Cooperative Tabletop working for Humans and Humanoid Robots: Early Investigations into Artifact Indication

H. Z. Hossen Mamode*, P. Bremner*, T. Pipe* and B. Carse*

*Bristol Robotics Laboratory, Bristol, United Kingdom
hamzah.hossenmamode@brl.ac.uk, paul.bremner@brl.ac.uk, tony.pipe@brl.ac.uk, Brian.Carse@uwe.ac.uk

Abstract— This paper represents the first steps in investigating issues emerging from a scenario where a robot is used as an avatar interacting around a tabletop surface with the robot being used to represent a remote member of a group. In particular, the effectiveness of the users, present in the immediate vicinity of the tabletop, at identifying which object on the tabletop is being indicated by a full size humanoid robot using our control scheme is investigated. The work is carried out in two phases; the first one focused on the accuracy in identification the object while the second phase focused on the response time of the identification process. The latter experiments are carried out with the introduction of a tabletop visual identification light while using the same robot. Two variants of the experiment were conducted for control and comparison. In variant 1, a human is used instead of the robot and in variant 2, no embodiment is used, that is, only the cursor is used to indicate the artifact. The results obtained indicate that the use of an embodiment, even with restricted degrees of freedom, when using our control scheme for designation has a significant positive effect on the artifact detection time.

I. INTRODUCTION

In recent years, there has been a substantial amount of research carried out on tabletop interactive devices: large multi-touch screens such as the Microsoft Surface® (Fig. 1). This research was not only concerned with how to successfully implement such devices but also on the impact of their use on users [1]. The focus of several of those projects has been on Human Tabletop Interaction (HTI) and how to improve on it while others were more focused on the impact that the use of such devices have on Human-Human Interaction (HHI) within groups of humans using them [1]. One issue with user collaboration on such a device is how to manage interactions if some of the users are remotely located, this is particularly relevant as an important aspect of collaboration (and communication in general) is the awareness of the actions of others, that is, collaboration is facilitated when the users are aware of each other’s actions [1]. It is therefore necessary for the actions that a user or entity is performing to be clearly visible to the others if collaboration is to be successful. Another important concept is that every entity should have a common understanding of the workspace in order to promote collaboration and minimize interaction errors [2].

In the context of a remote user wanting to participate in a group activity around a tabletop device a tele-operated humanoid robot could be used as an avatar. By using such an avatar, a physical presence is established in the local environment that the other users can relate to. This in turn will allow for the local users to collaborate more effectively with the remote user by being physically aware of the remote user’s actions; hence, it is suggested, communication and action comprehension will be improved among the different users. The use of the robot avatar in such a manner will allow sensor bridging [3] between the different groups as the robot will be able to convey the information in a way specific to its capabilities. This paper presents the beginning of a larger project; here we will be investigating some of the potential benefits of using a humanoid robot in tabletop interaction thus forming a fundamental building block toward our eventual goal. More specifically, we will be investigating the impact of having the robot using a particular control scheme has on artifact identification.

Figure 1. Microsoft Surface

In our observations of people collaboratively using a tabletop device we noted that they not only interact with virtual artifacts on a tabletop device, but they also use gestures to designate areas of interest or specific artifacts to one another. So a requirement for a robot operating as an avatar is to successfully convey induction of specific areas of the tabletop. It is thus necessary to investigate the accuracy of our chosen robot platform, Engineered Arts’ RoboThespian™, in concert with our joint coordination

1 Microsoft Surface® is a Product of Microsoft®. http://www.surface.com
2 RoboThespian™ is a product of Engineered Arts Limited. http://www.robothespian.co.uk/
method in the performance of this task. In particular, the robot is highly suited to Human-Robot Interaction (HRI) due to its compliant actuation method and light construction. Moreover since it does not operate with high precision and is limited in some key degrees of freedom, it makes for a good platform to test our non-trivial joint coordination scheme to produce a human-like motion. In the work of Riek et al. [4] it was shown that cooperative gestures need not be precisely humanlike to be useful, so we hypothesize that this will also be the case here. Furthermore, we have investigated whether similar benefits (in terms of action identification) might be realized with the robot as with a human, performing pointing gestures.

The key motivation for this paper is to investigate the efficacy of pointing gestures performed by a robot with limited range and speed of motion, for cooperative tabletop interaction. Doing so provides insight into how design compromises (whether they be for cost or safety reasons) affect robot utility. Furthermore, we aim to demonstrate the validity of our control scheme in producing effective pointing gestures.

II. BACKGROUND

Gesturing is an important element of Human-Human Interaction (HHI) [5]. For example, gestures are often used to emphasize certain elements of a spoken sentence and to convey thoughts when words are inadequate for the task [6]. The range of gestures is quite wide from a simple twitch of the face, to moving a limb to a specific position. Some gestures are easily recognized and widely accepted while others are harder to recognize and can lead to misinterpretation. It is therefore important for proper interaction to take place for gestures to fit in the correct context. In the case when humans are referring to a specific object in their environment they often point at it to add emphasis to what they are saying. This emphasis through pointing is however not limited to HHI in a local environment; in fact humans using avatars and interacting in virtual environments have demonstrated similar behaviours when referring to virtual objects [7]. Some gestures have even been successfully used between humans present in different physical locations [2] while other gestures have been proven unsuccessful [8].

Since gesturing seems to convey so much meaning, a lot of research has been carried out in Human-Computer Interaction (HCI) [9] and consequently Human Robot Interaction (HRI) to not only identify which gesture is being performed [10] [11] but also once identified how to deal with it appropriately [6] [12]. For instance, robots have been developed that are able to follow and interact with objects that a user is looking at [13]. However, it should be noted that since gestures have to be performed in a specific manner in order to convey the right meaning and decrease the possibility of a misunderstanding, it is critical for interaction when using a robot that it does it in a sufficiently humanlike way [14]. This is not only to prevent any misunderstanding of the gesture itself, if the gesture is wrongly performed but also due to the fact that the human interacting with the robot will expect the robot to follow the context of the communication just like a human would [15].

III. ROBOT PLATFORM

A. RoboThespian

RoboThespian (Fig. 2) is a humanoid robot that has been designed to be a robot actor, i.e., a robot designed specifically to perform gestures just like actors in plays or movies. It is therefore, well suited for gesturing. The robot contains 30 degrees of freedom with 10 in each arm and 4 in the head, and the rest in the hips and torso. It is powered by both pneumatic muscles to allow it to move in a manner similar to a human and make it compliant, as well as with motors for more precise rigid movements. The robot also has a light frame and has a limited speed of execution so it requires low actuator power. These features are highly beneficial within the context of close interaction with humans; they allow the robot to interact safely, as any possible impact will be of low energy. Additionally, as the robot is human-sized and of human-like appearance, the robot is more likely to be acceptable to the participant as long as it moves in a human-like manner [15].

Despite all of the capabilities mentioned previously, the fact that this is an off-the-shelf robot, and not purpose-designed for tabletop interaction, means that it suffers from some limitations. The limitations that are of particular significance here are the absence of some degrees of freedom, mainly in the wrists and hands, making the wrist limited to one plane of motion, and the limited range of motion of the elbows and shoulders, especially when compared to the equivalent human joints. Such limitations prevent some gestures that are possible for a human to be directly copied by the robot. Of particular importance, there is the inability of the arms to be at an acute angle with respect to the torso, i.e. limited adduction; this prevents the robot from pointing to anything directly in front of it using the arms alone. This is a problem particularly for tabletop interaction, as it severely limits the area in which the robot can indicate objects with each hand using only the arm joints. The
control scheme described in the next section, seeks to overcome this problem, while still achieving human-like motion, through the use of the hip and torso joints in the robot.

B. Control Scheme

1) Preliminary Study of HTI

In order to develop a control scheme that produces human-like motion, it is critical to understand how humans behaved when using an interactive tabletop. Thus, a preliminary study of participants was conducted using the Microsoft Surface. Participants were tasked with solving a digital jigsaw puzzle, requiring them to select pieces by touching them and dragging them across the surface. During the study, it was observed that while touch was used to move the artifacts around as instructed; pointing was also used to indicate other artifacts of interest, though the actual pose differed. On this note, the arm poses were observed to be dependent on the proximity of the artifact to be pointed to, and this was incorporated in the control scheme to increase the human-likeness of the motion of the robot.

2) Pointing Pose Generation.

Due to the limitations on the robot’s degrees of freedom and range of motion, it is unable to precisely reproduce human motion. However, our previous work has shown that effective gesturing can be produced with motions that are sufficiently human-like [14]. Here we present a method to produce human-like motion for pointing using the robot.

A series of key poses were identified from the preliminary HTI study. In particular the joint positions of the shoulder and elbow were noted to vary depending on the proximity of the identified object. Hence, joint positions for the robot’s shoulder (pronation/supination and abduction/adduction) and elbow joints were specified for the boundaries of the required work volume (i.e., the tabletop edges); these poses are shown in Figures 3 and 4. In order to point towards a specific position inside the boundary area interpolation between the boundary poses was used to find a pose that puts the target position within the range of locations that the robot is able to designate. With the interpolation generating a human-like position, it was necessary to ensure that the robot was actually pointing to the desired position on the tabletop device. In order to accomplish this, the interpolated pose was set on a kinematic model of the robot and the required position on the tabletop device was calculated in the robot’s coordinate frame. Hence, inverse kinematics could be used to calculate the hip and wrist joint values in order to align the robot’s forefinger with the target. By maintaining the shoulder and elbow joints in the interpolated positions the pose was kept as human-like as possible, while also reducing the problem to one of a two degrees of freedom with only one solution. The control scheme thus palliates for the limitations in the ranges of the joints in the arms by making use of joints in the hips to allow the robot to point to artifacts that would be inaccessible otherwise.

From the configuration obtained the position that the robot is pointing to was calculated, using forward kinematics. The position was then checked to be within the margin of acceptance of the target position and if it was the robot was set to that configuration.

![Figure 3. Robot designating a near point](image)

![Figure 4. Robot designating a far point](image)

Note left arm position between figure 3 & figure 4

3) Gaze.

In addition to the pointing gesture being implemented to designate a specific artifact on the tabletop surface, gaze was also implemented on the robot. Since the eyes of the robot are animations on LCD displays and thus cannot convey non-perpendicular angles due to the 2-D rendition, gaze was implemented using head orientation. Joints from the neck and torso were used in order to allow the robot to gaze at specific artifacts. In order to implement gaze, a virtual vector that ran from the back of the head to the middle of the eyes was used. This vector was aligned with the target position in a manner similar to the pointing gesture described previously while the eyes remained static. The gaze has been implemented in conjunction with the orientation of the finger to mirror the Human-Human Interaction (HHI), and in order to prevent adverse effects in the form of misdirected robot attention due to wrong gaze, as suggested in other studies [16][17]. Conversely while the gaze directs the attention of the participant to the specific artifacts, it alone does not provide enough information for the participant to identify the artifact uniquely.

IV. EXPERIMENTS

In order to investigate the performance of the proposed control scheme for our robot, in the production of comprehensible pointing gestures, an experiment was designed to evaluate the speed and of the robot’s pointing gestures.
In the experiment the robot was tasked with designating specific artifacts present on the tabletop device. The artifacts were displayed as a grid of alternating colored square tiles (Fig. 6). The experiment was conducted in two distinct phases. In the first phase the focus was solely on the accuracy in identification of the tile so, the task of the participants was to identify which part of an interactive tabletop was being indicated by the robot. In order to more finely assess the performance of this phase of the experiment, it was carried out using different granularities or grid sizes. One of which was a 4x4 arrangement of 10 cm squares, and the other 8x8 of 5 cm squares. In order to capture the participants’ responses a Launchpad\(^3\) was used as an interface, which was configured to display specific colors mirroring the grid of tiles on the surface (see Fig. 5). Each test was started with the robot selecting a tile at random and indicating it to the participant (see Fig. 6). After the robot indicated a tile, the participant was required to press the corresponding button on the Launchpad Controller to the area they thought was being pointed to.

For the second phase of the experiment the focus was placed on the response time and how the use of the robot affects the response time when an object is being designated. In this phase, an 8x8 grid was again used (see Fig. 6) but a cursor was added represented by a red 1 cm square that would appear in the middle of a specific square to designate that square. The participants were tasked to press a button as soon as they saw the cursor. The response time being captured in this experiment was the time between the cursor appearing on the Microsoft Surface and the participant pressing the button. While the cursor provided an accurate designation of which tile was to be selected, the robot was used to direct the attention of the participant to where the cursor was on the screen.

Both phases of the experiment were setup with the robot facing the tabletop device, with the hand about 100mm from the table providing more information on the which artifact is selected that the information obtained from the gaze, and the participant on the other side of the tabletop device with the input pad. This meant that the participant could make a direct mapping of the tiles on the tabletop device with the keys on the input device. Each participant took part in both phases of the experiment. A control condition was conducted in both phases of the experiment where a human was placed in the position of the robot and performed the same task as the robot. For the second phase of the experiment an additional control condition with no embodiment was carried out, in order to determine the response time when using only the cursor. In each phase of the experiment the conditions were varied between subjects using partial latin-squares.

---

\(^3\)Launchpad is a Product designed and developed by Ableton and Novation. http://www.novationmusic.com/products/midi_controller/launchpad

A. Results

A total of twelve participants took part in the experiment (4 females, 8 males), aged 21 – 37 (mean 27.1). For each of the experimental conditions, subjects were asked to identify 4 squares selected at random; only 4 squares were used to minimize learning effects from too many similar trials. This resulted in a total of 144 datasets (from the 12 participants across the 3 conditions and in identifying 4 squares) for the first phase of the experiment. These datasets were analyzed and the Figs. 7 &8 were produced.

<table>
<thead>
<tr>
<th>TABLE I.</th>
<th>MEAN EUCLIDEAN DISTANCE FOR BOTH GRID SETUPS USING THE ROBOT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Using 4x4 Grid</td>
</tr>
<tr>
<td>Mean Euclidean Distance (in squares)</td>
<td>0.7256</td>
</tr>
</tbody>
</table>

Table1 shows that using the robot even in an 8x8 grid setup provides enough indication for the participant to estimate the tile within an acceptable margin of 1square (1.414 or 7.07 cm on the Surface) to define where the indicated tile is.
The figures above represent the result of the analysis of the demand-response from the participant when the pointing was done on the 8x8 grids using the robot and the human. The horizontal axes of both graphs represent the tile indicated (indication) with respect to the components of the tile with the vertical axes representing the button pressed (response). The Columns and Rows represent the X and Y components respectively of a selected tile in the grid (Fig. 6). The analysis is done with respect to the columns and rows that the tile belonged to, with the vertical lines representing the range of the input from the participant. The boxes in the above figure represent the inter-quartile range of the input and the whiskers represent the standard deviation in the input.

By comparing the datasets from the first phase, there seems to be a similar trend between the datasets involving the robot and the ones involving the human. One of the trends, which were shared between the robot and human gesturing experiments, was that the accuracy was better when identifying which column is being designated by the embodiment (human or robot) with respect to the rows. After performing a statistical analysis of the data obtained, it was found that the trend using either the human or the robot seem to match closely, having a Pearson’s correlation coefficient of 0.991. When the number of errors from the robot and the human were compared, it was found that in the case of the robot, the mean number of errors was increased by 1.714. There was however no proportional increase in the error zone between the two conditions thus making the increase in error due to boundary delimitation between the tiles while the embodiment still defines the same area, this problem can be solved by the use of the cursor previously described.

A statistical analysis of the data using Student’s T Test was conducted (using a value of alpha at 0.017 due to the Bonferroni correction across three tests), the results of which showed a highly significant difference, \( t(18)=3.08, p=0.006 \), between the use of the robot and cursor, and the cursor only; a similar trend was found between the use of the human and cursor, and the cursor only, \( t(16)=3.72, p=0.002 \); finally there was no significant difference between the use of the robot and cursor, and with the human and cursor \( t(21)=0.75, p=0.461 \). These results show that presence of an embodiment has an impact on the response time while the nature of it (human or robot) does not.

V. DISCUSSION AND CONCLUSION

A. Discussion

From the data obtained from the first phase of the experiment it can be seen that even without the use of the cursor, the participants were able to correctly specify which column that a tile belonged to and were (on average) within one square of the target row. For the purposes of spatial and action awareness required in many HTI scenarios, this is likely to be sufficient for successful cooperation.

In order to better understand the issues with identifying the correct tile being designated by the robot, it is instructive to decompose the error into its components in terms of rows and columns of the grid (Fig. 5). From this the main source of error appears to be in properly identifying to which row the tile belongs (i.e., the Y value of the tile). From this, and the setup of the experiment, it can be concluded that determining a tile along the X axis of the grid is better than along the Y axis for a square grid shown by the slightly smaller inter-quartile boxes. The better recognition with respect to the side than distance from the participants has important implications for interface design. The control method implemented in the paper has thus shown that it can be used to allow the robot to point with an accuracy similar to a human being.

According to Fitts’ Law [18] the Average Time taken to select an object increases with the distance to target while keeping all the the other factors the same, a parallel can therefore be drawn to account for our finding. In our case the response time is a measure of the distance to target from the initial focus area. This together with the data obtained from the second phase of the experiment, showed that the presence of an embodiment, whether it being a robot or a human, has a positive impact on the reduction of the response time. That is, the use of an embodiment complements the use of the cursor by decreasing the time that is required to see the cursor while the cursor itself adds an improved accuracy. The use of embodiments, or representation of them, have also been proved to add trajectory information to interaction [19 - 21]. This is of particular importance, especially if the robot in this case is to eventually be used as an avatar for a remote person to interact with the ones present.

![Figure 7. Robot Pointing for 8x8](Image 64x680 to 287x767)

![Figure 8. Human Pointing for 8x8](Image 140x148 to 301x227)

Figure 7. Robot Pointing for 8x8 Figure 8. Human Pointing for 8x8
around the tabletop device. This is supported by the absence of statistically significant differences, when using a significance factor of 0.017, between the embodiment being a human or a robot, both in terms of accuracy and response time. In the case of the presence of an embodiment, the attention is drawn to a specific portion of the screen thus decreasing the distance between where the participant is focussing and the cursor.

B. Conclusion

In the work presented here we have shown that embodiment causes a significant improvement in indication recognition time over with the cursor alone. More importantly, there is no significant difference between the improvement found with a human or a robot with restricted range of motion. These results show that in the context of tabletop interaction precise humanlike behavior is not required, thus proving the hypothesis from section I. Further, our control scheme was able to overcome the limitations of the RoboThespian indicating its usefulness in our proposed avatar scenario.

Our findings show that our control scheme allows robots, even those with restricted range of motion, to be able to indicate artifacts successfully in the context of tabletop interaction. Further, the findings imply that close bio-mimicry is not required in a robotic avatar for tabletop interaction. Due to the inherent safety issues, and hence design compromises common in cooperative HRI, we believe this is a useful result. The ideas presented here could be extended to other robot platforms, and also have implications for robot avatar design. One such implication is the cost of a suitable robot platform, e.g., RoboThespian is around a tenth of the cost of the BERT robot platform we have used in our other HRI work [4][14]; hence, in the context of tabletop interaction a cheaper robot can be used.

C. Future Work

In order to find out what the effect on perception is with respect to the viewing angle, a further investigation will be carried out with the users placed at different angles to the robot (as opposed to just in front of it). Furthermore since the control scheme allows the robot to operate in an adequate manner for pointing, a future development will be the use of the robot to indicate tile selection and movement (dragging) which is a common feature in HTI. Expanding on this, another future development will be allowing the robot to operate in a two-mode system where the robot will have to perform other interactive gestures, thus further investigating the capabilities of the RoboThespian to produce comprehensible actions to an observer using the described control scheme.

ACKNOWLEDGMENT

This paper is part of a project of the Bristol Robotics Lab from funding from the University of the West of England and in collaboration with the University of Bristol.

REFERENCES