

## Reach Out and Touch Someone's Heart: Exploring the Essence of Communication to Create a Spiritually Rich Society

*Takeshi Yamada*

### Abstract

NTT Communication Science Laboratories has been exploring the essence of communication since its founding 30 years ago. With the aim of achieving communication that *reaches the heart*, its researchers have been creating innovative technologies that approach and exceed human abilities in fields such as media processing and data science. They have also been discovering basic principles that lead to a deeper understanding of humans in fields such as cognitive neuroscience and brain science. This article introduces key activities at NTT Communication Science Laboratories in pursuit of the essence of communication with a look back at past research.

*Keywords: artificial intelligence, machine learning, cognitive neuroscience*

### 1. Introduction

This year marks the 30-year anniversary of NTT Communication Science Laboratories (NTT CS Labs), which was founded on July 4, 1991. Throughout these 30 years, it has been exploring the essence of communication and conducting basic research to enable communication that *reaches the heart* [1]. The essence of communication is inherently multifaceted. In addition to (1) conveying information accurately and efficiently, it includes the (2) deepening of mutual understanding by sharing the meaning of information and (3) sharing of underlying intent and emotion by devising creative methods of conveying information, enabling the (4) creation of a spiritually rich society by fostering *heartfelt* contact. With a focus on these four viewpoints of communication, this article introduces important activities at NTT CS Labs in pursuit of the essence of communication while taking a look at past research.

### 2. Basic technologies for information transfer and speech coding

NTT has been continuously engaged in the research of basic communication technologies such as audio and voice processing and natural language processing since the Nippon Telegraph and Telephone Public Corporation era. The root of this research is speech coding technology, which is one of the most important technologies from the viewpoint of transmitting information accurately and efficiently. Line Spectrum Pair technology proposed by NTT in 1975 is still used in most mobile phones throughout the world as an international standard, and in 2014, it was recognized as an IEEE (Institute of Electrical and Electronics Engineers) Milestone marking a historic achievement in the field of telecommunications [2]. Inheriting this legacy, NTT CS Labs became a leading contributor in the development of Enhanced Voice Services (EVS) technology that was approved as a 3GPP (3rd Generation Partnership Project) standard in 2014. In Japan, this standard has been used since 2016 as

fourth-generation coding in a coding-transmission system shared by three mobile phone companies, and as of 2021, it has been used in smartphones throughout the world. NTT CS Labs has also been contributing to the EVS extension for Immersive Voice and Audio Services called IVAS. More recently, it developed Bitplane Rearrangement for Audio and Voice Encoding (BRAVE), an audio and voice codec featuring robust bit-error performance and low latency. BRAVE was adopted in wireless microphones commercialized by TOA Corporation in February 2021 [3].

### 3. Obtaining categories and concepts to share meaning

Humans learn by grouping similar things into categories, which enables advanced cognitive activities such as thinking, inferring, decision-making, and communication. It also makes learning itself more efficient. For example, when catching sight of an animal, if its form happens to be sufficiently similar to a category that one has already learned, say “cats,” that animal would be recognized as a member of that category, in other words, as a cat. While it is difficult to individually remember all objects one has seen up to the present, they can be remembered in a compact form by grouping them into categories. Later, when looking back, it may not be possible to remember the detailed features of that cat, but the fact that it was a cat will not be forgotten. In addition, if something that one encounters is different from any category that one has already learned, one can simply create a new category. Therefore, learning can be made efficient by flexibly increasing, or even decreasing, categories as needed in accordance with data characteristics even for large volumes of data.

NTT CS Labs has been working on achieving such flexible human-type category learning on computer. For example, the products that each customer has purchased can be recorded in matrix form as a history of purchased data that can be used to categorize both customers and products. This type of categorization corresponds to rectangular partitioning of a purchased data matrix. An efficient learning technique based on a Bayesian nonparametric model was proposed that adjusts the optimal partition in accordance with the given input data from among an infinite number of combination patterns in rectangular partitioning [4].

Thus, “cats” as a category is not a specific “cat” but rather an abstraction of “cats” in general. A concept,

on the other hand, is a mental representation of a category stored in memory. In other words, it is a set of information that a category points to and consists of what is known about that category [5]. The concept of “cats” that humans hold is not limited to the shape or form of cats. It is rather an integrated abstraction of various aspects of cats, such as the sounds they emit (meowing, etc.), their behavior, the feel of their fur, etc. and the language used to express such aspects. That is, a concept can be acquired by seeing a thing (its set) from different viewpoints (different types of media information or modalities) and be understood as abstract information independent of any viewpoint and expressed as coordinates in a common conceptual space. NTT CS Labs is researching the autonomous acquisition of concepts without having to train a system with correct answers. This can be done by focusing on the co-occurrence of different types of media information such as images and sounds of cats, that is, by using the fact that different types of media information originating from the same thing appear not in a random manner but with specific relationships [6].

### 4. Communication and language acquisition in infants

Do human infants autonomously learn from the co-occurrence of phenomena in the natural world? To comprehend the essence of communication, NTT CS Labs has been examining communication and language acquisition in infants. For infants, communication is an important means of recognizing objects and promoting the acquisition of knowledge, concepts, and vocabulary. Infants accumulate various types of knowledge from information obtained from the surrounding environment, such as by listening to a parent’s conversation, speech from a television, etc., and learn groups of syllables that co-occur with high frequency as words based on statistical learning. However, this does not mean that the infant indiscriminately processes a huge amount of information. Research conducted at NTT CS Labs has found that learning in infants is promoted by communication signals from a parent such as utterances directed toward the infant [7]. The infant uses such communication signals as a learning cue to focus appropriately on learning targets and sort out what to learn from the environment and how.

As explained above, parent-infant communication promotes brain development of the infant. It further affects subsequent vocabulary growth. NTT CS Labs

has been promoting research on language acquisition in infants, and from the results of that research, more than 280,000 copies of picture books for young children supervised by NTT CS Labs have been published. What is significant is that these books are not digital but rather printed material that children can interact with using the five senses. More recently, “personalized educational picture books” was proposed in collaboration with NTT Printing Corporation. These are educational picture books with pictures emphasizing new words for an individual child to learn on the basis of vocabulary checking conducted by the child’s parents and on a child-vocabulary-development database developed through research at NTT CS Labs. This venture began with picture books to be read out loud to children, but more recent research at NTT CS Labs revealed that an understanding of characters and their correspondence to sounds actually starts around three years old, slightly before the ability to read and write *hiragana* (Japanese syllabic characters). Thus “names-in-*hiragana/katakana* picture books” was proposed to generate interest in characters targeting children of about three years old. Personalized educational picture books can now be ordered online at [ehon.nttprint.com](http://ehon.nttprint.com) [8].

NTT CS Labs has also undertaken research on the interaction between a parent and infant focusing on the parenting side. Parenting stress and postpartum depression in mothers, child abuse and neglect, etc. have become problems throughout society. To study how mothers interact with infants, types of infant vocalizations and the manner in which a mother approaches her infant in response to those sounds were investigated. It was found that a mother would reflexively respond to the sound of crying and that her response in approaching her infant would become stronger the more urgent those sounds feel to her. In short, the sound of crying arouses a feeling of wanting to respond quickly (a sense of urgency) in the mother, who then approaches the child in a reflexive, unconscious manner.

Humans are equipped with a mechanism for suppressing this reaction. Specifically, there is a hormone called oxytocin. This hormone serves to secrete mother’s milk. It is also called a prosocial hormone since it is known to easily arouse positive emotions with respect to another person. The concentration of oxytocin is also known to have a positive correlation with caregiving motivation in the mother. Research at NTT CS Labs has found that oxytocin suppresses this reflexive impulse in the mother to approach her infant

at the sound of crying [9]. This research suggests that if the level of oxytocin is low, the mother loses her composure and wants to quickly stop her baby from crying, but if the level of oxytocin is high, parasympathetic nerve activity increases, resulting in making the mother more relaxed and suppressing a reflexive approach to her crying baby. This result may lead to knowledge on how to promote a sense of well-being in parenting.

## 5. Creating new forms of communication

NTT CS Labs is also working on ways of communicating, that is, on creating new forms of communication. “The medium is the message” is the famous phrase coined by Marshall McLuhan, a scholar of English literature. Through these words, McLuhan asserts that a message includes the means of conveying the message and stresses the importance of the medium that transmits the message in communication, that is, the sensory image that the medium itself possesses. In this regard, there is the famous slogan “Reach out and touch someone” used by AT&T, the American telecommunications company, in commercials in the 1970s with the intention of softening its stiff image [10]. McLuhan contributed to the creation of this catchphrase, which was novel for its time.

In line with this “reach out” point of view, NTT CS Labs once researched a room-sized remote communication system called “t-Room.” This is a system in which multiple geographically and temporally separated users share the “feeling of being in the same room” while being at remote locations [11]. However, t-Room did not include the sharing of the sense of touch. For this reason, NTT CS Labs is now researching new sensation-presentation technology using touch that would enable *kansei* (emotionally rich and sensitive) communication to convey deep feelings by touch. The “Mega-Futuristic Experiential Public Telephone” (versions 3 and 4) proposed in 2018 is a touch-based communication system in which pressing the push buttons of a telephone causes a variety of tactile sensations to stimulate the other party’s body. More recently, new systems such as Remote High Five and Public Booth for Vibrotactile Communication that truly share tactile sensations beyond distance have been proposed [12].

NTT CS Labs is also researching a means of speech conversion that would enable content that one would like to convey to be freely converted to one’s desired form of expression for transmitting and receiving. This research is expected to create new forms of

communication that extend human vocal and auditory functions.

To achieve communication that *reaches the heart*, methods are needed for picking up what a person is feeling from the outside without placing too much of a burden on that person. Regarding the well-known saying, “the eyes are the window to the soul,” NTT CS Labs discovered that a human’s pupil unconsciously constricts on seeing an attractive face. Consequently, if the size of a person’s pupil can be measured in real time, it would be possible to pick up what that person is feeling to some extent. At the same time, it was found that making a person’s pupil constrict through luminance/contrast changes actually enhanced the attractiveness of a face as seen by that person. This result suggests that controlling—as opposed to measuring—the size of a pupil could change to some extent that person’s preferences [13].

## 6. Spiritually rich society with diverse values in harmony

Finally, from the viewpoint of *heartfelt* contact, I would like to introduce research in pursuit of the essence of communication from a slightly different angle. It is often said that modern society is becoming increasingly divided. Due to the flood of information, people are becoming increasingly confrontational when they see two sides of an issue that appear to be at odds with each other. Instead of listening to different opinions, they take one side while sacrificing the other, as in globalism or nationalism, centralization or decentralization, and analog or digital. However, precisely for this reason, it is indispensable to create a spiritually rich society that allows for contradictions, recognizes diverse values, deepens mutual understanding through communication while protecting privacy, and nurtures *heartfelt* contact by promoting empathy.

In machine learning, especially deep learning, a massive increase in data and the need to protect privacy are generating a need to distribute and store training data on a group of local servers. However, if each server should train locally under these conditions, the end result will be trained models that are different and mutually contradictory. These models will not converge if they are poorly coordinated. Therefore, NTT CS Labs devised an asynchronous distributed deep-learning framework in which data dispersedly stored on a group of dispersedly located servers can be trained as a global model as if the data were consolidated at one location by having the serv-

ers communicate with each other to build a consensus [14].

NTT CS Labs also devised three-dimensional (3D) video-generation technology to enable clear viewing of the corresponding 2D image with the naked eye. In short, this is technology that enables people who prefer to view 3D images with 3D glasses and people who are uncomfortable with 3D and prefer 2D to enjoy the same image together without sacrificing the needs of the other [15].

## 7. Conclusion

In this article, I introduced key activities at NTT CS Labs in pursuit of the essence of communication with the aim of achieving communication that *reaches the heart*, or more exactly, that *reaches out and touches someone’s heart*. The way we interact with other people is changing due to countermeasures and restrictions related to the COVID-19 pandemic. It is especially necessary at this time to pursue the possibilities of new media leveraging the five senses while identifying and solving any problems that may arise in this pursuit. Before trying to convey one’s feelings to someone far away, one should realize that people do not sufficiently know themselves and their true feelings in the sense that the unconscious can be a stranger within. Deepening an understanding of oneself can improve the quality of one’s daily life. Going forward, NTT CS Labs will challenge the so-called technical limits of approaching human intelligence and, at the same time, will strive to obtain a new understanding of what it means to be human by incorporating diverse points of view from the social sciences, philosophy, and other fields.

## References

- [1] T. Yamada, “I Want to Learn More about You: Getting Closer to Humans with AI and Brain Science,” NTT Technical Review, Vol. 18, No. 11, pp. 11–15, Nov. 2020.  
<https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202011fa1.pdf>
- [2] T. Moriya, “LSP (Line Spectrum Pair): Essential Technology for High-compression Speech Coding,” NTT Technical Review, Vol. 12, No. 11, Nov. 2014.  
<https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201411in1.pdf>
- [3] Press release issued by TOA, “New Series of 800MHz Band Digital Wireless System with High Sound Quality that Can Use up to 15 Microphones at the Same Time,” Feb. 10, 2021 (in Japanese).  
<https://www.toa.co.jp/products/news/2021/news2021-02-10wm.htm>
- [4] M. Nakano, A. Kimura, T. Yamada, and N. Ueda, “Baxter Permutation Process,” Proc. of the 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Dec. 2020.
- [5] L. J. Rips, E. E. Smith, and D. L. Medin, “Concepts and Categories: Memory, Meaning, and Metaphysics,” in The Oxford Handbook of

- Thinking and Reasoning, ed. K. J. Holyoak and R. G. Morrison, pp. 177–209, Oxford University Press, 2012.
- [6] K. Kashino, “See, Hear, and Learn to Describe—Crossmodal Information Processing Opens the Way to Smarter AI,” NTT Technical Review, Vol. 17, No. 11, pp. 12–16, Nov. 2019.  
<https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201911fa2.pdf>
- [7] Y. Okumura, Y. Kanakogi, T. Kobayashi, and S. Itakura, “Ostension Affects Infant Learning More Than Attention,” Cognition, Vol. 195, 104082, Feb. 2020.
- [8] Personalized educational picture books (in Japanese), <https://ehon.nttprint.com/>
- [9] D. Hiraoka, Y. Oishi, R. Mugitani, and M. Nomura, “Relationship between Oxytocin and Maternal Approach Behaviors to Infants’ Vocalizations,” Comprehensive Psychoneuroendocrinology, Vol. 4, Nov. 2020.
- [10] C. S. Fischer, “‘Touch Someone’: The Telephone Industry Discovers Sociability,” Technology and Culture, Vol. 29, No. 1, pp. 32–61, Jan. 1988.
- [11] K. Hirata, Y. Harada, T. Takada, S. Aoyagi, Y. Shirai, N. Yamashita, and J. Yamato, “The t-Room—Toward the Future Phone,” NTT Technical Review, Vol. 4, No. 12, pp. 26–33, Dec. 2006.  
<https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr200612026.pdf>
- [12] “Public Booth for Vibrotactile Communication with Heightened Presence,” Bimonthly Magazine Furue, Vol. 26, Dec. 2019 (in Japanese), <http://furue.ilab.ntt.co.jp/book/201912/contents3.html>
- [13] H.-I. Liao, M. Kashino, and S. Shimojo, “Attractiveness in the Eyes: A Possibility of Positive Loop between Transient Pupil Constriction and Facial Attraction,” Journal of Cognitive Neuroscience, Vol. 33, No. 2, pp. 315–340, Feb. 2021.
- [14] K. Niwa, N. Harada, G. Zhang, and W. B. Kleijn, “Edge-consensus Learning: Deep Learning on P2P Networks with Nonhomogeneous Data,” Proc. of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD), pp. 668–678, Aug. 2020.
- [15] T. Fukiage, T. Kawabe, and S. Nishida, “Hiding of Phase-based Stereo Disparity for Ghost-free Viewing without Glasses,” ACM Transaction on Graphics, Vol. 36, No. 4, 147, July 2017.



### 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 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 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.