No.48 Jan.2014 ISSN 1883-1974 National Institute of Informatics News

This English language edition NII Today corresponds to No.62 of the Japanese edition

NII Today

National Institute of Informatics News

FEATURED **TOPIC**

Can a Robot Join an *Idobata Kaigi*?

INTERVIEW WITH

Mayumi Bono

Assistant Professor, Digital Content and Media Sciences Research Division, NII
Assistant Professor, Department of Informatics, School of Multidisciplinary Sciences
The Graduate School for Advanced Studies
Principal Investigator, Ido-Robo Project

I Want to Get to the Heart of Communication with the Ido-Robo Project

Can a robot join an *idobata kaigi* ("congregation at the well" – a Japanese concept that represents how Japanese women in a village used to chat, circulate gossip, and exchange community information as they gathered at the well to wash clothes and draw water)? This is probably as difficult as – no, it is even more difficult than – a robot passing the University of Tokyo's entrance exam. In the first place, it is not clear how individuals become a member of a well-side congregation, nor is it clear how human beings have natural conversations. As an NII Grand Challenge that seeks to deepen our understanding of human interaction, the project "Can a Robot Join an *Idobata Kaigi*?" ("Ido-Robo Project" for short) began in FY2012. Assistant Professor Mayumi Bono is its principal investigator. Professor Toyoaki Nishida, an AI researcher and a founder of the field of conversational informatics, interviews her about the concept of the project and its direction

Nishida "Ido-Robo," the project you've been working on, sounds fascinating. What led you to begin this research?

Bono The seeds of this project were planted when I was chosen as a Sakigake*1 researcher in October 2009. With the Sakigake grant, I pursued research on telecommunication with hearing-impaired people who use sign language. At the same time, this project involved trying to understand human interaction. As I was conducting my research, the Todai Robot Project*2 began in November 2011. This led me to want to further deepen my exploration of human communication, and so I proposed the Ido-Robo Project. The Todai Robot Project approaches the question of intellect and intelligence from the study of human abilities, the brain, logic, and so on. In contrast, the Ido-Robo Project approaches the question of human beings' social nature from the study of conversations. Until now, understanding interaction has been an area of research in the humanities, which is my background. Researchers in subjects like conversational analysis and interaction studies used their own original methods, which can be likened to their virtuoso performances. With the Ido-Robo Project, I think I can build a platform that treats the understanding of interaction as a science by incorporating the highly versatile methods of informatics.

Nishida Do you plan to build a conversational robot by yourself?

Bono I'm not creating an actual robot myself. However, we have the participation of some robotics engineers. They contribute to our discussions by taking a constructive approach – they seek to understand a subject by making actual things. There are many elements in interaction studies that I think robotics engineers will find fascinating, and I want to share these things with them. In the near future, I hope to be able to be directly involved in building a robot through joint research.

Nishida *Idobata kaigi* sounds more sophisticated than mere conversation.

Bono The key point is seeking to understand what happens when a conversation between two people changes to an interaction with three or more people. In some previous robot research efforts, robot design was based on the assumption that conversations take place between a person and the robot. However, when we focus on illuminating the mechanisms of interaction, we must think about conversation's surface phenomena a little deeper. For example, how humans act in response to something that occurs during a conversation, like a gesture or the timing of an utterance, takes on greater importance. I think such an emphasis will change the way robots are designed.

Nishida Did you have difficulty in figuring out how to

express the idea of "*idobata kaigi*" in English?

Bono In interaction studies, the term "multi-party" refers to the gathering of a large number of people, such as at a lecture. So, a more apt term for this research would be "small-party," but that has a negative connotation. The name "*idobata kaigi*" works well as it is, because the concept of a "well-side congregation" is so Asian. So we don't need to refer to it by another term. Such academic research is possible only in Japan, isn't it?

Nishida Your focus on Asian-style conversation is fascinating. What specifically are you highlighting in the Ido-Robo Project?

Bono First, a major pillar is research on robot theater as a field of study. I'm focusing on theater plays directed by Oriza Hirata in which robots and androids designed by Professor Hiroshi Ishiguro, a robotics researcher, perform. This research got its start when Professor Ishiguro said to me, "Your research is boring! What you want to do is already all in Oriza's head! In other words, the "patterns of human conversation" are in Oriza's head. This inspired me, and I decided to ask Oriza to show me the actual place where he produces robot theater. I video-recorded the rehearsals over the course of several months, and am now analyzing them. Another major pillar of research is conducting conversation analysis where we investigate conversational practices between the science communicator and visitors at the National Museum of Emerging Science and Innovation. I'm focusing on how we can effectively convey scientific knowledge. Furthermore, I've begun fieldwork in the village of Nozawa Onsen to examine actual *idobata kaigi*. I haven't been able to see many *idobata kaigi* conversations recently, so I've actually made villagers' gatherings to prepare for a fire festival as the actual object of my research.

Nishida The Todai Robot Project takes a goal-oriented approach where much effort is placed on putting together elemental technologies to realize an integrated system. In contrast, the Ido-Robo Project takes a rather exploratory approach that seeks to discover interesting phenomena in fields you can't even imagine before you begin exploring. A very interesting scientific challenge here is the dilemma where on the one hand, you must abstract the phenomenon in order to gain generalization, while on the other hand, you have to retain the details in order not to lose the essence of communication.

Bono Professor Charles Goodwin of the University of California, Los Angeles, is a pioneer in multimodal interaction studies. He incorporates gazes and body movements in his research. He usually doesn't conduct quantitative analysis. However, he carried out quantitative analysis from the outset in a paper, and what's more, he showed the importance of qualitative analysis by analyzing data in detail that was left out from quantitative analysis.

Nishida This is a very important point. Scientific and non-scientific fields are not two opposing sides, but instead they are brought together to produce new knowledge. I believe this direction will also pave the way to resolving the negative effects of achievement-oriented and efficiency-oriented thinking. What do you aim for as the final goals of the Ido-Robo Project?

Bono If the Todai Robot Project seeks to cleverly fill the brain of a single robot with knowledge, with the Ido-Robo Project I want to show how conversations and interactions play critical roles in supporting human society. I also want to integrate modalities of interaction that have been segregated until now, such as utterances, expressions, and gestures, and I want to reveal the essence of humanity by systematizing these different fields of study. Going forward, I want to focus on interactions to provide better service – in another word, *omotenashi* (Japanese-style hospitality).

Nishida The idea of integrating insights from multi-disciplinary approaches is something your project has in common with the Todai Robot Project. In this world, we need problem-solving expertise, but it would be a boring world without comedy like the Yoshimoto Shinkigeki troupe and *omotenashi*, wouldn't it? Grappling with *omotenashi* as a science is very challenging. You will have the opportunity to show off your skills by penetrating to the heart of the problem. I look forward to witnessing how the Ido-Robo Project develops.

(Written by Madoka Tainaka)

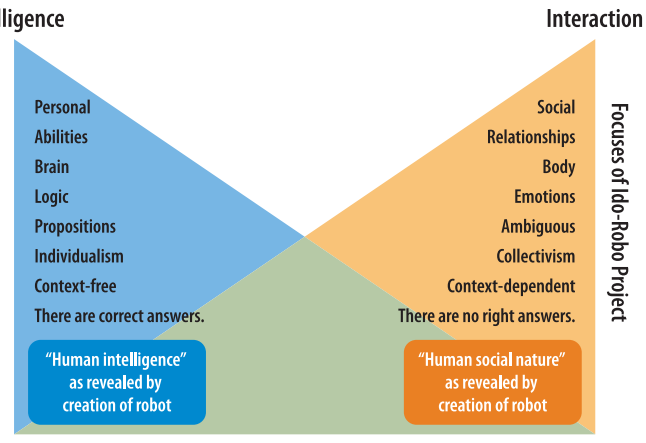


Figure: Comparison of Todai Robot Project and Ido-Robot Project

A Word from the Interviewer

Conversational informatics seeks to explore the past and future of our spiritual world by cross-fertilizing individual areas of science and technology centered on the phenomenon of conversation. Assistant Professor Bono's Ido-Robo Project has the strong potential to sharply carve the contours of the indivisible spiritual world not accessible by an approach solely guided by rationality. On the surface, this project stands in contrast to the Todai Robot Project in various respects. However, far beneath the surface, they both will eventually penetrate into the essence of humanity by integrating different areas of knowledge. I'm very much looking forward to this project's progress.

Toyoaki Nishida

Professor, Department of Intelligence Science and Technology Graduate School of Informatics, Kyoto University

Graduated from the Faculty of Engineering, Kyoto University in 1977. Obtained master's degree from the same university in 1979. He was appointed professor of the Nara Institute of Science and Technology in 1993; the School of Engineering, the University of Tokyo in 1999; and the Graduate School of Information Science and Technology, the University of Tokyo in 2001. Since April 2004, he has been serving as professor of the Graduate School of Informatics, Kyoto University. He is engaged in the research of conversational informatics, social intelligence design, and primordial knowledge models.

***1 Sakigage ("Pioneer") Research**
Personal research projects (PRESTO) sponsored by the Japan Science and Technology Agency. The purpose of the grants is to nurture future innovations based on strategic goals.

*** 2 Todai Robot Project**
An NII Grand Challenge AI project formally entitled "Can a Robot Get Into the University of Tokyo?" http://21robot.org.
Introduced in NII Today No. 60 (http://www.nii.ac.jp/muom2c5rm-4542/#_4542).

Exploring Interaction between Humans and Robots with Conversation Analysis

The Ido-Robo Project seeks to place the understanding of human beings' daily interactions, such as conversations and gestures, in informatics. It applies interdisciplinary efforts to discover knowledge that can be exploited in future robot design. In this respect it has deep connections with the field of conversational analysis (CA), the specialty of Professor Aug Nishizaka, a sociologist. Assistant Professor Mayumi Bono, who has a background in the humanities and is also engaged in CA, discusses the role played by sociology in human-robot interaction with Professor Nishizaka.

Investigating the Structure of Utterances

Bono Traditionally, research of conversational robots has proceeded within a framework that considers a robot as being able to talk if a speech-recognition system is created and incorporated into the robot. In contrast, the Ido-Robo Project is pursuing knowledge such as the "structure" of interaction with two or more people. In other words, I hope to offer another perspective for social robotics research by exploring the structure of human social exchange.

One pillar of this research is the study of "robot-human theater." I'm exploring the relationships between fine body movements and gaze movements and conversation as I observe director Oriza Hirata's production and collect a variety of data. As part of this study, I'm focusing on the start of conversations and communication. I think the methodology of conversational analysis (CA) can be used to analyze the flow of lines of script as they are repeated under Oriza's direction during rehearsals.

Nishizaka How fascinating. In CA, turn-taking rules on who speaks next in a conversation are formulated in such a way that at first glance, it looks as if they can be converted directly into a computer program. Because of this, at one time AI researchers thought they could use the rules to create conversational robots. But the results didn't turn out so well. This is because even if the rule can be formalized, actual conversations depend all too much on context.

Turn-taking rules mean prioritizing between two techniques that select the next speaker. Either the current speaker chooses the next speaker, or the next speaker begins to speak on his or her own. The first technique is composed of two elements. One, the current speaker addresses utterances to a particular person. Second, the speaker makes utterances that strongly elicit a particular behavior. For example, if the speaker addresses a question to a specific person, that person becomes the next speaker and is given a strong prompt to reply.

Bono But, in actual cases, what happens doesn't always work according to theory, right?

Nishizaka Right. Suppose the sentence that the speaker says is a question in Japanese. Even so, such a sentence doesn't always end explicitly with the sentence-ending particle "-ka" to indicate that it is a question. What's more, it is not always clear to whom a question is addressed. If I ask in Japanese, "Bono-san no shusshin wa doko?" ("Where's Bono-san from?"), if you are in front of me, then the question is addressed to you. But if you aren't there, then the same question can be directed to someone who knows you. In other words, it all depends on context. Formalizing this structure of "utterance design" to allow robots to have conversations is extremely difficult. To put it another way, this shows how flexible humans are in their adaptability to context. I think what is deeply intriguing about the Ido-Robo project is that it is not an attempt to create a robot that can converse like a human being, but rather it seeks to bring about perceptional changes by placing something in the likeness of humans in our midst to study how we converse.

Mayumi Bono
Assistant Professor, Digital Content and Media Sciences Research Division, NII
Assistant Professor, Department of Informatics, School of Multidisciplinary Sciences
The Graduate University for Advanced Studies
Principal Investigator, Ido-Robo Project

Robot-ness, Human-ness

Bono In the first place, engineers segregate conversations from context all too easily when they design robots. That's why I started the Ido-Robo project. I want to incorporate the complexity of conversation into the design of social robots a little more. Even at a performance of the robot-human theater, there are actually a variety of things that can start a conversation, such as the actors or robots' casual nodding to each other, making eye contact, and moving closer together. For example, imagine a scene where a robot that gathers data in a lab is giving a tour of a facility. As it is having a conversation with a character, another character walks by. The passerby's gaze is received, and a complex interchange of gazes between the three characters takes place. What's more, added into all this are Oriza's detailed directions, such as when to look at a character, how to return a gaze, and the robot's nod to the passerby. So, even when the passerby is a silent third party, his presence increases the interactions. The relationships clearly change when a three-person conversation takes place.

Nishizaka I'm excited to see how robot-ness is staged in the plays. After all, robots and humans are different in their

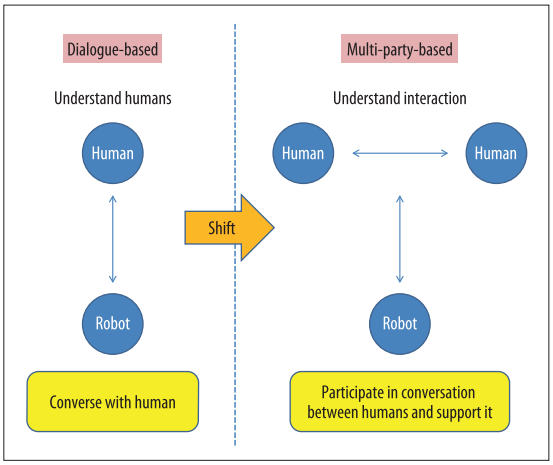


Figure 1 Core of Ido-Robo Project

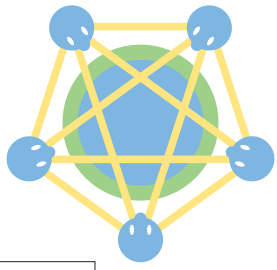


Figure 2 Director Oriza Hirata stages an exchange between robots and customers at a shopping mall to examine the interaction between humans and robots in the future. There are nearly 50 patterns of interaction. Here, human actors play both human and robot roles. (Photos courtesy of the Advanced Telecommunications Research Institute International [ATR], taken on May 14-15, 2013.)

essence. Oriza's project can be seen as an attempt to reveal the difference between humans and robots by showing just how humans and robots that behave like humans differ.

Giving the Insights of Conversation Analysis **Back to Society**

Nishizaka CA is originally a research area under sociology. It is a methodology to observe and analyze the interaction between human beings called "conversation." The concept of analyzing conversations arose as personal audio recorders became available in the early 60s. This made it easy to record and play back voice recordings. Next, the invention of compact video cameras made it convenient to record real-life scenes. So sociologists started paying attention to how visual information such as gazes and gestures function as resources to realize conversations. They found that while they could systematically extract elements that are in play in the exchange of words, formalizing conversation resources besides talk was difficult, because those other resources were too analog.

Bono Non-talk elements are still difficult to formalize, aren't they? But when it comes to talk, long years of CA research in that area have brought beneficial results to society. For example, I've heard that analyses of calls to the police and fire departments have proved useful as materials in court. Professor Nishizaka, you recently analyzed conversations by survivors of the Great East Japan Earthquake at footbaths in shelters in Fukushima, and wrote about your findings in a book, didn't you?

Nishizaka In the aftermath of the disaster, the shelters were still in chaos. The evacuees frequently didn't know what they themselves needed. One helpful piece of knowledge gained from the experiences of the Great Hanshin Earthquake and the Chūetsu Earthquakes was to set up footbaths. When the evacuees used them, they naturally talked about the things that were troubling them. So, I thought that by applying CA, we could explicate the structure of communication at the footbaths. I thought we could contribute to volunteering at the shelters this way.

Bono What did you find?

Nishizaka The evacuees may find it easier to tell their experiences thanks to hand massages and footbaths. But more than that, we understood how important the structure of the interaction was. Even if the conversation itself stops, the interaction continues. In other words, massaging is the base that supports interaction. It actually serves as the foundation to the very end of the interaction, but doesn't hinder conversation that is built on top of it. So, as the evacuees receive hand massages at the footbaths from volunteers, it's not unusual that they naturally tell about their experiences, even if the volunteers don't ask. This finding can show the way communication takes place in places like shelters and temporary housing if similar structures are present, even if the footbaths themselves are not there.

Bono New interactions arise as people do something while conversing—this story relates to the Ido-Robo Project, doesn't it?

Nishizaka The structures of human interaction are revealed by placing robots in the middle of interactions. That is precisely the appeal of the Ido-Robo Project. I look forward to this research's progress.

(Written by Shigeyuki Ohara)

Aug Nishizaka
Professor, Department of Sociology
Meiji Gakuin University

The Future of *Idobata Kaigi* ("Congregation at the Well"): Realizing Natural Conversations with Distance Access

Conversations in daily life consist not only of verbal information, but also nonverbal information frequently conveyed through eye gazes, gestures, and other body and hand movements. It is thus crucial in informatics to understand nonverbal communication to achieve natural conversation between humans and social robots. Meanwhile, efforts to develop video conferencing systems that allow us human users to exchange nonverbal as well as verbal information are considered as the forerunner of research in this domain. Dr. Kazuhiro Otsuka, senior research scientist of NTT Communication Science Laboratories, has worked on this challenging yet fascinating topic with NII for several years now. We interviewed him to hear about the latest developments in this area of research.

Daily Life Is Full of Nonverbal Information Exchange

When communication such as daily conversation and meetings takes place, verbal information conveyed by voice is not the only information being exchanged. Nonverbal information, such as gazes, facial expressions, gestures, body and hand movements, and the tone of voice, are also being exchanged. Communication in diverse forms occurs as a result of the comprehensive use of such information, and during the course of communication, conversation and topics change. Therefore, what is critical for realizing the Ido-Robo Project, which aims for coexistence with humanity, is not only understanding verbal information, but also elucidating communication processes that include non-verbal information.

The Real-time Multimodal System for Conversation Scene Analysis was developed by the NTT Communication Science Laboratories (NTT CS Labs) to study this theme from the standpoint of information engineering.

This system analyzes dynamically and in real time multiple people who are involved in a conversation. Specifically, in a small meeting room of about eight people, an integrated omnidirectional camera/microphone system placed on the table gathers image and voice information. The system then analyzes the state of the conversation by processing and integrating the obtained information, and presents the results on a display. By doing so, conference participants can monitor in real time the state of the conversation, answering questions such as "Who is speaking when?", "Who is talking to whom?", and "Who is paying attention to whom?"

NTT CS Labs Senior Research Scientist Kazuhiro Otsuka explains the background that led to the development of the system: "Videoconferencing systems hold great promise as a technology that allows people to communicate at a distance. However, they haven't advanced to the stage where smooth conversations like actual face-to-face conversations are possible."

Kazuhiro Otsuka
Senior Research Scientist, Supervisor/Distinguished Researcher
Sensory Resonance Research Group
Human Information Science Laboratory
NTT Communication Science Laboratories

Conventional Videoconferencing Systems Cannot Convey the Atmosphere of a Place

Conventional videoconferencing systems use flat displays. This makes it difficult to convey the atmosphere of the meeting. Problems include gaze mismatch, in which it is difficult to understand who is looking at whom. Therefore, to recreate more realistic conferences on videoconferencing systems, we must understand the processes and mechanisms by which communication between people is established. These elements include the atmosphere of the meeting.

Dr. Otsuka says: "The Real-time Multimodal System for Conversation Scene Analysis leverages imaging and voice technologies developed by NTT CS Labs to observe human behavior as visual and audio information. We believe we can elucidate the mechanisms of communication by analyzing the information we obtained with computers."

Nobuhiro Furuyama
Associate Professor, Information and Society Research Division, NII
Associate Professor, Department of Informatics, School of Multidisciplinary Sciences
The Graduate University for Advanced Studies
Visiting Associate Professor, Department of Computational Intelligence and Systems Science
Interdisciplinary Graduate School of Science and Engineering
Tokyo Institute of Technology

Assistant Professor Bono's research provided an inspiration for the development of the system, says Dr. Otsuka. "Communication research until now has prioritized language. But Assistant Professor Bono's research made us realize that in a conversation, the content is not just verbal, but also includes human behavior."

Assistant Professor Bono says: "There are few researchers in informatics who conceive research of communication from the standpoint of understanding human behavior. So I thought we should mutually support one another. The outstanding point of this system is its method of presenting images of the multiple participants in a conversation in a cylindrical form so that they can all be shown on the display (Photo 1). This makes it very easy to reproduce spatial information in a conference room with multiple video cameras. This information was lost with conventional video capture methods. Analyzing gazes and the direction of the

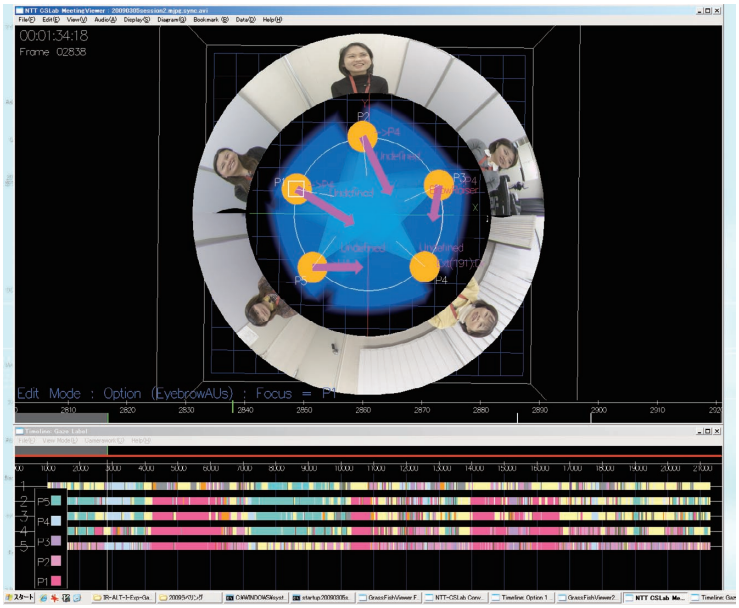


Photo 1: Real-time Multimodal Conversation Analysis System

Photo 2: t-Room

Photo 3: MM-Space

movement of the participants' bodies also becomes easy."

Associate Professor Nobuhiro Furuyama, who participates in the Ido-Robo Project with research of perception, behavior, and embodied communication, says: "The Real-Time Multimodal System for Conversation Scene Analysis is extremely effective as a tool for thoroughly investigating how people use gestures, which are often vague, as resources for communication."

In 2009, "Corpora of Four-Member Spoken Conversation and Four-Member Signed Conversation" was jointly developed by NII and NTT CS Labs using the Real-Time Multimodal System for Conversation Scene Analysis and video recording equipment installed in Dr. Otsuka's lab. Furthermore, in October of the same year, the two organizations signed a comprehensive institute-to-institute partnership agreement. Both research institutes are now conducting joint research to realize the Ido-Robo Project and create a "future *idobata kaigi*."

Mayumi Bono
Assistant Professor, Digital Content and Media Sciences Research Division, NII
Assistant Professor, Department of Informatics, School of Multidisciplinary Sciences
The Graduate University for Advanced Studies
Principal Investigator, Ido-Robo Project

Next-Generation Videoconferencing System "t-Room" and "MM-Space"

NTT CS Labs is working to bring about seamless communication between remote participants. Examples of the fruits of their research include the next-generation videoconferencing systems "t-Room" and "MM-space."

t-Room was developed with the purpose of achieving communication that allows each conference participant in their respective remote locations to feel as though everyone was in the same room together. Specifically, every wall in a polygonal room is made up of a display. Video cameras are placed on the top of each display. The captured video of each participant is shown on the displays of t-Rooms in other locations (Photo 2). If a participant moves, his or her image moves in sync from one display to the neighboring display. The t-Room system makes possible communication that allows the participants to experience being in the same room together. For example, if a participant points in a particular direction, all the other participants can look in that direction to a display that shows what the participant is pointing at.

MM-Space is a system that makes it possible to reproduce a conversation of multiple people in remote locations so that each participant feels as if they were involved in the conversation in the same place (Photo 3). The face and voice of participants are captured in the space where they are actually speaking and transmitted to another space where the conversation is reproduced. At the destination of the transmission, the presence and position of the conference participants in remote locations are reproduced by multiple sets of projectors, transparent screens, and speakers (a set for each remote participant) and actuators that move the screens. The facial image of each remote participant was projected on his or her transparent screen with the background removed. This screen is moved by actuators in synchronization with the movement of the participant's head. In other words, the MM-space system analyzes nonverbal behaviors that are produced in an actual conversation, such as a participant's gaze, tilting and turning of the head, nodding, and so on, and represents them as physical movements of the screen. It is as if the remote participant was actually physically present at the meeting, turning his or her head to others, nodding, and so on.

Dr. Otsuka stresses, "We have great expectations that this system will be able to convey more clearly the atmosphere of a place. Participants will be able to get a sense of who is focusing on whom and share empathy over a topic."

Currently, researchers are enhancing functions to enable more detailed transmission of nonverbal information in conversations. An example is using actuators to move a screen forward/backwards and left/right, so that it can appear, for example, that the participant is leaning forward in anticipation to hear about a topic that has drawn his or her interest.

Associate Professor Furuyama also sees great promise in the MM-Space system. "I think we can get closer to the essence of communication by analyzing a variety of information that is exchanged by many people in a conversation," he says. "An objective of the Ido-Robo Project going forward is to apply knowledge gained from designing MM-Space."

The *idobata kaigi* of the future is beginning to take shape with the evolution of videoconferencing systems. Their advances are spurring researchers to dream big.

(Written by Hideki Ito)

The Ido-Robo Project and the Todai Robot Project: Different Approaches to Natural Language Processing

"Can a Robot Get into the University of Tokyo?" (a.k.a. the "Todai Robot Project") and "Can a Robot Join an *Idobata Kaigi* ("Congregation at the Well")?" (a.k.a. the "Ido-Robo Project") are the themes of Grand Challenges underway at NII. While both projects belong to the field of AI, their research strategies seem diametrically opposed. Professor Yasuharu Den researches natural conversations, in which "modalities" such as speech, gestures, and facial expressions are integrated. Associate Professor Yusuke Miyao creates AI systems that can understand logical and correct propositions described by texts. Both researchers share a background in research of natural language processing. Where do they differ and where do they share commonalities?

Creating "Corpora" Is the Foundation of Natural Language Processing

Miyao I'm involved in the Todai Robot Project. To understand the University of Tokyo ("Todai")'s entrance exam questions, the robot we're creating utilizes "corpora," which are digitized resources of languages. Corpora became greatly enriched from the 1980s, and fields like machine translation have expanded greatly with them as the foundation. Professor Den, you have been researching corpora from the standpoint of linguistics. What role do they play in linguistics?

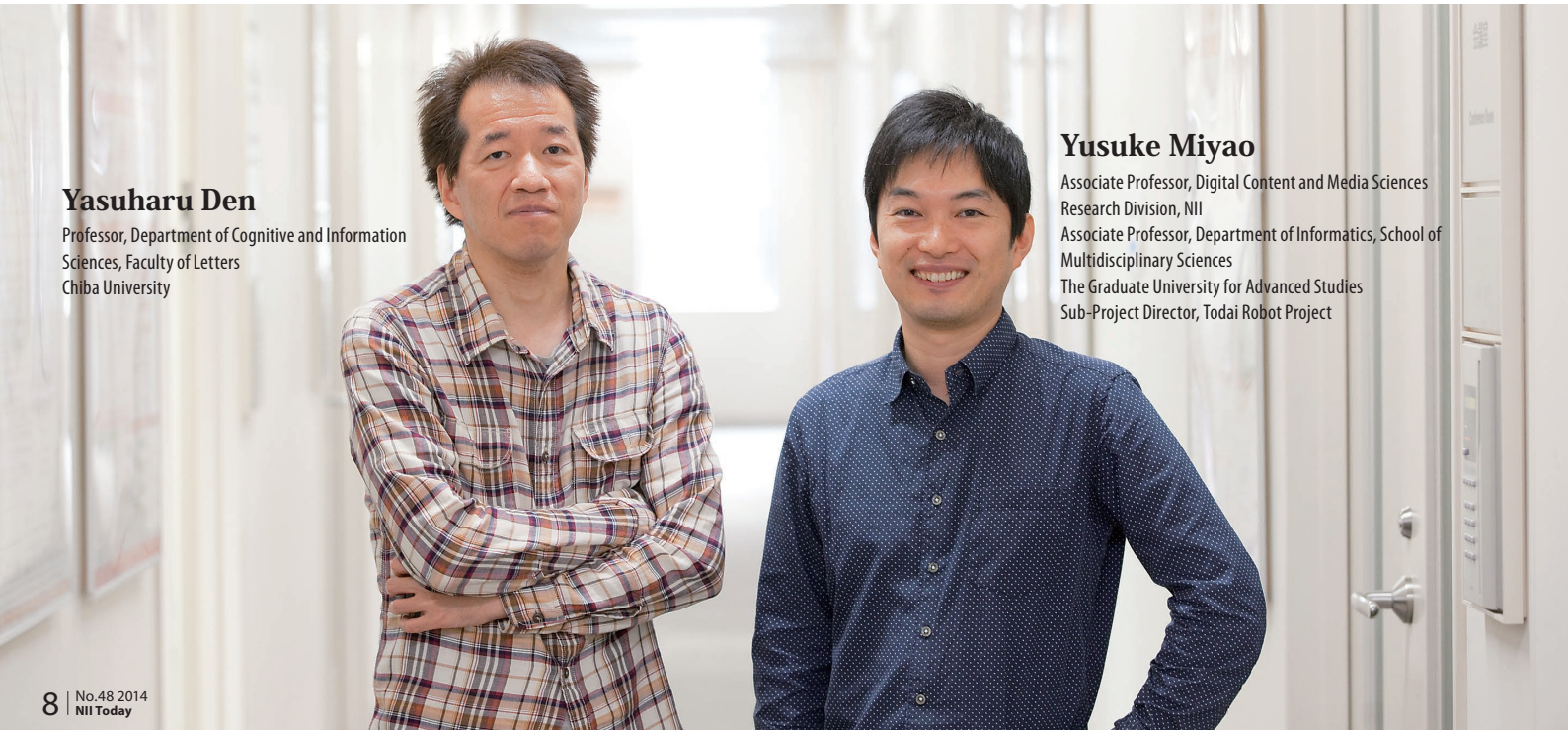
Den This may be a minor area of research. However, apart from linguistics research, corpora are being applied to a variety of systems, including machine translation software and automated answering systems. In the case of machine translation, before the 1980s, researchers tried to systematize methods they believed humans were using. But this approach was actually not very useful. Collecting language materials used in the real world, like newspaper articles and transcripts of conversations, into corpora has greatly advanced machine translation to a practical level. "Morphological analysis" is a fundamental technique in corpus linguistics. It breaks down a natural language sentence into its most basic units and distinguishes parts of speech. The accuracy of this technique has already reached 98.5 percent for newspaper articles and 95 percent for blogs. This is all thanks to actual examples of language use being collected in corpora.

Miyao Corpora are essential for natural language processing. The goal of the Todai Robot Project is to pass Todai's entrance exam by 2021. Actually, however, our research is on "thought processes." To make the research as simple as possible, we focus on text that can be clearly described. Vague linguistic elements are omitted. Because of this, we can make answering entrance exam questions our goal. This approach is quite different from the Ido-Robo Project's, isn't it?

From Monomodal Corpora to Multimodal Corpora

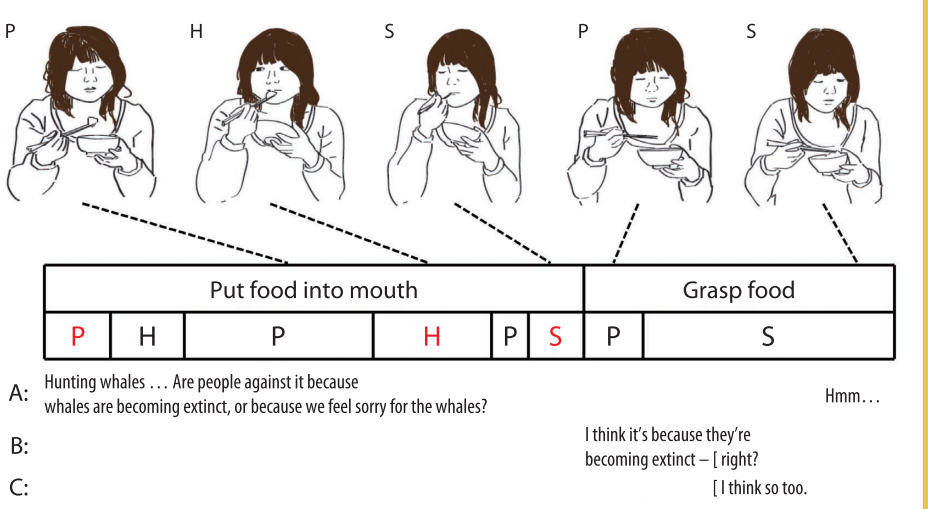
Den The goal of the Ido-Robo Project is to enable machines to enter into real conversations, which include nonverbal elements. Thus, gestures, facial expressions, gaze direction changes, and actions not directly related to the content of the conversation are also being researched. These elements are called "modalities," and a conversation is usually "multimodal."

Corpora of the past only extracted one modality from multimodal conversation. For example, corpora for task-oriented dialogues, such as a "dialogue for route directions," were created in the late 1990s. Before long, corpora of casual conversations and conversations with three or more



Yasuharu Den
Professor, Department of Cognitive and Information Sciences, Faculty of Letters
Chiba University

Yusuke Miyao
Associate Professor, Digital Content and Media Sciences Research Division, NII
Associate Professor, Department of Informatics, School of Multidisciplinary Sciences
The Graduate University for Advanced Studies
Sub-Project Director, Todai Robot Project



Explanation of example

- A does not bring food into her mouth after she finishes speaking. She is beginning to bring her chopsticks up while she is speaking. (P: Preparation)
- A stops bringing up her chopsticks, holding them in midair (H: Hold). She anticipates the timing of when she will finish speaking.
- As soon as A finishes speaking, she puts the food into her mouth (S: Stroke). She finishes chewing during the time others are responding, and immediately begins her next utterance.

→ Holding is connected to the action of quickly eating in preparation for the next utterance.

Illustration produced by Ayami Joh, Ph.D., Project Researcher, NII Bono Lab

partners were created. However, for really understanding conversations, we must collect and analyze conversations as-is in their multimodal state.

We made such an effort at Chiba University in 2003, resulting in the "Three-Party Conversation Corpus." Twelve groups of friends threw dice to determine the initial topic of conversion. They then conversed freely about the topic. Their conversation and non-verbal actions were recorded by headsets with microphones and video cameras; a portion of the collected data was annotated with additional information. However, creating multimodal corpora, including this example, has been limited to gathering people in a lab and recording conversations in a special environment. Conversations in everyday life are quite different from that.

Recording and **Annotating** Table Talk

Den Because of this difference, we next developed a corpus of a daily situation, specifically table talk by three people (video data provided by Professor Naoki Mukawa of Tokyo Denki University). Here, a person uses his or her mouth both to make utterances and to eat, and his or her hands to make gestures and use chopsticks and bowls. We are not usually conscious of how we eat and talk at the same time, and imitating this with AI is an immense challenge. I think multimodal conversation analysis provides new insights into solving this challenge.

Eating actions such as "putting food into mouth" were separated into cycles like "prepare – hold – stroke – retract" and recorded on a time-axis. The content of conversation was simultaneously recorded. We learned a lot of things from this. For example, when you yourself are speaking, you would speak while holding the food held by your chopsticks in the air. After you finish speaking, you would put the food into your mouth about just 0.2 seconds later. This is because you have already begun the action of eating timed to when you anticipate finishing speaking. When another person finishes speaking, the time until you start speaking next is also about 0.2 seconds. This is because you have already caught signs that your conversation partner is finishing up.

Before this analysis, we had hypotheses like "the time periods of speaking and eating alternate" and "the more a person speaks, the shorter his action of putting food into his mouth." However, these hypotheses were not supported by our findings. A person who speaks a lot speaks while temporarily stopping the action of "putting food into mouth," which he had begun before so he can both eat and talk. We have not conducted any analysis of the content of the conversation, but we have observed that if the speaker is talking about a personal experience, others tend to continue their eating actions but not interrupt the speaker.

Miyao That's fascinating. The English section of the National Center Test for University Admissions has fill-in-the-blank questions for a two-person conversation. For humans, these questions are perfectly ordinary, but for machines they are quite difficult to solve. I think this is because humans naturally understand the flow of a conversation without being aware of it. Machines, however, lack this common sense. Data that can teach this part of a conversation to machines would help our research a lot.

Even More Diverse **Situations** Found in Preparations for a Fire Festival Are Being Studied

Den At this stage of research, we want a corpus of an *idobata kaigi* for the Ido-Robo Project. However, we can't record such a meeting in real life due to tough restrictions on privacy. So, last year, we visited an onsite preparation for a fire festival (a Shinto festival for a guardian deity) and recorded the actions of teams called "san-ya-kou" in charge of the festival. This festival takes place every year in January in the town of Nozawa Onsen (Nagano Prefecture). San-ya-kou are composed of three groups separated by age: supervisors, principals, and apprentices. Many people cooperate together while divided in labor to carry out tasks like cutting sacred wood and building a shrine. Traditional knowledge and skills are passed down by showing how something is done besides using words.

The preparations can be considered to be typical scenes of goal-oriented interaction. Oftentimes a person would judge his or her own actions while monitoring others. Everyday conversation would also develop. I think the data we obtained has much in common with an *idobata kaigi* involving many people.

Miyao We are researching how humans think with the Todai Robot Project. So we can't immediately apply your research. But as we proceed with our work, I'm sure that before long we'll run into something we lack. At that time, I'm sure your findings will be useful. Twenty or thirty years down the road, our research may finally become integrated. I hope that you will break new ground in corpus linguistics, and systematize data that informatics researchers can utilize.

(Written by Masahiro Doi)

Creating Robots with the Ability to Understand the Role Played by Body and Hand Movements in Conversations

Assistant Professor Shogo Okada of the Tokyo Institute of Technology has been using machine learning in his research to analyze gestures in people's conversations. The objective of the Ido-Robo Project is to analyze the role of gestures in everyday conversation—for example, at an *idobata kaigi* ("congregation at the well")—by incorporating knowledge from communication science, Assistant Professor Okada's field of expertise. Professor Seiji Yamada of NII is a researcher of AI, HAI (human-agent interaction), and ISS (intelligent interactive systems). He comments on the engineering significance of this project from an outsider's perspective.

Utilizing the Knowledge of Communication Science for Research with Few Precedents

The Ido-Robo Project seeks to understand multimodal (i.e. use of many methods) interaction (conversation), which includes not only voice but gestures (body movements and body language). The gestures of other people besides the speaker are important for understanding conversations. Thus, researchers are including not only words but also gestures when recording and analyzing people's conversations, and seeking to connect the knowledge gained to the design of conversational robots.

Assistant Professor Shogo Okada of the Tokyo Institute of Technology is involved in the research of gesture analysis. From early on in his research until now, he has continued to use the methods of machine learning to analyze computer data representing human gestures.

A portion of the science of gesture recognition is already being commercialized.

Assistant Professor Okada says: "Video game controllers, like Nintendo's Wii Remote and Microsoft's Xbox Kinect, are using gestures for gameplay. For example, in a fishing game, if you make a gesture as if you're shaking a fishing rod, you make your fishing rod in the game shake in the same way. Because the protocol for control is established, many examples of research studies and commercialized products using gesture control have been produced."

On the other hand, there have been few studies of gestures that are a part of daily conversation, which are harder to study.

"Of the nonverbal language expressed in daily conversation," Assistant Professor Okada says, "a lot of research has been carried out on the effectiveness of 'gaze movement' and so on in identifying the conversation partner and the object of interest. However, gestures such as hand movements are 'noisy,' so it is difficult to attach meaning to them. For example, during a conversation, suppose you can't find the right word. So you move your hands around as you grasp for it. This gesture is a type of nonverbal communication. It would be useful to recognize meaningful information from such information, which has been discarded as noise."

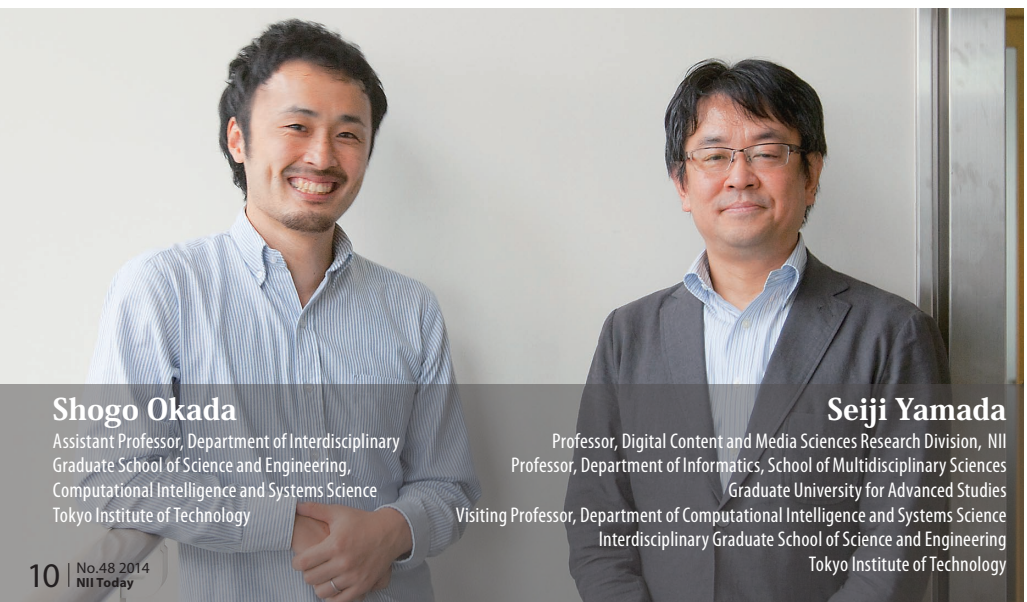
To accomplish this, the Ido-Robo Project seeks to incorporate knowledge from communication science, such as the knowledge that gestures co-occur with spoken words, to understand conversation at a deeper level.

Machines Are Successful in Recognizing Gestures Made during Meetings

An achievement of Assistant Professor Okada's research is the discovery of gesture patterns performed during everyday conversation. For example, suppose person A is conversing with person B. Here, data that include a series of conversation and gestures, such as "A is speaking," "A is looking at B," and "B is nodding," are called "multimodal time-series data."

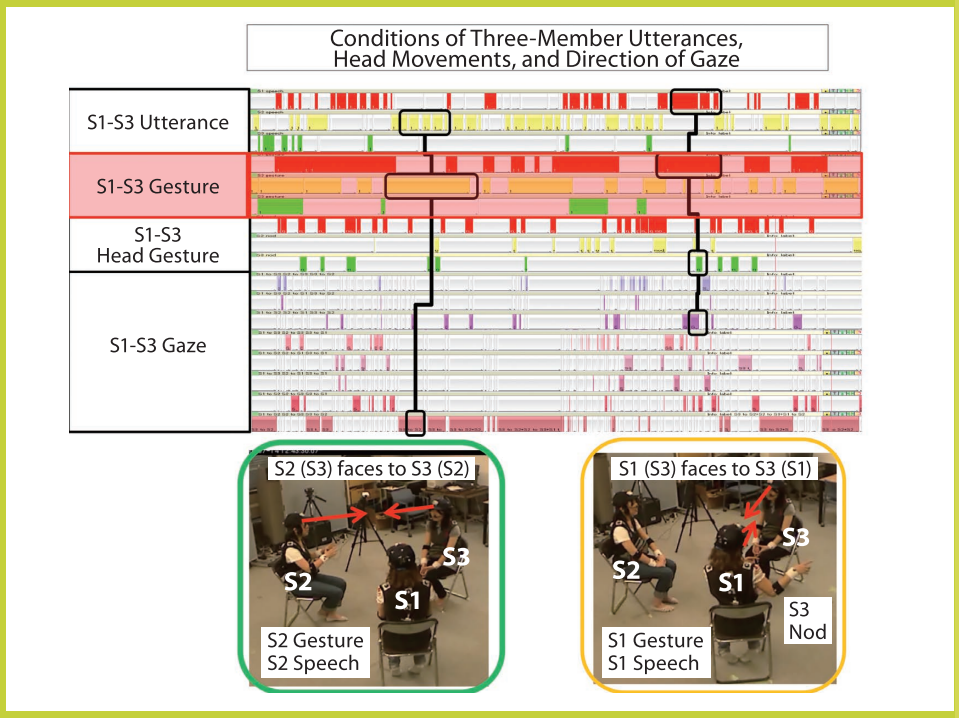
Assistant Professor Okada hypothesized that this multimodal time-series data contain meaning in the time-sequence of gestures that occur as the conversation progresses. Thus, he used multiple methods, such as microphones and motion capture of human movements, to detect the gestures—in other words, the nonverbal behavior—of people in meetings. Data on body movements, such as the direction a conversation partner is facing (the direction of gaze), the turning of the head, and the pattern of hand motion, are gathered. Using a machine learning approach called "unsupervised learning,"* patterns from the collected data are automatically extracted by computers.

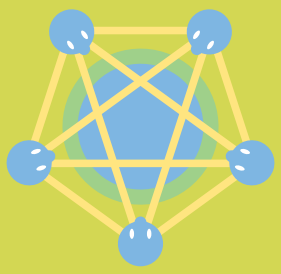
This resulted in the detection of patterns such as: "Gestures used when the speaker is speaking occur with the listener's



Shogo Okada
Assistant Professor, Department of Interdisciplinary Graduate School of Science and Engineering, Computational Intelligence and Systems Science
Tokyo Institute of Technology

Seiji Yamada
Professor, Digital Content and Media Sciences Research Division, NII
Professor, Department of Informatics, School of Multidisciplinary Sciences
Graduate University for Advanced Studies
Visiting Professor, Department of Computational Intelligence and Systems Science
Interdisciplinary Graduate School of Science and Engineering
Tokyo Institute of Technology





To recognize the roles and functions of gestures used in a conversation, it is effective to utilize the speakers' nonverbal behavior (utterances, head movements, direction of gaze). As shown by the two examples below, listeners are looking and nodding at the speaker, who is making gestures complementing an explanation. By observing nonverbal behavior that co-occurs with gestures, it is possible to improve the accuracy of gesture role recognition.

gaze," "Gestures during an explanation and nodding co-occur easily," and "Explanations using gestures take place before and after a listener's question."

We may feel that these patterns are a matter of course; however, it is a major step forward in AI research that they can be recognized by a computer. Professor Okada's achievement shows that machine learning methods can be used to understand more deeply great amounts of human conversation. The knowledge gained can also be applied to the design of robots.

Many Challenging Engineering Issues **Still Remain**

Meanwhile, Professor Seiji Yamada, who has been studying conversations between robots and humans for a long time, comments as an external observer of the Ido-Robo Project: "The Ido-Robo Project is conducting research on conversations without creating an actual robot. However, I want the researchers to build one. If a robot, a man-made object, can advance to the stage of joining old ladies in chitchat, that would be really wonderful."

In reality, even if a robot that resembles a human being perfectly is created, it would still have difficulty joining a human partner in conversation. This is because the actions that a robot can make are clearly different from the natural actions of humans. Thus, humans look at a robot's actions and feel that they are clearly unnatural. Professor Yamada observes, "It should be deeply interesting to research what actions and gestures performed by something with an appearance different from humans would humans find friendly." For example, the robot that appears in the animated film Castle in the Sky (director Hayao Miyazaki, 1986) differs from humans in its external appearance. However, the film convincingly depicts its ability to establish communication with humans to a certain extent with its hand and foot movements and blinking lights. Professor Yamada says that this suggests an avenue of exploration for robotics research.

Also, processing of multimodal time-series data by machine learning, which was carried out in the research described above, is difficult in the first place.

Professor Yamada says, "I anticipate Assistant Professor Okada's efforts will contribute to developing the essence of machine learning, such as knowing what feature quantity to focus on and how to create learning algorithms."

There Is Significance in Accumulating **Academic Research Experience**

In response to Professor Yamada's observations, Assistant Professor Okada comments on the gap between the Ido-Robo Project and real robots: "We are social creatures, so we understand nonverbal communication. You know, for example, that you can start speaking to a person if you receive his or her gaze. However, such a simple thing is difficult for a robot. We're not talking about reading people's mood or sensing the atmosphere in a room. Yet, it is hard to make machines understand the roles and social character of people gathered in a place."

Thus, the Ido-Robo Project is researching themes that are difficult for robots by incorporating not only knowledge from the "hard sciences" like engineering and computer science, but also knowledge from the "soft sciences" like communication science. For example, when using machine learning that assimilates knowledge from communication science to determine whether a gesture made by a speaker during a meeting pertains to explanations, the accuracy can be improved 2 to 16 percent by adding information about the speaker's movements (nonverbal information) besides the gesture in question.

If understanding gestures is advanced by computer processing as described above, innovations can also be brought to communication science's research methodologies. Greater amounts of knowledge can be mined by using machine learning to automate the task of analyzing behavior, which traditionally is performed by humans examining video clips.

Assistant Professor Okada says, "We will achieve new things thanks to interdisciplinary efforts." Professor Yamada encourages him, saying, "I wish for knowledge that isn't simply pure scientific knowledge, but knowledge from an engineering approach that can be applied." How research unfolds from here on is sure to be exciting.

(Written by Akio Hoshi)

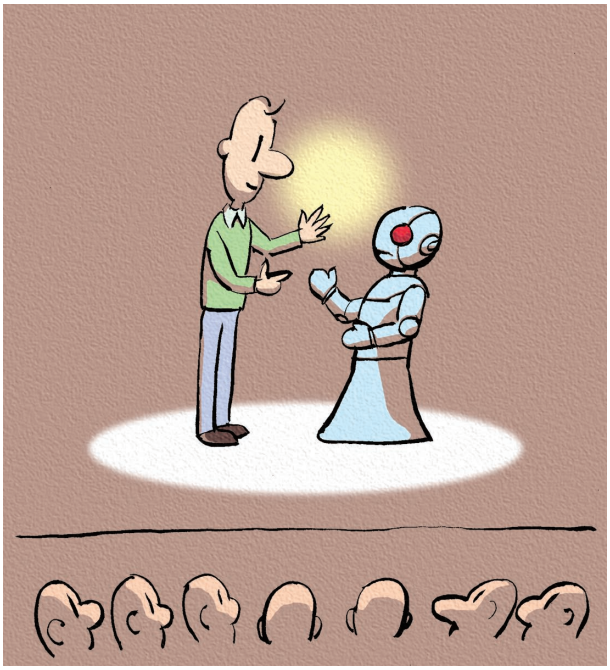
*** Unsupervised learning**
A machine learning technique for processing problems to which answers are not known. It is used to extract essential structures hidden in large volumes of input data.

NII Essay

How Robot-Human Theater Began

Oriza Hirata

Playwright, Director
Professor, Center for the Study of Communication-Design, Osaka University



It has already been six years since the Robot-Human Theater project with Professor Hiroshi Ishiguro of Osaka University began. Performances have been carried out in 33 cities in 15 countries (including 12 cities in Japan) to tremendous response. A Grants-in-Aid for Scientific Research project that produces visible results in such a short time is probably unusual.

About a year after I moved to Osaka University, I ran into the university's then president, Kiyokazu Washida, in the waiting room for a university PR event, and we chitchatted. I had been gradually drawn into the university's Communication Design Center, which began under the direction of Professor Washida. And I had given several lectures on theater there.

Professor Washida asked me, "Is there anything else you want to do?"

I shared with him an idea that had been brewing in my mind for quite some time. "I want to use robots to create theater."

Professor Washida immediately contacted Professor Ishiguro and Professor Minoru Asada, then Professor Ishiguro's superior. I still remember distinctly the day I first visited Professor Asada's lab, a week later. I first asked him the following.

"If I join the project, I can present the robots as being more advanced than they really are right now. Is it okay to do this?"

I was asking for something taboo in the academic world. If you do something like this at a conference presentation, you will be absolutely criticized for fabricating results. However, the first words from Professor Asada's mouth were: "That's exactly what I want." I received Professor Asada's endorsement and began my project with Professor Ishiguro. The two of us are not only of the same generation, but the extent to which we share the same view of humanity and the same thinking about communication is astounding. Our research accelerated at full throttle from the start. Professor Ishiguro aspired to be a painter in his youth, and I have written many plays with science and technology as the theme, so our mutual histories were a reason why the project proceeded so smoothly.

Only Professor Ishiguro had the concepts for not just what kinds of robots to create, but for how to present them and how to stage the production. Thus I assumed a new position at Osaka University, and helped realize this project. Our achievements are evidence that Japanese universities still have some value as universities.

I have discussed the history of this robot-human theater project around the world. The upshot of my story is that the greatest contributor to this project was a philosopher, Professor Kiyokazu Washida, who brought Professor Ishiguro and me together.

Weaving Information into Knowledge

NII

Cover Illustration

The communal well was a source of water for daily life in the past. But not only that. People gathered at the well to chitchat and trade gossip while drawing water and washing clothes. So the well was also an important venue for information exchange that supported the community. Can a robot ever join an *idobata kaigi* ("congregation at the well")? This may be as difficult as humans conversing with cats, but grand plans are afoot to realize this dream.

NII Today **No.48, Jan.2014** [This English language edition NII Today corresponds to No. 62 of the Japanese edition]

Published by National Institute of Informatics, Research Organization of Information and Systems http://www.nii.ac.jp/
Address: National Center of Sciences 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430
Chief editor: Yoh'ichi Tohkura **Cover illustration:** Makoto Komori **Photography:** Seiya Kawamoto / Yusuke Sato **Copy Editor:** Madoka Tainaka **Production:** Kurdyla and Associates Co., Ltd.
Contact: Publicity Team, Planning Division, General Affairs Department Tel: +81-3-4212-2164 Fax: +81-3-4212-2150 E-mail: kouhou@nii.ac.jp