

## Communication Science that Adapts to the Changing Present and Creates a Sustainable Future—Aiming to Create Technology that Brings Harmony and Symbiosis among People, Society, and the Environment

*Futoshi Naya*

### Abstract

NTT Communication Science Laboratories (CS Labs) celebrated its 30th anniversary in 2021. From the beginning, we have been engaged in research to construct basic theories that address the essence of human and information sciences and develop innovative technologies that can bring about changes in society. We aim to achieve communication that “reaches the heart” between humans and between humans and computers. In this article, I present the recent research results at CS Labs and efforts toward future advances.

*Keywords: communication science, artificial intelligence, brain science*

### 1. Introduction

NTT Communication Science Laboratories (CS Labs) was established on July 4, 1991, at Keihanna Science City, Kyoto. CS Labs, which was initially housed within the Advanced Telecommunications Research Institute International (ATR) and started from two research groups on machine learning and information theory, currently has about 150 researchers at Keihanna and Atsugi, Kanagawa Prefecture. Along with last year’s open house, we set up a website to commemorate the 30th anniversary of our founding and summarize our main research results [1].

Research themes at CS Labs, which began with the mission of understanding human-to-human communication, have been developed in the two areas. The first is *approaching and surpassing human abilities*

including media-information processing enabling computers to have the same abilities as humans, i.e., seeing, listening, and speaking, as well as quantum information theory and machine learning. The second area is *deeply understanding humans* including human information science for revealing various human perception and motor skills and brain science focusing on cognitive flexibility and diversity for elucidating excellent cognitive ability such as those of top athletes (**Fig. 1**).

Although these research areas have developed and changed to reflect the technological progress and social needs of each era over 30 years, we have continued to develop technologies that are close to people and society while adhering to the philosophy of *understanding the essence of communication*. I present our recent research results in the following areas.

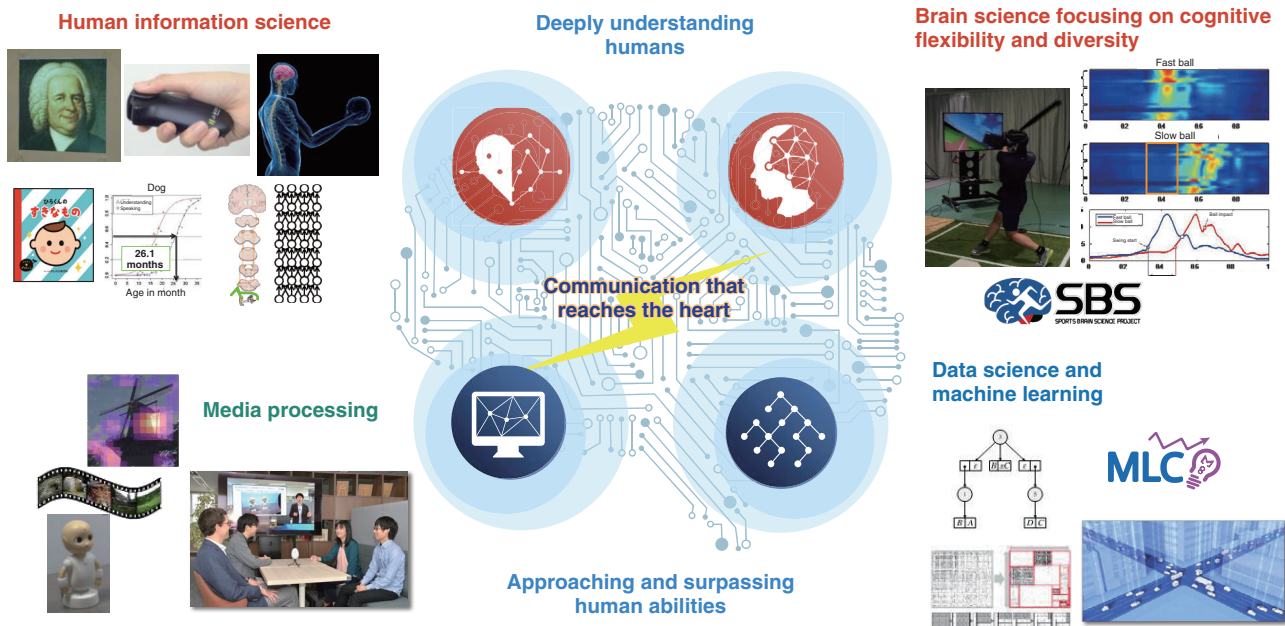


Fig. 1. Research areas of CS Labs.

## 2. Technologies that approach and surpass human abilities

With the advent of deep learning, artificial intelligence (AI) technology related to media processing, such as seeing, listening, and speaking, continues to make remarkable progress, and has already exceeded human discrimination performance in, for example, image-recognition competitions. However, to obtain high performance in deep learning, a huge amount of training data consisting of pairs of input data and their correct labels is generally required.

Even if we humans can obtain only a small amount of information, we can make flexible and sophisticated inferences from experience and supplementing missing information with other information. For example, although a photograph captures three-dimensional (3D) spatial information as a 2D image, humans can infer to some extent 3D information such as the shape and depth of the subject just by looking at the photograph based on experience.

Conventionally, to make computers learn such abilities, a large amount of data including 3D information measured with a special device, such as a depth sensor or stereo camera, is needed in addition to a 2D image taken as a photograph, but this is very costly. To overcome this problem of obtaining data, we proposed the deep learning technology aperture

rendering generative adversarial network [2] that takes into account the relationship between aperture and blur as an optical constraint of the camera. This technology can learn 3D information such as depth and bokeh effects from only general photographs (2D information) such as public images on the Internet, and generate new images with different blur and bokeh effects with different depths of field.

Humans also have the excellent ability of selective listening, i.e., focusing on the specific voice of a person in a multi-person talking situation, such as a party, or a noisy or reverberant environment. At CS Labs, we proposed SpeakerBeam, a technology that distinguishes the voice of the target person using only voice. In 2021, we proposed Multimodal SpeakerBeam [3], which enables selective listening from multiple clues like what humans can do by combining video information and audio. This technology achieves highly accurate speaker-voice separation when there are multiple speakers with similar voice qualities. The movement of the lips is used as the main clue, and conversely, when the image of the lips cannot be obtained, the voice is used. This research is developing into a universal sound extraction technology that can distinguish not only voice but also notable sounds such as sirens of fire engines and the barking of dogs. The Feature Articles in this issue introduce the latest speech enhancement technology

for extracting high-quality voice as if it was recorded with a microphone near the speaker from the voice recorded with a microphone located at a remote location [4].

Since the 1990s, CS Labs has been researching dialogue systems that can naturally communicate with people. We initially focused on researching dialogue systems for specific purposes such as reservations and searches, but we are now researching chat-dialogue systems that can respond naturally while handling a wide range of topics regardless of purpose [5]. In 2021, with the cooperation of Seika Town Hall in Kyoto Prefecture, where the Keihanna CS Lab is located, we started an AI demonstration experiment [6] in which visitors can enjoy chat dialogue while performing operations such as government office guidance and providing tourist information. We have also released the largest *Transformer* dialogue model in Japanese, which combines ultra-large-scale dialogue data collected on the web and deep learning techniques [7]. We have been building a more cohesive dialogue system by discussing image scenes seen from the window of a vehicle where the situation of the place changes from moment to moment along with the user. We are also engaged in research for constructing a system that can memorize the preferences of the user, maintain the consistency of the dialogue content, and continue dialogue closer to humans.

There has also been progress in the field of machine learning in terms of being close to people. For example, when making decisions for people such as loan approval and recruitment, conventional machine learning techniques that simply prioritize prediction accuracy can result in unfair predictions when using a variety of subtle human characteristics, such as gender, race, and disability. We proposed a machine learning method that is based on causal relationships [8] for achieving fair accurate prediction for each individual by modeling prior knowledge about unfairness as a causal relationship (causal graph) between features and prediction results.

This issue introduces, as a cutting-edge achievement of machine learning, CS Labs' distributed deep learning technology for achieving signal-free mobility using digital twins with an eye on a future autonomous driving society [9]. This optimization technology improves the average moving speed of all vehicles while preventing them from colliding through communicating and coordinating with one another in a road network without traffic lights.

### 3. Research to deeply understand humans

---

At CS Labs, we have been studying various illusions caused by the latent functions of the brain in exploring the mechanisms of human perception and movement. We have also published a website called *Illusion Forum* [10] where one can experience various visual and auditory illusions. We obtained results that clarify the various latent functions of the brain from conducting an illusion-presentation experiment using virtual reality, indicating, for example, that a batter feels as if the ball is rising when the pitcher's pitching motion becomes faster even if the ball trajectories and ball speeds are exactly the same.

In October 2021, we started joint research with Shizuoka General Hospital on voice and language recognition for cochlear implant wearers [11]. It is known that even deaf children can acquire the same level of spoken language as children with normal hearing by wearing cochlear implants at an early stage, but the mechanism of speech perception and language development in the brain remains unclear. By approaching this issue from both medicine and brain science, we are attempting to elucidate aspects of the hearing mechanism of the elderly and other hearing-impaired people as well as the mechanisms behind individual differences in speech perception and language development.

From research on language acquisition, CS Labs published a word-familiarity database that evaluated the degree of familiarity of about 80,000 Japanese words on a 7-point scale in 1999. In 2021, we built the Reiwa version [12] for re-examining the original and includes more than 160,000 additional words. The database is commercially provided by NTT Printing Corporation. We also developed a method for estimating a person's approximate size of vocabulary by answering whether the person knows 50 selected words using this word-familiarity database. The website is open to the public [13].

### 4. Toward the creation of technologies that bring harmony and symbiosis among people, society, and the environment

---

Research activities at CS Labs have been carried out by pursuing each area of expertise, but the amount of research that crosses two or more such areas has been increasing. For example, media-processing research that specializes in a single modality such as seeing or listening has evolved into *cross-modal* information processing that handles multiple media

at the same time, such as speaker identification using sound and video, as described above. In human science as well, research is shifting to elucidating the mechanism of multisensory integration. In 2021, the *pitcher simulator*, which is the result of research that brought together all the knowledge and techniques of brain science, human science, and media processing at CS Labs, contributed to winning the gold medal of the Japan Women's Softball National Team at the major international sporting event held in Tokyo and was featured as a "secret weapon" [14] in newspapers and television. As another example, this issue presents an approach to personal cardiac modeling that combines mobile sensing, media processing, and machine learning [15].

Recent advances in science and technology and their commoditization are accelerating rapidly. Furthermore, the COVID-19 pandemic, global climate change, increasing natural disasters, and international conflict have dramatically changed our daily lives and values. There is a limit to solving increasingly complex and diversified social issues simply by pursuing research within an area of expertise, and research that creates new value through synergies with research results in peripheral areas and research in different areas is becoming increasingly important to open up new interdisciplinary research areas.

Basic research also requires a research approach from a long-term perspective that is fundamentally different from the past. Therefore, the Institute for Fundamental Mathematics was established in October 2021 to study the fundamental theory of modern mathematics at CS Labs [16]. For more information on our research at the Institute for Fundamental Mathematics, please see the article in this issue [17]. Toward the implementation of IOWN (the Innovative Optical and Wireless Network), while making full use of the latest mathematical methods, we will tackle and solve problems that occur in various areas and continue to work on communication science research to develop technologies that bring harmony and symbiosis among people having diverse values, society, and the environment.

## References

- [1] Website of 30th anniversary of NTT Communication Science Laboratories (in Japanese), <https://www.kecl.ntt.co.jp/30th/>
- [2] T. Kaneko, "Unsupervised Depth and Bokeh Learning from Natural Images Using Aperture Rendering Generative Adversarial Networks," NTT Technical Review, Vol. 20, No. 7, pp. 89–95, July 2022. <https://doi.org/10.53829/ntr202207ra1>
- [3] M. Delcroix, T. Ochiai, H. Sato, Y. Ohishi, K. Kinoshita, T. Nakatani, and S. Araki, "Developing AI that Pays Attention to Who You Want to Listen to: Deep-learning-based Selective Hearing with Speaker-Beam," NTT Technical Review, Vol. 19, No. 9, pp. 39–45, 2021. <https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202109fa4.html>
- [4] T. Nakatani, R. Ikeshita, N. Kamo, K. Kinoshita, S. Araki, and H. Sawada, "AI Hears Your Voice as if It Were Right Next to You—Audio Processing Framework for Separating Distant Sounds with Close-microphone Quality," NTT Technical Review, Vol. 20, No.10, pp. 49–55, Oct. 2022. <https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202210fa7.html>
- [5] H. Sugiyama, M. Mizukami, T. Arimoto, H. Narimatsu, Y. Chiba, and H. Nakajima, "The Day a System Becomes a Conversation Partner—Exploring New Horizons in Social Dialogue Systems with Large-scale Deep Learning," NTT Technical Review, Vol. 19, No. 9, pp. 28–33, 2021. <https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202109fa2.html>
- [6] NTT press release issued on Nov. 12, 2021 (in Japanese). [https://group.ntt.jp/topics/2021/11/12/ai\\_seika.html](https://group.ntt.jp/topics/2021/11/12/ai_seika.html)
- [7] NTT press release issued on Sept. 30, 2021 (in Japanese). <https://group.ntt.jp/topics/2021/09/30/transformer.html>
- [8] Y. Chikahara, S. Sakaue, A. Fujino, and H. Kashima, "Learning Individually Fair Classifier with Path-specific Causal-effect Constraint," Proc. of the 24th International Conference on Artificial Intelligence and Statistics (AISTATS), Virtual, Aug. 2021.
- [9] K. Niwa, "Smart Traffic Coordination via Learnable Digital Twins—Future Possibilities of Distributed Deep Learning," NTT Technical Review, Vol. 20, No. 10, pp. 61–66, Oct. 2022. <https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202210fa9.html>
- [10] Website of Illusion Forum by CS Labs (in Japanese), <https://illusion-forum.ilab.ntt.co.jp/>
- [11] NTT press release issued on Oct. 19, 2021 (in Japanese). <https://group.ntt.jp/newsrelease/2021/10/19/211019a.html>
- [12] NTT press release issued on June 3, 2020 (in Japanese). <https://group.ntt.jp/newsrelease/2020/06/03/200603a.html>
- [13] Website of "Reiwa-version of Vocabulary Test" (in Japanese), [https://www.kecl.ntt.co.jp/icl/lirg/resources/goitokusei/vocabulary\\_test/php/login.php](https://www.kecl.ntt.co.jp/icl/lirg/resources/goitokusei/vocabulary_test/php/login.php)
- [14] M. Yamaguchi, D. Nasu, D. Mikami, T. Kimura, T. Fukuda, and M. Kashino, "Women's Softball × Sports Brain Science," NTT Technical Review, Vol. 20, No. 2, pp. 32–38, Feb. 2022. <https://doi.org/10.53829/ntr202202fa3>
- [15] K. Kashino, R. Shibue, and S. Tsukada, "Personal Heart Modeling Using Mobile Sensing," NTT Technical Review, Vol. 20, No. 10, pp. 56–60, Oct. 2022. <https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202210fa8.html>
- [16] NTT press release, "NTT has established the Institute for Fundamental Mathematics—Advancing the pace of exploration into the unexplored principles of quantum computing," Oct. 1, 2021. <https://group.ntt/en/newsrelease/2021/10/01/211001a.html>
- [17] M. Wakayama, "Number Theory and Quantum Physics Based on Symmetry—Themes from Quantum Optics," NTT Technical Review, Vol. 20, No. 10, pp. 67–76, Oct. 2022. <https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202210fa10.html>



**Futoshi Naya**

Vice President and Head of NTT Communication Science Laboratories.

He received a B.E. in electrical engineering, M.S. in computer science, and Ph.D. in engineering from Keio University, Kanagawa, in 1992, 1994, and 2010. He joined NTT Communication Science Laboratories in 1994. From 2003 to 2009, he was with Intelligent Robotics and Communication Laboratories, Advanced Telecommunications Research Institute International (ATR). His research interests include communication robots, sensor networks, pattern recognition, data mining in cyber-physical systems, and AI-based tailor-made education support. He is a member of the Institute of Electrical and Electronics Engineers (IEEE), the Robotics Society of Japan, the Society of Instrument and Control Engineers, and the Institute of Electronics, Information and Communication Engineers (IEICE).

---