Dialogue AI for Financial Services

Business Improvement through Use of Digital Dialogue Services

Dialogue AI, meaning the use of AI for interactive speech interfaces, has attracted interest along with the rise of smartphones and smart speakers in consumer markets and corporate initiatives such as chatbot and speech recognition services. The aim of dialogue AI is to create new services and improve business processes by converting human speech into digital form. This article looks at the practical business application of dialogue AI in finance institutions, describing the technologies and features that future dialogue AIs will required by considering the difficulties faced with existing chatbots and the approaches being adopted to overcome them.

Takeshi Shirai
Masaaki Yamamoto, Ph.D.
Yu Asano, Ph.D.
Yusuke Fujita
Katsuyuki Tsunami

1. Introduction

Breakthroughs over recent years in areas such as deep learning and improvements in computer performance have raised expectations for the corporate use of artificial intelligence (AI), with the subject featuring in newspapers and magazines. One use of this technology is dialogue AI, which combines speech processing (the conversion of analogue speech data into digital text) with language processing (the interpretation of what the text means) (see **Figure 1**).

Dialogue AI provides a way to improve business efficiency through its use in new types of interface, such as machines that accept verbal commands or frequently-asked-question (FAQ) chatbots that answer questions in place of people. Accordingly, it is seen as

Figure 1 — Overview of Dialogue AI
Speech processing and language processing are used to convert user speech into digital data in a form that a system is able to process.

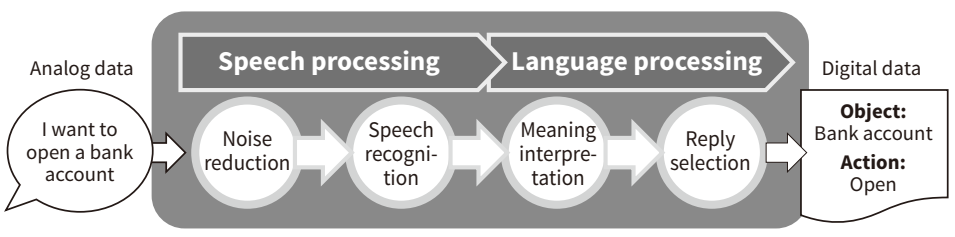
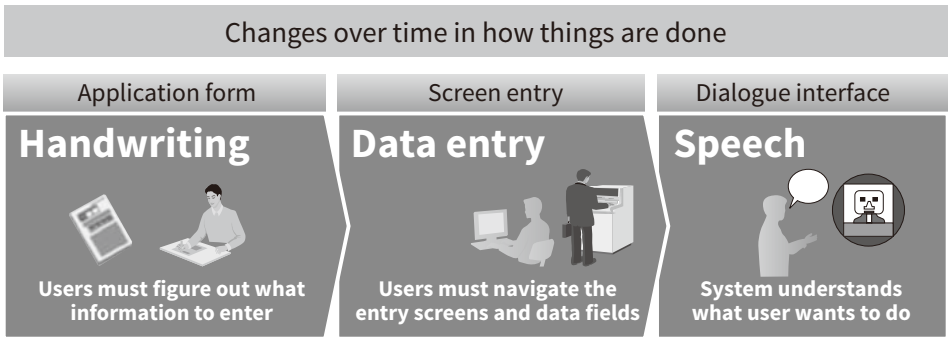


Figure 2 — Interface Evolution

There has been a major shift over recent years in how things are done, moving away from form filling and toward more interactive methods. This change has involved a progressive move from handwriting to mechanical data entry to speech, a shift toward user-centric interfaces that eliminate the difficulties that users have in understanding what is required and in supplying information.



a way of solving problems facing society in Japan, a nation that is dealing with a shrinking workforce due to its aging population and low birthrate.

This article describes the challenges that arise when dialogue AI is utilized in a financial institution and considers how to go about overcoming them.

2. Trends in Dialogue AI

Devices that use dialogue as an interface in applications such as smart speakers, chatbots, communication robots, and voice-operated assistants have become increasingly familiar in recent years, with rising market expectations for this type of interface.

2. 1
Expectations for Dialogue Interfaces

The use of dialogue AI as an interface is seen as having the potential to improve convenience for financial institution customers. Past practice when requesting a service from a financial institution, such as withdrawing money from a bank account, has typically involved filling out a form or using the screens on an automated teller machine (ATM), tasks that the customer is required to understand and complete by themselves. This can be difficult for people who are unfamiliar with such procedures.

The new interfaces are expected to make things easier for customers by using a dialogue AI to assist with the procedure. In other words, the machine is able to converse with the customer to determine what it is they want to do, and then execute the desired procedure for them. The potential benefits include saving the customer from having to find the appropriate form or screen themselves (see **Figure 2**).

2. 2
Spread of Digital Dialogue

These technologies can be put to a wide variety of uses, such as a dialogue AI chatbot that responds to inquiries or the use of speech processing on its own to take the minutes of meetings. **Table 1** lists examples of services already on the market that use chatbots or speech recognition. These are used across a wide range of activities at financial institutions, from front end to back end processes, with the technology seen as being able to deliver benefits that include not only a reduction in workloads but also assistance with making quantitative and qualitative improvements in service.

3. Challenges Associated with Business Applications for Dialogue Systems

One example of a dialogue system already in widespread use is the chatbot. The challenges in this case include poor reply accuracy and the considerable effort required for maintenance. Examples of speech recognition applications include its use for voice-activated appliances and to record the minutes of meetings. The

Table 1 — How Services Based on Chatbots or Speech Recognition are Used in Marketplace

Business services based on chatbots or speech recognition are appearing in diverse applications, with this use expected to grow in the future.

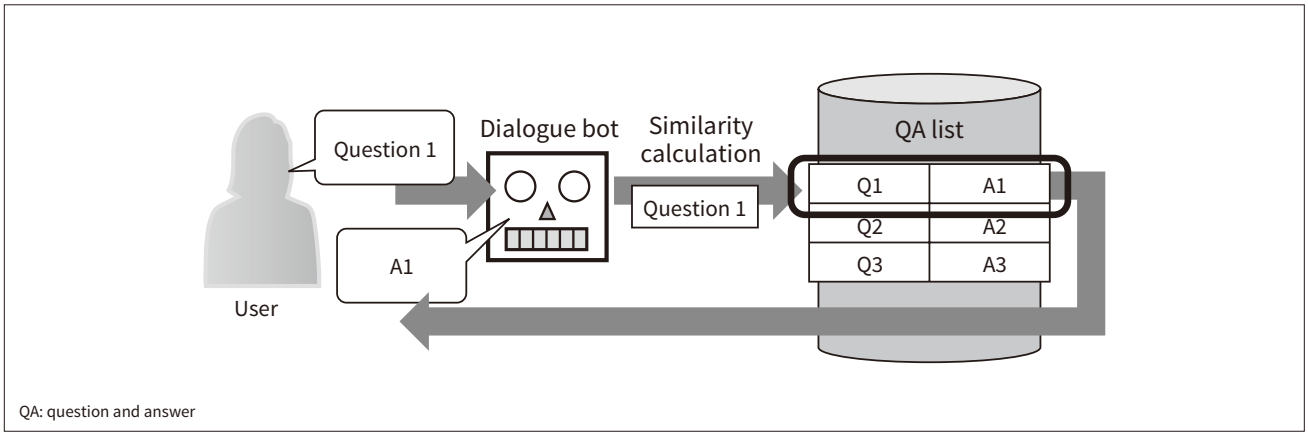
Role in business	Purpose	Example applications	Type		Benefits
			CB	Speech	
Front-end processes (dealing with customers, sales, etc.) ↑ ↓ Back-end processes (in-house processes, manual tasks, etc.)	24-hour operation	Provision of 24-hour telephone support	✓		Quantitative service improvement
		Concierge services		✓	
	Multilingual support	Real-time translation for SNSs or chat	✓		
		Real-time translation for Skype*		✓	
	Omni-channel	Change request handling via SNS or chat	✓		
		Verbal reordering		✓	
	Compliance risk mitigation	Support for monitoring and checks		✓	Qualitative service improvement
	Marketing differentiation	Coordination	✓		
		Verbal ordering instructions		✓	
	Interaction recording	Recording of meeting minutes		✓	
		Conversion of verbal orders to text		✓	
	Satisfaction level improvement	Purchasing via SNS or chat	✓		
		Minimizing cancellations at call center		✓	
	Call center and help desk support	Use of FAQs to screen customer inquiries	✓		Workload reduction
		Automation of in-house IT inquiry handling		✓	
	Hands-free operation	Sales records entry		✓	
		Work records entry		✓	

SNS: social networking service FAQ: frequently asked questions CB: chatbot

* Skype is a registered trademark of the Microsoft group of companies.

Figure 3 — Conventional (Table-lookup) Chatbot

This type of chatbot looks up the answer to each question in a table. Natural language processing can also be incorporated to narrow down the intended meaning of the question.



QA: question and answer

challenges here include the need for the microphone to be located close to the person speaking to achieve acceptable accuracy and that each person has to have their own microphone.

3. 1
Conventional (Table-lookup) Chatbots

Past dialogue systems (developed by Hitachi) have operated on the table-lookup principle whereby a single answer is provided for each user inquiry. This works by calculating the similarity between the actual text of the question from the user and the list of pre-defined questions stored in the table. If a match with high similarity is found, the system outputs the response associated with that question (see **Figure 3**).

Unfortunately, because the same question can be expressed in many different ways, a problem with the old system was an inability to find the pre-defined question that matches the actual question asked. In response, drawing on the idea that the forms in which

Table 2—Alternative Patterns of Expression

Hitachi has enumerated the alternative forms of speech (patterns) that people tend to use when asking about something that has happened. As patterns 6 and 7 in particular cover a wide variety of different cases, a long reply has to be generated to indicate that it is a correct answer. Unfortunately, longer replies take more effort for the user to read and this can result in their failing to understand.

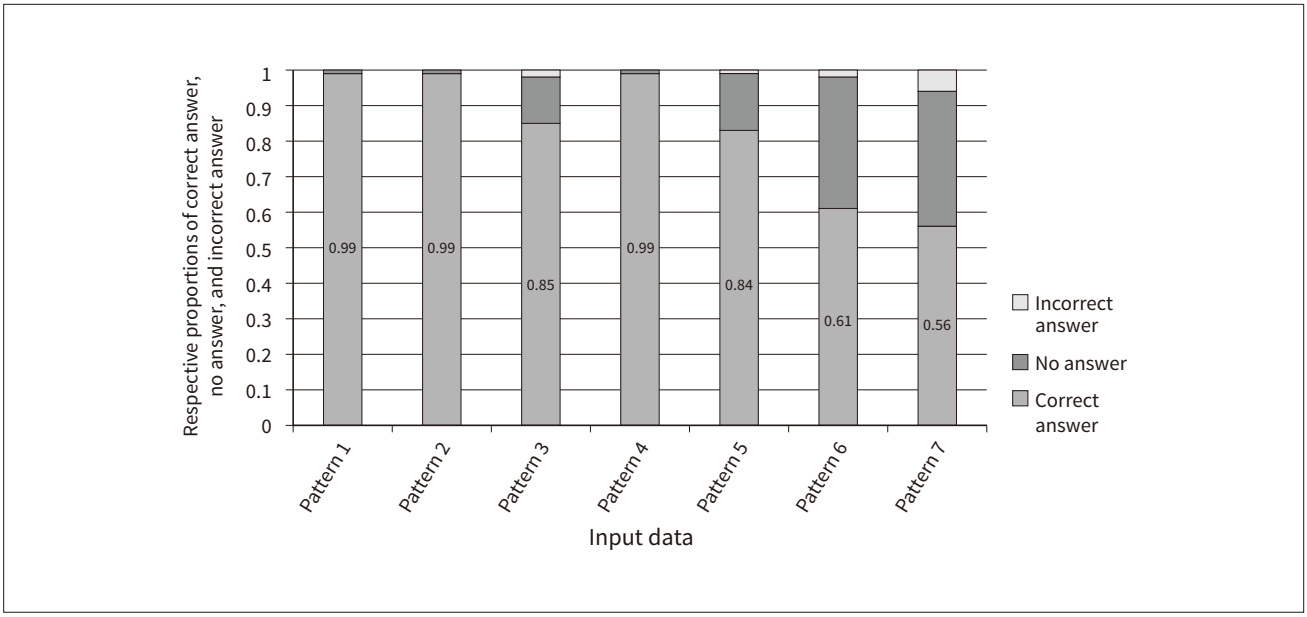
Pattern	Description
Event	Clicking the service menu print button on an operations management device does not work.
1	Use of formal Japanese
2	Use of different particles*
3	Different form of question (use of synonyms)
4	Use of different phrase ordering
5	Use of different vocabulary or syntax
6	Different form of question (such as omission of information assumed from context)
7	Expression of objective and means only

Accuracy is low for patterns 6 and 7

* In Japanese, particles indicate the role of phrases in a sentence and can be used in different ways.

Figure 4—Accuracy for Different Patterns of Expression

The graph shows the accuracy achieved using Hitachi's Natural Language Understanding (NLU) engine. The system is able to return the correct answer for most variations of phrase ordering or particle usage as occur in patterns 1 and 4. While the accuracy for patterns 3 and 5 was lower in cases when the words used were not registered in the system, the correct answer was returned once these words were added. The results also show that accuracy falls dramatically for patterns 6 and 7 in those cases when the questions were asked in different ways or when necessary information was missing.



people express themselves can be broadly categorized into seven patterns, Hitachi assessed the probability of the old system correctly identifying questions expressed in each form (see **Table 2**). The accuracy of the system for each pattern was determined using a data set made up of manually generated questions expressed in each of these forms for 195 FAQ entries (7 patterns × 195 entries) (see **Figure 4**).

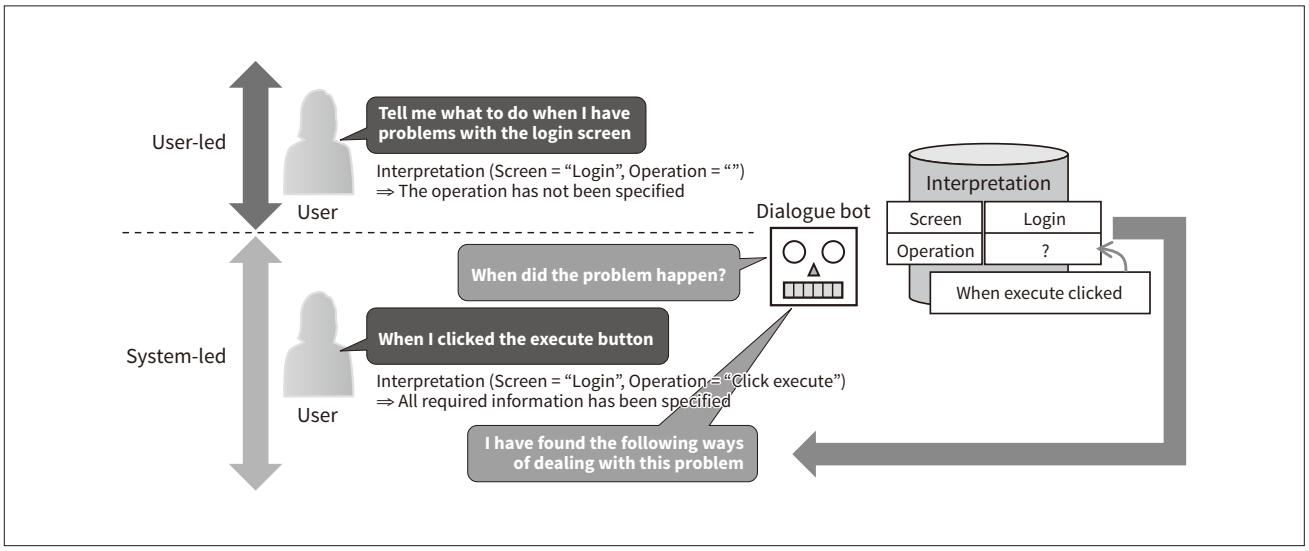
The results indicated that accuracy was especially low (39 to 44%) for patterns 6 and 7 where expected information (such as the activity being performed, device name, screen name, and so on) was omitted from the question.

3. 2
System-led Directed Dialogue

The technique devised to improve performance for the low-accuracy patterns 6 and 7 is called "system-led dialogue." It involves having the system ask clarifying questions so as to build up information on the form taken by the question and what is meant by it. This technique has the system rather than the user lead the dialogue, with the system asking questions of the user to obtain the information needed to narrow down the correct answer. That is, the system takes the initiative, asking questions of the user to obtain this information in cases when there is more than one candidate answer (see **Figure 5**).

Figure 5 — Proposed System-led Clarification Technique

The user will not necessarily ask the question in a way that provides all of the information needed to identify the answer, as is the case with patterns 6 and 7 in Figure 4. In such cases, the dialogue bot identifies the correct answer by first determining what information is missing and then prompting the user to provide it.



The system answers the question once the required information has been obtained and only one matching question remains. A form of machine learning called conditional random fields was used to obtain information from the user. When the reply accuracy was assessed again to determine the efficacy of this technique, it succeeded in reducing the proportion of dialogue failures for patterns 6 and 7 by about 20%.

3. 3

Lower Operating Costs from Using Self-growth

One of the issues with chatbots is the high cost of maintenance. Use of current chatbots to respond to user inquiries requires the collation of a wide variety of different question forms. The problem with this is the difficulty of producing a comprehensive list at the first attempt.

What happens instead is that a human must review the dialogue logs once the system is in operation and add question forms that were not previously included.

In response, Hitachi has adopted "growth dialogue" (sentence form extrapolation) systems that minimize maintenance costs by generating omitted forms automatically (see **Figure 6**). When Hitachi conducted simulations using operator logs from one of its call centers to assess the operational cost savings provided by this method, it achieved a reduction of 60% in the cost of adding each additional question form.

3. 4

Interactive Dialogue Implementation

Incorporating the system-led dialogue and growth dialogue techniques into an existing chatbot succeeded in improving response accuracy while also cutting operating costs, two of the problems faced by chatbots. However, because these changes mean there is no longer a one-to-one correspondence between questions and answers, obtaining the correct answer requires a means of selecting the appropriate form of interaction to use for the question.

Another requirement, meanwhile, is a way to deal with users who want to query any terminology that they do not understand while the dialogue is still in progress.

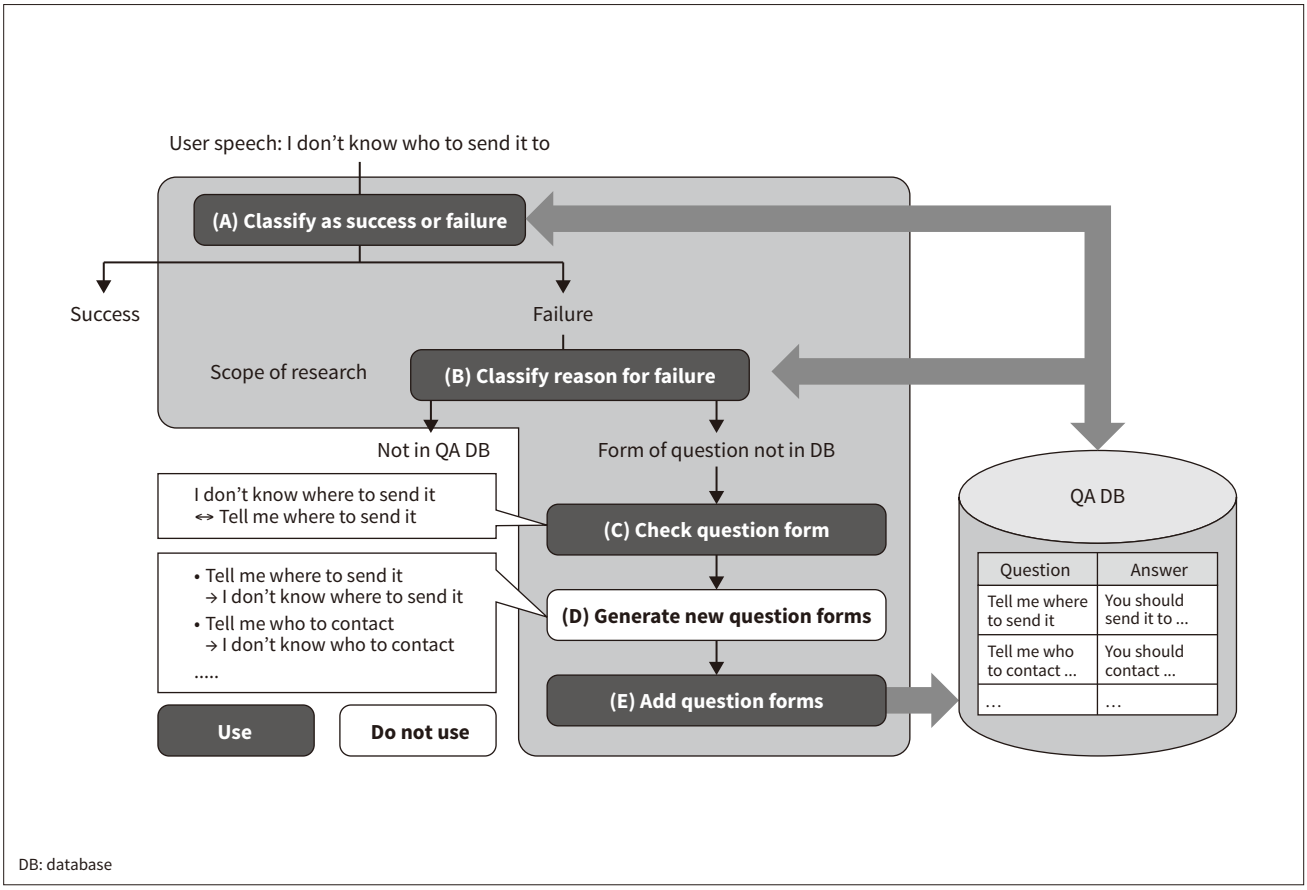
To satisfy these, Hitachi devised a dialogue system that is able to switch responses to suit the user by incorporating a dialogue task switching function (see **Figure 7**).

4. Speech Processing Challenges and Sound Source Separation

Influenced by factors such as working practice reforms and the arrival of smart speakers on the market, people feel comfortable about using speech recognition in routine meetings or when engaging with customers at a retail store, resulting in growing demand for

Figure 6 — Block Diagram of Growth Dialogue System

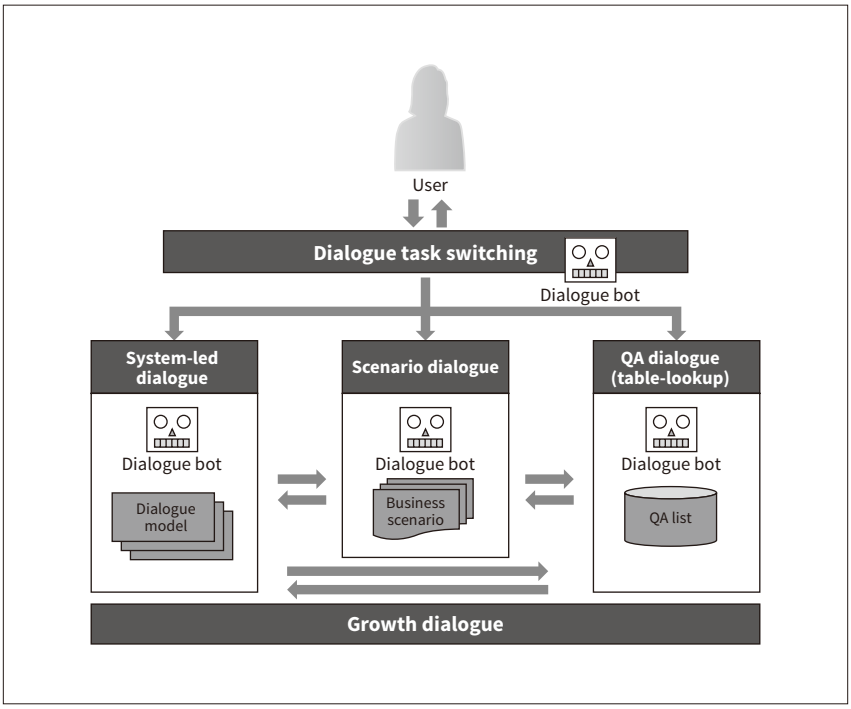
In step (A), user speech contained in dialogue logs is divided into questions for which the correct answer was obtained, and those for which it was not. Next, in step (B), the user speech for which no correct answer was obtained is classified by the reasons for failure. This involves categorizing the failures into those for which the QA DB does not contain the corresponding question and answer and those for which it does contain the answer but not a recognizable form of the question. Step (C) is performed for this latter category and involves checking the question forms so as to determine whether the question as spoken by the user has the same meaning as a question form in the QA DB. Step (D) (the process of generating new question forms) is performed for those cases when the meaning is the same and involves first generating rewording rules from cases when two forms have the same meaning and then applying these rules to all of the question forms in the QA DB to obtain new question forms. Finally, step (E) is to add the new question forms, which means updating the QA DB with the question as spoken by the user and the new question forms obtained in step (D).



DB: database

Figure 7 — Dialogue Task Switching

Dialogue task switching enables the dialogue bot to provide the best answer by directing the user's question to the dialogue model with the best answer based on the question content, without the user being aware of the different dialogue models.



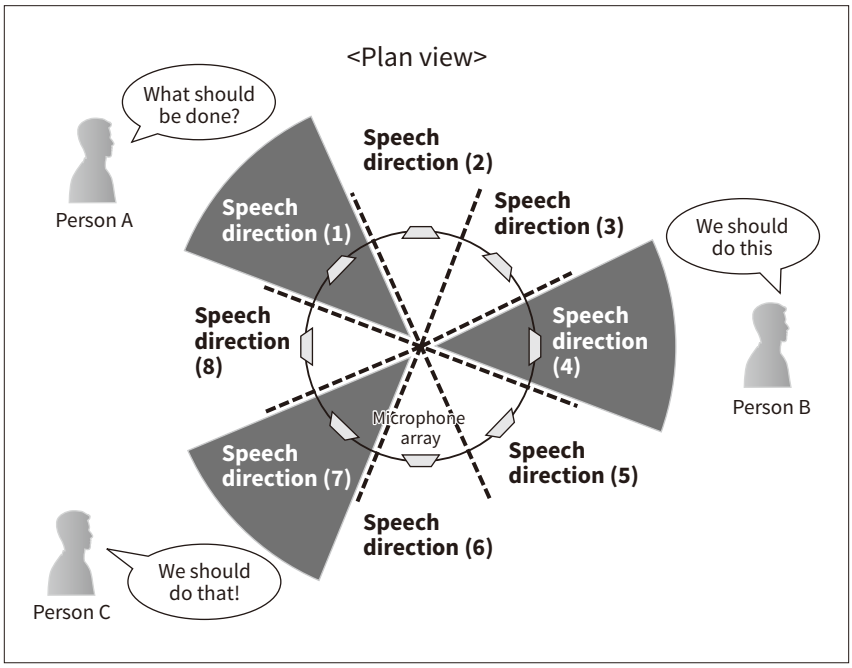


Figure 8 — Speech Recognition Based on Direction from which Person is Speaking

The ability to convert speech to text based on the direction from which the speech is received by a single microphone means that the person speaking can be identified in the text conversion process.

workload reduction and the use of accumulated dialogue records. However, because speech recognition accuracy falls away the further the microphone is from the speaker's mouth, most applications are found at call centers, where audio is captured by a telephone, or at large conferences where gooseneck microphones, for example, are used to provide each speaker with a high-quality microphone of their own.

For a variety of reasons, there are situations in business where it is desirable to record dialogues as text without those involved necessarily being aware of the microphones. Accordingly, Hitachi has looked at using technology that is able to convert the speech of a number of people at a venue into text using only a single microphone (see **Figure 8**).

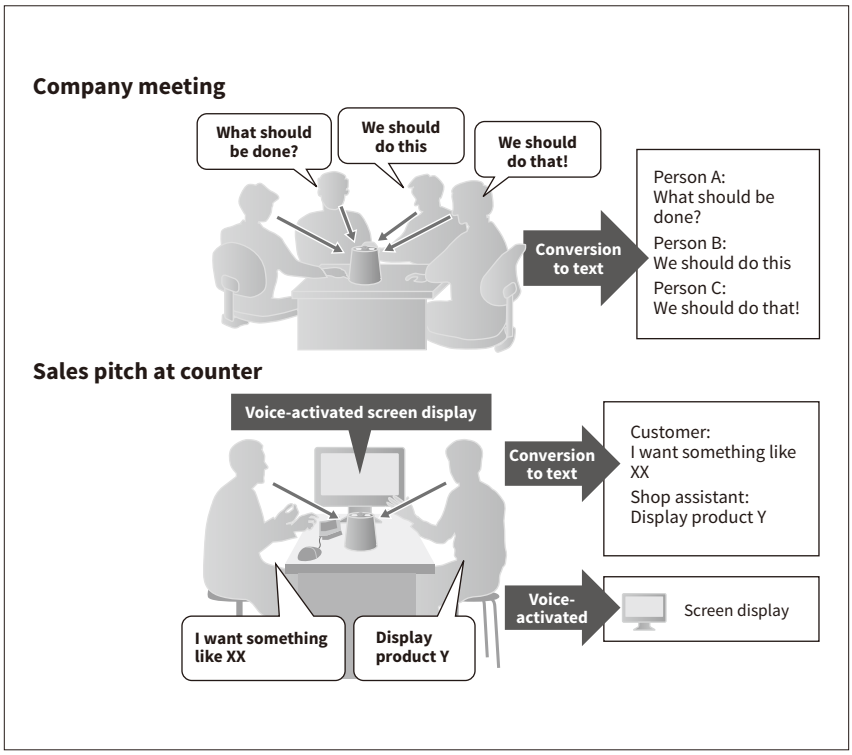


Figure 9 — Expanding Scope of Speech Applications

This helps company meetings to proceed more efficiently by using speech-to-text conversion to record the action points that arise during a meeting, thereby avoiding the need to take written notes. Speech-to-text conversion can also be used to help generate records of dealings with customers in a retail setting.

As Hitachi's proprietary sound source separation technique can use a microphone array to identify different voices based on the direction from which they are speaking, it is possible to separate the speech from a number of different people and convert it to text using a single microphone array.

This means it is possible to record dialogues without those involved being aware of the microphone, extending the scope of possible applications to include situations such as company meetings or talking to a customer at a shop counter (see **Figure 9**). There is also scope for using the technology to improve service by analyzing the collected data for such purposes as identifying customer preferences or performing compliance checks on the sales pitches used.

5. Conclusions

This article has looked at the use of dialogue AI and speech recognition to improve business efficiency.

Dialogue AI has the potential to extend beyond its role as a simple user interface and find uses in various different forms and different business situations. Hitachi intends to continue with this work and contribute to business innovation by customers through digital dialogue.

References

1) Hitachi News Release, "Development of Active-learning Dialogue Data-based AI Technology," (Sep. 2017), http://www.hitachi.com/New/cnews/month/2017/09/170928b.html

2) Y. Asano et al., "Dialogue System that Grows by Asking Unclear Points," Special Interest Group on Spoken Language Understanding and Dialogue Processing 81, pp. 66–71 (Oct. 2017).

Authors

Takeshi Shirai
Financial Innovation Center, Financial Information Systems Sales Management Division, Financial Institutions Business Unit, Hitachi, Ltd. *Current work and research*: Business planning and development for the financial industry.

Masaaki Yamamoto, Ph.D.
Media Intelligent Processing Department, Center for Technology Innovation – Digital Technology, Research & Development Group, Hitachi, Ltd. *Current work and research*: Research on dialogue systems. *Society memberships*: The Institute of Electronics, Information and Communication Engineers (IEICE).

Yu Asano, Ph.D.
Media Intelligent Processing Department, Center for Technology Innovation – Digital Technology, Research & Development Group, Hitachi, Ltd. *Current work and research*: Research and development of dialogue systems. *Society memberships*: The Japanese Society for Artificial Intelligence (JSAI).

Yusuke Fujita
Media Intelligent Processing Department, Center for Technology Innovation – Digital Technology, Research & Development Group, Hitachi, Ltd. *Current work and research*: Research and development speech recognition technology. *Society memberships*: The Acoustical Society of Japan (ASJ) and IEICE.

Katsuyuki Tsunami
IoT & Cloud Services Business Division, Service Platform Business Division Group, Services & Platforms Business Unit, Hitachi, Ltd. *Current work and research*: Software development for digital dialog services.