

NTT Research and Development— Leading the Way to B2B2X

Hikomichi Shinohara

*Senior Executive Vice President and Head of Research
and Development Strategy Department, NTT*



Overview

This article introduces NTT's research and development (R&D) activities for creating new value through collaboration with various partners to promote the B2B2X (business-to-business-to-X) business model under the NTT Group's medium-term management strategy, *Towards the Next Stage 2.0*. It is based on a lecture presented by Hikomichi Shinohara, NTT Senior Executive Vice President and Head of the Research and Development Strategy Department, at NTT R&D Forum 2017, which was held in February 2017.

Keywords: B2B2X, artificial intelligence, Internet of Things

1. Expansion of B2B2X model and roles of R&D

We believe that research and development (R&D) serves as an engine for driving the business-to-business-to-X (B2B2X) business model. This is because propelling innovation is the foundation of our ability to provide new value to service providers (the second B in B2B2X) and consumers and enterprises (X). On the basis of this understanding, the NTT R&D laboratories (hereinafter, NTT R&D) is pursuing co-innovation with various partners. This effort, over the last two years, has shown us that collaboration, especially with partners in unrelated industries, can generate hitherto unimagined concepts, so we are currently focusing on that type of collaboration (**Fig. 1**). While network, security, and cloud technologies are naturally important for driving B2B2X, this article focuses on artificial intelligence (AI), the Internet of Things (IoT), and media technology.

2. NTT Group's AI technology: corevo®

AI can be approached in two different ways. One is

to simulate human intelligence and thinking. The other is to supplement and draw forth human ability. NTT R&D is working on the latter, namely, AI that creates new value through coexistence and co-creation with humans.

In the spring of 2016, the NTT Group unified the group's AI-related activities under the brand name corevo®. It signifies our wish to bring about new, revolutionary developments in collaboration with a variety of players (*co-revolution*) by integrating different types of AI technology. Specifically, our R&D is focused on four categories of AI (**Fig. 2**).

Agent-AI is closest to the AI that is presently making news. It supports humans by interpreting the information that they generate. For example, Agent-AI makes it possible for a robot to conduct an intelligent conversation with humans. Ambient-AI interprets humans, objects, and the environment, and instantly forecasts and controls the immediate future. In this sense, it embraces the concept of IoT. Heart-Touching-AI interprets human emotions and physical conditions and understands our deep psyche, intellect, and instinct so that it can see things from a

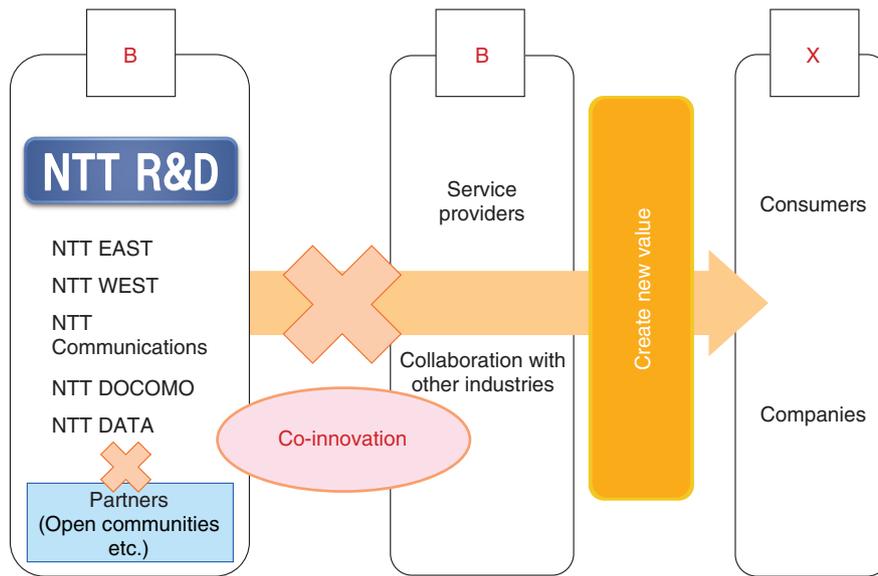


Fig. 1. Expansion of B2B2X model and roles of R&D.

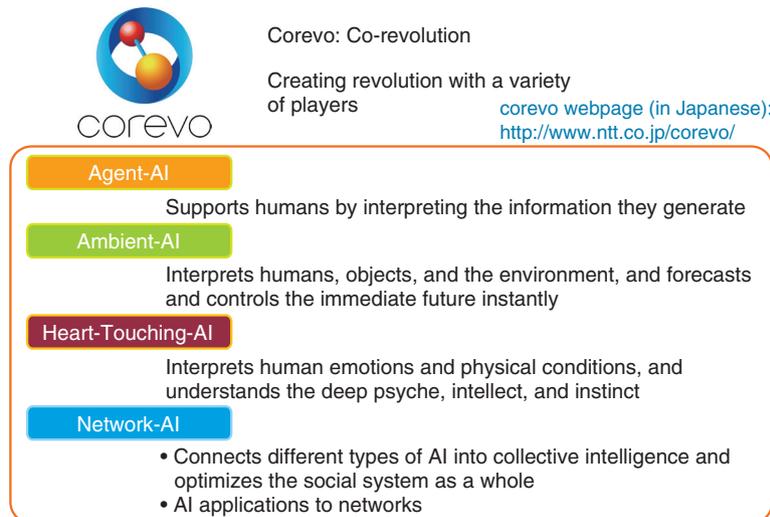


Fig. 2. NTT Group's AI Technology: corevo®.

human perspective. Network-AI embraces two concepts. One is to connect different types of AI into collective intelligence and optimize the social system as a whole. The other is to apply AI to networks in order to enhance their reliability, efficiency, and other qualities.

2.1 Agent-AI technologies

Agent-AI is founded on three types of technology:

auditory technologies; speaking technologies, which include the capacity to understand speech; and viewing technologies. Representative technologies are presented below.

(1) Auditory technologies

One auditory technology is noise suppression. This makes it possible to hear a human voice clearly, even in an environment filled with noise and cheering to a level of more than 100 dB. Speaker diarization

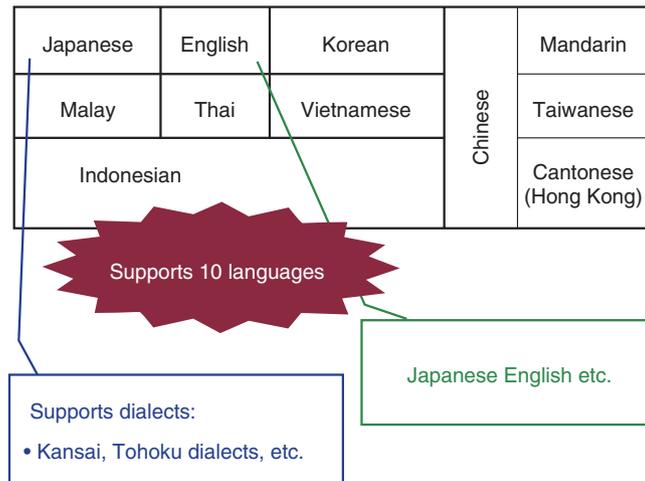


Fig. 3. Language-identified speech recognition.

technology allows the user to identify a speaker from among many people speaking simultaneously. Currently, one speaker can be identified from among up to six people talking at the same time. If this technology is used inside a vehicle, for example, the user can listen to the voice of just the driver.

In the CHiME-3 (the 3rd CHiME Speech Separation and Recognition Challenge), an international technical evaluation contest in which participating organizations competed in speech recognition in a variety of noisy environments, NTT’s distortionless speech enhancement and deep-learning speech recognition technology achieved a speech recognition rate of 94.2%, the highest in the contest. Language-identified speech recognition has been developed based on this technology. Currently, it identifies and recognizes speech in ten different languages (Fig. 3). It can even recognize the type of English typically spoken by Japanese people, and its coverage of representative Japanese dialects is being expanded.

Frequently asked question (FAQ) search is often employed when AI is used to answer questions. One challenge is that different people ask the same question using different expressions. AI must accurately understand which question in the FAQ list the questioner is referring to. NTT R&D’s utterance comprehension technology understands questions accurately, thanks to its machine learning that utilizes a text corpus containing extremely diverse expressions and a large-scale semantic dictionary that has been enriched over a long period.

We are also working on emotion recognition tech-

nology. This determines, for example, if a person is angry, satisfied, or worried from the way he or she speaks. In addition to *hot* anger characterized by shouting, this technology can identify the more subtle *cold* anger, which is characterized by the speaker sounding calm but being angry inside, an attitude said to be very common in Japanese people.

(2) Speaking technologies

We are studying technology that automatically constructs intonation information of a speech using deep learning (Fig. 4). AI needs to differentiate words that have the same sounds but different accents and meaning, such as “haSHI” (bridge) and “HAsHi” (chopsticks) in Japanese, or the noun “object” (ábdzikt) and the verb “object” (əbdzékt) in English. The aim of this technology is similar to that of the emotion recognition technology mentioned above. It is intended to help the user to communicate in a manner that is appropriate for the occasion, for example, speaking to someone in trouble in a sympathetic manner or speaking to an angry person in a soothing manner. This technology can efficiently construct prosodic information.

Free dialog technology is aimed at enabling AI not only to answer questions but also to conduct a free conversation. In March 2016, at SXSW (South by Southwest), a major business and content event held in Austin, Texas, an android built by Professor Hiroshi Ishiguro of Osaka University called “Geminoid” had a conversation with a woman, a total stranger to the android. This was made possible by incorporating our speech and dialog-related technologies into the

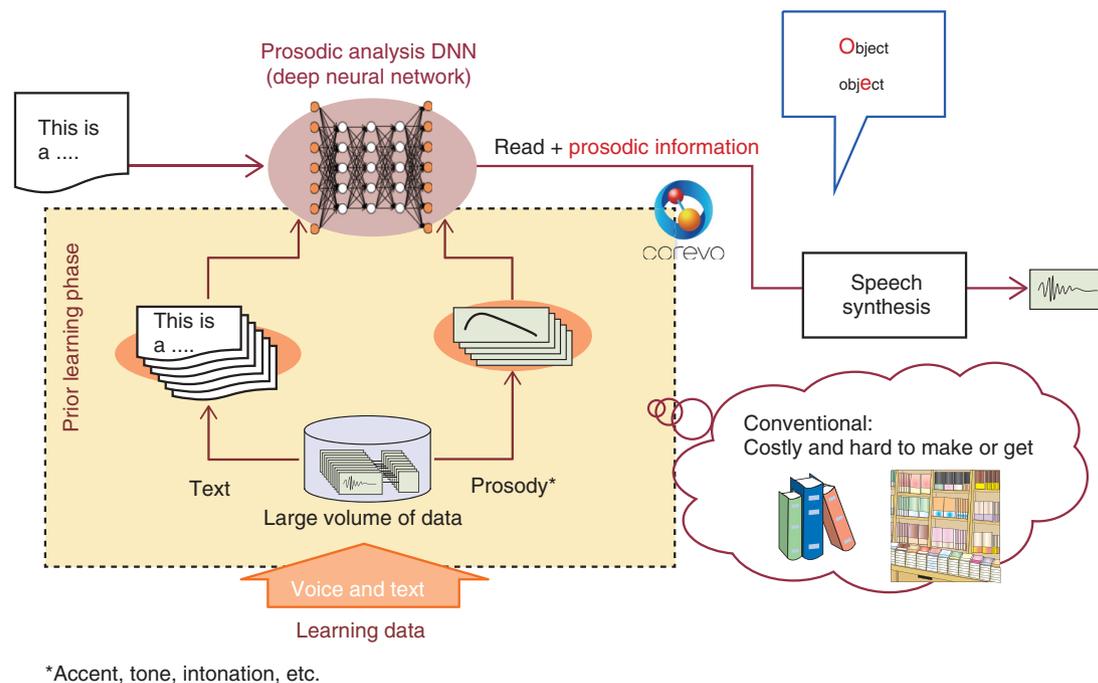


Fig.4. Automatic acquisition of prosodic information using deep learning.

android. We will seek to enable robots to perform something more challenging, such as carrying out a debate or improvising amusing dialog.

(3) Viewing technologies

We humans recognize an object by comparing what we see with what we have in our memory. Angle-free object search technology does just that. It searches images previously registered in a cloud for the object captured by a smartphone camera. Its forte is that it can recognize an object even if only one or two photos are pre-registered and even if the object is captured from an oblique angle or in close-up, or part of the object is obscured by someone standing in front of it.

We are studying video event detection technology, which searches for a video showing a scene of eating or a scene of driving a car, for example. At TRECVID 2016, a worldwide workshop hosted by the U.S. National Institute of Standards and Technology, NTT R&D won first to third places in several categories with this technology.

2.2 Application examples of Agent-AI

Using Agent-AI technologies, NTT Communications has initiated a commercial service using Communication Engine COTOHA™. This service understands natural language and answers questions from

users. If the question is vague, the engine attempts to understand what the user wants by asking specific questions for clarification. If it determines that it still cannot understand the question, it transfers the question to a human operator and learns from the way in which the operator interacts with the user.

In addition to introducing individual technologies, we are holding concept exhibits such as corevo for drivers and corevo for service desks so that people can get some idea of specific situations in which corevo technologies can be used.

2.3 Support for robots and sensors

A variety of services can be created by combining a group of some of the AI technologies mentioned above with robots and sensors. With a view to enabling customers to create their own services using our AI technologies, rather than leaving service creation to professionals, we have developed R-env®, a cloud-based human-machine interaction control. It provides a mechanism whereby the user can quickly construct a program using a web browser. NTT EAST, NTT WEST, NTT Communications, and NTT DOCOMO are conducting joint field trials on potential applications for medical care or invigoration of local economies. The trials have already yielded some practical services.

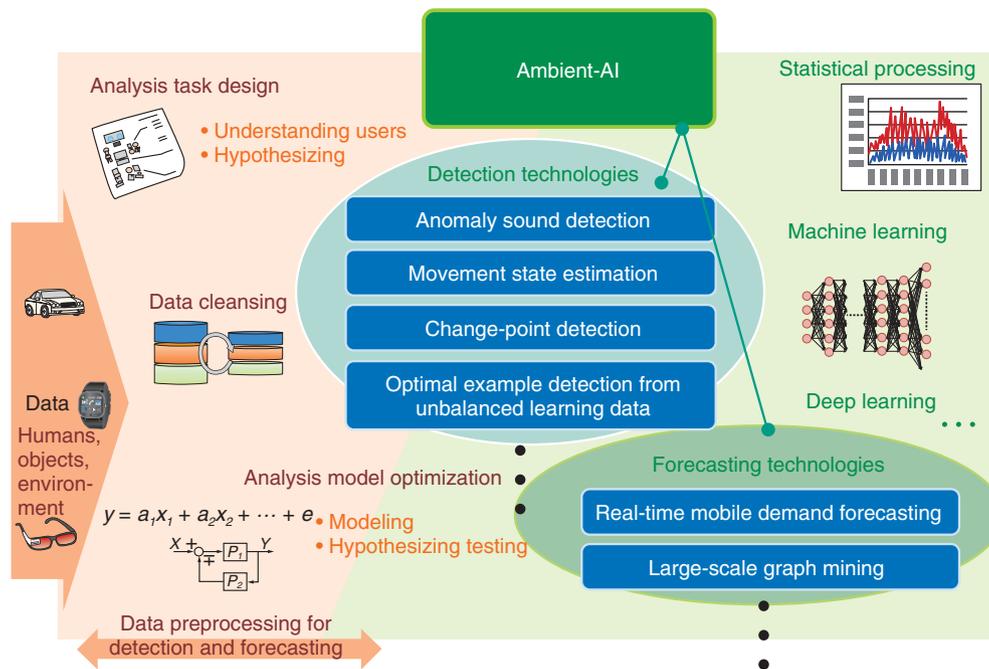


Fig. 5. Ambient-AI.

2.4 Ambient-AI technologies

Some of the Ambient-AI technologies being developed by NTT R&D are described below. These days, we frequently hear terms such as machine learning and deep learning. These are also utilized in Agent-AI. The most important factor in using machine learning, deep learning, or statistical processing tools is to optimize the analysis model used. How good or bad your analysis model is determines the value of your data. For instance, it has a significant impact on detection or forecasting performance. Thus, we are optimizing analysis models used by detection and forecasting technologies (Fig. 5).

We are developing technology that can quickly detect significant differences between two satellites or aerial photos of the same area taken at different times. It does so by measuring the difference in information quantity, called *entropy*, at each pixel after the photos are compressed using the video encoding applied to 4K and 8K. Currently, the rate of successful detection of change-points is about 90%, indicating that there is still room for improvement. We will conduct a field trial during this fiscal year with NTT GEOSPACE.

We are studying technology for learning with a high degree of accuracy just the relevant data from among a group of data of which only an extremely small fraction is relevant (Fig. 6). In collaboration

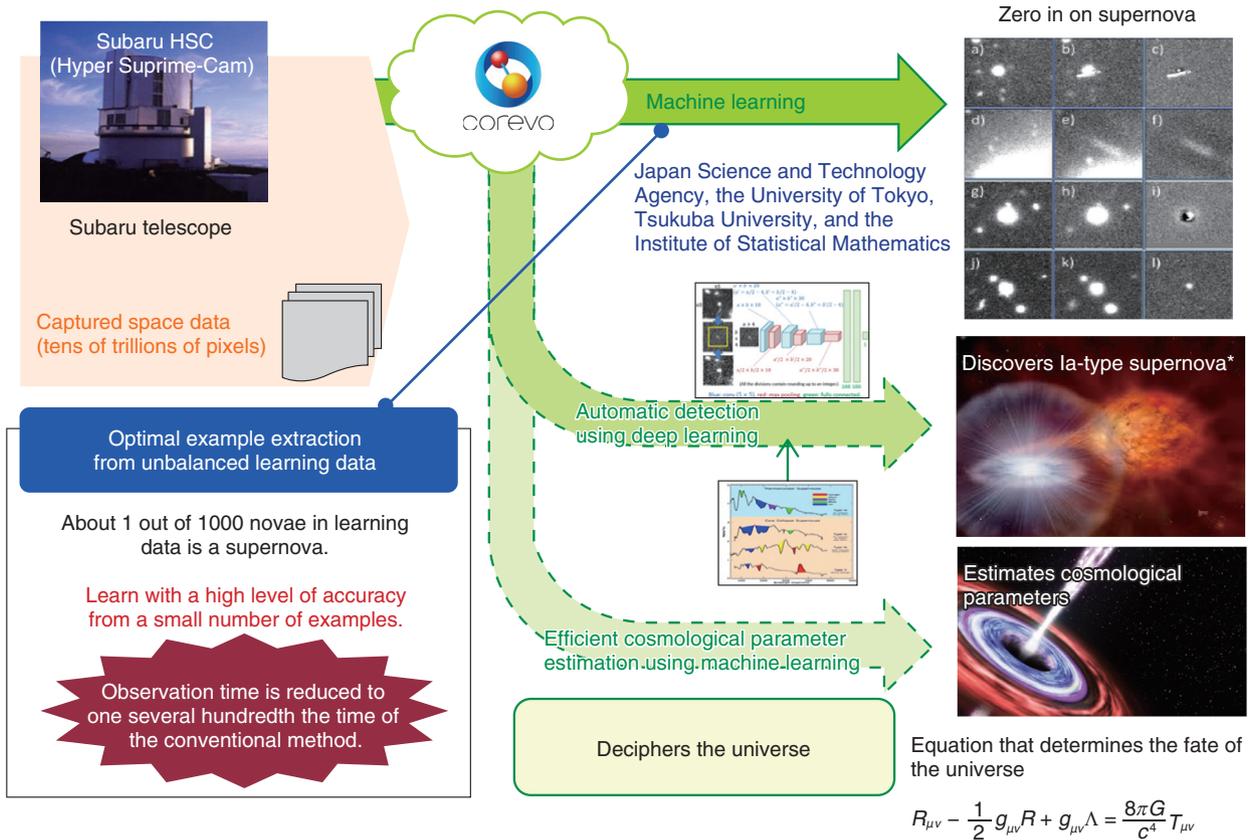
with the Japan Science and Technology Agency, the University of Tokyo, Tsukuba University, and the Institute of Statistical Mathematics, we are using this technology to zero in on supernovae from among a huge collection of space photos taken with the Subaru telescope. The number of pixels comprising these photos is on the order of several tens of trillions. Although only 1 in 1000 novae in the learning data is a supernova, this technology has reduced the observation time required to detect supernovae by a factor of several hundred compared to that of the conventional method. We next tried to discover Ia-type supernovae and found the first one at the end of 2016. Ultimately, we will try to use machine learning to estimate the parameters of the equation that determines the fate of the universe. Our aim is not exactly to decipher the universe but, rather, to refine our technologies through these activities.

2.5 Application examples of Ambient-AI

Some commercial applications of Ambient-AI technologies are described below.

(1) Detection of advance signs of machine failures

A technology that NTT DATA has jointly developed with Hitachi Zosen Corporation focuses on machine operation sounds in order to support stable operation of factories. It detects glitches and advance



*One of the subcategories of supernovae and cataclysmic variable stars

Fig. 6. Detection technology.

signs of failures using intelligent microphone technology, which picks up target sounds clearly even in noisy environments, and technology that distinguishes between normal and abnormal sounds.

(2) Detection of dangerous driving behavior

This technology detects dangerous driving of a car by analyzing multimodal information in video data stored in a drive recorder, and sensor data such as speed and acceleration information. In a trial jointly conducted by NTT Communications and Nippon Car Solutions Co., Ltd., dangerous driving was detected with about 85% accuracy. The information gained will be used for educating drivers and formulating measures to reduce accidents.

(3) Congestion forecasting

NTT DATA is seeking to help reduce traffic congestion by visualizing congested conditions and predicting imminent jams. This is achieved by applying large-scale graph mining to data from beacons and traffic information in text form. Field trials are being carried out in China, and another is planned for the

UK.

(4) Taxi demand forecasting

We are also using AI to combine and analyze a variety of information ranging from taxi operation data through demographic data, weather data, and event data, in an attempt to forecast areas that will see large taxi demand 30 minutes ahead. Allocating taxis based on such forecasting could boost taxi company sales and reduce waiting times for taxi users. NTT DOCOMO is conducting joint field trials with the Tokyo Musen Cooperative Association and Tsubame Taxi Group in Nagoya (Fig. 7).

2.6 Network-AI technologies

One category of Network-AI is application of AI to networks. We are working on technologies for detecting faulty parts in a network with a high degree of accuracy, for forecasting communication traffic by area, such as urban areas and residential areas, and for routing traffic to avoid faulty or congested points, if any. The other category of Network-AI is connecting

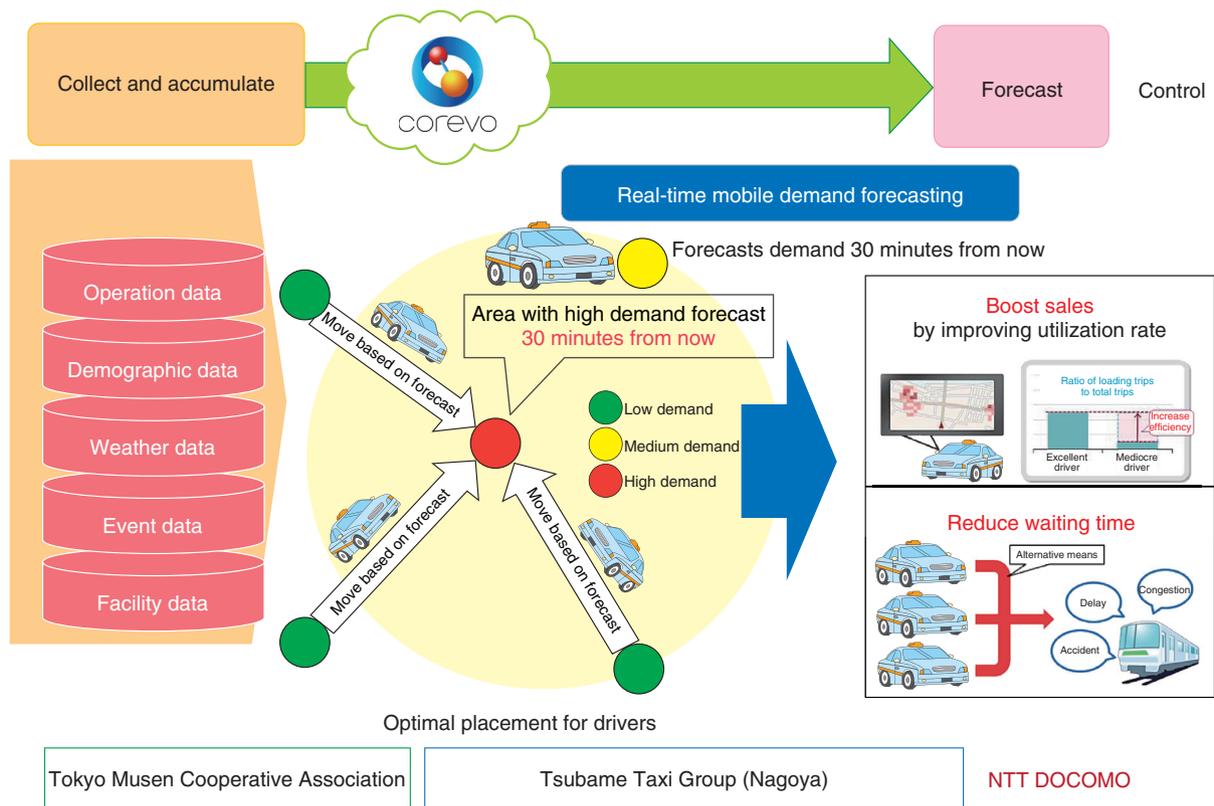


Fig. 7. Taxi demand forecasting.

different types of AI in order to achieve global optimization of a system. In collaboration with the Japan Agency for Marine-Earth Science and Technology, we are developing technology for combining wide-area simulation with local area-specific processing in order to improve the accuracy of weather forecasting.

2.7 Heart-Touching-AI technologies

We have recently initiated new activities in the field of sports brain science. According to specialists, physical performance is affected not only by physical conditions but also by the brain. The objectives of our activities are to elucidate brain mechanisms and to create effective training methodologies that can be used in the field.

3. IoT targeted by NTT R&D

The IoT world is aimed at creating new value by collecting data from various things in society, visualizing them, and analyzing them using AI technologies. NTT R&D is focusing on the following four IoT requirements.

First, for data collection, sensors need to permeate humans and objects *naturally*. Second, while some data processing, such as that for paddy field management in agriculture, has no time constraints, data processing for operation of machinery in factories must be carried out in real time in order to minimize delay. Third, for value creation, a mashup of diverse types of data is important. Lastly, a common requirement is secure handling of data.

3.1 IoT basic architecture

In line with these requirements, NTT R&D uses the following IoT basic architecture (Fig. 8).

If we are to promote circulation of information, the architecture needs to embrace the following parts: 1) an IoT gateway that unifies, at the entrance, communications standards that vary from industry to industry; 2) IoT data exchange that standardizes data formats and distributes data to different analysis functions; 3) software components and high-speed distributed processing needed to optimize data analysis performance; 4) library middleware and applications for turning data into value; and 5) management of the

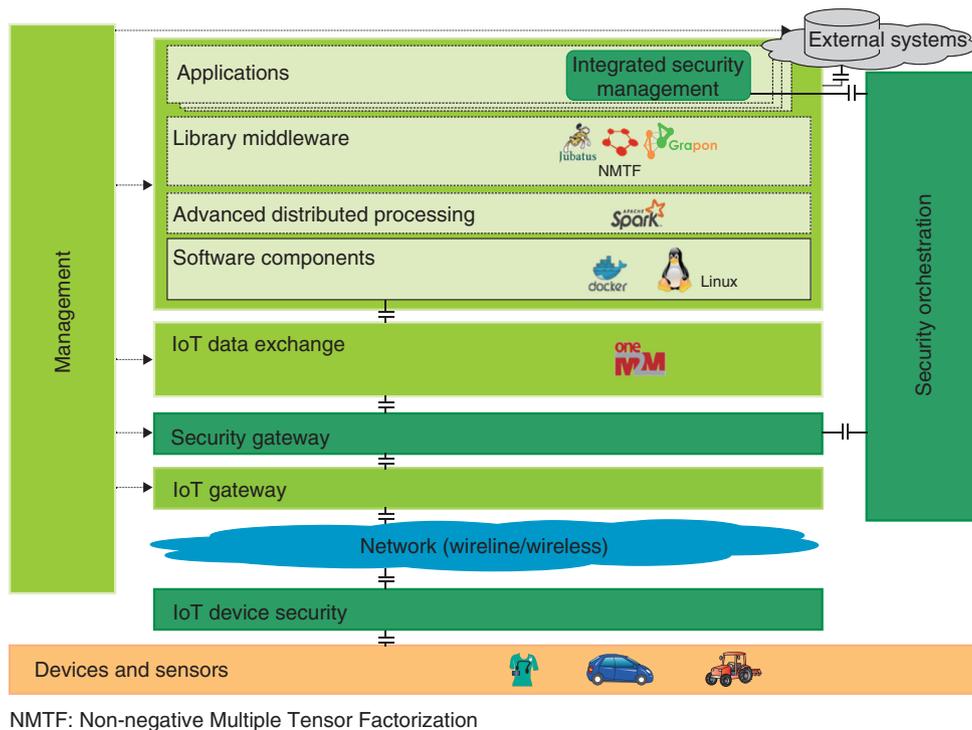


Fig. 8. IoT basic architecture.

entire operation. Furthermore, the following security measures should be added to the architecture: 6) IoT device security for protecting information at the communication layer; 7) a security gateway for detecting and blocking malicious data at the receiving points; and 8) integrated security management and security orchestration for detecting abnormalities in a system as a whole and immediately applying blocking rules.

We are developing functional units that implement NTT R&D technologies in the IoT basic architecture. We are also studying which of these functional units should be used in combination for a specific application area or usage, and how to allocate these functional units among clouds, edges near devices, and gateways if we wish to reduce delay or achieve compact packaging.

3.2 Sensing

In collaboration with Toray Industries, Inc., we have developed “hitoe,” a sensing fabric that can measure cardiograms and electromyograms. In addition to using the fabric to simply collect data, we have begun to study how to identify fatigue, heat stroke, and the user’s mental state from a cardiogram, and how to determine muscle fatigue and lactate thresh-

old from an electromyogram. For example, “hitoe” is used to estimate the fatigue level of long-distance bus drivers and to determine if outdoor workers are suffering from heat stroke. NTT WEST is employing “hitoe” to visualize the mental states of golfers.

Formerly, we were able to claim only that “hitoe” could measure heart rate. Since registering “hitoe” as a general medical device in August 2016 (Fig. 9), we have been permitted to state that the fabric can measure a cardiogram. Today, it is used in many hospitals. We are conducting a joint field trial with Fujita Health University to study how “hitoe” can be used to reduce the duration of hospitalization for patients undergoing rehabilitation.

We are also applying the fabric in professional sports. In the Indy 500 in the United States, NTT DATA uses “hitoe” to monitor the physical loads on racecar drivers in the extremely demanding environment. In bicycle races, “hitoe” is used to visualize not only road racers’ heart rates, speeds, and rotations but also their reserves of physical strength.

3.3 Application examples of IoT

Examples of the use of NTT R&D’s IoT technologies are presented below.

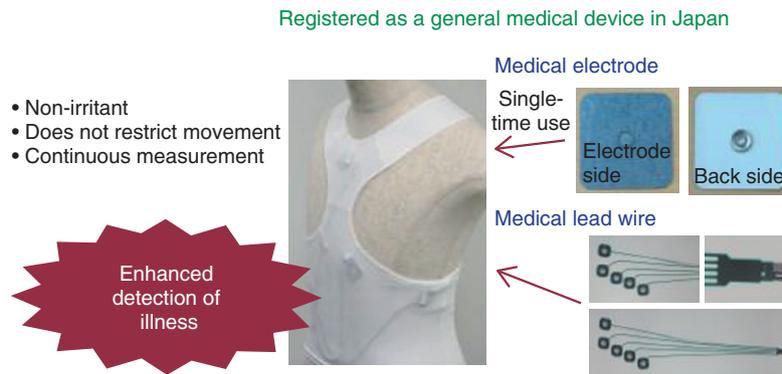


Fig. 9. Applications of “hitoe.”

(1) Support for vehicle operation

IoT technologies can support vehicle operation by estimating the fatigue levels of bus or truck drivers and recommending rest when a certain level of fatigue is suspected. In a field trial on a highway, conducted jointly with Keifuku Bus, data on the driver’s fatigue level were visualized. The data showed that after the start of the drive, the fatigue level increases gradually and that after resting at a service area, the driver recovers from the fatigue. Following this field trial, NTT Communications launched a service that supports vehicle operation. In addition to “hitoe” being used to obtain bioelectrical data, NTT R&D’s high-speed distributed processing plays a critical role in processing streaming data such as heart rate data in real time.

(2) Optimization of manufacturing/production in a factory

With a view to innovating manufacturing, we are seeking to raise production efficiency by getting various machine tools in a factory to work in coordination and by processing the data involved in real time (**Fig. 10**). A key to achieving this real-time processing is edge computing, in which the necessary processing functions are allocated at edges near the machine tools rather than in clouds. We have combined edge computing with an IoT data exchange function, which circulates information handled by various types of machine tools in the form of common data, and with a software component management function, which flexibly selects application programs according to the particular usage of machine tools in a factory. We are working with FANUC CORPORATION toward its planned launch of a service in autumn this year.

(3) Improvement of farming productivity

The NTT Group is collaborating with Kubota Corporation to improve productivity and competitiveness in agriculture by combining NTT R&D’s technologies with NTT GEOSPACE’s map data, which are NTT assets, Halex’s weather data, and JSOL’s yield forecast data.

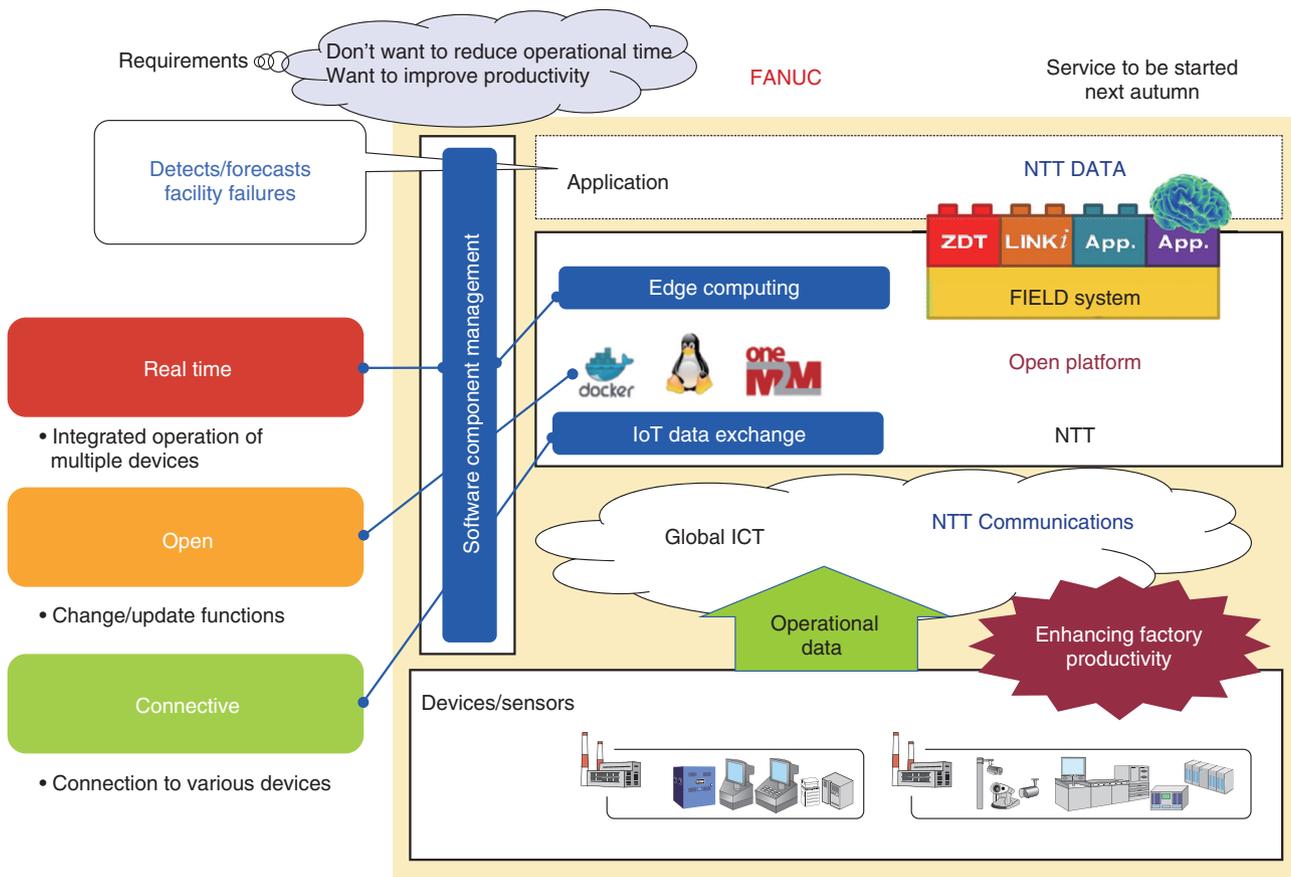
3.4 Security

An important consideration in promoting IoT is security. The IoT environment is characterized by an inability to execute complex processing due to limited processor power.

Our next-generation passwordless authentication technology does not require servers to manage information needed for authentication (**Fig. 11**). The user can securely authenticate an IoT device with only two items of information: a device identification and an item of secret information held in the IoT device. The server does not hold authentication information for each device. This technology has been developed and released as open source software. As a future application, we are studying the possibility of using this technology to prevent spoofing of IoT devices. For example, when a genuine home delivery drone and a fake home delivery drone arrive, it will be possible to authenticate only the real one.

We are researching technology for lightweight encryption, which is secure and does not impose a heavy processing load. It is important to confirm that the encrypted information will not be broken. We do so by creating encryption analysis methods ourselves and checking resistance to new attacks. We recently developed a new encryption analysis method called a nonlinear invariant attack and proved that some existing lightweight encryption systems are vulnerable.

We are also studying how to strengthen the security



ICT: information and communication technology

Fig. 10. Optimization of manufacturing/production in a factory.

of a system as a whole, especially critical infrastructure, in addition to the security of individual IoT devices. Cyber-attacks aimed at disrupting critical infrastructure can produce serious consequences. With Mitsubishi Heavy Industries, Ltd., we are jointly developing a system that determines the operating mode of a certain infrastructure from data collected from sensors, detects abnormalities, and restores the infrastructure. We are studying a way to minimize damage from novel cyber-attacks by combining security gateways, analysis applications, and security orchestration.

4. Visualizing the near future and beyond

Looking to 2020 and the creation of lasting assets, we believe that NTT should play three roles: providing stable high-quality network services, implementing reliable network security measures, and providing hospitality, deep positive impressions, and new experiences to visitors. This article focuses on our activities in respect to the third role.

riences to visitors. This article focuses on our activities in respect to the third role.

4.1 Hospitality

We have initiated a number of trials aimed at providing services for international tourists, who continue to increase in number. In field trials underway at a Tokyo Metro subway station and at Nissan Stadium, angle-free object search technology, mentioned above, and 2.5-dimensional (2.5D) map representation technology are used to provide intuitive navigation (Fig. 12). When a tourist points a smartphone at an information sticker pasted in a Tokyo Metro station, his/her current location is detected without any need for beacons. When the user inputs his/her destination information, a 2.5D map, which is a 2D map with height information added, is displayed on his/her smartphone so that he/she can easily find the way to the destination. In the field trial at Nissan Stadium, navigation using an accessibility map is also provided.

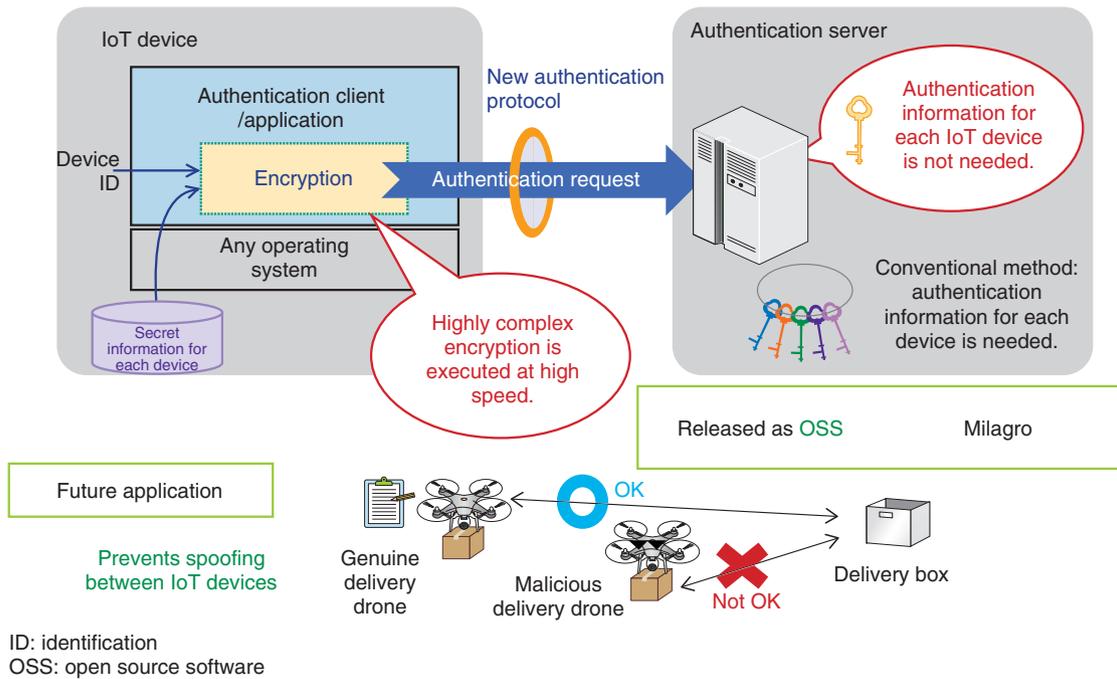


Fig. 11. Next-generation passwordless authentication.

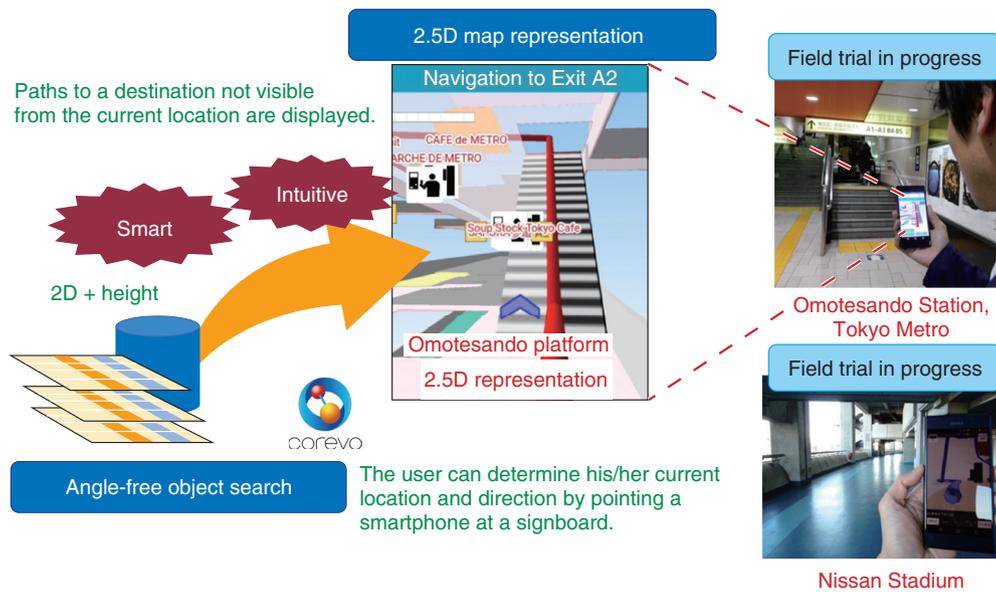


Fig. 12. Navigation for international tourists.

For example, when the system learns that a person is in a wheelchair, it provides navigation specifically designed for wheelchair-bound persons.

In the field trial at the Tokyo Metro subway station,

angle-free object search technology is used to direct passers-by using advertisements on walls. For example, when a person points a smartphone at an advertisement on a wall, he/she receives a coupon for a

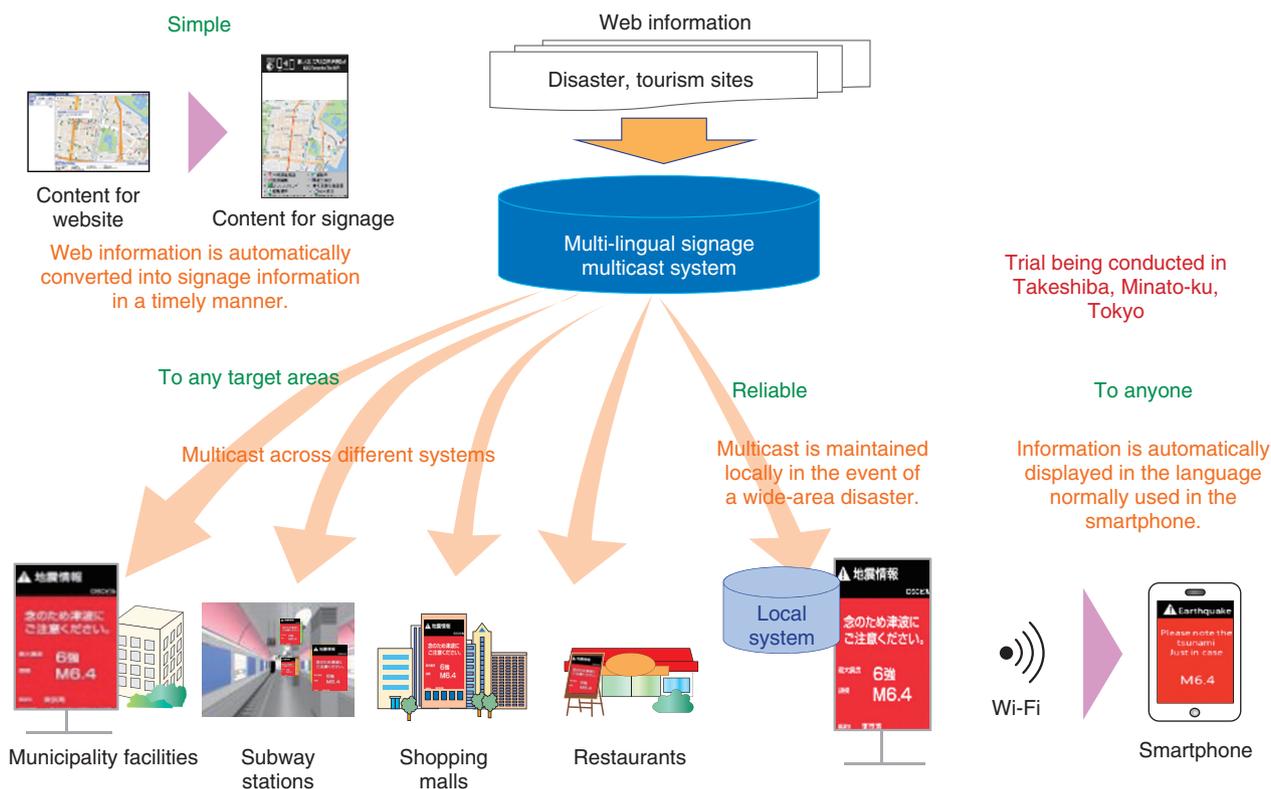


Fig. 13. Multicast of multi-lingual information on signage.

special offer as well as information on how to navigate to the shop concerned. This will be especially convenient for overseas visitors.

In a field trial conducted in Takeshiba, Minato-ku, in Tokyo, emergency information is multicast to various signage systems in the event of a disaster (Fig. 13). It is vital to ensure that information that has been multicast to signage units dispersed through a town is definitely displayed, irrespective of the sign owners or the types of signage units. We have developed a multi-lingual multicast system that multicasts information to different signage systems and directs those systems to send information to the smartphones of passers-by in such a way that the information is automatically displayed in the language used in the smartphone. The objective is to enable overseas visitors to promptly receive emergency information in the event of a disaster.

In the retail and distribution fields, NTT and Seven & i Holdings Co., Ltd. are carrying out a joint field trial in which overseas visitors can easily get product information using smartphones. For example, when a person points a smartphone at a rice ball, he/she can

immediately get information on the ingredients and allergic substances contained in the rice ball—currently, in any of 15 languages.

With a view to providing ease of use, we have developed a see-through device in collaboration with Panasonic Corporation (Fig. 14). Since the device is not as powerful as a smartphone, its operation is assisted by edge computing technology. When it is pointed at an object, it displays information about that object. It uses the latest transparent color display from Japan Display Inc. Widespread use of these technologies will expand the industry in Japan.

Since last year, we have been holding R&D Forum Showcase to enable visitors to the forum to have hands-on experience with our R&D technologies. In this event, visitors can try our official app. It recommends a route to an individual user based on his/her interests and the current distribution of people on the floor, and answers questions about recommended exhibits using AI. When the user points a smartphone at an exhibited panel, the “point and get information” function of the app displays information relating to that exhibit. Information is shown in English if the

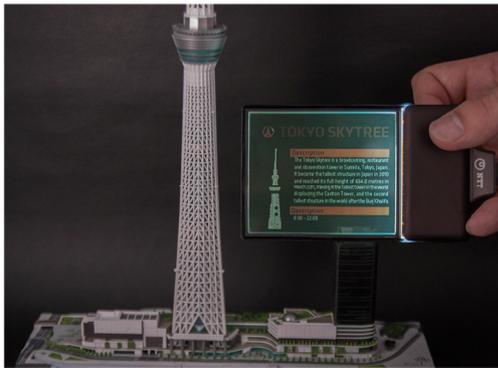


Fig. 14. See-through device.

user is from overseas.

4.2 Deep positive impressions and new experiences

The immersive telepresence technology called “Kirari!®” has been exhibited at NTT R&D Forum since 2015, when images of people were extracted from a recorded video and displayed in quasi-3D. In 2016, images of people were extracted from a video in real time for the first time during my keynote address. In 2017, we attempted a more challenging undertaking of extracting images of people from a wider area in a higher-definition video.

We have been experimenting with convergence of information and communication technology (ICT) and kabuki, a Japanese traditional performing art. At the Niconico Chokaigi (super conference) held in April 2016, “Cho Kabuki” was presented. This was followed by a performance of “KABUKI LION SHI-SHI-O” produced by SHOCHIKU Co., Ltd. in Las Vegas, in May. In March 2017, “Virtual Kabuki Theater” was staged in Kumamoto. The latest object extraction technology has been used to display the “swinging of head hair” by a kabuki actor. Details of his whirling hair were extracted and displayed clearly and naturally.

We will continue to improve Kirari!. Next year, we will develop technology for projecting a 3D image in such a way that it can be viewed from an arena-shaped spectator area. In other words, the image will be viewed not just from one direction but from all directions.

4.3 Convergence of sports and ICT

We are using ICT to assist the training of athletes and to captivate fans by giving them deep, positive

experiences of sports events.

Our efforts to help train athletes target both endurance-type sports such as swimming and cycling, and instantaneous and interpersonal sports such as baseball and badminton. For the latter, we are elucidating brain functions of athletes by collecting data from former professional baseball players and members of university baseball teams. For the former, we are studying analysis of heart rates and myoelectric information on swimmers because it has become technically possible to use the sensing fabric “hitoe” for underwater monitoring.

We are studying the use of virtual reality to offer new experiences in watching sports events. We have developed athlete first-person vision synthesis technology. This was initially targeted at baseball. It synthesizes a real baseball park video and a CG (computer graphics)-reproduced baseball so precisely that the user can view the ball and the park from any viewpoint, for example, from the position in the right batter’s box or from the position of the catcher. Our objective was to enable children to experience fast balls pitched by top players. However, systems that implement this technology are now also used by professional baseball teams for training.

We have expanded the application of this technology to other types of sports. For example, the user can experience a free kick by a soccer player or a drop shot serve by Kei Nishikori, a Japanese tennis player with whom we have concluded a sponsorship contract. Conveying the amazing skills of top athletes to children—the next generation of players—will help to expand sports-related business.

5. Conclusion

Although I have mainly introduced applied technologies in this article, we are also committed to basic research, knowing that the various AI technologies in our hands today have been built on the results of our past basic R&D spanning several decades. The lesson is that we should not confine ourselves to R&D of technologies that are useful today. We must also pursue basic R&D that may become applicable five or ten years from now.

Let me introduce two types of basic research. We have developed a new quantum computing principle that uses light to rapidly solve difficult mathematical problems that cannot be cracked using today’s super computers. This was published in *Science*, a U.S. scientific journal. In the quantum world, there is a realism problem whereby the state cannot be

determined before the time of observation. There has been intensive argument as to whether this is true only in the quantum world or also holds in the real world or in the macroscopic world. We have proven that breaking of realism occurs even in the macroscopic world. This was released in *Nature Communications*, a UK scientific journal.

We are convinced that if we want to promote B2B2X, we must garner abilities that enable us to

continue to be selected by our partners. If we are to succeed in collaboration, we must have the ability to integrate the strengths of all those involved in any such collaboration, and NTT R&D must consistently produce cutting-edge research results. We will not emphasize one or the other of the above but will commit ourselves to garnering the ability to collaborate and pursuing basic and fundamental research.