Indirect Human-Robot Task Communication Using Affordances

Seppo S. Heikkilä and Aarne J. Halme

***Abstract*— One problem in current human-robot task communication is the laborious need to define action and target object parameters for each task request. This paper's solution to the problem is to enable indirect task communication by mimicking the human cognitive ability to understand affordances, i.e. action possibilities in the environment with respect to different actors. This enables humans to communicate tasks using only the task-related action or target object names, and thus avoid the need to remember explicit task request utterances. The proposed task communication is integrated as a subsystem into an existing service robot, and its functionality is evaluated through a set of user experiments in an astronaut-robot task communication context. Affordance-based indirect task communication is shown to successfully reduce the workload experienced by the human and to decrease task communication times, while also being the preferred way to communicate tasks.**

I. INTRODUCTION

It is known that humans prefer to communicate with robots on a level where they think the robots will correctly understand them [14]. In most cases, this means short and simple utterances that do not leave room for misinterpretation. This reflects the basic requirement of human-robot task communication, which is that the task request utterances used need to be usable both for the human and the robot [1], [8].

This fundamental requirement is also a starting point for the so-called affordance-based task communication method, which is presented in this paper. The idea is that only a reference to a task-related target object or action can communicate the whole task for a robot that is capable of associating objects with actions that the robot can perform with those objects.

An example of this kind of indirect task communication is shown in Fig. 1. Instead of directly communicating all the task parameters, which in this case are the "analyse" action and the "rock" target object, the human can use affordance-based indirect task communication by stating only the task's action or target object name, i.e. in this case, "analyse" or "rock". The robot can then complete the task request by using the knowledge of what actions it can perform with the referred object.

The context menus in graphical user interfaces have long been used to provide context-related menu entries [13]. For example, the context menu entries could be "open" and "delete" actions if a pdf-document object is being selected. This means that a computer is utilising its knowledge of the selected object's affordances, i.e. what actions it can perform related to that object.

Department of Automation and Systems Technology, School of Electrical Engineering, Aalto University, P.O. Box 15500, 00076 Aalto, Finland seppo.heikkila@aalto.fi

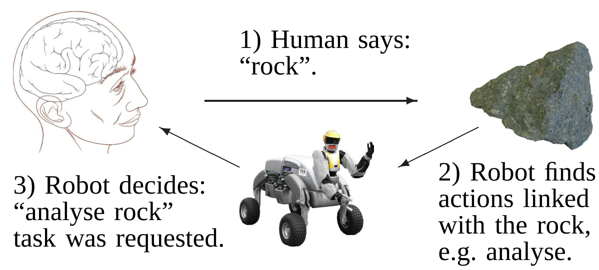

Fig. 1. Indirect human-robot task communication using only task-related object (or action) reference is possible if the robot can associate objects with actions that it can perform with the objects.

A. Concept of Affordances

The proposed indirect task communication method is derived from the theory of affordances [2], which defines affordances as "action possibilities in the environment in relation to the action capabilities of an actor". In other words, the theory of affordances proposes that all objects have a property called affordance that defines which actions are possible in relation to the actors.

Several subsequent studies provided further information on the role of affordances in human cognitive processes. For instance, it has been shown that human perception of objects enables a direct association with the possible actions that can be performed with those objects [3], [15]. This means that, for example, seeing a rock activates in the brain action presentations such as "analyse" and "pick up".

Furthermore, it has been shown that no other indication of action other than the object itself is required for the object action-association to occur, and that the perception can also be a form other than visual perception for the object-action association to work [3]. In addition, the object does not need to be visible to the human at the time of the action selection [15]. This means that any type of reference to an object is able to trigger the action presentations in the brain.

Thus, if a perception of object is received, for example through a pointing gesture, the human gets a number of possible actions that could be performed related to the object. This way the object alone can be used to communicate also the possible actions [16].

B. Task Communication Problem and Hypothesis

The core problem addressed in this paper is that current human-robot task communication is not efficient with regard to human workload and task communication time. The most important reason for this is that humans and robots do not communicate tasks the same way. Humans are not able to request tasks with the fixed communication utterances required by robots, while robots are not able to use complex natural human communication [1], [10], [8].

The hypothesis examined in this paper - which proposes a solution to the above-stated problem - is that humans are able to efficiently communicate tasks consisting of actions and target objects [12] to a robot in the same way that humans can communicate tasks indirectly to other humans using only the task-related action or object names.

This hypothesis is tested with two user experiments where participants are requested to communicate tasks to a robot by first using only direct task requests, consisting of action and target-of-action utterances, and then by using only indirect task requests, consisting only of task-related action or target object names.

The paper is structured as follows: Section II presents a user experiment performed with a fully autonomous centauroid robot in an environment where each object is linked to only one possible action. Section III extends the previous experiment into environments where all the objects generally have several possible actions that can be performed with them; and finally, Section IV presents the conclusion.

II. Unambiguous Task Communication Experiment

The first user experiment examines affordance-based task communication performance in a constrained case where each object and action is unambiguously linked to each other. The first experiment done with affordance-based task communication indicated that the method is able to decrease the subjective workload of the participants [6]. The experiment presented extends these results by increasing the number of participants and by measuring objective task communication performance using task communication times.

The proposed affordance-based method is integrated along with an existing human-robot interaction system, as shown Fig. 2. The key difference between the compared systems is that the information about object-action relationships enables the robot to interpret, for example, stand-alone object names as task requests. The dialogue structures of the task communication methods can be seen in Fig. 3.

A. Method

A total of 16 participants took part in the experiment. All the participants, excluding one high school trainee, were Aalto University staff or students. However, all of them were unfamiliar with the system tested and can therefore be considered novice users. The average age of the participants was 29.2 $\pm$ 5.8 years.

The overall scenario in the experiment was that the participant is an astronaut who is performing tasks with Aalto University's WorkPartner robot [5] next to a lander on Mars. The participant's goal was to fix any problems that emerged by requesting the robot to execute the correct task that solves the problem. The possible problems were jammed radio reception, sand build-up on the solar panel, or a sample stuck in the experiment unit. To fix these problems, the participant

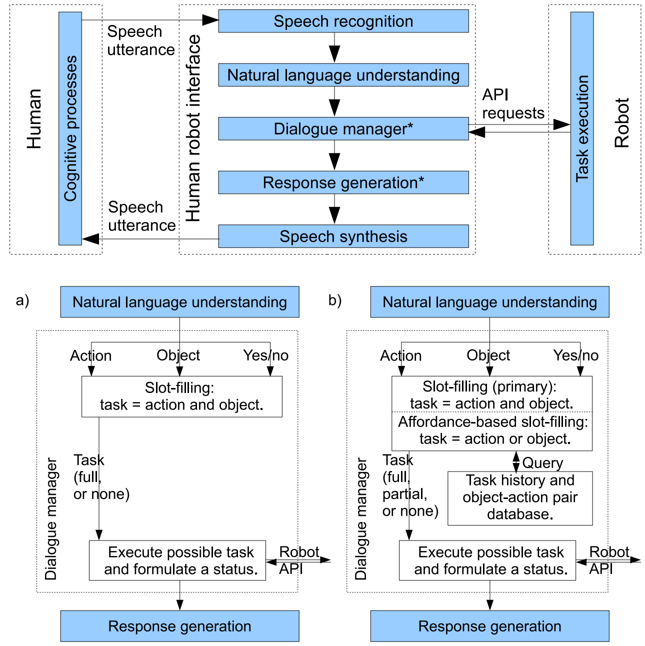

Fig. 2. The human-robot cooperation system high-level diagram (top), and dialogue manager subsystems for the compared (a) direct and (b) indirect cooperation systems. The differences between the compared systems are located in the marked* submodules, i.e. dialogue manager and response generation.

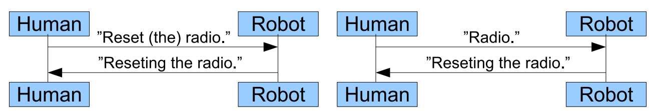

Fig. 3. Dialogue structure of direct (left) and indirect (right) task communication methods in the unambiguous experiment.

was able to request the robot to either reset the radio, clean the solar panel, or pick up the measurement unit, respectively.

The physical configuration of the experiment is shown in Fig 4. The participant and the WorkPartner robot were situated next to a lander mock-up, which had a radio transmitter, solar panel, and measurement unit on top of it, out of easy manipulation range of the participant.

The experiment was organized as follows. The experiment scenario was first explained to the participants, after which the CMU Sphinx-2 [7] speech recognition software was trained to correctly recognise the three object names and the three action names used in the experiment, i.e. reset, clean, take, radio, panel, and unit. The participants were told that their primary task was to focus on solving the problems that emerged as quickly as possible, and that they should work on a secondary inventory task, simulated by calculating arithmetic operations, only when they had free time.

The participants started the actual experiment by communicating the correct tasks required to solve the randomly occurring problems by first using only one of the examined communication methods. Each of the three problems was shown two times in random order in this first test round. The robot executed the requested tasks autonomously and always correctly, for instance, by sweeping the solar panel

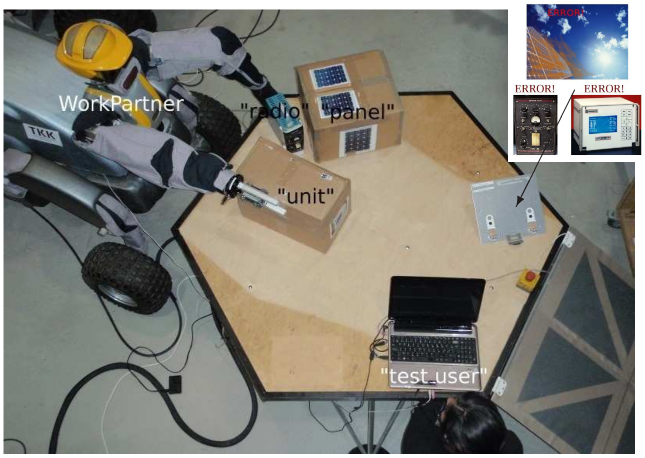

Fig. 4. The setup in the first experiment. The WorkPartner robot and the participant are located next to a lander mock-up with three items on top of it: a solar panel, a radio, and a measurement unit. The problems that emerge are displayed on paper sheets inserted into the stand in the top right corner.

with a brush. Next, the first test round was repeated using the other communication method, i.e. test round two. The experiment was counterbalanced, i.e. half of the participant started the first test round with the direct task communication method and half with the indirect method.

The two first test rounds were then repeated in an identical manner for test rounds three and four. This was done in order to get the actual experiment results from a higher point on the learning curve. Table I shows a short extract from one of the communication dialogues in the experiment.

TABLE I

EXAMPLE OF TYPICAL COMMUNICATION DIALOGUE BETWEEN THE HUMAN(H) AND THE ROBOT(R) IN THE UNAMBIGUOUS EXPERIMENT.

Event description	Direct method	Indirect method
1) Dusty solar panel. Request the robot to clean the solar panel.	H→R: Clean panel. R→H: Cleaning the panel.	H→R: Panel. R→H: Cleaning the panel.
2) Radio is jammed. Request the robot to reset the radio.	H→R: Reset radio. R→H: Resetting the radio.	H→R: Radio. R→H: Resetting the radio.

Finally, the fifth test round of the experiment had the participants communicate all the tasks three times by freely choosing which task communication method to use, i.e. direct or indirect method. The idea of the fifth test round was to show which communication method was preferred.

However, data were mostly collected from the third and fourth test rounds of the experiment. Firstly, the task communication time, i.e. the time from the emergence of the problem until the start of the human speech utterance, was measured. The purpose of this communication time was to measure how long it takes for the human to formulate the speech request after the problem has been noticed. Secondly, the NASA TLX questionnaire was filled in immediately after using the examined communication methods in order to evaluate the task communication workload experienced by humans.

The one-way within-subjects ANOVA was used to compare the data collected from the two examined communication methods in order to determine if the differences found were statistically significant. The ANOVA input data sphericity assumption was checked with Mauchly's sphericity test.

B. Results

The NASA-TLX subjective workload evaluation results for the direct and indirect task communication methods are shown in Fig. 5. The one-way within-subjects ANOVA showed that the difference between the averages is significant $F(1,15)=10.29$, $p=0.006$.

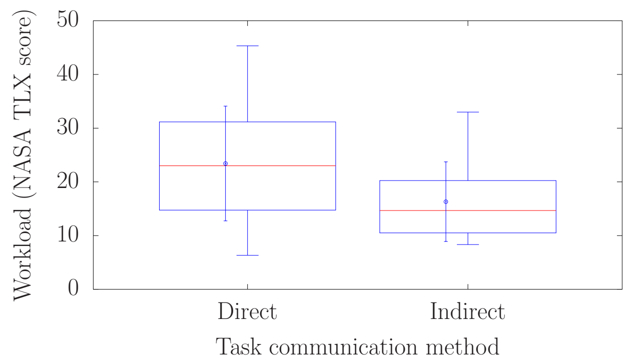

Fig. 5. Boxplot of the NASA-TLX workload, comparing direct and indirect task communication methods. The workload values range from 0 to 100, i.e. from no workload to full workload, respectively. Means and standard deviations are shown on the left sides of the box plots.

The communication times for using the indirect and direct communication methods are shown in Fig. 6. The one-way within-subjects ANOVA showed that the difference between the averages is significant $F(1,15)=8.027$, $p=0.013$.

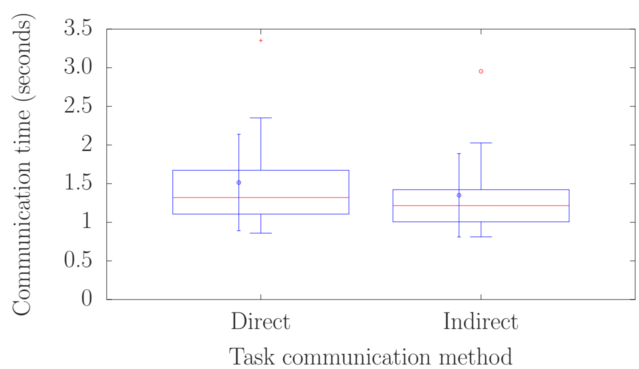

Fig. 6. Boxplot of task communication times for the compared direct and indirect task communication methods.

The participants' communication method preferences, measured in the fifth test round, can be seen in Fig. 7. The one-way within-subjects ANOVA showed that the difference between the averages is significant $F(1,13)=6.650$, $p=0.023$. However, two of the 16 participants did not take part in this fifth round of the experiment due to time constraints. One of these two participants started with direct task requests and the other with indirect task requests, so these results are also correctly counterbalanced.

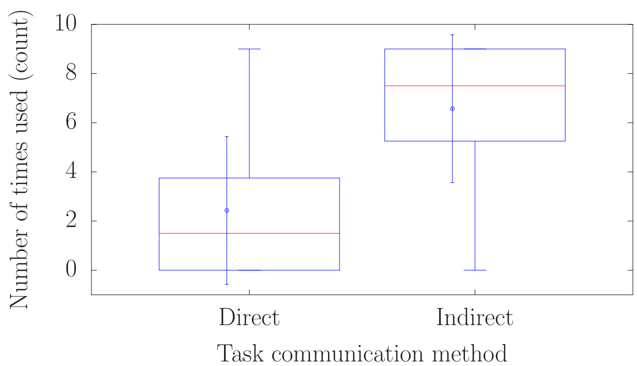

Fig. 7. Boxplot showing the utilisation of the communication methods, i.e. how many times both of the task communication methods were used by each of the participants. The maximum number of usages is nine because the total number of tasks that had to be requested was nine.

C. Discussion

The NASA-TLX workload analysis showed that the observed workload was lower with indirect task communication than with direct task communication. The measured communication times supported this observation, as it also took less time for the participants to communicate with indirect task requests. In addition, the results from the last "free choice" test round, shown in Fig. 7, indicate that participants prefer the indirect affordance-based task requests over the direct ones.

The possible explanation for these results is that with indirect task requests the human does not need to remember the action itself, but is only required to associate which object is at the core of the task. With direct communication, the human is instead required to also remember and formulate the action related to the task. The affordance-based task requests enable the human to leave the object-action association as a robot task.

III. Ambiguous Task Communication Experiment

Section II described a human-robot communication experiment where the robot was able to perform only one action with each object. Thus, the purpose of this section is to examine the problem of how affordance-based indirect communication could work in cases with ambiguous object-action relationships.

Current state-of-the-art task communication methods do utilise the concept of affordances at a certain level, because the robot is often able to explain what objects it recognises in the environment and what actions it can perform with certain objects [9], [11]. This type of special listing requests allows the human to learn or to remember again how to request a certain task from the robot. Nevertheless, the human is still always required to request the task directly by stating both the object and action name to the robot.

This kind of listing is already a powerful communication method in of itself as it is can likely enable the astronaut to communicate any task that might be required. Significant disadvantages with this kind of mechanical listing approach are the time required for listing and the unnecessarily high workload caused by the communication.

For this reason, the indirect affordance-based task communication method was formulated again for this experiment based on the experience gained from the unambiguous task communication experiment. Once again, the hypothesis is that only the object or action names can be used to efficiently communicate the tasks. The object-action association ambiguities are resolved by using past task requests to predict the most likely next task requests. These predictions are then accepted or rejected by the participant. The direct and indirect dialogue structures used in this experiment are shown in Fig. 8.

The algorithm used for task request prediction was the FxL sequence prediction algorithm [4], which is based on mixed-order Markov models. Based on the performance of the FxL algorithm with human-computer interaction predictions, such as Microsoft Word usage, the experiment was tuned so that 75% of the predictions were correct when accompanied by action or object name hints. The underlying assumption is that the tasks are often performed in relatively predictable sequences and the communication system can thus adapt online to the work by simply learning the tasks and the order in which they are performed.

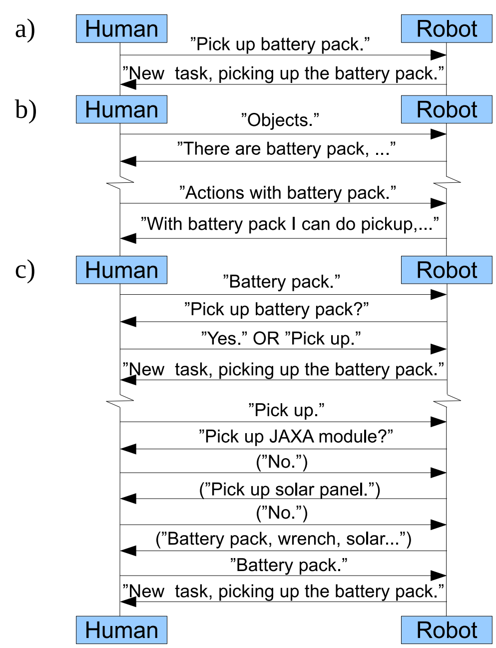

Fig. 8. Direct (b) and indirect (c) task communication method dialogue structures. The shared dialogue structure (a) is the one where a task is requested by using both an action and an object.

A. Method

A total of 18 participants were selected for the experiment. All of the participants were either Aalto or Helsinki University students or researchers. All of them were unfamiliar with the system examined and can therefore be considered novice users. The average age of the participants was 26.7 ± 5.5 years.

The overall scenario in the experiment was astronaut-robot lander preparation on Mars. The participant, who was acting as an astronaut, had to use speech to request 20 tasks from the robot in order to successfully complete the experiment.

The experiment was organized as follows. First the participant was given an explanation of the six objects that could be used and of the 21 actions that could be performed with the objects. However, there was a total of only 65 different tasks that could be requested, as all actions cannot be performed with all objects. Each task, consisting of an action performed on a certain object, was described to the participant with a comic strip-type of picture. After learning to recognise the tasks from these pictures, the participant trained the commercial Nuance Dragon NaturallySpeaking 10.0 speech recognition software to correctly recognise all the words used in the experiment dialogues.

Next, after explaining how the compared communication method dialogues worked, the participant tried all of the possible dialogue options a few times. Depending on the participant, this required five to ten rehearsal task communications.

The physical setup of the experiment is shown in Fig. 9. The task requests, of which an example is shown in Table II, were spoken into a wireless microphone, while the robot speech replies were output from the laptop speakers. The robot responses were simulated by playing video sequences on the laptop screen according to the corresponding task being executed by the robot. The participants were able to see a picture depicting the next task to be performed by pressing any key on the keyboard after the previous task had been executed.

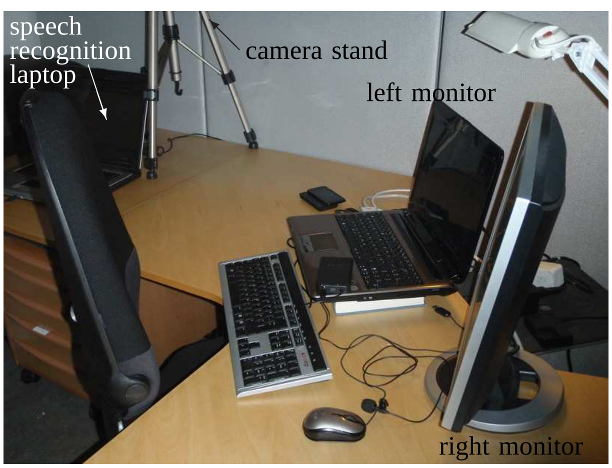

Fig. 9. Experiment setup for the ambiguous task communication experiment. The laptop in front of the chair was used to run the simulator (left monitor) and to show a picture depicting the next task to be requested (right monitor).

The participant's primary goal was to communicate the 20 tasks given - shown one by one on the monitor - like an astronaut would do when working on Mars. Between task communications, while the robot executed the requested task, the participant also calculated multiplications given on a sheet of paper as a secondary task. The 20 tasks

TABLE II

EXAMPLE OF TYPICAL COMMUNICATION DIALOGUE BETWEEN THE HUMAN(H) AND THE ROBOT(R) IN THE AMBIGUOUS EXPERIMENT.

Event	**Direct method**	**Indirect method**
1) Requesting the robot to pickup solar panel using action and object.	H→R: Pickup the solar panel R→H: New task, picking up the solar panel	H→R: Pickup the solar panel R→H: New task, picking up the solar panel
2) Requesting the robot to take image of wrench but without knowing the object's name	H→R: Objects R→H: There are wrench, battery... H→R: Image wrench R→H: New task, taking image of the wrench	H→R: Image R→H: Image wrench H→R: Yes R→H: New task, taking image of the wrench

were communicated three times: firstly, using only the direct method, then using only the indirect method, and lastly, making available both the direct and indirect methods at the same time. The experiment was counterbalanced, i.e. half of the participant started the first test round with the direct task communication method and half with the indirect method.

The main evaluation metrics were very similar to the ones in the unambiguous experiment. The participant workload was measured with the NASA TLX questionnaire; the time required to communicate and to perform all of the 20 tasks was measured; and finally, the participants' communication method preferences were collected by allowing them to freely choose between the direct and indirect methods.

B. Results

The NASA-TLX subjective workload evaluation results for using the direct and indirect task communication methods are shown in Fig. 10. The one-way within-subjects ANOVA showed that the difference between the averages is significant F(1,17)=11.70, p=0.003.

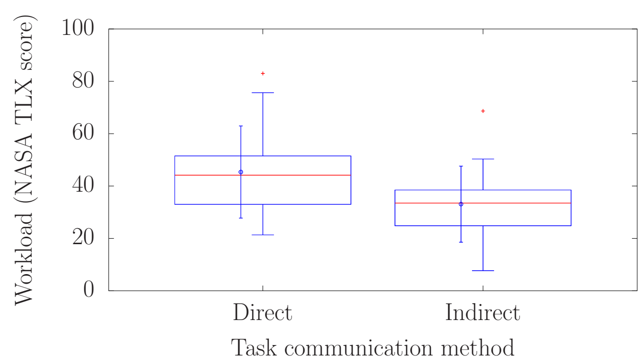

Fig. 10. Boxplot of the NASA-TLX workload for the compared direct and indirect task communication methods.

The test round execution times while using the indirect and direct communication methods are shown in Fig. 11. The one-way within-subjects ANOVA showed that the difference between the averages is significant F(1,17)=11.27, p=0.004.

The participants' communication method preferences, measured with the third test round, can be seen in Fig. 12. The one-way within-subjects ANOVA showed that the difference between the averages is significant F(1,17)=7.94, p=0.012.

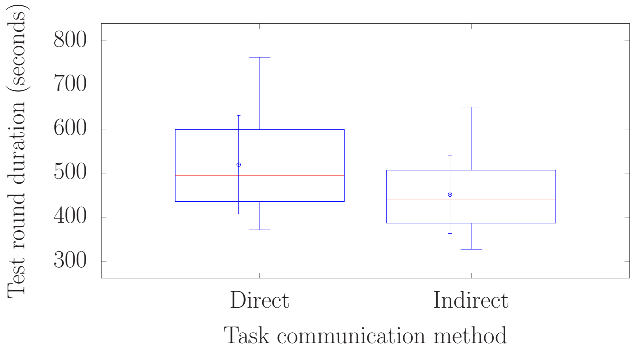

Fig. 11. Boxplot of the test round execution times for the compared direct and indirect task communication methods.

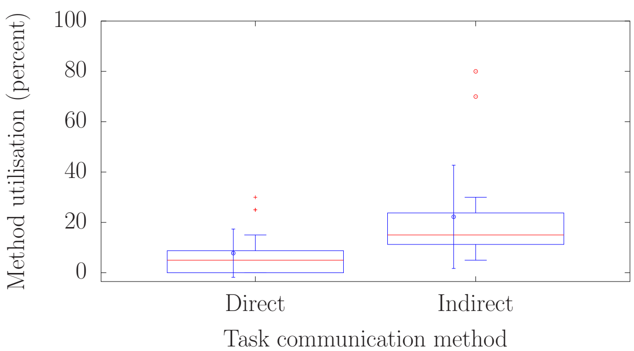

Fig. 12. Boxplot of the utilisation of the direct and indirect task communication dialogues in the third test round. The rest of the task requests were explicit task requests containing both actions and object names.

C. *Discussion*

The main finding of the second user experiment was that the formulated indirect task communication method was able to simultaneously decrease the subjective human workload and the total test round execution times, while also being the preferred way to communicate tasks. This is a clear indication that the proposed affordance-based indirect task communication method is a feasible and effective way to improve the conventional speech-based human-robot task communication in complex work environments as well.

This result is congruent with the unambiguous experiment result. The main argument in favour of indirect task communication was also the same in both experiments, i.e. it is easier to remember only task-related action or object names than both of them. There did not seem to be any significant additional mental processing, such as thinking about the object-action associations, that would have hindered task communication.

The formulated methods are also certain types of speech-based menus, which have long been used to communicate with both computers and robots. However, the novelty of the presented approach lies in the structuring of the menus using object-action associations. Context-menus, for example, display actions that are related to a selected object, but not usually the other way round, as is done here by displaying objects that are related to a certain action.

IV. CONCLUSIONS

This paper presented and evaluated human-robot task communication methods based on the concept of understanding action possibilities in the environment, i.e. affordances. Two user experiments showed that humans are able to communicate tasks indirectly using only the task-related object or action name utterances. Furthermore, the affordance-based task communication methods were able to reduce the human task communication workload and decrease the task communication times. This indicates that, in addition to explicit task requests, affordance-based task communication would be a feasible and effective alternative method for requesting tasks in work environments such as the examined astronaut-robot planetary exploration work environment.

V. ACKNOWLEDGMENTS

This work is supported by the European Space Agency's Networking/Partnering Initiative (NPI) programme and by the Finnish Centre of Excellence in Generic Intelligent Machines Research (GIM).

REFERENCES

[1] G. W. Furnas, T. K. Landauer, L. M. Gomez, and S. T. Dumais. The vocabulary problem in human-system communication. *Communications of the ACM*, 30(11):964–971, 1987.

[2] J. Gibson. The theory of affordances. *Perceiving, acting and knowing*, pages 67–82, 1977.

[3] J. Grezes and J. Decety. Does visual perception of object afford action? evidence from a neuroimaging study. *Neuropsychologia*, 40(2):212–222, 2002.

[4] M. Hartmann. *Context-Aware Intelligent User Interfaces for Supporting System Use*. PhD thesis, Technische Universität Darmstadt, 2010.

[5] S. Heikkilä, F. Didot, and A. Halme. Centaur-type service robot technology assessment for astronaut assistant development. In *Proc. of the 10th ESA Workshop on Advanced Space Technologies for Robotics and Automation (ASTRA)*, Noordwijk, Netherlands, November 2008.

[6] S. Heikkilä, A. Halme, and A. Schiele. Human-human inspired task and object definition for astronaut-robot cooperation. In *Proc. of the 10th International Symposium on Artificial Intelligence, Robotics and Automation in Space (i-SAIRAS)*, Sapporo, Japan, August 2010.

[7] X. Huang, F. Alleva, H. Hon, M. Hwang, and R. Rosenfeld. The SPHINX-II speech recognition system: An overview. *Computer, Speech and Language*, 7:137–148, 1993.

[8] T. Iio, M. Shiomi, K. Shinozawa, T. Miyashita, T. Akimoto, and N. Hagita. Lexical entrainment in human-robot interaction: Can robots entrain human vocabulary? In *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3727 – 3734, October 2009.

[9] H. Jones and S. Rock. Dialogue-based human-robot interaction for space construction teams. In *Proc. of IEEE Aerospace Conference*, volume 7, pages 3645–3653, 2002.

[10] V. Klingspor, J. Demiris, and M. Kaiser. Human-robot communication and machine learning. *Applied Artificial Intelligence*, 11(7):719–746, 1997.

[11] V. Kulyukin. Human-robot interaction through gesture-free spoken dialogue. *Autonomous Robots*, 16(3):239–257, 2004.

[12] R. Moratz and T. Tenbrink. Affordance-based human-robot interaction. *Lecture Notes in Computer Science*, 4760:63–76, 2008.

[13] M. S. Pandit and S. Kalbag. The selection recognition agent: instant access to relevant information and operations. *Knowledge-Based Systems*, 10(5):305–310, 1998.

[14] T. Tenbrink. Communicative aspects of Human-Robot interaction. In H. Metslang and M. Rannut, editors, *Languages in Development*. Lincom Europa, 2003.

[15] M. Tucker and R. Ellis. Action priming by briefly presented objects. *Acta Psychologica*, 116(2):185–203, 2004.

[16] K. Tylén, M. Wallentin, and A. Roepstorff. Say it with flowers! An fMRI study of object mediated communication. *Brain and language*, 108(3):159–166, 2009.