



Robot Feedback Design for Response Delay

Dahyun Kang¹ · Changjoo Nam² · Sonya S. Kwak¹

Accepted: 25 September 2023 / Published online: 22 December 2023
© The Author(s) 2023

Abstract

We investigated an effective human-robot interaction design to alleviate the user's negative experience caused by the response delay of a robot. The feedback regarding a robot's delayed response was designed using human-like and machine-like approaches. The first experiment was conducted in task-based interaction to explore the effects of the feedback type (human-like or machine-like feedback versus a baseline robot) and task type (a high-cognitive-demand task versus a low-cognitive-demand task) on the understandability, perceived waiting time, and service evaluation of the robot ($N = 36$). The robot with machine-like feedback was rated the most positive in terms of understandability, perceived waiting time, and service evaluation. The perceived waiting time mediated the effect of feedback type on service evaluation. There were significant effects between the feedback types and the task types on understandability, perceived waiting time, and service evaluation. In a low-cognitive-demand task, machine-like feedback was evaluated more positively than human-like feedback, while in a high-cognitive-demand task, there was no significant difference between human-like feedback and machine-like feedback. In addition, the second experiment was conducted in social interaction between humans and robots to explore the effect of the two feedback types on the understandability, perceived waiting time, sociability, and service evaluation of the robot ($N = 36$). The robots providing feedback were evaluated more positively than the baseline robot, which did not provide any feedback, and sociability mediated the effect of feedback type on service evaluation.

Keywords Response delay · Robot feedback design · Machine-like feedback · Human-like feedback · Task-based interaction · Social interaction

1 Introduction

Although robot reactions to user commands are becoming faster with the advancements in technology, they have yet to reach the speeds users expect [1, 2]. A slow response from a robot increases the likelihood that the user will abandon the interaction. According to Guynes [3], the user experiences negative emotions, such as anxiety, when the response delay time of an interacting object is high. However, such emotions

can be alleviated by effectively managing the waiting experience of the user [4–7]. Further, negative emotions increase when the waiting time is uncertain [5, 8]. In other words, if the uncertainty in waiting time is reduced, the level of stress experienced by the user decreases, while the willingness to wait for the response increases [9, 10]. Giving feedback to the user during response delay can reduce this uncertainty by leading to a positive user experience [11, 12]. The robot can prolong the waiting time of the user while avoiding negative emotions by providing feedback on the response delay [13]. The goal of this study is to figure out the effective robot feedback for response delay. Accordingly, this study suggests that the feedback on the delayed response of the robot helps to inform the user regarding the state of the robot, thereby reducing the perceived waiting time (PWT) and mitigating the negative user experience caused by the delayed response.

✉ Sonya S. Kwak
sonakwak@kist.re.kr

Dahyun Kang
dahyun@kist.re.kr

Changjoo Nam
cnam@sogang.ac.kr

¹ Center for Robotics Research, Korea Institute of Science and Technology, 5, Hwarang-ro 14-gil, Seongbuk-gu, Seoul 02792, Republic of Korea

² Department of Electronic Engineering, Sogang University, 35 Baekbeom-ro, Mapo-gu, Seoul 04107, Republic of Korea

2 Related Works

In this section, we introduce theoretical background on response delay and feedback design. In addition, we describe the task type as a factor that can influence the preferred response delay feedback type.

2.1 Response Delay

When people interact with web applications, three types of response time limits exist. If the web application responds within 0.1 s, users assume that they obtained an immediate response [14]. A 1.0 s delay is the maximum time during which the user does not think that there is a response delay [14]. Finally, 10 s is the maximum amount of time the user can focus on the interaction [14–16]. If a longer response delay is required, an estimated time to complete the task may be provided by the computer. The response time has been used as one of the service quality indicators [17]. Delays in system response and waiting times for services negatively affect user perception [18, 19]. Users can easily become annoyed and assume that a security problem has occurred when a response delay occurs [4, 20].

Providing appropriate visual feedback on what is happening in the system using user interface design is one of the most important and universal ways to reduce the negative perception of a response delay [13]. One of the most common forms of system feedback is an animated progress indicator. Using such an indicator, a computer can inform the user of its state, indicating it is in the middle of loading or processing information, and the user can be assured that the system understands the user's request and is processing it [13, 14].

In the field of human-computer interaction (HCI), progress indicators have been extensively studied to alleviate negative emotions caused by the system response delay. In Nah's [11] study, the experiences of two groups that were/were not provided with a progress bar were compared when clicking on a link and waiting for the page to load. Consequently, those who saw the progress bar felt more satisfied with their experience and waited three times longer than those who did not see any progress indicator [11]. Lee et al. [12] found that the progress bar shortened the perceived uncertainty of the delay and reduced the PWT.

In the study of human-robot interaction (HRI), this delay is utilized as a social strategy. Rich et al. [21] explored adequate time delay values for the robot to be perceived as responsive to the human counterpart. As a result, 3.0 s delay was the best time delay value when the robot gazes at the object after a person pointed at the object, while 1.8 s delay was the best time delay value when the robot mutually gazes at the counterpart after a person gazes at the robot [21]. In addition, Yamamoto et al. [22] found that a robot that starts talking about 0.6 s after the user greets is the most preferred in a greeting situation.

Kanda et al. showed that when a robot responds physically, it should have a delay of about 0.89 s to be perceived as natural [23]. These studies revealed appropriate response times for various situations on the premise that the robot may respond immediately. However, the robot may also respond slower than expected due to technical limitations. Therefore, it is necessary to study viable strategies to mitigate the effects of unavoidable response delays.

Several studies suggest that robots use conversational fillers to alleviate user complaints about delayed responses because robots that used conversational fillers were evaluated more positively in preference, aliveness, humanness, and perceived speediness than robots that did not use fillers, and users felt a higher social presence when interacting with a robot that used conversational fillers [24–26]. Robots can be perceived as both machines and social entities that interact with humans. Thus, both the feedback for the response delay provided by the devices such as computers, copiers, and cell phones and the feedback used in human-human interactions such as fillers can be applied to robots.

2.2 Human-Like Versus Machine-Like Robot Design

Due to the human-machine duality of robots, which can be both machines and social entities, some researchers suggest designing robots as machine-like, while others suggest designing robots as human-like.

Continuous efforts are being made in robotics to develop human-like robots so that the user can intuitively predict and understand the expressions of robots during HRI [27]. Researchers argued that robots can interact with humans most effectively when they are designed to communicate in the same way as humans [28]. In addition, a human-like robot design was evaluated to be friendlier and more intelligent than a machine-like robot design, and it has been asserted that robots should be designed in a form that mimics humans, allowing them to interact socially [29–31].

Contrarily, other studies suggest designing a robot as machine-like. Kwak et al. [32] examined the effect of a robot's organism-based versus object-based appearance design on user satisfaction. They found that when the function expected from the robot's appearance matched the actual function of the robot, user satisfaction increased. The results showed that the object-based appearance of a robot reduces the user expectation of the robot's function rather than the organism-based appearance, and the appearance that reduces user expectation leads to an increase in user satisfaction [32]. In addition, Woods found that human-like robots are perceived as more aggressive and less friendly than machine-like robots because human-like objects which are imperfectly resemble humans provoke uncanny feelings and aversion in observers [33]. Researchers who have argued that robots should be designed to be machine-like state that machine-like

(less human-like) robots can receive positive evaluations by lowering expectations and alleviating discomfort or repulsion evoked due to the uncanny valley in human-like robots [33–36]. Several robot engineers and designers agreed with this viewpoint and designed the robot's appearance as machine-like [37–39]. As such, opinions about robot design are divided because a robot has two characteristics as a machine and a social entity. In any case, the results of previous HRI studies show that the robot design type affects the evaluation of robots. In the field of HRI, many studies on robot state expressions are being conducted to apply human biological signals and social cues to robot state expressions for natural human-robot interaction [40, 41]. On the other hand, since a robot is also a machine, indicators used for machines, such as a bar-type indicator showing the remaining battery level and a stair-step indicator showing the network status, are being used in designing robots. In this study, we investigated which feedback type, machine-like or human-like, is preferred by the user when the robot's response is delayed.

2.3 High-Cognitive-Demand Task and Low-Cognitive-Demand Task

Studies have shown that the appropriate robot appearance varies depending on the task type. Parsonage et al. showed that a machine-like robot is predicted to perform physical labor rather than tasks requiring high cognition. Contrarily, intellectual work is also expected in addition to physical labor in the case of a human-like robot [42]. In addition, people tend to prefer human-like robots for creative or negotiation tasks. In contrast, machine-like robots are preferred for manual jobs or automation [43]. As such, the robot task type influences the preferred robot appearance by humans.

Moreover, task complexity is one of the factors that influence the user's tolerance to delayed systems [13]. When the system is performing a task that is difficult for humans, the user will be more willing to wait for the system to complete the task than when the system is handling an easy task. The tasks can be divided into a low-cognitive-demand task (LCDT) and a high-cognitive-demand task (HCDDT). An LCDT means describing a certain fact as it is and solving simple problems following known procedures, while an HCDDT involves connecting and analyzing information and drawing conclusions based on it [44, 45]. A robot can do both tasks. For example, PR2, developed by Willow Garage, can perform tasks such as plugging a cord into an outlet, walking a dog, and playing billiards [46, 47]. Among the tasks PR2 can perform, plugging a cord into an outlet is an LCDT that humans do daily without analyzing information and drawing conclusions. By contrast, playing billiards is an HCDDT that requires recognizing the situation, inferring the strategy the opponent will implement, and formulating and executing their own strategy.

However, from the robot's point of view, a task that is difficult for humans can be easy, or vice versa. When getting an order to move a small object from point a to point b, humans build action planning in a very short time. They can understand the order, recognize objects, calculate a travel path to grab the object, and calculate the path to move the object—all intuitively and customarily. Thus, it is an LCDT for humans. However, for today's robots in natural interaction with humans, it can be an HCDDT. When performing the same task, robots need longer time than humans because they must do sound processing, acoustic scoring, search algorithms using language dictionary databases, understand the order, identify the object, recognize the location of the object, find the most ideal path among the various possibilities based on its location before picking up an object and moving it. Meanwhile, robots can solve mathematical calculations, which are an HCDDT for humans, in seconds. As can be seen from the success of AlphaGo, which was developed by Google following the rapid development of artificial intelligence (AI), the robot's AI can guess an opponent's next move overwhelmingly faster than humans and solve very complex mathematical problems on the fly [48].

Because moving an object is an LCDT that humans can do faster and easier than robots, users may find it tedious to wait for the robot's response delay, whereas when object movement occurs as a result of mathematical calculation, the user will be more open to waiting for the robot's response delay because a real robot can perform mathematical calculations in seconds, but it is an HCDDT that takes several minutes or more for humans. At this point, it is possible to assume that a response delay may negatively affect the user for LCDTs, but a delay can be considered natural when the robot performs an HCDDT, although it only needs an additional split second. Therefore, we tried to determine whether there is a difference in user satisfaction between tasks that are easy for humans and tasks that humans consider difficult.

3 Study 1: Task-Based Interaction

3.1 Hypotheses

Based on the results of research on response delay in HCI and HRI, several hypotheses were proposed about how robot feedback types for response delay affect HRI. Hypothesis 1, Hypothesis 2, Hypothesis 3, and Hypothesis 4 predict the effect of the presence or absence of robot response feedback and feedback types on people's perceptions of the robots. Hypotheses 5, 6, and 7 predict the presence of interaction effects between the robot's task type and the feedback type on the robot's impressions.

3.1.1 Hypothesis 1

Robot response delay feedback helps the user to better understand its processing state. This prediction follows the design study of HCI that the user understands the state of the system delay better when there is a progress indicator than when there is not [13].

3.1.2 Hypothesis 2

Robot response delay feedback affects PWT. This prediction follows the findings of the study of Hui and Zhou [9] and Weinberg [10]. Hui and Zhou [9] stated that decreasing the degree of uncertainty reduces PWT. In addition, Weinberg [10] demonstrated that letting the computer indicate its state to the user is effective in decreasing uncertainty about the machine.

3.1.3 Hypothesis 3

Robot response delay feedback types affect user satisfaction with the robot. Our predictions come from the findings of Bar-Cohen and Breazeal [29], Hegel et al. [30], Kwak et al. [32], and Woods [33]. According to these studies, robots can be designed to be human-like or machine-like, and these appearances affect user evaluation of the robots [29, 30, 32, 33]. Accordingly, we predict that the user evaluates robots differently depending on the type of robot feedback design.

3.1.4 Hypothesis 4

The effects of the robot feedback type on user satisfaction with the robot is mediated by PWT. In Hypothesis 2, we predicted that robot response delay feedback would affect PWT, and in Hypothesis 3, we hypothesized that robot response delay feedback would affect user satisfaction. Guynes [3] found that as the waiting time prolongs, people feel more negative emotions. Therefore, it was predicted that the robot feedback, which reduces the PWT, provokes a positive evaluation of the user for the service provided by the robot.

3.1.5 Hypothesis 5

The type of robot, human-like or machine-like, can influence people's assumptions about the robot's role [42, 43]. Additionally, the nature of a robot's feedback, whether human-like or machine-like, can impact that users interpret the robot's state. Moreover, according to Brun and Teigen's research [49] on human verbal behavior, it has been shown that the same expression can be understood differently in various contexts. Considering these studies, we hypothesized that the user's level of comprehension regarding the robot's state, influenced

by the robot's feedback type, may vary depending on the task type.

3.1.6 Hypothesis 6

Depending on the task type, the feedback type that reduces the PWT is different. This hypothesis was established based on HCI studies by Nah's [11], Sherwin's [13], and Abbas et al.'s [50]. According to these studies, task type can affect users' tolerance, and the effective strategy to reduce PWT for each task type may be different.

3.1.7 Hypothesis 7

Goetz et al. [43] found that the expected role of a robot is different according to its appearance. In addition, Kwak et al. [32] showed that robots can be positively evaluated when they perform in accordance with the expected function. Based on their findings, we predicted that the preferred response delay feedback type for each task would be different. Therefore, we hypothesized that the preferred robot feedback design would vary depending on the task type.

To verify these hypotheses, we conducted experiments with mixed participants. We executed a 2 (task type: LCDT vs. HCDT) X 3 (feedback type: human-like feedback vs. machine-like feedback vs. baseline) mixed-participant experiment. Task type was a between participant variable. Feedback type was a within participant variable. That is, a participant experienced three types of robot feedback within one task.

3.2 Robot feedback design

We designed two types of robot feedback to explore the effect of robot feedback types on response delay- human-like feedback and machine-like feedback.

Since machines are artifacts created by humans as tools to achieve specific goals, most parts of machines are designed to be useful and efficient for users. The progress bar gives feedback on the response delay, providing a visual indication of how much longer the user will have to wait. It has already been found to be a useful and efficient status display method for response delays, as it makes users more willing to wait longer than other feedback [11, 13, 14]. On the other hand, many HCI and HRI studies have revealed that human-like expression of a machine is the most effective way to communicate with humans [27, 29–31]. It is difficult for the human to display the counterparts how much information they have processed and how long they have to wait. Thus, unlike machines, humans might not be designed to be efficient in indicating response delay. However, a robot's human-like feedback can intuitively inform the human of the robot's status since they are used to the social cues used

between humans [27]. Therefore, in order to find out how effective the human-like feedback, which gives feedback in the same way as humans, although the amount of information is insufficient, is compared to the progress bar, the feedback in this study is divided into two types as follows. A feedback that mimics human behavior when thinking and one of the representative machine-like feedback, the progress bar.

3.2.1 Human-Like Feedback

To design human-like robot feedback, we observed how people behaved before any action or speech, that is, when they were in a state analogous to a robot's processing state. We analyzed our observations and applied them to robot feedback design.

Data Collection

We recruited 10 participants (7 males and 3 females) in their mid-20s to early 30s. To observe what people do when they think, we gave people contemplative situations and observed their actions and voices. Each participant was asked 10 riddles. The riddles set by the examiner were, “What goes on four legs in the morning, on two legs at noon, and on three legs in the evening?” [51], “This thing devours all things—birds, beasts, trees, and flowers. It gnaws iron, bites steel, grinds hard stones to meal, slays kings, ruins towns, and beats high mountains down. What is it?” [52], and so on. The participants sat down face to face with the examiner and solved the riddles. The participants' voices and gestures were recorded and analyzed.

Observation and Analysis

Since the goal of the coding was to discover the most common thinking behaviors, we focused on capturing the types of thinking behaviors and selecting representative behaviors among them. We recruited two coders and conducted training sessions by asking them to code two video samples. In the training session, the coders segmented the two videos and then arbitrated the disagreement on the segmentation. Afterward, they coded the participants' behaviors within the agreed segmentation. Lastly, they arbitrated for disagreement on coding tags.

We used the proportional reduction in loss (PRL) reliability measure to check the intercoder reliability [53]. The PRL level of the video segmentation was .77 (22/33 = .67). Secondary training session for segmentation was conducted to perform reliable coding since the PRL level marginally exceeded the adequate PRL level of .70. The PRL reliability measure of the secondary segmentation was calculated as .89 (24/29 = .83). Because the PRL level exceeded the adequate PRL level of .70, the coders performed category coding on the videos after arbitration for segmentation. The PRL level of the two videos for category coding was .81 (21/29 = .72).

After the training session, the coders coded 10 videos following the same procedure used in the training session. The

PRL level of the video segmentation was .96 (200/215 = .93). The PRL level of the 10 videos for category coding was .94 (196/215 = .91). It supports the utility of video segmentation and behavior category coding since the PRL level of behavior segmentation and category coding exceeded the proper PRL level of .70.

The total segment was 210 related to human thinking from 10 videos. The 210 segments were categorized into 10 types: 1) tapping the table with fingers, 2) touching the face with the hands, 3) propping the chin in the hand, 4) one hand touching the other hand, 5) holding two hands, 6) touching the head, 7) banging the table with fists/palms, 8) touching the body, 9) depicting the object being thought about by hand, and 10) writing on the table with a finger. Tapping the table with fingers had the highest number of segments (46 segments), and all participants except one performed this behavior (see Fig. 1). Touching the face with the hands was observed in 40 segments, but only four participants performed this behavior. One hand touching the other (34 segments) was the third most frequently observed behavior, followed by propping the chin in the hand (31 segments), holding two hands (20 segments), touching the head (16 segments), banging the table with fists/palms (7 segments), touching the body (7 segments), depicting the object being thought about by hands (7 segments), and writing on the table with a finger (2 segments). Therefore, we applied tapping of the table with fingers, which was the most frequently observed behavior performed by almost everyone, to the robot feedback design for response delay.

In thinking situations, all participants used fillers, such as “uh” or “um.” Smith and Clark [53] argued that people use fillers when the response is delayed. When a longer time is required to answer, “um” is more likely to be used than “uh” as a filler ($M_{uh} = 2.65$, $M_{um} = 8.83$) [54]. “Um” and “uh” are the fillers most commonly used along with “ung” and “geu” by Koreans [55]. A filler with a vocalic-nasal form is universal as it is used in English, German, Dutch [56], Korean [55], and Chinese [57]. Therefore, we designed the robot's human-like feedback as follows: visual expression-tapping on the table with its fingers, and sound expression-the filler “um.”

3.2.2 Machine-Like Feedback

In HCI design, a progress indicator is recommended for tasks that take more than 1 s [13]. Progress indicators include looped animation, percent-done animation, and so on [13]. The recommended progress indicator differs depending on how long the process takes. Looped animation, which does not provide information about how long the user has to wait, is used only when there is a relatively short delay of 2–10 s [13]. For a response delay of more than 10 s, the percent-done animation, which displays the task's degree-completed and

Fig. 1 Most frequently observed thinking behavior: Tapping the table with fingers



Table 1 Examples of scripts by two task types

	Low-cognitive-demand task	High-cognitive-demand task
Participant	“Put the can in box B”	“Put the can in the answer box There are three people at the scene of the incident: a police officer, a criminal, and a witness. Police always tell the truth, criminals always lie, and witnesses tell the truth or lie depending on the situation. The three stated: Who is the criminal among A, B, and C? A: I am not a police officer B: I am not a criminal C: I am not a witness.”
Robot	(Expressing feedback for response delay for 10s) (Putting the can in box B)	(Expressing feedback for response delay for 10s) (Putting the can in box B)
Participant	“Put the can in box A”	“Put the can in the answer box Three people, A, B, and C, who were alleged to be suspects said A: I am not the culprit B: That is right. A is not the culprit C: I am the culprit. If one of the suspects is the culprit, the culprit is lying, and the witnesses’ words may or may not be lying, who is the culprit?”
Robot	(Expressing feedback for response delay for 10s) (Putting the can in box A)	(Expressing feedback for response delay for 10s.) (Putting the can in box A)

degree-remaining state, is recommended [13]. This progress indicator can comfort a user and increase the willingness to wait [13]. In general, robots spend 10s or more for simple tasks, such as moving an object from point a to point b. Therefore, we adopted a progress bar that shows percent-done animation for the robot’s machine-like feedback. In addition, as a sound expression corresponding to the filler of human-like feedback, the electronic sound “beep beep” was added to the robot feedback design. Thus, we designed the robot’s machine-like feedback as follows: visual expression-LED progress bar, and sound expression- “beep beep.”

3.3 Method

3.3.1 Participants

We recruited 36 participants, aged between the mid-20s to the mid-30s ($M = 28.89$; $SD = 3.32$; 17 males and 19 females). Participants received \$10 for the experiment.

3.3.2 Task

To figure out the appropriate robot feedback design according to task type, the following two tasks were devised.

In the LCDT, the user gave this voice command to the robot: “Put the can in box A/B/C,” and the robot picked up the can after one of the three types of feedback-human-like, machine-like, or baseline -and then put the can in the box the user said. In the HCDT, the participant presented a reasoning question and three choices of A, B, and C to the robot and asked the robot to put the can in the answer box. The robot exhibited one of the three kinds of feedback-human-like, machine-like, or baseline- in a counterbalanced order, picked up the can, and put it in the answer box. The tasks requested by the users in LCDTs and HCDTs are shown in Table 1.

3.3.3 Robot Type

This experiment was a study comparing the effects of human-like feedback and machine-like feedback. Therefore, it is not appropriate to use a humanoid that mimics the human body or

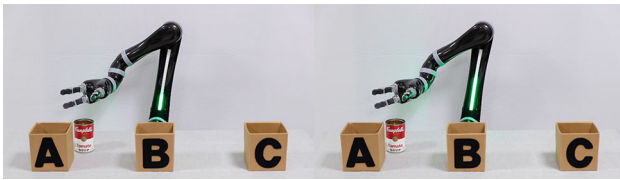


Fig. 2 Machine-like feedback design. During the response delay of 10s, the robot's progress bar gradually fills up



Fig. 3 Human-like feedback design. During the response delay of 10s, the robot taps the desk with its fingers

to use a robotized product that maintains the shape of a product for experiments. According to the observations in Sect. 3.2, since the most frequently observed behavior was tapping the table with fingers, we used a robot with an arm and fingers for human-like feedback in the experiment. Therefore, Kinova Gen2 robot, an arm robot that mimics human body parts (arms and hands) in a mechanical form, was used in this experiment [58]. This arm robot can manipulate objects in a three-dimensional workspace. Machine-like/human-like feedback design was applied to the robot. To implement machine-like feedback, an LED strap was attached between joint 2 and joint 3 of the robot. The user experienced visual feedback as a progress bar while the robot processed information and audible feedback as a “beep beep” sound (Fig. 2).

For human-like feedback, the robot behavior of tapping fingers on the table was implemented by programming it with Robot Operating System [59] to give the user visual feedback, and a filler sound was used to give audible feedback (Fig. 3).

3.3.4 Experimental Setup

In the pilot experiment, when the robot recognized the human voice and provided response delay feedback, the success rate of manipulation did not reach 80%. The robot did not always move in the same path, and sometimes, it failed to pick up the can. In particular, it was difficult to manipulate the time delay in seconds. Since a time delay of just a few seconds could affect user perception, we had to minimize this irregularity of the robot's time delay. In addition, several participants did not say the exact command given. By mistake or because of their own speech style, their use of predicate or arrangement of words differed, which allowed the robot to perceive the participant's speech differently. Therefore, we conducted this study with the Wizard of Oz technique to set up a fully

manipulated experiment [60]. The participant was asked to control the robot using voice commands in telecommunication. In the remote environment where the robot was installed, the robot took home motion. In front of the robot, there were three boxes: boxes A, B, and C. A camera was located in front of the table where the robot was installed. There were two spaces, a space for interacting with the robot remotely and a space for filling in a questionnaire. A monitor and a microphone were placed in the laboratory where the participant interacted with the robot remotely, and a chair was placed in front of the monitor for the participant. The participant sat in this chair and asked the robot to perform the task. Tables, chairs, and a tablet PC were placed in the survey space so that the participant could sit on the chair and complete the questionnaire using a tablet PC. The experimenter played the video with a robot that took home motion. When the user gave a voice command to the robot to perform a specific task, the experimenter played the video with the robot expressing response delay feedback while performing the task.

3.3.5 Measurement

We measured understandability to see if robot feedback helped users understand the state of the robot (4 items, Cronbach's $\alpha = .781$) [61]. To determine the extent to which the robot's response delay feedback could reduce the waiting time perceived by the user, the PWT was evaluated (3 items, $\alpha = .884$) [62]. The items of the PWT were drawn from the HCI study [62] and modified for this study. Service evaluation was measured to verify the impacts of the tasks performed by the robot and robot feedback design on user satisfaction (3 items, $\alpha = .939$) [63] (Table 2).

3.3.6 Procedure

After briefing the experiment, the participants filled out the experiment agreement document. The participant experienced a robot with human-like feedback and a robot with machine-like feedback, and a baseline robot that did not give any feedback for a particular task in a counterbalanced order. Participants rated the robot immediately after experiencing it.

3.4 Results

The Shapiro-Wilk test was performed to test the normality of the data obtained through the experiment. Our data met normality. In addition, an equal variation test was performed. As a result, the data was satisfied with homoscedasticity. The detailed normality and equal variation test results are shown in Table 3.

Therefore, we analyzed the data by performing parametric statistical analysis methods: paired t-test, two-way repeated

Table 2 Measurement

Scale	Items
Understandability	1: Confusing—7: Clear 1: Inconsistent—7: Consistent 1: Hard to Understand—7: Easy to Understand 1: Inexpressive—7: Expressive
PWT	The robot’s response to your request was fast The robot’s response to your request was speedy The robot’s response to your request was quick (1: Strongly disagree—7: Strongly agree)
Service evaluation	1: Very good—7: Very poor 1: Completely dissatisfied—7: Completely satisfied 1: Would avoid using the service—7: Would want very much to use the service

Note 1 PWT: perceived waiting time

measure ANOVA, and Tukey’s method. To test mediation effects, we used PROCESS model 4 of the SPSS Process macro to evaluate the mediation effects [64].

3.4.1 Effects of Response Feedback on Understandability, PWT, and Service Evaluation

H1 was supported by the data. There was a significant effect of robot feedback type on the understandability of the robot’s state ($M_{human-like} = 4.67, SD = 1.21$ vs. $M_{machine-like} = 5.30, SD = 0.81$ vs. $M_{baseline} = 4.00, SD = 1.06; F(2,68) = 15.383, p < .001$). As a result of the Post hoc analysis using a Tukey HSD test, there were significant differences between the evaluation of a robot with machine-like feedback and human-like feedback ($p = .004$), with machine-like feedback and without feedback ($p < .001$). However, there was no difference between the evaluation of a robot with human-like feedback and without feedback ($p = .101$). Participants understood the state of a robot with machine-like feedback better than the state of the robot with human-like feedback and that without feedback.

H2 was partially verified. The ANOVA showed that the effect of the robot’s feedback for response delay on the PWT was significant ($M_{human-like} = 3.19, SD = 1.50$ vs. $M_{machine-like} = 4.22, SD = 1.22$ vs. $M_{baseline} = 3.20, SD = 1.35; F(2,68) = 11.949, p < .001$). As a result of the Tukey HSD test, the Post-hoc analysis, there were significant differences between the evaluation of a robot with machine-like feedback and human-like feedback ($p < .001$), and with

Table 3 Results of the normality and equal variation tests in study 1

		Understandability			PWT			Service Evaluation		
		H	M	B	H	M	B	H	M	B
Normality	LCDT	Z(18) = .924 p = .154	Z(18) = .956 p = .526	Z(18) = .960 p = .592	Z(18) = .907 p = .077	Z(18) = .946 p = .366	Z(18) = .907 p = .076	Z(18) = .971 p = .808	Z(18) = .975 p = .891	Z(18) = .973 p = .846
	HCDT	Z(18) = .969 p = .773	Z(18) = .970 p = .790	Z(18) = .910 p = .086	Z(18) = .955 p = .504	Z(18) = .948 p = .400	Z(18) = .950 p = .425	Z(18) = .929 p = .184	Z(18) = .926 p = .162	Z(18) = .901 p = .061
Equal Variance		p = .503			p = .302			p = .591		

Note 1 PWT Perceived waiting time, LCDT Low cognitive demand task, HCDT High cognitive demand task, H Human-like feedback, M Machine-like feedback, B Baseline

Note 2 If the p-value is greater than 0.05, the data are normally distributed

Note 3 If the p-value is greater than 0.05, equal variance can be assumed

machine-like feedback and without feedback ($p < .001$). However, there was no difference between the evaluation of a robot with human-like feedback and without feedback ($p = .969$). Participants perceived waiting times shorter when interacting with a robot providing machine-like feedback compared to interacting with a robot without feedback or with human-like feedback.

H3 was partially supported by the data. The robot's feedback type had a significant effect on service evaluation ($M_{\text{human-like}} = 4.56, SD = 1.65$ vs. $M_{\text{machine-like}} = 5.44, SD = 1.00$ vs. $M_{\text{baseline}} = 4.55, SD = 1.20$; $F(2,68) = 17.064, p < .001$). As a result of the Post hoc analysis using a Tukey HSD test, there were significant differences between the evaluation of a robot with machine-like feedback and human-like feedback ($p < .001$), and with machine-like feedback and without feedback ($p < .001$). However, there was no difference between the evaluation of a robot with human-like feedback and without feedback ($p = .938$). Participants evaluated the service of a robot with machine-like feedback as more positive than a robot with human-like feedback and that with no feedback (Fig. 4).

3.4.2 Mediation Effect of PWT Between Response Feedback Type and Service Evaluation

H4 was supported by the data. A mediation analysis result indicated that the PWT fully mediated the influence of the response delay feedback type on service evaluation. The feedback type was affected by the PWT marginally ($b = 0.35, SE = 0.20, p < 0.1$). Moreover, the PWT was a significant positive predictor of service evaluation ($b = 0.67, SE = 0.13, p < 0.001$). After controlling by the mediator, the feedback type was no significant predictor of service evaluation ($b = -0.04, SE = 0.21, p = 0.8452$) (Fig. 5).

3.4.3 Effects of Response Feedback Type and Task Type on Understandability, PWT, and Service Evaluation

H5 was supported by the data. There was a significant interaction effect between feedback type and task type on understandability ($F(2,68) = 4.127, p = .025$). To be specific, there was significant interaction between feedback type (human-like feedback vs. machine-like feedback) and task type ($F = 8.442, p = .006$). However, there was no significant interaction between feedback type (machine-like feedback vs. baseline robot) and task type ($F = .317, p = .577$). In addition, there was no significant interaction between feedback type (human-like feedback vs. baseline robot) and task type ($F = 2.680, p = .111$). After conducting the Tukey HSD test, it was found that the robot with machine-like feedback provided users with a better understanding of its status ($M = 5.11, SD = 0.90$) compared to the robot with human-like feedback ($M = 3.89, SD = 1.36$) in LCDT ($p = .005$). However, in

HCDT, the difference between the robot with machine-like feedback and the robot with human-like feedback was not significant ($p = .993$).

H6 was supported by the data. There was a significant interaction effect between feedback type and task type on PWT ($F(2,68) = 4.174, p = .024$). To be specific, there was marginally significant interaction between feedback type (human-like feedback vs. machine-like feedback) and task type ($F = 3.133, p = .086$). In addition, there was significant interaction between feedback type (human-like feedback vs. baseline robot) and task type ($F = 8.423, p = .006$). However, there was no significant interaction between feedback type (machine-like feedback vs. baseline robot) and task type ($F = 1.322, p = .258$). The Tukey HSD verification indicated that the robot with machine-like feedback responded faster and reduced the PWT more effectively ($M = 3.65, SD = 1.16$) than the robot with human-like feedback ($M = 2.20, SD = 0.70$) in LCDT ($p = .001$). However, the difference between the robot with machine-like feedback and the robot with human-like feedback was not significant in HCDT ($p = .300$). Meanwhile, no significant difference was observed between the robot with human-like feedback and the baseline robot in both LCDT ($p = .166$) and HCDT ($p = .223$).

H7 was supported by the data. There was a significant interaction effect between feedback type and task type on service evaluation ($F(2,68) = 3.303, p = .049$). To be specific, there was significant interaction between feedback type (human-like feedback vs. machine-like feedback) and task type ($F = 5.883, p = .021$). In addition, there was significant interaction between feedback type (human-like feedback vs. baseline robot) and task type ($F = 5.808, p = .022$). However, there was no significant interaction between feedback type (machine-like feedback vs. baseline robot) and task type ($F = 0.105, p = .748$). After conducting the Tukey HSD test, it was found that the robot with machine-like feedback received more positive evaluations ($M = 5.06, SD = 1.03$) compared to the robot with human-like feedback ($M = 3.67, SD = 1.66$) in LCDT ($p = .010$). However, the difference between the robot with machine-like feedback and the robot with human-like feedback was not significant in HCDT ($p = .512$). Meanwhile, there was no significant difference between the robot with human-like feedback and the baseline robot in both LCDT ($p = .450$) and HCDT ($p = .158$) (Table 4; Fig. 6).

3.5 Discussion

This study explored how feedback type and task type affect impressions on robots in a response delay. Through the verification of H1, H2, and H3, we found that overall machine-like feedback was evaluated more positively than the other two feedback. The mediating effect of the PWT was confirmed between the feedback type and service evaluation. This sug-

Table 4 Statistical results in task-based interaction

Understandability			
Main effect	H ($M = 4.67, SD = 1.21$)	M ($M = 5.30, SD = 0.81$)	B ($M = 4.00, SD = 1.06$)
	H in LCDT ($M = 3.89, SD = 1.36$)	H in HCDT ($M = 5.44, SD = 1.07$)	B in LCDT ($M = 3.88, SD = 1.04$)
Interaction effect	Feedback type ($F = 15.383, p < .001$)	M in HCDT ($M = 5.49, SD = 0.72$)	B in HCDT ($M = 4.50, SD = 1.42$)
	H versus M ($F = 9.676, p = .004$)	H versus B ($F = 2.842, p = .101$)	M versus B ($F = 25.032, p < .001$)
PWT	Feedback type (H versus M versus B) x Task type (LCDT versus HCDT) ($F = 4.127, p = .025$)	H versus B ($F = 2.680, p = .111$)	M versus B ($F = .317, p = .577$)
	H versus M ($F = 8.442, p = .006$)	-	-
Main effect	H ($M = 3.19, SD = 1.50$)	M ($M = 4.22, SD = 1.22$)	B ($M = 3.20, SD = 1.35$)
	H in LCDT ($M = 2.20, SD = 0.70$)	M in LCDT ($M = 3.65, SD = 1.16$)	B in LCDT ($M = 2.91, SD = 1.45$)
Interaction effect	Feedback type ($F = 11.949, p < .001$)	H versus B ($F = 17.776, p < .001$)	M versus B ($F = 17.776, p < .001$)
	H versus M ($F = 19.061, p < .001$)	-	-
Service evaluation	Feedback type (H vs. M vs. B) x Task type (LCDT vs. HCDT) ($F = 4.174, p = .024$)	H versus B ($F = 8.423, p = .006$)	M versus B ($F = 1.322, p = .258$)
	H versus M ($F = 3.133, p = .086$)	H versus B in LCDT ($p = .166$)	-
Main effect	H ($M = 4.56, SD = 1.65$)	M in HCDT ($M = 5.81, SD = 0.83$)	B ($M = 4.55, SD = 1.20$)
	H in LCDT ($M = 3.67, SD = 1.66$)	M in LCDT ($M = 5.06, SD = 1.03$)	B in LCDT ($M = 4.22, SD = 1.36$)
Interaction effect	Feedback type ($F = 17.064, p < .001$)	H versus B ($F = .006, p = .938$)	M versus B ($F = 26.947, p < .001$)
	H versus M ($F = 16.576, p < .001$)	H versus B ($F = 5.808, p = .022$)	M versus B ($F = .105, p = .748$)
Interaction effect	Feedback type (H vs. M vs. B) x Task type (LCDT vs. HCDT) ($F = 3.303, p = .049$)	H versus B in LCDT ($p = .450$)	-
	H versus M ($F = 5.883, p = .021$)	H versus B in HCDT ($p = .158$)	-

Note 1 PWT Perceived waiting time, LCDT Low cognitive demand task, HCDT High cognitive demand task, H Human-like feedback, M Machine-like feedback, B Baseline

Fig. 4 Evaluation of robot feedback for response delay in task-based interaction. Note 1: Error bars indicate standard deviation

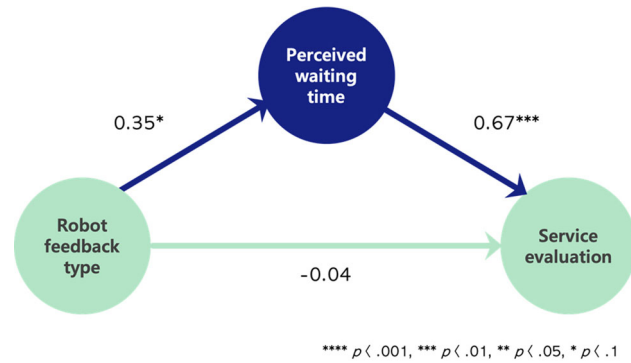
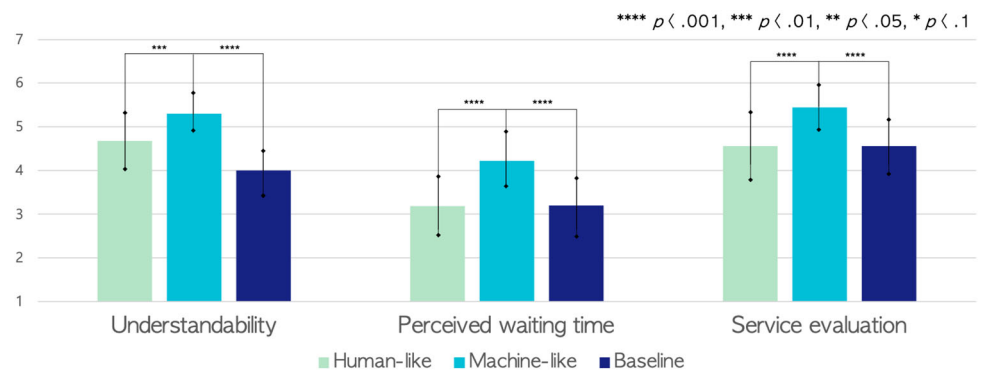
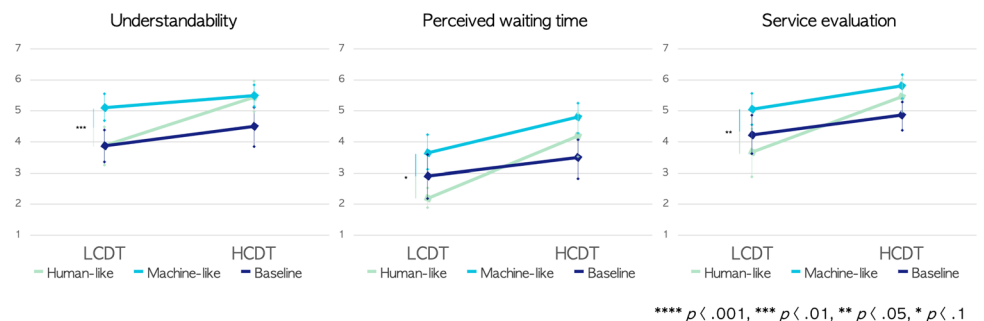


Fig. 5 Mediation effect of PWT between feedback type and service evaluation

gests that the robot feedback that can reduce the PWT should be designed to obtain a positive evaluation of the robot.

Surprisingly, we initially expected that a robot with feedback would be evaluated more positively than a baseline robot. However, the results showed that, in most cases, a robot with human-like feedback had no significant difference from a baseline robot in LCDT. Descriptive data also imply that in the situation of performing an LCDT, a robot with human-like feedback enabled the user to understand its state to a level similar to that of a robot without feedback. On the other hand, in the situation of performing an HCDT, it allowed users to comprehend its state at a level similar to machine-like feedback, which is perceived as making users understand its state the best. Furthermore, in the situation

Fig. 6 Interaction effect of feedback type and task type on understandability, PWT, and service evaluation. Note 1: LCDT: low cognitive demand task, HCDT: high cognitive demand task. Note 2: Error bars indicate standard deviation



of performing an LCDT, a robot with human-like feedback received lower ratings for PWT compared to a robot providing no feedback according to the descriptive data. Conversely, in the situation of performing an HCDT, a robot with human-like feedback was evaluated to respond faster than a baseline robot. These results suggest that the evaluation of robot feedback varied significantly depending on the task type, and the user’s perception toward inappropriate robot feedback was similar to that of receiving no feedback at all.

In this study, humans and robots had task-based interactions. The user gave a command to the robot, and the robot performed what the user commanded, similar to the interactions that humans already have with other machines. Therefore, the machine-like feedback may have been familiar to the robot user, and the user accepts a robot with machine-like feedback better than a robot with human-like feedback. Since a robot can have both the properties of a machine and that of a social entity, interaction with a robot can be social interaction as well as task-based interaction. Therefore, we additionally conducted an experiment to see what kind of feedback is preferred in social interaction.

4 Study 2: Social Interaction

Response delay occurs not only in human-computer or human-machine interactions but also in human-human interactions. In this case, humans consciously or unconsciously use fillers and gestures [65–67]. Studies related to the

response delay of a robot in conversation with a human were conducted. During conversation, the robot's backchannels such as nodding or saying "Umm" made HRIs more natural and enjoyable [68, 69].

In the first study, a robot performed its task that the user asked for, and the participants preferred a robot with machine-like feedback rather than a robot with human-like feedback and a baseline robot. Unlike task-based interaction, when a robot interacts with a human in a social interaction, it is possible that human-like feedback might be preferred to machine-like feedback because humans are accustomed to the social cues, which are used when socially interacting with other humans [27]. Furthermore, reducing uncertainty makes the PWT shorter [9, 10], and human-like characteristics reduce uncertainty about the behavior of nonhuman social agents and increase confidence in predicting what to do next [70]. For successful social interaction between a robot and a human, it is important for the robot to be recognized as a social being by humans. To be seen as a social being, the robot must communicate with humans socially in accordance with the human social norm and context [71] and achieve high-level social outcomes [72]. In addition, the high sociability of nonhuman social agents increases user satisfaction and system acceptance [73, 74]. Thus, human-like feedback makes the robot perceived as more social and can increase service evaluation. We predict the robot's social cues for response delay including conversational fillers and gestures can allow humans to understand the state of the robot better, can reduce the PWT, and increase perceived robot sociability, and user satisfaction.

4.1 Hypotheses

4.1.1 Hypothesis 8

In a social interaction, a robot with human-like feedback will help the user to better understand its processing state than a robot with machine-like feedback and a baseline robot. Breazeal [27] found that humans treat robots in a similar way as they treat humans. Moreover, Dittmann and Llewelyn [67], Butterworth and Beattie [65] argued that in human-human interactions, people use conversational fillers and gestures when they have response delays. Accordingly, it was predicted that users would be able to best understand the state of robots using conversational fillers and gestures when the robot has a response delay.

4.1.2 Hypothesis 9

A robot with human-like feedback will shorten the PWT rather than a robot with machine-like feedback and a baseline robot. This prediction follows Epley et al.'s finding, which revealed that human-like features reduce uncertainty about

the action of nonhuman social agents, and Hui and Zhou and Weinberg's findings, which showed that the PWT decreases when uncertainty is reduced [9, 10, 70].

4.1.3 Hypothesis 10

Human-like feedback will lead to an increase in the perceived robot's sociability. This assumption is based on Mutlu et al. [72]'s study that human-like robots achieve high levels of social performance.

4.1.4 Hypothesis 11

A robot with human-like feedback will be rated as more positive than a robot with machine-like feedback and a baseline robot. Our prediction follows the findings of a study by Bar-Cohen and Breazeal [29] that people evaluated human-like robots as more positive than machine-like robots when they interact with them socially.

4.1.5 Hypothesis 12

Sociability will mediate between the robot feedback type and product evaluation. Wang et al. [73] found that nonhuman social agents were evaluated more positively when they were perceived as more social. Therefore, if the robot's feedback enhances the perceived robot's sociality, we predict that this will increase the satisfaction with the robot.

To test the hypotheses, the participants experienced three types of robot feedback in a counterbalanced order.

4.2 Method

4.2.1 Participants

We recruited 36 participants who did not participate in Study 1, aged between the mid-20s to the mid-30s ($M = 27.25$, $SD = 2.65$; 19 males and 17 females). Participants received \$10 for the experiment.

4.2.2 Interaction Type

Weather talk has the property of neutrality and is frequently dealt with in human-to-human social interaction. It is also an infinitely expandable conversation topic [75]. Thus, in social interaction, the experiment was set up so that the participant and the robot talked about the weather. Participants asked the robot questions about the current weather and then talked about clothes and food appropriate for the weather. Since this experiment was conducted in late summer-autumn, it was assumed that the weather would be either sunny, cloudy, or rainy, as shown in Table 5.

Table 5 Examples of scripts by social interaction

	Social interaction
Participant	“Put the can in the answer box How is the weather today?” A: Sunny B: Cloudy C: Rainy
Robot	(Expressing feedback for response delay for 10 s) (Putting the can in the box that corresponds to the weather on that day.)
Participant	“Put the can in the answer box What should I wear to go out in the weather?” A: Short sleeve sweatshirt B: Long sleeve sweatshirt C: Sweatshirt with a coat
Robot	(Expressing feedback for response delay for 10 s) (Putting the can in the box corresponding to weather-appropriate clothing)

4.2.3 Measurement

Understandability (4 items, $\alpha = .760$) [61], PWT (3 items, $\alpha = .936$) [62] were measured as in the first study. In addition, we used a product evaluation scale (4 items, $\alpha = .955$) instead of the service evaluation used in the first study [76]. In the first study, to explore the effective delay feedback according to the task types, we investigated whether the user perceived that the robot had done the task right by measuring service evaluation. On the contrary, in the second study where we aimed to explore the effective robot feedback in social interaction, we found the overall impression of the robot by measuring product evaluation [76]. Moreover, since participants and robots had social conversations in this experiment, sociability was expected to have a great effect on the impression of the robot, which affected service evaluation. Thus, sociability was measured in the second study (4 items, $\alpha = .902$) [77] as shown in Table 6.

4.2.4 Robot Type and Procedure

The robot and procedure used in this experiment were the same as those used in the first experiment.

Table 6 Sociability measurement

Scale	Items
Sociability	I consider the robot a pleasant conversational partner I find the robot pleasant to interact with I feel the robot understands me I think the robot is nice (1: Strongly disagree—7: Strongly agree)

4.3 Results

As in the first experiment, the Shapiro-Wilk test was performed to check the normality of the acquired data. Our data met normality. In addition, as a result of the equal variance test, our data satisfied equal variance except for understandability. The detailed normality test and equal variation test results are shown in Table 7.

Therefore, we analyzed the data by performing a non-parametric statistical analysis method, Welch’s ANOVA and Dunnett’s test for understandability, a parametric statistical analysis method, one-way ANOVA, and Tukey’s method for other measures. In addition, to test the mediation effect, we used PROCESS model 4 of the SPSS Process macro to evaluate the mediation effects [64].

4.3.1 Effects of Response Feedback on Understandability, PWT, Sociability, and Product Evaluation

H8 was partially supported by the data. There was a significant effect of the robot’s feedback type on understandability ($M_{\text{human-like}} = 5.43$, $SD = 0.84$ vs. $M_{\text{machine-like}} = 4.99$, $SD = 0.92$ vs. $M_{\text{baseline}} = 4.31$, $SD = 1.45$; $F(53) = 4.765$, $p = .013$). The Post-hoc analysis using Dunnett’s test revealed that there was a significant difference between the evaluation of a robot with human-like feedback and a robot without feedback ($p = .010$). A robot with human-like feedback was evaluated as more understandable than a robot without feedback. However, there was no significant difference between the evaluation of a robot with machine-like feedback and without feedback ($p = .163$) and between the evaluation of a robot with machine-like feedback and human-like feedback ($p = .452$).

H9 was partially supported by the data. The significant effect of the robot’s feedback type on the PWT was observed ($M_{\text{human-like}} = 4.06$, $SD = 1.25$ vs. $M_{\text{machine-like}} = 3.83$, $SD = 1.58$ vs. $M_{\text{baseline}} = 2.43$, $SD = 1.24$; $F(53) = 7.533$, $p = .001$). In Post hoc comparisons using a Tukey HSD test, there were significant differences between the evaluation of a robot with human-like feedback and without feedback ($p =$

Table 7 Results of the normality and equal variation tests in study 2

	Understandability			PWT			Sociability			Product evaluation		
	H	M	B	H	M	B	H	M	B	H	M	B
	Normality	Z(18) = .944 p = .337	Z(18) = .914 p = .100	Z(18) = .959 p = .590	Z(18) = .908 p = .094	Z(18) = .908 p = .218	Z(18) = .933 p = .098	Z(18) = .913 p = .098	Z(18) = .933 p = .217	Z(18) = .968 p = .753	Z(18) = .930 p = .196	Z(18) = .930 p = .372
Equal variance	p = .035			p = .594				p = .993			p = .221	

Note 1 PWT: perceived waiting time, H: human-like feedback, M: machine-like feedback, B: baseline.

Note 2 If the p-value is greater than 0.05, the data are normally distributed.

Note 3 If the p-value is greater than 0.05, equal variance can be assumed.

.002), and the evaluation of a robot with machine-like feedback and without feedback ($p = .009$). However, there was no significant difference between the evaluation of a robot with machine-like feedback and human-like feedback ($p = .877$). It was evaluated that a robot with human-like feedback and a robot with machine-like feedback responded faster than a robot without feedback.

H10 was partially verified by the data. There was a significant effect of the robot’s feedback type on sociability ($M_{human-like} = 4.94, SD = 1.06$ vs. $M_{machine-like} = 4.18, SD = 1.14$ vs. $M_{baseline} = 2.99, SD = 1.26; F(53) = 13.060, p < .001$). As a result of the Post hoc analysis using a Tukey HSD test, there were significant differences between the evaluation of a robot with human-like feedback and without feedback ($p < .001$), and the evaluation of a robot with machine-like feedback and without feedback ($p = .009$). However, there was no significant difference between the evaluation of a robot with machine-like feedback and human-like feedback ($p = .128$). A robot with human-like feedback and a robot with machine-like feedback were perceived as more social than the baseline robot.

H11 was partially supported by the data. There was a significant effect of the robot’s feedback type on product evaluation ($M_{human-like} = 5.29, SD = 0.89$ vs. $M_{machine-like} = 4.93, SD = 0.94$ vs. $M_{baseline} = 4.04, SD = 1.23; F(53) = 7.014, p = .002$). The Post hoc analysis using a Tukey HSD test revealed that there were significant differences between the evaluation of a robot with human-like feedback and without feedback ($p = .002$), and between a robot with machine-like feedback and without feedback ($p = .033$). However, there was no significant difference between the evaluation of a robot with machine-like feedback and human-like feedback ($p = .548$). A robot with human-like feedback and a robot with machine-like feedback were rated as more positive than a baseline robot (Fig. 7).

Overall, although human-like feedback compared to baseline was evaluated positively, there was no significant difference between machine-like feedback and human-like feedback (Table 8).

4.3.2 Mediation Effect of Sociability Between Response Feedback Type and Product Evaluation

H12 was supported by the data. There was a mediation effect of sociability between the robot feedback type and service evaluation. The feedback type affected the perceived robot sociability ($b = 0.98, SE = 0.19, p < 0.001$). Furthermore, sociability was a significant positive predictor of product evaluation ($b = 0.46, SE = 0.10, p < 0.001$). After controlling by the mediator, the feedback type was no significant predictor of product evaluation ($b = -0.02, SE = 0.16, p = 0.8901$). Thus, a mediation analysis result indicated that sociability

Fig. 7 Evaluation of robot feedback for response delay in social interaction. Note 1: Error bars indicate standard deviation

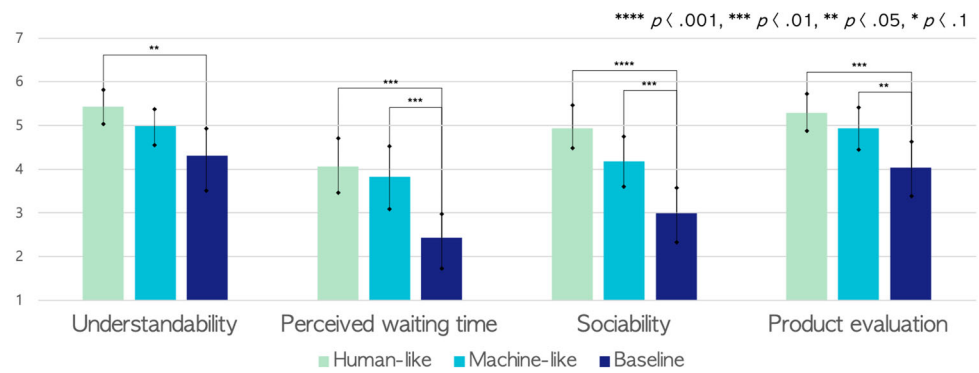


Table 8 Statistical results in social interaction

	Human-like	Machine-like	Baseline
<i>Understandability</i>			
Main effect	H ($M = 5.43, SD = 0.84$)	M ($M = 4.99, SD = 0.92$)	B ($M = 4.31, SD = 1.45$)
	H vs. M vs. B ($F = 4.765, p = .013$)		
	H vs. M ($p = .452$)	H vs. B ($p = .010$)	M vs. B ($p = .163$)
<i>PWT</i>			
Main effect	H ($M = 4.06, SD = 1.25$)	M ($M = 3.83, SD = 1.58$)	B ($M = 2.43, SD = 1.24$)
	H vs. M vs. B ($F = 7.533, p = .001$)		
	H vs. M ($p = .877$)	H vs. B ($p = .002$)	M vs. B ($p = .009$)
<i>Sociability</i>			
Main effect	H ($M = 4.94, SD = 1.06$)	M ($M = 4.18, SD = 1.14$)	B ($M = 2.99, SD = 1.26$)
	H vs. M vs. B ($F = 13.060, p < .001$)		
	H vs. M ($p = .128$)	H vs. B ($p < .001$)	M vs. B ($p = .009$)
<i>Product evaluation</i>			
Main effect	H ($M = 5.29, SD = 0.89$)	M ($M = 4.93, SD = 0.94$)	B ($M = 4.04, SD = 1.23$)
	H vs. M vs. B ($F = 7.014, p = .002$)		
	H vs. M ($p = .548$)	H vs. B ($p = .002$)	M vs. B ($p = .033$)

Note 1 PWT: perceived waiting time, H: human-like feedback, M: machine-like feedback, B: baseline

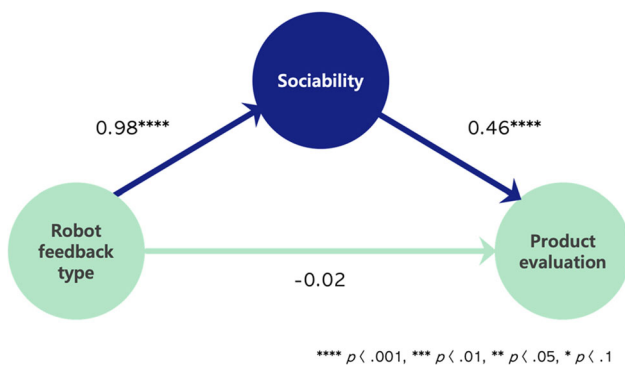


Fig. 8 Mediation effect of sociability between response feedback type and product evaluation

fully mediated the influence of the robot feedback type on product evaluation (Fig. 8).

4.4 Discussion

All the hypotheses of Study 2, which investigated the effect of robot feedback types on user perception towards the robot

in social interaction, were partially supported. However, the effect of machine-like feedback, which was overwhelmingly positively evaluated compared to human-like feedback, was weakened. In addition, the human-like feedback that was rated lower than the baseline in terms of service evaluation in Study 1 was evaluated positively compared to the baseline in all measurements in Study 2.

We predicted that a robot with human-like feedback would be more effective than other robots in social interaction based on literature review. Even though a robot with human-like feedback was evaluated positively compared to a baseline robot which had no feedback, the difference between a robot with machine-like feedback and a robot with human-like feedback was not significant. This may have been influenced by the appearance of the robot. In general, social robots have a human-like appearance with a face and body, such as PR2 [47], Pepper [78], QTrobot [79], and Quori [80]. However, the robot used in this paper was a robotic arm modeled after a part of the human body. Because of the dissonant between the robot’s non-anthropomorphic appearance and the human-

like feedback, human-like feedback may not be accepted as meaningfully by users as we expected.

5 General Discussion

5.1 Summary and Interpretation of Results

The first study investigated robot feedback for response delay that is acceptable to a user when having task-based interactions with a robot. In task-based interaction, the robot with machine-like feedback was evaluated as more positive than the other two robots, the robot with human-like feedback and the baseline robot, regardless of the task it performed. Furthermore, the PWT positively mediated the robot's feedback type and service evaluation in task-based interaction. Besides, the evaluation between the robot with human-like feedback and the baseline robot was different depending on the task type. When performing LCDT, the service of the robot with machine-like feedback was evaluated as more positive than the robot with human-like feedback. By contrast, when performing an HCDD, there was no significant difference between the service provided by a robot with human-like feedback and that with machine-like feedback.

The second study investigated robot feedback for response delay accepted by users when humans and robots interact socially. As a result, overall, robots giving feedback were evaluated more positively than the baseline robot, which did not give any feedback. The result of this study revealed no significant difference between human-like feedback and machine-like feedback for understandability, PWT, sociability, and product evaluation. However, descriptive data imply that unlike in the first study, a robot with human-like feedback was the highest on most of the scales including understandability, PWT, sociability, and product evaluation. In addition, the robot's sociability positively mediated the robot's feedback type and service evaluation in social interaction.

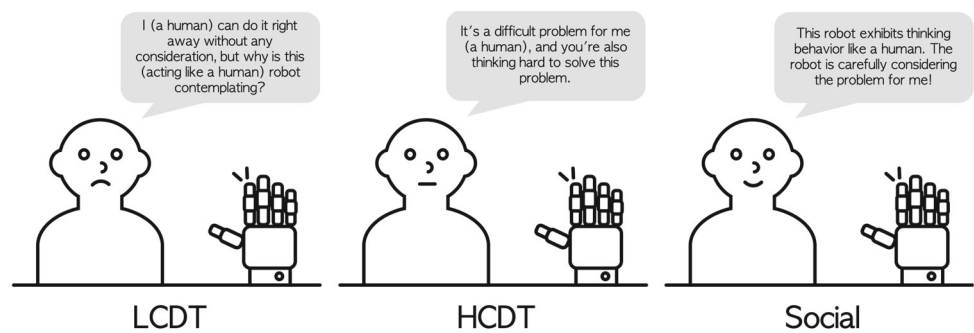
These results suggest that the appropriate robot feedback for response delay differs depending on the role the robot plays. As a machine that executes human commands, the robot's machine-like feedback was most positively accepted by the user. This implies that because people are already familiar with the way they interact with machines, machine-like feedback is accepted for robots in task-based interaction. In contrast, when interacting with users socially, a robot's human-like feedback was evaluated more positively rather than when performing tasks. It is inferred that the human-like feedback was accepted by users most by functioning as a social cue because people use social cues when they socially interact with other people.

Regardless of interaction types, either task-based interaction or social interaction, when interacting with a baseline robot during two experiments, most users did not understand

that the robot was in processing and asked the experimenter, "Should I say it again?" "I think the robot is broken. It is not doing anything." Nevertheless, in the case of the robot performing an LCDT, a baseline robot was evaluated more positively than a robot with human-like feedback. In the case of LCDT, since it is a task that humans can immediately execute without requiring time for additional complicated calculations and action planning, a robot providing human-like feedback that merely taps the table with a finger and makes an "um" sound without promptly executing the requested task may give the impression that the robot is not intentionally fulfilling the user's request. This perception of intentional delay by the robot, as if it does not want to work, can increase the PWT and have a negative impact on the overall service evaluation. Previous studies have demonstrated that the configuration of a robot's mental attribution can influence users' perceptions of the robot's internal state and their reactions to it, even when the robot displays the same expression [81, 82]. When interpreting this study in this context, users tend to perceive the robot's expression as reflecting its internal state if they believe there is a match. However, if users perceive a misalignment between the robot's expression and its internal state, they may develop doubts about the robot's mental attribution. In such cases, users might attribute a special intention or hidden agenda to the robot's expression. The study's results suggest that whether the user perceives alignment between the robot's expression and its internal state may depend on the type of task the robot is performing.

The amount of information provided by human-like feedback and machine-like feedback differed. Machine-like feedback included information about the time taken for the robot to make a decision, which is not present in human-like feedback. In other words, machine-like feedback was more informative regarding the response delay compared to human-like feedback. However, when a user engages in social interaction with a robot, machine-like feedback does not enhance the user's understanding of the robot's state or reduce the PWT. This implies that the user's understanding of the robot's state might not be solely determined by the amount of information provided by the robot. In social interactions, the robot's ability to leverage social cues as a social entity can be just as important as the computational information provided by the robot as a machine in aiding the user to understand the robot's state.

When combining the results of Study 1 and Study 2, it became evident that the robot's feedback was evaluated differently based on the task it performed. In the scenario where the robot executed the LCDT, the understandability of human-like feedback was assessed at a level similar to that of a robot without any feedback. Additionally, the PWT of human-like feedback was rated lower than that of a robot without any feedback. This implies that when the robot deals

Fig. 9 Interpretation of results

with the LCDT, which generally does not require considerable effort for a human to solve it, if the robot shows a thinking gesture, the robot's thinking gesture would hinder the user's understandability of the robot's state. On the other hand, in the scenario of performing HCDT, the robot with human-like feedback received significantly higher evaluations in terms of understandability compared to the robot without any feedback and was rated similarly to the robot with machine-like feedback. This can be interpreted as the human-like feedback, expressing thinking behavior, effectively helps users understand the robot's state which is solving complex problems, even on par with the machine-like feedback, which provides more information. In social interactions, there was a trend that the robot with human-like feedback received the most positive evaluations in the data, unlike in task-based interactions. In task-based interactions, it can be crucial for the robot to inform the user about its progress. This is because, in task-based interaction, the user's goal is clear, which is to complete the given task, and there is a specific answer to achieve the user's goal usually. In such cases, the robot can calculate how much time is needed to complete the task and efficiently inform the user of its progress. However, social questions often lack definitive answers. In such cases, offering social cues through human-like feedback might become effective in conveying the robot's attentiveness and concern for the user. This because, in social interactions, when the robot encounters a question that has no fixed or predefined response, the robot's human-like thinking behavior can be perceived as a careful consideration to provide valuable advice to the user (Fig. 9).

5.2 Implications for Research

This paper suggests a user-centered design for robot response delay feedback. User-centered design research on the feedback for response delay of robots is less explored than that of web pages and applications. Unlike devices such as computers, smartphones, and tablet PCs, the appearance and interaction methods of robots are becoming increasingly similar to humans due to the development of technology. Accordingly, users can expect human-level response speed

from the robots. However, robots do not yet have human-level response speed in interactions that are not preprogrammed. Therefore, we conducted this experiment to alleviate the negative impression of the robot caused by the robot's response delay that does not meet the user's expectations.

We explored factors that influenced users' expectations by focusing on robot feedback design and task (e.g., a condition that increases user expectation-i.e., LCDTs and human-like feedback vs. a condition that lowers user expectation-HCDTs and machine-like feedback). Moreover, we investigated the effects of these variables on the impression made by the robots in the first study. By comparing these variables through user evaluation, we found that providing inappropriate feedback can be worse than providing no feedback at all. Therefore, the study paves the way for robot feedback design considering user expectations.

In addition, we found that the appropriate feedback type can vary depending on the role the robot plays. When the user interacts with the robot as a being that executes the user's commands, the robot's machine-like feedback was positively evaluated. By contrast, when the user interacted with the robot as a social interactant, the robot's human-like feedback was evaluated positively. Accordingly, it will be a cornerstone that opens the way for research on the various state expression methods of robots depending on their various roles.

5.3 Implications for Robot Design

When performing an action-based task, a robot inevitably has a response delay. Realistically, there is a limit for some robots to have the response speed that satisfies the user. Therefore, our research explores variables that mitigate users' negative emotions toward the response delay of robots.

Robot feedback can reduce or increase the PWT during the same waiting time. We found that the PWT affects service evaluation in task-based interaction. It would be best if the actual user's waiting time could be reduced through technological development, but technology development to reduce the robot's response delay time requires a lot of money, time, and effort. If the response delay cannot be managed immedi-

ately, users will give up interacting with the robots, making them unviable. Therefore, for a robot to be accepted in the market, it is suggested that when it has a response delay, the interaction design that reduces the PWT should be considered first in task-based interaction.

Furthermore, in social interaction, it is suggested that the robot's feedback is designed as human-like. When robots interact with users as social beings, not as tools or machines, human-like expressions could improve the robot's sociability and service evaluation. We confirmed that when the robot functions as a social interactant, its sociability has a positive impact on user satisfaction, as tested in H12. Therefore, when developing a robot that provides social interaction to the user, the interaction design that enhances the robot's sociability should be carefully considered.

5.4 Limitations

First, the shape of the robot we used in this study was a mechanical arm. This may have had an impact on evaluating the type of feedback that is effective for the robot. In future study, if machine-like feedback and human-like feedback for a robot with a human-like appearance are compared and analyzed, results can be different from those of this study.

Second, for rigorous experimental manipulation, the robot's action was limited to grabbing a can, moving it, and putting it into one of the three boxes. The way the robot responds to the user's question in social interaction was not as natural as in real social interaction. Therefore, there is a limit to applying the current experimental results to social robot development directly. If diverse robot actions are designed according to the situation, it will be possible to generalize the study results and suggest design guidelines that are applicable to robot development.

Third, the number of participants and our participant pool were limited. To develop universally commercialized robots, the end-users' perception of robots should be investigated. The results of the study would be more generalized if larger number of participants with various ages and levels of knowledge can be recruited.

Forth, if we study feedback on response delay in more diverse tasks, and use more detailed service evaluation and understandability evaluation measurements in the future, the results of the study can be extended and expanded and can propose more sophisticated design guidelines for robot response delay feedback.

Fifth, apart from the distinction between human-likeness and machine-likeness, various other factors can influence the user's perception of the robot's response delay. For instance, comparing the hedonic strategy, which offers a gaming service to the user, with the practical strategy, which provides information about the current level of action planning while

awaiting the robot's response, we can identify the HRI strategy that effectively reduces PWT and enhances the overall evaluation of the robot. If these response delay-related factors that can additionally impact the user's perception of the robot are explored in diverse ways, more sophisticated HRI guidelines for robot response delay feedback can be developed.

Finally, a short-term experiment was executed in a laboratory setting. Long-term experiments conducted in a natural environment are desirable for future work. Furthermore, some of the statistical results in this study exhibit marginal significance. Given the variability in opinions regarding the reporting of marginally significant statistical values, future work which investigate factors that can substantially influence the marginal segment may benefit from addressing this issue. Such investigations could be expected to enhance the robustness and interpretability of the findings in this paper.

Acknowledgements We would like to thank Daseul Yang for his technical support.

Author Contributions Dr DK—Conceptualization, Methodology, Formal analysis, Data collection, Original Draft Writing, Manuscript Editing. Pfc CN—Conceptualization, Methodology. Dr SSK—Conceptualization, Methodology, Manuscript Editing, Supervision.

Funding This work was supported by the Korea Institute of Science and Technology (KIST) Institutional Program under Grant (2E32302).

Declarations

Conflicts of interest The authors declare that they have no conflict of interest.

Ethical Approval The Institutional Review Board (IRB) of the Korea Institute of Science and Technology approved the study (IRB approval number: 2021-049).

Consent to Participate Informed consent was obtained from the participants to whom the possible consequences of the studies were fully explained.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Luo S, Cheng M, Ding R, Wang F, Xu B, Chen B (2022) Human-robot shared control based on locally weighted intent prediction for a teleoperated hydraulic manipulator system. *IEEE/ASME Trans Mechatron*. <https://doi.org/10.1109/TMECH.2022.3157852>
- Rajapaksha U, Jayawardena C, MacDonald BA (2022) Design, implementation, and performance evaluation of a web-based multiple robot control system. *J Robot* 2022:55. <https://doi.org/10.1155/2022/9289625>
- Guynes JL (1988) Impact of system response time on state anxiety. *Commun ACM* 31(3):342–347
- Dellaert BG, Kahn BE (1999) How tolerable is delay?: Consumers' evaluations of internet web sites after waiting. *J Interact Mark* 13(1):41–54
- Hui MK, Tse DK (1996) What to tell consumers in waits of different lengths: an integrative model of service evaluation. *J Mark* 60(2):81–90
- Larson R, Larson B, Katz K (1991) Prescription for waiting-in line blues: entertain, enlighten and engage. *Sloan Manag Rev* 32(2):44–55
- Taylor S (1994) Waiting for service: the relationship between delays and evaluations of service. *J Mark* 58(2):56–69
- Osuna EE (1985) The psychological cost of waiting. *J Math Psychol* 29(1):82–105
- Hui MK, Zhou L (1996) How does waiting duration information influence customers' reactions to waiting for services? 1. *J Appl Soc Psychol* 26(19):1702–1717
- Weinberg BD (2000) Don't keep your internet customers waiting too long at the (virtual) front door. *J Interact Mark* 14(1):30–39
- Nah FF-H (2004) A study on tolerable waiting time: how long are web users willing to wait? *Behav Inf Technol* 23(3):153–163
- Lee YG, Chen AN, Hess T (2017) The online waiting experience: using temporal information and distractors to make online waits feel shorter. *J Assoc Inf Syst* 18(3):1
- Sherwin K (2014) Progress indicators make a slow system less insufferable. <https://www.nngroup.com/articles/progress-indicators/>. Accessed 22 Oct 2023
- Nielsen J (1994) Usability engineering. Morgan Kaufmann, San Francisco
- Card SK, Robertson GG, Mackinlay JD (1991) The information visualizer, an information workspace. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 181–186
- Miller RB (1968) Response time in man-computer conversational transactions. In: *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I*, pp. 267–277
- Liikkanen LA, Gómez PG (2013) Designing interactive systems for the experience of time. In: *Proceedings of the 6th international conference on designing pleasurable products and interfaces*, pp. 146–155
- Rose GM, Evaristo R, Straub D (2003) Culture and consumer responses to web download time: a four-continent study of mono and polychronism. *IEEE Trans Eng Manag* 50(1):31–44
- Rose GM, Straub DW (2001) The effect of download time on consumer attitude toward the e-service retailer. *e-Service* 1(1):55–76
- Dabrowski J, Munson EV (2011) 40 years of searching for the best computer system response time. *Interact Comput* 23(5):555–564
- Rich C, Ponsler B, Holroyd A, Sidner CL (2010) Recognizing engagement in human-robot interaction. In: *2010 5th ACM/IEEE international conference on human-robot interaction (HRI)*, pp. 375–382. IEEE
- Yamamoto M, Watanabe T (2003) Time lag effects of utterance to communicative actions on robot-human greeting interaction. In: *The 12th IEEE international workshop on robot and human interactive communication*, 2003. *Proceedings. ROMAN 2003.*, pp. 217–222. IEEE
- Kanda T, Kamasima M, Imai M, Ono T, Sakamoto D, Ishiguro H, Anzai Y (2007) A humanoid robot that pretends to listen to route guidance from a human. *Auton Robot* 22(1):87–100
- Goble H, Edwards C (2018) A robot that communicates with vocal fillers has... uhhh... greater social presence. *Commun Res Rep* 35(3):256–260
- Shiwa T, Kanda T, Imai M, Ishiguro H, Hagita N (2009) How quickly should a communication robot respond? Delaying strategies and habituation effects. *Int J Soc Robot* 1(2):141–155
- Wigdor N, de Greeff J, Looije R, Neerincx MA (2016) How to improve human-robot interaction with conversational fillers. In: *2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN)*, pp. 219–224. IEEE
- Breazeal C (2003) Toward sociable robots. *Robot Auton Syst* 42(3–4):167–175
- Zlatev J (2001) The epigenesis of meaning in human beings, and possibly in robots. *Mind Mach* 11(2):155–195
- Bar-Cohen Y, Breazeal C (2003) Biologically inspired intelligent robots. *Smart Struct Mater 2003 Electroact Polym Actuators Devices* 5051:14–20
- Hegel F, Krach S, Kircher T, Wrede B, Sagerer G (2008) Understanding social robots: a user study on anthropomorphism. In: *RO-MAN 2008-The 17th IEEE international symposium on robot and human interactive communication*, pp. 574–579. IEEE
- Walters ML, Koay KL, Syrdal DS, Dautenhahn K, Te Boekhorst R (2009) Preferences and perceptions of robot appearance and embodiment in human-robot interaction trials. *Procs of New Frontiers in Human-Robot Interaction*
- Kwak SS, Kim JS, Choi JJ (2017) The effects of organism-versus object-based robot design approaches on the consumer acceptance of domestic robots. *Int J Soc Robot* 9(3):359–377
- Woods S (2006) Exploring the design space of robots: children's perspectives. *Interact Comput* 18(6):1390–1418
- Dautenhahn K (2013) Human-robot interaction. *The Encyclopedia of Human-Computer Interaction*, 2nd Ed
- Lee H, Kang H, Kim M-G, Kwak JLSS (2016) The impact of robot design approach on product evaluation, product usefulness and purchase intention. *Adv Sci Technol Lett* 129:93–97
- Minato T, Shimada M, Ishiguro H, Itakura S (2004) Development of an android robot for studying human-robot interaction. In: *International conference on industrial, engineering and other applications of applied intelligent systems*, pp. 424–434. Springer
- Hinds PJ, Roberts TL, Jones H (2004) Whose job is it anyway? A study of human-robot interaction in a collaborative task. *Hum-Comput Interact* 19(1–2):151–181
- Woods S, Dautenhahn K, Schulz J (2004) The design space of robots: investigating children's views. In: *RO-MAN 2004*. In: *13th IEEE international workshop on robot and human interactive communication (IEEE Catalog No. 04TH8759)*, pp. 47–52. IEEE
- Zlotowski J, Sumioka H, Nishio S, Glas DF, Bartneck C, Ishiguro H (2016) Appearance of a robot affects the impact of its behaviour on perceived trustworthiness and empathy. *Paladyn J Behav Robot*. <https://doi.org/10.1515/pjbr-2016-0005>
- Lee D, Kwak S, Kim M (2007) Application of biological signal as an idle feedback for emotion expression of robot. *Arch Des Res* 20(4):17–28
- Ohshima N, Kimijima K, Yamato J, Mukawa N (2015) A conversational robot with vocal and bodily fillers for recovering from awkward silence at turn-takings. In: *2015 24th IEEE international symposium on robot and human interactive communication (RO-MAN)*, pp. 325–330. IEEE
- Parsonage G, Horton M, Read JC (2020) Designing experiments for children and robots. In: *Proceedings of the 2020 ACM interaction design and children conference: extended abstracts*, pp. 169–174

43. Goetz J, Kiesler S, Powers A (2023) Matching robot appearance and behavior to tasks to improve human-robot cooperation. In: The 12th IEEE international workshop on robot and human interactive communication, 2003. Proceedings. ROMAN 2003. pp. 55–60, IEEE
44. Smith MS, Stein MK (1998) Reflections on practice: selecting and creating mathematical tasks: from research to practice. *Math Teach Middle Sch* 3(5):344–350
45. Van den Walle J, Karp K, Bay-Williams J (2013) Elementary and middle school mathematics teaching developmentally, Eight. Pearson Educations, Upper Saddle River, NJ
46. Meeussen W, Wise M, Glaser S, Chitta S, McGann C, Mihelich P, Marder-Eppstein E, Muja M, Eruhimov V, Foote T et al. (2010) Autonomous door opening and plugging in with a personal robot. In: 2010 IEEE international conference on robotics and automation, pp. 729–736. IEEE
47. PR2 Kernel Description. <https://www.wevolver.com/article/seeing-through-a-robots-eyes-helps-those-with-profound-motor-impairments>. Accessed 10 Aug 2022
48. AlphaGo Kernel Description. <https://www.youtube.com/watch?v=vFr3K2DORc8>. Accessed 10 Aug 2022
49. Brun W, Teigen KH (1988) Verbal probabilities: ambiguous, context-dependent, or both? *Organ Behav Hum Decis Process* 41(3):390–404
50. Abbas T, Gadiraju U, Khan V-J, Markopoulos P (2021) Making time fly: Using fillers to improve perceived latency in crowd-powered conversational systems. In: Proceedings of the AAAI conference on human computation and crowdsourcing, vol. 9, pp. 2–14
51. Berg S, Clay D et al (1988) *Oedipus the King*. Oxford University Press, New York
52. Tolkien JRR (2007) *The hobbit*. Houghton Mifflin, Boston
53. Rust RT, Cooil B (1994) Reliability measures for qualitative data: theory and implications. *J Mark Res* 31(1):1–14
54. Smith VL, Clark HH (1993) On the course of answering questions. *J Mem Lang* 32(1):25–38
55. Kim H, Kang B (2011) Gender and usage of discourse markers in spoken korean. In: Proceedings of the 16th conference of pan Pacific association of applied linguistics. Chinese University of Hong Kong, Hong Kong, pp. 30–31
56. De Leeuw E (2007) Hesitation markers in English, German, and Dutch. *J Ger Linguist* 19(2):85–114
57. Yuan J, Xu X, Lai W, Liberman M (2016) Pauses and pause fillers in mandarin monologue speech: the effects of sex and proficiency. *Proc Speech Prosody* 2016:1167–1170
58. Gen2 Kinova. <https://www.kinovarobotics.com/en/products/gen2-robot>. Accessed 10 Aug 2022
59. ROS. <https://www.ros.org/>. Accessed 10 Aug 2022
60. Riek LD (2012) Wizard of oz studies in hri: a systematic review and new reporting guidelines. *J Hum-Robot Interact* 1(1):119–136
61. Moshkina L (2012) Reusable semantic differential scales for measuring social response to robots. In: Proceedings of the workshop on performance metrics for intelligent systems, pp. 89–94
62. Chen AN, Lee Y, Hwang Y (2018) Managing online wait: designing effective waiting screens across cultures. *Inf Manag* 55(5):558–575
63. Lee MK, Kiesler S, Forlizzi J, Srinivasa S, Rybski P (2010) Gracefully mitigating breakdowns in robotic services. In: 2010 5th ACM/IEEE international conference on human-robot interaction (HRI), pp. 203–210. IEEE
64. Hayes AF (2013) *Mediation, moderation, and conditional process analysis. Introduction to mediation, moderation, and conditional process analysis: a regression-based approach* edn. Guilford Publications, New York
65. Butterworth B, Beattie G (1978) Gesture and silence as indicators of planning in speech. In: *Recent advances in the psychology of language*, pp. 347–360. Springer, Boston
66. Campbell R (1978) *Recent advances in the psychology of language: formal and Experimental Approaches*. Plenum Press, NewYork
67. Dittmann AT, Llewellyn LG (1969) Body movement and speech rhythm in social conversation. *J Pers Soc Psychol* 11(2):98
68. Inden B, Malisz Z, Wagner P, Wachsmuth I (2014) Micro-timing of backchannels in human-robot interaction
69. Yu G (2019) *Effects of timing on users' perceived control when interacting with intelligent systems*. PhD thesis, University of Cambridge
70. Epley N, Waytz A, Cacioppo JT (2007) On seeing human: a three-factor theory of anthropomorphism. *Psychol Rev* 114(4):864
71. Bartneck C, Forlizzi J (2004) A design-centred framework for social human-robot interaction. In: RO-MAN 2004. 13th IEEE international workshop on robot and human interactive communication (IEEE Catalog No. 04TH8759), pp. 591–594. IEEE
72. Mutlu B, Kanda T, Forlizzi J, Hodgins J, Ishiguro H (2012) Conversational gaze mechanisms for humanlike robots. *ACM Trans Interact Intell Syst (TiiS)* 1(2):1–33
73. Wang J, Yang H, Shao R, Abdullah S, Sundar SS (2020) Alexa as coach: leveraging smart speakers to build social agents that reduce public speaking anxiety. In: Proceedings of the 2020 CHI conference on human factors in computing systems, pp. 1–13
74. De Graaf MM, Allouch SB (2013) Exploring influencing variables for the acceptance of social robots. *Robot Auton Syst* 61(12):1476–1486
75. Coupland N, Yi'anne-McEwen, V (2014) *Talkabout-theweather: Smalltalk, leisuretalkandthetravelindustry*. In: *SmallTalk*, pp.163–182. Routledge
76. Zhao, M., Hoeffler, S., Dahl, D.W. (2009) The role of imagination-focused visualization on new product evaluation. *J Market Res* 46(1): 46–55
77. Heerink M, Kröse B, Evers V, Wielinga B The influence of social presence on acceptance of a companion robot by older people
78. Pepper. <https://www.aldebaran.com/en/pepper>. Accessed 22 Dec 2022
79. QTrobot. <https://luxai.com/>. Accessed 22 Dec 2022
80. Quori. <http://www.quori.org/about>. Accessed 22 Dec 2022
81. Duarte NF, Raković M, Tasevski J, Coco MI, Billard A, Santos-Victor J (2018) Action anticipation: reading the intentions of humans and robots. *IEEE Robot Autom Lett* 3(4):4132–4139
82. Zonca J, Folsø A, Sciutti A (2021) The role of reciprocity in human-robot social influence. *Iscience* 24(12):103424

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Dahyun Kang received B.A. degrees from the Department of Industrial Design and Department of Consumer Science at Ewha W. University, South Korea, in 2013. She received her M.S. and Ph.D. degrees from Ewha W. University in 2016 and 2019, respectively. She is currently a Post Doc. Researcher in the Center for Intelligent and Interactive Robotics at the Korea Institute of Science and Technology. Her research interests include robot UX design and robotic product design.

Changjoo Nam received the B.S. and M.S. degrees in electrical engineering from Korea University, Seoul, South Korea, in 2009 and 2011, respectively, and the Ph.D. degree in computer science from Texas A&M University, College Station, TX, USA, in 2016. He is an Assistant Professor with the Department of Electronic Engineering at Sogang University. His research interest includes task and motion planning for multiple mobile robots and manipulators for their coordination.

Sonya S. Kwak received a B.S. degree from the Department of Industrial Design at KAIST, South Korea, in 2001. She received her M.S. and Ph.D. degrees from KAIST in 2003 and 2008, respectively. She is currently a Senior Research Scientist in the Center for Intelligent and Interactive Robotics at the Korea Institute of Science and Technology. Her research interests include human-robot interaction design and robotic product design.