Toward bridging the gap between Social Robots and Industrial Robots

Alap Kshirsagar
Sibley School of Mechanical Engineering
Cornell University
Ithaca, NY 14850
Email: ak2458@cornell.edu

Julia Chen
Sibley School of Mechanical Engineering
Cornell University
Ithaca, NY 14850
Email: tc468@cornell.edu

***Abstract*—Most of the collaborative robot arms working alongside humans on factory floors do not have mechanisms for communicating with the human teammates. On the other hand, most of the socially expressive robots do not have capabilities to perform industrial tasks with high accuracy and repeatability. In this work we evaluate the effect of combining a social tablet-head robot with a collaborative robot arm. The tablet-head communicates with the human teammate through gazes and nods when they execute a collaborative task with the robot arm. We evaluate the effect of a tablet-head on likability and efficiency of the robot. In our experiments, human participants perform a collaborative task with the robot arm in the presence and absence of this tablet-head. Our quantitative analysis shows that the likability of robot arm increases if it is accompanied by a tablet-head. Also, we find that the task efficiency is not affected by the inclusion of a tablet-head.**

Keywords—Collaborative Robots, Social Robots, Human-Robot Interaction, Non-verbal communication, Teamwork and Collaboration

I. INTRODUCTION

Industries are rapidly introducing small table-top robotic arms to work alongside human workers on factory floors. Examples of such robots are Universal Robot-5, Sawyer, Kuka LBR IIwa. These collaborative robotic arms or ‘cobots’ are designed to meet safety, accuracy and repeatability requirements. However, the likeability of cobots is usually not a design consideration. There have been reports of human workers’ unwelcoming attitude towards these collaborative robots. This may result in human capital loss and reduced productivity for the industry. Thus, there is a need to make likable collaborative robots.

At the other end of human-robot interaction, there are social robots or ‘sobots’ which are designed primarily for a better user experience. So, they can communicate verbally or through gestures, detect emotions and provide entertainment. But they are not designed for accuracy and repeatability required in an industrial application.

In this project, we build and evaluate a ‘co-so-bot’ (hybrid of collaborative and social robots) by combining a robotic arm and a tablet-on-neck social robot. Our hypothesis is that adding a simple tablet-head, which is capable of non-verbal communication with the human-teammate, improves the likability of a collaborative robotic arm, even though the arm and tablet-head are not physically connected as a single entity. In industrial tasks the efficiency of task execution is of prime importance. So, our second hypothesis is that the ‘co-so-bot’ approach increases the efficiency of teamwork. To evaluate these hypotheses, we run a within-subject experiment with human participants for two conditions. In the first condition, the participants perform a joint task with robotic arm alone and in the second condition, they perform the same task with our ‘co-so-bot’ configuration. We measure the likeability of each system using the Godspeed Questionnaire [1]. Also, we measure the efficiency of teamwork from video recordings of the experiments.

II. RELATED WORK

A. Rise of Industrial Collaborative Robots

The world’s first industrial collaborative robot was introduced by Universal Robots in 2008 [2]. After 2012, larger robot manufacturers such as Kuka, ABB and Fanuc launched their own products in this new class of ‘cobots’. However, all these robots are designed as simple robotic arms without any mechanism for social interaction with the human teammates. In 2015, Rethink Robotics released their collaborative robot called ‘Sawyer’ which has animated eyes displayed on a screen attached to the robot arm. Our ‘co-so-bot’ configuration differs from this design as we place the simple tablet-head outside of the workspace, so that it does not affect the reachability and task capabilities of the robot arm. Another advantage of our approach is that it is independent of the design of robot arm itself and hence compatible with different models of collaborative robot arms.

B. Evaluation of Human-Robot Teamwork

Hoffman and Breazeal (2004) [3] first developed the framework for joint-task execution by a human-robot team. Breazeal et al. (2005) [4] evaluated the effects of nonverbal communication on efficiency and robustness in human-robot teamwork. S. Lee et al. (2011) [5] evaluated the effects of appearance and functions on likability and perceived occupational suitability of robots. K Fischer et al. [6] evaluated the effects of social gaze in human-robot collaborative assembly. We compare the likeability and efficiency of a collaborative robot arm with and without the accompanying social tablet-head.

III. SYSTEM OVERVIEW

We built the hybrid ‘co-so-bot’ for performing joint tasks with a human by combining two independent robots and a vision sensor. As Aristotle rightly said ‘Whole is greater than the sum of its parts’, the synergy between these parts enables the robot to perform complex tasks with a human teammate. The system is shown in Figure 1.

Figure 1: System consists of a robot arm, tablet-head social robot and vision sensor

A. *Hardware*

a) Robotic Arm: We use the ‘WidowX Mark II’ robot manipulator built by Interbotix Labs. This robot arm is suitable pick-and-place tasks with medium lifting strength and high repeatability. It weighs 1.33 kg and has horizontal reach of 41 cm and vertical reach of 55cm. It can lift weights up to 0.4kg at 30 cm and 0.8kg at 10cm. This robot arm has 5 degrees of freedom (DoF). The actuators consist of two MX-64 Dynamixel Motors for shoulder and elbow, one MX-28 Dynamixel Motor for base rotation, one MX-28 Dynamixel Motor and one AX-12A Dynamixel Motor for the wrist. It is also equipped with one DoF parallel gripper with an AX-12A Dynamixel Actuator.

b) Tablet-head: Our system includes a simple robot head to communicate with the human teammate through non-verbal cues like gazes and nods. It is developed by HR2C Lab at Cornell and consists of a Tablet Computer (Samsung Galaxy 3 Tab) mounted on a 4 DoF platform. The platform has four MX-28 Dynamixel actuators for pan, N-tilt, H-tilt and base rotation. In our experiments, we display a static image of two eyes on the tablet and control the actuators for signaling gazes and nods. In future, the eyes can be animated for better interaction.

c) Vision sensor: The functional ‘eye’ of the robot is a ‘Orbbec Astra’ sensor. It is a powerful standalone 3D camera with VGA color and range of 0.6m to 8m.

d) Electronics: The robot arm and simple head are powered using 12V adapters. The arm is controlled using a ArbotiX-M Robocontroller which is connected to the Laptop running ROS through a USB.

B. *Software*

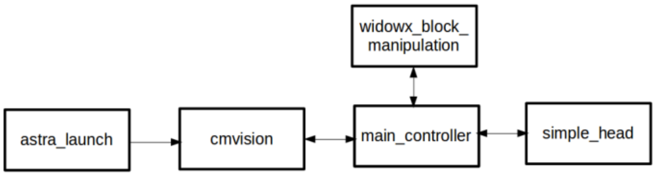

Figure 2: Modular software using ROS framework

In recent years the robotics community has made tremendous progress in software development for robots using Robot Operating System (ROS). ROS is an open source software framework that provides operating-system like capabilities for controlling low level hardware. By making the work of robotics community easily accessible to everyone, ROS enables us to ‘stand on the shoulders of giants’.

As shown in Figure 2, we developed the software for controlling the robot in a modular way using the ROS framework. The low-level control of each component takes place in separate nodes. A *main_controller* node communicates with these nodes via ROS topics and ROS action server mechanism.

a) Controlling the Robot Arm: Interbotix has developed a ROS package called ‘*widowx_arm*’ for controlling the WidowX arm. This provides a ROS action server for commanding the robot arm to pick up objects and place them at desired locations. We wrote a client node to send position goals to this action server. We can also configure the gripping width to match the dimensions of objects to be picked up.

b) Controlling the Tablet-Head: We use the '*simple_head*' package developed by Guy Hoffman at Cornell, to control the motion of tablet-head robot. The desired joint-state poses are stored in a YAML file and the robot can be commanded by publishing the desired pose on '*\goto_pose*' topic.

c) Tracking colored blocks: Orbbec has released the '*ros_astra_camera'* package to interface Orbbec Astra vision sensor with ROS. Their '*ros_astra_launch'* package loads all nodelets to convert raw depth/RGB/IR streams to depth images, disparity images and registered point clouds. We use the '*cmvision*' package developed by Nate Koenig to track different colored blobs in the RGB data from camera.

IV. COLLABORATIVE EXECUTION

A. *Sample Task Description*

The robotic system described in previous section can perform complex structured tasks involving pick-and-place operations of colored blocks. For this project, we consider this sample task: *Different colored blocks are randomly scattered in the workspace. They must be picked up and placed in the target bins. Each bin corresponds to one color and the blocks can be picked up in any order. The robot is required to do this task jointly with a human*. There can be multiple real-life counterparts of this task. For example, in medicine factory a collaborative robot and human partner need to sort medicines into different boxes or in kitchen a domestic robot and human need to put spices or grains in different containers.

The high-level decision-making system of the robot is shown in Figure 3. It is implemented as a state-machine in the '*main_controller*' node. The states are defined as follows:

- *LookAtWorkspace*: Simple-head turns towards workspace
- *LookAtTarget*: Simple-head turns towards target area
- *LookAtPerson*: Simple-head turns towards person
- *LookAtRobot*: Simple-head turns towards the robot
- *Nod*: Simple-head nods in agreement
- *Anti-nod*: Simple-head nods in disagreement
- *MoveBlockNext*: Robot arm moves block to target bin

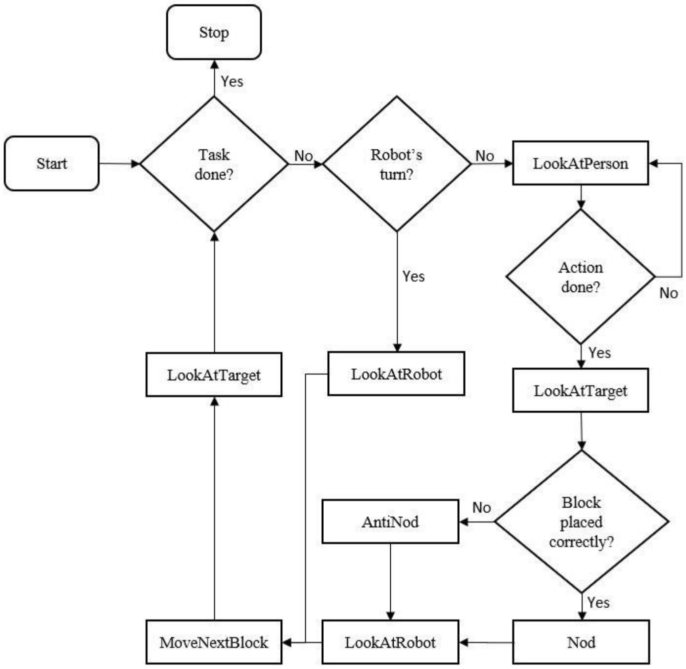

Figure 3: High-level decision making system of robot (implemented as a state-machine in *main_controller* node)

The behavior of the robot satisfies these teamwork requirements [3]:

- Turn taking: Robot looks at human after completing own turn and waits for human to finish their turn.
- Dynamic meshing of sub plans: Robot chooses its actions based on the common goal and other teammates' actions. The robot keeps track of completed subtasks (correctly sorted blocks) and decides the next action based on unfinished sub-tasks.
- Mutual support: If the human places a block in incorrect target bin, in the next turn the robot places that block in correct target bin. Currently we do not have any gesture for requesting support from the human.
- Mutual belief: Robot looks at the target area after human finishes their action. Robot nods in agreement if the human moved block to correct target bin. Robot nods in disagreement if the human moved block to incorrect target bin.

V. EXPERIMENT

We conduct a within-subject experiment with human participants to test following hypotheses:

1. Robot is more likable in Condition-H than Condition-NH
2. Task efficiency is more in Condition-H than Condition-NH

Here conditions are defined as:
Condition-H: Robotic arm and the tablet-head both (shown in Figure 4)
Condition-NH: Only the robotic arm (shown in Figure 5)

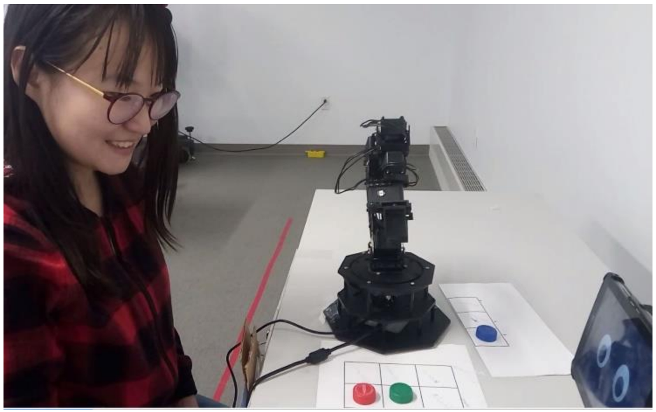

Figure 4: In Condition-H participants perform the block sorting task with the robot arm and the tablet-head

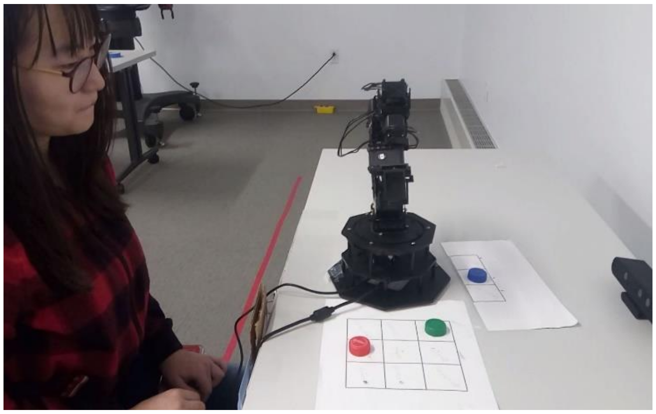

Figure 5: In Condition-NH participants perform the block sorting task with the robot arm

A. *Procedure*

Three blocks of different colors (Red, Green, Blue) are placed on the table in some initial configuration. The target bins are marked by boxes. The human participants are first given a description of the task. They are asked to take alternate turns with the robot to move the blocks to target bins. It is explained that they can place the blocks in incorrect bins if they forget the correct target bin. Each participant performs this task for both Condition-H and Condition-NH. At the end of each task, they are asked to fill up a questionnaire. The questionnaire is based on Godspeed Questionnaire series [1]. It covers concepts such as likability, anthropomorphism, perceived intelligence and perceived safety. The users are asked to give ratings for impression of the robot on a scale of 1 to 5 for different questions which measure the same variable. The questions are arranged in a random order.

Since this is a within-subject experiment, the order of Condition-H and Condition-NH is kept random to reduce the ordering effect. In the pilot experiment, we found that the block-tracking algorithm was not robust and affected the outcome of the experiment. So, to remove the effect of this additional independent variable, the initial positions of blocks are kept same in all the experiments. Also, one of the experimenters manually enters the user's action using keyboard. To eliminate delays in this manual input, we use shorthand notations for each action of the user. For example, 'r1' is used for user's action of moving 'Red' block to bin '1'.

B. *Participants*

A total of 7 participants completed the experiments. None of them were part of the HRI class and had never interacted with the robot before. They were nearly evenly mixed gender (4male, 3 female) and ranged in age from 22-29. No monetary compensation was given to the participants.

C. *Data Analysis and Results*

1. *Measuring likability*

We use the data gathered from questionnaires to measure likeability of each condition. There are 5 questions in the questionnaire that measure 'likability'. We take arithmetic mean of each user's ratings for these 5 questions for Condition-H and Condition-NH. Table 1 shows the mean, standard deviation, minimum and maximum likability ratings for Condition-H and Condition-NH.

TABLE 1: USER LIKEABILITY RATINGS FOR TWO CONDITIONS

	Condition-H	**Condition-NH**
N	7	7
Mean	3.971	3.264
Std. Deviation	0.4821	0.7341
Std. Error	0.182	0.277
Minimum	3.200	2.250
Maximum	4.600	4.200

We run a one-sided t-test to test the hypothesis that mean likability is greater in Condition-H than Condition-NH. Table 2 shows the results of this t-test. On average, participants liked the combination of robot arm and tablet-head (M = 3.971, SE = 0.182) than only the robot arm (M = 3.264, SE = 0.277). Thus Hypothesis-1 is confirmed: t (6) = 2.395, p = 0.027 and Effect size = 0.905 (Cohen's d).

TABLE 2: PAIRED SAMPLES T-TEST (LIKABILITY OF ROBOT)

		t	df	p	Cohen's d
Condition H	- Condition NH	2.395	6	0.027	0.905

2. *Measuring efficiency*

We analyze the video recordings of each experiment to measure time taken by robot and human for completing each turn. One interesting observation is that some participants intentionally move the blocks to incorrect bins to test the robot. So total task completion time is not a good measure of task efficiency in this case. Therefore, we use the average time taken per turn by each teammate to measure the task efficiency. It is calculated by taking arithmetic mean of time taken for completing each turn by each agent. We also calculate the net average time per turn for the entire task by dividing task completion time with total number of turns taken by both the agents. The efficiency is interpreted as the inverse of average time taken per turn. So lower average time per turn means higher efficiency. Table 3 shows the average time per turn taken by each agent and the net average time per turn for the entire task.

TABLE 3: AVERAGE TIME PER TURN (S)

Agent	**Robot**		**Human**		**Net**	
Condition	**H**	**NH**	**H**	**NH**	**H**	**NH**
N	7	7	7	7	7	7
Mean	26.72	26.28	5.589	4.615	16.15	15.45
Std. Deviation	2.160	2.015	1.479	0.7259	1.702	1.120
Std. Error	0.817	0.762	0.559	0.274	0.643	0.423
Minimum	22.53	22.89	2.735	3.750	12.63	13.70
Maximum	29.35	28.82	7.245	5.560	17.44	17.19

We run a one-sided t-test to check whether the average time taken per turn by robot is more in Condition-NH than Condition-H. The results of this t-test are shown in Table 4.

TABLE 4: PAIRED SAMPLES T-TEST (ROBOT'S AVERAGE TIME PER TURN)

		t	df	p	Cohen's d
Condition H	- Condition NH	0.523	6	0.690	0.198

The difference between average time taken per turn by the robot in Condition-H (M= 26.72s, SE = 0.817s) and Condition-NH (M = 26.28s, SE = 0.762s) is not significant: t (6) = 0.523, p = 0.690, Effect size = 0.198 (Cohen's d). This means that the robot takes same average time per turn in both conditions. So, hypothesis-2 is not confirmed.

We run a one-sided t-test to check whether the average time taken per turn by human is more in Condition-NH than Condition-H. The results of this t-test are shown in Table 5.

TABLE 5: PAIRED SAMPLES T-TEST (HUMAN'S AVERAGE TIME PER TURN)

		t	df	p	Cohen's d
Condition H	- Condition NH	1.453	6	0.902	0.549

The average time taken per turn by the human in Condition-H (M= 5.59s, SE = 0.559s) is not less than Condition-NH (M = 4.62s, SE = 0.274s). Thus, the hypothesis-2 is not confirmed: t (6) = 1.453, p = 0.902.

To evaluate if there is any difference between the net average time per turn for entire task, we run a two-sided t-test. The results of this t-test are shown in Table 6.

TABLE 6: PAIRED SAMPLES T-TEST (NET AVERAGE TIME PER TURN)

		t	df	p	Cohen's d
Condition H	- Condition NH	1.018	6	0.348	0.385

The difference between net average time taken per turn in Condition-H (M= 16.15s, SE = 0.643s) and Condition-NH (M = 15.45s, SE = 0.423s) is not significant: t (6) = 1.018, p = 0.348. This means that the efficiency is same in both conditions.

D. *Discussion*

This experiment evaluates the effect of combining a social robot with a collaborative robot arm. The robot arm alone can also perform a joint task with human teammate. Though addition of a social-robot does not give additional functional capabilities to the arm, our results confirm that the likability of the robot increases if it is accompanied by a simple-head robot. This simple-head robot performs non-verbal communication with the human-teammate through gazes and nods. We also find that the average time taken by robot and the human is not significantly different in Condition-H than Condition-NH. This means that the efficiency of robot and human is almost the same in both conditions. However, participants informally reported that the time taken by robot was perceived to be less in Condition-H than Condition-NH. This may be attributed to the fact that in Condition-H participants were engaged with the social-robot head when the robot arm was in idle state.

Another important observation is that the average time taken by human per turn is significantly less than the average time taken by robot per turn. The limitations on robot's speed come from the sequential execution of actions in a pick-and-place operation. The *pick_and_place_action_server* of *widowx_arm* package executes a pick-and-place operation in this sequence: start from idle position, open gripper, hover over the starting position, go down in z-direction, close the gripper, go up in z-direction, hover over the target position, go down in z-direction, open the gripper, go up in z-direction, go back to idle position. This sequential method is not an efficient way of implementation. The speed can be significantly increased by performing parallel operations.

VI. CONCLUSION

In the first phase of this project, we built a platform for human-robot collaboration by combining off-the-shelf robot arm, simple-head robot and a vision sensor. Individually, the robot arm performs pick-and-place operations, tablet head carries out non-verbal communication with the human teammate and vision sensor tracks state of the environment. We connected these individual components through a main controller in ROS framework. To make the robot autonomous, we developed its high-level decision-making algorithm based on teamwork requirements found in HRI literature. Currently, this robotic system can perform joint tasks with a human that involve pick and place operations. In future, we would like to extend it to other dexterous manipulation tasks with the use of appropriate gripper.

In the second phase of this project, we conducted human-subject experiments to compare the likeability and task efficiency of robot arm with and without the tablet-head. Human participants were asked to perform a block-sorting task by taking alternate turns with the robot. Analysis of the participants' responses to questionnaire confirms our hypothesis that a tablet-head increases the likeability of a collaborative robot arm. Also, the analysis of video recordings shows that there is no significant difference in the efficiency of two conditions. The major drawback in this experiment is 'novelty effect'. As none of the participants had interacted with the robot before, they may have given high likeability ratings in the experiments. It is possible that once the novelty associated with the robot wears off, the likeability of both the conditions (with and without the social-head robot) becomes same. Thus, future work includes a long-term experiment of collaborative robots working with human partners.

REFERENCES

[1] Bartneck C., Croft E., Kulic D. & Zoghbi S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. International Journal of Social Robotics, 1(1) 71-81

[2] History of the Cobots, Universal Robots, 2016

[3] Hoffman G, & Breazeal C (2004). Collaboration in Human-Robot Teams. In Proc. of the AIAA 1st Intelligent Systems Technical Conference, Chicago, IL, USA. [PDF]

[4] Breazeal C., Kidd C., Thomaz A, Hoffman G. & Berlin M. (2005), Effects of nonverbal communication on efficiency and robustness in human robot teamwork, Intelligent Robots and Systems (IROS)

[5] Lee S., Lau I. & Hong Y. (2011), Effects of appearance and functions on likability and perceived occupational suitability of robots, Journal of Cognitive Engineering and Decision Making, v.5, issue 2, pp.232-250

[6] Fischer K, Jensen L., Kirstein F., Stabinger S., Erkent O., Shukla D. & Piater J., (2015), The Effects of Social Gaze in Human-Robot Collaborative Assembly, International Conference on Social Robotics, Lecture Notes in Computer Science, vol 9388. Springer, Cham