FULL PAPER

Improving user’s sense of participation in robot-driven dialogue

Makoto Kawamoto, Masaki Shuzo and Eisaku Maeda

School of System Design and Technology, Tokyo Denki University, Adachi-ku, Tokyo, Japan

ABSTRACT

In this study, we aimed to enhance the sense of dialogue participation (SDP), which reflects an individual’s conscious engagement in dialogues, and investigated the multimodal dialogue elements that contribute to its improvement. To achieve this, we developed a dialogue system utilizing an android robot specifically for tour spot recommendations at a travel agency. The system was designed with various components aimed at enhancing SDP and was evaluated by 29 participants during the Dialogue Robot Competition 2022. After conceptually defining SDP, we employed the author’s subjective evaluation to visualize the time-series changes in SDP at approximately 20-second intervals, using video recordings of the dialogues. Initially, we expected an overall decrease in SDP, as the robot had a speaking duration of around two minutes during the middle of the dialogue session when presenting the tour plan. However, the analysis of five participants’ data revealed an increase in SDP during segments that involved ‘Dealing with content that increases the user’s interest’ and ‘Multimodal response requests.’ Consequently, we can conclude that these elements have a discernible impact on enhancing SDP.

1. Introduction

In a human-robot symbiotic society, various dialogue systems are being introduced. However, in many cases, the robot takes the initiative in driving the dialogue and choosing the topics. Particularly in task-oriented dialogues, the purpose of the dialogue is predetermined by the developer, making the dialogue inevitably robot-driven. In such robot-driven dialogues, the user’s freedom in terms of speech timing and content is restricted, leading to a decrease in the user’s sense of participation. For instance, users whose speech is limited or feels restricted may experience dissatisfaction or discomfort when interacting with the robot as a dialogue partner. Consequently, the user’s familiarity with and trust in the robot as a dialogue partner decrease, resulting in reduced satisfaction with the dialogue itself.

In this study, we refer to the user’s conscious stance toward dialogue participation as the ‘sense of participation in dialogue (hereafter, SDP).’ We investigated technological elements that enhance the sense of participation in multimodal dialogue with a robot. Additionally, we explored evaluation methods for assessing the SDP.

From a third-party perspective, spatial information between participants is useful for assessing an individual’s engagement in the dialogue. It is well-established that in human communication, the arrangement of participants is influenced by situational cues such as head and body orientation, distance from the speaker, and other factors [1]. Additionally, the form of participation can vary depending on the participants’ relationship, including their roles and social positions. The SDP can manifest from both the speaker’s and the listener’s perspectives.

Quality in dialogue has been explored not only in sociology, linguistics, and psychology, but also in engineering, where researchers have actively investigated objective evaluation methods utilizing sensors and devices. Estimating and visualizing various psychological internal states of individual participants, such as interest [2], concern [3], tension [4], concentration [5], understanding [6], liking [7], and motivation [8], are valuable for enhancing individual communication skills. Moreover, there have been attempts to enhance the communication environment by analyzing externally observable information, such as participants’ speech quantity and the quality of the outcomes generated through discussions, as evaluation metrics in group dialogues. The interdisciplinary field has seen the widespread adoption of research techniques that measure and estimate participants’ internal and external states, leveraging advancements in wearable sensing technology and deep learning.
Against this background, Inoue et al. developed an information processing system that utilizes real-time acquisition of facial expressions, gestures, nods, and eye contact to estimate the level of dialogue engagement for a user listening to a robot’s explanation [9]. However, a common issue with data collection relying on subjective evaluations from users or annotators is its susceptibility to individual differences, requiring further enhancement in accuracy. Moreover, it is important to acknowledge that experiments conducted in large-scale laboratory settings may not fully replicate real-world environments.

Therefore, drawing inspiration from the concept of ‘engagement’ originally used in sociology [10] and the definition of ‘engagement’ provided by Inoue et al. that emphasizes the quality of the connection between dialogue participants, we defined the SDP as the conscious stance of the interlocutors themselves. The main focus of this study is not to conduct a precise quantitative evaluation of the SDP but rather to examine the effectiveness of various dialogue strategies implemented in the robot aiming at improving the SDP. To achieve this, we analyzed data from a five-minute interactive dialogue [11] between an ordinary visitor and an android robot employed at a travel agency for providing tourist information [12, 13]. From the extensive body of previous research, we selected several psychological measures that are considered to contribute to the SDP. These measures were consistently labeled to visualize the time-series changes in the SDP within short time intervals. In this study, as shown in Figure 1, we subjectively evaluated five distinct constructs—interest, concentration, comprehension, liking, and motivation—as a realistic means of evaluation, with the expectation that machine evaluation will be incorporated in future research.

This paper is organized as follows: Section 2 provides an overview of the experimental conditions for an android dialogue system that provides information on tourist spots, while Section 3 introduces the developed system. In Section 4, we discuss five ideas that are aimed at enhancing the SDP. The results of the experiment, including dialogue logs, are presented in Section 5. Section 6 presents a time-series analysis of the evaluation criteria for the SDP, and the findings are further discussed in Section 7.

2. Travel destination recommendation task by android robot

As part of Grant-in-Aid for Scientific Research on Innovative Areas ‘Communicative intelligent systems towards a human-machine symbiotic society’, the first Dialogue Robot Competition was held in 2020-2021 (hereafter, DRC2021). The second Dialogue Robot Competition (DRC2022) followed in 2022. Interactive services using humanoid robots are expected to be easier to use and more hospitable by using various modalities such as robot facial expressions and gestures. In the competition, an android robot (Android I [14, 15]) plays the role of a counter salesperson for a travel agency and interacts with a visitor who wishes to choose one of two pre-selected sightseeing destinations. The task of the android robot is to stimulate the interest of the visitor in the spot designated by DRC organizers through a five-minute dialogue with the visitor. The dialogue condition is a one-on-one interaction between the android robot and the participant (Figure 2). The six tourist attractions are located in the vicinity of the venue.

The effectiveness of a sales promotion task using a dialogue agent has been shown in existing studies [16, 17]. The travel agent task of the DRC2022 is a type of sales promotion task. Therefore, it is expected that most of the participating users will have a certain level of interest in

**Figure 1.** A component of the SDP. Interest, Concentration, Understanding, Liking, and Motivation are evaluated based on the user’s multiple behaviors. The SDP is calculated using these five evaluations.

**Figure 2.** Android I (right) performs dialogue with customer (left). Pictures of two tourist spots chosen by customer are shown on display.
the android robot that they are interacting with. However, further efforts are required to increase the user's SDP.

3. Development of dialogue system

3.1. Overview of dialogue flow

As shown in Figure 3, the flow of the dialogue is to first acquire user information through interviews, and then introduce tourist attractions based on the information. Then, the user is interviewed again about the lunch location, and finally, the user is introduced to the recommended sightseeing spot and lunch. Finally, a sightseeing route including recommended sightseeing spots and lunch is proposed, and the dialogue ends.

The robot's behavior was designed to align with the intended flow of the dialogue. We utilized the basic motion commands provided by DRC organizers, including smiling, bowing, and nodding. However, we recognized that using these commands as they were could potentially create an unsettling impression due to the robot's humanoid appearance, as discussed in the uncanny valley phenomenon.

Therefore, we established our own parameters for the three elements of 'displaying natural smiles,' 'performing polite bows,' and 'adopting an attentive posture while listening' in order to make the robot's behavior more closely resemble that of a genuine salesperson while maintaining a consistent dialogue flow. The following subsections provide details. Additionally, we incorporated parameters that we found effective, such as natural fluctuations and eye tracking, which had been validated during previous competition, DRC2021.

3.2. Design a natural smile for the android

We implemented a natural smile that is independent of the state of the corners of Android I's mouth. While Android I is talking, its corners of the mouth cannot be adjusted because it has a built-in lip-sync function that moves the corners of its mouth in accordance with the content of its speech. Therefore, if the corners of Android I’s mouth were raised to express a smile, it would not be able to maintain a smile in situations where the corners of the mouth are forced down, such as when speaking the word 'o.' Therefore, the corners of Android I’s mouth were raised to express a smile. Therefore, we adjusted the movements of the eyebrows and eyes to convey a smile regardless of the state of the corners of the mouth of Android I. We also adjusted the facial expressions of Androids I and II so that they would smile regardless of the state of the corners of their mouths. Sudden changes in facial expression can make Android I appear inhuman [18]. Therefore, we created and implemented several types of smiles to give Android I complex human-like facial expressions, and combined these facial expressions when changing the facial expressions of Android I to give Android I natural facial changes. The facial expressions are shown in Figure 4.

3.3. Represent politeness by bowing

Bowing, an important gesture for travel agents in Japan and other countries, is represented by the android robot’s upper body movements, gaze, and facial expressions. Bowing increases a person’s attractiveness and gives the impression of being polite and obedient. In our system, the android robot begins bowing after the greeting speech is complete. In order to emphasize politeness, the android robot slowly turns its gaze toward the user after a short pause (about one second) following the raising of
its upper body. By creating a gap between the android robot’s movements and its gaze, we gave the user the impression that the android robot was more polite and human-like. We also made the android robot smile before it finished bowing, so that the change in facial expression would be natural. Figure 5 shows the transition of the upper body, gaze, and facial expression during bowing.

3.4. Understandable listening attentively

To facilitate turn-taking from the android robot to the user, we implemented a behavior in which the android robot always leans forward when asking questions to the user (Figure 6). This visual cue assists the user in determining when to respond. The leaning forward behavior was accomplished by employing the standard command provided by DRC organizers to tilt the robot’s back forward. The robot takes approximately two seconds to return to its original position.

When a robot asks a question solely through voice, the user may have difficulty hearing certain parts of the sentence or grasping the question’s intent. Consequently, users often find themselves confused about when it is their turn to speak. If the robot remains silent after asking a question, the user may suspect a malfunction or encounter trouble. To address this issue, we have incorporated nonverbal elements into the robot’s questioning process. Among the simplest actions, the robot’s forward-leaning posture proves effective in indicating turn-taking. We have implemented this forward-leaning posture using the ‘lean’ function provided by DRC organizer, starting from the completion of the question until the user’s response is received (Figure 6). By observing this distinct posture, users can easily discern when it is their turn to speak.

4. Dialogue strategies to improve user’s SDP

If android robots are used for dialogue services, incorporating multiple modalities can enable the provision of friendlier and more hospitable services. General design guidelines for robots that require hospitality are provided in a survey paper [19]. Recent advances in large-language model technology have led to a flurry of efforts in scenario construction using neural speech generation. In DRC2022, Team LINE [20] won the first place by achieving the most natural speech response using its own language model. Although our system is a rule-based dialogue system and may not be as natural as other systems, we aimed to give a sense of familiarity and hospitality by using our own strategy. In addition to the previous section, five specific strategies that were experimentally introduced to improve the SDP are described.

4.1. Proposal of a reasonable communication goal

In the DRC2022, the participants acted as users of a travel agency. The purpose of the dialogue was taught in advance: ‘While conversing with the android robot, decide on a destination from two choices of sightseeing destinations.’ The android robot’s behavior was designed
to meet this dialogue objective. However, there is a possibility that there will be a discrepancy between the android robot’s assumed dialogue objective and the user’s true intention. In such cases, it will be difficult for the android robot to behave appropriately. As a result, the user may feel discomfort or distrust toward the android robot, and may be reluctant to participate in the dialogue. For this reason, in this system, the android robot confirms the purpose of the dialogue with the participant at the beginning of the dialogue. By having the user and the android robot agree on the purpose of the dialogue, the android robot’s behavior is prevented from diverging from the user’s desires, and the user’s desire to participate in the dialogue is not compromised.

4.2. Dealing with content that increases the user’s interest

When a robot suddenly describes a sightseeing spot that the user has never been to, it is difficult for the user to correctly recognize and remember all the information given. In addition, if the user’s interest in a sightseeing spot is not very high, the user’s interest in the dialogue may decrease if only topics related to the sightseeing spot are discussed, leading to a decrease in the user’s SDP. Therefore, this system incorporates topics related to ‘meals,’ which are likely to be of high interest to users. First, we tried to secure the user’s interest in the dialogue by handling the topic of ‘cafes’ before starting the explanation of sightseeing spots. In the latter half of the dialogue, we considered that the user would become bored with the dialogue by handling the topic of ‘cafes’ before starting the explanation of sightseeing spots. In the latter half of the dialogue, we considered that the user would become bored with the dialogue, and we tried to improve the user’s interest by discussing ‘lunch’.

4.3. Sympathetic behavior in response to the user’s utterances

There have been many studies on dialogue systems for listening comprehension, such as the text-based study [21], the study involving video projection robots [22], and the study using virtual agents [23]. In these studies, researchers found that techniques such as repetition, probing questions, evaluative responses, and lexical
responses were effective in improving dialogue interactions. We used these studies as references when developing our own dialogue system.

The robot’s facial expressions and emotional statements lead to the user’s interest in and liking of the robot. Therefore, in this system, when the robot affirms the user’s opinion or when the user agrees with the robot’s opinion, the robot’s facial expression is ‘smiling’ and emotional robot speech is incorporated, such as ‘This looks like fun.’ This conveys to the user that the robot has favorable feelings toward the user, thereby increasing the user’s interest and liking. As described in the next subsection, the user’s responses to the robot’s suggestions and questions were made by both the user’s gesture and utterance, but the dialogue program continued only with the content of the user’s speech responses.

In addition, when a robot and a user are conversing, situations in which the robot asks the user for his/her choice or opinion frequently occur. In such situations, what is important is the robot’s utterance after listening to the user’s opinion. This is because the user judges whether his/her opinion is properly and favorably received from the robot’s utterance. This leads to a favorable impression of the robot and interest in the dialogue. Therefore, when the robot asks the user to make a choice or give an opinion, the user’s opinion should be affirmed, and the robot should explain why it is affirming the user’s opinion. For example, when we asked the user which restaurant he/she most wanted to go to, the robot said, ‘I was thinking of going to an Italian restaurant,’ and ‘I recommend this restaurant because it is delicious and reasonably priced.’

4.4. Providing choice opportunities

The dialogue domain used in this study is a persuasive dialogue at a travel agency. Therefore, the robot provides explanations to the user in order to interest the user in the recommended sightseeing spots. However, one-sided explanations may cause a decrease in user understanding and motivation. As demonstrated by the previous studies mentioned in the preceding section, it is crucial to
paraphrase and restate the user’s input. Therefore, when explaining tourist attractions, the robot emphasizes the user’s own choice, such as ‘I will introduce the tourist attraction that you said you wanted to visit the most,’ or ‘Why don’t you go to the other tourist attraction that you chose next?’ In this way, the robot indicated that the explanation it was about to give was based on the user’s choice, rather than one-sidedly, and this was intended to motivate the user. The purpose was also to encourage the user to organize the information by reminding the user of his/her choice, and to improve the user’s understanding of the content of the subsequent dialogue. Here, our dialogue system was carefully designed with rule-based programs to avoid dialogue breakdowns in situations where the robot repeats the user’s selections.

4.5. Multimodal response request

This system incorporates speech based on the android robot’s emotions and subjectivity into the recommendation dialogue. The purpose of this is not only to select information from a database and provide it to the user, but also to create a close relationship between the user and the android robot by making the user feel the android robot’s character [24]. However, android robots that talk about themselves in this way may reduce the user’s SDP. Therefore, we asked users interacting with the android robot to engage in multimodal dialogue by combining utterance input with actions such as pointing and raising hands (Figure 7). While it is possible to capture the user’s pointing and hand-raising states through gesture recognition techniques ([25–27]), our system during DRC2022 solely relied on speech information as the input method.

Our goal was to give the user a stronger impression that the dialogue changes according to the user’s own choices by having the user interact in a combination of multiple modalities rather than having the user interact only with voice input. In addition, by having the user perform actions toward the android robot, we made the user’s body turn toward the android robot and the display. In this way, the user, whose attention was focused on something unrelated to the dialogue, was encouraged to concentrate on the dialogue.

5. Interactive experiment

In the preliminary round of DRC2022, we demonstrated our system at the National Museum of Emerging Science and Innovation (Miraikan). A total of 29 participants, males and females, aged 10–50 years, were included in the experiment. After obtaining consent to participate in the experiment, the dialogue was recorded as video data. As shown in the figure (Figure 8), some participants joined with their families or friends, but they were instructed to interact with the robot one-on-one during the experiment. The dialogue proceeded as shown in Figure 8 over a five-minute period. The dialogue log is shown in Table 1 as an example.

Since the experiment was conducted as part of a competition, each user was asked to fill out a questionnaire after the experiment. Users were asked to rate their overall impression of the dialogue on a 7-point scale according to the following 9 items: (1) satisfaction with selection, (2) sufficiency of information, (3) naturalness of dialogue, (4) appropriateness of dialogue, (5) likability of dialogue, (6) dialogue satisfaction, (7) trustworthiness, (8) reliability of information, and (9) utility. In addition, changes in the desire to visit the designated spots were obtained using an analog slider.

![Figure 7. Indication of android intent through subtle behavior. Forward leaning posture shows intention to give utterance turn.](image-url)
Figure 8. A photo of the conversation taken with a 360-degree camera.

The results of the questionnaire survey are shown in Table 2. This table includes the results of systems built by DRC organizers for comparison. Our proposed system outperformed the baseline system in terms of satisfaction with choice (Sat/c), $4.72 \pm 1.93 > 4.19$. It also obtained a higher rating than the baseline system for dialogue favorability (Lik), $5.07 \pm 1.78 > 4.59$. Dialogue satisfaction (Sat/d) was also higher than the baseline at $4.72 \pm 1.96 > 4.15$. Thus, the results suggest that the proposed system was favorably accepted by the participants. Statistical tests are not shown because individual results of other teams are not disclosed by DRC organizers. For reference, the table also shows the simple average of the results of the 12 participating teams. Our team did not rank in the top three but was placed in the second group, above the average of the participating teams, and received an honorable mention award. In the review by DRC organizers, our dialogue strategy of eliciting pointing gestures from the user was highlighted as unique.

6. Evaluation of the SDP

6.1. Evaluation method

The subjective evaluation of the SDP was conducted solely by the author in a collaborative process with assistance from two students from our laboratory. The dialogue videos used in the analysis were five DRC2022 dialogue videos. These five videos were selected from those that had achieved the task of ‘guiding to recommended sightseeing spots,’ which was set in DRC2022, and had a small number of system troubles. The attributes of the five participants are shown in Table 3.

First, from the video recordings of the dialogues, we segmented the points corresponding to the points of contrivance. We prepared 20 segmented videos, each lasting approximately 20 s, for each session. While watching the above videos, we examined the five items (interest, concentration, understanding, liking, and motivation) in turn whether there was an effect or not. The criteria for each item are described below.

6.1.1. Interest

This is an evaluation of the user’s interest in the current topic and the android robot’s speech. The factors we focused on in this evaluation were ‘Utterance,’ ‘Head nodding,’ ‘Facial expression,’ ‘Gaze,’ and ‘Tone of voice.’ First, users who were interested in dialogue showed vocal responses such as ‘yes’ and ‘huh’ even when they were not explicitly given the right to speak. Users who responded to questions from the android robot with sentences rather than simple answers were considered to be highly interested. Similarly, users who were interested in the dialogue tended to make head movements. However, nodding at certain times regardless of the android robot’s utterance, or excessive nodding, is likely to indicate a decrease in interest. Next, we considered that users who did not frown or had other negative facial expressions, such as a furrowed brow, or who had clearly positive facial expressions, such as a smile, were likely to be interested in the dialogue. However, if such expressions were observed when the user was talking with a companion rather than with the android robot, they were considered invalid. Finally, we considered the user to be uninterested in the dialogue if he or she looked at something unrelated to the dialogue (e.g. a companion, behind the user’s back, etc.). However, we assumed that a brief
### Table 1. An example of a dialogue log.

| Speaker | Speech Contents | Elapsed Time |
|---------|-----------------|--------------|
| S       | Welcome.        | T1 (0:01–)   |
| U       | Oh, hi, yes.    |              |
| S       | My name is Ai. Nice to meet you … . |              |
| U       | Oh, it's a pleasure to meet you. |              |
| S       | I am sure that you are asking about which of the two tourist attractions shown on this screen you would like to visit. | T2 (0:14–) |
| U       | Yes, that is correct. Yes, that is correct. That is correct. |              |
| S       | Thank you very much. By the way, which one are you more interested in? | T3 (0:28–) |
| U       | Uh, yes.        |              |
| S       | Could you tell me the name of the tourist attraction while pointing to this screen here? |              |
| U       | Ok, then you are more interested in the Tokyo Trick Art Labyrinth Museum. |              |
| S       | Are you sure it is the Tokyo Trick Art Labyrinth Museum? | T4 (0:47–) |
| U       | Yes, that is correct. |              |
| S       | Thank you much. Tokyo Trick Art Labyrinth Museum. |              |
| S       | The Tokyo Trick Art Labyrinth Museum is a very nice tourist attraction, and I was hoping that you would visit there. | T5 (0:58–) |
| U       | Then, please let us work with you to plan your trip. Do you like to make travel schedules? | T6 (1:07–) |
| S       | Oh, not really, it’s a bit of a hassle, and I’m just about here. |              |
| S       | Actually, there are quite a few customers who find it troublesome to make schedules. I, Ai, am a robot, so I am good at this kind of thing. Today, I would be happy to talk with you and work with you to come up with a travel schedule. | T7 (1:23–) |
| S       | Actually, there are quite a few customers who find it troublesome to make schedules. | T8 (1:40–) |
| U       | Ah, to complement— |              |
| S       | As for the means of transportation for this trip, we recommend that you take the train as a measure against global warming. | T9 (1:57–) |
| U       | Ah, yes, we can take the train. |              |
| S       | Thank you very much. Next, I would like to suggest which specific sightseeing spots to visit in and in what order. I would like to say that the first sightseeing spot we will visit is the Tokyo Trick Art Labyrinth Museum, which you were most interested in, but in fact, Odaiba has some great cafes that you can only find here! | T10 (2:18–) |
| S       | When I travel, I have a cup of tea at a cafe first before going to the sightseeing spots. I would like to start this customer’s travel plan with a nice cafe! | T11 (2:30–) |
| S       | It will definitely be a great trip plan. The cafe I would like to recommend to our guests visiting the Tokyo Trick Art Labyrinth Museum is Island Vintage Coffee. This is one of my favorite cafes and I would like to recommend it to our guests. | T12 (2:49–) |
| S       | After enjoying a relaxing cup of tea here, why not visit the Tokyo Trick Art Labyrinth Museum, which you have been interested in? You can walk from the cafe to the Tokyo Trick Art Labyrinth Museum. | T13 (3:03–) |
| S       | This is a trick art museum where visitors can enjoy hands-on art using optical illusions. The museum is full of various surprises, such as paintings that appear three-dimensional, works that change impressions depending on the angle from which they are viewed, and rooms where visitors can experience optical illusions by entering. The museum is composed of five areas, including the ‘Edo Area’ where visitors can slip back in time to the streets of Edo, and the ‘Masterpiece Area’ where visitors can enjoy numerous trick art masterpieces. | T14 (3:30–) |
| S       | Photography is allowed, so don’t forget to bring your camera. I heard that it is OK to take pictures. This sounds like fun … . | T15 (3:38–) |
| S       | Why don’t we have lunch after this? There are two recommended restaurants near here. I will ask you which one you are interested in, so please listen carefully. | T16 (3:49–) |
| U       | Uh, yes.        |              |
| S       | The first one is a set meal restaurant called ‘Oshokudo Yuki’ and the second one is a steakhouse called ‘Good Meat by Ogawa Dex’. Both restaurants are within walking distance from Tokyo Trick Art Labyrinth Museum. Which one is your favorite restaurant? | T17 (4:06–) |
| S       | If you are interested in the first one, please answer with your right hand, and if you are interested in the second one, please answer while raising your left hand. |              |
| U       | Well, yes, with my left hand, yes, please, at the steak house. | T18 (4:21–) |
| S       | Thank you very much. I will be happy to add the restaurant you have chosen to our schedule. I am sure you will be satisfied with this restaurant because it is a really nice place, very tasty and reasonably priced! | T19 (4:37–) |
| S       | It is almost time to go. Let me introduce our travel plan one last time. First, have a relaxing cup of tea at Island Vintage Coffee, then enjoy the Tokyo Trick Art Labyrinth Museum, followed by a delicious meal. We hope you will have a wonderful holiday with this plan! | T20 (4:58–) |
| S       | Next time you visit us, we would be happy to work with you on a schedule to visit the Tokyo Metropolitan Water Science Museum. The Tokyo Metropolitan Water Science Museum is also an interesting place that introduces the wonders and importance of water from a scientific perspective, and is a hands-on museum that deepens interest in water and waterworks. Thank you very much for visiting us today. |              |
| U       | Thank you.     |              |

Note: In ‘Speaker’ column, S is the system and U is the user. The segments in the ‘Elapsed Time’ column are labeled T1 to T20, separated by approximately 20 s. (Translated version)

---

1.2. Concentration

This is an evaluation of whether or not the user is conscious of the dialogue. The elements we focused on in this evaluation were ‘Utterance,’ ‘Nodding,’ and ‘Gaze.’ First, we judged whether the user was able to respond to the android robot’s questions. This was based on the idea that a user who is concentrating on his/her own activity is listening to the android robot’s utterance and can respond in some way when the android robot asks him/her a question. Next is nodding. As mentioned earlier for the ‘interest’ item, nodding at certain times regardless of the robot’s utterance, or nodding excessively, is likely to indicate a lack of focus on the dialogue. Conversely, if the number of nods is decreasing, it is possible that the android robot is concentrating on its own utterance and not on nodding. Finally, there is eye contact. Users glance at the companion did not affect the fluctuation of interest.
who are focused on the dialogue are likely to gaze at the android robot or the display used for the dialogue. Conversely, if the user is gazing at something unrelated to the dialogue (e.g. a companion, behind the user’s back, etc.), the user is not concentrating on the dialogue. In addition, if the user’s eye movement was too frequent, we judged that the user was not concentrating on the dialogue.

### 6.1.3. Understanding

This is whether the user understands the android robot’s utterance. The factors we focused on in this evaluation were ‘Utterance,’ ‘Head nodding,’ and ‘Facial expressions.’ First, we judged whether or not the robot was able to respond appropriately to the android robot’s questions. This is because users who understand the android robot’s utterances listen to the android robot’s utterances and are likely to be able to respond correctly even when the android robot is asking a question or requesting a special response method. This is because the user hears the android robot’s utterance. If the android robot responded aloud with a response such as ‘I see’ or ‘huh’ after hearing the android robot’s utterance, it was judged to have understood the android robot’s utterance. Next, in terms of head movements, users who nodded in response to the android robot’s utterance were judged to have a high level of understanding, such as nodding during the brief silence that occurs during the android robot’s speech.

Finally, we judged that users who frowned or looked at their companion with a wry smile had a low understanding of the android robot’s facial expressions. This is because the user could not understand the content of the android robot’s utterance.

### 6.1.4. Liking

This measures the user’s positive feelings toward the dialogue. The factors we focused on were ‘Utterance,’ ‘Head nodding,’ ‘Facial expression,’ ‘Body pose,’ and ‘Tone of voice.’ First, we judged the android robot’s responses to the questions it was asked. Excessive use of filler words such as ‘uh’ and ‘ah’ was judged to be unfavorable. Conversely, if the android robot responded in a clear voice and was quick to respond, it was judged to be favorable. Next, users who showed head nodding behavior were judged to be favorable. However, as with the other indicators, nodding at a certain timing regardless of the android robot’s utterance, or excessive nodding, was considered highly likely to indicate that the user was not favorable. In terms of facial expressions, users who frowned or looked at their companion with a wry smile were judged to be less favorable. Finally, we considered the posture (body pose) of the users. Finally, the body posture was judged to be favorable when the user was leaning forward with a straight back. On the other hand, the body pose of the user with cheekbones or leaning back excessively was judged to be less favorable.

### 6.1.5. Motivation

This measures the user’s willingness to engage in dialogue. The factors we focused on were ‘Head nodding,’ ‘Facial expression,’ ‘Body pose,’ and ‘Tone of voice.’ Highly motivated users nodded their heads when listening to the android robot’s utterance. They also tend to have neutral or positive facial expressions. Users with a positive body pose, such as stooping or straightening their backs, were judged to be highly motivated.

### 6.2. Evaluations for each strategies

First-person evaluations were conducted on a 5-minute dialogue video using the video annotation tool, ELAN. The session included 13 checkpoints designed to improve the SDP (Figure 9). The specific contents of the checkpoints are shown in Table 4. For these checkpoints, the user’s interest, concentration, understanding, liking, and motivation were rated as high or low, respectively. A timing chart, created based on the results of the first-person evaluation, is shown in Figure 10. The horizontal axis indicates the checkpoints where the evaluations were performed (see Table 1).
Figure 9. Checkpoints in a 5-minute dialogue (‘Th’ indicates the relevant section).

Table 4. The specific dialogue strategies implemented to improve the SDP are shown.

| ID | Strategy | Details of Implementation | Part Name |
|----|----------|---------------------------|-----------|
| T2 | 4.1 | Confirm the dialogue objectives of DRC2022 | Hearing |
| T3 | 4.5 | Multimodal response (utterance + pointing) request |  |
| T4 | 4.3 | Expresses a big smile in response to the user’s utterances |  |
| T5 | 4.3 | Strongly affirmative in response to the user’s opinion |  |
| T6 | 4.1 | Proposal of communication goal ‘creating a travel plan’ |  |
| T7 | 4.3 | Affirmative with a big smile in response to the user’s utterances |  |
| T9 | 4.4 | Highlighting the user’s choices |  |
| T10 | 4.2 | Presenting the keyword ‘cafe’ | Planning |
| T12 | 4.4 | Highlighting the user’s choices |  |
| T14 | 4.3 | Emotional expression using subjective speech |  |
| T15 | 4.2 | Presenting the keyword ‘lunch’ |  |
| T17 | 4.5 | Multimodal response (utterance + raising one’s hand) request | Wrap-Up |
| T18 | 4.3 | Strongly affirmative in response to the user’s opinion |  |

Note: The ‘ID’ column indicates the checkpoint where the strategies were incorporated. The ‘Strategy’ column indicates the subsection of the main text where the specific description is provided. The ‘Part Name’ column indicates the approximate place in the overall dialogue.

Next, based on these results, we examined the effectiveness of the strategy for improving SDP. We calculated the sum of the evaluation axes assuming the presence of an effect as 1, and 0 in cases where it did not appear. The summary is provided below.

6.2.1. Analysis results: proposal of a reasonable communication goal

The effects of the proposed rational dialogue objectives are shown in Table 5. First, the intended effect was observed for all participants of T2, in which the participants were asked to ‘decide which of the two selected sightseeing spots to visit,’ which was set by the DRC2022 administrator. For T6, in which the purpose of the dialogue was to ‘make a travel plan,’ the intended effect was observed in only four participants.

6.2.2. Analysis results: dealing with content that increases the user’s interest

The effects of using content to increase user interest are shown in Table 6. For T10, which provides topics related

Figure 10. The following are the evaluation results regarding SDP for the checkpoints where dialogue strategies were incorporated. A-E are the IDs given to each user. I, C, U, L, and M on the vertical axis indicate the evaluation axes, meaning interest, concentration, understanding, liking, and motivation, respectively. The horizontal axis indicates the ID (T1-T20) of the evaluated dialogue segment. Hearing, Planning, and Wrap-up in the figure are the part names of the dialogues, respectively. The graphs are filled in with 1 for improvement in each axis, and 0 for no improvement. Blank areas are dialogue sections in which no improvement was incorporated.
to ‘cafes,’ the intended effect was observed in four users. For T15, which provides topics related to ‘lunch,’ the intended effect was observed for all users.

6.2.3. Analysis results: sympathetic behavior in response to the user’s utterances
The effects of the robot’s empathetic behavior toward user utterances are shown in Table 7. First, we examined the robot’s attempts to express emotion using facial expressions and subjective utterances. For T4, which was implemented in the Hearing Part, three users showed the intended effect. Similarly, all users showed the intended effect for T7 in the Hearing Part. For T14, which was implemented in the Planning Part, the intended effect was observed in three users.

As for the robot’s attempts to affirm the user’s opinion using the speech text, the intended effect was observed in all the users for T5 in the Hearing Part. The intended effect was observed for T5 in the Hearing Part by three users. The intended effect was also observed for T18, which was conducted in the Wrap-Up Part.

6.2.4. Analysis results: providing choice opportunities
The effect of providing the opportunity to make a choice is shown in Table 8. For T9, which was conducted in the Hearing Part, the intended effect was observed in three users. The effect of the robot’s continuous speech is shown in the following table. For T12, which was implemented in the Planning Part where one-way robot speech is continuous, only one user showed the intended effect.

6.2.5. Analysis results: multimodal response request
The effects of the multimodal response requests are shown in Table 9. The intended effect was observed for all participants for both T3 in the Hearing Part and T17 in the Planning Part, where the robot’s speech continued continuously.

6.3. Evaluation results summary
Next, we summarize the effectiveness of the strategies for improving SDP for each participant (Table 10). Among the five participants, participant A (a male in his 20s) showed the highest effectiveness, while participant E (a woman in her 50s) showed the lowest effectiveness. While it is not possible to draw definitive statistical conclusions regarding individual differences, the results suggest that the strategies tend to work more effectively with the younger generation.

Despite the individual differences among the participants, the effects of the dialogue strategies designed in Section 4 are summarized in Table 11 for each of the five components of the SDP. Simple average rates for the five participants are shown only for the indicators that each strategy aimed at.

7. Discussion
In the previous section, we visualized the time-series change in the SDP. The results show that the SDP, which was high at the beginning of the dialogue, tends to
decrease when the robot retains the speaker’s right to speak for approximately two minutes (T10 to T16) during the route planning part in the middle of the dialogue. We focused on the time before and after the interval in which the user’s SDP decreased, which was intentionally placed, and compared the effects of the strategies we implemented.

First, the behavior that shows empathy for the user, which is often used in listening-oriented robots, does not work effectively in the proposed part, although the results in 6.2.3 indicate that it contributes to the SDP. In addition, although giving the user freedom of choice generally improves the SDP, the results of 6.2.4 show that the robot’s emphasis on ‘the customer’s chosen ___’ does not work effectively in the proposal part. For users who meet the robot for the first time, the robot’s facial expressions and detailed expressions of speech content are unlikely to lead to direct recognition.

On the other hand, contents that stimulate the user’s interest can appeal directly to the user without increasing the user’s recognition cost. The results of section 6.2.2 suggest that the user’s SDP is generally enhanced when the topic is related to food, such as ‘cafe’ and ‘lunch.’ Since this strategy element was made in the proposal part, it is suggested that it is effective even in a situation where the SDP may decrease.

Furthermore, the results of 6.2.5 suggest that having the user speak with body movements immediately after the user’s right to speak is restricted for a certain period of time contributes to improving the SDP. This is an interesting result, although quantitative research is needed to sort out the effects of other factors.

8. Conclusion

In this paper, we proposed indices of user’s sense of participation in dialogue (SDP) from the viewpoint of how to make the user engage in the dialogue with the robot. Interest, concentration, comprehension, liking, and motivation, which were estimated by using sensors and devices in previous studies, were considered to be the components of the SDP, since these values can be obtained by video evaluation by others.

The authors’ team, which participated in the DRC-2022, constructed an android dialogue system that incorporates many strategies that contribute to improving the SDP.

The system won an honorable mention award after 29 people experienced a dialogue scenario about guiding them to tourist attractions. We investigated the effects on the SDP of five strategies: (1) proposing reasonable dialogue objectives, (2) using content that increases the user’s level of interest, (3) showing the robot’s empathy for the user’s utterance, (4) providing opportunities to make choices, and (5) expressing intentions through user actions.

We extracted dialogue data of five people with different attributes, assigned effect labels to each of the five components by video annotation based on the author’s subjective evaluation, and visualized the time-series changes in the SDP. In the middle part of the scene where the robot guided the user along a sightseeing route, the robot retained the right to speak for about two minutes, and the user’s SDP was supposed to decrease. Although the overall trend during this period was a decrease in SDP, the effects of (2) the use of content that increased the user’s interest and motivation, and (5) the expression of intention by the user’s actions, which aimed to increase interest and understanding, continued to be effective. Consequently, we can conclude that these methods have a discernible impact on improving the user’s SDP. Moving forward, increasing the sample size will be imperative for the validation of quantitative evaluations.

Acknowledgments

We would like to express our gratitude to Dr. T. Minato and the development team members at ATR for their invaluable support of the Android platform. Their expertise and dedication have greatly contributed to the success of this paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was supported by JSPS Grant-in-Aid for Scientific Research on Innovative Areas, Communicative intelligent systems towards a human-machine symbiotic society [grant number JP19H05693].

Notes on contributors

**Makoto Kawamoto** received a Master’s degree in Engineering from Tokyo Denki University in 2023. She served as the leader of Team DSML-TDU in DRC 2021 and as the leader of Team MIYABI in DRC 2022. In each instance, she achieved the Excellence Award and the Honorable Mention Award, respectively. Her main areas of research focus on human-robot interaction. She is currently working at TOPPAN.

**Masaki Shuzo** obtained a Ph.D. degree in engineering from The University of Tokyo in 2003. Since October 2017, he has been working in Tokyo Denki University. His research interests include human-robot interaction and android science.

**Eisaku Maeda** received the MS degrees in zoology in 1984 and the PhD degree in mathematical engineering in 1993 from the University of Tokyo respectively. He joined NTT (Nippon Telegraph and Telephone Corp.) in 1986, and has been involved in a wide range of research fields including pattern recognition,
machine learning, image recognition, natural language processing, and neuroscience. In 2017, he moved to the faculty of system design engineering at Tokyo Denki University as a professor, and has served as dean of the faculty since 2020.

References

[1] Bono M, Suzuki N, Katagiri Y. An analysis of participation structure in conversation based on interaction corpus of ubiquitous sensor data. In: Rauterberg M, Menozzi M, Wesson J, editors. Human-Computer Interaction INTERACT’03: IFIP TC13 International Conference on Human-Computer Interaction, 2003, Zurich, Switzerland. Amsterdam, Netherlands: IOS Press; 2003.

[2] Kawahara T, Hayashi S, Takanashi K. Estimation of interest and comprehension level of audience through multi-modal behaviors in poster conversations. In: Proceedings Interspeech 2013, Lyon. 2013. p. 1882–1885. https://www.isca-speech.org/archive/interspeech_2013/

[3] Miyahara M, Aoki M, Takiguchi T, et al. Tagging video contents with positive/negative interest based on user’s facial expression. In: Proceedings Advances in Multimedia Modeling 2008. Berlin, Heidelberg: Springer-Verlag; 2008. p. 210–219.

[4] Lopez G, Ide H, Shuzo M, et al. Workplace stress estimation from physiological indices in real situation. In: Cipresso P, Matic A, Lopez G, editors. Pervasive Computing Paradigms for Mental Health. Cham: Springer International Publishing; 2014. p. 13–22.

[5] Uema Y, Inoue K. JINS MEME algorithm for estimation and tracking of concentration of users. In: Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers. New York (NY): Association for Computing Machinery; 2017. p. 297–300.

[6] Hamaguchi N, Yamamoto K, Iwai D, et al. Subjective difficulty estimation for interactive learning by sensing vibration sound on desk panel. In: de Ruyter B, Wichert R, Keyson DV, Markopoulos P, Streitz N, Divitini M, Georgantas N, Mana Gomez A, editors. Ambient Intelligence. Berlin, Heidelberg: Springer Berlin Heidelberg; 2010. p. 138–147.

[7] Uchida T, Minato T, Nakamura Y, et al. Female-type android’s drive to quickly understand a user’s concept of preferences stimulates dialogue satisfaction: dialogue strategies for modeling user’s concept of preferences. Int J Soc Robot. 2021;13:1499–1516. doi:10.1007/s12369-020-00731-z

[8] Maruyama K, Yamato J, Sugiyama H. Analysis of how motivation to conversation varies depending on the number of dialogue robots and presence/absence of gestures. IEICE Trans Inf & Syst (Japanese Ed.). 2021;J104-D(1):30–41.

[9] Inoue K, Kala D, Takanashi K, et al. Engagement recognition by a latent character model based on multimodal listener behaviors in spoken dialogue. APSIPA Trans Signal Inf Process. 2018;7:e9. doi:10.1017/ATSIP.2018.11

[10] Goffman E. Behavior in public places: notes on the social organization of gatherings. New York: Free Press; 1963.

[11] Kawamoto M, Shuzo M, Maeda E. Improving user’s sense of participation in robot-driven dialogue. Preprint, 2022. arXiv:2210.09746. cs.RO.

[12] Higashinaka R, Minato T, Sakai K, et al. Dialogue robot competition for the development of an android robot with hospitality. In: 2022 IEEE 11th Global Conference on Consumer Electronics (GCCE), Osaka. 2022. p. 357–360. https://ieeexplore.ieee.org/document/10014078

[13] Minato T, Higashinaka R, Sakai K, et al. Overview of dialogue robot competition 2022. Preprint, 2022. arXiv:2210.12863. cs.RO.

[14] Nishio S, Ishiguro H, Hagita N. Geminoid: teleoperated android of an existing person. In: de Pina Filho AC, editor. Humanoid Robots. Rijeka: IntechOpen; 2007. Chapter 20.

[15] Glas DF, Minato T, Ishi CT, et al. ERICA: the ERATO intelligent conversational android. In: 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). New York (NY): IEEE Press; 2016. p. 22–29.

[16] Matsumura R, Shiomi M, Hagita N. Does an animation character robot increase sales? In: Proceedings of the 5th International Conference on Human Agent Interaction. New York (NY): Association for Computing Machinery; 2017. p. 479–482.

[17] Iwamoto T, Baba J, Nakanishi J, et al. Playful recommendation: sales promotion that robots stimulate pleasant feelings in stead of product explanation. IEEE Robot Autom Lett. 2022;7(4):11815–11822. doi:10.1109/LRA.2022.3189149

[18] Sato W, Namba S, Yang D, et al. An android for emotional interaction: spatiotemporal validation of its facial expressions. Front Psychol. 2022;12. doi: 10.3389/fpsyg.2021.80657

[19] Collins GR. Improving human-robot interactions in hospitality settings. Int Hosp Rev. 2020;34(1):61–79.

[20] Yamazaki T, Yoshikawa K, Kawamoto T, et al. Tourist guidance robot based on HyperCLOVA. Preprint, 2022. arXiv:2210.10400. cs.CL.

[21] Han S, Bang J, Ryu S, et al. Exploiting knowledge base to generate responses for natural language dialog listening agents. In: Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue; September 2015. Prague, Czech Republic: Association for Computational Linguistics; 2015. p. 129–133.

[22] Johansson M, Hori T, Skantz G, et al. Making turn-taking decisions for an active listening robot for memory training. In: Agah A, Cabibihan JJ, Howard AM, Salichs MA, He H, editors. Social Robotics. Cham: Springer International Publishing; 2016. p. 940–949.

[23] Schröder M, Bevacqua E, Cowie R, et al. Building autonomous sensitive artificial listeners (Extended abstract). In: 2015 International Conference on Affective Computing and Intelligent Interaction (ACII), Xi’an. 2015. p. 456–462. https://www.computer.org/csdl/proceedings/acii/2015/120mNyugyR5

[24] Watanabe M, Ogawa K, Ishiguro H. At the department store—can androids be a social entity in the real world? In: Geminoid Studies: Science and Technologies for Humanlike Teleoperated Androids. 2018 Apr, Springer. p. 423–427. https://link.springer.com/book/10.1007/978-981-10-8702-8
[25] Wu Y, Huang TS. Vision-based gesture recognition: a review. In: Braffort A, Gherbi R, Gibet S, Teil D, Richardson J, editors. Gesture-Based Communication in Human-Computer Interaction. Berlin, Heidelberg: Springer Berlin Heidelberg; 1999. p. 103–115.

[26] Mitra S, Acharya T. Gesture recognition: a survey. IEEE Trans Syst Man Cybern Syst. 2007;37(3):311–324. doi:10.1109/TSMCC.2007.893280

[27] Suarez J, Murphy RR. Hand gesture recognition with depth images: a review. In: 2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication, Paris. 2012. p. 411–417. https://www.rsj.or.jp/event/international/roman/