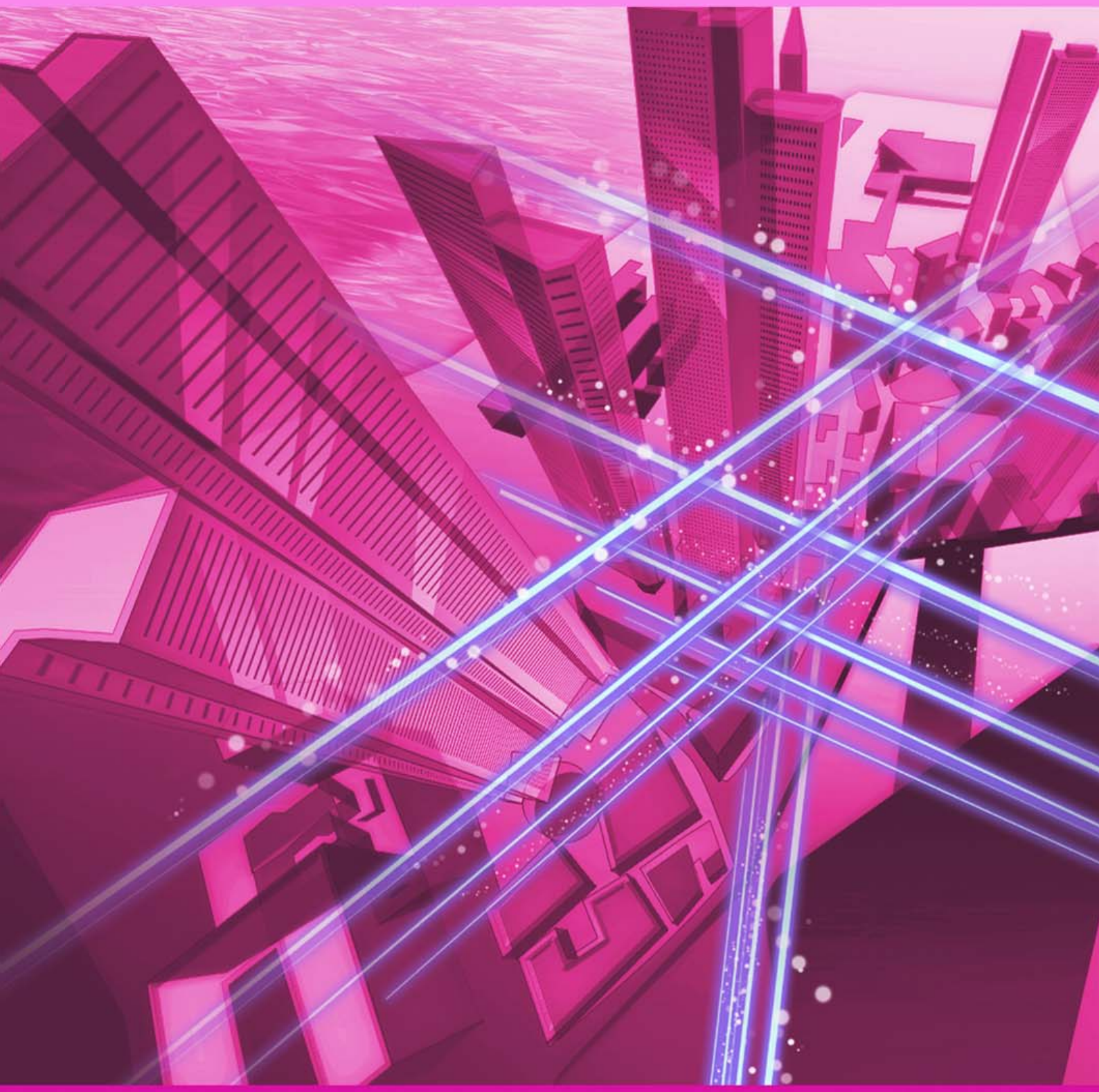
NTT Technical Review

11
2016



November 2016 Vol. 14 No. 11

NTT Technical Review

November 2016 Vol. 14 No. 11

Short Reports

- Arkadin’s Unified Communications & Collaboration Portfolio and Service Strategy Position it for Leadership in the Virtual Workplace

External Awards/Papers Published in Technical Journals and Conference Proceedings

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

Establishing a Global and Local Brand of NTT DATA with an Eye to Our Clients' Future

Eiji Ueki
Executive Vice President, NTT DATA

Overview

Eiji Ueki, NTT DATA Executive Vice President, believes in the importance of strengthening global brand power by developing trustworthy relationships with clients and providing new value with an eye to their future. We asked Mr. Ueki how NTT DATA plans to face the challenges ahead.

Keywords: management plan, global brand, client relations

Innovative production technologies and advanced technologies as keys to the future

—Mr. Ueki, as Executive Vice President in charge of the financial segment and technology strategies, please tell us about the current state of business and the areas of focus of NTT DATA.

In the Medium-Term Management Plan running from fiscal year (FY) 2012 to FY2015 (the period from April 1, 2012 to March 31, 2016), we set specific targets of becoming a global top 5 enterprise (net sales of 1.5 trillion yen) and achieving earnings per share (EPS) of 200 yen. To this end, we defined three areas of action: expansion of new fields and reinforcement of product competitiveness; expansion, enhancement, and reinforcement of global business; and pursuit of overall optimization. In domestic operations, our remarketing initiatives and control of unprofitable projects put us on an upward trend from FY2013 in both sales and profit, and in our overseas operations, steadily increasing sales enabled us to achieve our net sales and EPS targets in FY2015.

The business of NTT DATA has traditionally been in the public and financial segments, which have been the company's dominant fields since the Nippon Telegraph and Telephone Public Corporation era. Against this background, we have been accelerating our activities in the corporate domain with an eye to expanding business even further. Moreover, to establish good relations with Japanese enterprises that are aggressively expanding their business globally, we ourselves have been actively pursuing globalization through merger and acquisition (M&A) activities over the last ten years. Here, we have been targeting overseas companies that have a close affinity with the business of NTT DATA or that have know-how related to enterprise resource planning (ERP) packages. As a result, our global business volume has been increasing year-to-year. However, in comparison with its name recognition in Japan, NTT DATA has yet to establish itself as a globally recognized brand. For this reason, the new Medium-Term Management Plan calls for business growth and greater local presence in every region of the world and establishment of a truly global brand toward the Global 2nd Stage,

as a basic policy called NTT DATA: ASCEND (Rise and grow as a global brand).

Two major pillars of the new Medium-Term Management Plan are creating value through a "Game-Changing Approach" and "Breakthrough Technology." The latter seeks to enhance technological competence—a major element of NTT DATA's business—through innovation in production technology and application of cutting-edge technologies to provide clients with high-value systems.

As for innovation in production technology, we are promoting the automation of software development, the accumulation and expansion of legacy modernization technologies for smooth upgrading of large-scale legacy systems, and an agile development approach. In particular, we have been engaged in the automation of software development for some time, but further efforts will be made. This is because automation enables us to focus as never before on what kind of system should be created to grow the client's business and provide the client with value. Consequently, the automation of software development helps us to provide our clients with high-value systems.

Regarding the application of cutting-edge technologies, we are constantly keeping an eye on new technology trends and researching and developing new technologies in up-and-coming fields such as artificial intelligence (AI), Internet of Things (IoT), and robotics in collaboration with NTT laboratories. Plus, in fields that are difficult to go into alone or that demand rapid development, we are pursuing open innovation through efforts such as forming tie-ups with start-up companies. In this way, given that information technology (IT) is continuously evolving, we seek early adoption of the latest developments to provide clients with high-added-value systems.

—How are you working, in particular, to increase brand power?

From the viewpoint of technology development, I believe that using new technologies in actual business and establishing a proven track record in this regard can help increase brand power. For example, there's Sota®*, a communication robot that performs advanced intelligent processing using a cloud robotics platform. A Sota robot can be positioned at a bank's reception desk to provide guidance to visitors and direct them to the teller counter. The finance industry, which is generally thought to be somewhat conservative, is now expressing interest in new technologies such as financial technology (FinTech) and robots.

Within Japan, I would like to enhance the NTT DATA brand by providing solutions to the public, finance, and enterprise segments. In the public segment, I would like to provide solutions related to the My Number (social security and tax number) system. In the finance segment, I want to provide FinTech-related solutions and core systems as well as solutions that help improve the client's top line. In the enterprise segment, I would like to build up our track record in providing omni-channel retailing- and IoT-related solutions.

In the overseas market, where the NTT DATA name was once unknown for the most part, we have obtained a certain amount of name recognition through corporate sponsorship of the British Open in golf and IndyCar Series racing in the United States. Nevertheless, we still have to improve our brand power by promoting our successes in the actual business domain. In this sense, let me give you two examples of business achievements that have propelled us forward.

The first is our digital archive project with the Vatican Apostolic Library, which I believe will have great appeal to the many followers of the Roman Catholic religion throughout the world and the many people that visit the Vatican every year. Furthermore, as an additional benefit, being involved in such a magnificent project has had a very positive effect on NTT DATA employees.

The second example is our business with globally expanding automobile manufacturers. Thanks to our M&A of Cirquent, an information systems subsidiary of BMW in Germany, we have received orders from

* Sota is a registered trademark of Vstone Co., Ltd.

BMW and also from Daimler and Volkswagen, thereby expanding our global support business on a scale of 10 billion yen annually in sales.

I would like to strengthen the NTT DATA brand by improving our track record of such successes one project at a time.

Building trustworthy relationships with our clients at every level

—What is essential to developing business relationships with clients?

To begin with, we must keep in mind that the client comes first. One positive quality of NTT DATA that has been true from the start is the building of long-term relationships. In other words, NTT DATA places importance not on the short term but on loyalty based on trustworthy relationships that may span five to ten years. While the world may change because of IT, relationships with business-to-business (BtoB) clients do not change greatly. We are expected to contribute to the growth of our clients' businesses as part of such a trustworthy relationship.

At the same time, it is essential that we provide our clients with know-how and solutions that we possess whenever such changes in technology occur. NTT DATA annually updates its NTT DATA Technology Foresight report that describes technology trends and future uses of IT for the benefit of its clients. While discussing what that outlook means for our clients' businesses and proposing new solutions, we develop forward-looking business.

Long-term relationships are necessary not just at the top executive level but at every level of a company including management, system development, and on-site operation. For example, in the construction of an integrated services center for banks, system development should only start after coordinating the interests of the stakeholders and creating a win-win relationship among all clients in the planning stage. In addition, if certain conditions should arise that can affect the system construction schedule, an effort must be made to revise the development scheme and minimize the impact while asking the clients to make adjustments to their requirements if necessary. Furthermore, the occurrence of a problem in a bank's core system—something that should never be allowed to happen—can have a huge impact not only on the client, that is, the bank, but also on ordinary people who use that bank. At such a time, it is essential that the client's understanding and cooperation be obtained in hammering out a solution to minimize the impact as much as possible. A trustworthy relationship with the client is vital at practically all stages of system provision, from system planning and development to system operation.

—What are you doing to help employees broaden their perspective?

Our business is project based, so employees cannot help but be concerned only about their current project. Similarly, executive managers supervise business efforts on a department-by-department basis such as whether they concern public services or financial services, and as a result, they tend to fall into a vertical and compartmentalized frame of mind. For this reason, we have set up a system called "executive caravan" in which executive managers make on-site visits to other departments. I have heard from the person in charge of this system that managers encounter all kinds of interesting challenges in these visits and always discover something new.

In addition, NTT DATA opened the "From the Toyosu Port" Open Innovation Forum three years ago. The idea here was to provide a forum for NTT DATA, NTT DATA clients, and start-ups with groundbreaking technologies to come together and combine novel ideas and technologies, all with the aim of producing innovative services quickly. This

forum has also come to be used as a place where NTT DATA employees can step outside the departments they belong to and propose new business ventures in a cross-disciplinary, autonomous manner. Moreover, in cooperation with overseas start-ups and the embassies of their countries, it has become a very energetic site fostering a lively exchange of ideas based on the interesting technologies of those start-ups and themes proposed by NTT DATA clients as well. However, these activities are not only meant to produce new ideas. For ideas that are original and interesting but that lack a means of realization, the forum is starting to produce results by applying such ideas to actual services. For example, linking Japan's ANSER Internet banking service with a household budget ledger app makes for a service that enables the user to access information from multiple financial institutions without having to worry about identifications and passwords.

—What is the role of NTT DATA within the NTT Group?

In 1988, NTT DATA was spun off from NTT as an independent company, but at that time, the relationship between telecommunications operators and system integrators was not that deep. In addition, the majority of current NTT DATA employees joined the company after its founding, which meant that they had only slight exposure to the corporate culture at Nippon Telegraph and Telephone Public Corporation and NTT. As a result, NTT DATA seemed to have a somewhat separate existence within the NTT Group.

However, the recent spread of digital technologies such as IoT and cloud computing has made it important that we coordinate with the NTT Group to provide our clients with full-support services. In this regard, I believe that we are already producing results. Particularly, in our overseas operations, where the brand power of NTT DATA and even that of NTT is weak, such coordination is becoming increasingly important to establish a successful track record on a global basis.

In research and development too, the direction of research since the completion of the DIPS mainframe development has been somewhat different from the business of NTT DATA. This created the impression of two companies that are distant from each other, but in these last few years, we have become much closer. The robot dialogue control and AI applications that I mentioned earlier are good examples of this trend. Based on the basic technologies coming out of NTT

laboratories, we are becoming increasingly successful in arranging and providing NTT technologies to our clients.

—In light of this situation, what would you like to say to NTT DATA researchers?

The IT domain is extremely broad and I'm sure researchers too are often at a loss as to what should be researched and developed. So I would ask researchers in their daily work to constantly think about what areas of IT might be important for our clients instead of being obsessed with immediate matters. If a researcher is not contemplating what clients want and what their latent needs might be, and how technology can be used to meet those needs, it is easy to fall into a state of complacency. Those of us in the NTT DATA business division can provide valuable information that we obtain from our clients, so I would like to produce good results by establishing clear channels of communication between us and NTT DATA's researchers to make the world a better place.

—Mr. Ueki, please leave us with a message for all NTT DATA employees too.

To move business forward, it is important that we exchange all kinds of ideas and that we share our problems and search for solutions together. Instead of worrying only about one's own project, working in a synergetic manner across multiple projects can create new business opportunities or provide clients with new value. To this end, I would like to ask all employees to make an effort to widen their perspective.

Also, for those employees who are active overseas, I emphasize that relationships with clients in every region of the world are of prime importance. So to

begin with, please work diligently to build up business in your regions. At the same time, please adopt a mindset that sees solutions as something that can be cultivated by applying the synergy of the NTT DATA Group, and carry out your work while looking at things from the viewpoint of the entire group.

Interviewee profile

■ Career highlights

Eiji Ueki entered Nippon Telegraph and Telephone Public Corporation (now NTT) in 1981. Following the spinoff of the NTT Data Communications division as NTT DATA Corporation in 1988, he served as Senior Vice President, Head of the Second Financial Sector from 2009, Director and Senior Vice President, Head of Group Strategy Headquarters from 2013, and Director and Executive Vice President from 2015. He took up his current post as Representative Director and Executive Vice President in June 2016.

Research is Like Child Rearing—Many Failures Hold the "Seeds of Research"

Tomohiro Nakatani
Senior Distinguished Researcher, NTT Communication Science Laboratories

Overview

Interfaces using speech recognition have become common practice these days. However, commonly used technologies for this purpose can suffer a drop in recognition performance in noisy environments or when the microphone is too far from the speaker. There is therefore a growing need for technologies that can provide more robust and accurate speech recognition. We asked Dr. Tomohiro Nakatani, Senior Distinguished Researcher at NTT Communication Science Laboratories, whose technology last year achieved the world's highest performance for speech recognition in noisy environments, to tell us about his recent research results and his approach to research.

Keywords: speech recognition, speech recognition interface, dereverberation

Achieving a speech recognition interface that can understand human conversation in diverse environments

—Dr. Nakatani, please tell us about your current area of research.

I am researching natural speech recognition interfaces (**Fig. 1**). For example, some of you may have experienced operating a smartphone by speech recognition. Speech recognition is a simpler and more convenient way of inputting information than using a keyboard. In recent years, we have seen wider use of voice-operated smartphones and tablets, so the usefulness of speech recognition interfaces has become well known. Current smart devices, however, require the user to bring the microphone up close and to speak distinctly. But in the real world, people communicate freely with each other without worrying about the existence of a microphone.

I wanted to create a mechanism that enables anyone to access information or to converse with robots simply by speaking without paying attention to electronic devices, so I have been doing research in this field ever since I joined NTT.

This research has been advancing, and as a result, the smart home is on its way to becoming a reality. At the present stage, the goal is to enable users in a smart home with a microphone installed in the living room to operate home appliances through vocal commands. The key to this mechanism is technology that can recognize speech even for utterances made at a

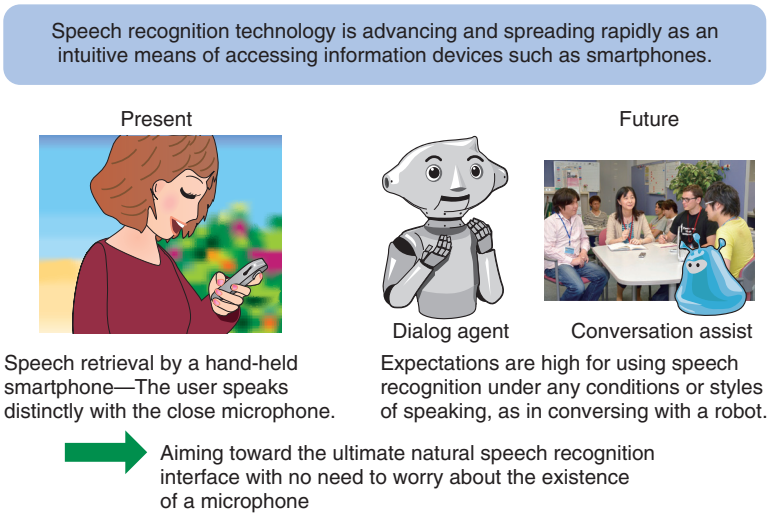


Fig. 1. Natural speech recognition interfaces.

distance from the microphone. An era is approaching when a computer will be able to accurately recognize the speech of each person and extract information in a situation in which a number of people are sitting around a table and conversing.

New challenges and friendly competition among experts who develop world-class technologies

—This sounds like technology that can have a profound effect on our lives.

Actually, NTT has already developed advanced technologies that can accurately recognize and process speech even in noisy environments or when there is reverberation. In fact, NTT has been a world leader in these technologies. For example, in last year's international technology evaluation competition of mobile speech recognition in noisy public spaces (CHiME-3 Challenge), NTT's system won first place in speech recognition accuracy at a level that was significantly higher than the second-place system. This is a result of developing a number of key technologies. These include technology for reducing noise and reverberation without distorting the user's speech, and deep-learning speech recognition technology for accurate modeling of speech even under noisy conditions (**Fig. 2**). Our word error rate (WER) in this competition was about 5%. In contrast, the WER of speech recognition by a conventional deep neural network (a machine learning technique simulating the human brain) was 33%. In addition, an investigation by a major Chinese search engine company found that the rate of mistakes made by people when listening to speech was about 11%. You can see that the WER that we achieved significantly outperformed other systems. This NTT technology should provide a foundation for speech recognition that exceeds human capabilities even in noisy environments.

Let me talk about these technologies in easy-to-understand terms. Please visualize a scene in which a number of people are talking freely inside a room. If an audio recording was made at this time without placing a microphone near the speakers, the accuracy of speech recognition would dramatically drop. There are two main factors behind this deterioration. One is that the quality of the speech recorded via the microphone is poor. This is because of background noise such as from air conditioning and the effects of reverberation caused by speech waves arriving slightly late at the microphone after reflecting from walls or other objects. In addition, overlapping speech from more than one person may be recorded. The other factor is that pronunciations are often ambiguous, and words are often omitted since speakers are talking freely without paying attention to the microphone.

To deal with such factors, we need speech enhancement technologies such as denoising, dereverberation, and speech separation to reduce the effects of noise, reverberation, and other people's voices, respectively. We also need natural-language speech recognition technologies for accurately recognizing unconstrained speech. I myself am mainly in charge

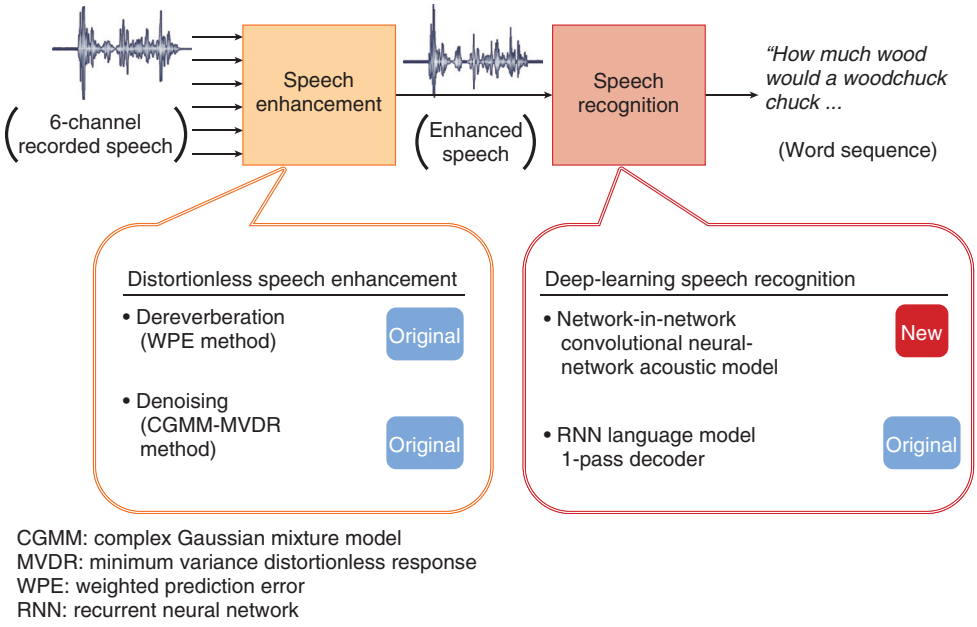


Fig. 2. NTT system features.

of speech enhancement technologies for distinguishing a person's voice.

Since we came out first in the CHiME-3 Challenge, we can now look back with a sense of joy, but at the time, I can honestly say that we were quite apprehensive about what kind of evaluation our technology would receive. However, I felt that we had to live up to the NTT name, so I wanted to do whatever it took to take first place. We established a WER target within 8% and took up the challenge of a new and difficult task. Somehow, we achieved good results thanks to our efforts.

—There must have been some drama behind the birth of this technology.

Yes, there were all kinds of dramatic stories, but let me tell you about our Signal Processing Research Group, where I am currently group leader. In 2000, the year following the reorganization of NTT, this group was set up in NTT Communication Science Laboratories (NTT CS Labs) thanks to the efforts of the late Dr. Yoichi Tokura, the director of NTT CS Labs at the time, Dr. Shigeru Katagiri, the first group leader (currently Doshisha University professor), and others. The mission established for this group was to remake speech processing technology starting from the basics and discover new possibilities that exceed conventional limits. The group brought together an elite team of basic researchers in two different technologies—audio processing and speech recognition. I joined this group as a researcher in 2001.

From the start, the aim of this group was to develop natural speech recognition interfaces. To develop this "dream technology," we would have to solve a number of challenging problems that would take a long time to overcome. These include development of speech separation technology to distinguish speaker voices when more than one person is speaking at the same time, dereverberation technology to remove reverberation from speech recorded in a reverberant room and make it easier to hear that speech, speech activity detection technology for detecting speech intervals from recorded sound mixed with noise, speaker diarization technology for estimating who among multiple speakers is speaking and when, acoustic modeling technology for accurately modeling and classifying speech even under noisy conditions, and computationally highly efficient decoding technology for real-time continuous speech recognition that can handle an ultra-large vocabulary of more than 10 million words. These technologies reflect the many problems that had to be overcome.

It's been about 15 years since then, and I can say that many of these problems that we took to be very challenging at the time have been solved. We have

Fig. 3. Demonstration of on-site recognition of recorded conversation.

come a long way, and at the NTT Communication Science Laboratories Open House 2016 held earlier in the year, we introduced a demonstration system that performs on-site computer recognition of the speech of multiple speakers sitting and conversing at a table with a microphone placed at the center (**Fig. 3**). Please take a listen!

—This speech is really clear after removing noise and reverberation!

A driving force behind this development and one of its features is an assembly of researchers from the two different fields of audio processing and speech recognition who work side by side in our group. These fields may appear similar at first glance, but their basic technologies and application targets are quite different. They have progressed along different paths. Actually, our group that is carrying out these two lines of research together is unique even on a worldwide basis. Exchanges between these fields in terms of products or academic activities are not that frequent. Through pioneering research that places importance on this boundary region, I believe that we have given birth to a string of new ideas that have helped propel speech research at NTT.

To give an example of just how important the intersection between these two fields has been, let me tell you about dereverberation technology, which I have been deeply involved in. Research into removing reverberation is an area of audio processing that has been progressing for quite some time. The idea here is to mathematically model the propagation of sound in a room and to use that model to remove unnecessary sound such as reverberation from recorded sound. However, this method has not been effective for adequately removing reverberation if the conditions of the room in which sound is to be recorded are unknown.

In speech recognition, in contrast, there is a commonly used technique for learning patterns in speech signals and for processing signals to identify those patterns. (In modern parlance, this process would probably be called a machine learning framework.) With this in mind, we proposed a framework called "pattern-oriented audio processing" that incorporates the idea of pattern processing developed in speech recognition into audio processing. We then proceeded to develop a variety of new algorithms based on this framework. In this way, we developed a technology that automatically recognizes what is happening within recorded sound under unknown conditions and that uses a mathematical model from audio processing to accurately break down the features of the

target sound. Amid this flow of developments, a new dereverberation algorithm was intensively researched around 2006 here at NTT CS Labs in Keihanna (Kyoto-Osaka-Nara) Science City, resulting in a world-first technology.

Another reason as to why the Signal Processing Research Group has been able to produce such breakthrough results is that our research has continued along the path of basic research pioneered by many great senior researchers at NTT. Since the invention of audio coding for mobile phones in the Nippon Telegraph and Telephone Public Corporation era, NTT has a history of being a world leader in speech research. There has also been a strong, trustworthy relationship between NTT researchers and worldwide researchers for many years. This research environment that enables us to share a view of the world with these great senior researchers (at least in part) has been a driving force behind our basic research endeavors.

Finding the "seeds of research" through many failed attempts

—What kind of mindset is important when beginning a research project?

I would like researchers to take three points to heart. First, when trying something for the first time, be prepared for things to go against your plan. Second, even if you face up to the fact that the result was not what you expected, keep in mind that the result may be useful in later research in some form. Third, after long and careful attempts to achieve results, the result that you are seeking may suddenly appear.

For example, we initially experienced a number of failures in our research on dereverberation technology that we eventually achieved as a world-first. Dereverberation had been a major problem for some time in the field of audio signal processing, but a definitive solution had eluded researchers. In the long history of this research, it was believed that the biggest problem in achieving effective dereverberation was determining how to prevent the occurrence of speech whitening, a phenomenon in which the results of processing are a flat speech spectrum and a loss of natural speech characteristics. We ourselves tried all sorts of ideas to solve this problem. However, while applying countermeasures to speech whitening had a somewhat positive effect on improving dereverberation, we were still short of a fundamental solution. It was really one step at a time in the dark.

Amid these trials and tribulations, the pattern-oriented audio processing that I mentioned earlier became our research compass. In particular, we introduced a technique commonly used in speech recognition that learns and distinguishes temporal changes in sound as patterns, and we used it to approach the dereverberation problem. After beginning with some primitive trials, we failed repeatedly in our experiments. But as partial successes began to appear, we started to make discoveries a little at a time. One day, however, after a long period of trial and error, we suddenly arrived at an important discovery. Keeping the temporal structure of the speech from collapsing is more important than preventing spectrum flattening. I really felt as if the veil had fallen from my eyes, even with all the experience I had had as a researcher. With this discovery, we shifted our research to creating a mechanism for removing reverberation without destroying the temporal structure of the speech. Our research into dereverberation began to accelerate from that point on.

—As an active researcher, what does it take to produce results? What would you say to young researchers?

Thoroughly testing results that go against your expectations will reveal the next "research seed." Experiencing many failures is an interesting thing. It gives you an opportunity to think about what went wrong. Indeed, from the viewpoint of always failing, I think research and child rearing can be very similar! I myself have two sons in primary school in the third and fifth grades. For example, thinking that I want my kids to be more careful with their time, I promise to give them a reward if they follow a rule that I set down. However, while they may observe the rule perfectly the first time, it is not uncommon for them to completely ignore it the second time. I then wonder why it is that what I expected to happen turned out to be so different from what actually occurred. This sensation is similar to searching for the cause of a failure in research. If I think carefully about it, an idea as to what to do next will eventually come to me. If things go well using this idea, that's great, but if things go bad, I can put another idea to the test. Progress can still be made with failures, though it may happen a little at a time. Conversely, if you think that to fail is common and treat research as something like a game after failing, it might prove to be enjoyable, just like raising kids.

A researcher with little experience will not have

much background in performing tests and may find it difficult to make good decisions. I would therefore like researchers to get in the practice of thinking more about their results. Having one failure after another is not good, but a success should also be questioned as to why it occurred.

Researchers should also make an effort to find topics that they find interesting. Although it is often said, it is nevertheless true that you cannot continue what doesn't interest you for a long period of time. Being a successful researcher requires that one has a clear interest in something and knows one's own strengths.

—Dr. Nakatani, what is your outlook for the future?

I envision that the technologies introduced here will be used to develop speech recognition systems that can operate smoothly even in noisy locations where many people are speaking. These might be speech recognition interfaces for public spaces such as cafés and airports and conversation recognition products for offices and home living rooms. We can expect these developments to contribute greatly to the expanded use of smartphone speech agents and communication robots. For example, I would like to see growth in technologies that can support the hearing-impaired or people speaking different languages for whom communication is difficult, and technologies that can help people connect with each other. I am committed to continuing my research efforts while keeping in mind that these future visions will be achieved not that far off—maybe no more than a few years from now!

■ **Interviewee profile**
Tomohiro Nakatani
Senior Distinguished Researcher, Supervisor, Signal Processing Research Group, Media Information Laboratory, NTT Communication Science Laboratories.

He received his B.E., M.E., and Ph.D. from Kyoto University in 1989, 1991, and 2002. He joined NTT Basic Research Laboratories in 1991 and moved to NTT Communication Science Laboratories in 2001. During 2005–2006, he was a visiting scholar at Georgia Institute of Technology, USA. Since 2008, he has been a visiting assistant professor in the Department of Media Science, Nagoya University, Aichi. His research interests include speech enhancement technologies for intelligent human-machine interfaces. He received the 1997 JSAI (Japanese Society for Artificial Intelligence) Conference Best Paper Award, the 2002 ASJ (Acoustical Society of Japan) Poster Award, the 2005 IEICE (Institute of Electronics, Information and Communication Engineers) Best Paper Award, and the 2009 ASJ Technical Development Award. During 2009–2014, he was a member of the IEEE (Institute of Electrical and Electronics Engineers) Signal Processing Society Audio and Acoustics Technical Committee (AASP-TC) and has been an associate member since 2015. He has been a member of the IEEE Signal Processing Society Speech and Language Processing TC since 2016. He served as the Chair of the review subcommittee of AASP-TC during 2013–2014, an associate editor of the IEEE Transactions on Audio, Speech, and Language Processing during 2008–2010, the Chair of the IEEE Kansai Section Technical Program Committee during 2011–2012, a Technical Program co-Chair of the IEEE WASPAA-2007, and as a member of the IEEE Circuits and Systems Society Blind Signal Processing Technical Committee during 2007–2009. He is a member of IEEE, IEICE, and ASJ.

From Information Transmission to Mutual Understanding: Paradigm Shift in the Age of Data

Eisaku Maeda

Abstract

The paradigm shift from information transmission to communication is taking place amid technological advancements in artificial intelligence (AI) and their attendant expectations. Against this background, I introduce NTT's AI-related research and development strategy and the research currently being pursued by NTT Communication Science Laboratories. I also examine the role played by communication science and present a vision of its future.

Keywords: communication, artificial intelligence, human-computer interaction

1. NTT's AI technology "corevo™"

In recent years, interest in artificial intelligence (AI) has risen sharply. AI is wrapped up in excessive expectations and disenchanted criticism. However, at the very least, new AI-related discoveries and technologies are undoubtedly transforming systems of the world. To make the right investments in the future, we must properly understand the technological foundation of AI and its latent possibilities.

At NTT, the NTT Group's AI technology is given the brand name corevo™. The name incorporates the idea of collaborating with a variety of players to bring about a revolution. However, corevo does not emulate human intelligence or thinking. Instead, it encompasses AI technologies that seek to overcome social issues and strengthen industrial competitiveness by supporting human activities and complementing and drawing out human abilities.

To develop the AI-related elemental technologies that make up corevo, NTT is leveraging its strength as a communication carrier to conduct AI research in four directions [1] (**Fig. 1**). NTT has developed groups of diverse technologies in each of the four AI directions: Agent-AI, which seeks to understand the intentions and emotions of people active at work and play and to lend assistance for contact centers and for support of the elderly; Heart-Touching-AI, which supports improvement in sports and mental wellness by enhancing the human essence; Ambient-AI, which governs intelligent sensors as the brain of IoT (Internet of Things) that lie at the core of operations such as healthcare and traffic control; and Network-AI, which supports ultra-distributed real-time processing of massive systems, for example total optimization on a global scale (**Fig. 2**). AI does not refer to several specific technologies. Instead, it should be viewed as the name of functions or services realized by combining a variety of technologies related to human beings and communication.

2. Market economics of intelligence

Undoubtedly, what led to the start of immense power being wielded by AI as a practical technology is the gathering and utilization of big data. To be sure, there are fields in which it is difficult to gather and utilize big data such as in the discovery and analysis of rare illnesses in medicine. There are also AI technologies that can demonstrate their effectiveness without big data. However, major technological tools that leverage big data, as exemplified by deep learning, are becoming critical elements of AI.

Through the process of information circulation,

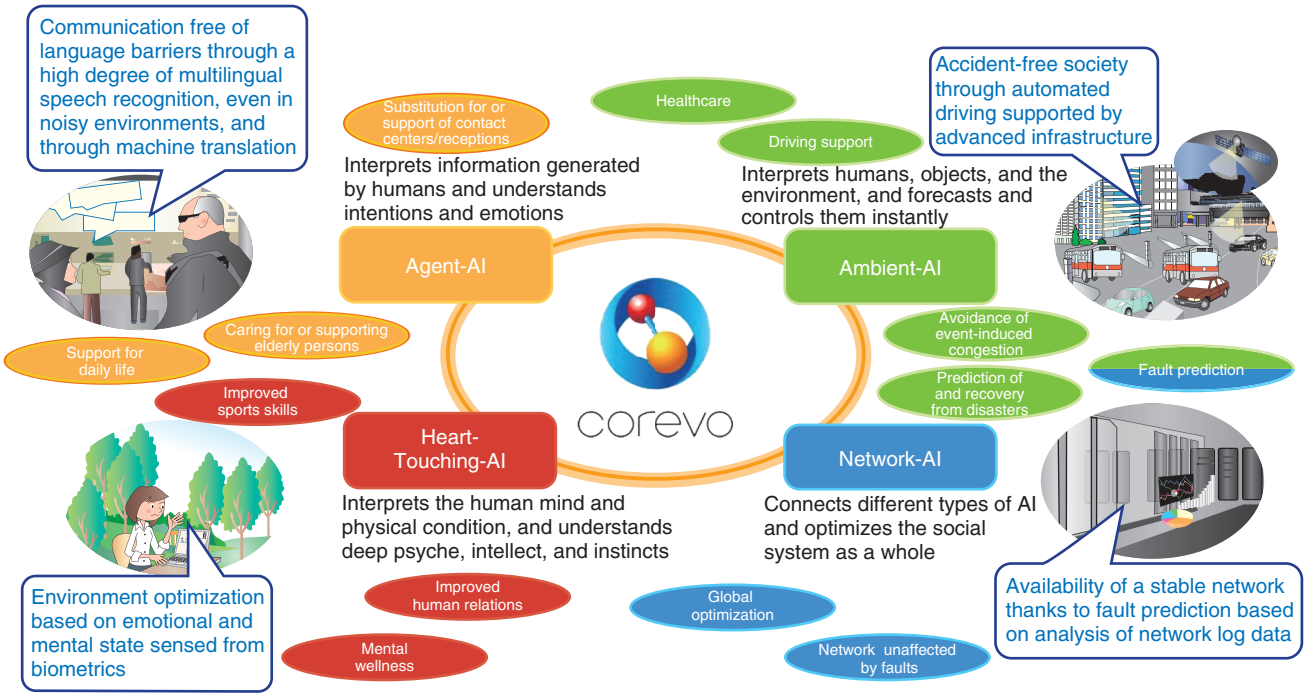
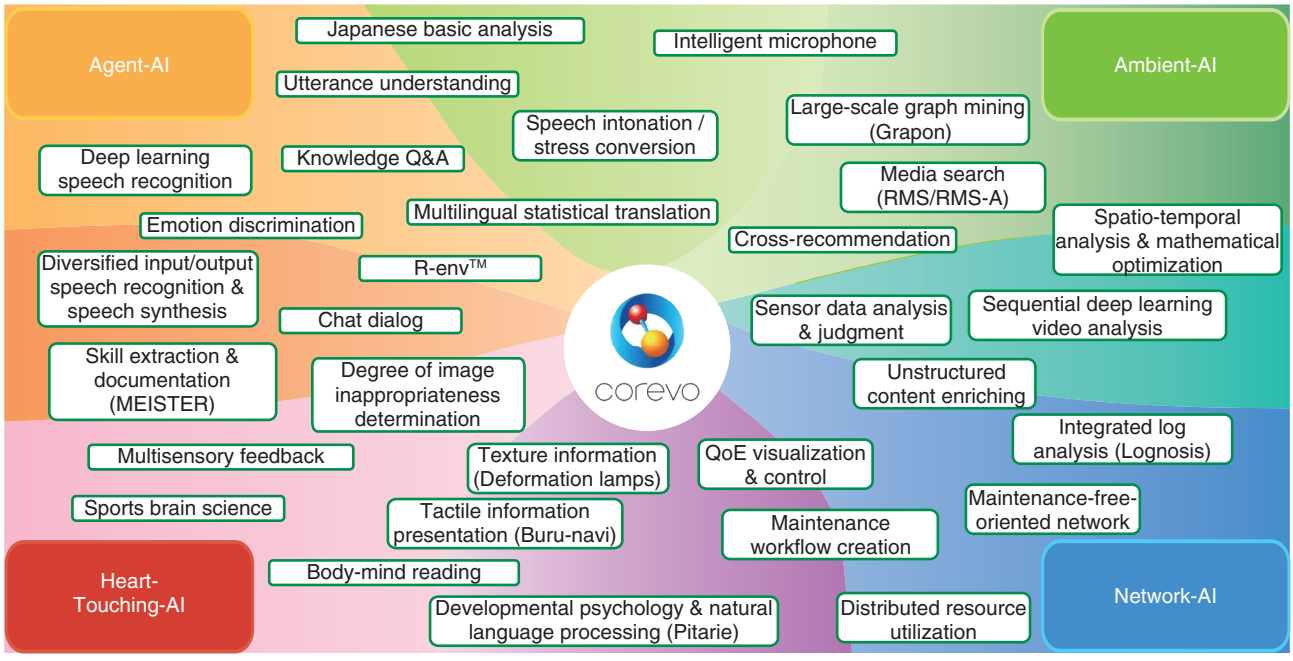


Fig. 1. Four types of AI that make up corevo™.



RMS: robust media search
QoE: quality of experience

Fig. 2. Constitutive technologies and modules of corevo™.

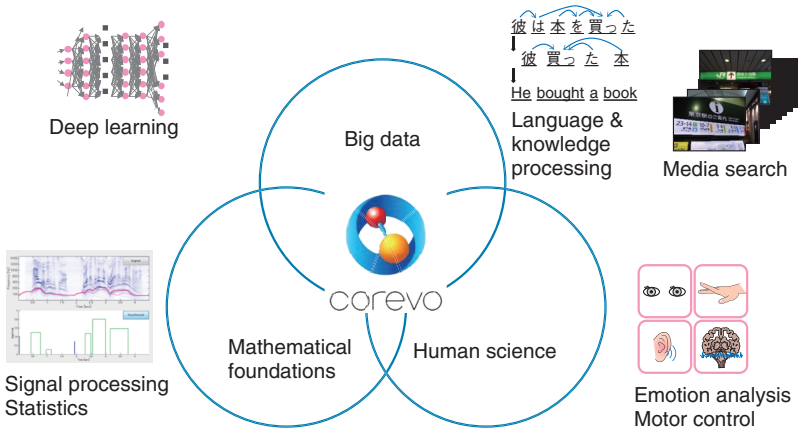


Fig. 3. Technological foundation of AI.

which involves obtaining data from just about anything (people, objects, environments) in the real world and decoding, searching (exploring), and designing (implementing) elements in the world, the results are fed back to the real world [2]. Models of intelligence will be created from all sorts of data related to people, including their written words, actions, speech, images, voices, vital statistics, and muscle and brain activity. Through this process, intelligence in a variety of forms is broken down and componentized. Sooner or later, an era in which these components are treated as products in the market economy will surely arrive. At that time, the product cost of intelligence as components will be influenced by the cost of gathering and analyzing big data.

3. Exploring the essence of communication

Research on communication involving human beings is becoming ever more important. This includes research that forms the foundation of communication transmission and information, for example, research on voice communication to make VoLTE (voice over Long Term Evolution) a practical reality [3] and research on extracting necessary knowledge from vast amounts of information [4]. Other areas include research on machine intelligence that handles information in place of humans [5], research on how human beings transmit and receive information between two parties [6], research on diverse communication styles among people [7], and research on mechanisms to support people [8].

Research on AI in the broad sense is composed of technologies that act as substitutes for human intelligence and technologies that enhance human intelligence. The three major areas for advancing research in these technologies are big data, mathematical foundations, and human sciences (**Fig. 3**). AI technologies in diverse forms, from practical voice recognition and machine translation to sensing technologies that can decode human emotions and actions, are being incorporated into real life. In the midst of this trend, the time has come to reconsider the essence of *communication*. The environment surrounding human beings is greatly changing despite fixed constraints; the flow of physical time is constant, and the lifespan of human beings as organisms remains largely unchanged. I wish to sketch out a design plan of the future that broadly and deeply explores the essence of communication between person and person, and person and machine.

References

[1] Feature Articles: "Artificial Intelligence Research Activities in NTT Group," NTT Technical Review, Vol. 14, No. 5, 2016. https://www.ntt-review.jp/archive/2016/201605.html
[2] E. Maeda, "Embracing Information Science and Technology—Decoding, Exploring, and Designing the World," NTT Technical Review, Vol. 13, No. 11, 2015. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201511fa1.html
[3] Y. Kamamoto, T. Moriya, and N. Harada, "Transmission of High-quality Sound via Networks Using Speech/Audio Codecs," NTT Technical Review, Vol. 14, No. 11, 2016. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201611fa2.html
[4] M. Blondel, A. Fujino, and N. Ueda, "Learning from a Large Number of Feature Combinations," NTT Technical Review, Vol. 14, No. 11, 2016. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201611fa3.html

[5] H. Sugiyama, R. Higashinaka, and T. Meguro, "Towards User-friendly Conversational Systems," NTT Technical Review, Vol. 14, No. 11, 2016.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201611fa4.html
[6] Y. Okumura, T. Kobayashi, and Y. Oshima-Takane, "Child Language Development: The Differences between Japanese and English," NTT Technical Review, Vol. 14, No. 11, 2016.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201611fa5.html
[7] T. Kimura, K. Takahashi, D. Mikami, and M. Kashino, "Elucidating the Brain Processing Mechanisms of Athletes Using Virtual Reality Technology," NTT Technical Review, Vol. 14, No. 11, 2016.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201611fa6.html
[8] N. Yamashita, "How Tracking Technologies Improve Family Communication," NTT Technical Review, Vol. 14, No. 11, 2016.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201611fa7.html

Eisaku Maeda
Vice President, Head of NTT Communication Science Laboratories.
He received a B.E. and M.E. in biological science and a Ph.D. in mathematical engineering from the University of Tokyo in 1984, 1986, and 1993. He joined NTT in 1986. He was a guest researcher at the University of Cambridge, UK, in 1996–1997. He was awarded the Information Processing Society of Japan (IPSJ)'s 45th anniversary best paper on the next 50 years of information science and technology for his paper "Resurgence of Fairies and Goblins—A Proposal for the New Vision of 'Ambient Intelligence.'" His research interests are statistical machine learning, intelligence integration, and bioinformatics. He is a fellow of IEICE (Institute of Electronics, Information and Communication Engineers) and a senior member of IEEE (Institute of Electrical and Electronics Engineers) and IPSJ.

Transmission of High-quality Sound via Networks Using Speech/Audio Codecs

Yutaka Kamamoto, Takehiro Moriya, and Noboru Harada

Abstract

This article describes two recent advances in speech and audio codecs. One is EVS (Enhanced Voice Service), the new standard by 3GPP (3rd Generation Partnership Project) for speech codecs, which is capable of transmitting speech signals, music, and even the ambient sound on the speaker's side. This codec has been adopted in a new VoLTE (voice over Long-Term Evolution) service with enhanced high-definition voice (HD+), which provides us with clearer and more natural conversations than conventional telephony services such as with fixed-line/land-line and 3G mobile phones. The other is MPEG-4 Audio Lossless Coding (ALS) standardized by the Moving Picture Experts Group (MPEG), which makes it possible to transmit studio-quality audio content to the home. ALS is expected to be used by some broadcasters, including IPTV (Internet protocol television) companies, in their broadcasts in the near future.

Keywords: audio/speech coding, data compression, international standards

1. Introduction

Many audio and speech codecs are available, and we can select the most suitable one for different usage scenarios ranging from those requiring reasonable quality with low bit rates to ones demanding original signal quality with high bit rates. With the increases in network capacity that have been achieved, content that requires high bit rates such as 4K television (TV) and high-resolution audio can also be transmitted. However, the first priority is to transmit speech signals in ordinary telephony without congestion. Therefore, speech codecs for telephony should use as low a bit rate as possible. In addition, they must have lower algorithmic delay because the longer the codec delay is, the more difficult it becomes for people to communicate with each other.

In contrast, one-way transmission such as broadcasting is less sensitive to delay. Most audio codecs utilize the advantages of longer delay and then efficiently compress audio signals by means of signal processing with sufficient frame length. Moreover, speech codecs use a human phonation model, so they are not suitable for music. When clean speech items are coded by audio codecs at low bit rates, we get the impression that a machine is talking. Speech and audio compression schemes have these kinds of trade-offs. To achieve the best quality of speech and music content with less delay, experts in speech and audio coding around the world have been working together to develop new codecs. Furthermore, lossless coding, which refers to compression without any loss, has been standardized to ensure that the quality of the original content is maintained.

This article presents an overview of the 3rd Generation Partnership Project (3GPP) Enhanced Voice Services (EVS) codec, which has been newly standardized for mobile communications, and MPEG-4 Audio Lossless Coding (ALS) for high-resolution audio, which was standardized by the Motion Picture Experts Group (MPEG).

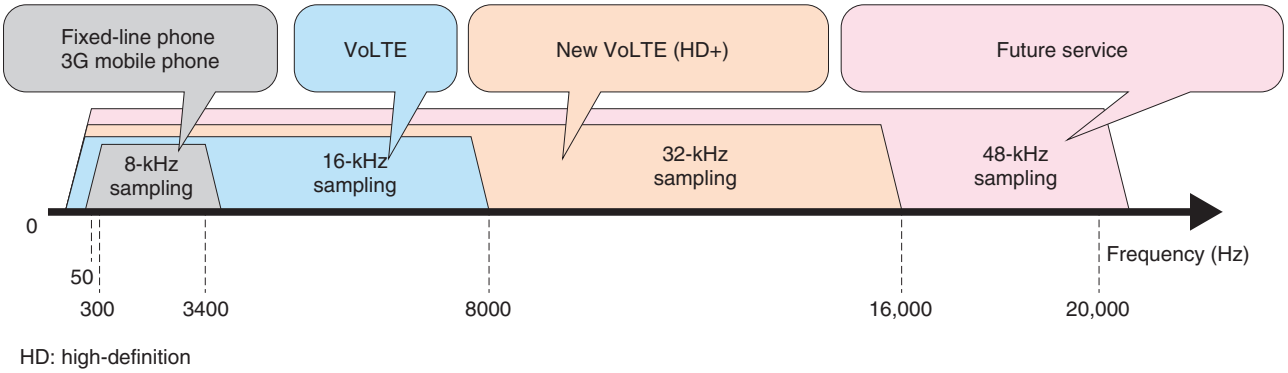


Fig. 1. Supported audio bandwidth in EVS.

2. 3GPP EVS codec for mobile communications

3GPP is the international standardization consortium for mobile communications. It has newly defined EVS speech and audio coding standards for voice over Long-Term Evolution (VoLTE) [1, 2]. Conventional speech coding schemes for mobile phones have been based on code excited linear prediction (CELP). These schemes have utilized a human voice production model and achieved high-quality speech transmission with very low bit rates. EVS consists of newly developed low-delay and low-bit-rate audio coding modules in addition to CELP, and it achieves high-quality transmission of various types of input signals, including speech, audio, background noise, and background music [3, 4].

EVS uses new bandwidth extension technologies to support signals with higher sampling rates up to 48 kHz, in contrast to the narrowband signal (8-kHz sampling rate) of conventional fixed-line/land-line telephones and 3G mobile phones and the wideband signal (16-kHz sampling rate) of VoLTE. Note that the wideband signal is used for AM (amplitude modulation) radio, the super-wideband signal (32-kHz sampling rate) is used for FM (frequency modulation) radio, and the full-band signal (48-kHz sampling rate) is used for digital broadcasting, as shown in **Fig. 1**.

EVS has been optimized for VoLTE with a frame length of 20 ms and algorithmic delay of 32 ms. It has been designed to minimize perceptual distortion against packet loss, whereas coding schemes for conventional 3G mobile phones were optimized for robustness against bit errors. In addition, EVS covers a wide range of bit rates from 5.9 kbit/s to 128 kbit/s and enables frame-by-frame selection of bit rates. This enables smooth migration from the conventional VoLTE system since EVS has inter-operability with AMR-WB (Adaptive Multi-Rate Wideband).

During the standardization process, a huge number of subjective quality evaluations were conducted for various coding conditions, input items, and languages. The results of the evaluations indicated that EVS outperformed conventional speech and audio coding schemes in terms of quality [5]. NTT used a similar procedure to conduct listening tests on Japanese materials [6]. The results of the tests confirmed the superiority of EVS over the coding schemes for conventional mobile communications systems.

Note that all EVS development has been carried out by 12 organizations[*1] based mainly in Europe, North America, and East Asia, including Japanese companies. EVS has been deployed in commercial services such as VoLTE (HD+) by NTT DOCOMO since the summer of 2016 [7] and by some operators in the USA and Europe as well [8]. We believe that EVS will allow billions of people around the world to enjoy high-quality communication in the near future.

3. MPEG-4 ALS for high-resolution audio transmission

MPEG has standardized many useful audio and video codecs that are commonly used in daily life. The MPEG audio subgroup standardized the lossless compression scheme MPEG-4 ALS, which can perfectly reconstruct original signals. NTT is one of the

*1 In alphabetical order, Fraunhofer IIS, Huawei Technologies Co. Ltd, Nokia Corporation, NTT, NTT DOCOMO, INC., ORANGE, Panasonic Corporation, Qualcomm Incorporated, Samsung Electronics Co., Ltd., Telefonaktiebolaget LM Ericsson, VoiceAge Corporation, and ZTE Corporation.

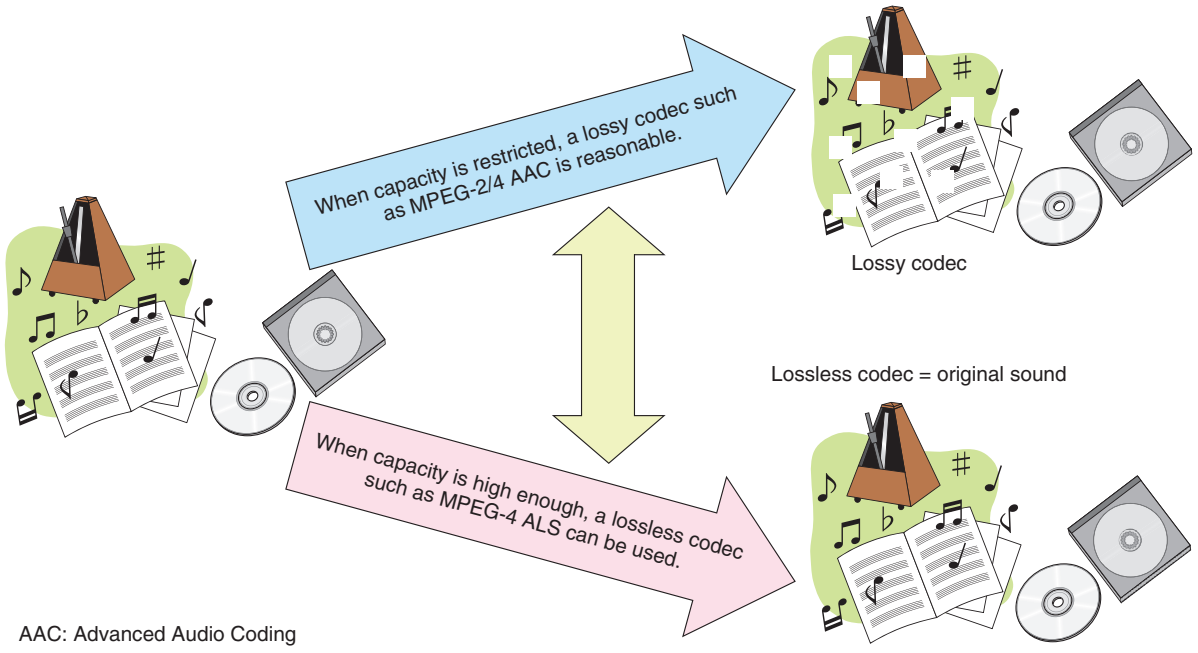


Fig. 2. Selection of audio codec according to the allowed bit rate.

contributors to this standard [9, 10]. Although the compression ratio depends on the input signal, a file is normally compressed to around 30% to 70% of its original size [11, 12]. High-resolution audio*2 has become more popular recently and enables precisely digitized music. Lossy codecs often cannot convey the fidelity of high-resolution audio, so a lossless codec such as MPEG-4 ALS is necessary.

For example, the audio signal in TV broadcasting is usually produced in a studio or broadcasting station in a 48-kHz, 24-bit format, which is referred to as high-resolution audio. Since access to radio waves is limited, we cannot assign a high bit rate to the content, and it is necessary to compress the audio signal at a loss of quality. This lossy codec enables us to enjoy broadcasting content because the codec reduces the bit rate remarkably without any noticeable difference from the original signal.

The ultrahigh-definition (or super-high-definition) TV service called 4K/8K TV has now started, and it uses very high bit rates. The audio signal is also expected to use higher bit rates. The Association of Radio Industries and Businesses (ARIB), which defines the standards for radio-wave systems in Japan, standardized ARIB STD-B32 for the 4K/8K TV system. This standard enables the use of MPEG-4 ALS as one of the audio codecs [14]. MPEG-4 ALS can reconstruct in the home music content that was produced in a broadcasting studio—with the quality of the original signal—because the lossless codec guarantees bit-exactness over the entire transmission. We can enjoy high-resolution audio in our living rooms when a sufficient bit rate is assigned to the audio signal (**Fig. 2**). IPTV (Internet protocol TV) services, which use optical-fiber lines, may introduce MPEG-4 ALS before the radio-wave services do.

In order to support practical deployment of MPEG-4 ALS, NTT has prepared related standards such as MPEG-4 ALS Simple Profile and IEC 61937-10 Edition 2 by the International Electrotechnical Commission (IEC) [15]. MPEG-4 ALS Simple Profile restricts the parameters of the input signal such as the sampling frequency, number of channels, bit depth, and frame size, and also restricts some processing tools and functionalities that require higher computational complexity such as compression for floating-point format signals. ARIB STD-B32 recommends the use of MPEG-4 ALS Simple Profile with LATM/LOAS (Low-overhead Audio Transport Multiplex/Low Overhead Audio Stream)*3 capsuling. To facilitate

*2 High-resolution audio: The Japan Electronics and Information Technology Industries Association (JEITA) defines high resolution audio as an audio signal with a sampling frequency higher than 48 kHz and a bit depth greater than 16 bits [13].

*3 LATM/LOAS: A transmission scheme of the header and stream defined in MPEG-4 Audio.

the connection of a TV and a digital audio device, IEC 61937-10 Edition 2 newly supports the bitstream of MPEG-4 ALS Simple Profile with LATM/LOAS, which can be transmitted via radio waves with 4K/8K video. Then, with 4K/8K TV, we can listen to high-quality music through a high-resolution TV or by using a digital amplifier that can decode MPEG-4 ALS.

4. Future work

The EVS codec enables high-quality communications by means of speech and music even when delay and bit rates are low. MPEG-4 ALS can transmit original audio content losslessly. The delivery of high-quality music has now been achieved, so we will start to consider ways to achieve even more realistic audio transmission. Basic research on interactive communication may achieve a synergistic effect between live venues and reception sites. We will continue to develop speech and audio codec schemes to make timely contributions to new services.

References

[1] 3GPP TS26.441: Codec for Enhanced Voice Services (EVS); General overview, 2015.

[2] 3GPP TS26.445: Codec for Enhanced Voice Services (EVS); Detailed algorithmic description, 2015.

[3] M. Dietz, M. Multrus, V. Eksler, V. Malenovsky, E. Norvell, H. Pobloth, L. Miao, Z. Wang, L. Laaksonen, A. Vasilache, Y. Kamamoto, K. Kikuiri, S. Ragot, J. Faure, H. Ehara, V. Rajendran, V. Atti, H. Sung, E. Oh, H. Yuan, and C. Zhu, "Overview of the EVS Codec Architecture," Proc. of the 40th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2015), pp. 5698–5702, Brisbane, Australia, Apr. 2015.

[4] S. Bruhn, H. Pobloth, M. Schnell, B. Grill, J. Gibbs, L. Miao, K. Järvinen, L. Laaksonen, N. Harada, N. Naka, S. Ragot, S. Proust, T. Sanda, I. Varga, C. Greer, M. Jelínek, M. Xie, and P. Usai, "Standardization of the New 3GPP EVS Codec," Proc. of ICASSP 2015, pp. 5703–5707, Brisbane, Australia, Apr. 2015.

[5] 3GPP TR26.952: Codec for Enhanced Voice Services (EVS); Performance characterization, 2015.

[6] Y. Kamamoto, T. Moriya, and N. Harada, "Subjective Evaluation of 3GPP EVS Codec for Japanese Speech Items," The Acoustical Society of Japan 2016 Spring Meeting, 3-2-9, 2016 (in Japanese).

[7] Website of NTT DOCOMO, VoLTE/VoLTE (HD+) service (in Japanese),
https://www.nttdocomo.co.jp/support/area/volte/

[8] Press release issued by Ericsson on June 14, 2016.
https://www.ericsson.com/news/160614-evolved-hd0-voice-for-volte_244039855_c

[9] ISO/IEC14496-3: Information Technology—Coding of Audio-visual Objects—Part 3: Audio, fourth edition, 2009.

[10] T. Liebchen, T. Moriya, N. Harada, Y. Kamamoto, and Y. A. Reznik, "The MPEG-4 Audio Lossless Coding (ALS) Standard - Technology and Applications," Proc. of the 119th Audio Engineering Society Convention, New York, USA, Oct. 2005.

[11] Y. Kamamoto, T. Moriya, N. Harada, and C. Kos, "Enhancement of MPEG-4 ALS Lossless Audio Coding," NTT Technical Review, Vol. 5, No. 12, 2007.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr200712sp2.html

[12] N. Harada, T. Moriya, and Y. Kamamoto, "MPEG-4 ALS: Performance, Applications, and Related Standardization Activities," NTT Technical Review, Vol. 5, No. 12, 2007.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr200712sp3.html

[13] JEITA, "Definition of High-resolution Audio," Mar. 2014 (in Japanese).
http://home.jeita.or.jp/page_file/20140328095728_rhsiN0Pz8x.pdf

[14] ARIB STD-B32: Video Coding, Audio Coding, and Multiplexing Specifications for Digital Broadcasting, 2011.

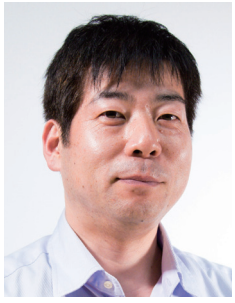
[15] IEC 61937-10: Digital Audio - Interface for Non-linear PCM Encoded Audio Bitstreams Applying IEC60958 – Part 10: Non-linear PCM Bitstreams According to the MPEG-4 Audio Lossless Coding (ALS) Format, 2011.

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

He received a B.S. in applied physics and physico-informatics from Keio University, Kanagawa, in 2003 and an M.S. and Ph.D. in information physics and computing from the University of Tokyo in 2005 and 2012. Since joining NTT Communication Science Laboratories in 2005, he has been studying signal processing and information theory, particularly speech/audio coding and lossless compression of time-domain signals. He was also with NTT Network Innovation Laboratories from 2009 to 2011, where he developed the audio-visual codec for ODS (Other Digital Stuff/Online Digital Source). He has contributed to the standardization of coding schemes for MPEG-4 ALS, the International Telecommunication Union - Telecommunication Standardization Sector (ITU-T) Recommendation G.711.0: Lossless compression of G.711 pulse code modulation, and 3GPP EVS. He received the Telecom System Student Award from the Telecommunications Advancement Foundation (TAF) in 2006, the IPSJ Best Paper Award from the Information Processing Society of Japan (IPSJ) in 2006, the Telecom System Encouragement Award from TAF in 2007, the Awaya Prize Young Researcher Award from the Acoustical Society of Japan (ASJ) in 2011, and the Technical Development Award from ASJ in 2016. He is a member of IPSJ, ASJ, the Institute of Electronics, Information and Communication Engineers (IEICE), and the Institute of Electrical and Electronics Engineers (IEEE).

Takehiro Moriya
NTT Fellow, Head of Moriya Research Laboratory, NTT Communication Science Laboratories.

He received his B.S., M.S., and Ph.D. in mathematical engineering and instrumentation physics from the University of Tokyo in 1978, 1980, and 1989. Since joining NTT laboratories in 1980, he has been conducting research on medium- to low-bit-rate speech and audio coding. In 1989, he worked at AT&T Bell Laboratories as a visiting researcher. He has contributed to the standardization of coding schemes for the Japanese Public Digital Cellular System, ITU-T, ISO/IEC MPEG, and 3GPP. He has received a number of awards including the IEEE James L. Flanagan Speech and Audio Processing Award in 2016. He is an IEEE Fellow, IEICE Fellow, and a member of IPSJ and ASJ. He is also a member of the IEEE Speech Technical Committee of the Signal Processing Society (SPS) and a chair of the IEEE SPS Tokyo Joint Chapter.



Noboru Harada
Senior Research Scientist, Supervisor, Moriya Research Laboratory, NTT Communication Science Laboratories.

He received his B.S. and M.S. in computer science and systems engineering from Kyushu Institute of Technology, Fukuoka, in 1995 and 1997. He joined NTT in 1997. His main research area is lossless audio coding, high-efficiency coding of speech and audio, and their applications. He was also with NTT Network Innovation Laboratories from 2009 to 2011, where he developed the audio-visual codec for ODS. He is an editor of numerous international standards specifications including ISO/IEC 23000-6:2009 Professional Archival Application Format and ITU-T G.711.0 and has contributed to the standardization of MPEG-4 ALS and 3GPP EVS. He is a member of IEICE, ASJ, Audio Engineering Society (AES), and IEEE.

Learning from a Large Number of Feature Combinations

Mathieu Blondel, Akinori Fujino, and Naonori Ueda

Abstract

Second-order polynomial regression can often outperform simple linear regression by making use of feature combinations. However, when the number of feature combinations is large, second-order polynomial regression quickly becomes impractical. In this article, we present convex factorization machines, a new technology developed by NTT Communication Science Laboratories, which can cope with a large number of feature combinations and guarantees globally optimal model parameters.

Keywords: machine learning, regression analysis, feature combinations

1. Introduction

With the democratization of the Internet, social media, and connected devices, the amount of data that can be used for scientific or business purposes is ever growing. In this context, machine learning has recently attracted a lot of attention due to its ability to leverage large amounts of data for predictive analytics. In particular, regression analysis is a frequently used predictive technology in machine learning.

We present regression analysis by using house price prediction as a running example (**Fig. 1**). House price is typically determined by numerous features such as whether the house is detached or terraced (adjoined to other homes), the number of rooms, and whether it has a garden. We can use regression analysis to obtain from past examples of sold houses an equation that relates these features to the house price. In linear regression, the relationship between the features $x = (x_1,...,x_d)$ and the house price y is modeled by $y = \sum_{\{j=1\}}^{d} w_j x_j = w^T x$, where $w = (w_1,...,w_d)$ is a weight vector estimated from previously sold houses. By inspecting the estimated weights, we can infer what features influence house price the most. In addition, by using the aforementioned model equation, we can predict the price of new houses, given their features.

However, while linear regression is very simple, it has some limitations. For example, while the price of both detached and terraced houses decreases with distance from the city center, we expect the price of terraced houses to decrease faster than that of detached houses. In this case, since linear regression estimates a weight for the distance from city center independently of whether a house is detached or terraced, it cannot achieve high predictive accuracy. To solve this problem, it is necessary to estimate different weights for the distance to the city center, depending on whether a house is detached or terraced. In other words, it is necessary to introduce feature combinations in the model equation. This is called second-order polynomial regression.

Second-order polynomial regression can estimate models that fit the data better than linear regression. However, because the number of feature combinations is quadratic in the number of features, the number of feature combinations can quickly explode. For example, in genomic selection, which is the task of predicting grain yield from the DNA (deoxyribonucleic acid) of cereal plants, the number of genes is very large, and therefore, using feature combinations in the model equation can become impractical. Factorization machines (FM) [1] are a recently proposed method that can deal with a large number of feature combinations. Unfortunately, with FM, the quality of the estimated model strongly depends on the parameter initialization. To address this issue, we at NTT Communication Science Laboratories developed convex factorization machines (CFM), a new

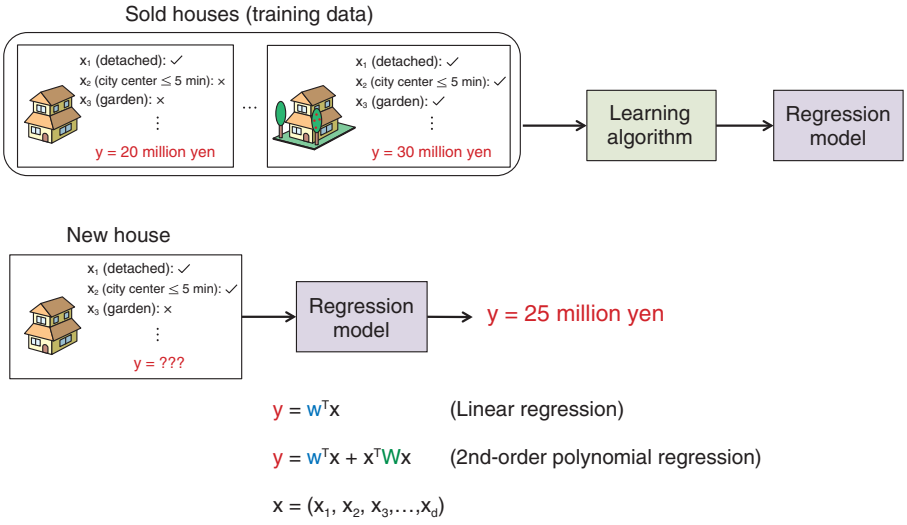


Fig. 1. Application of regression analysis to house price prediction.

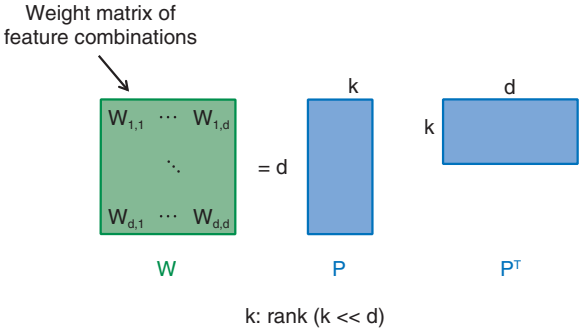


Fig. 2. Matrix decomposition obtained by the original FM.

technology that can both cope with a large number of feature combinations and guarantees a globally optimal model regardless of the initialization [2].

2. CFM

In second-order polynomial regression, the relationship between house features and house price is modeled by the equation $y = w^Tx + x^TWx$, where again, w is a weight vector, and W is a matrix whose elements correspond to the weights of feature combinations. When the number of features d is large, estimating W can quickly become impractical because W is a d x d matrix. To address this issue, both CFM and the original FM reduce the number of parameters to be estimated by assuming that W is a low-rank matrix. With the original FM, W is replaced by PP^T, where P is a d x k matrix (k << d) and k is a user-defined rank hyper-parameter. The original FM then use training data to estimate P instead of W (**Fig. 2**). However, because the estimation of P involves a non-convex optimization problem, the quality of the obtained parameters greatly depends on the initialization. In practice, it is therefore necessary to try different initializations in order to obtain good results.

In contrast, our proposed technology, CFM, is guaranteed to obtain globally optimal model parameters regardless of the initialization. We developed an efficient algorithm to learn W in eigendecomposition form. We can use our algorithm to estimate the k eigenvalue-eigenvector pairs of W (**Fig. 3**). In addition, our algorithm automatically determines the rank

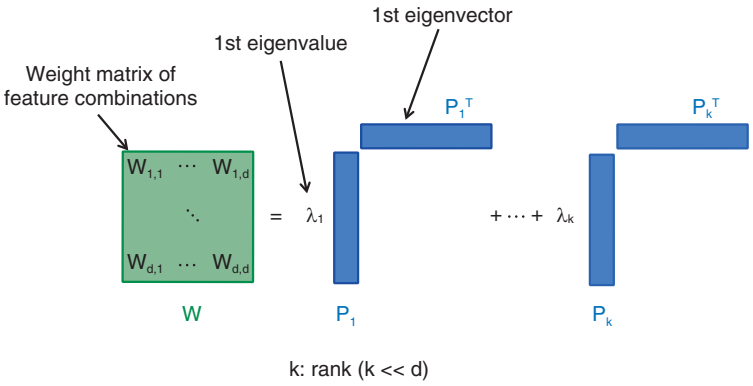


Fig. 3. Eigendecomposition obtained by CFM.

Table 1. Example of application of CFM to genomic selection.

	2nd-order polynomial regression	FM	CFM
Wheat 1	0.397	0.376	0.402
Wheat 2	0.471	0.501	0.526
Rice	0.660	0.656	0.662

k of W from data.

In **Table 1**, we empirically compare ordinary second-order polynomial regression (i.e., without a low-rank constraint), FM, and CFM on genomic selection (the task of predicting grain yield from the DNA of cereal plants). The values in the table indicate the Pearson correlation between the true grain yield and the grain yield predicted by the three methods (higher is better) on test data. Results for FM were obtained by trying several possible initializations. These results show that CFM can achieve higher predictive accuracy than FM. In addition, the CFM results are also better than those for ordinary second-order polynomial regression. In machine learning, it is generally known that a model can overfit the data if the number of parameters is too large. We believe that CFM can mitigate this issue thanks to the reduced number of parameters to be estimated.

An important property of the low-rank constraint used in FM and CFM is that it enables the weights of feature combinations that were not observed in the training set to be estimated. This property is particularly useful in implementing recommender systems, a domain where FM have been particularly popular in recent years.

3. Higher-order extensions

We presented CFM, a new technology capable of efficiently leveraging second-order feature combinations. To further improve predictive accuracy, it is sometimes useful to consider third-order or higher-order feature combinations. We recently proposed new efficient algorithms for this purpose [3, 4].

References

[1] S. Rendle, "Factorization Machines," Proc. of ICDM 2010 (the 10th IEEE International Conference on Data Mining), pp. 995–1000, Sydney, Australia, Dec. 2010.
[2] M. Blondel, A. Fujino, and N. Ueda, "Convex Factorization Machines," Proc. of ECML PKDD (European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases) 2015, Vol. 9285, pp. 19–35, Porto, Portugal, Sept. 2015.
[3] M. Blondel, M. Ishihata, A. Fujino, and N. Ueda, "Polynomial Networks and Factorization Machines: New Insights and Efficient Training Algorithms," Proc. of ICML 2016 (the 33rd International Conference on Machine Learning), pp. 850–858, New York, USA, June 2016.
[4] M. Blondel, A. Fujino, N. Ueda, and M. Ishihata, "Higher-order Factorization Machines," Proc. of NIPS 2016 (the 30th Annual Conference on Neural Information Processing Systems), to appear, Barcelona, Spain, Dec. 2016.

Mathieu Blondel

Research Scientist, Ueda Research Group, NTT Communication Science Laboratories.

He received an engineering diploma from Telecom Lille, France, in 2008 and a Ph.D. in engineering from Kobe University, Hyogo, in 2013. He joined NTT Communication Science Laboratories in 2013. His current research interests include machine learning, mathematical optimization, the design of efficient machine learning software, and the application of these areas to real-world applications.

Akinori Fujino

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

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

Naonori Ueda

Head of Ueda Research Laboratory, NTT Communication Science Laboratories and NTT Fellow.

He received his B.S., M.S., and Ph.D. in communication engineering from Osaka University in 1982, 1984, and 1992. He joined Yokosuka Electrical Communication Laboratories of Nippon Telegraph and Telephone Public Corporation (now NTT) in 1984. In 1994, he moved to NTT Communication Science Laboratories in Kyoto, where he has been researching statistical machine learning, Bayesian statistics, and their applications to web data mining and big data analysis. From 1993 to 1994, he was a visiting scholar at Purdue University, Indiana, USA. He is a guest professor at the National Institute of Informatics and a visiting professor at Kyoto University. He is a Fellow of the Institute of Electronics, Information and Communication Engineers and a member of the Information Processing Society of Japan and the Institute of Electrical and Electronics Engineers. In July 2013, he became Director of Machine Learning Data Science Center, after serving as Director of NTT Communication Science Laboratories for three years. Since April, 2016, he has been head of Ueda Research Laboratory, NTT Communication Science Laboratories, and an NTT Fellow.

Towards User-friendly Conversational Systems

Hiroaki Sugiyama, Ryuichiro Higashinaka, and Toyomi Meguro

Abstract

Remarkable progress has been made with conversational systems in recent years, and they are becoming much more common. However, many problems remain to be solved such as errors in speech recognition and the narrow range of tractable dialogue topics. In this article, we introduce our efforts to improve the dialogue quality of our dialogue systems and to prevent dialogue breakdown using multiple dialogue robots.

Keywords: conversation, dialogue robots, dialogue breakdown detection

1. Introduction

Many robots and applications have been developed recently that are designed to converse with people. A few years ago, most conversational systems were implemented on smartphone applications. Some companies working in this field, notably Pepper (SoftBank Group Corp.) and OHaNAS (TOMY Company, Ltd.), changed direction and began developing technology to enable conversations between people and robots. Robots that can communicate with people through conversation are expected to be used as a natural interface between people and information and also as a way to improve human communication skills.

Yet how fluently can such conversational systems talk with people? People who have actually had conversations with them may have been disappointed if the robots output meaningless utterances because the robots did not understand some aspects of the human voice or did not have detailed knowledge of certain dialogue topics.

Current (especially commercial) conversational systems are developed with many hand-crafted response rules assuming that correct texts are obtained from speech recognition. This approach enables us to create appropriate and interesting response rules for frequently used user utterances. Furthermore, we can reduce the cost of developing such rules by dissociating textual appropriateness from speech recognition performance. However, it is obvious that not all of the topics of user utterances are covered by hand-crafted rules, and inappropriate system utterances can be generated when speech recognition fails. In this article, we introduce our recent work to overcome these problems.

2. Automatic response generation for various topics

Rule-based utterance generation is widely used in conversational systems. In this method, we first construct a dialogue example database that consists of pattern-response utterance pairs (called *rules*). The rules are created manually or gathered from actual dialogues. Then, a system applying this approach retrieves patterns that match a user utterance and outputs responses associated with the retrieved patterns. This rule-based approach works well when the range of dialogue topics is narrow. For example, recent rule-based systems such as A.L.I.C.E. (Artificial Linguistic Internet Computer Entity) have repeatedly won the Loebner Prize (an artificial intelligence competition for chatterbots). However, to generate utterances for conversational systems, the huge variety of topics in conversations means that substantial

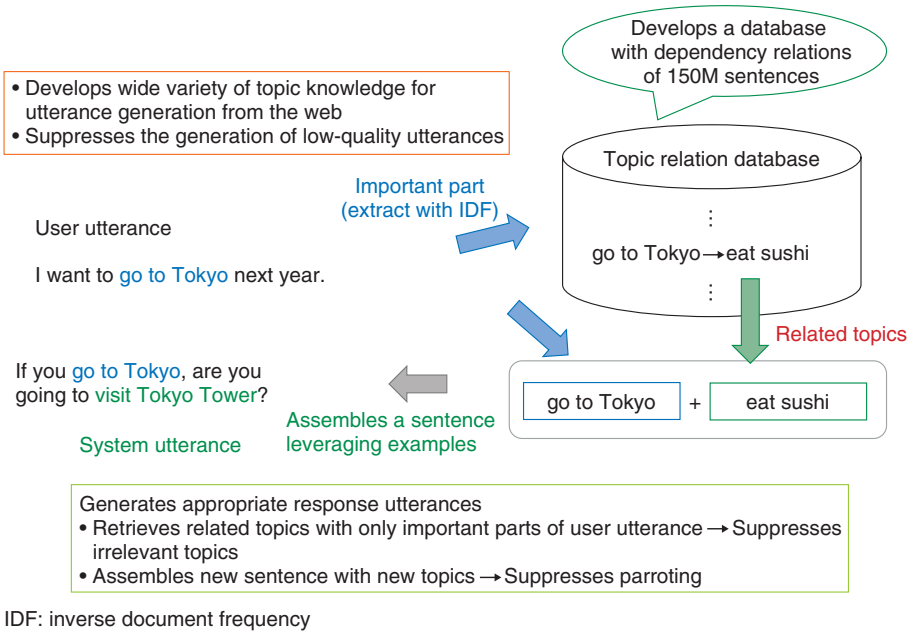


Fig. 1. New utterance generation method.

resources are required to build enough rules to cover all topics and to maintain the developed rules without contradiction.

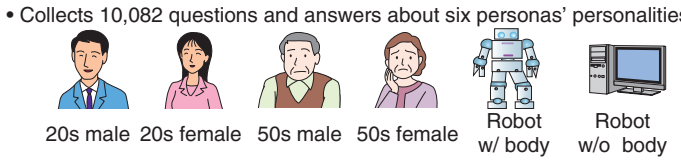
To make it feasible to automatically generate system utterances that are relevant to such a wide variety of topics of user utterances, a retrieval-based approach has been proposed. This approach retrieves sentences from the web or microblogs as system utterances by word matching with user utterances. This approach can generate responses relevant to user utterances by leveraging a wide variety of topics of web articles. However, since the retrieved sentences include the inherent contexts of the document in which the sentences originally appeared, the retrieved sentences may contain information that is irrelevant to user utterances.

To automatically define the relevancy between topics, we utilize dependency relations that express more specific relationships than normal co-occurrence. We propose an utterance generation method that combines two strongly related semantic units (phrase pairs with dependency relations that represent the topics of utterances) to create a system utterance; here, the first semantic unit is the one found in the user utterance, and the second semantic unit is the one that has a dependency relation with the first one in a large text corpus (**Fig. 1**) [1]. Our method generates utterances that have new information relevant to the current topics, which makes it easier for users to continue talking about the topic than with conventional methods.

3. Design of system *personality*

Using the method explained above, we can automatically obtain system utterances that relate to user utterances. However, is the development of conversational systems all that is required? Actually, in conversations, people often ask questions related to the specific personality or characteristics of the person with whom they are talking, for example, questions about their favorite foods or their experience playing sports. Such personality questions have reportedly appeared in conversations with task-oriented dialogue systems [2]; therefore, it is necessary to respond to such questions to achieve conversational systems. However, these questions cannot be answered with the prior utterance generation method. Moreover, if we develop question-answering systems based on information on the web, it will be difficult to maintain consistency among answers.

Therefore, we developed a question-answering system for questions that ask about an agent's specific personality, using manually created large-scale question-answer pairs. We first developed a Person Database (PDB) with large-scale personality question-answer



Question	Question category	Topics	Persona	Answer
What color do you like?	Favorite color	Favorite color	20s female	Pink
What's your favorite color?			50s male	Green
Where do you go on trips?	Where you want to travel to	Places you want to go	50s female	Bhutan
What countries do you go to?			50s male	France
Where do you go on domestic trips?	Where you go on domestic trips		20s female	Azumino

Fig. 2. Development of Person Database (PDB).

pairs for six personas gathered from many questioners and a few answerers and categorized the questions manually (**Fig. 2**) [3]. Our question-answering system responds to about 60% of personality questions and improves user satisfaction of dialogues.

4. Implementation of dialogue systems in actual robots

We have so far developed conversational systems for the text-chat format, but it is becoming more popular to use robots with speaking capabilities as a dialogue interface. To examine how our system can talk with people naturally, we collaborated with Professor Hiroshi Ishiguro at Osaka University and implemented our conversational systems in Geminoids, which are robots with human-like appearance. The architecture of this system is as follows. The system first captures user voices with a microphone and converts the voices to text using speech recognition technology. Our conversational system generates response texts for the user utterances. Finally, the text-to-speech system converts the response texts to system voices. We demonstrated this robot system at a well-known event called South by South West (SXSW) [4], and on a TV program titled "Matsuko x Matsuko."

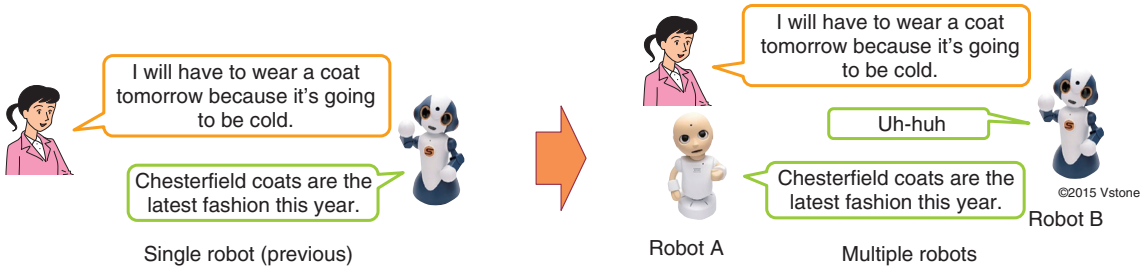
When we talked with the robot using only voice, the robot sometimes gave inappropriate utterances because of speech recognition errors, which was as we expected. Moreover, if a user said multiple sentences to the robot in rapid succession, the robot was unable to keep up with the user utterances. These difficulties were also expected. In contrast, though, some problems were unexpectedly resolved through voice conversation. For example, users that talked with the robot using voice only were more insensitive to breakdowns in dialogue logic than when text chats were used. Additionally, when a robot generated inconsistent utterances, users tended to continue the dialogue if they were facing an actual robot. This behavior was totally opposite to that observed in text chats. It has been reported that people tend to maintain relationships with dialogue partners, and we assume that this effect exists even with robots when they have a human-like appearance [5]. We are currently investigating these advantages and trying to incorporate them as fundamental techniques of dialogue systems.

5. Dialogue with multiple robots

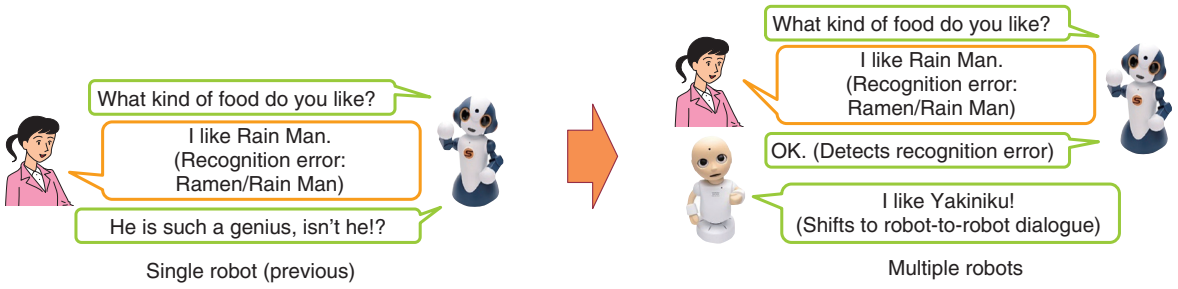
Multiple robots or computer-generated agents were reported to be effective for maintaining active and natural dialogues in system-initiative dialogue systems such as those for museum audio tour guide systems [6]. We therefore developed techniques to avoid dialogue breakdown through collaborative interaction between robots when utterance generation or speech recognition returns the wrong results (**Fig. 3**).

5.1 Utterance generation errors

An example is given in Fig. 3(a) in which the robot understands only part of the user utterance (*coat*) and generates system utterances with slightly wrong



(a) Avoiding dialogue breakdown in case of utterance generation error



(b) Avoiding dialogue breakdown in case of speech recognition error

Fig. 3. Conversation with multiple robots.

dialogue topics (*The trend this year is Chesterfield coats*). Robots that talk with users have a duty to respond appropriately to user utterances. However, in this case, since the robot cannot generate an appropriate response utterance, the user is disappointed with the robot response.

In contrast, when there are two robots, if robot A responds to a user utterance with fillers, the duty is partially fulfilled. At that time, if robot B generates the earlier utterance, this utterance can be construed as a new dialogue topic that is introduced based on the previous dialogue topic that is reacted to by robot A; therefore, the user does not sense the slight inappropriateness of the utterances and easily continues talking.

5.2 Speech recognition errors

When a critical error occurs in speech recognition, it is expected that the system and user utterances will be completely inconsistent (Fig. 3(b)). We developed a technique to avoid speech recognition errors using dialogue breakdown detection technology that identifies inconsistent utterances. When a speech recognition error is detected, our robots have a conversation according to the dialogue topic that contained the previous dialogue history. In this case, although the user feels that the robots are ignoring the user, since the dialogue topics are consistent and the dialogue itself is continuing, it is more natural than when the robots generate utterances using the wrong results of the speech recognition. We found that this technique significantly improved user satisfaction compared to the case when a single robot tried to avoid this type of dialogue breakdown with the same approach.

6. Conclusion

In this article, we introduced our text-chat based conversational systems and its implementation with one or more actual robots. We are also tackling other issues such as voice synthesis that expresses utterance intentions, automatic evaluation of conversational systems, and improvements in turn-taking to achieve user-friendly conversational robots.

Acknowledgments

This article describes collaborative work done with

Professor Hiroshi Ishiguro at Osaka University and with NTT Media Intelligence Laboratories.

References

[1] H. Sugiyama, T. Meguro, R. Higashinaka, and Y. Minami, "Open-domain Utterance Generation Using Phrase Pairs based on Dependency Relations," Proc. of the 2014 IEEE Workshop on Spoken Language Technology, South Lake Tahoe, Nevada, USA, Dec. 2014.

[2] L. C. Tidwell and J. B. Walther, "Computer-mediated Communication Effects on Disclosure, Impressions, and Interpersonal Evaluations: Getting to Know One Another a Bit at a Time," Human Commun. Res., Vol. 28, No. 3, pp. 317–348, 2002.

[3] H. Sugiyama, T. Meguro, R. Higashinaka, and Y. Minami, "Large-scale Collection and Analysis of Personal Question-answer Pairs for Conversational Agents," Proc. of IVA 2014 (the 14th International Conference on Intelligent Virtual Agent), pp. 420–433, Boston, MA, USA, Aug. 2014.

[4] L. Ulanoff, "Eerie Geminoid Robot Can Now Carry on a Conversation," Mashable, Mar. 2016. http://mashable.com/2016/03/13/geminoid-robot-conversation/?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+Mashable+%28Mashable%29#IT9WelTlkkqR

[5] A. Baylor and S. J. Ebbers, "Evidence that Multiple Agents Facilitate Greater Learning," Artificial Intelligence in Education: Shaping the Future of Learning through Intelligent Technologies, pp. 377–379, 2003.

[6] S. Takeuchi, T. Cincarek, H. Kawanami, H. Saruwatari, and K. Shikano, "Construction and Optimization of a Question and Answer Database for a Real-environment Speech-oriented Guidance System," Proc. of the 9th International Committee for the Co-ordination an Standardization of Speech Databases and Assessment Techniques, pp. 149–154, Penang, Malaysia, Dec. 2007.

Hiroaki Sugiyama

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

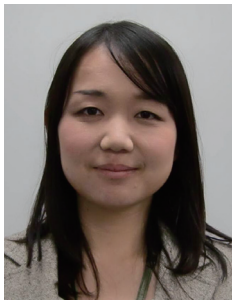
He received a B.E. and M.E. in information science and technology from the University of Tokyo in 2007 and 2009, and a Ph.D. in engineering from the Nara Institute of Science and Technology. He joined NTT Communication Science Laboratories in 2009 and studied chat-oriented dialogue systems and language development of human infants. He is a member of the Japanese Society for Artificial Intelligence (JSAI).

Ryuichiro Higashinaka

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

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

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



Toyomi Meguro

Senior Research Scientist, NTT Communication Science Laboratories.

She received a B.E. and M.E. in electrical engineering from Tohoku University, Miyagi, in 2006 and 2008. She joined NTT in 2008. Her research interests are building listening agents and recommendation systems.

Child Language Development: The Differences between Japanese and English

Yuko Okumura, Tessei Kobayashi, and Yuriko Oshima-Takane

Abstract

In early language development, it is known that Japanese-speaking children acquire words in a more gradual manner and have smaller productive vocabulary sizes compared with English-speaking children. On the other hand, Japanese-speaking children have an ability to learn new words correctly from earlier stages of lexical development than English-speaking children. Why do Japanese-speaking children have smaller productive vocabulary sizes despite this ability to learn words correctly? To explore this riddle, we compared parental input between Japanese and English and examined the relationship between parental input and child vocabulary development.

Keywords: language acquisition, parental input, cross-linguistic study

1. Introduction

Children begin to produce their first words around one year of age, and their productive vocabulary increases rapidly from about 18–20 months on. Just how children acquire words is a major topic of study in the fields of psychology, linguistics, cognitive science, and education as well as a question of great concern to child rearers.

At NTT Communication Science Laboratories, we have been studying how children learn a language from a human science perspective in relation to information engineering. A key project in this research aims to clarify what types of words children can speak and when they begin speaking them [1]. In this project, we succeeded in creating a child vocabulary development database by having about 1300 mothers fill in a checklist on what words their children could comprehend and produce. On the basis of the checklist data, we estimated the age for children to comprehend/produce each of about 2700 words. In particular, on closely examining the data for words of infant-directed speech (IDS) such as onomatopoeic words (e.g., *bow-wow* for a dog and *vroom* for a car), it was found that children could speak IDS words quite early compared with words of adult-directed speech (ADS).

We also proposed a picture book search system using the child vocabulary development database to facilitate the retrieval of picture books that match the child's interests and developmental stage [2]. It is well-known that reading picture books to children promotes their vocabulary development [3]. Finding just the appropriate picture book for a child should have the effect of both expanding vocabulary and fostering emotional development.

We have recently come to focus on cross-linguistic studies exploring differences in language development between Japanese and English. The goal of cross-linguistic studies is to elucidate the mechanism that drives vocabulary learning in children by clarifying what is common and different to both languages. In the following, we introduce our latest findings in these Japanese-English cross-linguistic studies.

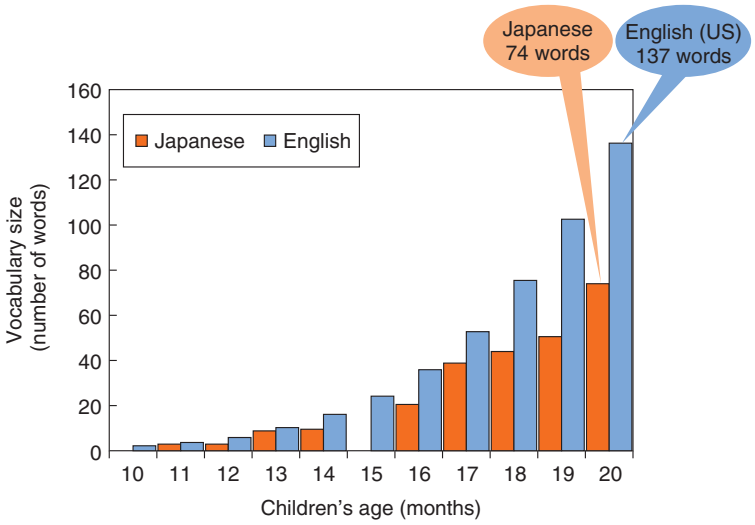


Fig. 1. Children's productive vocabulary size.

2. Comparison of productive vocabulary between Japanese and English

With regard to differences in language development between Japanese and English, it has been known that Japanese-speaking children have smaller productive vocabulary sizes than English-speaking children [4]. This tendency was reported in previous research using a vocabulary checklist method in which parents are asked about words that their children can produce. However, the previous cross-linguistic finding on vocabulary development was based on a very small sample, so we resurveyed the vocabulary development by collecting a large sample using an internationally standardized vocabulary checklist called the MacArthur-Bates Communicative Developmental Inventories (CDI) [5, 6].

For Japanese-speaking children, we extracted data for about 1700 children and estimated the number of productive words at the age of 10 to 20 months. For English-speaking children, we used data for about 1800 mother-child pairs obtained from the publically released United States version of the CDI and investigated the number of productive words for the same months. The results of children's vocabulary size are shown in **Fig. 1**. It can be seen that Japanese-speaking children acquired words in a more gradual manner than English-speaking children. At 20 months, Japanese-speaking children were able to produce 74 words on average, while English-speaking children were able to produce 137 words on average, or nearly twice as many. These results show that Japanese-speaking children have a smaller productive vocabulary than English-speaking children in the early lexical development of one-to-two-year-olds.

3. Comparison of word learning between Japanese and English

Can the finding that Japanese-speaking children have a smaller productive vocabulary than English-speaking children be attributed to the differences in their ability to learn new words? To answer this question, we investigated Japanese- and English-speaking 20-month-old children's ability to learn new words by conducting an experiment based on the habituation method that used a child's looking time as an index [7]. Specifically, we investigated whether a child accurately associated a novel verb (i.e., an unfamiliar verb) with an action and not an object. In the learning phase, we presented the child with a movie clip of a rabbit performing action A on object A (knocking down a blue object) together with speech containing a novel verb (it is *seta*-ing a toy) and a movie clip of a rabbit performing action B on object B (jumping on a red object) together with speech containing a novel verb (it is *moke*-ing a toy), as shown in **Fig. 2(a)**.

At first, the child looks at the two movie clips. In a while, however, the child becomes habituated to the repeated movies, and their looking times tend to decline upon repeated presentations. This was the

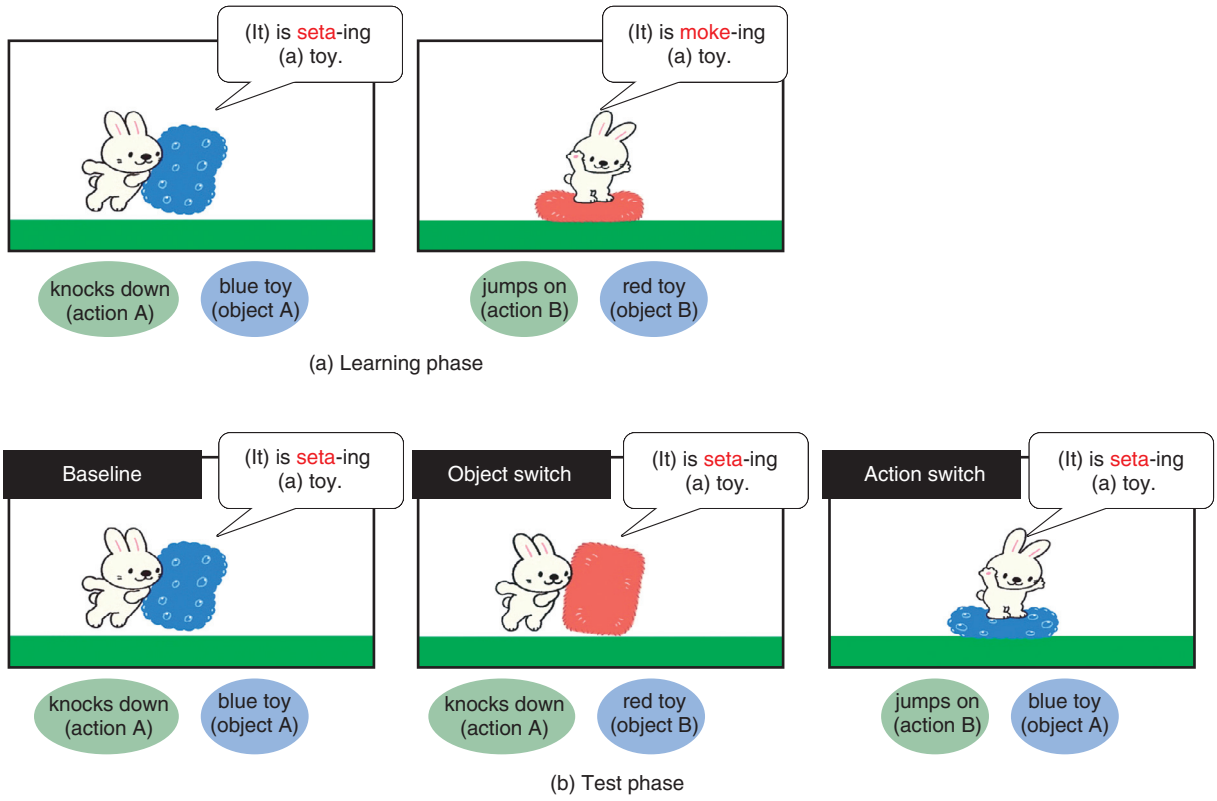


Fig. 2. Word learning experiment.

beginning of the test phase. Here, we presented the child with a movie clip showing the same combination of action and object as in the learning phase (a rabbit performing action A (*seta*-ing) on object A) as a baseline trial plus movie clips switching those action/object combinations: a rabbit performing action A (*seta*-ing) on object B (object-switch trial) and a rabbit performing action B (*seta*-ing) on object A (action-switch trial) (**Fig. 2(b)**).

Now, if the child has correctly associated the verb presented in the learning phase (*seta*-ing) with action A, it can be predicted that the child will notice a change in the word-action combination in the action-switch trial and that the child's degree of attention toward the movie clip will increase, resulting in a rise in looking time (dishabituation). Additionally, if the child has mistakenly associated the verb presented in the learning phase (*seta*-ing) with object A, it can be predicted that looking time will likewise rise in the object-switch trial since the word-object combination has changed. Thus, this method can be used to experimentally judge whether a child has associated a novel verb with an action or object.

The results of this word-learning experiment showed that Japanese-speaking children increased their looking time only in the action-switch trial, indicating that they can correctly associate verbs only with actions (**Fig. 3**). In contrast, English-speaking children exhibited a rise in looking time in both the action-switch and object-switch trial, indicating that they associate verbs with both actions and objects. These results provide strong evidence that Japanese-speaking children have a greater ability to correctly associate words with their referents than English-speaking children at 20 months of age. It is therefore unlikely that the reason why Japanese-speaking children have a smaller productive vocabulary than English-speaking children is that their ability to learn new words is not sufficiently developed.

4. Comparison of parental input between Japanese and English

Despite the fact that Japanese-speaking children develop an ability to learn words more correctly than English-speaking children, why do Japanese-speaking

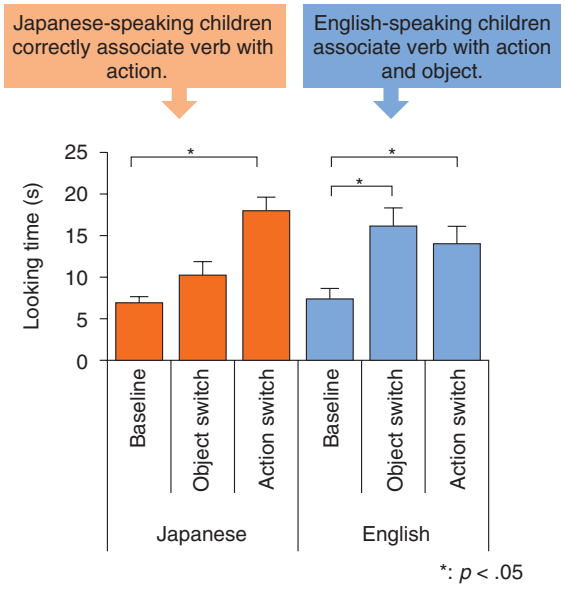


Fig. 3. Results of word learning experiment.

children have a smaller productive vocabulary? It is known that a variety of factors can affect vocabulary development such as the child's social competence (joint attention etc.) [8] and the environment in which the child is raised (family conditions etc.) [9]. Of these, we focus here on parental speech to children. This is because children are thought to gradually acquire words through repeated language input from their parents and because differences in that parental input may have an influence on the vocabulary development of Japanese-speaking and English-speaking children.

With this in mind, we conducted an experiment on parental input in Japan and Canada targeting mothers having Japanese- and English-speaking 20-month-old children, respectively [10]. In the experiment, each mother sat her child on her lap in a laboratory and described to the child scenes from 15 movie clips (a dog eating, a pig tumbling, etc.) shown on a monitor (**Fig. 4**). On completing this experiment, we prepared and analyzed the full transcript of what the mothers said.

The results showed that Japanese-speaking and English-speaking mothers differed significantly in their use of IDS words. Such words generally refer to onomatopoeic words and repetitive vocal sounds, which constitute a special type of vocabulary that adults use with younger children. In particular, the Japanese-speaking mothers frequently used IDS words, which occupied 26% of their input on average. In contrast, only 8% of the English-speaking mothers' utterances contained IDS words. In addition, the Japanese-speaking mothers would tend to mix both IDS words and ADS words for a referent in a movie clip, such as "It's a dog. It's a bow-wow! Mr. Doggie is eating some food. Munch-munch!" In this example, Japanese mothers used "dog," "Mr. Doggie," and "bow-wow" to teach her child about the referent "dog," and "eating" and "munch-munch" to teach about the referent "to eat." As a result, they switched labels when referring to the same referent, and their input ranked low in consistency.

In contrast, the English-speaking mothers would tend to use a single word for a referent in a movie clip, such as "Dog! It's eating. The dog is eating." In other words, English-speaking mother's utterances were more consistent rather than using multiple labels. This analysis of parental input revealed that the Japanese-speaking mothers tend to use multiple labels consisting of both IDS and ADS words and that their input is low in consistency compared with English-speaking mothers.

On the basis of the results presented above, we consider that the low-consistency input style of Japanese mothers has an effect on word acquisition in their children and slows down vocabulary development. A conceptual diagram of the relationship between parental input and child's word acquisition is shown

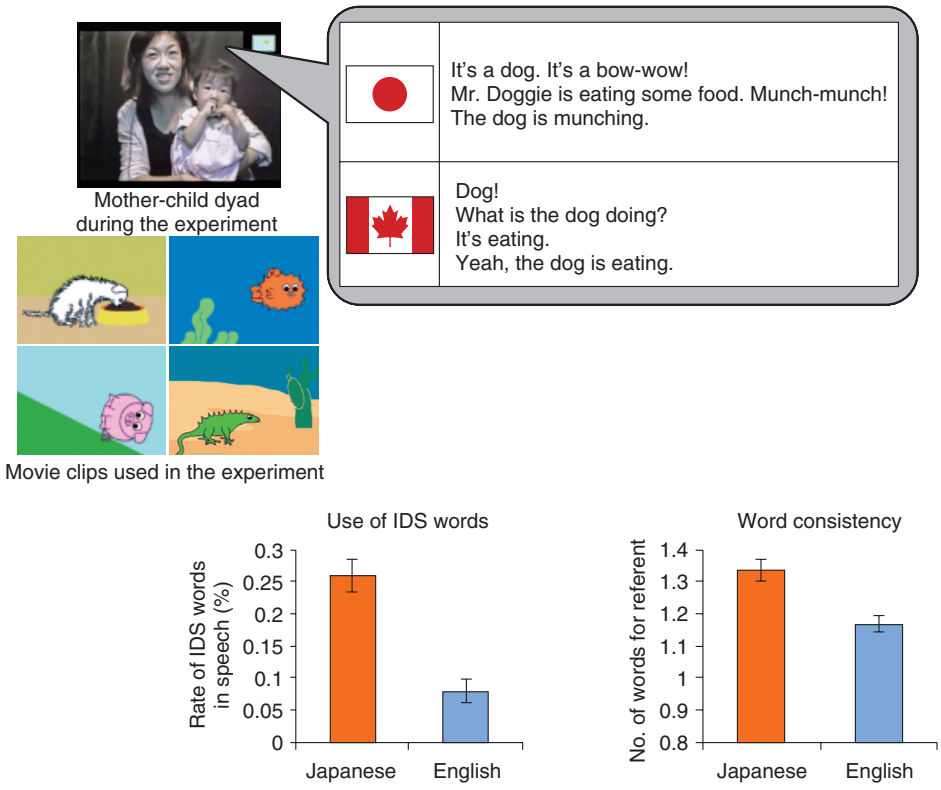


Fig. 4. Parental input experiment.

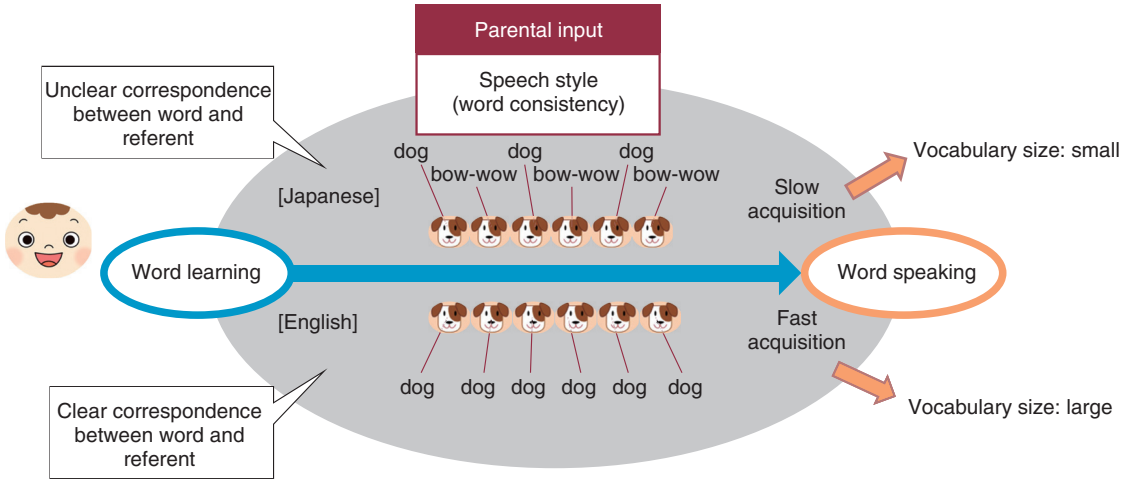


Fig. 5. Parental input and child's word acquisition.

in **Fig. 5**. For Japanese-speaking children, the parent's use of multiple words for a certain referent can make the correspondence between those words and the referent unclear, which could be strongly related to a delay in word acquisition and a small productive vocabulary. In contrast, for English-speaking children, we consider that the consistent input they receive for a certain referent helps to make

the correspondence between that word and the referent clear, resulting in faster word acquisition and a larger productive vocabulary. These differences in parental input observed by cross-linguistic studies are providing very interesting evidence for elucidating the mechanism of vocabulary development.

5. Future outlook

In this article, we introduced research exploring the mechanism of word acquisition in children through a cross-linguistic study between Japanese and English. We described the way in which Japanese-speaking parents use more than one word for a single referent and that such parental input with low consistency may contribute to relatively slow word acquisition in Japanese-speaking children. In contrast, the ability to learn new words develops earlier in Japanese-speaking children than English-speaking children. The reason may be that Japanese parental input with such low consistency has a different effect on the child at the time of learning words than at the time of producing the words. That is, the input style may give the child some sort of positive effect when learning words. In addition, prior research reported that American mothers place value on fostering linguistic competence in their children while Japanese mothers place importance on establishing affectionate communication with their children [4]. This way of interacting with children in Japan makes frequent use of IDS words, which may have positive effects outside of word acquisition. In future research, we aim to provide a comprehensive explanation of the various ways in which parental input can have an effect on a child's development.

If the effects of parental input can be clarified by such a series of studies, it may be possible to derive more effective methods for supporting the overall development of communication skills. Going forward, our plan is to propose guidelines for communication environments that support the education of children by linking such methods with information and communications technology.

References

[1] T. Kobayashi, Y. Okumura, and T. Hattori, "Exploring Differences in Learnability of IDS and ADS Words in Children," NTT Technical Journal, Vol. 27, No. 9, pp. 26–29, 2015 (in Japanese).

[2] T. Hattori, T. Kobayashi, S. Fujita, Y. Okumura, and K. Aoyama, "Pitarie: Picture Book Search with Interdisciplinary Approach," NTT Technical Review, Vol. 14, No. 7, 2016.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201607ra1.html

[3] G. J. Whitehurst, F. L. Falco, C. J. Lonigan, J. E. Fischel, B. D. DeBaryshe, M. C. Valdez-Menchaca, and M. Caulfield, "Accelerating Language Development through Picture Book Reading," Dev. Psychol., Vol. 24, No. 4, pp. 552–559, 1988.

[4] A. Fernald and H. Morikawa, "Common Themes and Cultural Variations in Japanese and American Mothers' Speech to Infants," Child Dev., Vol. 64, No. 3, pp. 637–656, 1993.

[5] L. Fenson, P. S. Dale, J. S. Reznick, E. Bates, D. J. Thal, and S. J. Pethick, "Variability in Early Communicative Development," Monogr. Soc. Res. Child Dev., Vol. 59, No. 5, pp. 174–185, 1994.

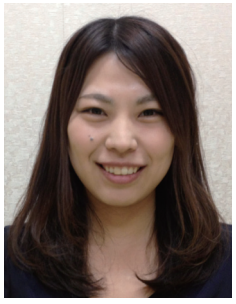
[6] T. Ogura and T. Watamaki, "The Japanese MacArthur Communicative Development Inventory: Words and Gestures," Kyoto International Social Welfare Exchange Center, 2004.

[7] Y. Oshima-Takane, T. Kobayashi, and E. Chan, "Learning Novel Transitive Verbs in Japanese-, French- and English-speaking Infants: A Cross-linguistic Study," IASCL 2014 (the 13th International Congress for the Study of Child Language), Amsterdam, Netherlands, July 2014.

[8] Y. Okumura, Y. Kanakogi, T. Kobayashi, and S. Itakura, "Individual Differences in Object Processing by Gaze Following Predict Language Development," SRCD (Society for Research in Child Development) Biennial Meeting 2015, Philadelphia, USA, Mar. 2015.

[9] E. Hoff, "How Social Contexts Support and Shape Language Development," Dev. Rev., Vol. 26, No. 1, pp. 55–88, 2006.

[10] T. Kobayashi, Y. Oshima-Takane, Y. Okumura, K. Hirama, and B. Melhem, "A Cross-linguistic Study on Maternal Input Style and Child Word Acquisition," JSLS (Japanese Society for Language Sciences) Annual Conference, Tokyo, Japan, June 2016 (in Japanese).



Yuko Okumura
Research Associate, Interaction Research Group, Innovative Communication Laboratory, NTT Communication Science Laboratories.
She received a Ph.D. in psychology from Kyoto University in 2014. She is currently conducting research on social cognition and language development in young children.

Tessei Kobayashi
Senior Research Scientist (Distinguished Researcher), Interaction Research Group, Innovative Communication Laboratory, NTT Communication Science Laboratories.
He received a Ph.D. in psychology from the University of Tokyo in 2004. He is currently researching child language development, especially vocabulary spurts and syntactic bootstrapping.

Yuriko Oshima-Takane
Emeritus Professor, McGill University and Research Professor, Interaction Research Group, Innovative Communication Laboratory, NTT Communication Science Laboratories.
She received a Ph.D. in psychology from McGill University, Canada, in 1985. She is researching child language development, including a cross-linguistic study on early lexical and syntactic development.

Elucidating the Brain Processing Mechanisms of Athletes Using Virtual Reality Technology

Toshitaka Kimura, Kosuke Takahashi, Dan Mikami, and Makio Kashino

Abstract

In addition to fitness, physical skill and state of mind are important factors for athletes to achieve sporting success. These factors are mainly determined by information processing mechanisms in the brain, but the potential for clarifying them with conventional measurement techniques is limited. NTT is developing a virtual reality (VR) system for sports measurement that can provide a highly realistic sports experience. We aim to use this novel VR system to extract key features related to an athlete's skill and mental state and to establish systematic methods for sports training and coaching.

Keywords: sports, brain science, VR

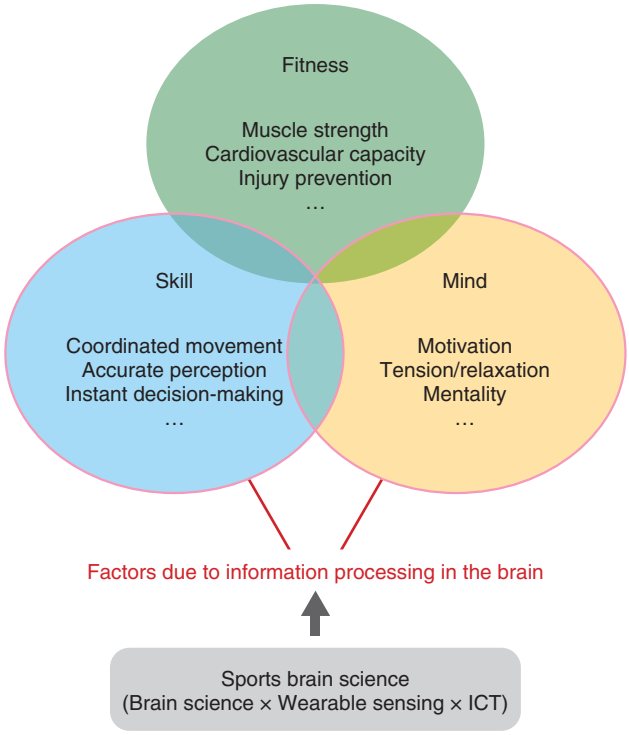
1. Introduction

An athlete's physical fitness, skill, and state of mind are essential factors in sporting performance (**Fig. 1**). Fitness, including muscle strength and cardiovascular capacity, has long been studied in the fields of exercise physiology and sports medicine, and these findings have been utilized in various ways in sports coaching and training. The athlete's skill and mental state are also critical in terms of success in actual sports games. Examples of these include dexterously coordinating the various parts of the body, adequately recognizing a given situation, and overcoming mental pressure. These abilities are mainly determined by information processing in the brain. However, the mechanisms involved are poorly understood because work in this area faces a number of technical limitations. Thus, systematic training methods designed to enhance an athlete's skill and state of mind have not been fully established.

NTT is engaged in a sports brain science project to improve athletic performance (Fig. 1), where the aims are to read the key brain processing features related to an athlete's skill and mental state, develop assistive methods and devices based on the obtained findings, and finally, to improve the athlete's brain functions. Information and communication technology (ICT) know-how as well as neuroscientific techniques are required for the success of this project. For example, wearable sensing and information engineering technologies will be important if we are to effortlessly obtain behavioral and biological data and effectively feed the obtained findings back to the athlete. The virtual reality (VR) technology introduced in this article will also be a powerful tool for facilitating our project, and especially in elucidating the brain processing mechanisms related to an athlete's skill and mental state.

2. Utilizing VR for sports brain science

To identify the key features of brain processing related to skill and mental state, it is first necessary to measure the behaviors and biological signals (e.g., movements, muscle activity, heart rate, and respiration) of an athlete engaging in a sport. There are two possible ways to do this. One is to perform measurements in a laboratory, which is the most commonly



ICT: information and communication technology

Fig. 1. Scope of sports brain science project.

used approach. Such measurements enable the conditions to be strictly controlled, but there is a potential limitation that the target action will be unlike the actual sports action.

The second way is to perform measurements during an actual sporting event. Recent rapid advances in various types of wireless and wearable sensors have gradually enabled us to obtain behavioral data in actual sports fields [1]. We also attempt to obtain biological data during sporting activity using wearable bioelectrodes such as the conductive fabric called "hitoe" [2]. However, unlike in the laboratory, it is hard to control the measurement conditions and situations during an actual sporting event. Since there are usually many factors in actual sports environments that can potentially influence an athlete's performance, it may be difficult to interpret the observed phenomena. Consequently, there is a trade-off between controllability and reality with both types of measurement. Therefore, the measurement approach should be selected according to the aim of the study.

Here, we propose the use of VR technology as a third approach to sports measurement that can compensate for the trade-off between controllability and reality. VR has the advantage of providing a highly realistic visual experience. Various experience-based services and exhibits have recently been introduced and are often combined with the use of a highly immersive head-mounted display (HMD) to enable the user to experience situations such as standing in a high place or riding a roller coaster, and they have successfully monitored user responses. There have also been some sports-related applications, including an HMD-based VR system that enables the user to experience a live 360-degree view of a football player (STRIVR Labs, Inc.). Another advantage is that the technique allows us to arbitrarily manipulate a visual environment combined with computer graphics (CG), which makes it possible to configure a visual environment that is inconceivable in the real world (**Table 1**). With these advantages of VR technology, we are trying to construct VR environments for achieving sports measurement capabilities with both sufficient reality and flexible controllability of the visual experience so as to capture behavioral data that were formerly hard to observe.

Table 1. Advantages and disadvantages for sports measurement environments.

	Laboratory	VR	Real field
Controllability	◎	○	×
Reality	×	○	◎

◎: excellent
○: good
×: poor

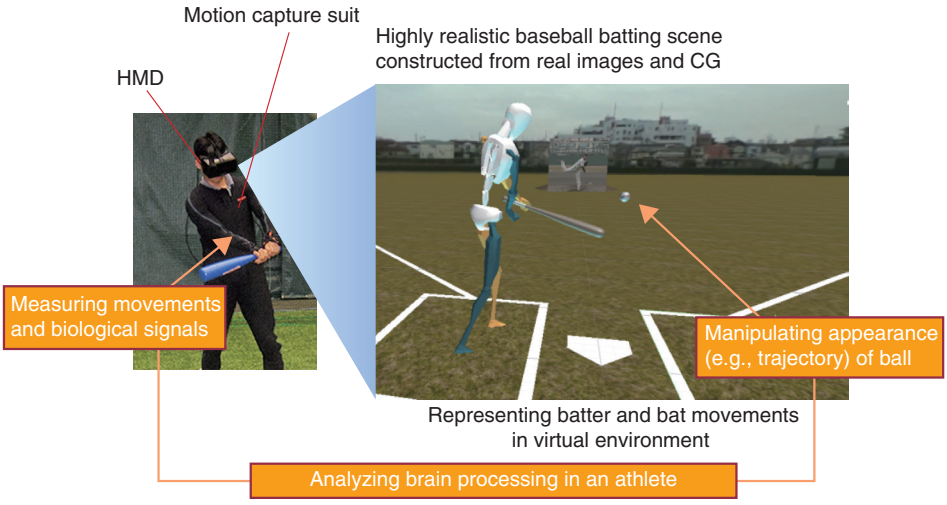


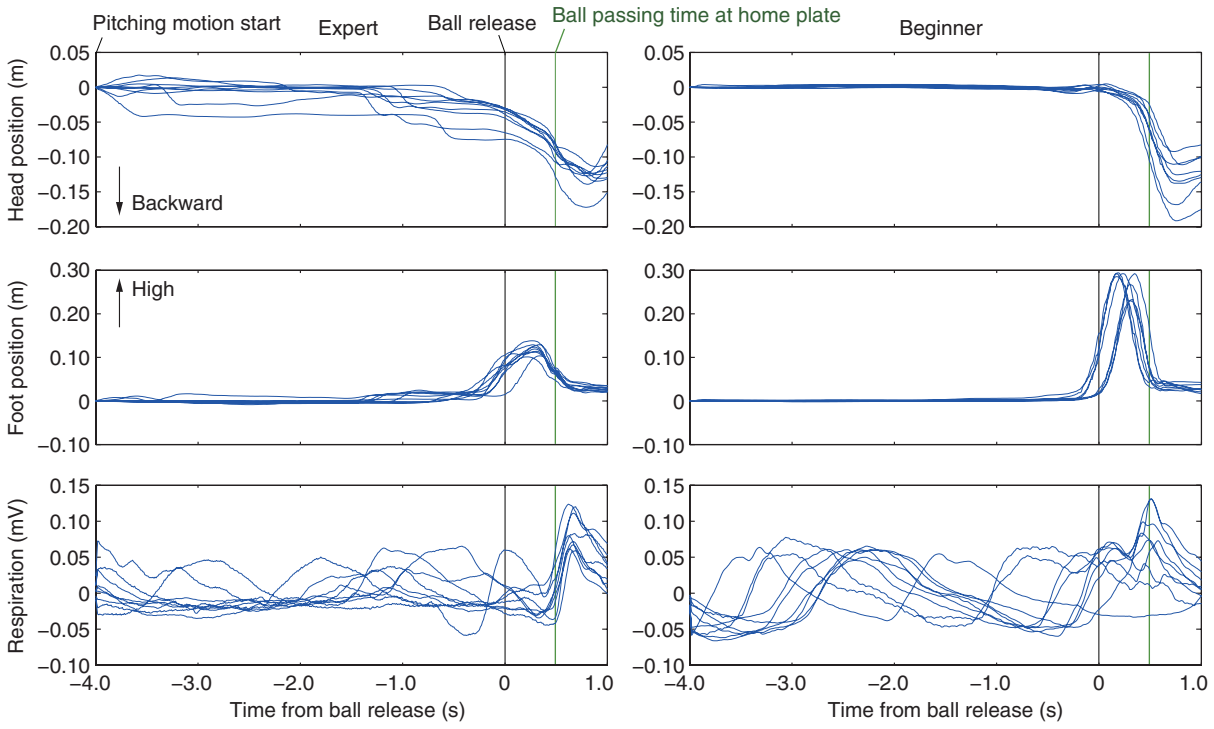
Fig. 2. VR system for experiencing baseball batting.

3. Measuring behavioral data with sports VR system

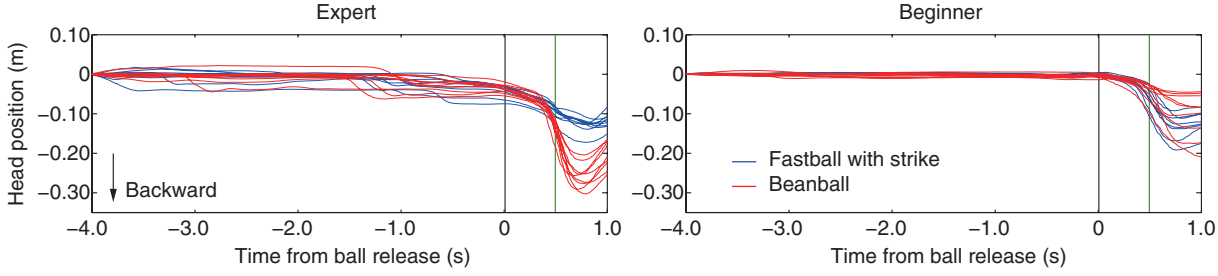
A baseball batter has to manipulate his/her bat within very tight time constraints (within about 0.5 s) from the release of the ball by the pitcher to the moment of impact [3]. However, it would appear to be impossible to achieve appropriate batting by employing the relatively slow conscious (explicit) processing in the brain, because it takes 0.5 s or more between recognizing the ball trajectory to swinging the bat. A key requirement is fast unconscious (implicit) processing in the brain without conscious perception or judgment. Instant decision-making and the quick execution of an action are critical in many sports in addition to batting in baseball. We are studying such implicit brain processing mechanisms with VR technology using baseball batting as an example.

NTT Communication Science Laboratories and NTT Media Intelligence Laboratories have developed an HMD-based VR system for baseball batting (**Fig. 2**), in which a user can experience virtual batting from a batter's viewpoint in a virtual baseball stadium constructed from a combination of real images and CG [4]. The pitched balls are depicted by CG based on previously recorded ball trajectories and are thrown in time with the motion of a pitcher in a simultaneously recorded video. The batter and the bat movements are measured using the nine axes of inertial sensors attached to various parts of the body and the bat and are represented in a virtual avatar in real time. Visual images presented to the right and left eyes are independently displayed on the HMD in tune with the position and direction of the user's head, which allows the user to obtain a virtual view with appropriate depth sensation in an arbitrary location and orientation in the batter's box. These VR system settings enable a user to experience a strong feeling of reality, as if standing in the batter's box, and to perform virtual batting in which the ball bounces depending on its interaction with the bat.

Examples of behavioral responses to virtual batting by a single expert and a beginner are shown in **Fig. 3**. The participants were instructed to swing at fastballs



(a) Athlete's respiration and batting movement during fastballs with strikes



(b) Batting movements in fastballs with strikes and beanballs

Fig. 3. Behavioral responses to virtual baseball batting.

or curve balls that were thrown in the strike zone. The head and foot movements and the respiratory waveforms for each of 10 trials when randomly presenting fastballs that were strikes are shown in Fig. 3(a). The head movement of the expert participant changed smoothly before the pitcher released the ball, while this was not the case with the beginner. This suggests that the expert prepares in anticipation of the ball release, while the beginner reacts to the ball release. The expert also synchronized his respiration and his foot stance to his estimated timing of the ball's impact. Further, we found that when a ball was coming toward the head, which is known as a beanball (red lines in Fig. 3(b)), the expert moved his head backward and avoided the ball, but the beginner moved too late or insufficiently to avoid the ball.

These behavioral data not only reveal certain characteristics of an athlete but also provide unique findings via VR measurement. For example, it is not ethically possible for researchers to have a pitcher intentionally pitch a beanball at a batter in a real game, and hence, to the best of our knowledge, there is no way of finding out about such occurrences. We also want to obtain novel findings about how a batter predicts a

ball and how emotional experiences (e.g., fear) affect batting history by occasionally removing the ball and manipulating the sequence and variety of pitches.

4. Future prospects

Work has begun recently on evaluating sports performance in virtual environments [5]. However, most of these studies have aimed at approximating the real world, while there have been few studies that purposely manipulated visual environments. Also, the observed objects are mainly behavioral outcomes (e.g., movement and force) with few underlying biological data. Although the VR system we describe in this article is confined solely to limb movements and respiration, its combination with other measurements such as muscle activity, eye movements, and heart rate should enable us to understand multiple aspects of sports performance such as how force is exerted, lines of sight, and the athlete's mental state. We anticipate that our work using VR technology will clarify the key physical and mental features of an athlete and the underlying information processing by the brain, which have not been determined with existing approaches in the laboratory or in real sports fields.

Our VR system will be useful not only as a tool in sports neuroscience but also in sports training and coaching. For example, although a batter can carefully study opposing pitchers using video, realistic experiences of pitches from the batter's viewpoint would be more useful. Furthermore, sharing these VR images with coaches can result in more practical and effective coaching. A batter can also use batting training to deal with weak pitchers and pitches. Moreover, this interactive VR technology will be able to provide new value and content in relation to other fields such as rehabilitation and entertainment.

Acknowledgment

This research is supported by the Japan Science and Technology Agency (JST) CREST project.

References

[1] D. Anzaldo, "Wearable Sports Technology - Market Landscape and Compute SoC Trends," Proc. of ISOCC 2015 (the 12th International SoC Design Conference), Gyungju, Korea, Nov. 2015.

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

[3] T. Ijiri, M. Shinya, and K. Nakazawa, "Interpersonal Variability in Timing Strategy and Temporal Accuracy in Rapid Interception Task with Variable Time-to-contact," J. Sports Sci., Vol. 33, No. 4, pp. 381–390, 2014.

[4] D. Mikami, M. Isogawa, K. Takahashi, H. Takada, and A. Kojima, "Immersive Previous Experience in VR for Sports Performance Enhancement," Proc. of icSports2015 (the 3rd International Congress on Sport Sciences Research and Technology Support), Lisbon, Portugal, Nov. 2015.

[5] B. Bideau, R. Kulpa, N. Vignais, S. Brault, F. Multon, and C. Craig, "Using Virtual Reality to Analyze Sports Performance," IEEE Comput. Graph. Appl., Vol. 30, No. 2, pp. 14–21, 2010.

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

He received a Ph.D. in human neuroscience from the Graduate School of Arts and Sciences, the University of Tokyo, in 2003. He joined NTT Communication Science Laboratories in 2003. He has been engaged in research on sensorimotor control and action-perception interaction in humans. He has recently been focusing on brain functions in athletes and the use of ICT in sports. He is a member of the Society for Neuroscience, the Japan Neuroscience Society, and the Institute of Electronics, Information and Communication Engineers (IEICE).

Kosuke Takahashi
Research Scientist, Visual Media Project, NTT Media Intelligence Laboratories.

He received a B.Sc. in engineering and an M.Sc. in informatics from Kyoto University in 2010 and 2012. He joined NTT Media Intelligence Labs in 2012. His research interests include computer vision. He received the Best Open Source Code award Second Prize at the 2012 IEEE Conference on Computer Vision and Pattern Recognition.

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

He received a B.E. and M.E. from Keio University, Kanagawa, in 2000 and 2002, and a Ph.D. from Tsukuba University in 2012. He joined NTT in 2002. His current research activities are mainly focused on computer vision and virtual reality. He was awarded the Meeting on Image Recognition and Understanding 2009 Excellent Paper Award 2009, the IEICE Best Paper Award 2010, the IEICE KIYASU Zen'iti Award 2010, the IPSJ SIG-CDS Excellent Paper Award 2013, and the IEICE Human Communication Award 2015.

Makio Kashino
Senior Distinguished Scientist/Executive Manager of Human Information Science Laboratory, NTT Communication Science Laboratories.

He received his B.A., M.A., and Ph.D. in psychophysics from the University of Tokyo in 1987, 1989, and 2000. He joined NTT in 1989. From 1992 to 1993, he was a visiting scientist at the University of Wisconsin. He is currently a visiting professor in the Department of Information Processing, Tokyo Institute of Technology (2006–). He has been conducting psychophysical experiments and using neuroimaging, physiological recordings, and computational modeling to investigate the functional and neural mechanisms of human cognition, especially auditory perception, crossmodal and sensorimotor interaction, and interpersonal communication. His other recent research interests include sports, music, and autism.

How Tracking Technologies Improve Family Communication

Naomi Yamashita

Abstract

In this article, we explore how health-tracking technologies could be designed to support family caregivers to better cope with the unexpected behaviors of a depressed family member. We designed a tracking tool called Family Mood and Care Tracker (FMCT) and deployed it for 6 weeks in the homes of 14 family caregivers looking after depressed family members. FMCT is a web-based tracking tool designed specifically for family caregivers to allow them to record their caregiving activities and the sufferers' health conditions. Our findings demonstrate how the family caregivers made use of FMCT to better cope with depressed sufferers and how it improved the communication between family caregivers and sufferers.

Keywords: depression, family caregivers, tracking technology

1. Introduction

The onset of mental illness in a family places a significant burden on its members and causes a great deal of stress. In most cases, mental illness—as well as its medication—causes unexpected behaviors in the sufferer. For example, sufferers may become demanding and disruptive, or they might express extremely negative attitudes or excessive anger. When family caregivers face such circumstances, they often become puzzled and do not know how to react.

In addition to the difficulty of coping with such unexpected behaviors, social stigma is attached to such illnesses, making it difficult for family caregivers to consult with others. Under such circumstances, most family caregivers tend to gather information about the illness themselves, although finding useful information is complicated due to the huge individual differences among symptoms. Consequently, family caregivers generally have little choice but to rely on discoveries based on their own experiences.

Despite the significant impact of mental illness on families, much of human computer interaction and computer-supported cooperative work on mental illness has focused on either the patient or the clinician, aiming for better patient care. Little research has focused on supporting family caregivers.

Our goal is to design new technologies that help family caregivers develop strategies to better interact with sufferers. In this article, we focus on computer diary based tracking technology and explore its potential to aid family caregivers. For example, by recording and tracking the sufferer's moods and caring activities, family caregivers may be able to analyze how their behavior affected the sufferer's moods/symptoms and thus discover effective coping strategies. Such behavioral change in caregivers may lead to better sufferer-caregiver relationships and prevent an escalation in stress between the two.

Meanwhile, previous research also warns that family caregivers' use of tracking technologies to monitor sufferers' health may create sufferer-caregiver conflicts caused by the sufferer's perception of surveillance, that is, an uncomfortable feeling of being monitored. Since individuals suffering from mental illness tend to have a high perception of surveillance, we need to be aware of its risks and pay extra attention to how tracking technologies affect domestic relationships.

We explore the following research questions in this article: (1) how tracking technology might assist

family caregivers' development of coping strategies, and (2) whether/how it affects the relationships between family caregivers and sufferers. We investigate these research questions by focusing on caregivers who look after family members diagnosed with depression.

To answer these research questions, we first conducted a preliminary interview study with 14 family caregivers who were looking after a depressed family member. The aim of this study was to identify the basic design features (e.g., items family caregivers wish to track, family caregivers' concerns of using tracking technologies at home) of our tracking tool. Based on the interview study, we developed a web-based tracking tool for recording data and deployed it with 14 family caregivers (not members of the preliminary interview study) of depressed family members. During the deployment study, the family caregivers recorded their caregiving activities and the sufferer's behaviors/moods for six weeks. The focus of the deployment study was to investigate the impact of tracking technology on the family caregivers' development of coping strategies and the relationships between family caregivers and sufferers.

2. Preliminary interview study

To identify the ways in which tracking technologies could be designed to support family caregivers, we interviewed 14 family caregivers looking after a depressed family member. From the interviews, we identified the basic design requirements of our tracking tool. Much information was gleaned from the interviews, including different ways family caregivers currently handle their caregiving tasks, and how technology might assist them in caregiving.

2.1 Current practices

Some family caregivers mentioned that they kept a caregiving journal. However, none wrote on a daily or regular basis. Furthermore, they typically only wrote about negative events such as arguments they had had with the sufferer, and they described their own feelings about those events as a way to let off steam. Although they sometimes reviewed their journal, they felt it discouraging and unhelpful because it was filled with negative and emotional descriptions.

According to the family caregivers, some sufferers also kept a diary about their health conditions as part of therapy or treatment suggested by their doctors. However, most of the family caregivers mentioned that they avoided reading what the sufferers wrote due to privacy concerns.

2.2 Items to record

When family caregivers were asked what items they wanted to track, they listed a range of things that could be consequential to sufferer moods: weather, medication, amount of sleep, outdoor activities, meals, drinking (i.e., consumption of alcohol), their own caregiving activities, and unexpected events such as phone calls from a friend or fighting and/or arguments. For example, many mentioned that the sufferer's mood became worse when it was rainy or when the atmospheric pressure decreased. They also mentioned that sufferers either ate too much or did not eat at all when they were in a bad mood. Some seemed to notice a change in sufferer behavior after a change in medication. The articulation of unexpected events was expected to help them better understand the disease.

2.3 Concerns about using tracking technologies at home

Although many participants showed interest in detecting the signs of mood changes in the sufferers or in developing coping strategies to achieve more positive interactions with them, they also expressed some concerns about using tracking technologies at home. Many worried that sufferers might perceive surveillance by having their details recorded. They also worried that sufferers might accidentally read the recordings and take them negatively. Overall, the family caregivers were concerned that tracking the sufferer's mood/behavior might cause sufferer-caregiver conflicts. Finally, the family caregivers worried about the additional workload caused by using the technology.

2.4 Technology requirements

The current practices suggest that our tracking tool should:

- promote unbiased recording (i.e., not just negative events) in order to support positive self-reflection by the family caregivers.
- not take advantage of recorded data of sufferers, meaning that data input will be done solely by family caregivers.
- include the following items: weather, medication, amount of sleep, outdoor activities, consumption of meals and alcohol, their own caregiving activities, and unexpected events.
- work not only on ordinary personal computers but also on mobile devices to support family

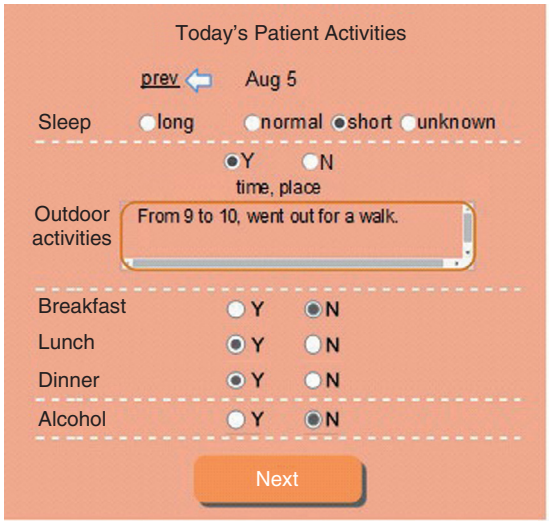


Fig. 1. User interface in recording section.

caregivers' private data input.
- minimize the burden of data input.

3. Tracking tool design

The Family Mood and Care Tracker (FMCT) tracking tool consists of two sections: recording and reviewing. The recording section allows the family caregivers to record data, and the reviewing section facilitates reflective analysis by family caregivers through visualization of the recorded data in a chart.

3.1 Recording section

The recording section allows the family caregivers to record all of the items identified in the preliminary interview study: the sufferer's mood, medication, amount of sleep, outdoor activities, consumption of meals and alcohol, their own caregiving activities, and unexpected events.

One of the pages of the recording section is shown in **Fig. 1**. To minimize family caregivers' input burden, most recording items consist of multiple-choice questions. For example, for the sufferer's sleeping, the caregivers enter the amount of sleep by selecting from four choices: long, normal, short, and unknown. The weather information (atmospheric pressure, highest/lowest temperatures, and actual weather conditions) was automatically retrieved from a weather information site.

Free-form text boxes were provided for three items: outdoor activities, own caregiving activities, and unexpected events. For outdoor activities, the caregiver could enter details of where the sufferer went and for how long. For own caregiving activities, three free-form text boxes were prepared. Each box prompted input on positive and negative outcomes of their caregiving activities as well as lessons learned. We expected these three items to facilitate caregivers' unbiased reflection on their daily caregiving activities and help them develop coping strategies for the future. Finally, a free-form text box was prepared to facilitate reflection prompted by unexpected events. Since unexpected events happen on an irregular basis, the data input to this field was optional.

3.2 Reviewing section

The reviewing section displays a chart that facilitates the caregiver's reflection on his/her caregiving activities and the sufferer's conditions (**Fig. 2**). All of the data were displayed in a single chart so that caregivers could explore their data in a holistic manner. The chart shows the sufferer's mood (red line) and caregiver burden (blue line) as graphed lines; the closer to the left a graph point is, the more negative it is. Weather information and other data are shown as icons.

Family caregivers can also see the unexpected events and the caregiving activities recorded on a particular day by clicking on a date, which makes a textbox pop up that shows the recorded data. Another textbox for the details of outdoor activities also pops up by clicking on the shoe icon. We expected this

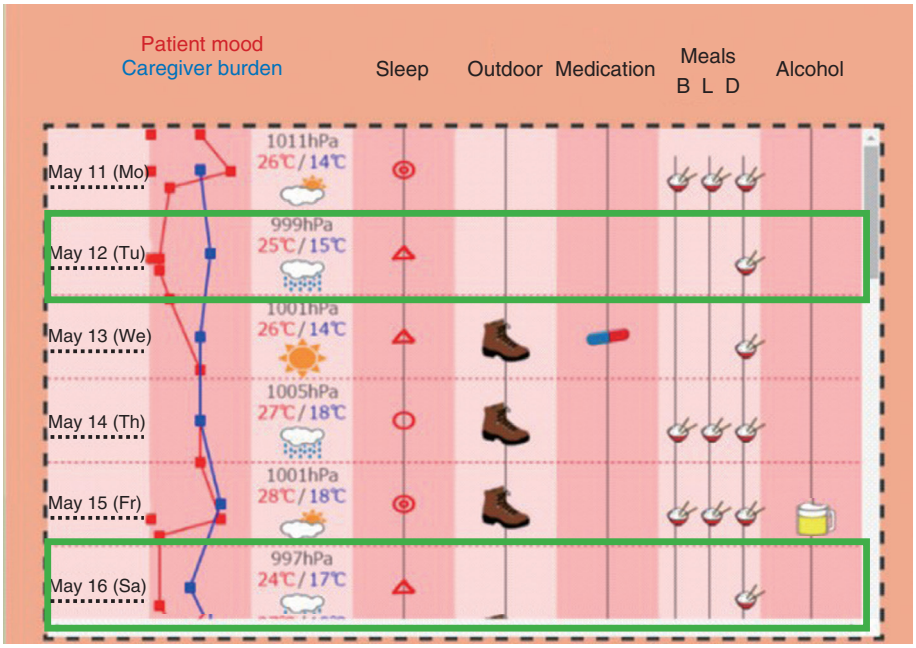


Fig. 2. User interface in reviewing section.

chart to help family caregivers identify the sufferer's health condition patterns and develop coping strategies. From Fig. 2, for example, we see that the sufferer tends to stay indoors or skip breakfast/lunch when he is in a bad mood (as highlighted in green boxes).

4. Deployment study

We conducted a six-week deployment study with FMCT to investigate how family caregivers' tracking behaviors affected their caregiving activities and interaction with sufferers. The study was reviewed and approved by the ethics committee of NTT (ethics review ID: H27-011).

4.1 Participants: family caregivers

We recruited 15 family caregivers (12 females and 3 males, mean age = 43.0) for the deployment study from the same consumer marketing company used for the preliminary interview study. In the recruiting process, any family caregivers who were themselves diagnosed with major depression were excluded. Thus, we ended up with 14 family caregivers, as mentioned above, who were currently looking after a depressed family member. Nine caregivers were spouses of sufferers, two were parents, two were daughters, and one was a sibling. All of the caregivers lived with the sufferer and held primary responsibility for looking after them. Six caregivers were full-time homemakers, six had full-time jobs outside the home, and two worked at home.

4.2 Results

Although this study was meant to be the first step to reach our goal (i.e., to develop new technologies for better patient-caregiver relationships), our tracking also received a surprisingly high evaluation from the family caregivers. Indeed, most participants (12) noted in the interview that they became more actively engaged in the care, gained better control over the sufferer's moods, and increased/improved communication with the sufferers. Below, we analyze how such changes were made through the data obtained by FMCT.

5. Interview findings

We report here some of the caregivers' views and comments on the use of FMCT to support their caregiving activities.

5.1 Becoming attentive to sufferer's behaviors

In general, the survey results showed that family caregivers perceived themselves as becoming more attentive to the sufferer's behaviors (mean (M) =

1.93, standard deviation (SD) = 0.65 on a 5-point Likert scale: 1 = became very attentive, 5 = did not become attentive at all). One participant described how using FMCT produced greater attentiveness toward her sufferer's moods and behaviors:
[Identification (ID) 7, full-time employee, daughter of sufferer] "Since I'm always busy in the morning, I didn't really care if she was in a good or bad mood. But after I started to use this [FMCT], I paid more attention to her condition to find something to record. Like her mood while eating breakfast and her tone of voice."

Another participant, whose husband experienced a relapse two years ago, mentioned that she started to pay extra attention to his medication:
[ID 1, homemaker, wife of sufferer] "I feel like I'm now more involved with my husband's illness. I started to watch him more carefully, and I realized that he often forgot to take some of his medication."

Accordingly, although the family caregivers' access to FMCT was limited (1.52 times per day), the recording items seemed to remind them of the things to which they should pay attention when they were with the sufferers. Consequently, most family caregivers became more attentive to their sufferer's behavior, and such attentiveness led them to new findings about the sufferer's behaviors.

5.2 Making better sense of sufferer's behavior

By becoming more attentive to sufferer moods and behaviors, family caregivers noticed subtle things that they never noticed before. These details seemed to provide cues to better understand mood changes and behavioral patterns. For example, one participant realized that her mother's mood was greatly affected by the weather:
[ID 2, full-time employee, daughter of sufferer] "I wasn't even aware that weather affected my mother's mood. During the study, I realized a couple of times that when she was out of sorts, the weather was also bad."

As with *ID 2*, many family caregivers realized that there are underlying reasons for the mood changes of sufferers. Although our study gathered no concrete evidence that their individual discoveries were true, the family caregivers seemed to feel that they could make better sense of the sufferer's behaviors by recording and reviewing the sufferer's moods and activities every day.

5.3 Developing concrete action plans

Such discoveries of sufferer moods/behaviors affected the family caregivers in multiple ways. For example, the discoveries helped them develop concrete action plans to effectively support the sufferers. In other words, the family caregivers gained a clearer sense of what they should and should not do/say to the sufferers. *ID 1*, whose husband often forgot to take his medicine, described how she came to provide help:
[ID 1, homemaker, wife of sufferer] "I talked to my husband when I noticed that he hadn't taken all of his medication. I thought he knew, but to my surprise, he didn't. Since he is a well-organized person, I had been assuming that he could manage by himself. But I realized that there were certain things that he couldn't manage. I decided to support him in these areas."

Other family caregivers reviewed the FMCT graph and searched for clues to better cope with their sufferers. For example, one family caregiver described how she successfully handled her mother's unexpected behavior by reviewing the FMCT record:
[ID 7, full-time employee, daughter of sufferer] "My mother was in a good mood while eating supper, but she suddenly got very quiet. I wondered why and reviewed the [FMCT] record. While tracking back, I started to wonder what things made her happy. Then I found a case where my mom became happy when I praised her cooking. I instantly thought this might be the case. I realized that in this case, I had just eaten silently without praising her cooking. I went to her and said, "Your cooking was so delicious that I completely forgot to thank you." I was relieved to see her returning to normal."

As with *ID 7*, some caregivers reviewed the FMCT record when things went wrong, developed a hypothesis and an action plan to improve the situation, and tested it on the sufferers. Again, we do not know if *ID 7*'s hypothesis was correct—the sufferer may have had different reasons for her change in mood. However, the important point is that FMCT appeared to be useful from the family caregiver's perspective (*ID 7* in the case above). Through building and testing hypotheses, family caregivers seemed to realize that there are some connections between their own behavior and the sufferer's mood/behavior. Some family caregivers even noticed patterns between the two:
[ID 4, full-time employee, husband of sufferer] "Previously, I couldn't understand why her mood went down. Although I had a feeling that I had said something wrong, it quickly faded from my memory. By tracking and reviewing every day, I detected certain patterns. Similar things happened over and over. I

knew that praising was good in my head, but I came to realize how important it is to praise her as a real experience."

Such discoveries seemed to help the family caregivers realize that there are certain things they, as family caregivers, could do to improve the current situation. According to the family caregivers, the *lessons learned* section was particularly useful for keeping track of the trial-and-error results of their hypotheses and for developing better coping strategies. The holistic view of sufferer mood and caregiving activities also seemed to help them analyze the connection between the two, and to think of a new coping strategy. Throughout the process, many family caregivers seemed to perceive themselves as becoming attentive to their own behaviors (M = 2.28, SD = 0.78 on a 5-point Likert scale: 1 = became very attentive, 5 = did not become attentive at all).

5.4 Changing views on sufferers

The family caregivers' discoveries about sufferer moods/behaviors also seemed to affect their views on the sufferers. Many realized that their views were biased or unfair. For example, one family caregiver who had experienced depression himself described how he came to realize that giving his daughter advice based on his own experience could be inappropriate:

[ID 10, working at home, father of sufferer] "I recently realized that my daughter's experience of depression might be different from mine. Hers is seasonal, but mine isn't. I knew this for a fact, but I actually didn't understand the difference. (…) I tended to give her advice based on my own experience. I guess I was being over-possessive at times. (…) While reviewing the graphs, I noticed a correlation between the weather and her moods, and I started to feel that her depression might be totally different from mine."

As with *ID 10*, by reviewing the FMCT records, some family caregivers realized that their views had been unfair. Some also mentioned that their views about the sufferers gradually changed as they entered the *positive outcomes* field of their caregiving activities every day:

[ID 11, full-time employee, wife of sufferer] "I realized that he himself is making an effort. For example, one day when he went to the library, he borrowed a book for me because he thought I might like it. By keeping records, I realized how much he cares about me, the small things that I would normally pay no attention to. (…) I used to think that it was always me, the caregiver, who provides support, but after using this tool, I felt as if we could care for each other."

Overall, the recording items of FMCT (particularly filling out the positive outcomes field) seemed to facilitate positive self-reflection on the family caregivers' caregiving activities. It provided them a chance to reconsider their caregiving habits and objectively think about how they had been communicating with the sufferers. As a result, many family caregivers realized that their views on sufferers had been biased, which helped them mitigate their biased thoughts of the sufferers.

5.5 Being generous to sufferers

Such changes in the family caregivers' views affected their attitude toward the sufferers. Aside from some specific coping strategies to deal with unexpected sufferer behaviors, many family caregivers mentioned that they became more generous to them:

[ID 10, working at home, father of sufferer] "I don't think this [FMCT] tool is critical, but it does have its uses. It changed the way I communicate with my daughter. I used to brush her off and criticize her opinions. But I started to feel that I should be more generous and listen to her until she was finished without criticizing her, even when I think she's wrong."

[ID 11, full-time employee, wife of sufferer] "I can now see that the disease is the troublemaker, not him. (…) I started to avoid saying negative things to him. Consequently, FMCT seemed to give family caregivers breathing room to take a step back and communicate with the sufferers in a more relaxed manner."

5.6 Achieving better communication with sufferers

Finally and most importantly, although we were concerned that family caregivers' use of FMCT might create patient-caregiver conflicts, none of the family caregivers reported such an effect. In fact, most reported the opposite:

[ID 2, full-time employee, daughter of sufferer] "The amount of communication with my mom definitely increased. Previously, communication often started with a complaint, for example, "Why did you do this or that?", and so our communication tended to be antagonistic. (…) During this study, I had to ask her some questions to record her daily activities. After I returned from work, I asked her about her day, whether she had eaten lunch, how she was feeling, etc. There was no reason to become antagonistic about these questions. They can be asked in a normal tone. My mom also looked pleased when I asked them. I think she felt cared for."

Some even reported that the sufferers became more proactive when talking about their health conditions: *[ID 3, homemaker, wife of sufferer]* "To record his condition, I often asked him some questions such as whether he had slept well last night and so on. Gradually, he started to talk about his condition before I asked him. (…) Previously, it was comments like "I'm tired" or "I'm exhausted," and nothing else. But he's started to elaborate on his condition and show that he is thinking about the future. (…) I felt like he was indirectly participating in the study. It's good to have a common goal and to collaborate in keeping the record and fighting the disease."

As with *IDs 2* and *3*, although family caregivers initially asked questions of the sufferers (e.g., whether they had eaten lunch or not) to maintain the FMCT records, those questions seemed to please the sufferers and open up thoughtful communication between the sufferers and the family caregivers. Overall, the recording items seemed to serve as a spring board to start a safe conversation with the sufferers. Together with the family caregivers' relaxed attitudes, such conversation led them to a positive communication cycle.

6. Conclusion

This article explained the potential of tracking technologies to support family caregivers in interacting more effectively with sufferers of depression. Our findings are significant in three ways. First, the manual input of sufferer mood/behavior by family caregivers promoted higher attentiveness to the sufferers and helped them understand the sufferer better. Second, the selected recording items and the holistic view of sufferer mood/behavior and caregiving activities helped the family caregivers develop effective coping strategies. Third, the recording items served as a spring board to initiate a safe conversation with the sufferers. Altogether, our technology probe changed the ecology of the family caregivers, sufferers, and the disease from *sufferer with disease* vs. *family caregiver* to *sufferer and family caregiver* vs. *disease*. We believe our study will open up new ways of using tracking technologies in the field of mental healthcare.

Naomi Yamashita
Primary Researcher, Communication Environment Research Group, Media Information Laboratory, NTT Communication Science Laboratories.
She received a B.Eng. and M.Eng. in applied mathematics and physics and a Ph.D. in informatics from Kyoto University in 1999, 2001, and 2006. She joined NTT in 2001. Her primary interests lie in the areas of computer-supported cooperative work and computer mediated communication. She aims to use a combination of quantitative and qualitative research methods to uncover the nature of human discourse/interaction and propose guidelines for designing new communication technologies.

Indium Phosphide-based Heterojunction Bipolar Transistors with Metal Subcollector Fabricated Using Substrate-transfer Technique

Yuta Shiratori, Takuya Hoshi, Norihide Kashio, Kenji Kurishima, and Hideaki Matsuzaki

Abstract

We fabricated indium phosphide (InP)-based heterojunction bipolar transistors (HBTs) with a highly thermal conductive gold (Au) subcollector on a silicon carbide substrate using a substrate-transfer technique. The fabricated HBTs show good electrical characteristics without any degradation caused by the transfer process. In addition, they exhibit about a 50% reduction in thermal resistance (R_{th}) compared with conventional HBTs on an InP substrate. The reduced R_{th} enables us to increase collector current density without a rise in the junction temperature of HBTs, which improves the HBT high-frequency performance. The fabricated Au-subcollector HBTs have great potential for boosting the operation speed of future telecommunications integrated circuits.

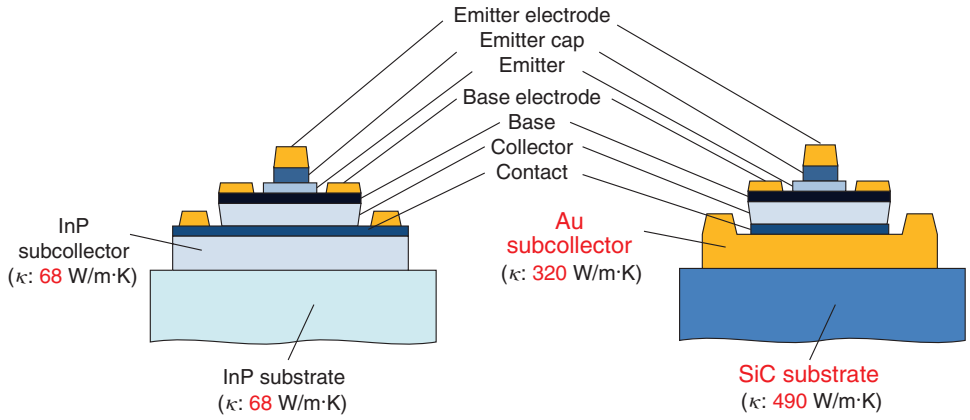
Keywords: InP-based HBT, thermal resistance, wafer bonding

1. Introduction

Indium phosphide (InP)-based heterojunction bipolar transistors (HBTs) have excellent high-frequency performance suitable for large-bandwidth integrated circuits (ICs). Increasing the current density of HBTs is very effective for improving the operating speed of these ICs because the increased current density helps to reduce the charging time. We demonstrated a current-gain cutoff frequency of 513 GHz and a maximum oscillation frequency of 637 GHz for a 0.25-μm-wide-emitter HBT at the collector current density (J_C) of 9.5 mA/μm^2 [1]. However, the increased current density results in higher power consumption per unit area, which leads to a rise in junction temperature and degraded current gain of the HBTs.

To suppress these undesirable effects, it is necessary to improve the heat transfer from inside the HBTs to the backside of the substrate. A basic approach to suppressing them is to thin the substrate or change the substrate to one with high thermal conductivity. However, drastically improving the heat transfer is a major challenge because of the possible degradation of the electrical performance associated with the fabrication process.

To improve the heat transfer, we propose a novel HBT structure with a gold (Au) subcollector fabricated on a silicon carbide (SiC) substrate using a substrate-transfer technique. The proposed HBT successfully suppresses the degradation of the current gain at high current density by reducing the thermal resistance, thanks to the highly thermal conductive SiC substrate and the simultaneously introduced highly thermal conductive Au subcollector.



(a) Conventional HBT on InP substrate (b) Metal-subcollector HBT

Fig. 1. Cross-sectional view of (a) conventional InP-based HBT on InP substrate and (b) metal-subcollector HBT on SiC substrate.

2. Device structure and fabrication process

A cross-sectional view of the proposed HBT structure is shown in **Fig. 1**. For comparison, a conventional InP-based HBT structure is also shown. The conventional HBTs are formed on an InP substrate, and the subcollectors are also made of InP. In contrast, the substrate for the proposed HBT is SiC, and the subcollectors are not a semiconductor but a metal (Au). We therefore call the proposed HBT a metal-subcollector HBT. Because SiC and Au have much higher thermal conductivity (490 and 320 $Wm^{-1}K^{-1}$, respectively) than InP (68 $Wm^{-1}K^{-1}$), metal-subcollector HBTs can enhance the heat transfer compared with HBTs on an InP substrate. In addition, the collector current of metal-subcollector HBTs flows vertically through an indium gallium arsenide (InGaAs) contact layer because the Au subcollector also acts as a collector electrode. This is a clear advantage for heat transport in the metal-subcollector HBTs because the thickness of the InGaAs contact layer, which has very low thermal conductivity of 5 $Wm^{-1}K^{-1}$, can be reduced without increasing its electrical resistance.

The epitaxial layers for metal-subcollector HBTs are designed for high-current-density operation (> 10 $mA/\mu m^2$). In addition, we prepared three kinds of HBTs with different collector contact layer (t_{cc}) thicknesses (50, 30, or 0 nm) to investigate the effect of the thickness of the InGaAs t_{cc} on the thermal resistance (R_{th}) of HBTs. The emitter is 20-nm-thick InP. The gallium arsenide antimonide (GaAsSb) base is 30 nm thick and highly C-doped at 8×10^{19} cm^{-3} (for t_{cc} of 50 nm) or 4×10^{19} cm^{-3} (for t_{cc} of 30 and 0 nm). The collector is 100-nm-thick InP.

The fabrication process of HBTs started with the epitaxial growth of HBT layers on a 3-inch InP substrate by metal-organic chemical vapor deposition. After Au was deposited as an adhesive and subcollector material on both the HBT layers and a 3-inch SiC substrate, they were bonded using a surface activated bonding (SAB) method as shown in **Fig. 2(a)**. SAB prevents thermal degradation of the epitaxial layers due to the sufficiently low bonding temperature (~150°C) [2]. Note also that the growth sequence of the epitaxial-layer structure was inverted for the wafer bonding, which started from the emitter cap layer.

A scanning acoustic microscope image of the bonded wafer is shown in **Fig. 3(a)**. The InP and SiC substrates were bonded with almost no voids except at the wafer edge, thanks to the excellent physical properties of Au (low Young's modulus and non-oxidizing). After the removal of the InP substrate as shown in **Fig. 2(b)**, HBT epitaxial layers were completely transferred onto the SiC substrate as shown in **Fig. 3(b)**. Finally, HBTs with an emitter width of 0.25 μm were fabricated using the SiO_2/SiN (silicon dioxide/silicon nitride) sidewall process [3] as shown in **Fig. 2(c)**. A scanning electron microscope image of a fabricated metal-subcollector HBT with an emitter size of 0.25 μm × 4.0 μm is shown in **Fig. 4**. HBTs were successfully fabricated on a SiC substrate without any serious problems related to the bonding process.

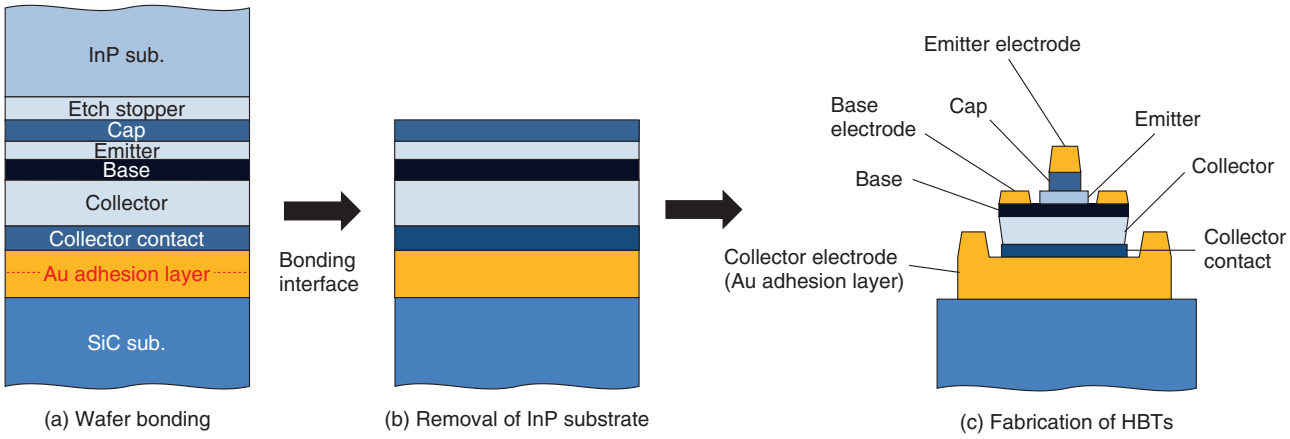
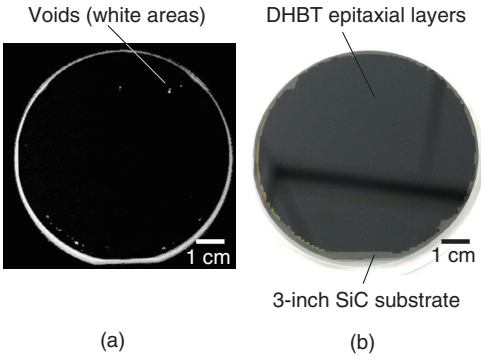


Fig. 2. Fabrication steps for metal-subcollector HBT.



DHBT: double heterojunction bipolar transistor

Fig. 3. (a) Scanning acoustic microscope image of HBT epitaxial layers on an InP substrate bonded to an SiC substrate. (b) Photograph of the bonded wafer after removal of InP substrate.

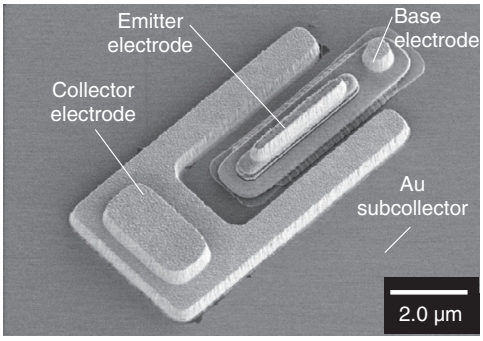


Fig. 4. Scanning electron microscope image of a fabricated metal-subcollector HBT with an emitter size of 0.25 μm × 4.0 μm.

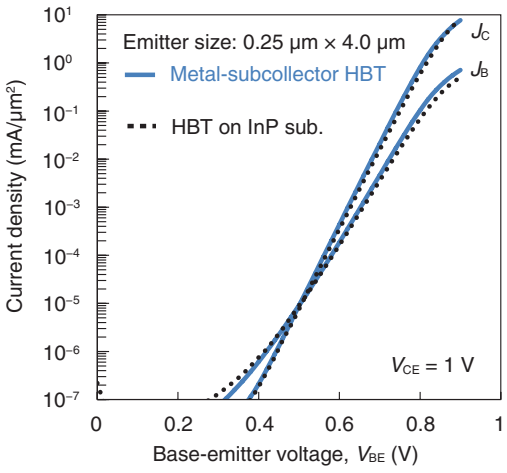
3. Device characteristics

We measured the electrical characteristics of the fabricated HBTs to investigate how the bonding process affected the HBTs. Gummel plots for the metal-subcollector HBT on a SiC substrate (hereinafter referred to as an HBT on SiC) with a 50-nm-thick InGaAs collector contact layer are shown in **Fig. 5**. For comparison, Gummel plots for a conventional HBT on an InP substrate (hereinafter referred to as an HBT on InP) with the same epitaxial layer structure are also shown (black dotted lines). As seen in the graph, the two HBTs had almost the same current-transfer characteristics. We confirmed from high-frequency measurements that the HBT on SiC exhibits a current-gain cutoff frequency of 337 GHz and a maximum oscillation frequency of 400 GHz at J_C of 15 mA/μm^2. These values are comparable to those of the HBT on InP. These results indicate that the carrier transport properties were not degraded by the bonding process.

Next, to investigate heat transport properties, we estimated the R_{th} for the HBTs from the electrical characteristics in a common-base configuration. R_{th} is expressed by the following equation [4].

$$R_{th} = \frac{1}{\phi}\frac{\Delta V_{BE}}{\Delta P},$$

where ΔV_{BE} is a voltage shift in the base-emitter voltage for a given collector current when the collector-base voltage increases, ΔP is an associated increase in power dissipation, and ϕ is a thermo-electric feedback coefficient, experimentally obtained from the



J_C: collector current density
J_B: base current density

Fig. 5. Gummel plots of fabricated HBTs.

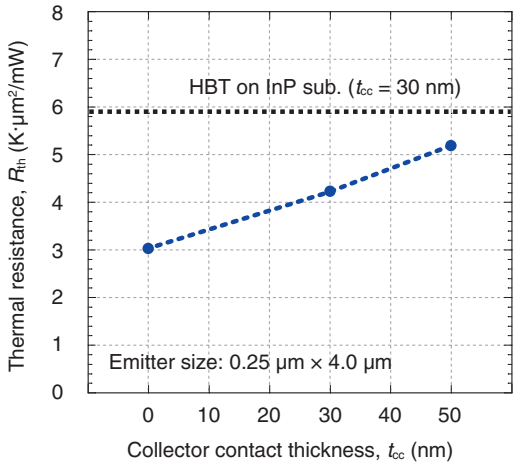


Fig. 6. Thermal resistance as a function of collector contact thicknesses of fabricated HBTs.

Gummel plots at various ambient temperatures. For the InP/GaAsSb HBTs, the measured ϕ is about 0.9 mV/K. In this article, the R_{th} is normalized by the total emitter junction area.

The R_{th} as a function of t_{cc} for an HBT on SiC is plotted in **Fig. 6**. For comparison, the R_{th} of an HBT on InP with t_{cc} of 30 nm is also plotted (black dotted line). When t_{cc} is 30 nm, R_{th} for the HBT on SiC is about 30% lower than that for the HBT on InP due to the effect of the highly thermal conductive SiC substrate and Au subcollector. In addition, the R_{th} for the HBT on SiC decreases with t_{cc}. In particular, the HBT on SiC without the InGaAs contact layer (t_{cc} = 0 nm) exhibits R_{th} of 3.02 Kμm^2/mW, which is about a 50% reduction compared to the HBT on InP. The obtained R_{th} values are reasonable compared with those estimated from the analytical model of R_{th} [5]. It is expected that the R_{th} can be further reduced by optimizing the thickness of the Au subcollector and utilizing substrates with higher thermal conductivity such as diamond.

Collector *I-V* curves for the HBTs without the InGaAs contact layer (t_{cc} = 0 nm) on SiC are shown in **Fig. 7**. Junction temperatures (T_j) were estimated from the measured R_{th} and are plotted as dotted curves. Due to the reduced R_{th}, the T_j of the HBT on SiC is less than about half that of the HBT on InP. As a result, the former successfully suppresses a reduction in current-gain in the high-V_{CE} and high-I_C bias regions (~20 mW/μm^2). The proposed HBT is therefore very effective for increasing the operation current density without degradation of the current gain due to device self-heating.

4. Summary

We fabricated InP-based HBTs with a Au subcollector on a SiC substrate using the substrate-transfer technique. From a comparison of electrical characteristics between the metal-subcollector HBT on SiC and a conventional HBT on an InP substrate, we confirmed that the wafer-bonding process did not degrade the carrier transport properties of the HBT. In addition, the HBTs on the SiC substrate exhibited about a 50% reduction in R_{th} compared with those on the InP substrate. The proposed metal-subcollector HBT technology is very promising for improving the speed performance of ICs to support future large-capacity telecommunications network systems.

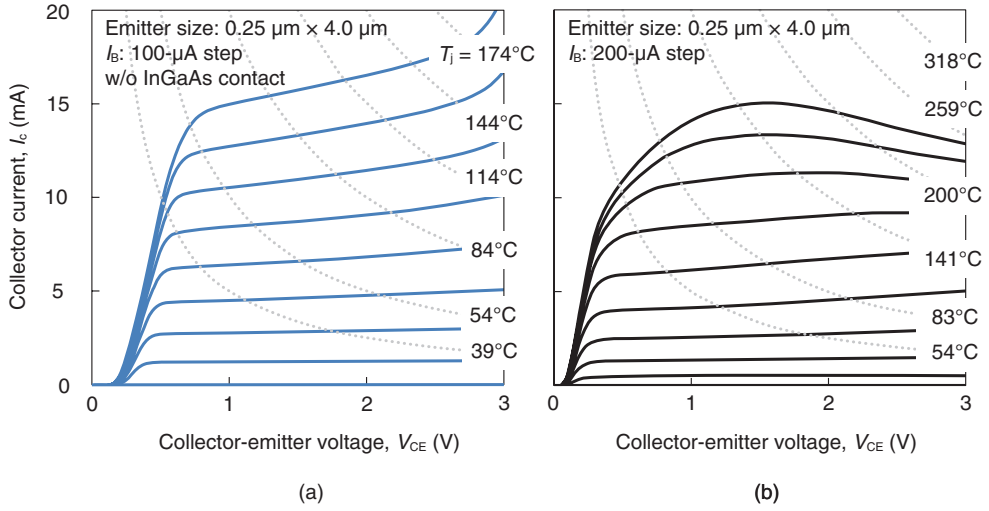


Fig. 7. I_C-V_{CE} characteristics of HBTs fabricated on (a) SiC substrate and (b) InP substrate. Dotted lines show junction temperatures estimated from R_{th}.

References

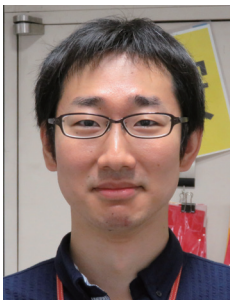
[1] N. Kashio, T. Hoshi, K. Kurishima, M. Ida, and H. Matsuzaki, "Improvement of High-frequency Characteristics of InGaAsSb-base Double Heterojunction Bipolar Transistors by Inserting a Highly Doped GaAsSb Base Contact Layer," IEEE Electron Device Lett., Vol. 36, No. 7, pp. 657–659, 2015.

[2] Y.-H. Wang, J. Liu, and T. Suga, "Low-temperature Wafer Bonding Using Gold Layers," Proc. of ICEPT-HDP '09 (International Conference on Electronic Packaging Technology & High Density Packaging), pp. 516–519, Beijing, China, Aug. 2009.

[3] N. Kashio, K. Kurishima, M. Ida, and H. Matsuzaki, "Over 450-GHz f_t and f_{max} InP/InGaAs DHBTs with a Passivation Ledge Fabricated by Utilizing SiN/SiO_2 Sidewall Spacers," IEEE Trans. Electron Devices, Vol. 61, No. 10, pp. 3423–3428, 2014.

[4] W. Liu, "Handbook of III-V Heterojunction Bipolar Transistors," p. 363, Wiley-Interscience, New York, 1998.

[5] F. N. Masana, "A New Approach to the Dynamic Thermal Modeling of Semiconductor Packages," Microelectron. Reliab., Vol. 41, No. 6, pp. 901–912, 2001.



Yuta Shiratori

Researcher, Materials and Devices Laboratory, NTT Device Technology Laboratories.

He received his B.E., M.E., and Ph.D. in electrical engineering from Hokkaido University in 2006, 2008, and 2011. He joined NTT Photonics Laboratories in 2011, where he conducted research on InP-based double heterojunction bipolar transistors (DHBTs) and wafer bonding techniques. He is a member of the Japan Society of Applied Physics (JSAP).

Takuya Hoshi

Researcher, Materials and Devices Laboratory, NTT Device Technology Laboratories.

He received a B.E. from University of Tsukuba, Ibaraki, in 2007, an M.E. from Tohoku University, Miyagi, in 2009, and a Ph.D. in engineering from University of Tsukuba in 2015. He joined NTT Photonics Laboratories in 2009, where he was engaged in research on metal organic chemical vapor deposition (MOCVD) growth and InP-based DHBTs. He is a member of JSAP and the Institute of Electrical Engineers of Japan (IEEJ).

Norihide Kashio

Senior Research Engineer, Photonic Network Device Project, NTT Device Innovation Center.

He received his B.E., M.E., and Dr. Eng. from Waseda University, Tokyo, in 2000, 2002, and 2012. He joined NTT Photonics Laboratories in 2002, where he researched ultrahigh-speed InP-based optical and electrical devices for future optical fiber communications systems. He is a member of JSAP and IEEJ.

Kenji Kurishima

Senior Research Engineer, Materials and Devices Laboratory, NTT Device Technology Laboratories.

He received his B.S., M.S., and Dr. Eng. in physical electronics from Tokyo Institute of Technology in 1987, 1989, and 1997. He joined NTT Atsugi Electrical Communications Laboratories in 1989, where he was involved in research and development of InP-based HBTs and MOVPE growth. His current research interests include the design and fabrication of high-speed electronic devices for future communications systems. He is a member of JSAP.

Hideaki Matsuzaki

Senior Research Engineer, Supervisor, Materials and Devices Laboratory, NTT Device Technology Laboratories.

He received a B.S. and M.S. in physics from Kyoto University in 1993 and 1995. In 1995, he joined NTT Atsugi Electrical Communications Laboratories, where he was engaged in researching and developing fabrication technologies for high-speed resonant-tunneling diode (RTD)/high-electron mobility transistors (HEMT)- and photo-diode/RTD/HEMT-based circuits. Since 2003, he has studied and developed ultrahigh-speed and low-noise heterostructure field-effect transistors and their monolithic microwave integrated circuits on InP substrates. He served as a senior manager in the Technical Division at NTT EAST from 2006 to 2008 and as a senior research engineer in the Research Planning Division from 2009 to 2011. He is a member of the Institute of Electrical and Electronics Engineers and the Institute of Electronics, Information and Communication Engineers.

Standardization Activities in International Electrotechnical Commission Technical Committee 86 (Fiber Optics)

Takashi Matsui, Noriyuki Araki, and Hisashi Izumita

Abstract

The International Electrotechnical Commission Technical Committee 86 (IEC TC86) is an international standardization organization that prepares and decides on international standards in relation to products used for optical fiber telecommunication. In this article, we provide an overview of standardization activities, introduce topics discussed at meetings in 2015 and 2016, and describe the Japanese standardization strategy in IEC TC86.

Keywords: IEC, optical fiber, international standard

1. Introduction

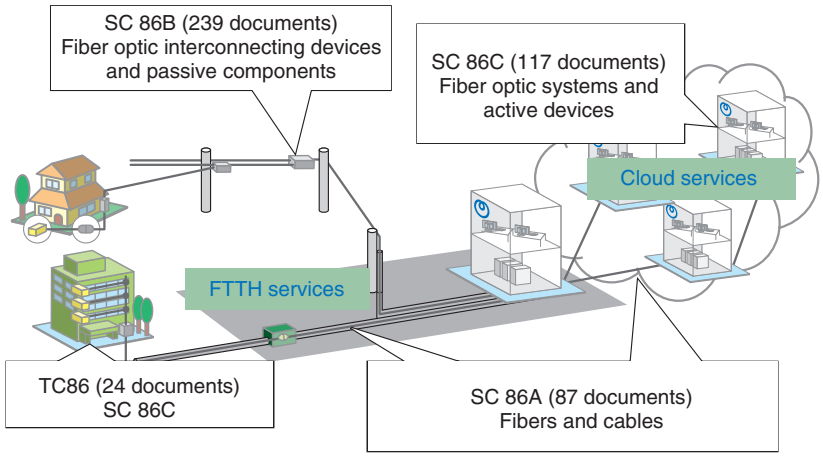
The International Electrotechnical Commission (IEC) is a standardization organization that sets all of the required standards for electrical and electronics technologies. It develops international standards for product specifications and provides a conformity assessment regarding the safety and quality of products.

The scope of Technical Committee 86 (TC86) is to prepare standards for fiber optic systems, modules, devices, and components intended primarily for use with communications equipment. This activity covers terminology, characteristics, related tests, calibration and measurement methods, functional interfaces, and optical, environmental, and mechanical requirements to ensure reliable system performance.

The technical scope of IEC TC86 is shown in **Fig. 1**, and its structure is indicated in **Fig. 2**. TC86 has three subcommittees (SCs) that can make decisions: SC 86A concerns fibers and cables, SC 86B concerns fiber optic interconnecting devices and passive components, and SC 86C concerns fiber optic systems and active devices. The TC and SCs have 13 Working Groups (WGs) that engage in technical discussions regarding standards documents. The number of published documents and the documents under discussion in the TCs are indicated in **Fig. 3**. TC86 is one of the largest TCs in the IEC and is responsible for the preparation of 400 or more standards documents. More than 300 experts from 30 countries are participating in the related discussions.

2. Recent fiber optics standardization topics

The standardization of optical fiber cables and connecting devices for fiber optics has been aggressively discussed for use in access network systems because of the rapid growth of FTTH (fiber-to-the-home) services throughout Europe and developing countries. SC 86A has established a new international standard for optical fiber ribbon, which included the Japanese proposal of partially bonded ribbon. New test procedures for optical fiber cables have been proposed by several (mainly European) countries that take into account the different environments and installation methods used in those countries. Japan also proposed standardizing a new test procedure for the friction



FTTH: fiber to the home

Fig. 1. Technical area of IEC TC86.

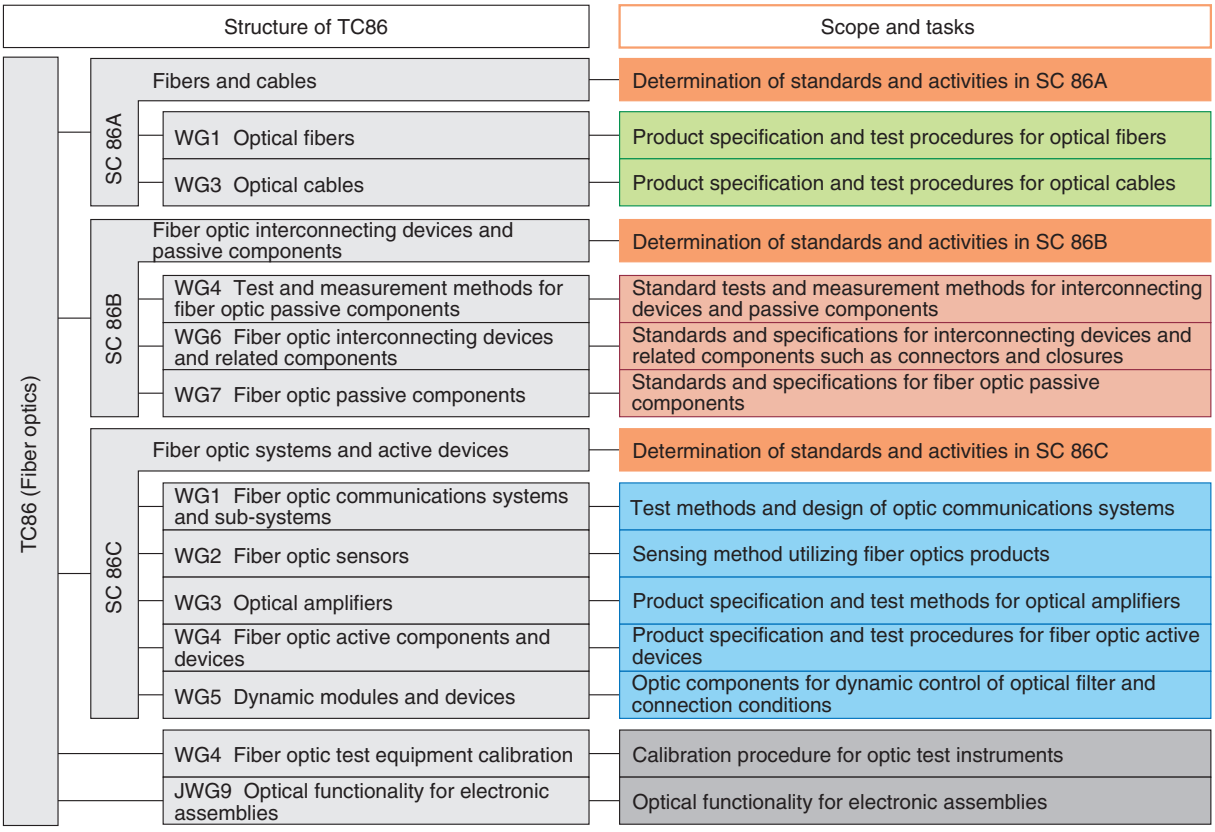


Fig. 2. Structure of IEC TC86.

coefficient between cables, and this proposal has been accepted. SC 86B holds active discussions on standardizing field mountable connectors, which are fusion splice connectors for use in Europe and

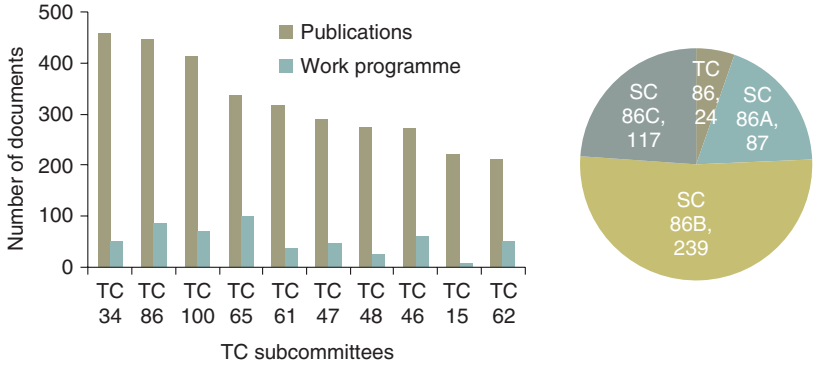


Fig. 3. Number of documents published and under discussion in each TC.

mechanical splice connectors for use in Japan.

IEC TC86 works closely with the International Telecommunication Union - Telecommunication Standardization Sector (ITU-T) Study Group 15 (SG15) to define the functional requirements for public fiber telecommunication networks. The development of a standard for optical connectors is being discussed separately under collaboration between ITU-T and IEC, and the functional requirements are being considered in ITU-T and the product specifications in IEC.

In view of the huge growth of datacenter technology recently, increasing attention has been focused on developing standards that take datacenter use into consideration. In particular, in SC 86C, the standardization of photonic integrated circuits (PICs) has been discussed, and TC86 decided to develop new standards for PICs in SC 86C/WG4 in 2015. SC 86A has accelerated discussions related to revising product specifications and test procedure standards for multimode fibers, which are used for short reach communications over less than several hundred meters in datacenters. SC 86B has discussed performance specification standards for multiple-fiber push-on/pull-off (MPO)-type optical connectors, which are used to connect fiber ribbons and optical transceivers for parallel transmission.

The fiber optics technologies discussed in TC86 cover many industries including automotive communication, sensing, smart grids, and industrial automation. TC86 is proceeding with standardization efforts in collaboration with various standardization organizations. The collaboration structure in TC86 is shown in **Fig. 4**. ITU-T SG15 is very closely related to TC86 in dealing with systems and products used in public telecommunication. ITU-T SG15 and TC86 share information and exchange inquiries and respectively discuss functional requirements and product specifications. For smart grid technology, which uses both power lines and telecommunication lines, TC86 established a liaison with TC46, which discusses standards for coaxial cables. In addition, SC 86A established a joint WG with SC 46C.

SC 86C established WG2 to study optical sensors in 2013 in order to address the use of fiber optics in sensing applications. Academia and measurement instrument manufacturers in Japan and Italy play a central role in discussing the standardization of optical sensors in SC 86C/WG2, and WG2 has discussed establishing standards for optical temperature and strain sensors as an urgent issue.

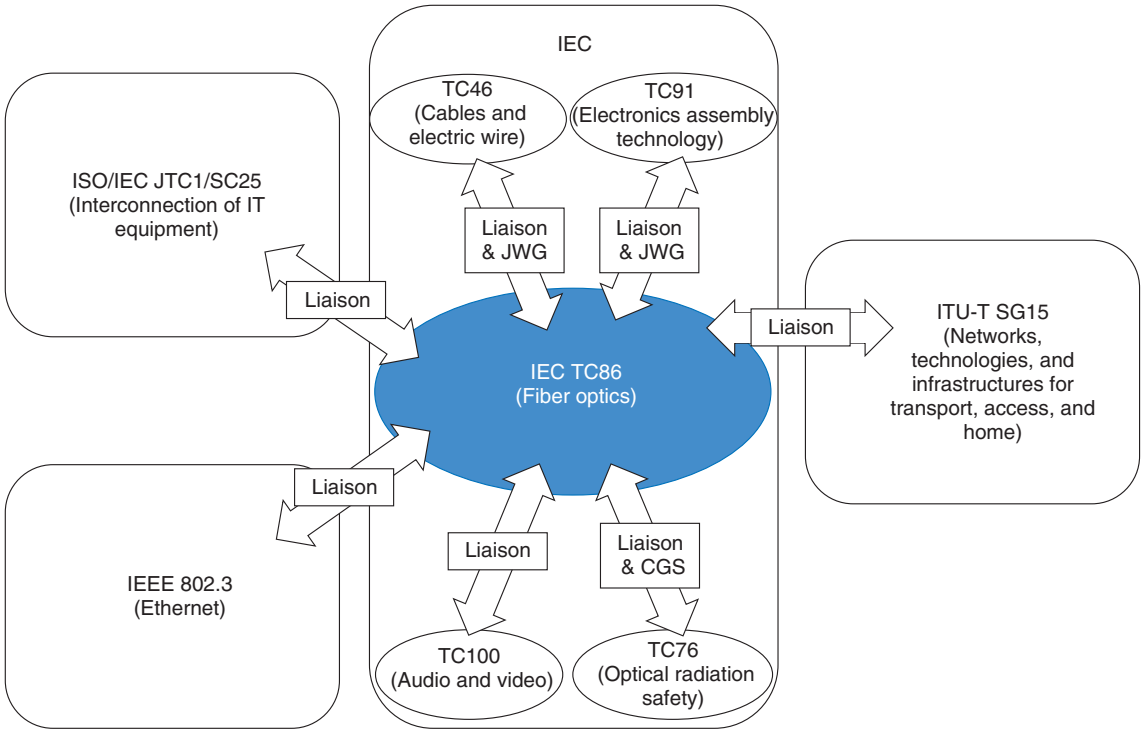
3. Latest topics

(1) SC 86A/WG1: Optical fibers

WG1 has finished the comment resolution for new product specifications and test procedures for a polarization-maintaining optical fiber proposed by Japan and has agreed to accelerate discussion toward establishing new standards. WG1 has also agreed to revise the product specifications of A1 type multimode optical fiber (IEC 60792-2-10) to include a new product, namely, wide-band multi-mode optical fiber for wavelength-division multiplexing transmission.

(2) SC 86A/WG3: Optical cables

WG3 agreed to circulate a modified draft of a new standard for optical fiber ribbon that Japan had proposed. There was an active discussion on adding new test procedures and improving the existing tests for optical cables, and WG3 agreed to include unchanged the procedure proposed by Japan for testing the friction coefficient between cables. WG3 also agreed to



CGS: Coordination Group on Safety
ISO: International Organization for Standardization
IT: information technology
JTC: Joint Technical Committee
JWG: Joint Working Group

Fig. 4. Collaboration structure of TC86.

initiate a new topic concerning optical cable for shared use both inside and outside.

(3) SC 86B/WG4: Standard tests and measurement methods for fiber optic interconnecting devices and passive components

Japan proposed a method for testing the strength of a mounted adapter, and WG4 agreed to circulate the revised draft. Many proposals were made that were related to new topics such as the visual inspection of fiber optic connectors and the measurement of fiber angle and position in a multi-fiber connector ferule by utilizing a computed tomography scan.

(4) SC 86B/WG6: Standards and specifications for fiber optic interconnecting devices and related components

WG6 agreed to change the spring strength to 20N in the interface specifications of the 2-fiber-row MPO connector. WG6 is continuing to discuss the loss of a field mountable connector at a wavelength of 1625 nm since Japan insisted on considering existing products.

(5) SC 86B/WG7: Standards and specifications for fiber optic passive components

Japan proposed changing the document structure of the generic standard for optical isolators. SC 86B held a joint meeting of WG6 and WG7 to consider a generic guideline for the performance standard of optical passive components, and discussed the definition of the closure and strength test for fiber clamps.

(6) SC 86C/WG1: Fiber optic communications systems and sub-systems

WG1 discussed revising the generic standard for the optic sub-system that Japan proposed and also discussed a method to measure loss in multi-mode fiber transmission lines, a method to measure polarization mode dispersion in transmission lines, chromatic dispersion compensation methods, and PON (passive optical network) system loss measurement.

(7) SC 86C/WG2: Fiber optic sensors

New standards for fiber Bragg grating based strain sensing (IEC61757-1-1) and distributed temperature

sensing (IEC61757-2-2) were published in 2016. Japan proposed standardizing electric current sensing, and WG2 agreed to discuss it as a new issue. WG2 plans to discuss fiber Bragg grating based temperature sensing, distributed vibration sensing, and distributed strain sensing.

(8) SC 86C/WG3: Optical amplifiers

WG3 discussed a new draft and the subject of a new project. SC 86C held a joint meeting of WG3 and WG4 on the standardization of the semiconductor optical amplifier and discussed in particular a gain ripple measurement method.

(9) SC 86C/WG4: Fiber optic active components and devices

WG4 discussed new drafts concerning PIC packages and radio over fiber. WG4 also discussed a performance template for a semiconductor optical amplifier in a joint meeting with WG3.

(10) SC 86C/WG5: Dynamic modules and devices

WG5 discussed new drafts for the performance template of a multi-cast optical switch and the transient crosstalk measurement of a wavelength selective switch. WG5 agreed to investigate tests concerning the reliability and high power resistance of dynamic modules.

(11) TC86/WG4: Fiber optic test equipment calibration

Comments and voting results of a draft method for calibrating an optical spectrum analyzer were confirmed and published. WG4 agreed to develop the technical specifications for the optical frequency analyzer calibration method as an international standard in the future.

(12) TC86/JWG9: Optical functionality for electronic assemblies

JWG9 (Joint Working Group 9) confirmed a liaison request from the Consortium for On-Board Optics (COBO) and agreed to establish the liaison after a decision had been reached in TC86.

4. Further standardization activities of TC86

The secretary of SC 86B, the chair of SC 86C, the conveners of several WGs, and many project leaders of standards discussions are from Japan. The Japanese influence in TC86 is strong, and we lead the standardization discussions in order to communicate the superiority of Japanese technologies. We are also proceeding with domestic collaboration between ITU-T and IEC for strategic standardization activities involving both standardization organizations. Recently, the discussion has spread to new application areas such as datacenters and automotive applications, and collaboration is under way with participants in other standardization bodies such as ITU-T and the Institute of Electrical and Electronics Engineers (IEEE).

In June 2016, IEC President Junji Nomura, IEC General Secretary Frans Vreeswijk, and James Shannon, the IEC President-Elect, visited NTT headquarters in Tokyo and discussed a master plan for a standardization strategy for the next five years. At this meeting, they discussed the importance of both the standardization of industrial products and the Internet of Things, the system standard for smart cities, and conformity assessment. TC86 is planning to deal with these topics strategically and to strengthen collaboration with standardization bodies such as ITU-T that discuss system requirements.

Takashi Matsui
Senior Research Engineer, Access Media Project, NTT Access Network Service Systems Laboratories.

He received his B.E., M.E., and Ph.D. in electronic engineering from Hokkaido University in 2001, 2003, and 2008. He attained the status of Professional Engineer (P.E.Jp) in electrical and electronic engineering in 2009. In 2003, he joined NTT Access Network Service Systems Laboratories and has been researching optical fiber design techniques. He has been contributing to the activities of IEC TC86 (fiber optic systems) as an expert of SC 86A/WG1 and WG3 since 2009. Dr. Matsui is a member of the Institute of Electronics, Information and Communication Engineers (IEICE).

Noriyuki Araki
Senior Research Engineer, Access Media Project, NTT Access Network Service Systems Laboratories.

He received his B.E. and M.E. in electrical and electronic engineering from Sophia University, Tokyo, in 1993 and 1995. He joined NTT Access Network Service Systems Laboratories in 1995, where he researched and developed operation and maintenance systems for optical fiber cable networks. He has been contributing to standardization efforts in ITU-T SG6 since 2006. He was the rapporteur of Question 6 of ITU-T SG6 from 2006 to 2008 and the rapporteur of Question 17 of ITU-T SG15 from 2008 to 2012. He also served as the chairman of the ITU-T Focus Group on Disaster Relief Systems and Network Resilience and Recovery. He has been the vice-chairman of ITU-T SG15 since 2013. He also contributes to the activities of IEC TC86 (fiber optic systems). He received the ITU-AJ award from the ITU Association of Japan in 2012. He is a member of IEICE.

Hisashi Izumita
Senior research engineer, supervisor at NTT Access Network Service Systems Laboratories and a visiting professor at Ibaraki University.

He received a B.S. and M.S. in physics and a Ph.D. in engineering from Waseda University, Tokyo, in 1987, 1989, and 2008. He joined NTT Transmission Systems Laboratories in 1989 and engaged in research on high performance optical time domain reflectometry using coherent lightwave technologies and optical fiber distributed sensing applications. He also developed an optical fiber distribution system for central offices, an optical fiber maintenance and support system, and optical fiber cable systems. He has contributed to the activities of IEC TC86 WG4 since 1999 and IEC SC 86A/WG1 and WG3 since 2012. Dr. Izumita is a member of IEEE and the Optical Society of America, and a senior member of IEICE.

Event Report: NTT Communication Science Laboratories Open House 2016

Jun Suzuki, Norimichi Kitagawa, Masaru Tsuchida, Katsuhiko Ishiguro, and Shinobu Kuroki

Abstract

NTT Communication Science Laboratories Open House 2016 was held in Keihanna Science City, Kyoto, on June 2 and 3, 2016. Nearly 1300 visitors enjoyed 6 talks and 29 exhibits, which focused on our latest research activities and efforts in the fields of information and human sciences.

Keywords: information science, human science, artificial intelligence

1. Overview

At NTT Communication Science Laboratories (NTT CS Labs), two kinds of fundamental research are underway. The first is research to create the future. We are leading the world in preparing for a new era by developing technologies that will bear fruit five to ten years in the future and by accumulating knowledge. The second is research that addresses the needs of today. We are utilizing the technologies and knowledge we have cultivated to tackle the societal challenges directly confronting us now.

NTT CS Labs Open House has been held annually with the aim of introducing the results of the labs' basic research and innovative leading-edge research to both NTT Group employees and visitors from various industries, universities, and research institutions who are engaged in research, development, business, and education.

This year, the open house was held at the NTT Keihanna Building in Kyoto on June 2 and 3, and nearly 1300 visitors attended it over the two days. We prepared many hands-on exhibits to allow visitors to intuitively understand our latest research results and to share a vision of the future where new products based on the research results are widely used. We also organized an invited talk on the theme of communication. This article summarizes the event's research talks and exhibits.

2. Keynote speech

The open house started with a speech by the Vice President and head of NTT CS Labs, Dr. Eisaku Maeda, entitled "From information transmission to mutual understandings—Paradigm shift in the age of data," (**Photo 1**).

Dr. Maeda introduced NTT Group's artificial intelligence technology called corevo™, which creates

Photo 1. Dr. Eisaku Maeda, Vice President, Head of NTT CS Labs, giving the keynote speech.

models of intelligence from all kinds of data related to people, for example, written words, speech, images, music, vital signs, muscle activity, and brain activity. Through this process, intelligence is componentized by disassembling it into a variety of forms and then reassembling it. An era will certainly arrive sooner or later in which intelligence is treated as a product in the market economy. In this talk, he also pointed out the fact that the flow of physical time does not change, and the lifespan of human beings as organisms does not drastically change. However, environments that surround people are undergoing radical transformation despite these fixed limitations. He stated that NTT CS Labs is pursuing and exploring deeply and broadly the essence of communication between people and between people and machines in order to draw this future blueprint.

3. Research talks

Four talks were given, as summarized below, which highlighted recent significant research results and high-profile research themes. Each presentation introduced some of the latest research results and provided some background and an overview of the research. All of the talks were very well received.

(1) "Avenues toward super-human speech recognizers—Advances in deep learning and signal processing that are making speech recognition leap forward," by Dr. Takuya Yoshioka, Media Information Laboratory

Dr. Yoshioka described how deep learning has been used and how it has changed the way speech recognition is performed and described microphone array processing technologies that make speech recognition available in different situations. He also talked about the future of speech recognition technology and how it will impact our lives (**Photo 2**).

(2) "How to transmit high-quality sound via networks—Research and deployment of speech codec EVS for VoLTE and lossless audio codec ALS for broadcasting," by Dr. Yutaka Kamamoto, Moriya Research Laboratory

Dr. Kamamoto introduced two recent enhancements in speech and audio codecs. One was EVS (Enhanced Voice Service) standardized by 3GPP (3rd Generation Partnership Project), the new standard for speech codec, which is capable of transmitting even the ambient sound on the speaker's side. The other was MPEG-4 ALS (Audio Lossless Coding) defined by the Moving Picture Experts Group, which enables us to perfectly transmit high-quality audio content such as that produced in a studio to our home (**Photo 3**).

Photo 2. Research talk by Dr. Takuya Yoshioka.

Photo 3. Research talk by Dr. Yutaka Kamamoto.

(3) "Hacking human vision, controlling material appearance—Liquid material impression and Hen-GenTou," by Dr. Takahiro Kawabe, Human Information Laboratory

Dr. Kawabe described an unsolved but important question, "How do human observers perceive non-rigid materials such as liquid, smoke, and jelly?" He explained the finding that the human visual system utilizes the dynamic pattern of image deformation to see these materials, and introduced Hen-Gen-Tou, a light projection technique based on this finding (**Photo 4**).

(4) "Creating a computer that can perform casual conversation—Advances and challenges in open domain natural language processing," by Dr. Ryuichiro Higashinaka, Innovative Communication Laboratory / Media Intelligence

Photo 4. Research talk by Dr. Takahiro Kawabe.

Photo 5. Research talk by Dr. Ryuichiro Higashinaka.

Laboratories

Dr. Higashinaka explained current topics related to open domain natural language processing for building casual conversation systems. He also introduced NTT's casual conversation system that is now able to perform casual conversation with users to some extent (**Photo 5**).

4. Research exhibits

The open house featured 29 exhibits displaying NTT CS Labs' latest research results.

We categorized them into four areas: Science of Machine Learning, Science of Computation and Language, Science of Media Information, and Science of Human and Communication. Each exhibit was housed in a booth and employed techniques such as slides on a large-screen monitor or hands-on demonstrations, with researchers explaining the latest results directly to visitors (**Photos 6** and **7**). The following list gives the titles of the research exhibits in each category. More details including the names of researchers associated with each exhibit can be found on the website [1, 2].

4.1 Science of Machine Learning

(1) Learning from a large number of feature combinations—CFM: low-rank regression with global optimality guarantees

(2) Find a good number of salient patterns in a matrix—Infinite Plaid Models for Infinite Bi-clustering

(3) Relational data analysis by infinite trees—R-tree structure grows through stochastic process

Photo 6. Lively discussions were held at the exhibition booths.

Photo 7. Visitors enjoyed hands-on demonstrations.

(4) Navigate people with comfortable traveling route—Dynamic migration scheduling for greater visitor satisfaction
(5) How did you get here? Where will you go?—Trajectory analysis and prediction using deep learning
(6) Where are hotspots of "2020" traffic?—Predict the traffic of future events based on the individual behavior
(7) Fast and accurate deep learning—Efficient learning utilizing directions of past gradients

4.2 Science of Computation and Language

(1) Error correction, lossy compression ... as you like it—Multipotential coding method achieving the Shannon limit
(2) Quantum search over huge network for hidden structures—High-speed search over hypergraphs via quantum walk
(3) Assuring next generation web security—Formal verification of the QUIC and TLS protocols
(4) Generating true randomness for secure ICT—Gbps streaming of physical random numbers
(5) Pitarie: Find a picture book just right for a child—Picture-book search with interdisciplinary approach
(6) Analyze and translate Japanese in various domains—Written and spoken language analysis and machine translation
(7) Find words from age and proportion of acquisition—Creating child language development database by checklist method
(8) We want to talk with you!—Encouraging speech dialogue using multiple robots

4.3 Science of Media Information

(1) Reading designed words appearing in scenes—Optical word recognition with CNN features and WFST decoding
(2) Intonation morphing from one speaker to another—Prosody conversion with generative voice F0 contour model
(3) Additional finding beyond the real thing—Adding information to art using high-definition multiband image
(4) Recognizing your voice even in noisy environments—Advances in distant speech recognition technologies
(5) Speech/audio coding for telephone and broadcasting—Low-bit-rate speech coding EVS and lossless audio coding ALS
(6) Hen-Gen-Tou 2.0 (Animation Lamps)—Automatic projection of motion impression using image search
(7) How to change a dry scene into a scene after the rain—Material editing that makes object surfaces look wet

4.4 Science of Human and Communication

(1) English is OK!—Assistive technology for non-native speakers
(2) Different word acquisition of Japanese and English—Cross-linguistic study of parental input and word acquisition
(3) Making eye-contact with people in teleconference—Social telepresence with eye-contact using kinetic avatars
(4) Ear's delicate coding mechanism—Conversion from amplitude modulation to frequency modulation in the inner ear
(5) Mobile gadget for tugging at your hand—Evolution and application of "Buru-Navi"
(6) Shaping the athletic brain!—Sports performance improvement system based on brain science
(7) Ticket-to-talk: improving intra-family communication—How tracking technologies support caregiving at home

5. Invited talk

This year's event also featured an invited talk by playwright and director Mr. Oriza Hirata. The title of his talk was "Standing together on no common ground," and it explored the idea of what the essence of communication ability is, which is also an important research topic at NTT CS Labs. He pointed out that understanding the context is essentially the most important part of communication, and understanding the context enables us to design communication. To prove this, he showed evidence that was induced from his activities such as directing a play in which both human actors and robots took part and holding workshops using theatrical methods, and from cross-cultural understanding and communication in the biomedical area.

6. Information transmission using web

The members of NTT CS Labs have made a continuous effort to inform a lot of people both domestically and internationally about their research activities and results. As part of the ongoing effort, they simultaneously released both Japanese and English

websites [1, 2] for NTT CS Labs Open House 2016, which included a booklet, exhibition posters, and reference information. They plan to upload photos of the presentation and exhibition halls and videos of the director's keynote speech and the three research talks.

The members of NTT CS Labs consider it important to disseminate high-quality and valuable information on a timely basis. They also have a clear awareness that it is important to convey information on the professional research activities that are underway at the laboratories to a wide range of people in ways that are easy to understand. NTT CS Labs is therefore continuously trying to improve the ways of disseminating their research activities and results.

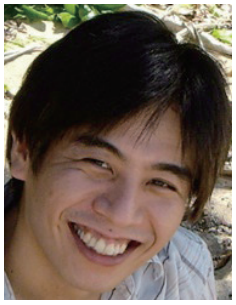
7. Concluding remarks

Just as they did last year, many visitors came to NTT CS Labs Open House 2016 and engaged in lively discussions on the research talks and exhibits and provided many valuable opinions on the presented results. In closing, we would like to offer our sincere thanks to all of the visitors and participants who attended this event.

References

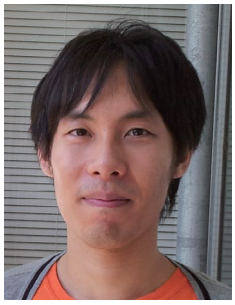
[1] Website of NTT Communication Science Laboratories Open House 2016 (in Japanese), http://www.kecl.ntt.co.jp/openhouse/2016/index.html
[2] Website of NTT Communication Science Laboratories Open House (in English), http://www.kecl.ntt.co.jp/openhouse/2016/index_en.html

Jun Suzuki
Senior Research Scientist, Linguistic Intelligence Research Group, Innovative Communication Laboratory, NTT Communication Science Laboratories.
He received a Ph.D. in engineering from the Graduate School of Information Science, Nara Institute of Science and Technology in 2005. He joined NTT Communication Science Laboratories in 2001. He is currently researching machine learning, natural language processing, and artificial intelligence areas.



Norimichi Kitagawa
Senior Research Scientist, Sensory Resonance Research Group, Human Information Science Laboratory, NTT Communication Science Laboratories.
He received a Ph.D. in psychology from Tokyo Metropolitan University in 2003. His research interests include multisensory perception and perception-emotion interactions.

Masaru Tsuchida
Research Scientist, Recognition Research Group, Media Information Laboratory, NTT Communication Science Laboratories.
He received a Ph.D. from Tokyo Institute of Technology in 2002. His research interests include multispectral imaging, three-dimensional image processing, and computer vision.



Katsuhiko Ishiguro
Research engineer, Learning and Intelligent Systems Research Group, Innovative Communication Laboratory, NTT Communication Science Laboratories*.
He received a Ph.D. in engineering from University of Tsukuba in 2010. He is conducting research on machine learning, especially modeling of structured data.
*He is currently with Mirai Translate, Inc.

Shinobu Kuroki
Research Scientist, Sensory Representation Research Group, Human Information Science Laboratory, NTT Communication Science Laboratories.
She received a Ph.D. in information science and technology from the University of Tokyo in 2011. Her research is focused on human tactile processing, in particular, frequency perception, time perception, and motion perception.

Arkadin's Unified Communications & Collaboration Portfolio and Service Strategy Position it for Leadership in the Virtual Workplace

1. Introduction

The virtual workplace is fast becoming the norm. With trends driven by the cloud, mobility, and an emerging generation of millennials who are used to working on their own terms, most major organizations are implementing virtual work policies that offer more flexibility in how and where work is conducted. Arkadin, an NTT Communications company, is extremely well positioned for leadership in this environment. Its collection of cloud-based unified communications and collaboration (UC&C) technologies that enable anytime, anywhere, any device communications and virtual workforce engagement are essential enablers.

Cloud-based UC services that combine audio, web, and video collaboration in particular, in a convenient all-in-one solution, will take on more importance. Businesses will benefit from the scalability, simplicity, and cost benefits compared to on-premise UC services. Recent developments underscore Arkadin's leadership in UC and its commitment to having the most advanced cloud services available for customers.

2. Collaboration Service Provider of the Year

For the 4th consecutive year, Arkadin was selected as Collaboration Service Provider of the Year at Frost & Sullivan's prestigious Asia Pacific ICT (information and communication technology) Awards in 2016. Arkadin received the highest score in various performance indicators, including revenue and market share growth, leadership in product innovation, breadth of services, customer acquisitions, and business strategy [1, 2].

According to Jessie Yu, Research Manager, Digital Transformation, Asia-Pacific at Frost & Sullivan, Arkadin has strong and sustainable momentum in Asia-Pacific: "In a highly competitive and fast-changing business environment Arkadin has achieved significant customer and revenue growth. Their service quality, delivery and customization, coupled with flexible pricing policies and fast, efficient technology enhancements, are contributing to the value they deliver to their customers, spanning enterprises, SMBs and the public sector."

3. Enabling Singapore businesses to leverage the full power of the virtual office

After its initial debut in North America and Europe, Arkadin Total Connect, a Microsoft Skype for Business* hosted service, is now available to Singapore businesses that require a fully integrated UC ecosystem with voice-enabled Office 365*, audio/web/video conferencing, and a contact center. Professional project management, migration services, end-user training, support, and change-management capabilities ensure efficient onboarding and seamless long-term client care.

Arkadin is one of only a few service providers operating on a global scale capable of meeting Microsoft's requirements and functioning as a fully integrated platform for meetings and voice. Through Arkadin Total Connect, businesses can harness the full power of UC for greater workplace productivity in a convenient, all-in-one cloud-based offer.

* Microsoft, Skype, and Office 365 are registered trademarks of Microsoft Corporation in the United States and/or other countries.

Photo 1. Arkadin's Andrea Carboni (left) accepts a Silver Stevie trophy from Michael Gallagher, president of the Stevie Awards, at a reception in Rome on October 21st. Didier Jaubert, CEO of Arkadin, won the Stevie Award for Telecommunications Executive of the Year but was unable to attend the reception.

According to Bill Haskins of Wainhouse Research, the service positions Arkadin well for delivering a superior overall experience: "Going beyond the standard benefits customers should expect for hosted UC deployments, Arkadin adds code-level knowledge of the hosted Skype for Business platform, global infrastructure, robust networks, and multi-lingual local support. This combination of services, expertise, and infrastructure positions Arkadin to deliver a superior overall experience."

4. Acquisition enabling large enterprises to transform to the virtual workplace

Arkadin's September acquisition of Applicable, a leading provider of cloud UC and enterprise voice services, extends Arkadin's capabilities beyond its SMB stronghold to large global businesses. The combination of Applicable's extensive 360-degree expertise in successfully delivering integrated Microsoft UC services that match the requirements of very large customers, with Arkadin's telephony integration and local approach to service, creates an unbeatable offering, and it positions Arkadin solidly as a global force in deploying Cloud Microsoft UC services for businesses of all sizes, from SMBs to very large global organizations [3].

5. Winning prestigious industry awards for executive leadership in telecommunications and video services

Recent awards conferred on Arkadin underscore leadership in the virtual workplace. Arkadin CEO Didier Jaubert was awarded a Stevie Award in the 2016 International Business Awards for Telecommunications Executive of the Year. The Stevie Awards are considered one of the world's premier business awards, attracting hundreds of organizations worldwide. Didier earned a Silver trophy in the Telecommunications category for his bold vision and leadership in digital workplace transformation initiatives. The awards were presented at a gala reception in Rome on October 21st (**Photo 1**).

Arkadin achieved a Silver award in the Best in Biz 2016 International competition for video services in the Product Line of the Year category. Arkadin's extensive video services product line enables businesses of all sizes to have face-to-face interactions anywhere, anytime, and from any device.

6. Future prospects and strategies

Arkadin is very excited about trends in the virtual workplace. Working together with NTT Communications,

it has an attractive value proposition to enable customers to succeed in a digitally transformed work environment. Global businesses are offered the most sophisticated infrastructure, networking, telecommunications, and interoperability services anywhere in the industry.

Arkadin will continue to align with NTT Communications on product development and business strategies to ensure long-term leadership in cloud-based UC&C services, which are so essential to the virtual workplace culture. Additionally, Arkadin will continue to collaborate with its network of partners, including technology companies such as Microsoft, Adobe Systems, Cisco, Vidyo, and BlueJeans Network to ensure it has the most advanced UC&C services. Arkadin's comprehensive portfolio enables it to compete through a single provider strategy for meeting the needs of all businesses.

For further information on specific business strategies, please refer to the article in the September issue of NTT Technical Review "Arkadin: Global Unified Communications & Collaboration Champion in a Dynamic Digital Workplace" [4].

References

[1] Link to video by Frost & Sullivan, "Movers & Shakers with Serge Genetet, Executive Vice President - Managing Director Asia Pacific, Arkadin."
https://vimeo.com/171506579/5e81e13733

[2] Press release issued by Arkadin on June 17, 2016.
https://www.arkadin.com.sg/about-us/our-awards/ntt-communications-and-arkadin-honored-frost-sullivan-asia-pacific-ict-awards

[3] Press release issued by Arkadin on September 7, 2016.
https://www.arkadin.com/about-us/news-media/arkadin-acquires-applicable-enable-large-enterprises-leverage-power-unified

[4] D. Jaubert, "Arkadin: Global Unified Communications & Collaboration Champion in a Dynamic Digital Workplace," NTT Technical Review, Vol. 14, No. 9, 2016.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201609ga1.html

External Awards

Certificate of Appreciation
Winner: Yoshitaka Shimizu, Yasuo Suzuki, Satoshi Kotabe, Tetsuro Komukai, and Atsushi Yamamoto, NTT Network Innovation Laboratories; Hironori Kuroki and Yuki Takamiya, NTT Advanced Technology Corporation
Date: July 29, 2016
Organization: Municipality of San Remigio, Cebu, Philippines

For their great efforts in demonstrating the effectiveness of MDRU (movable and deployable information and communication technology resource unit) application.

Certificate of Appreciation
Winner: Nei Kato, Hiroki Nishiyama, Daiki Murayama, Naoki Miyashita, Naoto Yamada, Yuki Takahashi, and Hiroaki Takagi, Graduate School of Information Sciences, Tohoku University; Yoshitaka Shimizu and Yasuo Suzuki, NTT Network Innovation Laboratories
Date: August 2, 2016
Organization: Municipality of San Remigio, Cebu, Philippines

For their great efforts in demonstrating the effectiveness of Relay-by-Smartphone technology.

IEICE SRW Young Researcher's Award
Winner: Doohwan Lee, NTT Network Innovation Laboratories
Date: August 22, 2016
Organization: The Institute of Electronics, Information and Communication Engineers (IEICE) Communications Society, Technical Committee on Short Range Wireless Communications (SRW)

For "An Overhead-reducing Channel Estimation Sequences Design for Millimeter-wave Short-range MIMO Wireless Communication Systems."
Published as: D. Lee, K. Hiraga, K. Sakamoto, and T. Nakagawa, "An Overhead-reducing Channel Estimation Sequences Design for Millimeter-wave Short-range MIMO Wireless Communication Systems," IEICE Tech. Rep., Vol. 115, No. 474, SRW2015-75, pp. 29–34, 2016.

Outstanding Contributor
Winner: Tetsuro Inui, NTT Network Innovation Laboratories
Date: September 7, 2016
Organization: Open Networking Foundation (ONF)

Mr. Inui has spent substantial amounts of time working in different areas to support the efforts of the Carrier Grade SDN Working Group since its establishment. He is one of the editors of the carrier grade software-defined networking (SDN) use case document and has contributed valuable and influential use cases. He also works as a key reviewer of the carrier grade SDN group charter and other documents. His diligence and support are very valuable to the group.

Papers Published in Technical Journals and Conference Proceedings

Effects of Speaking Rhythm Naturalness on the Neural Basis of Speech Perception
S. Hiroya, K. Jasmin, S. Evans, S. Krishnan, M. Ostarek, D. Boebinger, and S. K. Scott
Proc. of Neuroscience 2015, Chicago, USA, October 2015.

Manipulation techniques of speech sound naturalness will give us a new possibility to investigate the neural correlates of speech perception. However, few studies have investigated the neural mechanisms of temporal information underlying speech perception, i.e., speaking rhythm. We first developed a novel method for decomposing speech signals into a speaking rhythm and phonetic information. Next we performed fMRI (functional magnetic resonance imaging) scans during passive listening that investigated the neural basis of speech perception. Result showed that left-lateralized premotor cortex and supplementary motor area (SMA) were more activated for Japanese rhythm than for English rhythm, and the areas overlapped that of vowel production. A series of our studies suggests that greater premotor cortex activation during speech perception would explain speech sound unnaturalness of both frequency and temporal information.

The Neural Basis of Perceiving Speech with a Non-native Rhythm
K. Jasmin, S. Hiroya, S. Evans, S. Krishnan, C. Lima, M. Ostarek, D. Boebinger, and S. K. Scott
Proc. of Cognitive Neuroscience Society 2016 Annual Meeting, New York, USA, April 2016.

Rhythm is a natural part of speech. We developed a novel method for decomposing speech signals in order to separate phonetic information from rhythmic structure. Audio recordings of English sentences spoken by a Japanese native speaker were manipulated such

that their rhythm was stress timed (like English), mora timed (like Japanese) or had phonemes with equal durations. Twenty-one healthy right-handed participants underwent behavioural testing and functional magnetic resonance imaging (fMRI) scans. The results confirmed subjects judged English sentences as being most natural. FMRI was used to image the brains of participants while they listened to the sentences. The result showed that the supplementary motor area (SMA), a region involved in speech production, was sensitive to rhythm naturalness. This suggests that integrating non-native speech rhythm with native language speech may rely on increased auditory-motor processing.

FRT-Skip Graph: A Skip Graph-style Structured Overlay Based on Flexible Routing Tables

M. Hojo, R. Banno, and K. Shudo

Proc. of ISCC (the 2016 IEEE Symposium on Computers and Communication), pp. 657–662, Messina, Italy, June 2016.

Structured overlays enable a number of nodes to construct a logical network autonomously and search each other. Skip Graph, one of the structured overlays, constructs an overlay network based on Skip List structure and supports range queries for keys. Skip Graph manages routing tables based on random digits; therefore, the deviation of them disturbs effective utilization of the routing table entries and increases path length more than the ideal value. We therefore propose FRT-Skip Graph, a novel structured overlay that solves the issues of Skip Graph and provides desirable features not in Skip Graph. FRT-Skip Graph is designed based on Flexible Routing Tables (FRT) and supports range queries similarly to Skip Graph. Furthermore, it provides features derived from FRT, namely, dynamic routing table size and high extensibility.

Proposal of a Simple Ultra-low Contention CD ROADM

A. Iwaki, A. Sahara, and M. Fukutoku

IEICE Transactions on Communications, Vol. E99-B, No. 8, pp. 1772–1779, August 2016.

We propose a simple configuration for colorless and directionless (CD) reconfigurable optical add/drop multiplexers that enables ultra-low contention add/drop operation to be achieved. In the configuration, we apply a combination of multiple small-port-count CD add/drop banks (CD banks) and round-robin CD bank assignment. Evaluation results show that the proposed configuration can substantially reduce intra-node contention rate, which is less than 0.1%. We also find that the proposed configuration can improve the utilization efficiency of wavelength resources and transponders. We discuss the mechanism of how the proposed configuration reduces intra-node contention by analyzing the status of wavelength assignments in direction ports and CD banks.

Speech Rhythm Measure of Non-native Speech Using a Statistical Phonemic Duration Model

S. Hiroya, K. Jasmin, S. Evans, S. Krishnan, C. Lima, M. Ostarek, D. Boebinger, and S. K. Scott

Proc. of the 8th Annual Meeting of the Society for the Neurobiology of Language, London, UK, August 2016.

Rhythm is a natural part of speech. English speech sounds spoken by a Japanese native speaker were manipulated such that their rhythm was stress-timed and mora-timed. FMRI (functional magnetic resonance imaging) experiments showed that the left-lateralized supplementary motor area (SMA), a region involved in speech production, was more activated for mora-timed rhythm (non-native rhythm) than stress-timed rhythm. However, a difference between mora-timed and stress-timed rhythm in English should be quantified for further analysis. In this study, we developed a statistical model of phonemic duration in English to be independent of a type of interval. An expectation-maximization algorithm created a state-transition model of the phonemic duration. Results showed that the variability among states of self-transition probability for the native Japanese speaker was significantly larger than for the native English speaker. This suggests that these structures of phonemic duration affected activity in the speech perception network.

A Study of Non-synthetic Region for Texture Synthesis Based Image Coding

T. Sasaki, R. Tanida, and A. Shimizu

IEICE Transactions on Information and Systems (Japanese Edition), Vol. J99-D, No. 9, pp. 865–867, September 2016.

In the coding method using cartoon-texture image decomposition and texture synthesis, a region not suitable for texture synthesis is also included in the texture components. It is possible to improve the subjective quality if this region is excluded from texture components. In this paper, we propose a method to determine the region.

Generation of a Frequency Comb Spanning More Than 3.6 Octaves from Ultraviolet to Mid Infrared

K. Iwakuni, S. Okubo, O. Tadanaga, H. Inaba, A. Onae, F.-L. Hong, and H. Sasada

Optics Letters, Vol. 41, No. 17, pp. 3980–3983, September 2016.

We have observed an ultra-broadband frequency comb with a wavelength range of at least 0.35 to 4.4 μm in a ridge-waveguide-type periodically poled lithium niobate (PPLN) device. The PPLN waveguide is pumped by a 1.0–2.4 μm wide frequency comb with an average power of 120 mW generated using an erbium-based mode-locked fiber laser and a following highly nonlinear fiber. The coherence of the extended comb is confirmed in both the visible (around 633 nm) and the mid-infrared regions.

Adaptive and Efficient Multilayer Elastic Optical Network Planning

T. Tanaka, T. Inui, A. Kadohata, A. Hirano, and W. Imajuku

Proc. of the 42nd European Conference on Optical Communication (ECOC 2016), Th.3.D.1, Düsseldorf, Germany, September 2016.

We overview our current works on the heuristic IP-over-elastic optical network (EON) planning algorithms including modulation-aware virtual topology planning in the planning phase and multiperiod multilayer network planning which adapts to dynamic traffic conditions in the operational phase.

32-core Inline Multicore Fiber Amplifier for Dense Space Division Multiplexed Transmission Systems

S. Jain, T. Mizuno, Y. Jung, Q. Kang, J. R. Hayes, M. N. Petrovich, G. Bai, H. Ono, K. Shibahara, A. Sano, A. Isoda, Y. Miyamoto, Y. Sasaki, Y. Amma, K. Takenaga, K. Aikawa, C. Castro, K. Pulverer, Md Nooruzzaman, T. Morioka, S. U. Alam, and D. J. Richardson

Proc. of ECOC 2016, Th.3.A.1, Düsseldorf, Germany, September

2016.

We present a high-core-count SDM amplifier, i.e., 32-core multi-core-fiber amplifier, in a cladding-pumped configuration. An average gain of 17 dB and NF of 7 dB is obtained for –5 dBm input signal power in the wavelength range 1544–1564 nm.

On Compositional Reasoning about Anonymity and Privacy in Epistemic Logic

Y. Tsukada, H. Sakurada, K. Mano, and Y. Manabe

Annals of Mathematics and Artificial Intelligence, Vol. 78, No. 2, pp. 101–129, October 2016.

In this paper, we exploit epistemic logic (or the modal logic of knowledge) for multiagent systems to discuss the compositionality of several privacy-related information-hiding/disclosure properties. The properties considered here are anonymity, privacy, onymity, and identity. Our initial observation reveals that anonymity/privacy properties are not necessarily sequentially compositional. This means that even though a system comprising several sequential phases satisfies a certain unlinkability property in each phase, the entire system does not always enjoy a desired unlinkability property. We show that the compositionality can be guaranteed provided that the phases of the system satisfy what we call independence assumptions. More specifically, we develop a series of theoretical case studies of what assumptions are sufficient to guarantee the sequential compositionality of various degrees of anonymity, privacy, onymity, and/or identity properties. Similar results for parallel composition are also discussed. Further, we use the probabilistic extension of epistemic logic to consider the compositionality of probabilistic anonymity/privacy. We show that the compositionality can also be guaranteed in the probabilistic setting, provided that the phases of the system satisfy a probabilistic independence assumption.
