On the Scalability of Spatially Embedded Human Multi-Robot Interfaces

Shokoofeh Pourmehr, Jens Wawerla, Richard Vaughan and Greg Mori
School of Computing Science, Simon Fraser University, Burnaby, BC, Canada
{spourmeh, jwawerla, vaughan, mori}@sfu.ca

Abstract— **We review three spatially embedded interface designs for multi robot systems in real world environments. We show that adding the ability to create and command groups of robots improves interface effectiveness. We use the amount of time a human requires to interact with multiple robots as a measure of interface efficiency. A taxonomy for human multi-robot interaction design is presented, and validated by experimental evidence, for different modes of interacting with multiple robots. We present *Sequential Selection Concurrent Commanding (SSCC)* and *Concurrent Selection Concurrent Commanding (CSCC)* as two possible methods of robot group selection and commanding. We also show that in real world settings, the interaction time is affected by the spatial configuration of robots with respect to each other and to the user.**

I. INTRODUCTION

The number of robots that can be concurrently controlled by a single human operator has been of much interest in the human robot interaction community [1], [2]. The motivation is straightforward: more robots per human means more work can be done for the same human effort. The robot/human ratio is affected by the robots' degree of autonomy, assigned task, complexity of the task environment, interface design, human skill level, and workspace constraints. To date, most of the models proposed to predict an upper-bound on this ratio have been evaluated in simulation without considering the challenges that real robots and spatially limited workspaces impose [3], [4]. In this paper, we focus on scalability in *spatially embedded interfaces* for human multi-robot systems (HMRS). By spatially embedded interfaces, we mean those in which a human operator interacts with an individual or group of robots in a shared physical workspace, mediated at least partly by the robot's sensors. This is an important class of interfaces, since it allows what are arguably 'natural' means of communication such as speech, gestures and face engagement. We have demonstrated several such HMRS interfaces, and suggested that they provide engaging and intuitive means of interaction, plus other advantages [5]–[7]. However, embodied, spatially embedded, sensor-mediated systems have fundamental scalability constraints. In every case there will be a practical upper-bound of the robot/human ratio due to workspace and sensor limitations. For example, using computer vision the distance and angle of incidence inside which a robot can detect a human is limited by the chosen lens, camera resolution and algorithms, and a finite number of robots can be physically located inside this space.

Following Olsen [2] and others, we use the amount of time it takes for a human to interact with a single or a group of robots to indicate the human-robot interface efficiency. The smaller this time, the better. We decompose interaction time into three components: (i) the amount of time required to **select** a single or multiple robots, i.e. to acquire their attention and make them ready to receive control commands; (ii) the time required for the user to **switch** her attention from one robot or group to the next robot or group; and (iii) and the time required to **command** a robot or group, including acknowledgement from the robot(s). This simple model can be mapped onto most interaction designs, and captures enough information to classify designs in terms of their scalability. We introduce below a taxonomy of HMRS interfaces, distinguished by the growth rate of the individual time components with robot population size. We then classify systems from the literature in these terms, and show empirical evidence that the interaction time grows as predicted by the model. Of immediate practical interest is the result that selecting and commanding groups of robots concurrently can be shown to decrease the total per-robot interaction time in our examples, though the degree of concurrency is strictly limited by the available workspace.

The contributions of this paper are: (i) an analysis of the limitations on scalability of spatially embedded HRI methods; (ii) a taxonomy of possible HRI designs classified by scalability; and (iii) evaluation of previous systems in these terms, including original empirical measurements that (mostly) agree with the model. It should be possible to classify future systems in these terms, and make reliable predictions about their scalability.

II. BACKGROUND

One of the emerging efforts in human-robot interaction research is establishing standards and evaluation metrics [8]–[14]. Murphy and Schrechenghost [15] identify 42 proposed metrics which are categorized based on the object being directly measured: the human, the robot, or the system.

This paper builds on work that predicts system efficiency as a function of the number of robots that a single human can interact with. Olsen et al. [2] introduced the concept of *Fan-out* which posits a model-based upper-bound on the number of independent homogeneous unmanned vehicles (UVs) that a single human can interact with. This model has been modified to include wait times [11]. Goodrich et al. [16] extended it to the domain of heterogeneous robots and included switching cost. Zheng et al. [17] used *Fan-out* to evaluate different interaction models for supervisory multirobot control in real-world settings. Our paper extends this idea to explicitly consider spatial constraints as well as time.

There have been several studies [18]–[21] analyzing how the human-robot interface and team size can affect system effectiveness and performance. These models have been evaluated in simulation with traditional human-computer interfaces without taking into account the challenges presented by real world settings.

A key enabler to allowing a human operator to control a large population of robots is the ability to command multiple robots in parallel. Multiple robots can be selected and identified as a group, and the whole group is commanded with a single interaction. Parasuraman et al. [20] compares individual robot selection and group selection interfaces in a computer game scenario. They show that comparing command of individual robots versus command of groups, the participants won significantly more games without a statistically significant difference in workload. Another study [22] shows that for tasks including two robots participants use significantly fewer group selections than individual selections; in tasks with three or more robots, they use significantly more group selections than individual selections: at some critical population size, users appear to switch from favouring individual to group control to reduce their workload.

In this paper, we focus on assessing scalable embodied, sensor-mediated interfaces for human multi-robot systems (HMRS) in real world environments while examining how the mechanics of composing teams for concurrent control will affect interface effectiveness. To this end, we will discuss two distinct methods of creating groups of robots which we name: *Sequential Selection* and *Concurrent Selection*. These are directly analogous to the familiar CTRL-click and SHIFT-click group selection methods used in desktop computer graphical user interfaces (GUIs). To evaluate these methods, Mizobuchi and Yasumura [23] compared tapping with circling for multi-target selection, regarding accuracy, execution time and shape complexity. In circling, the targets must be surrounded to be selected. In tapping each target must be clicked on to become part of the selected group. They showed that circling is faster than tapping for highly cohesive targets and it is relatively insensitive to changes in the size of the individual targets. However, tapping selection time is significantly affected by size and spacing of the targets.

III. INTERACTION TIME

Olsen et al. [2] introduced Fan-out, $(F)^1$ as a measure for the number of robots a human operator can command. Fan-out is defined as the ratio of activity time (A), the time a robot operates autonomously, and interaction time (T), the expected amount of time that a human must interact with one or a group of robots.

$$F = \frac{A}{T} \tag{1}$$

Activity time (A) is generally a function of the robot's degree of autonomy faced with a certain task complexity, while interaction time (T) is proposed as an essential metric for human-robot interaction efficiency [18], where shorter interactions are more efficient than longer ones. Consequently, designing interfaces that produce small interaction times is a strong theme in HRI [13].

In multi robot systems, interaction time (T) can be decomposed into three components: i) switching time (W); ii) robot monitoring and selection time (L); and iii) command expression time (C):

$$T(n) = \sum_{i=1}^{n}(W_i + L_i + C_i) \tag{2}$$

where:

- $T(n)$ is the amount of time the operator needs to interact with n robots,
- W_i is the amount of time the user requires to switch her attention to robot i,
- L_i is the amount of time required to select robot i,
- C_i is the amount of time required to issue the command to ith robot.

Note that when the user switches her attention to a particular robot and selects it, she must take a moment to receive feedback from the newly-selected robot to confirm selection before controlling the robot. This time is included in the selection time (L).

In a real world setting, with embodied interfaces, W_i, L_i and C_i are functions of the interface design, communication method, physical workspace, spatial arrangement of the user and robots, and the amount of time needed by the robot to analyze the input signal. Also, it may be necessary to repeat the selection and command phases as necessary to compensate for sensing or processing failures in the robot. Therefore a human operator can have substantially different interaction times with individual robots, or the same robot at different times. However, under the assumptions of homogeneous robots and identical relative user-robot positions, Eq. (2) can be simplified to:

$$T(n) = n \times (W + L + C) \tag{3}$$

This interaction mode is *Sequential Selection Sequential Commanding (SSSC)*. Interaction time scales linearly with robot population size, assuming robots are on average evenly distributed in the environment.

We can use SSSC and Eq.(3) as the baseline performance for the interaction time of a HMRS. Intuitively, interaction designs that add concurrency should scale better than this baseline. In the following sections, we explain how adding the ability to form and command robot teams can achieve this.

---

[1] For descriptive purposes, we have modified the variables in the Fan-out equation.

IV. CONCURRENT COMMANDING

In SSSC Eq. (3), the user can control one robot at a time. However, by dynamically forming groups of robots, the operator is able to command all selected robots at once [5]–[7], [22], [24], [25]. This theoretically can reduce the amount of time needed to command n robots from linear time (nC) to constant time (C). As a result of this concurrent commanding, the overall interaction time should decrease. To evaluate how team make-up will improve system efficiency, we will examine two methods of selecting robots using spatial embodied interfaces in real world settings.

V. A TAXONOMY

We can now distinguish four classes of interaction that differ in their scalability:

- SSSC: Sequential Selection Sequential Command
- SSCC: Sequential Selection Concurrent Command
- CSCC: Concurrent Selection Concurrent Command
- CSSC: Concurrent Selection Sequential Command

CSSC seems unused, so we omit it. We examine how different selection modes will affect overall interaction time using available interfaces in the literature.

A. Sequential Selection Concurrent Command (SSCC)

In computer user interfaces, typically there are two ways to select multiple files or folders using keyboard and mouse. One way is to hold down the CTRL key, and then click each desired item. In this method, the user can select a nonconsecutive group of files or folders. Similarly, in HMRS, the user can sequentially select a desired robot and add it to the group. Once the team is formed, the operator can issue a command for the robots to perform a common task. Since commanding is simultaneous for all selected robots, the time required is just C. We call this method *Sequential Selection Concurrent Command (SSCC)*. The interaction time is:

$$T_{sscc}(n) = W_{sscc} + L_{sscc} + C_{sscc} \tag{4}$$

where

$$\begin{aligned} W_{sscc} &= n \times W \\ L_{sscc} &= n \times L \\ C_{sscc} &= C \end{aligned}$$

As shown in 4, we predict that switching time W_{sscc} and selection time L_{sscc} scale linearly with the number of robots. However, the command time C_{sscc} will be same as baseline (i.e. the required time for commanding a single robot and a group of robots will be the same). Here, we assume that there is no failure in receiving selection or commanding signals by the robots. In the literature, we identified two interfaces for team make-up using this selection method. Monajjemi et al. [7] used face engagement and gestures to add/remove robots to/from a group. Similarly, in our previous work, where we extended single-robot selection by face engagement [26], we proposed a system [6] which integrates spoken commands and face engagement to dynamically create and modify teams of robots.

For both of these interface designs, C_{sscc} is almost the same for interacting with one robot or with a team. As a result, due to simultaneous commanding, the model predicts a reduction in T_{sscc} compared to the baseline.

B. Concurrent Selection Concurrent Command (CSCC)

The second method for selecting a consecutive group of multiple folders or files in computer interfaces, is to drag the mouse pointer to create a selection rectangle around the outside of all the items to be included. The selection result is identical to clicking the first item, holding down the Shift key and then clicking the last item. In their work, Milligan et al. [5], Skubic et al. [24] and Micire et al. [22] applied *Concurrent Selection Concurrent Commanding* where the user can make a group selection by circling around the desired group of robots. We call this method *Concurrent Selection Concurrent Commanding (CSCC)*. The interaction time of this interface is composed of the time to switching to a group of desired robots, selecting them and send them a command:

$$T_{cscc}(n) = W_{cscc} + L_{cscc} + C_{cscc} \tag{5}$$

Similar to SSCC, the operator issues only one command to the formed group. Therefore the predicted interaction time of this interface is smaller than of the baseline. In principle, the selection time L_{cscc} is a function of the number of robots to be selected and their spatial layout (i.g. for selecting more robots, the user has to draw a larger circle). However, for most settings, this increase is relatively insignificant and we therefore assume L_{cscc} to be independent of the number of robots.

VI. SPATIAL CONSTRAINTS FOR REAL WORLD HMRS

The preceding discussion has introduced a model of the interaction time for HMRS without considering the challenges offered by spatial constraints for real world human multi-robot systems. Real robots are embodied and have to share their physical space with co-located human operator(s) and other robots; they are also situated and their abilities to deal with the world are limited by sensors and actuators.

There are situations where a user cannot interact with an individual or a group of robots without changing her location. If the user has to turn or walk through the workspace to be able to interact with a new group of robots, she needs to spend some time to switch attention to the robots in the new location. These switches require substantial movement by the human, and the resultant switching times are significantly different from the ones in Eqs. (4) and (5). In this case, we split the workspace into sub-spaces that the user can stand still in and interact with all robots in the sub-space. Formally, we divide the whole workspace into M sub-spaces. The maximum number of robots that fits in sub-space j is denoted by N_j. Therefore, Eqs. (4) and (5) are valid for $1 \leq n \leq N_j$. By summing over all the interaction times and switching times, we can calculate the general interaction time as:

$$T_G = \sum_{j=1}^{M} (T_j^* + W_j^*) \tag{6}$$

where:

- T_j^* is the amount of time needed to interact with a group of robots in sub-space j ,
- W_j^* is the amount of time the user needs to change locations to start interacting with a group of robots in sub-space j .

VII. EXPERIMENTAL EVALUATIONS

We studied three different interface designs to evaluate our hypothesis that composing robot teams can improve interaction time and system efficiency compared to a per-robot baseline.

The amount of time needed to switch attention to a robot, select it and receive its feedback, is considered as $(W + L)$. (C) is measured from the moment the user starts expressing the command (by gesture or speech) until the moment she gets feedback from the robots. (T) is the sum of the terms mentioned as in Eqs. (4) and (5).

A. Baseline

In experiments 1, 2 and 3, we compare interaction time of different group selection methods with our sequential baseline SSSC (Eq. (3)). For every interaction method examined, we measure the interaction time with one robot (T_1), and multiply it by the number of robots in the team n, i.e. the baseline for interacting with n robots will be nT_1.

B. Experiment 1

In the first experiment, we examine how an instance of *Sequential Selection Concurrent Commanding* $(SSCC)$ affects the interaction time (T). According to Eq. (4), in interacting with n robots, the human operator will spend n switching and selection times $(W + L)$ to select each robot and add it to the group. In addition, one command time (C) is required for issuing a command to the newly created group. The resulting interaction time (T) is less than the baseline, which is the time needed to interact with n robots individually (as explained above).

Using the multimodal interface proposed in our previous work [6], we reviewed video footage of experimental trials and measured the components of the interaction time (T). In this system, we integrate spoken commands and face engagement to create, modify, and command teams of robots. The user stands in front of a population of robots and designates a subgroup by looking at them and saying the desired number of robots, e.g. "You three." Then the operator sequentially makes face engagements with the robots of interest. One challenge with this approach is to determine which robot is selected when the user's face is visible to multiple robots at the same time. To solve this problem, we used a method developed earlier by our group [26] where by counting candidate face detections, we arrived at a "face-score." This score is higher for the robot that is directly being looked at. In this interface, during every face engagement, the robot with highest "face score" is selected. Once all the desired robots are selected, the user commands the group to perform an action. To isolate the group selection method's effects on interaction time, we assume that all interface modules work perfectly.

Fig. 1: Exp. 1: Face engagement and indirect speech interface for HMRS as an example of *SSCC* method [6].

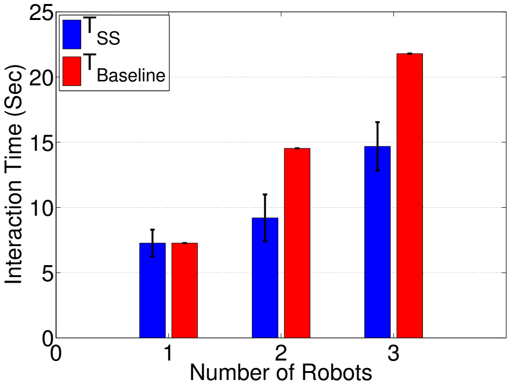

Fig. 2: Exp. 1: Comparison of interaction time of the face engagement and indirect speech interface with the baseline. $(d \approx 2.5m, \theta \approx 30deg.)$ (Sample size = 6)

TABLE I: Exp. 1: Components of interaction time of the face engagement and indirect speech interface. (Sample size = 6)

# of robots	$(\mathbf{W_{sscc} + L_{sscc}})$**(Sec)** $\mathbf{Mean(SD)}$	$\mathbf{C_{sscc}}$**(Sec)** $\mathbf{Mean(SD)}$	$\mathbf{T_{sscc}}$**(Sec)** $\mathbf{Mean(SD)}$
1	5.44 (1.05)	1.82 (0.14)	7.26 (1.03)
2	7.31 (1.77)	1.88 (0.11)	9.21 (1.80)
3	12.77 (1.96)	1.90 (0.15)	14.67 (1.86)

?? shows the average time spent to interact with multiple robots over 6 trials and compares it with the baseline. The robots were located 2.5m from the user with 30 degrees of separation and the user at the centre (Figure 1). As expected, the interaction time (T) increases as the number of robots in the team increases. However, a statistically significant difference (paired-sample t-test: $df = 5$, $p = 0.0012$ for 2 robots and $p = 0.0021$ for 3 robots) exists between the time needed to interact with multiple robots individually as in the baseline case and the interaction time (T) of this *Sequential Selection Concurrent Commanding* method. The main difference comes from the fact that the operator issues the command once for selected group, so there is a reduction in general interaction time (T) as in Eq. (**??**).

?? shows the different components of interaction time (T). It can be seen that the command time (C) remains the same for different numbers of robots in line with our model in Eq. (**??**). However, switching and selection times ($W+L$) do not increase linearly with the number of robots. This is caused by this particular interface design scheme. In this interface, switching and selection time ($W+L$) are composed of four components: i) time needed to express the spoken command, ii) processing the spoken command, iii) sequentially making the face engagement with all robots in the group and iv) iterative selection of the robot with highest "face score". Since in interacting with multiple robots, the user announces the desired number of robots once, the first two components are not affected by the number of robots. However, the last two components of switching and selection time ($W+L$) depend on the number of robots being selected. These two components are also sensitive to the spatial arrangement of the robots and the user, which will be examined later in experiment 4.

C. *Experiment 2*

In a second experiment, we examine the effect of *Sequential Selection Concurrent Commanding (SSCC)* on the interaction time and system efficiency with another spatially embedded interface design for HMRS proposed previously by our group [7]. For this interface design, the operator has to change his location to be able to interact with a new robot. Our hypothesis is that reduction in interaction time is not substantial due to the fact the user needs to change his location to initiate an interaction with another robot.

Using this interface, the user is able to compose a multi-robot team from a population of robots. The user starts the interaction with each robot, by standing in front of it and making a face engagement with it. When the robot confirms the engagement, the user adds it to the team by a right hand wave gesture (Figure 3). Waving both hands, he commands the entire group to execute a mission. Similar to the previous interface for *Sequential Selection Concurrent Commanding*, and based on Eq. (**??**), we expect to see a reduction in the interaction time. This is due to the time being saved by sending commands to the whole group at once. To show this reduction in interaction time we again compare against the baseline.

The average time spent on interaction with multiple robots over 5 trials, along with the baseline, is illustrated in Figure 4. The result of paired-sample t-test on this small sample size shows that there is no statistically significant difference ($df = 4$, $p = 0.3371$ for 2 robots and $p = 0.2615$ for 3 robots) between the individual and group selection. The reason is that the switching time W for changing the workspace is large and it can not be subsumed in the baseline. Table II shows that the command time C does not depend on the number of robots which agrees with our model in Eq. (**??**). The switching and selection time ($W+L$) on the other hand depends on the number of robots since the user has to walk to each robot and perform a waving gesture. The effect of the spatial embeddedness is large in this case because the robots in this experiment are quadcopters which require substantial free space for safe operation. As a result, to model the interaction time of this interface, Eq. (6) is more suitable.

Fig. 3: Exp. 2: Waving gesture interface for HMRS as an example of *SSCC* method [7].

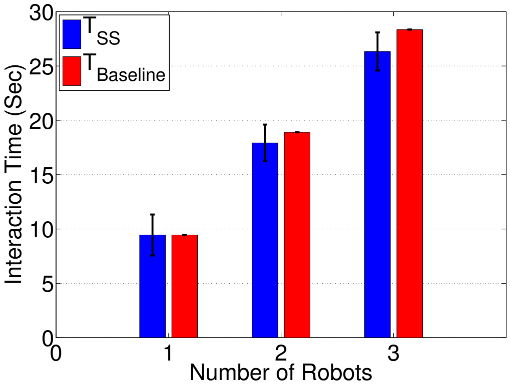

Fig. 4: Exp. 2: Comparison of interaction time of the waving gesture interface with the baseline. (Sample size = 5)

TABLE II: Exp. 2: Components of interaction time of the waving gesture interface. (Sample size = 5)

# of robots	$(\mathbf{W_{sscc}+L_{sscc}})$**(Sec)** **Mean(SD)**	$\mathbf{C_{sscc}}$**(Sec)** **Mean(SD)**	$\mathbf{T_{sscc}}$**(Sec)** **Mean(SD)**
1	5.10 (1.09)	4.35 (0.91)	9.45 (1.88)
2	13.54 (1.41)	4.38 (0.67)	17.92 (1.69)
3	21.70 (1.91)	4.64 (0.33)	26.34 (1.76)

D. *Experiment 3*

In this experiment, we examined the effect of *Concurrent Selection Concurrent Commanding* on the interaction time. We showed in Eq. (**??**) that using $CSCC$, the interaction time T will be reduced to one switching and selection time ($W+L$) and one command time C and it does not depend on the number of robots in the team. To analyze this hypothesis, another method of selecting groups of robots from a population proposed by Milligan et al. [5] was examined. In this interface multiple robots can be selected and commanded concurrently to perform a task using a vision based approach. To select robots, this method calls for the user to draw a circle around all robots she wants to select

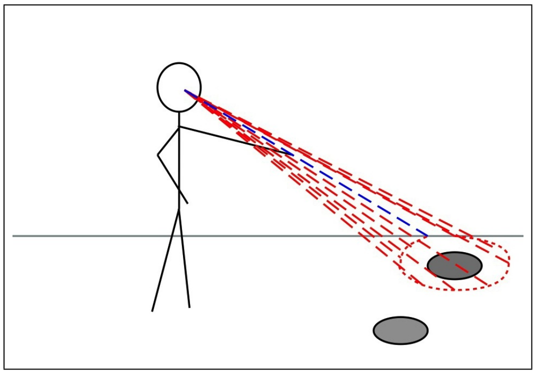

Fig. 5: Exp. 3: Circling gesture interface for HMRS as an example of CSCC method [5]

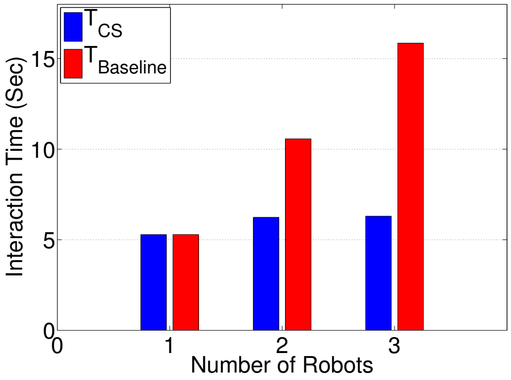

Fig. 6: Exp. 3: Comparison of interaction time of the circling gesture interface with the baseline. (Sample size = 1)

TABLE III: Exp. 3: Components of interaction time of the circling gesture interface. (Sample size = 1)

# of robots	$(\mathbf{W_{cscc}} + \mathbf{L_{cscc}})$**(Sec)**	$\mathbf{C_{cscc}}$**(Sec)**	$\mathbf{T_{cscc}}$**(Sec)**
1	1.82	3.47	5.29
2	2.59	3.65	6.24
3	2.58	3.72	6.29

(Figure 5). In this way, robots in the circle get selected and assigned to a common group. The robots can determine whether they are circled by the user by tracking the user's hand and face. After selection, a command is issued to the team using a pointing gesture.

The data from reviewing the video footage of this interface is presented in Table III and Figure 6. The baseline is measured in similarly to the previous experiments. For this experiment only one sample was available for different numbers of robots. Figure 6 shows the amount of time the user spends to select and command different size groups. Comparing the interaction times with the related baselines, we can see a reduction in the interaction time T, which agrees with our hypothesis. Table III shows that for interacting with multiple robots, both commanding time C and switching and selection time $(W + L)$ do not depend on group size, which supports our hypothesis.

E. Experiment 4

Our hypothesis for this experiment is that the amount of time a user spends to interact with a group of robots is affected by the spatial configuration of the robots with respect to each other and the user.

To test this hypothesis, we use the experimental results of the interface in Exp. 1 [6], where a user can select multiple robots by announcing the number of desired robots and sequentially looking at them. The amount of time that the user spends to interact with three robots is measured in various spatial arrangements of robots with respect to the user and each other. In every experiment, robots are located d meters from the user on a circle with θ degrees of separation. d ranges from 1 to 2.5 m with 0.5 m steps while θ ranges from 30 to 90 degrees with steps of 15 degrees. Each experiment is repeated five times. The data are presented in Figure 7. The results show that the interaction time T is affected by the user-robot distance as well as the angle between the robots. When the distance increases, it takes more time for robots to get selected because the user's face is harder to detect. This is caused by the limitation of camera resolution. The increase in the interaction time with the robots' separation is caused by the fact that the user has to switch between robots and directly look at them. When the robots are close to each other, the user's face is visible to all robots so when the user switches between robots, there is no need for robots to start detecting the user's face. However, by increasing the angle between robots, the interaction time increases because it takes more time for robots to detect the user's face and become selected.

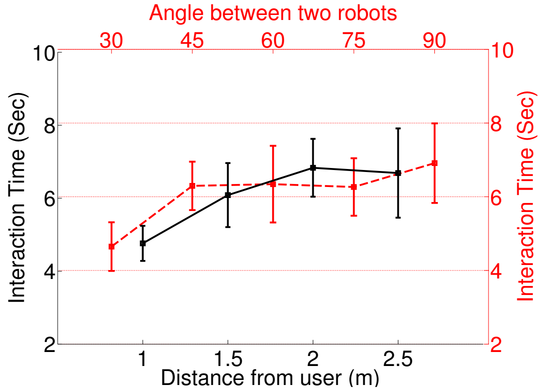

Fig. 7: Exp. 4: Comparison of interaction times for various user-robot distances and different angles from between robots. (Sample size = 20)

We previously showed that this interface design has inherent limits on the usable range and bearing between human and robot [6]. This experiment confirms that the spatial arrangement of robots also effects the interaction times.

VIII. Discussion and Future Work

In interacting with multi robot systems, a useful class of interfaces provide the ability to dynamically create, modify and command groups of robots, ideally scaling to large populations of robots and overcoming the time limitations a user has when interacting with large numbers of robots. We examined two different classes of group selection methods: *Sequential Selection Concurrent Commanding (SSCC)* and *Concurrent Selection Concurrent Commanding(CSCC)*.

In the examples we studied, the *CSCC* method exhibits a shorter interaction time than the *SSCC* method. This is due to the time reduction in concurrently selecting robots. However, *SSCC* has the advantage that the selection does not have to be consecutive. In *CSCC* method all robots to be selected have to be physically close and the cluster of robots has to be free of unwanted robots. This makes *CSCC* method good for homogeneous groups but possibly limits its application in heterogeneous groups. This method can be combined with sequential selection method for deselecting unwanted robots in the selected group.

We further showed that the spatial arrangement of robots and the relative position of the user, influence the effectiveness of the interface. These also limit the scalability of the interface. We also pointed out two types of switching costs: directing attention between robots in the local workspace, and moving to attend to another group of robots that were previously out of range. To the best of our knowledge, this is the first taxonomy of embedded HMRS interfaces.

In future work, we will expand our interaction time model to include the effects of spatial configuration of HMRS. In addition, our current model does not consider variance between interaction instances due to sensor failures or user errors. A more advanced model that incorporates the probability of success of each interaction, and the delay caused by repeats, could be useful.

IX. CONCLUSION

We have demonstrated that the amount of time required to interact with multiple robots can be reduced by different methods of creating groups of robots and doing concurrent interactions that exploit locality. Our analysis of previous experiments show that interaction time can be used to compare the efficiency of HMRS interfaces. We showed that interaction time equations can model the improvement of spatially embedded HRI designs. We also propose, and demonstrate, using experimental evidence, that in real world settings the interaction time is affected by the spatial arrangement of the workspace. This will affect the upper-bound of the number of robots that a single human can interact with.

ACKNOWLEDGEMENTS

This work was supported by the NSERC Canadian Field Robotics Network.

REFERENCES

[1] M. Lichtenstern, M. Angermann, M. Frassl, G. Berthold, B. Julian, and D. Rus, "Pose and paste; an intuitive interface for remote navigation of a multi-robot system," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, Nov 2013, pp. 1632–1639.

[2] D. R. Olsen, Jr. and S. B. Wood, "Fan-out: measuring human control of multiple robots," in *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, ser. CHI '04. New York, NY, USA: ACM, 2004, pp. 231–238. [Online]. Available: http://doi.acm.org/10.1145/985692.985722

[3] B. Trouvain and H. Wolf, "Evaluation of multi-robot control and monitoring performance," in *Robot and Human Interactive Communication, 2002. Proceedings. 11th IEEE International Workshop on*, 2002, pp. 111–116.

[4] P. Velagapudi, P. Scerri, K. Sycara, H. Wang, M. Lewis, and J. Wang, "Scaling effects in multi-robot control," in *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, Sept 2008, pp. 2121–2126.

[5] B. Milligan, G. Mori, and R. T. Vaughan, "Selecting and commanding groups of robots in a vision based multi-robot system," Video Proceedings of the 6th ACM/IEEE International Conference on Human-Robot Interaction, Lausanne, Switzerland, March 2011. [Online]. Available: http://www.youtube.com/autonomylab#p/u/0/Fq2qTlSuGqs

[6] S. Pourmehr, V. M. Monajjemi, R. T. Vaughan, and G. Mori, "You two! take off! creating, modifying and commanding groups of robots using face engagement and indirect speech in voice commands," in *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS'13)*, Tokyo, Japan, November 2013.

[7] V. M. Monajjemi, J. Wawerla, R. T. Vaughan, and G. Mori, "Hri in the sky: Creating and commanding teams of uavs with a vision-mediated gestural interface," in *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS'13)*, Tokyo, Japan, November 2013.

[8] M. A. Goodrich and A. C. Schultz, "Human-robot interaction: a survey," *Found. Trends Hum.-Comput. Interact.*, vol. 1, no. 3, pp. 203–275, Jan. 2007. [Online]. Available: http://dx.doi.org/10.1561/1100000005

[9] D. Olsen and M. Goodrich, "Metrics for evaluating human-robot interactions," in *Proc. NIST Performance Metrics for Intelligent Systems Workshop*, 2003.

[10] A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, and M. Goodrich, "Common metrics for human-robot interaction," in *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, ser. HRI '06. New York, NY, USA: ACM, 2006, pp. 33–40. [Online]. Available: http://doi.acm.org/10.1145/1121241.1121249

[11] M. Cummings and P. Mitchell, "Predicting controller capacity in supervisory control of multiple uavs," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, vol. 38, no. 2, pp. 451–460, 2008.

[12] J. W. Crandall and M. L. Cummings, "Developing performance metrics for the supervisory control of multiple robots," in *Human-Robot Interaction*, 2007, pp. 33–40.

[13] M. Goodrich and D. Olsen, "Seven principles of efficient human robot interaction," in *Systems, Man and Cybernetics, 2003. IEEE International Conference on*, vol. 4, 2003, pp. 3942–3948 vol.4.

[14] J. W. Crandall, M. A. Goodrich, D. R. Olsen, Jr., and C. W. Nielsen, "Validating human-robot interaction schemes in multitasking environments," *Trans. Sys. Man Cyber. Part A*, vol. 35, no. 4, pp. 438–449, July 2005. [Online]. Available: http://dx.doi.org/10.1109/TSMCA.2005.850587

[15] R. Murphy and D. Schreckenghost, "Survey of metrics for human-robot interaction," in *Human-Robot Interaction (HRI), 2013 8th ACM/IEEE International Conference on*, 2013, pp. 197–198.

[16] M. Goodrich, M. Quigley, and K. Cosenzo, "Task switching and multi-robot teams," in *Multi-Robot Systems. From Swarms to Intelligent Automata Volume III*, L. Parker, F. Schneider, and A. Schultz, Eds. Springer Netherlands, 2005, pp. 185–195. [Online]. Available: http://dx.doi.org/10.1007/1-4020-3389-3_15

[17] K. Zheng, D. Glas, T. Kanda, H. Ishiguro, and N. Hagita, "Supervisory control of multiple social robots for navigation," in *Human-Robot Interaction (HRI), 2013 8th ACM/IEEE International Conference on*, March 2013, pp. 17–24.

[18] J. Crandall and M. Cummings, "Identifying predictive metrics for supervisory control of multiple robots," *Robotics, IEEE Transactions on*, vol. 23, no. 5, pp. 942–951, 2007.

[19] C. Humphrey, C. Henk, G. Sewell, B. Williams, and J. Adams, "Assessing the scalability of a multiple robot interface," in *Human-Robot Interaction (HRI), 2007 2nd ACM/IEEE International Conference on*, 2007, pp. 239–246.

[20] R. Parasuraman, S. Galster, P. Squire, H. Furukawa, and C. Miller, "A flexible delegation-type interface enhances system performance in human supervision of multiple robots: empirical studies with roboflag,"

Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, vol. 35, no. 4, pp. 481–493, 2005.

[21] P. Squire, G. Trafton, and R. Parasuraman, "Human control of multiple unmanned vehicles: effects of interface type on execution and task switching times," in *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, ser. HRI '06. New York, NY, USA: ACM, 2006, pp. 26–32. [Online]. Available: http://doi.acm.org/10.1145/1121241.1121248

[22] M. Micire, M. Desai, A. Courtemanche, K. M. Tsui, and H. A. Yanco, "Analysis of natural gestures for controlling robot teams on multi-touch tabletop surfaces," in *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ser. ITS '09. New York, NY, USA: ACM, 2009, pp. 41–48. [Online]. Available: http://doi.acm.org/10.1145/1731903.1731912

[23] S. Mizobuchi and M. Yasumura, "Tapping vs. circling selections on pen-based devices: evidence for different performance-shaping factors," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '04. New York, NY, USA: ACM, 2004, pp. 607–614. [Online]. Available: http://doi.acm.org/10.1145/985692.985769

[24] M. Skubic, D. Anderson, S. Blisard, D. Perzanowski, and A. Schultz, "Using a hand-drawn sketch to control a team of robots," *Auton. Robots*, vol. 22, no. 4, pp. 399–410, May 2007. [Online]. Available: http://dx.doi.org/10.1007/s10514-007-9023-1

[25] J. Nagi, A. Giusti, L. Gambardella, and G. Di Caro, "Human-swarm interaction using spatial gestures," in *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*, Sept 2014, pp. 3834–3841.

[26] A. Couture-Beil, R. T. Vaughan, and G. Mori, "Selecting and commanding individual robots in a vision-based multi-robot system," in *Proceedings of the Canadian Conference on Computer and Robot Vision*, May 2010.