

# Service Robot Platform Technologies that Enhance Customer Contact Points

● Toru Kamiwada ● Takashi Imai ● Toshikazu Kanaoka ● Riichirou Take

Communication robots capable of talking to users have been put into practical use, and a movement is underway to introduce them into businesses. However, because most communication robots work only in response to user instructions, users have to learn the instruction set of each robot. To overcome this problem, Fujitsu Laboratories developed service robot platform technologies that enable robots to actively talk to users and introduce potentially useful services. With this platform, once many applications have been deployed, a robot chooses one that matches the interests and concerns of the customer and the circumstances of the interaction. The robot senses the customer's reactions and utilizes them to introduce more suitable choices. This platform enables service providers to enhance customer contact points, where customers and services come together, by presenting various services. This paper explains the service robot platform technologies developed by Fujitsu Laboratories that lead a customer to appropriate services by using the customer's preferences and circumstances as obtained through interactions between the user and the robot.

## 1. Introduction

The relationship between people and information and communications technology (ICT) has long utilized the call and response model. For example, to collect information about restaurants serving Japanese food in Shiodome, a user inputs keywords such as "Japanese food Shiodome" in the search box (i.e., the call), and the ICT responds by returning character strings containing related information (i.e., the response). However, because there is a lot of information and many services in the world, some services are not known to the ICT. Moreover, the usage of ICT is limited by the user's knowledge and experience.

However, the latest smartphones can show information such as "There is congestion in the vicinity" and "The temperature here is 10°C" even if the user has not entered anything. Such services can be referred to as "sense and recommendation services" because ICT senses a user's circumstance (the user's location in the example above) and provides appropriate information and a relevant service. However, such services can be problematic. For instance, although it is easy to sense a user's location, it is difficult to know whether a user

has a headache or not. The range of circumstances that can be sensed is limited, as this example shows; therefore, there is no guarantee that the provided information is appropriate for the user. The continued presentation of inappropriate information would certainly annoy the user.

Fujitsu Laboratories has been researching and developing "proactive interaction technologies" that enable ICT to actively sense users through dialogs and, as a result, provide appropriate information and services to users. To sense whether a user has a headache, ICT simply asks the user, "Do you have a headache?" We have also been researching and developing platform technologies for robots with the aim of providing interactive robots that can have such dialogs with users in practice. With this platform, a robot would understand a user's preferences and circumstances by talking actively to the user, leading to the provision of information and services suitable for the user. We aim to develop a sense and recommendation method that can be implemented in such robots.

## 2. Technological trends of communication robots

Communication robots that talk with users have recently become practical, and a movement is now underway to introduce them into businesses. The representative robot of this movement is named Pepper.<sup>1)</sup> Pepper was developed by SoftBank Robotics Corp. and has various means of communication such as voice dialog, arm and finger motion, and touch panel input. It has already been introduced into various commercial applications. The robot's movements and devices are controlled by a local system in its main body, and advanced information processing like voice recognition is done using a cloud service. This use of a cloud for communication robots has been widely adopted and is becoming a fundamental structure. Moreover, a software development kit (SDK) for creating application programs is available to the public, so third parties can develop and distribute application programs for various uses. More than 100 application programs are now available in the application program store for Pepper.

RoBoHoN<sup>2)</sup> of Sharp Corporation is another humanoid robot on the market. Because it was developed as an Android smartphone, it is small enough to carry for mobile use. RoBoHoN also uses a cloud, so the more the function of the cloud service is improved, the more RoBoHoN evolves. An SDK for RoBoHoN is also publicly available, so application programs for RoBoHoN can be freely developed and distributed in the Android application market.

Widely open application ecosystems such as Pepper and RoBoHoN are supported by many developers who accelerate the development and expansion of the system, as was the case with smartphones, resulting in increased usage.

Though a voice interface is expected to be the first user interface for robots, it has already become common in smartphones. Siri,<sup>3)</sup> which answers questions by voice and retrieves information, is a standard application program on the iPhone by Apple Inc. Microsoft Corporation also provides a voice assistant, Cortana,<sup>4)</sup> for PCs as a standard function of Windows. The search application of Google Inc. can also accept speech input, and a search is executed in accordance with the instruction following "OK, Google."<sup>5)</sup> This application has evolved into Google Now, which recommends an application program and service useful for users by utilizing

the history of operations and activities on the user's smartphone.

Google Now aims to provide a sense and recommendation service for the home that is already being used by many smartphone users. However, what can be sensed is limited to within the capabilities of the smartphone. Activities that occur while the smartphone is in the charging cradle cannot be captured, so recommendations will not be appropriate inside the home.

A new movement to make voice commands relevant in the home is rapidly developing. Amazon Echo,<sup>6)</sup> a speaker-type terminal of Amazon.com, Inc., enables the use of Amazon's online shopping and music distribution services in accordance with voice instructions processed by cloud-side recognition software. The engine for this voice command processing is named Alexa,<sup>7)</sup> and third parties can develop functions (skills) for Alexa by using the SDK. Several skills for various services have already been published by several developers.

Google is also offering a similar voice interface terminal named Google Home.<sup>8)</sup> Although it was originally planned to be used for controlling appliances in the home via voice, it has the potential to be applied to a wide range of applications, similar to Alexa. However, the services provided by both Alexa and Google Home are call and response services that work in accordance with the received command. Therefore, the user has to master a command set.

## 3. Realization of communication to connect services and users

A method that actively senses a user's preferences and circumstances through dialog has various advantages over existing methods that sense user circumstances externally using GPS and the smartphone's acceleration sensor.

First, users would be aware that they are providing private information through dialog, thereby avoiding problems such as use of location data without permission. Second, it would be possible to collect information about a user's preferences and physical condition that cannot be measured with sensors. In addition, some users may enjoy the dialog and the information and services thereby provided. Robots can learn user preferences and circumstances from user

reactions when talking about new topics and answering questions each day. The more the robot understands the user, the more suitable the topic that the robot selects, and the more the user becomes interested in talking with the robot. Thus, communication with the robot becomes like net surfing, and the user enjoys the new information provided by the robot based on his/her interests.

However, the difficulty lies in the realization. First, to sense a user's preferences, which are unlimited in variety, what dialog should be designed? Moreover, even if the robot determines a user's preferences, can it determine suitable services for them? Do such services exist?

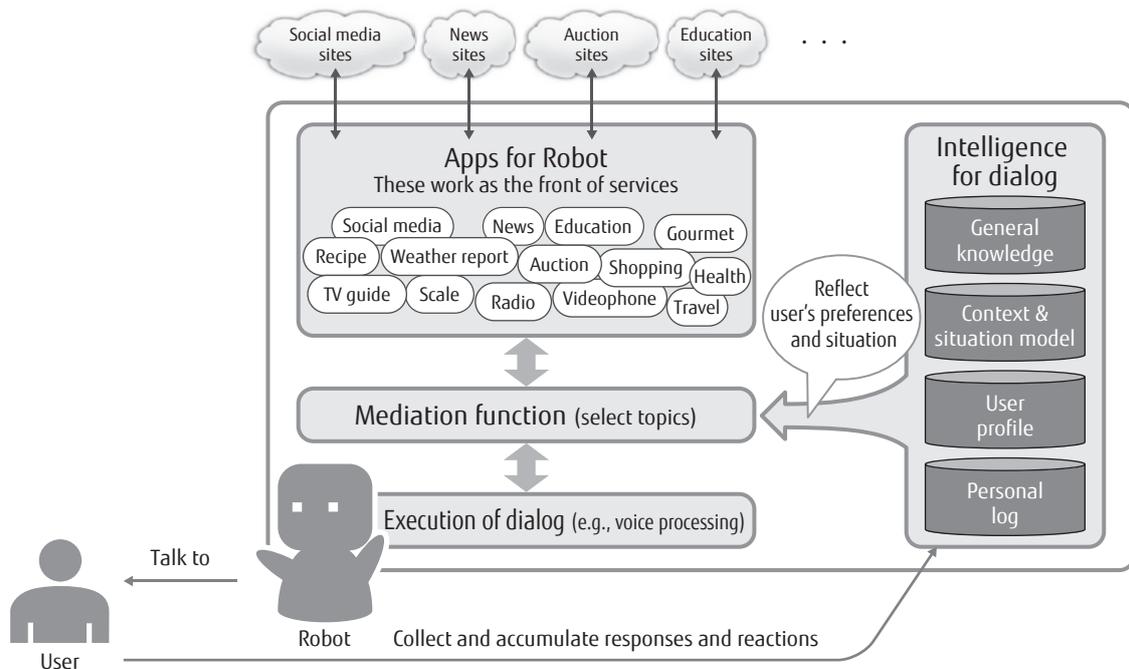
We are dealing with these problems by following the idea "Let information and services talk." The concept of our dialog technology is shown in **Figure 1**. Various application programs for robots are connected with terminal devices and robots through a mediation function. These application programs have the corresponding information and services. For instance, the news application program will connect to the corresponding news site and read out a news article on the site, and the scale application program will talk to the user while he/she stands on a scale.

This technology will support various conversations and questions that maintain a user's interest in the dialog, thus avoiding having the robot search through the unlimited world of human preferences and circumstances. Robot communication based on this method is achieved with the support of innumerable application programs that are proxy ICT services. A dialog system with such a composition has the following problems compared with existing dialog systems.<sup>9)</sup>

- 1) A mechanism is needed that realizes sense and recommendation in conversation by using an application program group.
- 2) Scalability and response are required when the number of application programs increases.
- 3) Development of the robot application program is expensive.

In the next section, we explain the technologies that we are developing to solve these problems.

In addition, we think the appearance of embodied robots is very important to the concept of "Let information and services talk." When information is to be provided to a user through the robot application program, the robot must adjust the time it starts speaking to match the user's receptivity so as to not annoy the user. For example, if the robot wants to say something,



**Figure 1**  
Concept of dialog technology to connect services and users.

it should not suddenly start speaking. Instead, it should begin by imitating human communication behavior, such as actively glancing toward the person or making eye contact with the person. Such behaviors naturally attract a person's attention and elicit reactions corresponding to the robot's action.<sup>10)</sup> An embodied robot that imitates human communication behaviors would facilitate use of the robot's services by naturally attracting a person's attention.

Personal conversation is performed by continuous communication in which a speaker and a listener alternate (generally speaking, "turn-taking"). The participants are satisfied with the conversation if each listener in turn responds appropriately to the speaker's utterances, and conversation is promoted.<sup>11)</sup> Appropriate responses are not simply verbal responses. They also include nonverbal ones such as facial expressions, glance movements, and head nodding and shaking.<sup>12),13)</sup>

Advanced information processing for speech recognition and conversation processing is performed using cloud computing, and its accuracy has improved greatly in recent years. However, delayed Internet access and enormous processing loads can result in communication delays in speech recognition, creating silent periods during turn-taking. Such silent periods can make the user lose interest in the conversation or cause him/her to speak repeatedly, assuming that the robot did not receive their utterance. As we mentioned above, the robot's non-verbal response behaviors show their understanding and promote the conversation; they can also fill in the silent period during turn-taking. In the next section, we explain more about non-verbal behavior technologies.

## 4. Developed technologies

In this chapter, we explain five technologies that we developed. Three technologies are about handling application programs and making conversation. The others are about non-verbal behavior in human robot interaction.

### 4.1 Handling application programs and making conversation

To realize conversation in a sense and recommendation system with various application programs, we developed an execution control technology for

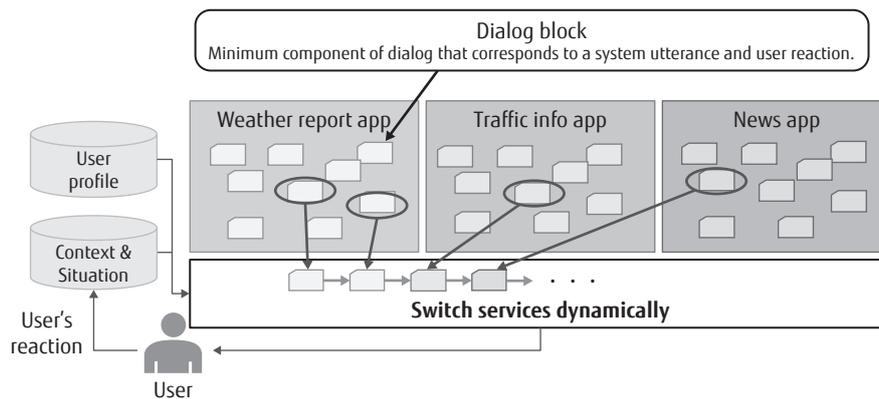
conversation application programs, a dialog service platform technology, and a development support technology for conversation application programs that addresses the three problems described in the previous section. Appropriate services for each user circumstance can be provided by these three technologies and various application programs deployed on the dialog service platform. Each problem and the corresponding technology are described in detail as follows:

#### 1) Execution control technology

A basic idea to realize sense and recommendation in conversation is for the system to actively talk to the user; it then senses the reaction of the user to its utterance and generates an utterance on the basis of the reaction. The reaction to the system utterance includes not only the user's utterance but also the user's movement and expression, if they are captured. In the case that the dialog application is developed with a dialog scenario script, it is difficult to prepare a system utterance pattern corresponding to various user reactions beforehand. Our system solves this problem by preparing various types of information and services instead of preparing system utterance patterns corresponding to various user reactions. It can then select services that match the user's reactions and recommend those services.

To make a conversation while switching between two or more conversation applications, our system divides the structure of the conversation application program into a small execution block and, for each utterance, it selects an execution block across the applications. We consider each pair of system utterance and user reaction during the conversation to be an execution block, which is called a "dialog block." Each conversation application program is composed of several dialog blocks, and the conversation of an application can be multiple dialog responses. If the user's reaction does not match the application's assumption, or a more suitable utterance exists, another dialog block can be executed.

The recommendation of a service with an appropriate timing can be achieved by preparing a dialog block that leads the conversation to the application's topic while sensing the user's reaction. (Figure 2). With such a mechanism, conversation may be unnatural compared with the conversations generated by a well-designed dialog scenario script or conversations



**Figure 2**  
Mechanism of dialog switching apps.

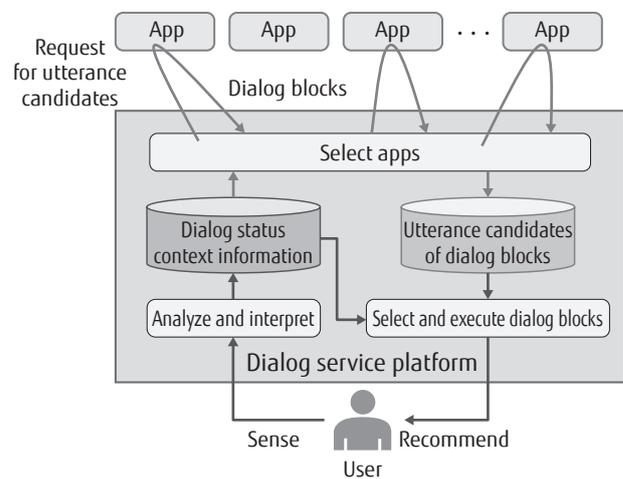
generated by dialog systems created from a large number of conversation examples of a specific situation.

The present system selects dialog blocks by matching words in the dialog blocks with words such as those in the last utterance of the user and in the system and user preferences.

2) Dialog service platform technology

Our sense and recommendation method was mainly described in the previous section. To enable the system to deliver appropriate ICT services, a great number of application programs should be supported by the system. In a previous dialog system, scalability with respect to the number of users was examined.<sup>14)</sup> In our system, we need to consider scalability with respect to the number of application programs; in other words, even if the number of application programs and the number of users increase, the system must be able to continue operation while maintaining a constant response. Therefore, the system is structured as shown in **Figure 3**. We call this system the “dialog service platform.”

The application programs in this system provide dialog contents composed of dialog blocks for requests from the dialog service platform. The dialog service platform acquires dialog blocks from application programs beforehand and maintains them. Moreover, it interprets and analyzes the contents of the user’s utterances, selects dialog blocks from the maintained candidate dialog blocks on the basis of its analysis, and uses them in its utterances to the user. Because the dialog blocks are acquired beforehand, neither large processing cost nor accesses to many application



**Figure 3**  
Structure of dialog service platform.

programs occur, so constant responses can be provided.

The dialog service platform executes the application programs on the basis of the conversation context and the user situation, so the utterance candidate dialog blocks change dynamically. To establish compatibility between securing a response and utterance selection from a great number of application programs, two processes are executed asynchronously: one in which selection of dialog blocks from candidates and utterances are executed with high frequency; and the other in which selection of applications and acquisition of candidate dialog blocks are included and executed with low frequency. The computational complexity for selecting one dialog block from innumerable application programs is normally proportional to the number

of dialog blocks in all applications. In our system, selection candidates are limited by another process, so the computational complexity of the selection is constant.

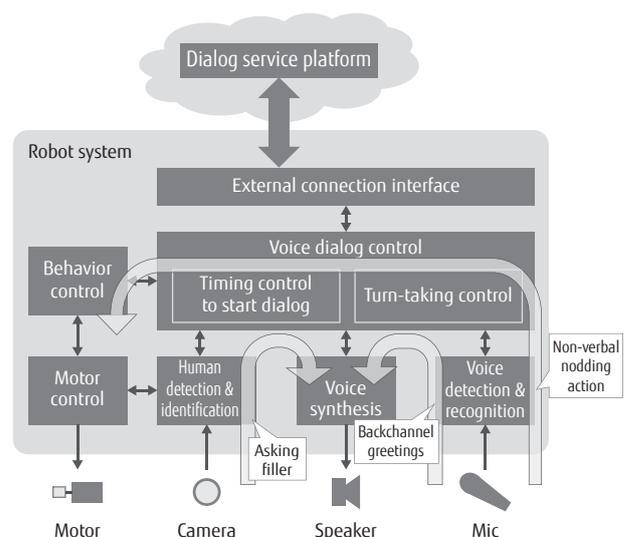
### 3) Development support technology

As mentioned above, the larger the number of application programs, the more effectively our system performs. Users can receive various services provided by the dialog service applications. For providers of ICT services, services are appropriately delivered to users simply by preparing application programs for this platform. However, the cost to develop applications afresh for this platform is high, which is a barrier to service providers. Existing web content and web application programming interfaces (APIs) are thus used to support the development of application programs for this platform, and we are developing a tool that will reduce the cost of developing application programs. The problem is how to convert web contents that have a structure designed for viewing elements to conversation applications that have a structure for conversation. Our application program development support technology supports mapping specific structure patterns in the layout view to the conversation structure pattern. The current prototype tool offers a function to output the contents of the conversation by modifying the structure (document object model: DOM) of the web contents into a conversation template.

Functions such as passing on newly received information by using the RSS (Really Simple Syndication) feed, selecting a list, and reading out a menu are provided in the conversation template. This template will be enhanced to make it easy to generate application programs for this platform by increasing the number of supported layout patterns. When a conversation application program is developed from scratch, the general development process is to design the response to all conversations and repeat the tuning. Our development support technology aims to reduce the time for design and implementation to about 10% that of development from scratch. After the first integration, there are no development costs even if there is a change in the web contents because the application program automatically generates dialog blocks from web contents.

## 4.2 Producing non-verbal behavior in human-robot interaction

Next, we explain the two nonverbal behavior technologies we developed for turn-taking. As a spoken dialog system technology that connects applications and users, we have developed a turn-taking control technology that reduces the length of silent periods and a conversation starting time control technology that determines the conversation starting time on the basis of sensor information. We describe the flow of the conversation here. A block diagram of our robotic system is shown in **Figure 4**. Information is pushed from the conversation service up to the cloud through the external interface between the local device and the conversation service in the cloud and is processed in the voice conversation control part of the service. The voice of the robotic system is synthesized from the speech text, and the speech timing is controlled by the conversation starting time in accordance with the user's situation as obtained by human sensing. The response from the user is input into the spoken dialog controller through the speech detection and speech recognition parts of the service. The conversation turn-taking controller responds to the user to show acceptance of the user's utterance and sends the user's utterance information, which has been translated into text, to the dialog service platform via the external interface.



**Figure 4**  
Block diagram of robot system.

### 1) Conversation starting time control technology

The conversation starting time control technology performs timing control. It determines whether to provide the information received from the dialog service platform. It understands the status of the interaction between the user and the robot by using information obtained from sensors for human detection, speech detection, and so on. On the basis of the interaction status, the robot provides application services along with various nonverbal behaviors that attract the user's attention such as actively glancing toward the user or expressing a desire to talk to the user.

In our technologies, we use the hierarchical architecture<sup>16)</sup> we developed for our "bear cub social robot,"<sup>15)</sup> which was designed to be a friendly robot with life-like appearance and behaviors that attract a user's attention. Evaluation showed that these non-verbal behaviors increase recognition of the robot's intentions by about 60%.<sup>10)</sup>

### 2) Turn-taking control technology

As mentioned above, personal conversation is performed by continuous communication in which a speaker and a listener alternate. The participants are satisfied with the conversation if each listener in turn responds appropriately to the speaker's utterances, and conversation is promoted. Nonverbal behavior helps personal conversation. A person using nonverbal behaviors in their responses promotes good rhythmical communication without disturbing the conversation. Our turn-taking control technology has three features that support good rhythmical communication.

First, the technology detects pauses in the user's speech on the basis of the user's utterances and the speech context and quickly responds with supportive backchannel cues such as "Unn, Unn" (like "Ah-hah" in English). Such quick response not only shows the acceptance of the user's utterance but also reduces the length of the silent periods during turn-taking. In addition, use of this technology in a robot with moving parts enables behaviors to be shown more naturally by combining the cues with nonverbal behaviors such as nods and glances.

Second, if the robot's utterance may be so long as to risk losing the user's interest, the technology can attract the user's attention and provide information<sup>10)</sup> by inserting filler words, such as "anoo" or "etto" (like "Ahh" and "Uhh" in English), at the beginning of the speech.

Third, the technology can anticipate frequent user utterances and quickly respond to them by using local speech recognition. This enables the robot to react to the user's utterances immediately without speech recognition and linguistic processing in the cloud. Such utterances include greetings and short responses like "yes" and "no."

These technologies reduce the essential silent periods between the ending of a user's utterance and the beginning of the robot's response to less than 1 s.

## 5. Application examples

A service robot mainly used for conversation should not only be equipped with technologies such as ones for providing information through utterances and recognizing a user's utterances but also ones for simulating the social skills needed for effective personal conversation. In simulating the social skills, a human-like face and an anthropomorphic body are important components. In addition, a mechanism for producing nonverbal behavior, such as making eye contact and nodding, is an important element to supplement verbal conversation and to naturally attract the user's attention. For this reason, several robots have only a head mechanism.

An anthropomorphic robot can give the impression of being a social entity like a living thing with consciousness. This helps it to smoothly provide services from a cloud application. Additionally, the behavior of the robot during a speech dialog promotes conversation with the user. An example of a conversation using the sense and recommendation method implemented in a robot is shown below. The system utterances are labeled S1 to S8, and the user utterances are labeled U1 to U8.

### 5.1 Conversation Example 1

S1: I heard that a new ramen (noodle) shop opened in Kawasaki. (restaurant retrieval application program)

U1: Really? Give me the details.

S2: The XX shop serves thick ramen. It opened last week. (restaurant retrieval application program)

U2: I'll go some time.

S3: Shall I give you a map? (restaurant retrieval application program)

U3: No, not now.

- S4: Ramen... Aren't you hungry? Let's make some. (recipe retrieval application program)  
U4: OK! I have some pork we can use.  
S5: Shall I retrieve a pork recipe? (recipe retrieval application program)  
U5: Thank you.  
S6: How about green pepper steak? (recipe retrieval application program)  
U6: There is no green pepper in the recipe.  
S7: Shall I retrieve one with green pepper? (recipe retrieval application program)  
U7: No thanks.  
S8: Green pepper is on sale at the YY store today. (YY store application program)  
U8: Let's go buy some.

In this example, utterance candidate dialog blocks are acquired beforehand on the basis of the current history of the conversation and the user's preference information. In S1, a dialog block about a newly opened shop generated by the restaurant retrieval application program is selected on the basis of the user's preference information. In S2, the system provides detailed information in response to the user's utterance, and in S3, the system offers to provide additional information. In S4, because its offer was refused, the system selects a dialog block generated by the recipe retrieval application that has a similar context. The recipe retrieval application program suggests recipes from S4 to S6. In S7, a dialog block related to green peppers is selected in response to U6. The offer in S7 is rejected in U7. Next, in S8, a dialog block generated by the application program of the YY store is selected and uttered. Thus, a sense and recommendation conversation is generated while sensing the user's reactions.

When the user begins to use this system, a trial and error conversation similar to Conversation Example 1 may be repeated. By repeating such conversations, the robot can obtain information about the user's preferences and personal information. The information included in the user's responses will likely become richer, enabling the system to make proposals that go beyond the user's expectations by offering topics related to personal memories and events. An example is shown in Conversation Example 2.

## 5.2 Conversation Example 2

- S1: By the way, you went to see a J League football

game with your family this time last year. (topic offering function of platform)

- U1: Yes, we went to see Kawasaki Frontale play.  
S2: Would you like me to check the Kawasaki Frontale schedule? (J League application program)  
U2: No thanks. My kids did not enjoy it much. They seem to enjoy playing in the park more.  
S3: You went to XX Park last week. (topic offering function of platform)  
U3: That's right. I like to take them to a different park once in a while. Is the YY Park good?  
S4: Yes, and there's a shop selling delicious pasta there. (restaurant retrieval application program)  
U4: OK, I'll spend the day there.

In this example, a topic based on information about the user's past is offered in S1 by the topic offering function of the platform. This is because a topic concerning a specific event can more easily draw the user into a conversation than a general topic. In U1, a specific team name is drawn out in response to S1. In S2, the system offers to provide information about the team schedule. However, in U2, the user responds negatively and utters a new word, "park." The system assumes that "park" means a specific park in the vicinity of the user's home rather than "park" in the general sense; therefore, in S3, the system mentions the specific park it knows to be the user's favorite from past conversations. Then, in S4, because "YY Park" was drawn out from the user in U3, the system is able to introduce a restaurant in the YY district from the restaurant retrieval application program.

This example shows that abundant information reflecting the user's intentions can be drawn out from the user by mentioning not only a general topic but also a personal event. In this way, the system can propose more appropriate services.

## 6. Conclusion

This paper described proactive interaction technologies for service robots, which are platforms for delivering various types of information and services to users, in which the sense and recommendation method is implemented.

Once a service provider prepares application programs to be run on this platform, topics that lead to its service are inserted into the dialog between robots and users, and a chance to create a contact point with a

new customer is provided. For users, a fresh topic that may be suitable for them is offered, and an encounter with a possibly useful new service useful is provided. If offering fresh and suitable topics is of benefit to the user, the platform can obtain various types of preference data, and the service provider is then able to develop its new business utilizing user data. However, whether a robot's appearance is favorable depends on the customer or the trend at that time. Although the appearance of robots is important for them to be accepted by customers, it is possible to offer various appearances depending on the use and the situation. With this platform, service providers can leave the management of data to the platform and focus on adjusting the robot's appearance to make it attractive to the target user group.

We will use these technologies for various applications to enhance customer contact points and to expand business utilizing user data.

## References

- 1) SoftBank Robotics Corp.: Robot.  
<https://www.softbank.jp/en/robot/>
- 2) Sharp Corporation: RoBoHoN.  
<https://robohon.com/global/>
- 3) Apple Inc.: Siri.  
<https://www.apple.com/ios/siri/>
- 4) Microsoft Corporation: What is Cortana?  
<https://support.microsoft.com/en-us/help/17214/windows-10-what-is>
- 5) Google Inc.: Google Search.  
<https://www.google.com/intl/en/search/about/>
- 6) Amazon.com, Inc.: Echo & Echo Dot.  
<https://developer.amazon.com/echo>
- 7) Amazon.com, Inc.: Alexa.  
<https://developer.amazon.com/alexa>
- 8) Google Inc.: Google Home.  
<https://madeby.google.com/home/>
- 9) T. Kawahara: A Brief History of Spoken Dialogue Systems – Evolution and Recent Technical Trend. *Journal of the Japanese Society for Artificial Intelligence*, Vol. 28, No. 1, 2013.
- 10) T. Kanaoka et al.: People-friendly communication robot having human-like attentiveness. *The 26th Annual Conference of the Japanese Society for Artificial Intelligence*, 2012.
- 11) I. Daibo: *Shigusa no Communication - Hito ha Shitashimi wo Dou Tsutaeruka -*. Science Sha, 1998 (in Japanese).
- 12) S. Andrist et al.: Conversational Gaze Aversion for Humanlike Robots. In Proc. HRI '14, 2014.
- 13) A. Terrell et al.: A Regression-based Approach to Modeling Addressee Backchannels. In Proc. SIGDIAL '12, 2012.
- 14) Y. Aragane et al: A Study for a Scalability Evaluation Model of Spoken Dialogue System. *IPSJ Journal*, Vol. 46, No. 9, 2005.
- 15) Fujitsu: People-friendly Terminal “Bear Cub Social Robot” – Interaction Technology for Building Relationship of Affinity with Users. *FUJITSU JOURNAL*, Vol. 36, No. 3, pp. 14–15, 2010 (in Japanese).
- 16) T. Imai et al: Bear cub social robot providing people-friendly interactive services. *Proceedings of the 12th Conference of The Society of Instrument and Control Engineers System Integration Division*, 1C2-2, 2011.



**Toru Kamiwada**

*Fujitsu Ltd.*

Mr. Kamiwada is currently engaged in the research and development of interaction technologies for service robots.



**Takashi Imai**

*Fujitsu Laboratories Ltd.*

Mr. Imai is currently engaged in research on distributed collaborative systems and robot service platforms.



**Toshikazu Kanaoka**

*Fujitsu Ltd.*

Mr. Kanaoka is currently engaged in the research and development of human-robot interaction technologies.



**Riichirou Take**

*Fujitsu Laboratories Ltd.*

Mr. Take is currently engaged in the research, development, and commercialization of human centric computing.