Human to Robot Demonstrations of Routine Home Tasks: Acknowledgment and Response to the Robot's Feedback

Aris Alissandrakis and **Yoshihiro MIYAKE**[1]

Abstract. This paper investigates the possible role of the robot's feedback in Human-Robot Interaction (HRI) from the *human perspective*, and attempts to highlight some important conceptual and practical issues such as the lack of explicitness and consistency on people's demonstration strategies. More specifically, any changes that can be expected on the part of a human (teacher), in the teaching of a task, when a robot (student) declares that the given demonstration was not understood. The findings from such studies can help in turn, from a *system perspective*, towards the design of HRI systems that are able to better anticipate and behave according to human expectations.

Partly intended as a replication and verification of a previous study, the everyday domestic task of setting a table, both in Japanese and in non-Japanese (or "western") style, is taught to a humanoid robot by the participants of the currently conducted user study. The participant's acknowledgment and responses to the robot's feedback are discussed in regard to demonstration changes and consistency, based on a HRI gesture classification.

1 INTRODUCTION AND OVERVIEW

As robots integrate more into human society and domestic/public environments (in contrast to industrial/factory environments), they can be expected to behave more like 'companions', i.e. a) make themselves 'useful' (being able to carry out a variety of tasks in order to assist humans in domestic home environments), and b) behave socially (possess social skills in order to be able to interact with people in a socially acceptable manner). Such robots should take into consideration individual human likes, dislikes and preferences in order to adapt their behavior accordingly.

Thus, issues of agency, believability and sociality become very important [5]. Humans have expectations about the way artifacts should react and perform in social situations and the design of robots needs to conform to some of these expectations. The design of such sociable robots needs input from research concerning social learning and imitation, gesture and natural language communication, emotion and recognition of interaction patterns [6].

Our broader research goal is investigating social learning in HRI and as part of this, how people explain routine home tasks to robots and how these human-robot interactions unfold by using speech and gestures. Using that as a starting point, we want to examine the possible role of feedback by a robot acknowledging (or stating failure to understand) the instructions of a human teacher.

[1] Miyake Lab, Dept. of Computational Intelligence & Systems Science, Tokyo Institute of Technology, Japan, email: alissandrakis@myk.dis.titech.ac.jp

This role would need to be systematically investigated, and examine how if affects the cycle of human-robot interactions. The overall teaching strategy would also need to be informed by studies investigating the characteristic ways of how people tend to segment tasks, forming goals and sub goals explicitly, or implicitly.

We would like to also highlight the dialogic nature of the problem and try to consider two perspectives: (a) *human perspective* – how people do teaching (more-or-less) naturally and (b) *system perspective* – how can hardware and sensor requirements be met and extensive (or even impossible) assumptions/pre-knowledge about the world be addressed. By examining the human perspective, we can better inform the system perspective; this line of research will enable researchers to design robots that can pick "close enough initial metrics of similarity" depending on context, kick-starting the robot's understanding of taught instructions and process of common ground negotiation with the human.

The paper is structured as follows. Section 2 introduces the four research questions. Section 3 details the methodology, section 4 discusses the results from the current user study, while section 5 discusses some design implications from the system perspective. Finally, section 6 offers some conclusions and possible future work directions.

2 RESEARCH QUESTIONS

The issue of how people explain routine tasks to robots has already been addressed in [13], suggesting that people's assumptions about the way a robot "should" be able to understand their demonstration, the implicit knowledge about the tasks that people tend to take for granted, and the robot's behavior legibility can and will influence the HRI in this type of task.

The purpose of the conducted user study is to examine the possible role of feedback (by the robot, to the human) regarding teaching a simple domestic everyday task. This feedback can be either *positive* or *negative*, i.e. indicating understanding (or not) of the human demonstration[2], by the robot.

The work presented here is also intended as a replication and extension of the work presented in [12], and as such shares three main research questions:

Q1 Do the human participants change their instruction(s) when the robot provides negative feedback (i.e. declares inability to understand the participant's gestures and/or speech)?
Q2 What is the nature of the change of the instructions (if any).
Q3 To what extent do people maintain the changes for the remaining of the teaching task.

[2] This can include both verbal explanations, and also non-verbal gestures by the participant.

In addition, given the two variations of the task (Japanese and non-Japanese style of setting a table), we add another:

Q4 Does the nature of the task to be taught (both known, but one not so frequently practiced by the participant) have any influence on the above questions?

On a related note, it is worth mentioning that [14] makes some observations about the way people tend to administer their own feedback when teaching a Reinforcement Learning agent:

(a) they use the reward channel not only for feedback, but also for future-directed guidance;
(b) they have a positive bias to their feedback, possibly using the signal as a motivational channel; and
(c) they change their behavior as they develop a mental model of the robotic learner.

These last points (about people's feedback) are outside the current scope of the work (focusing more on the robot's feedback), but very much of interest in the broad context of this work.

3 METHODOLOGY

3.1 Participants

The sample consisted of 11 native Japanese participants (6 female and 5 male), all of them university students in their early twenties. None of them had any computer science or robotics background, or (more importantly) had any previous experience of interacting with a robot, and could therefore be considered 'naive' in respect to their expectations of a robot behaving in a social manner.

For the purposes of the current exploratory study (as for the previous study presented in [12]) this sample size was considered suitable, in order to provide observations for future studies involving larger number of participants.

3.2 Materials/Apparatus

The current user study was conducted at the Miyake Lab, Tokyo Institute of Technology, in Japan.

Each session consisted of a participant teaching the humanoid communication robot Wakamaru (Mitsubishi Heavy Industries) two variations, Japanese and non-Japanese, of the everyday domestic task of laying a table.

The utensil objects were a plate, a bowl, a cup and chopsticks[3] for the Japanese style, and a plate, a glass, a fork and a knife for the non-Japanese style of laying a table. In each case, they were set on a side-table, from which they should be picked up by the participants and placed on another table, in front of the robot.

A video camera was used to capture the participant's demonstrations and interaction with the robot.

3.3 Procedure

In order to capture the most 'natural' responses and behavior by the participants, no specific instructions were given to them as to how to interact with the robot, besides the single restriction that they must only use one object at the time (but in any order), and hold it using one hand; as there were four objects to be used each time, each task was decomposed to four 'sub-tasks'.

[3] The participants were instructed to consider the two chopsticks together, as a single item. However, one participant (5) also considered the fork and the knife as a single 'item'.

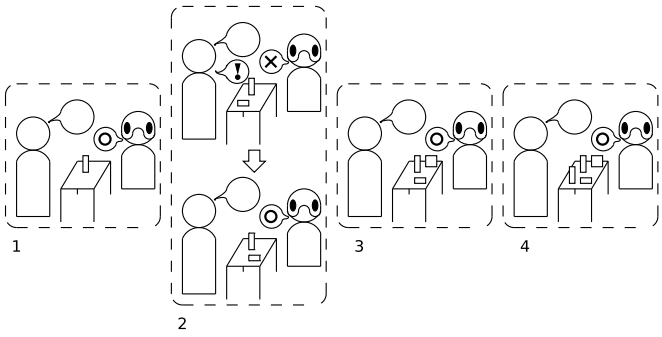


Figure 1. Robot's feedback for each sub-task. Irrespective of what the actual participant demonstration is, the robot should give positive feedback on the first, third and fourth demonstration, and (initially) give negative feedback on the second demonstration. Only after (and if) the participant repeats their second demonstration, the robot should then give positive feedback. See Fig. 2 for the actual gestures used by the robot.

In each session, a participant, after completing a background questionnaire, was asked to demonstrate to the robot how to lay the table twice, using a different set of utensils each time. The order for using the Japanese and the non-Japanese utensils was reversed for each successive participant.

In each of the tasks, the robot would provide positive feedback for the first, third and fourth sub-task demonstrations given by the participant, but (at first) provide negative feedback for the second demonstration (see Fig. 1), irrespective of what the actual instructions by the participants are.

The change from positive to negative feedback between the first and second demonstrations allows us to examine Q1 (and, in the cases that Q1 happens to be true, Q2); the change back to positive feedback for the third and fourth demonstrations allows us to examine Q3, assuming Q1 occurred. By comparing the interaction during the entire task in the Japanese and non-Japanese cases, we can examine Q4.

Finally, each participant was asked in a post-session semi-structured interview to subjectively recall the events as they occurred in their session, and comment on their impression of their interaction with the robot.

3.4 The "Wizard of Oz" Methodology

Due to technological issues (current state-of-the-art robots are not yet able to detect and understand *unrestrained* behavior by humans), plus the fact that for our purposes here the robot does not need to detect or respond to the actual participant's behavior (as all responses are predetermined), the robot can be controlled using the Wizard-of-Oz methodology.

The Wizard-of-Oz has come in common usage in the fields of experimental psychology, human factors, ergonomics, and usability engineering to describe a testing or iterative design methodology wherein an experimenter (the "wizard"), in a laboratory setting, simulates the behavior of a theoretical intelligent computer application or robotic system (often by remaining hidden, intercepting the communication and using teleoperation, with the participant having no a-priori knowledge of this – thus creating the illusion of autonomy). For more details about the Wizard-of-Oz methodology see e.g. [4].

The use of this methodology is very powerful for studies of this

kind; it is important for the current user study as we do not need to do an evaluation from the *system perspective* (therefore the actual robot performance only needs to be consistent and legible throughout) but rather from the *human perspective* observe the participant's behavior to the robot feedback in a controlled way (by having defined predetermined robot reactions for each sub-task).

Although the order and type of the responses is predetermined (as per Figs. 1 and 2), the precise timing depends upon each participant completing their current demonstrations. The end of a demonstration can be defined as when e.g.

- the participant stops talking and/or gesturing,
- the participant starts to interact with the next[4] object,
- the participant waits for the robot's acknowledgment,
- etc.

These criteria are more-or-less subjective, and currently far easier to be detected by humans (sensitive to a variety of cues across different modalities) than by an artificial system; therefore it is important that the wizard is the same person and s/he tries to be consistent throughout the experiment sessions (by using the same criteria in similar situations).

The precise response timing is an issue that cannot be easily controlled in a Wizard-of-Oz study. It has been shown (see [9, 10] for examples in HRI) that besides the explicit verbal part of communication (formal semantics) an implicit non-verbal part (expressed as e.g. the delay between an utterance and its response, but also the coordination of gesture and speech) also plays an important role, both in human-human and human-robot interaction.

3.5 Gesture Classification

The conceptual framework presented in [11] can be used to capture requirements for *contextual interpretation* of body postures and human activities for purposes of HRI. It defines five functional classes of gestures:

Manipulative gestures These are gestures that involve the displacement of objects (e.g. picking a cup), or miming such displacements.

Symbolic gestures These are gestures that follow a conventionalized signal. Their recognition is highly dependent on the context, both current task and cultural milieu (e.g. the thumbs up or thumb-index finger ring to convey "OK").

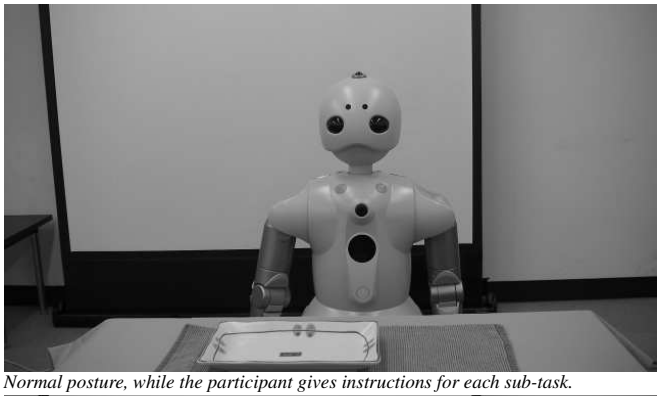
Interactional gestures This category classifies gestures used to regulate interaction with a partner. These can be used to initiate, maintain, invite, synchronize, organize or terminate an interaction behavior between agents (e.g. head nodding, hand gestures to encourage the communicator to continue).

Referencing/pointing gestures (Deictics) The gestures that fall into this category are gestures used to indicate objects or loci of interest.

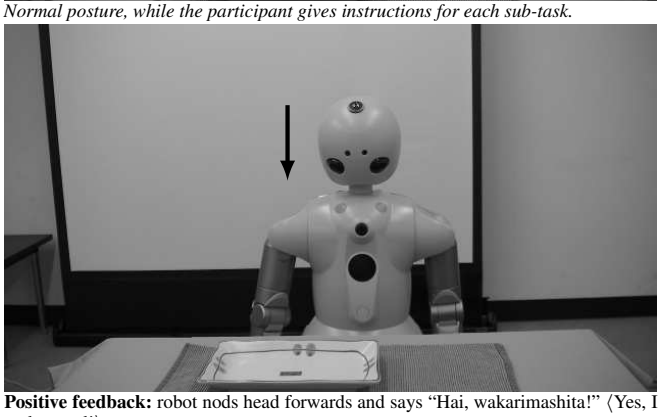
Side effect of expressive behavior These are gestures that occur as side-effects of people's communicative behavior. They can be motion with hands, arms, face etc but without specific interactive, communicative, symbolic or referential roles.

Irrelevant These are gestures that do not have a primary communicative or interactive function, e.g. adjusting one's hair or rubbing the eye.

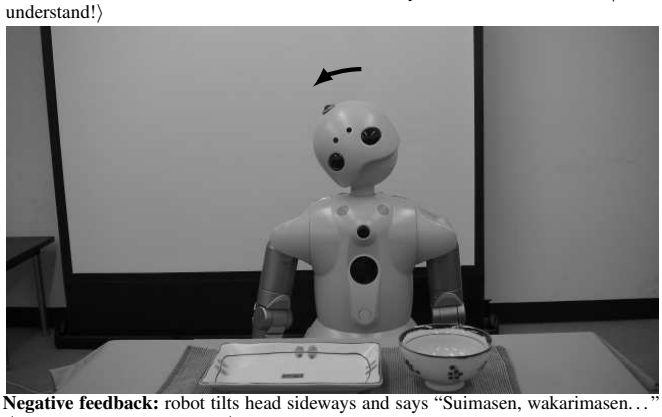
[4] Note that transporting and placing an object is usually not the end of a demonstration, and should not be used as such in general.



Normal posture, while the participant gives instructions for each sub-task.



Positive feedback: robot nods head forwards and says "Hai, wakarimashita!" ⟨Yes, I understand!⟩



Negative feedback: robot tilts head sideways and says "Suimasen, wakarimasen…" ⟨Sorry, I don't understand…⟩

Figure 2. Robot feedback. From top to bottom: normal posture, positive and negative feedback.

But it is important to note that certain gestures in particular situations might be multipurpose. Nehaniv et al. [11] stress the importance of knowing the context in which gestures are produced since it is crucial to disambiguate their meaning. In practice, data on the interaction history and context may help the classification process. The above framework is intended to complement existing and more detailed speech-focused classification systems (see for example [2, 3, 7]).

Towards examining the research questions posed in section 2, one can analyze the video obtained from the user study sessions, using the above classification. Doing this, a qualitative and descriptive picture of the interactions that took place in the sessions can be obtained, which can be also used to inform the requirements, from the system perspective, for identifying the human instructions and responding

accordingly.

For the task involved in this study, similar as in [12], only three of the previous categories were hypothesized to be relevant: *interactional gestures*, *manipulative gestures* and *referencing/pointing gestures*.

As utterances provide contextual information, one can also consider the following classification:

Interactional utterances produced to initiate, maintain, regulate or terminate the social part of the interaction.
Non-interactional utterances produced for the actual explanation of the proposed task. They can be further sub-divided into *indication of object*, *action* and/or *location*.

The objects' localization type can be also classified as

Relative reference, when the person indicates other objects or landmarks as reference points to the location where a specific object should be laid.
Absolute reference, when the person does not use other objects or landmarks as reference points. For example, the person might say "place the glass here" without making any additional gestures, apart from the actual transportation, to indicate possible relations among objects or landmarks regarding location.

Note that this localization classification is related (from the system perspective) to *effect metrics* for robot imitation (cf. [1]) i.e. measures of similarity between positions, orientations and states of external objects, but also changes to the body-world relationship of the agent, that can be used to achieve the same goal(s) in context.

4 RESULTS

The given task had no 'right' or 'wrong' way to be accomplished. The layout of the eating utensils on the table was done according to the personal preferences of each participant, with no objective measure of the task performance. Instead, this section presents a descriptive analysis of the data obtained from the user study; these include the recorded video from the sessions as well as the post-session interviews with each the participants.

As the current user study is (intentionally) very similar to the one presented in [12], we will also contrast some of the respective findings; that study will be referred to in the text as "previous", while the presently reported study as "current".

One general observation that is very contrasting to the previous study was the almost total lack of explicit referencing/pointing gestures. Only a single participant (3, in Japanese utensil task) used some explicit pointing gestures to indicate (relative) locations, but not to indicate an object. Four participants (4,7, 8 and 10, in both tasks) used a single, continuous, gesture to transport each object from the side table to the table in front of the robot, while at the same time verbally identified the object – sometimes adding localization information (4 and 10 only). The rest of the participants (1, 2, 5, 6, 9, 11 and including 3) used an implicit type of pointing gesture, typical, among other places, in Japan for presentation: the object was held at the end of the extended arm[5] of the participant, towards the robot, for a short[6] time before placing it on the table. This kind of gesture could be interpreted as both referencing the object itself but also its intended destination, although for the later interpretation, the extended arm's direction would not give precise information. Therefore, similar to the previous study, there was a lack of clear gesture segmentation between identifying an object and indicating its intended placement.

In some cases there were some subsequent correction gestures, moving the other objects on the table to improve the available space or to better orient the current utensil, but with no added verbal explanation as to what they were doing or why. Overall, there were no distinct interactional gestures observed. Two of the participants (4 and 6) produced an excessive amount of irrelevant type gestures throughout their sessions. Also, although most participants faced the robot across the table, participants 8 and 10 chose instead (for both tasks) to approach the side of the table, and lay the table while standing on the right side of Wakamaru.

4.1 Acknowledgment of the Robot's Feedback

In respect to Q1, every participant acknowledged the robot's feedback, and repeated (modified in some respect) their demonstrations for the second sub-task, in both tasks, when the robot stated that it does not understand their initial instructions.

4.2 Modification of Sub-Task Demonstration

The majority of the participants (1, 2, 4, 8, 9 , 10 and 11) when asked about it in their interviews, stated that the misunderstanding during the second sub-task (for either one, or both of the tasks) occurred due to their own fault (e.g. they did not speak loud enough for the robot to hear, or they did not hold the object within the robot's point of view); therefore, they repeated their demonstration more-or-less as before, both in terms of gesture and utterance, but speaking in a more loud and clear manner.

In terms of object localization, participants used both absolute (e.g. "I place it *at the center*.", "I place the plate *here*.") and relative (e.g. "I place the glass *on the right from your view*.", "I place the fork *on the opposite side*.", "I place the fork vertically *on Wakamaru's right*, *next to* the knife.") referencing, although in some cases they simply omitted any spoken placement instructions. Table 1 shows the object localization classification results for the current user study, indicating only three cases where the type of reference changed, as a response to the robot's misunderstanding. Note that for classification purposes, we considered the occurrence of an *absolute* reference when the participant explicitly used either "here" or "center", with no additional points of reference. In that sense, the absence of an explicit utterance but instead the performance of an implicit (*manipulative* or *deictic*) gesture could be also classified as an instance of absolute reference. However, in Table 1 we differentiate between absolute and absence-of (explicit) localization ("none").

In terms of instruction detail, all seven participants that used relative localization (1, 2, 3, 4, 6, 9 and 10) modified their instructions by changing or clarifying the reference points used (e.g. "[. . .] place it on the right back." → "[. . .] a little more in the back than the plate, to the right as much as it can be." or "I place it on the right." → "I place it on the right of Wakamaru."); three of those participants (1, 6 and 9) stated in their interviews that they thought the misunderstanding for the second sub-task occurred because they did not explicitly specify whether by "left" or "right" they meant from their own or the robot's perspective. In general, the instruction detail increased, with only a single participant (6) choosing to use 'simpler' referencing ("[. . .] on

[5] Some participants, contrary to the user study instructions, used both hands.
[6] Only one participant (6, both tasks) waited long enough after he presented a utensil in this manner until he explicitly received acknowledgment that the robot perceived the object, before demonstrating where the object should be placed.

#	task type	sub-task 1^{st}	2^{nd}	3^{rd}	4^{th}
1a	J	absolute	relative	relative	relative
1b	nJ	absolute	relative	relative	relative
2a	nJ	absolute	relative	relative	relative
2b	J	relative	absolute→relative	relative	relative
3a	J	relative	relative	relative	relative
3b	nJ	relative	relative	absolute	relative
4a	nJ	absolute	relative	none	none
4b	J	relative	relative	relative	absolute
5a	J	absolute	none	none	none
5b	nJ	none	absolute	relative	–
6a	nJ	absolute	relative	absolute	relative
6b	J	absolute	relative→absolute	absolute	absolute
7a	J	none	none	none	none
7b	nJ	none	none	none	none
8a	nJ	absolute	absolute	none	none
8b	J	absolute	absolute	none	none
9a	J	relative	relative	relative	relative
9b	nJ	absolute	relative	relative	relative
10a	nJ	absolute	relative	relative	relative
10b	J	relative	absolute→relative	relative	relative
11a	J	absolute	absolute	absolute	absolute
11b	nJ	absolute	absolute	absolute	absolute

Table 1. Object localization classification results (based on each participant's utterance). Task type order as given to each participant (*J*=Japanese, *nJ*=non-Japanese utensils). *None* indicates no explicit reference (i.e. only utensil name used). Only in 2b, 6b and 10b the participants changed the localization for the second sub-task. In all other cases, it remained the same. In task 5b, participant 5 incorrectly used both knife and fork together as the first item, similar to the chopsticks.

this side." → "[. . .] here." and "[. . .] on the right from your view." → "[. . .] on this side.")

One participant (7), when faced with the robot's misunderstanding in the second sub-task, she placed the utensil back on the side table, chose a different one instead, and then later came back to that utensil for the next sub-task. This occurred for both tasks, although there was no mention that the presentation order was important in the pre-session instructions given to the participants.

4.3 Consistency of Task Demonstration

As seen in Table 1, in every case when a participant changed their object localization reference type (2b, 6b and 10b), they remained consistent through the remaining sub-tasks; this is in contrast to the observations of the previous user study, where only few participants did. However, in the majority of cases in the current study, the participants did not change their referencing at the second subtask (in contrast to the previous study, where a larger number did), and besides participants 7 and 11 (which used "none" and absolute referencing, respectively, for the entirety of both tasks) a combination of localization types was used.

Although it could be argued that most participants in the current study were more-or-less consistent on their own, it is evident that there is still no consistency between them; even a simple assumption like e.g. "absolute reference is most likely to be used for the first object, in the absence of task-related reference points", cannot be generalized.

In comparison to the previous study, the current study results seem to indicate a higher degree of consistency, especially in the cases when a change was made in response to the robot's feedback; however the observation that in general humans lack consistency in HRI teaching tasks remains an issue, especially when (as in these initial user studies) the robot's feedback is not more informative besides a simple indication of understanding (or not).

4.4 Influence of Task Familiarity

In respect to Q4, in the current study, no differences were observed in the demonstration either of the whole task or in specific sub-tasks, between the two tasks, in terms of gestures. However, four participants (2, 3, 5 and 9) commented on the particular function of the Japanese utensils (e.g. "This is a *soup* bowl [. . .] it is used to eat miso soup.", "This is a *rice* bowl. I use it when I eat rice.")[7] compared to simply labeling the non-Japanese utensils. None of the participants added this description detail to their instructions for any of the non-Japanese utensils.

5 SOME STRATEGIES FROM A SYSTEM PERSPECTIVE

For currently available algorithmic and machine learning methods, the demonstrations by the participants of this study were far from clear or easy to identify. The setting of the study was quite freeform – the participants were given minimal instructions. The system/robot's capabilities were purposefully underpublicized (and we deliberately used 'naive' participants), in order to elicit a wide range of 'natural' responses from the participants, especially when the robot states that it can not understand their demonstration. We believe that investigating what people say and do while interacting with (here demonstrating to) robots, as well as what the people think about the robot's understanding is quite important for HRI. This line of research can enable researchers to design robots that are able to engage in a process of common ground negotiation with humans, depending on the context, towards achieving a variety of tasks/goals.

From a human perspective, we note that people are willing to interact with robots, but most importantly that they appear willing to accommodate the robot's feedback and modify their teaching demonstrations to facilitate its understanding. However, consistency of teaching style aside, without knowing what was unclear about the initial example, the additional example might contradict task-specific knowledge indicated in the previous example – or again be unclear. Given the initial character of the current study, the feedback was rather simple – an acknowledgment of understanding (or not), but not *what* was understood. A system should be able to communicate the nature of the source of the misunderstanding; one way this could be achieved would be for the robot to (partially) reproduce the current sub-goal. Even without verbal comments from the robot, such a reproduction attempt would implicitly advertize the robot's capabilities, and could provide the human with an insight on the source of the failure, to be addressed in their next demonstration attempt. This 'interactive', turn-taking approach to teaching would allow robots to take full advantage of the social environment for learning, and humans to increase their communication potential.

One of the reasons for this misunderstanding might be that people tend to combine information about the object and the manipulation/transportation/location. Unprompted, there is usual no clear segmentation, either by a distinct pause or particular gestures that emphasize e.g. the object or the location. From a system's perspective then, it is important for the robot to be able to prompt people to

[7] Note that both participants are here referring to the same utensil.

be more explicit about the sub-task segmentation of their demonstrations. This would possibly help isolate and identify any sources of misunderstanding, and target them specifically.

6 CONCLUSIONS AND FUTURE WORK

Based on the results from the current user study, all[8] participants did acknowledge the robot's feedback on their demonstration of teaching the task of laying a table (Q1); they appreciated the positive feedback, but most importantly they responded to the negative feedback as well, by repeating their demonstration. Unfortunately, the participants used very few distinct gestures, and, as the negative feedback was not very informative, the majority assumed that e.g. they had to speak louder, so they did not greatly modify their demonstrations (Q2). However, in the few cases they did (change the object localization), they were consistent for the rest of the task (Q3) – this observation was more promising than the findings of the previous study. Concerning the influence of task familiarity on the current task teaching (Q4), there were no definitive results from the current study, but it would appear that the participants tended to add extra details about the object's intended use and function in their verbal instructions.

One of the main motivations for the current user study was to confirm whether the initial observations of the previous study [12] on the acknowledgment of the robot's feedback were valid in a Japanese cultural setting (also, using a humanoid[9] robot), and inform the follow-up studies.

Next research topics would include further studies on to what extent are people willing to engage in interactions with robots that actively seek to show their understanding by demonstrating, and whether this strategy would be realistic in a wide range of domestic situations. Given the limitations of state-of-the-art active manipulation, what other mechanisms could perhaps put into place and serve as replacements?

Are people willing to accept the robot's requests regarding the way a task can be performed? What form should these requests take? Regarding these last issues, in the follow-up study we intend to increase the details of the robot's feedback, by adding details about the object's localization, followed by a pointing gesture. This could be either absolute ("here") or relative (e.g. " to the left of [name of other utencil]"), and instead of positive or negative, the feedback could be 'correct' (according to the participant's perceived intentions by the 'wizard') or not. A possible strategy to explore would then be to have Wakamaru use only one of either localization styles and make 'mistakes' if the participant uses a different one (or 'none').

Another observation we made on the current study is that most participants appeared during their session (and some made a remark in their post-session interviews) to expect more interactional gestures on the part of Wakamaru, instead of the robot maintaining a passive non-moving posture while they were performing their demonstrations. Therefore, we also aim to have Wakamaru actively indicate that it is being attentive (e.g. by regularly uttering "Hai ⟨Yes⟩" and nodding), while the sub-task demonstration takes place, before the feedback – this sort of indication of attention is very common (and expected) in Japanese culture.

One participant also remarked in her interview that the robot having the same response time for both positive and negative feedback did not feel right; in her opinion, the expression of misunderstanding should be more delayed (indicating perhaps a longer and more careful thinking process, or even politeness). In a follow-up system, more autonomous and less Wizard-of-Oz, the response timing of the feedback (time between the end of the participant's utterance and beginning of the robot's utterance) should be controlled as a function of the duration of the participant's utterance (similar to the approach in [9, 10]). This timing synchronization would facilitate a "co-creation" process [8], meaning the co-emergence of real-time coordination by sharing subjective time and space between different persons (and robots).

[8] But this is not always true; e.g. in the previous study a single participant chose to completely ignore the robot's feedback. More study is needed regarding this hypothesis in general.

[9] In the previous study, a robot of mechanistic appearance (PeopleBot, MobileRobots Inc.) was used.

ACKNOWLEDGEMENTS

The work presented in this paper was supported by a FY2007 (P07369) Postdoctoral Fellowship for Foreign Researchers from the Japan Society for Promotion of Science (JSPS).

The first author would like to acknowledge the hosting by Prof. Yoshihiro MIYAKE in Tokyo Institute of Technology. Both authors would like to thank Shohei YOSHIDA, Tatsunori NISHI and Kengo OKITSU for help in conducting the user study, Yukiko Thuresson for transcription and translation of the interviews, and Yumiko MUTO for additional translation and help.

REFERENCES

[1] A. Alissandrakis, C. L. Nehaniv, and K. Dautenhahn, 'Action, state and effect metrics for robot imitation', *Robot and Human interactive Communication, 2007. RO-MAN 2006. The 15th IEEE International Symposium on*, 232–237, (Sept. 2006).

[2] Justine Cassell, 'Nudge nudge wink wink: elements of face-to-face conversation for embodied conversational agents', in *Embodied conversational agents*, ed., J. Cassell et al., 1–27, MIT Press, Cambridge, MA, USA, (2000).

[3] Justine Cassell, Stefan Kopp, Paul Tepper, Kim Ferrimanand, and K. Striegnitz, 'Trading spaces: How humans and humanoids use speech and gesture to give directions', in *Conversational Informatics*, ed., T. Nishida, John Wiley & Sons, New York, (2007).

[4] Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg, 'Wizard of oz studies—why and how', in *Readings in intelligent user interfaces*, 610–619, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, (1998).

[5] Kerstin Dautenhahn, 'The art of designing socially intelligent agents: science, fiction and the human in the loop', *Applied Artificial Intelligence Journal, Special Issue on Socially Intelligent Agents*, **12**(7-8), 573–617, (1998).

[6] Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn, 'A survey of socially interactive robots', *Robotics and Autonomous Systems*, **42**(3-4), 143 – 166, (2003).

[7] D. McNeill, *Hand and Mind*, University of Chicago Press, Chicago, 1992.

[8] Y. Miyake, 'Co-creation in man-machine interaction', *Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003. The 12th IEEE International Workshop on*, 321–324, (2003).

[9] Y. Muto, S. Takasugi, T. Yamamoto, and Y. Miyake, 'Timing control of utterance and gesture in human-robot interaction'. submitted, 2009.

[10] K. Namera, S. Takasugi, K. Takano, T. Yamamoto, and Y. Miyake, 'Timing control of utterance and body motion in human-robot interaction', *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on*, 119–123, (Aug. 2008).

[11] C. Nehaniv, K. Dautenhahn, J. Kubacki, M. Haegele, C. Parlitz, and R. Alami, 'A methodological approach relating the classification of gesture to identification of human intent in the context of human-robot interaction', in *Proc. IEEE Ro-Man, 2005*, pp. 371–377, (2005).

[12] Nuno Otero, Aris Alissandrakis, Kerstin Dautenhahn, Chrystopher Nehaniv, Dag Syrdal, and Kheng Lee Koay, 'Human to robot demonstrations of routine home tasks: Exploring the role of the robot's feedback', in *Proc. 3rd ACM/IEEE International Conference on Human-Robot*

Interaction (HRI2008), Amsterdam, Holland, March 12-15, 2008, pp. 177–184, (2008).

[13] J. Saunders, N. Otero, and C.L. Nehaniv, 'Issues in human/robot task structuring and teaching', *Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on*, 708–713, (Aug. 2007).

[14] Andrea L. Thomaz and Cynthia Breazeal, 'Teachable robots: Understanding human teaching behavior to build more effective robot learners', *Artif. Intell.*, **172**(6-7), 716–737, (2008).