I Want to Learn More about You: Getting Closer to Humans with AI and Brain Science

Takeshi Yamada

Abstract

NTT Communication Science Laboratories aims to achieve communication that *reaches the heart* by pursuing innovative technologies that approach and exceed human abilities such as media processing, data analysis, and machine learning as well as studying cognitive neuroscience and brain science for obtaining a deeper understanding of people. It also aims to deliver concrete results to society through collaboration with its business partners. This article introduces the efforts to achieve these aims.

Keywords: artificial intelligence, communication science, brain science

1. Introduction

In 1985, the Nippon Telegraph and Telephone Public Corporation was privatized, which led to the founding of NTT. Before privatization, each home only had a traditional black telephone, which was rented from the corporation and mainly located in the entrance hall. After privatization, however, one could freely buy a telephone of different colors and with more functions such as cordless reception that enabled having a receiver in each room. The popular love song titled "I want to learn more about you"* focused on an intimate conversation between a girl and her (possibly) boyfriend over the phone with her asking "where are you now and what are you doing" and achieved estimated sales of 391,000 copies. Today, various forms of social media have almost entirely subsumed this role and even been extended to learning about people who are not close friends and basically strangers. A smartphone obtains a great deal of information about its user throughout the day. It is so smart that it may obtain more details about its user than what the user knows about him/herself. On the other hand, the traditional black telephone, while simple and not smart, might be more comfortable to use than current smartphones. As technology continues to develop, how will communication change? It is all the more important to identify the *core of communication* in this era when the physical distance between people must be maintained to prevent the spread of the novel coronavirus while avoiding social disconnection from our close friends and community.

Next year, NTT Communication Science Laboratories (NTT CS Labs), which was founded in Keihanna Science City, Kyoto, in 1991, will mark its 30-year anniversary. Since its founding, it has undertaken fundamental research based not only on the principle of conveying information accurately and efficiently but also on deepening mutual understanding, sharing feelings, and making genuine contact. Even though the research in the beginning focused on person-to-person communication, its aim today is sincere and *heartfelt* communication in both person-to-person and person-to-computer contexts. To this end, we at NTT CS Labs are pursuing innovative technologies for approaching and exceeding human abilities such as media processing, data analysis, and machine learning as well as studying cognitive neuroscience and brain science for obtaining a deeper understanding of people and their cognitive flexibility and diversity

* "I want to learn more about you" was NTT's "TALK ON THE PHONE" campaign song that was released immediately after NTT's privatization in 1985.

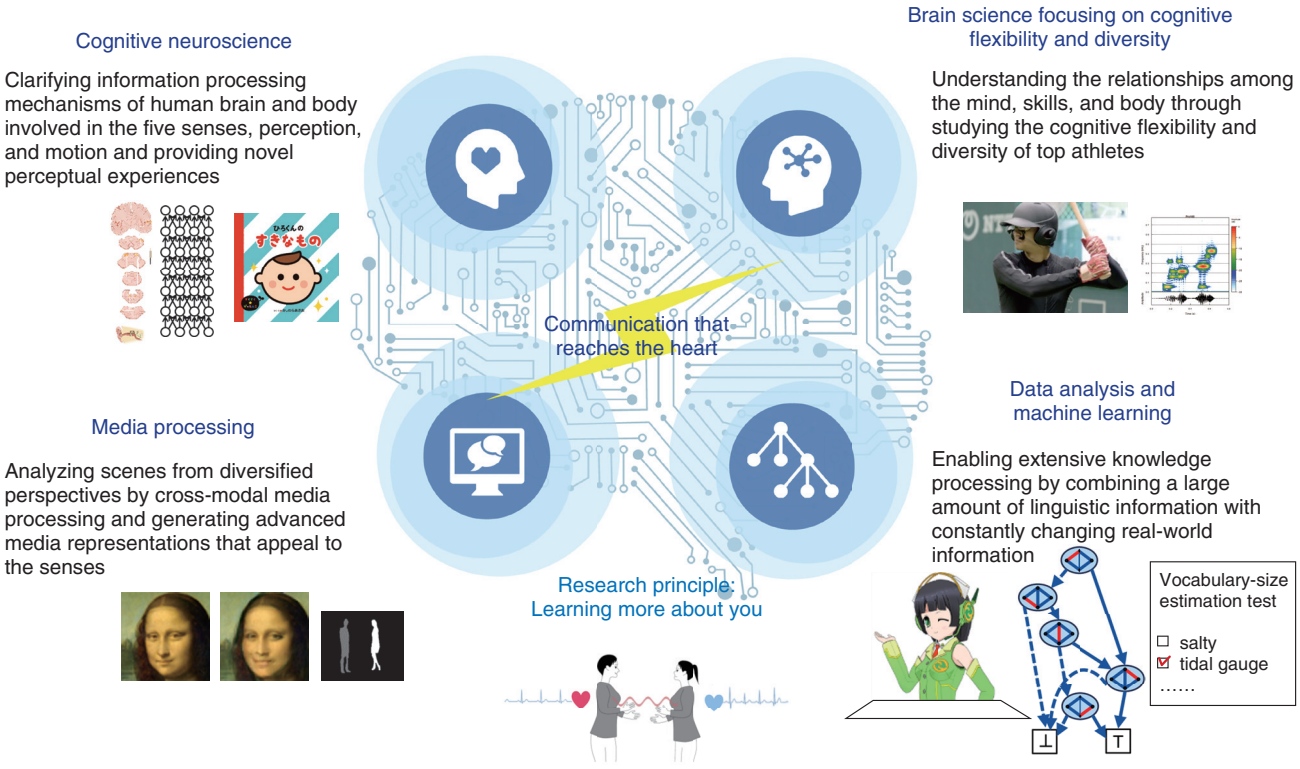

Fig. 1. Research areas of NTT CS Labs.

(**Fig. 1**) [1]. We are also continuing to work on the development of more fundamental mathematical theories [2]. Our primary mission is to carry out these basic research topics with unlimited curiosity to learn more about us humans. We are also delivering concrete results to society through collaboration with our business partners. Several examples of these research endeavors are introduced below.

2. Technologies that approach and exceed human abilities

Communication is first and foremost the recognition and understanding of spoken words. On our Illusion Forum website, which provides information on a variety of visual and auditory illusions, noise-vocoded speech (so-called mosaic speech) is introduced [3]. Mosaic speech, which is analogous to visual mosaic images, is a distorted speech sound where the detailed spectral information in the original clean speech sound is manipulated and destroyed. When one listens to it, he or she can somehow understand what is being said, even though the sound is unnatural and hard to hear. In this way, humans can hear the content of speech sounds even if their fine temporal structure is substantially degraded. One may fail to understand the sound at first, but if he or she listens multiple times, or listens to the original sound then listens to the converted one again, he or she will understand it. One has just experienced the depth of the human auditory system and sound recognition ability.

Mammals, including humans, have developed brain functions that recognize such natural sounds as speech and environmental from evolution. In the mammalian brain, the auditory system converts the physical properties of a sound stimulus to neural activities and processes them through a cascade of brain regions. Interestingly, the characteristics of amplitude modulation (AM) tuning in temporal and rate coding are transformed systematically along the processing stages from the periphery to the cortex. The AM rate at which neurons are synchronized gradually decreases and the number of neurons that perform rate coding gradually increases. We have discovered that when a deep neural network (DNN) is trained to classify raw sound data using raw waveforms that are directly input, the trained DNN resembles the

brain's auditory system throughout the entire cascade [4]. Such similarity to the auditory system increases as the classification accuracy improves. These results suggest that AM tuning in auditory systems emerged through evolution as a result of adapting to sound recognition in the real world.

Through evolution, humans have acquired the ability to select the specific voice of a person they want to hear and understand what he or she is saying in a meeting or at a party when several people are talking or when music is playing in the background. This skill is called *selective listening*. We have applied a proprietary DNN and developed technology called SpeakerBeam that enables computers to perform selective listening [5]. However, extracting a target voice is difficult if speakers with similar voices are included in the data. Therefore, in addition to voice, lip movements are used as features to distinguish among speakers even with similar voices. We are also developing DNN-based voice-conversion technology that makes it possible to freely change such voice features as voice quality and intonation while preserving the speech's linguistic content. Further development of these technologies will enable natural communication that overcomes disabilities or age-related decline in speaking or hearing functions and support conversations in foreign languages [6].

We are also investigating the language-acquisition mechanism in children, who basically acquire language by talking with their parents. Language and language-based oral communication are basic human functions that have evolved over the past 150,000 years. Since written language emerged relatively recently, only 50,000 years ago, the ability to read is not an inherent function of the human brain, which did not evolve for reading. Reading is achieved through a flexible combination of basic brain functions such as vision, audition, language, and cognition, which is called *neural recycling* [7].

To understand the language-acquisition mechanism, we constructed the Child Vocabulary Development Database by conducting a large-scale survey of what words children can understand and say at what particular development times and modeling the results [8]. With this database, we created personalized educational picture books in collaboration with NTT Printing Corporation for encouraging children to read. The books' contents were customized to the vocabulary development of each child. In cooperation with Onna Village in Okinawa and Tokushima City, we worked with NTT Printing Corporation to deliver these books to children and to encourage them to read from an early age [9, 10].

As an approach to understanding the advanced human abilities of both language and knowledge processing, we are participating in the artificial intelligence (AI) project "Todai Robot Project—Can a robot get into the University of Tokyo?" led by the National Institute of Informatics (NII). The project is researching the extent to which AI can solve the same problems that humans can solve. A team made up of members of the NTT CS Labs. and other project members took up the challenge of developing an AI system that can take and pass the English written exam administered by the National Center for University Entrance Examinations. The AI system achieved an extremely high score of 185 out of 200 points (64.1 T-score) in the exam's 2019 version [11]. English exams contain problems that require integrating both natural language processing and knowledge processing to solve. We exploit the knowledge gained in tackling these problems for our conversational AI research, which involves chatting with people (in chat-oriented dialogue systems) and providing information and guidance (in task-oriented dialogue systems) to achieve more natural and mutually understandable conversations between AI and humans.

Based on research on conversational AI, NTT is developing a role-play-based conversational AI called Narikiri AI, which reproduces the behaviors of a celebrity or a character in a novel or a game. Narikiri AI's dialogue data are collected through sets of questions and answers posted by online users, where one user asks a certain character a question, and another user mimics its personality and answers it. A few such Narikiri AIs have been constructed by NTT in cooperation with NTT DOCOMO and DWANGO Co., Ltd. Recently, we have begun a new Narikiri AI project in cooperation with Seika Town, Kyoto, involving its official public-relations character (mascot) Kyomachi Seika; thus, the project is called Narikiri AI Kyomachi Seika [12]. The concept of Kyomachi Seika is that of "a young agent is dispatched from the future to protect the world from a time paradox" [13]. It has attracted many fans through social networking services and is a perfect match with Narikiri AI.

We are conducting basic research on technologies to achieve natural human-like conversations by smoothly switching between chat-oriented and task-oriented dialogues. To demonstrate this technology, we collected dialogue data, including a large amount of knowledge and experience about Seika Town, through cooperation with its residents. We will build

a dialogue system that provides appropriate information and guidance regarding tourism and town administration, understand user's intensions, and flexibly answers user questions and requests through Kyomachi Seika.

3. Studying cognitive neuroscience and brain science for a deeper understanding of people

AI development is also intensifying the importance of obtaining a deeper understanding of people. For example, when searching on the Internet, advertisements suddenly appear that match the search words. Sometimes users click on such links and impulsively make purchases. Many users might argue that they purchased the product completely voluntarily, downplaying any third-party manipulation. As AI technology expands, the risk increases for such manipulation, which resembles an AI version of subliminal effects.

To avoid such risks, it is important to obtain a deeper understanding of the preconceptions held by people and how and when they will behave in a certain way. We are conducting research to clarify the information-processing mechanisms of the human brain and body involving the five senses, perception, and motion and providing novel perceptual experiences [14, 15]. We are also carrying out brain science research to understand the relationships among the mind, skills, and body of top athletes through studying their cognitive flexibility and diversity such as how they obtain and judge information from the outside world. Regarding baseball, for example, we are exploring the differences between great and mediocre hitters by investigating how well good hitters actually see the ball or whether a fastball really travels in a straight path. The plan is to use the knowledge gained from our research to provide feedback to athletes to sharpen their brain functions.

As societies worldwide are forced to coexist with the novel coronavirus, establishing profound contact has become more complicated. The economist and philosopher Adam Smith wrote that "we often derive sorrow from the sorrow of others" in his discussion of the importance of sympathy [16]. Sharing sentiment and emotions with others and feeling their experiences as one's own is the essence of sympathy and leads to genuine interaction. In collaboration with Tatsuya Kameda Laboratory at the University of Tokyo, we are studying how pain is shared during face-to-face interactions [17]. In one of the experiments carried out through this collaboration, a pair of strangers (participants) were exposed to identical pain-provoking (thermal) stimuli. The blood-volume pulse of both participants is recorded to measure their acute sympathetic responses while they both simultaneously experienced the stimuli. Under a face-to-face condition, participants with weaker reactions elevated their physiological reactivity to the stimulus based on their partner's reactions but not under a shielded, non-face-to-face condition. These results suggest that during face-to-face interactions, sharing such negative emotions as pain occurs at the unconscious physiological level. Another experiment of a two-player eSports competition showed that the heart rates of both players were synchronized in high-level close battles but not in intermediate one-sided battles.

4. Future prospects

The Japanese word for "happiness" is "shi-awase," where "shi" means "action" and "awase" means "match." This combination of meanings suggests that one way to achieve individual happiness is through successful interaction or communication with another person [18]. In other words, the synergistic effect of sincere and *heartfelt* contact improves human happiness, or in more modern terms, the well-being of people. In Japanese, "shi-ai" denotes games such as baseball or tennis, and the heart-rate synchronization in the eSports experiment seems to exhibit such a "shi-awase" or "action-matching" state. The mission of NTT CS Labs is to pursue sincere and *heartfelt* communication that literally achieves the following lyric from "I want to learn more about you"*: "Even if I'm far away, my heart is with you."

References

[1] T. Yamada, "Processing Like People, Understanding People, Helping People—Toward a Future Where Humans and AI Will Coexist and Co-create," NTT Technical Review, Vol. 17, No. 11, pp. 6–11, Nov. 2019.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201911fa1.html

[2] G. Kato, "Quantum Information Processing via Indirect Quantum Control," NTT Technical Review, Vol. 18, No. 11, pp. 32–35, Nov. 2020.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202011fa5.html

[3] Website of Illusion Forum by NTT CS Labs, Noise Vocoded Speech (in Japanese), http://www.kecl.ntt.co.jp/IllusionForum/a/noise_vocodedSpeech/ja/index.html

[4] Press release issued by NTT, "A Deep Neural Network Trained for Sound Recognition Acquires Sound Representation Similar to that in Mammalian Brains," July 10, 2019.
https://www.ntt.co.jp/news2019/1907e/190710a.html

[5] M. Delcroix, K. Zmolikova, K. Kinoshita, S. Araki, A. Ogawa, and T. Nakatani, "SpeakerBeam: A New Deep Learning Technology for

Extracting Speech of a Target Speaker Based on the Speaker's Voice Characteristics," NTT Technical Review, Vol. 16, No. 11, pp. 19–24, Nov. 2018.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201811fa2.html
[6] K. Tanaka, T. Kaneko, N. Hojo, and H. Kameoka, "Communication with Desired Voice," NTT Technical Review, Vol. 18, No. 11, pp. 27–31, Nov. 2020.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202011fa4.html
[7] S. Dehaene, "Reading in the Brain: The New Science of How We Read," Penguin Putnam Inc, 2010.
[8] S. Fujita, "Measuring Textual Difficulty—Estimating Text Readability and a Person's Vocabulary Size," NTT Technical Review, Vol. 18, No. 11, pp. 36–42, Nov. 2020.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202011fa6.html
[9] Press release issued by NTT Printing, "A Three-party Experiment Using a Personalized Educational Picture Book Has Started!", Feb. 6, 2020 (in Japanese).
https://www.nttprint.com/company/itemid419-000048.html
[10] Press release issued by NTT Printing, "Conclusion of the Joint Experiment Agreement on Supporting Picture Book Reading Activities at Home Using Personalized Educational Picture Books with Tokushima City and NTT," July 2, 2020 (in Japanese).
https://www.nttprint.com/company/itemid419-000053.html
[11] Press release issued by NTT, "AI Achieved a Score of 185 on the English Written Exam of the National Center Test for University Admissions in 2019," Nov. 18, 2019.
https://www.ntt.co.jp/news2019/1911e/191118a.html
[12] Press release issued by NTT, "Starting a Joint Experiment with Seika Town on a Conversational AI System," July 3, 2020 (in Japanese).
https://www.ntt.co.jp/news2020/2007/200703a.html
[13] What is Kyomachi Seika (in Japanese). https://kyomachi-seika.jp/profile/
[14] S. Kuroki, "Towards Understanding Human Skin Sensations," NTT Technical Review, Vol. 18, No. 11, pp. 16–20, Nov. 2020.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202011fa2.html
[15] S. Itoh, "Brain-information Processing for Quick and Stable Human Movements—Stretch-reflex Regulation Based on Visually Updated Body Representation," NTT Technical Review, Vol. 18, No. 11, pp. 21–26, Nov. 2020.
https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202011fa3.html
[16] A. Smith, "Theory of Moral Sentiments," 2 ed., Strand & Edinburgh: A. Millar; A. Kincaid & J. Bell, 1761.
[17] A. Murata, H. Nishida, K. Watanabe, and T. Kameda, "Convergence of Physiological Responses to Pain during Face-to-face Interaction," Scientific Reports, Vol. 10, 450, Jan. 2020.
[18] S. Genyu, "Zen Happiness Theory," KADOKAWA SSC Books, 2010 (in Japanese).

Takeshi Yamada
Vice President and Head of NTT Communication Science Laboratories.
He received a B.S. in mathematics from the University of Tokyo in 1988 and a Ph.D. in informatics from Kyoto University in 2003. He joined NTT Electrical Communication Laboratories in 1988. He was a visiting researcher at the School of Mathematical and Information Sciences, Coventry University, UK, from 1996 to 1997. He was a group leader of the Emergent Learning and Systems Research Group from 2006 to 2009 and an executive manager of Innovative Communication Laboratory from 2012 to 2013 at NTT Communication Science Laboratories. His research interests include data mining, statistical machine learning, graph visualization, metaheuristics, and combinatorial optimization. He is a Fellow of the Institute of Electronics, Information and Communication Engineers, senior member of the Institute of Electrical and Electronics Engineers, and a member of the Association for Computing Machinery and the Information Processing Society of Japan.