Utrecht University

Communication and gaze behavior in a dyadic collaborative task

Applied Cognitive Psychology Master thesis

Author: Ellen Verbunt (6488250)

Supervisor: Roy Hessels

Second assessor: Niilo Valtakari

ECTs: 27.5

Date: 08-07-2022

Abstract

Humans and robots continue to share an increasing number of spaces, which calls for safe and intuitive interfaces of these robots. Humans and robots interact in numerous ways, an important way is through collaboration. However, this part of human-robot interaction (HRI) is still underdeveloped. This research aims to give more insight into how robots should communicate with humans during collaboration by observing human-human interaction in a collaborative task. Observations were made on how people communicate with each other during a collaborative task and whether they look at their partner's face when they do. It was also looked at whether verbal communication has influences on task performance in a collaborative task. An experiment was designed in which two participants were asked to work together to recreate a Duplo figure while their gaze behavior was being tracked using the Tobii Pro Glasses 2. Four conditions were used, two in which the figure was either completely visible for both participants or partly hidden by other blocks, and two in which the participants were allowed or not allowed to verbally communicate with each other. Results show that very few the gaze fixations were on the partner's face. Participants verbally communicated a lot more during the obstructed conditions, and most of the utterances were comments, reactions and questions. The times that the partner's face was fixated on were not necessarily during communication, but the partner's face was looked at more often during nonverbal than verbal communication. No significant effect for verbal communication on task performance could be found. Based on these results, robots should not necessarily look at their human partner often. When they do, it could be for nonverbal communication. Verbal communication should be informative of nature, to keep the equal collaboration roles.

Introduction

Human-robot interaction has become a major area within robotics in the past couple of decades. Humans and robots share increasing amounts of spaces, for which it is essential that safe and intuitive systems are developed (Bütepage & Kragic, 2017). Because of the increase in shared spaces between humans and robots, it is necessary that the two can easily communicate (Krämer et al., 2012). Often, the integration of human actions in the decision-making process of robots is avoided. Robots are regularly designed to follow the master-slave principle, in which the robot is compliant towards the human's actions and acts as a support in achieving the human's goal (Bütepage & Kragic, 2017).

However, during interaction in daily life there is often a need for collaborating partners that can individually commit to a shared task, and not necessarily only follow commands (Bütepage & Kragic, 2017). For human-robot interaction (HRI), there are three types of interaction: instruction, cooperation, and collaboration. Instruction is largely described by the master-slave principle. Cooperation indicates how both robot and human work towards a shared goal, each by performing their own subtasks. Collaboration entails the interaction in which both human and robot work together towards a shared goal through interdependent actions. The last interaction type is the base for reciprocal learning and adaptation, and calls for trust (Bütepage & Kragic, 2017).

For a robot to be able to successfully collaborate with humans, it should have sufficient social skills. Krämer et al. (2012) show that humans will readily engage in communicative behavior and interaction with a robot in the same way that they do with other humans once they notice that the robot's social skills appear to be sufficient. There is an emphasis on the need for human-robot interaction to be very similar to human-human interaction. Humans still try to apply their own forms of communication that they are used to, even when the robot uses different forms, so it is important these human communication forms are applied (Krämer et al., 2012). Even though verbal communication is the primary way to convey messages in human-human interaction, nonverbal communication – including gaze behavior – is an important way to support the verbal message (Admoni & Scassellati, 2017). Aside from the fact that gaze behavior is important to support a verbal message, gaze can say a lot about a person's mental state. It is thus important to know when it is appropriate for a robot to look at the human with which they are collaborating.

One way to successfully achieve this similarity of HRI and human-human interaction is through the observation of human-human behavior. This research will investigate the ways humans verbally communicate in a dual collaboration task, whether gaze is supportive of this communication, and if verbal communication has influences on the collaborative task performance, which may elucidate how robots should behave socially with regards to communication to obtain successful collaboration with humans.

Ruesch and Kees (1969) describe communication as “all of the procedures by which one mind may affect another” (p. 1). The emphasis lies on the influence of the behavior with which one person influences the behavior of another (Buck & VanLear, 2002). Within communication, there is often a distinction made between verbal and non-verbal communication (Jones & LeBaron, 2002). Verbal communication describes vocal behaviors, primarily talking. Non-verbal communication includes visible behavior, namely facial expressions, eye gaze, hand gestures, and posture (Davis et al., 2006). Often, verbal and nonverbal communication happen at the same time, to support the message that is being conveyed (Jones & LeBaron, 2002).

Communication is an important part of collaboration (Hughes, 2008). Collaboration describes the situation in which people share roles, work together cooperatively towards a common goal, and share the responsibility of decision-making. Within a team, it has been shown that with increasing communication, collaboration also increases. In the medical field, most errors are a result of poor collaboration through lack of communication. To be able to perform a collaborative task, communication is of crucial importance (Hughes, 2008).

When performing a collaborative task, shared visual information can make communication a lot more efficient (Gergle et al., 2004a). Shared visual information allows people to view roughly the same objects at roughly the same time. It can decrease the amount of effort both parties would have to put in to complete a task. During a task in which there is shared visual information, planning and coordinating is at its most efficient. When there is discrepancy in the visual information between the two parties, more feedback is necessary to indicate progress, which would make communication less efficient. Pairs can work more accurately and faster when they have shared visual information. Shared visual information is critical for collaborative visual puzzle tasks because it aids in diminishing errors and gives ground for more efficient use of communication (Gergle et al., 2004a).

Research by Gergle et al. (2004b) shows that pairs alter their communication dependent on whether there is shared visual space available. In the case where there is shared visual information, participants would let their actions speak for them, and were less likely to verbally communicate. So, not only is task performance enhanced by working faster and making less errors, but there is also less verbal communication needed to complete a task when there is shared visual space (Gergle et al., 2004b).

Dual eye-tracking research by Macdonald and Tatler (2018) also shows that during a collaborative task, in which participants were asked to bake a cake, looking at the partner is does not occur frequently. When both participants share equal roles in decision-making, around 3% of the time is dedicated to looking at the partner. The conclusion was that looking at the partner rarely happened (Macdonald & Tatler, 2018).

To further aid in the development of robotics for human-robot interaction, this research aims to give insight into how humans communicate verbally and non-verbally during a collaboration task. For the present study, a collaborative puzzle task was designed, in which participants would have to copy a Duplo figure together, while either being allowed to verbally communicate or not. Eye-tracking glasses were used to measure gaze behavior to determine how people look at each other to communicate, and how they switch between copying the model and communicating with their partner. Based on this study, there will hopefully be more clarity on how humans verbally communicate during a task, whether this communication is accompanied by fixations on the face, and if the verbal communication has any influence on the collaborative task performance. With this information, it should become clearer which features to integrate in robots with regards to communicating during a collaboration with humans.

The first research question is: *How do people communicate during a collaborative task, and do they look at their partner's face when they do?* This research investigates whether, and how, participants switch between communicating and working on the task. With this, it will be looked at whether gaze behavior is an indicator of communication. One expectation is that participants will rarely look at their partner's face, based on the results Macdonald and Tatler (2018) found.

The second research question is: *How does verbal communication have an influence on task performance on a collaborative task?* The expectation is that the ability to communicate should not have an influence on task performance for the condition in which all of the blocks are completely visible for both participants, based on the research by Gergle et al. (2004b), in which they showed that when visual space is shared, participants would let their actions speak for themselves, instead of verbally communicating. Because of this, the expectation is that task performance will decrease when verbal communication is absent in the condition in which both participants do not have the same view of the figure and blocks are hidden, as opposed to when verbal communication is allowed.

Methods

Participants

For the experiment, twenty-four healthy participants were recruited through personal connections, to form twelve pairs during the experiment. All of the participants were students at Utrecht University. Of the twelve pairs, three were male-male pairs, four were mixed pairs (male-female), and five were female-female pairs. Prior to the experiment, written informed consent was collected. Participants who wore glasses were asked if they could remove them for the experiment if they could still see the blocks and their partner well. This research was reviewed and approved by the FETC; the approval is filed under protocol number 22-0974.

For the first research question, out of the twenty-four participants, five were excluded because of missing or incomplete data, three were excluded because the data quality was not high enough, one more participant was excluded because they reported to have nystagmus. The final group for the first analysis consisted of sixteen participants.

For the second research question, out of the twenty-four participants, none were excluded from the data analysis. Out of the twelve pairs, two were missing a complete trial, but were still included, because the other trial from the same conditions was still present.

Materials

For this experiment the Tobi Pro Glasses 2 were used to record the gaze behavior. Gaze behavior was recorded at 50Hz, and a scene camera in the glasses recorded the point of view of the participant at 25Hz.

For the collaboration task, Duplo blocks were used to build the figures, both the model and the copy of participants.

Procedure

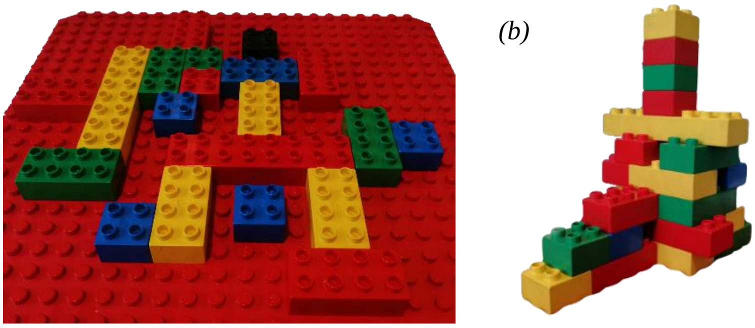
After the arrival of the two participants, they were asked to take place on both sides of the table. Before the experiment began, they received an information letter about the experiment and a consent form, which they were asked to sign if they wanted to continue with the experiment.

Once the consent form was signed by both participants, the participants were allowed to put on the Tobii Pro Glasses 2, with which their gaze behavior would be tracked. Both glasses would then be calibrated to guarantee accurate recordings. After the calibration, the participants were informed of the exact instructions for the task. There was room for questions and after that, the recording of the Tobii Pro Glasses 2 was started, and with it the experiment.

Each experiment had four conditions: verbal communication allowed vs. no verbal communication allowed and obstructed vs. non-obstructed figures. For the experiment, four figures were built per condition of obstructed and non-obstructed. Figure 1a shows a non-obstructed figure and figure 1b shows an obstructed figure. For the non-obstructed figures, the participants shared a visual space, so none of the blocks were obstructed (by other blocks) for either of them. As for the obstructed figures, the participants did not have a shared visual space, because not all of the blocks were visible for both participants, because they had been obstructed by other blocks.

Figure 1

(a) *(b)*



Note. Figures used in the task. (a) shows a non-obstructed figure, (b) shows an obstructed figure

At the beginning of the task, a practice trial was included. In this, a simplified, non-obstructed Duplo figure was presented. For this practice trial, participants were allowed to verbally communicate with each other. If there were no further questions after the practice trial, the experiment was commenced. It was made certain that conditions and Duplo figures were rotated between pairs to prevent bias and make sure no effects came from same sequence of conditions/trials.

There were four sets of condition and figure combinations, as displayed in table 1. The grouped figures (i.e., figures 5&8) were grouped to be matching in amount of time they took to complete (which was evaluated before the experiments were performed). These four sets were repeated three times over all the participants.

Table 1

	Non-obstructed		Obstructed	
Set 1	No verbal communication (figures 5&8)	Verbal communication (figures 6&7)	No verbal communication (figures 1&4)	Verbal communication (figures 2&3)

Set 2	Verbal communication (figures 5&8)	No verbal communication (figures 6&7)	Verbal communication (figures 1&4)	No verbal communication (figures 2&3)
Set 3	No verbal communication (figures 6&7)	Verbal communication (figures 5&8)	No verbal communication (figures 2&3)	Verbal communication (figures 1&4)
Set 4	Verbal communication (figures 6&7)	No verbal communication (figures 5&8)	Verbal communication (figures 2&3)	No verbal communication (figures 1&4)

Note. All possible combinations of conditions, grouped into sets which were used for the experiment.

When the experiment was completed, the recording was stopped, and participants would be debriefed. They could choose whether they wanted to receive 0.5 PPU (study credit) or not.

Data analysis

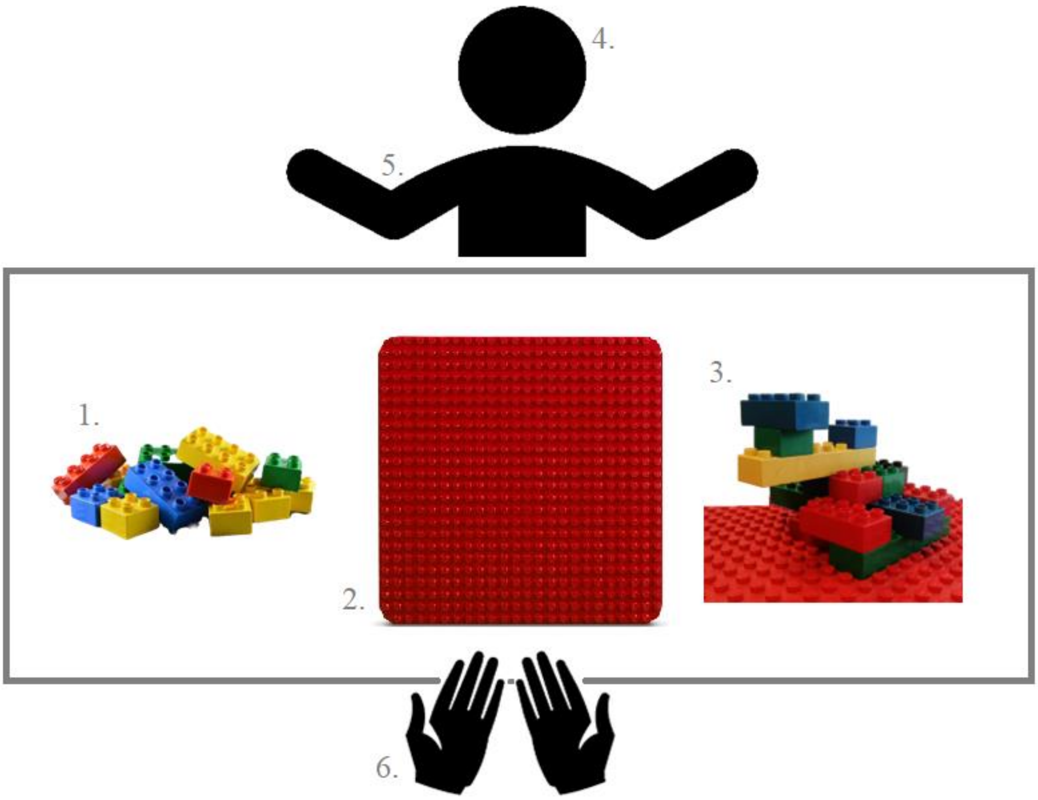
Firstly, the eye-tracking data was put into Glassesviewer (Niehorster et al., 2020) to determine whether the quality was sufficient for further analysis. The data quality was determined using the RMS-S2S and the data loss percentage. The data loss percentage indicates by the number of samples without a gaze coordinate, which can either be caused by blinking or by technical problems. The RMS-S2S (root mean square sample-to-sample) deviation is a measure for the precision of the eye tracking data. A large value for the RMS-S2S deviation indicates a noisy signal, which is not reliable enough to put through fixation categorization. When a recording has either too large of a data loss percentage, or a too large value for RMS-S2S deviation, the recording must be excluded from data analysis.

For the observation as to where participants look during the task, and where they look during communication, eye-tracking data was used. Measured with the Tobi Pro Glasses 2, the recordings were analyzed to assign where the participant looked. This was done using Gazecode (Benjamins et al., 2018). Using Gazecode, fixations in the eye-tracker data could be annotated to areas of interest (AOIs). In this research, there were six AOIs into which the fixations would be categorized (see also figure 2):

1. The Duplo blocks
2. The workspace
3. The Duplo model
4. The face of the partner
5. The body of the partner

6. The own hands

Figure 2



Note. Areas of interest (AOIs): 1. Duplo blocks 2. Workspace 3. Duplo model 4. Face of partner 5. Body of partner 6. Own hands.

To determine how participants communicate during the collaborative task, verbal communication was annotated. From the audio recordings the eye-tracker glasses took, every message transferred to the partner was put into one of seven categories:

1. Order (e.g., "You start with this one.")
2. Announcement (e.g., "I'll start building this part.")
3. Suggestion (e.g., "Maybe we could start here?")
4. Comment (e.g., "There's a long yellow block here.")
5. Question (e.g., "How many of these green blocks are there on your side?"
6. Reaction (any answer or reaction to a question or remark)
7. Other (often small talk unrelated to the task)

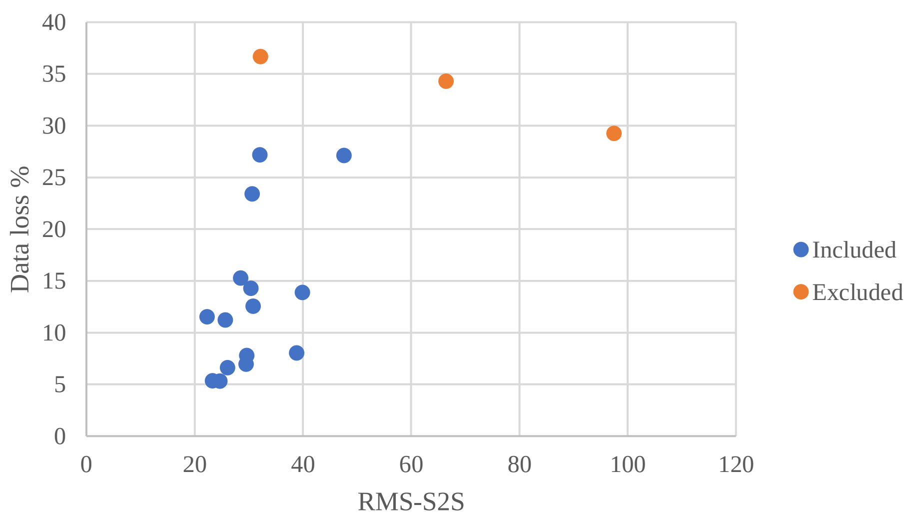
In this manner, it could be determined which kind of messages were more often communicated, and how often communication took place. To see if participants looked at their partner's face when either of them was talking, the fixations which were labeled with 'the face of the partner' were compared to the audio recordings. This way it would be clear whether people chose to alternate their gaze pattern to look at their partner's face during verbal communication.

To look at how verbal vs. nonverbal communication influenced the task performance on the collaborative task, task performance was measured by taking the amount of time it took the participants to complete the figure, and the number of mistakes that were made (and not corrected). The lower the time and the fewer mistakes were made, the better the task performance. These were then compared to each other over the different conditions (verbal vs. nonverbal and obstructed vs. non-obstructed). To compare these, the data was put into JASP and put through a repeated measures ANOVA.

Results

First, the data quality was assessed in Glassesviewer (Niehorster et al., 2020) to determine whether to include the recording in the further data analysis. For each participant – for which the data was as complete as possible, with at most one trial missing – the recording in Glassesviewer returned the data loss percentage immediately. The RMS-S2S deviation was calculated by taking the root of the sum of the squared values of the gaze point video X and gaze point video Y. For the data loss percentage, the cut-off to exclude a recording was 30% or higher. For the RMS-S2S deviation, the cut-off was 60 pixels or higher. Figure 3 displays the total amount of recordings that were included and excluded from the annotation in Gazecode Benjamins et al., 2018) based on the data loss percentage and the root square mean sample-to-sample deviation.s

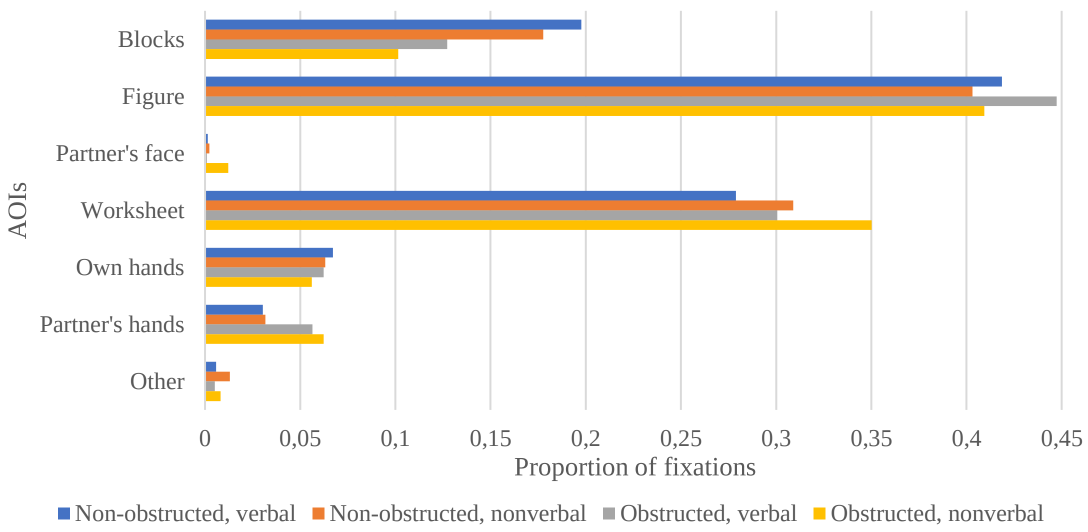
Figure 3



Note. Inclusion and exclusion of participants' eye-tracking data based on the data loss percentage and RMS-S2S. Cut off scores were >30% data loss and >60 pixels RMS-S2S. Each dot represents one participant. Out of the eighteen participants with complete enough data, three were removed from further analysis through Gazecode.

First, to see where participants look during the task, first all the fixations were calculated as proportions of the total fixations per condition. To offer a clear visual, these results were plotted in a bar chart. Figure 4 shows the proportions of fixations for each area of interest. The figure was looked at most often (~42%), followed by the worksheet (~30%) and the blocks (~15%). Then the own hands (~6%) and the partner's hands (~4%) were looked at. The least amount of fixations was on other areas (~0,7%) and the partner's face (~0,2%). Only for the AOI 'blocks' there is a difference how often participants look at it for each condition. Blocks are looked at relatively more often during the non-obstructed trials than the obstructed trials. This analysis showed that almost none of the fixations are on the face of the partner, and most of the fixations are on the figure or the worksheet.

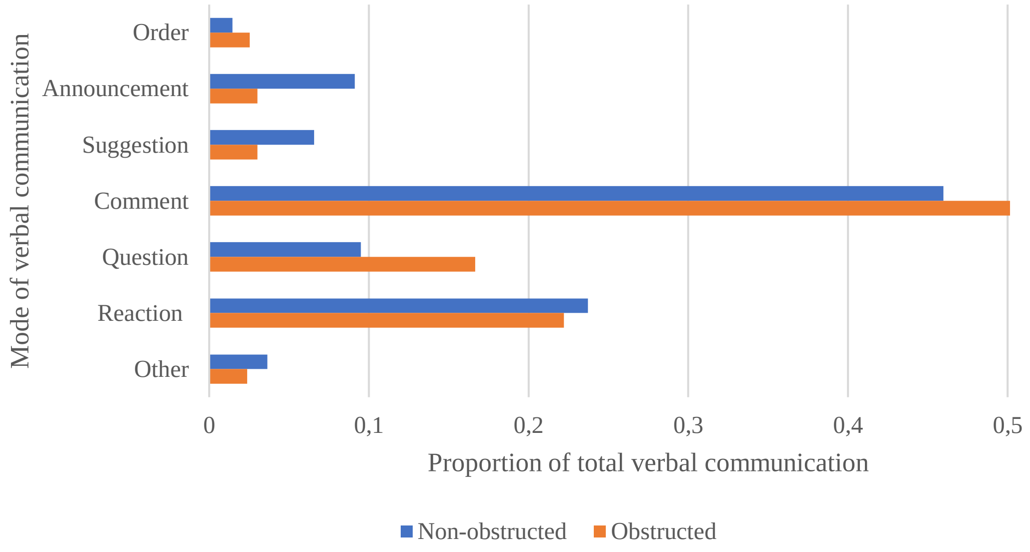
Figure 4



Note. Proportions of fixations per AOI per condition.

Secondly, all the labeled (verbal) communication was also plotted in a bar chart to show the relative use of each mode of communication for both obstruction conditions. Figure 5 illustrates these results. During the non-obstructed conditions, a total of 274 utterances was made, and during the obstructed conditions a total of 630 utterances was made. Most of the utterances made were comments (~48%). Then reactions were most used (~23%), followed by questions (~13%). The least amount of utterances were announcements (~5%), suggestions (~4%), other remarks (~3%), and orders (~2%). No big differences can be found for the different modes of communication between conditions. So, there is more communication during the obstructed conditions. For both conditions, most of the utterances were comments, followed by reactions and questions.

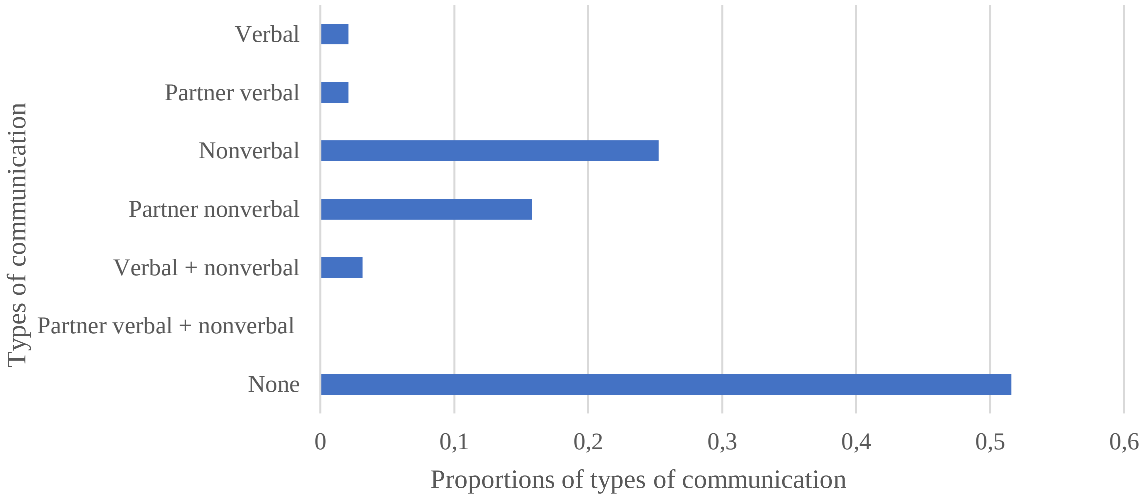
Figure 5



Note. Proportions of mode of verbal communication per obstruction condition.

For the last part of this research question, the number of times participants communicated with each other while fixating on the face of their partner were observed. Once again, these results were put into a bar chart. Figure 6 shows the proportions of types of communication during fixation on the partner's face. Most fixations on the face (~51%) happened without any communication surrounding the fixation. There is a clear difference between the verbal and nonverbal communications, as participants fixate more often on the face during nonverbal communication (~25% and ~15%) than verbal communication (both ~2%). There were a few instances in which the observed person looked at their partner's face while communicating both verbally and nonverbally (~3%), in all instances conveying a message with a question (i.e. "Is that a square blue one?") or a comment (i.e. "Like this.") and at the same time either pointing towards the figure or the own copy. There were no instances in which the participant looked at their partner's face when the partner was communicating verbally and nonverbally at the same time. Concluding, when participants look at their partner's face, it is mostly without communicative reasons, and when communication does happen during these fixations, it is nonverbal.

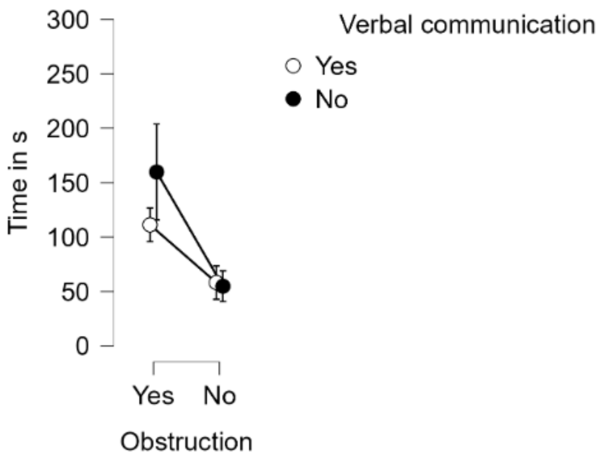
Figure 6



Note. Proportions of the types of communication during fixations on the partner's face. 'Partner verbal' and 'Partner nonverbal' indicate that the communication was made by the partner of the observed participant, while 'Verbal' and 'Nonverbal' were communications made by the observed participant. 'Verbal + nonverbal' is the communication that happened both verbally and nonverbally at the same time by the observed participant, and 'Partner verbal + nonverbal' is the communication that was both verbal and nonverbal by the partner. 'None' indicates the fixation on the face went without any additional communication.

To give insight into how verbal communication has an influence on task performance in the collaborative puzzle task, the means of the completion times (in seconds) and mistakes per pair per condition were calculated. Figure 7 shows the effect of verbal communication on the completion times in the task in both the obstructed and non-obstructed conditions. The data was put through a repeated measures ANOVA in JASP (JASP, 2022). The four conditions (obstructed vs non-obstructed, and verbal vs nonverbal) were put as the factors and levels; 'Obstruction' and 'Verbal

Figure 7



Note. The effect of verbal communication on the completion time (in s) in the obstructed and non-obstructed conditions.

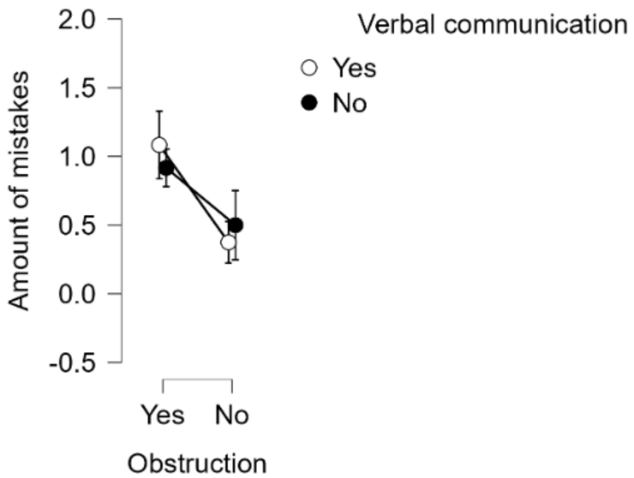
communication' as the factors and 'Yes' and 'No' for the levels in both factors. The repeated measures ANOVA showed that there was a statistically significant effect of obstruction on completion time, $F = 9.604$, $p = .01$ ($p<.05$). No significant effect of verbal communication on completion time could be found, $F = 0.742$, $p=0.407$ ($p>.05$). Obstruction seems to have an influence on completion time. Verbal communication does not seem to have an influence on completion time.

The interaction effect between obstruction and verbal communication was also not found to have a significant effect on completion time, $F = 1.082$, $p = .321$ ($p>.05$). Verbal communication does not have an influence on completion time, regardless of obstruction condition.

Another repeated measures ANOVA was run with the number of mistakes instead of completion time in seconds.

Again, a significant effect of obstruction on number of mistakes was found, $F = 6.410$, $p = .029$ ($p<.05$). Also, no significant effect of verbal communication on number of mistakes was found, $F = 0.007$, $p = .935$ ($p>.05$). Figure 8 shows the effect of verbal communication on number of mistakes made in the task for the obstructed and non-obstructed conditions. Again, only obstruction has an influence on number of mistakes, but not verbal communication.

Figure 8



Note. The effect of verbal communication on the number of mistakes made in the obstructed and non-obstructed conditions.

There was no significant interaction effect of obstruction and verbal communication on amount if mistakes made in the task either, $F = 1.803$, $p = .206$ ($p>.05$). There is no influence of verbal communication on number of mistakes, regardless of obstruction.

It can be concluded that only obstruction determines the task performance (both completion times and number of mistakes). Obstruction does, however, not determine verbal communication. It seems that verbal communication is not necessarily needed to complete the task better.

Discussion

To make a beginning towards discovering more about human-human interaction to incorporate into human-robot interaction, this thesis focused on discovering 1) how people communicate during a collaborative task and whether they looked at their partner's face while doing that, and 2) whether verbal communication has any effect on task performance in a collaborative task. During the collaborative task, participants were grouped in pairs and asked to copy a Duplo figure as fast and accurately as possible. With eye-tracking glasses (Tobii Pro Glasses 2) both participants' gaze behavior was recorded, as well as the environment from their point of view, and the audio during the experiment. The conditions were either non-obstructed or obstructed figures – in which participants did or did not have a shared visual space – and verbal or nonverbal – in which participants were either allowed or not allowed to talk with each other. All the participants had the same set of conditions, only alternating the sequence of the verbal conditions.

Through annotation of the eye-tracking data, most and least fixated on areas were discovered. From most to least fixated areas are the figure, the worksheet, the blocks, the own hands, the partner's hands, other areas, and the partner's face. Almost none of the fixations were on the partner's face (~0,2%). Reason for this could be that participants did not expect to find relevant information to the task from their partner's face. Especially compared to other AOIs like the figure, worksheet, or blocks, where almost all relevant information could be found. This is in line with the results that Macdonald and Tatler (2018) found that looking at the partner was rare during a collaborative task.

To see how people communicate during a collaborative task and whether they look at their partner's face when they do, first the proportion of fixations per area of interest was calculated. First of all, there was a clear difference between the non-obstructed and obstructed conditions with the number of utterances made. During the non-obstructed 274 utterances were made in total, while this number was 630 for the obstructed conditions. This indicates that obstructed conditions, with no shared visual space, need more communication. Almost half of the utterances were comments, a quarter reactions and an eighth questions. The rest of the types of verbal communication were used 5% of the time or less. No big differences could be found within the types of verbal communication between the obstruction conditions. There is clearly a bigger need for verbal communication during the obstructed conditions. This follows the theory of Gergle et al. (2004b) that with shared visual space, less communication is needed, because participants will let their actions speak for themselves. When there is no shared visual space, however, most of the utterances are comments, because participants like to keep their partner (and possibly themselves) in the loop on where they are looking and what they are seeing, as this is not something their partner can see. The goal there seems to be informing the partner on the observations. This appears to be the same case for both obstruction conditions. Orders appear the least of all, probably because when collaborating, both parties are equal within the decision-making process and should be able to trust each other equally (Hughes, 2008). When orders are given, this equal distribution of roles is cannot be assumed anymore, resulting in more instruction, instead of collaboration.

To understand whether participants looked at their partner's face while communicating, the proportion of times they looked at their partner's face during communication was calculated. Results showed that participants looked the least at their partner's face while communicating verbally. Most times, the fixations on the other's face were unrelated to any communication. During nonverbal communication, participants looked relatively more often at their partner's face than during verbal communication. During nonverbal communication, participants looked more often at their partner's face when trying to convey a message themselves, than when their partner was conveying a message. Possibly this is because they want to check if the message is received when participants communicate themselves, and because when the partner relays a nonverbal message, the participants more often have to look at their partners gestures and what they mean by that. As for the cases in which both verbal and nonverbal communication accompanied each other (which was only the case for observed participants, and not for the partner), it might be that when conveying the message, it is important to see if the partner understood it. As a receiver of the message, it might be more important to see what the participant is referring to, than the facial expressions or eye movements.

The results from the JASP analysis in which the effect of verbal communication on task performance was assessed show that there is only a significant effect of obstruction on both completion time of the task and number of mistakes made in the task. Verbal communication did not turn out to have a significant effect on either of the two task performance measures. This means that only obstruction had an influence on task performance, and not verbal communication. These findings are not in line with the expectations that were set up, this means that the null hypothesis cannot be rejected.

Also, no significant interaction effect of obstruction and verbal communication on both completion time and number of mistakes made was found. This indicates that verbal communication does not have more effect on task performance in the obstructed than in the non-obstructed conditions. The hypothesis that verbal communication would have a positive influence on task performance on the obstructed conditions can therefore be rejected. Verbal communication does not seem to have an influence on task performance on a collaborative task, regardless of the obstruction condition.

The current research's finding that verbal communication does not influence the task performance for both obstruction conditions, does not follow the discoveries of previous research by Gergle et al. (2004a). A factor that could have had an influence on these results, is that four pairs touched the Duplo figure in a way that could have positively influenced their time and accuracy, especially during the obstructed conditions, this could have had an impact on the scores. Also, two pairs verbally communicated during a trial when this was not allowed, and this could have influenced their scores positively as well. These participants were not excluded from the data, as their results were not outliers because of the possible advantage they had, but it is definitely useful to note that this could have influenced their results somewhat.

One other limitation that is worth mentioning, is the fact that after obtaining the data from the Tobi Pro Glasses 2, at least twelve recordings were unable to open. This meant that five recordings were not complete enough to include in the data analysis. These five excluded participants could have made for a more complete picture of the results. However, these missing recordings only impacted the data analysis in Glassesviewer (Niehorster et al., 2020) and Gazecode (Benjamins et al., 2018). For the other parts of the analysis, the recordings of the other participant in the pair could be used, so the one missing recording did not matter here.

In conclusion, there was a lot more verbal communication during obstructed conditions than during non-obstructed conditions. During obstructed conditions, more questions were asked and less announcements were made than during non-obstructed conditions, but no other clear distinctions were observed. It was shown that participants did not necessarily look at their partner's face when communicating with them, but participants did look relatively more often at their partner's face during nonverbal communication. No clear conclusion could be made for the influence of verbal communication on task performance in a collaborative task, as there did not seem to be any effect of verbal communication on task performance in both obstruction conditions.

Even though the results of this research are not entirely in line with the hypothesis, it does serve to contribute more to the relatively uncharted area of collaboration within human-robot interaction. The insights the observations gave are also valuable to learn more of communication and collaboration between humans. There are a few suggestions that can be made based on this research.

First of all, it is suggested that during collaboration, robots are not expected to look at their human counterpart's face that often. It should not necessarily be avoided, but fixations on the face, according to this research, are the least occurring during a collaborative task. Fixations on the face rarely happen, and even less during communication. The fixations on the partner's face that are made should mostly be made to accompany nonverbal communication. Verbal communication seems to call less often for support of gazing at the partner's face. In fact, people rarely seem to look at their partner's face during verbal communication.

For the part of verbal communication, the results suggest a more informative approach. Most of the utterances were comments, reactions, and questions. The comments were made to update the partner on what the participant was seeing, and both questions and reaction were often made to inform themself or their partner. The focus when implementing this in human robot collaboration should thus be informative.

Future research into human robot collaboration could look more into nonverbal communication, in particular gestures and body language, as to provide more information on how the robot should behave aside from verbal communication and gaze behavior. Also looking more into the way people seem to approach collaborative tasks like the one in the current study could be helpful to look at to integrate a more human strategy into the robot.

References

Admoni, H., & Scassellati, B. (2017). Social eye gaze in Human-Robot interaction: A review. *Journal of Human-Robot Interaction*, *6*(1), 25–63. https://doi.org/10.5898/jhri.6.1.admoni

Benjamins, J. S., Hessels, R. S., & Hooge, I. T. C. (2018). Gazecode. *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*. https://doi.org/10.1145/3204493.3204568

Buck, R., & VanLear, C. A. (2002). Verbal and Nonverbal Communication: Distinguishing Symbolic, Spontaneous, and Pseudo-Spontaneous Nonverbal Behavior. *Journal of Communication*, *52*(3), 522–541. https://doi.org/10.1111/j.1460-2466.2002.tb02560.x

Bütepage, J., & Kragic, D. (2017). Human-Robot collaboration: From psychology to social robotics. *ArXiv*. https://doi.org/10.48550/arXiv.1705.10146

Davis, S. J., Fullwood, C., Martino, O. I., Derrer, N. M., & Morris, N. (2006). Contemporary Ergonomics 2006. In P. D. Bust (Ed.), *Here's looking at you: A review of the nonverbal limitations of VMC* (pp. 290–292). Taylor & Francis.

Gergle, D., Kraut, R. E., & Fussell, S. R. (2004a). Language efficiency and visual technology. *Journal of Language and Social Psychology*, *23*(4), 491–517. https://doi.org/10.1177/0261927x04269589

Gergle, D., Kraut, R. E., & Fussell, S. R. (2004b, November). Action as language in a shared visual space. In J. Herbsleb & G. Olson (Eds.), *CSCW '04* (pp. 487–496). Association for Computing Machinery.

Hughes, R. G. (2008). *Patient safety and quality: An Evidence-Based handbook for nurses*. United States Department of Health and Human Services.

JASP (0.16.3). (2022). [Computer software]. JASP Team.

Jones, S. E., & LeBaron, C. D. (2002). Research on the Relationship between Verbal and Nonverbal Communication: Emerging Integrations. *Journal of Communication*, *52*(3), 499–521. https://doi.org/10.1111/j.1460-2466.2002.tb02559.x

Krämer, N. C., von der Pütten, A., & Eimler, S. (2012). Human-Agent and Human-Robot interaction theory: Similarities to and differences from Human-Human interaction. *Studies in Computational Intelligence*, 215–240. https://doi.org/10.1007/978-3-642-25691-2_9

Macdonald, R. G., & Tatler, B. W. (2018). Gaze in a real-world social interaction: A dual eye-tracking study. *Quarterly Journal of Experimental Psychology*, *71*(10), 2162–2173. https://doi.org/10.1177/1747021817739221

Niehorster, D. C., Hessels, R. S., & Benjamins, J. S. (2020). GlassesViewer: Open-source software for viewing and analyzing data from the Tobii Pro Glasses 2 eye tracker. *Behavior Research Methods*, *52*(3), 1244–1253. https://doi.org/10.3758/s13428-019-01314-1

Ruesch, J., & Kees, W. (1969). *Nonverbal Communication*. University of California Press.