laboratories; Shin-ichi Minato, Hokkaido University; Osamu Akashi, NTT Network Innovation Laboratories
Date: March 2, 2015
Organization: Technical Committee on Information Network, IEICE

For "Packet Classification for Global Network View of SDN with MDDs."

In software-defined networking, applications are allowed to access a global view of the network so as to provide sophisticated functionalities. They commonly rely on packet classification, but unlike the conventional classification problem of searching for the action taken at a single switch, the global network view requires them to identify the network-wide behavior of the packet. Conventional classification methods, however, fail to well support network-wide behaviors, since the search space is complicatedly partitioned.

This paper proposes a novel packet classification method that efficiently supports network-wide packet behaviors. Our method utilizes a compressed data structure named the multi-valued decision diagram, allowing it to manipulate the complex search space with several algorithms. Experiments with real network datasets show that our method identifies the packet behavior at 20.1 Mpps on a single CPU core with only 8.4 MB memory.

Published as: T. Inoue, T. Mano, K. Mizutani, S. Minato, and O. Akashi, "Packet Classification for Global Network View of SDN with MDDs," IEICE Technical Report, Vol. 114, No. 207, IN2014-50, pp. 1–6, Sept. 2014 (in Japanese).

Young Researcher's Award
Winner: Kazumitsu Sakamoto, NTT Network Innovation Laboratories
Date: March 11, 2015
Organization: IEICE

For "Reduction Effect of the Number of Quantization Bits Required in ADC by MIMO Detection before AD Conversion" and "Experimental Evaluation of Broadband Short-range MIMO Transmission Using Simple Decoding Method."

Young Researcher's Award
Winner: Maki Arai, NTT Network Innovation Laboratories
Date: March 11, 2015
Organization: IEICE

For "Study on Higher Modes of Antennas for MIMO Transmission by Orthogonal Directivities" and "Experimental Study on Higher-order-mode Antennas for Simple Spatial Division Transmission."
Published as: M. Arai, T. Seki, K. Hiraga, K. Sakamoto, H. Toshinaga, T. Nakagawa, and K. Uehara, "Study on MIMO Transmission Using Orthogonal Directivities Obtained from Higher Order Microstrip Antenna Modes," Electronics Letters, Vol. 50, No. 7, pp. 562–564, Mar. 2014.

Maejima Hisoka Award
Winner: Masahito Tomizawa, NTT Network Innovation Laboratories; Hiroshi Onaka, Fujitsu Limited; Takashi Mizuochi, Mitsubishi Electric Corporation; and Kiyoshi Fukuchi, NEC Corporation
Date: March 20, 2015
Organization: The Tsushinbunka Association

For "Research and Development on Fast Wavelength Switching Digital Coherent Technology for 100G Optical Network."

Papers Published in Technical Journals and Conference Proceedings

Unsupervised Non-parametric Bayesian Modeling of Non-stationary Noise for Model-based Noise Suppression
M. Fujimoto, Y. Kubo, and T. Nakatani

Proc. of 2014 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp. 5562–5566, Florence, Italy, May 2014.

The accurate modeling of non-stationary noise plays an important role in model-based noise suppression for noise robust speech recognition. We have already proposed methods for unsupervised noise modeling with a Gaussian mixture model or a hidden Markov model by using a minimum mean squared error estimate of the noise. However, our previous work fixed the structure of the noise model empirically without any consideration of noise characteristics; thus, optimization of the noise model structure is required if we are to obtain further improvements. Although the Bayesian information criterion (BIC) has been widely used as a conventional approach to model structure estimation, it is not always the optimal criterion. Therefore, this paper presents a way of modeling non-stationary noise with a non-parametric Bayesian approach that estimates the model structure depending on the characteristics of given observations. The proposed method provided improved results for the evaluations of two different speech recognition tasks compared with results obtained using the conventional BIC-based approach.

A Cross-layer Switching of OFDMA and MU-MIMO for Future WLAN Systems

T. Murakami, Y. Takatori, M. Mizoguchi, and F. Maehara

IEICE Communications Express, Vol. 3, No. 9, pp. 263–268, September 2014.

We propose a cross-layer switching method of orthogonal frequency division multiple access (OFDMA) and multiuser multiple input multiple output (MU-MIMO) for future wireless local area network (WLAN) systems. The proposed method, employed on the medium access control layer, switches between OFDMA and MU-MIMO as the transmission method after processing by using physical layer information such as an overhead of channel state information feedback, station number, and data length in order to improve the transmission efficiency. Simulation results show that the proposed method achieves higher total throughput than conventional OFDMA or MU-MIMO where switching is not performed.

Dual-band Nested Waveguide Antenna Employing Frequency-selective Surfaces

H. So, A. Ando, and T. Sugiyama

Proc. of 2014 IEEE International Symposium on Antennas and Propagation Society, pp. 2080–2081, Memphis, USA, July 2014.

This paper proposes a waveguide antenna that employs frequency-selective surfaces (FSSs). The proposed antenna nests one waveguide inside the other. The top surface of one waveguide acts as the top surface of the other. Waveguide walls are formed by FSSs, which reflect electromagnetic waves at each frequency band. An example implementation of this proposal, described in this paper, is a waveguide antenna designed to radiate at 2 GHz and 4 GHz. Electromagnetic field simulation results demonstrate that the fractional bandwidth is 29% (2 GHz) and 16% (4 GHz) for VSWR values under 2.0. The radiation patterns of the proposed dual-band waveguide antenna match well those of the corresponding single-band waveguide antennas, and the peak gain of each frequency band is the same.

Multiband Sector Antenna with the Same Beamwidth Employing Multiple Woodpile Metamaterial Reflectors

H. So, A. Ando, T. Seki, M. Kawashima, and T. Sugiyama

IEICE Transactions on Electronics, Vol. E97-C, No. 10, pp. 976–985, October 2014.

This paper proposes a sector base station antenna for mobile wireless communication systems employing multiple woodpile metamaterial reflectors and a multiband radiator that establishes the same beamwidth in the horizontal plane for more than two frequency bands. Electromagnetic band gap (EBG) characteristics of each metamaterial reflector can be controlled through structural parameters of the woodpile reflector, e.g., the rod width and rod spacing. A design for a triple-frequency-band antenna that radiates at 800 MHz, 2 GHz, and 4 GHz is shown as an example of the proposed antenna. The algorithm used to adjust the beamwidth of the proposed antenna is newly introduced and adjusts the beamwidth to be the same for each band using the rod width of the woodpile. A prototype of the proposed antenna has the approximately 90[°] beamwidth in the horizontal plane at the three frequencies, and the measurement results agree well with the electromagnetic field simulation results.

Spatial Division Transmission without Signal Processing for MIMO Detection Utilizing Two-ray Fading

K. Hiraga, K. Sakamoto, M. Arai, T. Seki, T. Nakagawa, and K. Uehara

IEICE Transactions on Communications, Vol. E97-B, No. 11, pp. 2491–2501, November 2014.

This paper presents a spatial division (SD) transmission method based on two-ray fading that dispenses with the high signal processing cost of multiple-input and multiple-output (MIMO) detection and antennas with narrow beamwidth. We show the optimum array geometries as functions of the transmission distance for providing a concrete array design method. Moreover, we clarify achievable channel capacity considering reflection coefficients that depend on the polarization, incident angle, and dielectric constant.

Online Gain Update for Manual Following Response Accompanied by Gaze Shift during Arm Reaching

N. Abekawa and H. Gomi

Journal of Neurophysiology, Vol. 113, No. 4, pp. 1206–1216, November 2014.

Visual motion applied during a reaching movement induces a rapid and automatic manual following response (MFR) in the direction of the visual motion. The MFR amplitude is modulated by the gaze direction relative to the reach target location (i.e. foveal or peripheral reaching). However, the time or state point at which the brain specifies this visuomotor gain remains unclear. In the present study, we measured MFR amplitudes during a task where the participant performed a saccadic eye movement that altered the gaze-reach coordination during reaching. The results indicate that the MFR amplitude immediately after the saccade termination changed according to the new gaze-reach coordination, suggesting a flexible online updating of the MFR gain during reaching. Our findings suggest that the brain flexibly updates the visuomotor gain for an online controller even during reaching movements based on continuous monitoring of the gaze-reach coordination.

Emphasized Accent Phrase Prediction from Text for Advertisement Text-To-Speech Synthesis

H. Nakajima, H. Mizuno, and S. Sakauchi

Proc. of PACLIC 28 (28th Pacific Asia Conference on Language, Information and Computation), pp. 170–177, Phuket, Thailand,

December 2014.

Realizing expressive text-to-speech synthesis requires both text processing and the rendering of natural expressive speech. This paper focuses on the former as a front-end task in the production of synthetic speech and investigates a novel method for predicting emphasized accent phrases from advertisement text information. For this purpose, we examine features that can be accurately extracted by text processing based on current text-to-speech synthesis technologies.

SFB: A Scalable Method for Handling Range Queries on Skip Graphs

R. Banno, T. Fujino, S. Takeuchi, and M. Takemoto

IEICE Communications Express, Vol. 4, No. 1, pp. 14–19, January 2015.

Skip graph is a promising candidate algorithm for large scale distributed systems. The principal feature is range query functionality, but skip graph does not have a definite method of multicasting inside ranges designated by query issuers. Even though several simple ways can be considered, they are inefficient regarding the latency or traffic volume. In this letter, we first introduce multi-range forwarding (MRF) used in multi-key skip graph. MRF can be used even in normal skip graph, and is efficient compared to the simple ways. Second, we propose a method named split-forward broadcasting (SFB). We analytically evaluate SFB and explain that SFB can roughly halve the average number of hops of MRF.

Spatiotemporal Analysis of Rambling Activities: Approach to Inferring Visitor Satisfaction

M. Ohta, Y. Watanabe, and T. Miyazaki

Proc. of ENTER2015, pp. 551–563, Lugano, Switzerland, February 2015.

A method for investigating trajectories of rambling objects is proposed. The goal of this study is to infer persons' satisfactions with their experiences using their trajectories. Two aspects of rambling activities: multi-stop and multi-purpose trips, and unplanned stopping-by at destinations are examined by mathematical knot theory.

Resolution Warped Spectral Representation for Low-delay and Low-bit-rate Audio Coder

R. Sugiura, Y. Kamamoto, N. Harada, H. Kameoka, and T. Moriya

IEEE Transactions on Audio, Speech, and Language Processing, Vol. 23, No. 2, pp. 288–299, February 2015.

We have devised a high-quality frequency-domain audio coder based on the state-of-the-art monaural wide-band coder aiming at its use in low-delay and low-bit-rate conditions. The coder efficiently represents frequency spectral envelopes of the target signals with low computational complexity using optimally prepared non-negative sparse matrices. The experimental results reveal that this representation has positive effects on the objective and subjective quality of the coder resulting in the comparable quality to the same bit rate of 3GPP Extended Adaptive Multi-Rate WideBand (AMR-WB+), a coder which permits more than four times longer delay compared with the proposed coder. Consequently, this coder is suitable for applications in mobile communications, which require low delay and low complexity.

Sub-lambda and Wavelength Path Reconfiguration in Multi-layer Transport Networks

A. Kadohata, A. Watanabe, and A. Hirano

Journal of Optical Communications and Networking, Vol. 7, No. 3, pp. A432–A439, March 2015.

We investigate adaptive reconfiguration of sub-lambda and wavelength paths to deal with unpredictable traffic demands. To accommodate traffic effectively, wavelength paths need to be reconfigured cooperatively with sub-lambda paths while minimizing the number of migration sequences. Therefore, we propose a scheme for sub-lambda and wavelength path reconfiguration based on wavelength resource management in multi-layer transport networks. Three steps comprise the proposed scheme: wavelength resource management, reconfiguration trigger, and reconfiguration. First, management of the number of wavelength paths that can be accommodated between all node pairs is calculated for each path accommodated. Second, if that number is less than the threshold value set in advance for some node pair, reconfiguration is performed. Third, the sub-lambda and wavelength paths are reconfigured. In this scheme, we also propose a sub-lambda path reconfiguration algorithm that reconfigures multi-hop sub-lambda paths to single-hop paths first, and two wavelength path reconfiguration algorithms that reduce wavelength fragmentation and the number of migration sequences considering the accommodation rate. Numerical evaluation shows that the number of fibers is suppressed by 12%–14% and that the equipment cost is reduced by approximately 8%–9% compared to when no reconfiguration is used. The results also show that the number of migration sequences for the wavelength defragmentation algorithms is reduced by approximately 23%–35% compared to that for the conventional algorithm.
