Remote Collaboration on Physical Tasks: The Influence of System Mobility

Manja Lohse[1], Chris Overgaauw[1], and Ronald Poppe[2]

[1] University of Twente, Enschede, The Netherlands, m.lohse@utwente.nl
[2] University of Utrecht, Utrecht, The Netherlands

Abstract. People perform many physical tasks such as assembling furniture. Often they could use support from an expert, however, it is expensive to send a person to help. Mobile telepresence robots could fill this gap by allowing a remote expert to have a conversation with a local person. This paper researches how the mobility of the telepresence robots influences collaboration in a physical task. We present first results from an experiment comparing a stationary and a mobile condition. Our results indicate that the robot's mobility did not influence task performance. However, local users were found to spend more time on visual grounding activities such as holding up objects in the mobile condition.

1 Introduction

People often perform physical tasks that they have little experience with. Typical examples are assembling furniture or installing internet wiring. In these tasks, mistakes can easily be made that an expert could prevent. However, it is expensive and time-consuming to have an expert physically present. One solution is to have a video-conferencing connection between the local user and a remote expert. However, various studies (e.g., [2, 4, 5]) concluded that a video feed provides no significant improvement in performance of the worker. These studies have in common that they used video sources from either a stationary camera or a dynamic camera mounted on the worker. Thus, the remote collaborators had no control over the content they viewed. Mobile telepresence robots present this opportunity. This paper sets out to discover how the mobility of telepresence robots influences collaboration in terms of task performance and collaborative activities on a physical task. One way the collaborators could achieve increased task performance is to produce grounding references in order to ensure that both participants are talking about the same objects. Clark and Brennan identified four different techniques of grounding references in interactions that are used to identify objects quickly and securely [1]. While three of these techniques are verbal (alternative descriptions, referential installments, and trial references), we are particularly interested in a visual reference because the mobility of the robot could potentially contribute to the exchange of visual references between the users and, hence, contribute to better task performance. Such visual grounding activities are indicative gestures (pointing, looking, touching, moving). We

Fig. 1. Interaction scenario and Double robot (www.doublerobotics.com)

believe that it is easier to establish common ground on the objects visually in the mobile condition where the robot operator can orient the robot towards the local users. Thus, users operating a mobile robot could adjust the field of view themselves, putting less effort for visual grounding on the person who is working physically on the task. This added control comes at a cost. As, among others, Kristoffersson et al. [3] point out, moving a mobile robot is a challenge which might prevent some remote operators from doing so. However, we expect that the more they do, the less time the local operators will have to spend on visual grounding. Our respective hypotheses are the following:

H1: The task performance will be better if the robot is mobile than when it is stationary.

H2: The more the remote user drives the robot around, the less time will the local user spend on visual grounding activities in the mobile condition.

2 Method

We designed a controlled user experiment in which remote participants collaborated with a local participant using a telepresence robot (see Fig. 1). We manipulated the mobility of the robot: in one condition the remote participants were allowed to move the robot around (mobile condition), in the other they were not (stationary condition). The system used in the study was a Double Robot.

The participants had to build two structures of PVC pipes, one for each condition. The two participants in each team were assigned different roles: local user (or builder) or remote operator. The remote operator had a picture of the structure that had to be assembled. They were told that they were not allowed to show the picture to the local users via the video stream. After the instructions, the participants had 10 minutes to build each structure. We counterbalanced the order of the conditions and the two structures. All participants needed the full time and most of them did not complete the task, thus, when the 10 minutes were over, they were asked to stop. After building each structure, the participants filled in a questionnaire. The whole procedure took approximately 45 minutes. We recorded data from a ceiling camera for later analysis. We also used screen capture software to record everything that the remote users saw on their screens, including the video from the robot.

We recruited 30 *participants*. The remote operators' age ranged from 21 to 53 years (mean = 29.9 years, sd = 7.6). The local users were between 22 and 61 years old with a mean of 32.1 years (sd = 12.5). Three remote operators (20%) and five local users (33%) were female.

3 Results

We used the video material of the interactions recorded with the screen capture software to obtain **quantitative performance measures**. To evaluate users' **task performance**, we compared the number of correct connections that the users made. In the mobile condition, the mean numbers of correct connections was 12.07 (sd=7.35). In the stationary condition, it was 13.47 (sd=7.30). We did not find any statistical effects. With regard to our first hypothesis, we conclude that we did not find the hypothesized effect.

Given the collaborative nature of the task, we were also interested in aspects of **how** it was performed. This leads us to the results concerning the **visual grounding activities performed by the collaborators**. Given the nature of the remote interaction, both participants could contribute to visual grounding in different ways: the remote operator could drive the robot to a position where he/she had a better view of the interaction. The local user could hold objects up and turn them for the remote operator to have a better view on them. We analyzed both aspects and their relation with each other. For the visual grounding activities performed by the local operator, our analysis showed an interaction effect between the conditions and the order of the conditions, $F(1, 13) = 16.309$, $p = .001$. The analysis also revealed a main effect of the condition on the duration of these visual grounding activities, $F(1, 13) = 7.92$, $p = .015$. Overall, the mean duration of the visual grounding activities performed by the local user was higher in the mobile condition (m=128s, sd=21.27s compared to m=101s, sd=20.31s). Independent of the order, the duration of the visual grounding activities initiated by the local user in the mobile condition was about the same (131s (sd=29.07s) when the stationary condition was first; 125s (sd=31.07s) when the mobile condition was first). However, the order mattered for the stationary condition. If the participants conducted the stationary condition first, the duration of visual grounding activities by the local user was significantly higher (m=143s, sd=27.75s) than if they conducted the stationary condition second (m=60s, sd=29.66s). In fact, the time reduced to less than half.

With respect to the remote operators' efforts on visual grounding, i.e., the time they drove the robot around, we did not find a significant effect. The time they drove the robot was not influenced by the order of the conditions (m=49.36, sd=29.77). Overall, there was a lot of variation in how much the remote operators drove the robot (min=7s and max=98s). Also, there was no relation between time driving the robot and task performance. Furthermore, we did not find a relation between time driving the robot and the time that the local users spent on visual grounding activities. Thus, we cannot confirm our second hypothesis.

4 Discussion and Conclusion

In our study we found that the users' task performance was about the same in both the stationary and the mobile condition, even though we had hypothesized that it would be better in the mobile condition. Also, we hypothesized that the local users would spend less time on supporting visual grounding in the mobile task if the remote users could take on some of this effort. However, we did not find this effect. A more qualitative look at the data provides us with some implication on why this might be the case. Generally, there was a lot of variation in how much visual grounding the pairs aimed for. It seems that some pairs relied more on verbal grounding such as the local user repeating the information that he/she received about the part (e.g., remote operator: "Please take the short pipe with the black and red tape", local user: "Ok, I have a short pipe with black and red tape"). This was probably due to the fact that participants perceived this strategy as more efficient than driving the robot around.

We also observed an interesting order effect for the stationary condition. The duration of visual grounding activities by the local users was significantly lower in the stationary condition, if they performed the mobile condition first. We expect that this is due to two reasons. First, building the structures requires common terminology and a certain strategy. For example, over time, it becomes more obvious how pipes can be distinguished. We expect that, generally, the efficiency of building structures increases, partly because of the more streamlined communication. However, these strategies are partly depending on the communication possibilities. The added possibilities of a mobile telepresence robot invite an exploration of its use. Consequently, we expect that both participants try to figure out how the remote operator can best support the building process. The interaction effect that we found could be explained by these two opposing trends. An analysis taking a closer look at the sequences of actions could reveal whether such a learning or habituation effect is actually present. Overall, further research is needed to explain which aspects influence the collaboration in order to understand and exploit the mobility of telepresence robots and to fully understand the potential of employing telepresence robots for physical tasks.

References

1. Clark, H., Brennan, S.: Grounding in communication. In: Resnick, L., Levine, J., Teasley, S. (eds.) Perspectives on Socially Shared Cognition, pp. 127–149 (1991)
2. Kraut, R.E., Fussell, S.R., Siegel, J.: Visual information as a conversational resource in collaborative physical tasks. Human-Computer Interaction 18(1), 13–49 (2003)
3. Kristoffersson, A., Coradeschi, S., Loutfi, A.: A review of mobile robotic telepresence. Advances in Human-Computer Interaction 2013 (2013)
4. Masoodian, M., Apperley, M.: User perceptions of human-to-human communication modes in CSCW environments. In: Proceedings of the World Conference on Educational Multimedia and Hypermedia (ED-MEDIA). pp. 17–21 (1995)
5. Whittaker, S.: Rethinking video as a technology for interpersonal communications: Theory and design implications. Human-Computer Studies 42(5), 501–529 (1995)