Comparing Heads-up, Hands-free Operation of Ground Robots to Teleoperation

Matthew Marge[*], Aaron Powers[†], Jonathan Brookshire[†], Trevor Jay[‡], Odest C. Jenkins[‡], Christopher Geyer[†]

[*]Language Technologies Institute
Carnegie Mellon University
Pittsburgh, PA 15213
mrmarge@cs.cmu.edu

[†]iRobot Research
iRobot Corporation
Bedford, MA 01730
aaronpowers@gmail.com
{jbrookshire, cgeyer}@irobot.com

[‡]Department of Computer Science
Brown University
Providence, RI 02912
trevorjay@gmail.com
cjenkins@cs.brown.edu

Abstract—Today, most commercially available UGVs use teleoperation for control. Under teleoperation, users' hands are occupied holding a handheld controller to operate the UGV, and their attention is focused on what the robot is doing. In this paper, we propose an alternative called Heads-up, Hands-free Operation, which allows an operator to control a UGV using operator following behaviors and a gesture interface. We explore whether Heads-up, Hands-free Operation is an improvement over teleoperation. In a study of 30 participants, we found that when operators used these modes of interaction, they performed missions faster, they could recall their surroundings better, and they had a lower cognitive load than they did when they teleoperated the robot.

Keywords-human-robot interaction; person detection; person following; gesture interaction; teleoperation

I. INTRODUCTION

As of 2011, thousands of unmanned ground vehicles (UGVs) are in use in Iraq and Afghanistan, and many hundreds of others are used in law enforcement and rescue operations. At present, however, most commercially available UGVs are teleoperated. The operator typically controls the robot using an operator control unit (OCU), often a game controller, tablet, or laptop. An operator's singular focus on the robot narrows his situational awareness and causes him to incur a high cognitive load. An operator's situational awareness is so hampered, in fact, that in battlefield situations, another soldier is typically assigned to guard the UGV operator. The goal of our research is to transform the way in which we interact with UGVs so that an operator's hands are free and head is up: Heads-up, Hands-free Operation (HHO). Fig. 1 illustrates such a transformation. On the left, the operator teleoperates the robot; his head is down and his hands are occupied with the OCU. On the right, using HHO, his head is up and he is able to carry something in his hands, here a clipboard. We have realized HHO using a combination of autonomous person following and gesture control simulated using Wizard of Oz (WoZ) methodology.

In this paper we report on the results of a study comparing HHO to teleoperation while walking. We claim that compared with teleoperation, when both operator and robot are on foot and copresent, operators using HHO conduct missions faster; they can better recall their surroundings later; and they have

(a) Teleoperation   (b) Heads-up, Hands-free Operation

Figure 1. (a) Control of the robot using teleoperation. The operator's hands are tied-up, and his head is down. (b) Control of the robot using Heads-up, Hands-free Operation. The operator can pay attention to his surroundings, and can hold something, here a clipboard. In both cases we used commercially available, military-grade robots.

lower cognitive load. We evaluated these claims in a simulated building clearing operation—a mobile procedure typical of military operations [1] (depicted in Fig. 2) and an intended use case for the UGV. In real building clearing operations, when a team enters a structure using this tactic, some members enter and clear rooms. Other members remain in corridors to prevent adversaries from slipping away. In the study we conducted, the UGV's function was to monitor corridors while the operator enters and exits rooms. Each participant in our study attempted to complete this simulated operation using teleoperation and using HHO. We then evaluated each operator's time-on-task, cognitive load, and ability to recall surroundings (situational recall), using standard measures found in the human-robot interaction (HRI) literature [2]. We found that compared with teleoperation, on average, participants' time-on-task was 12% lower; their situational recall was 18% higher; and their cognitive load was 14% lower. These findings support our intuitive expectations and demonstrate that HHO is a promising alternative to teleoperation.
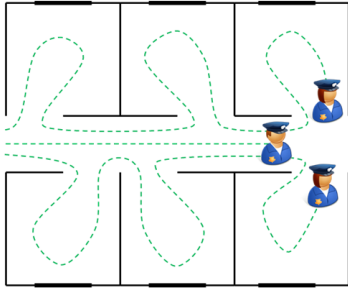
Figure 2. In this room-by-room clearing tactic, adapted from [1], some team members enter rooms while others monitor the hallway.

In summary, in this paper we make the following novel contributions to the HRI literature:

- A formal study comparing heads-up, hands-free robot operation to teleoperation in a mobile task.
- An investigation into how the intended users (military personnel) operate a heads-up, hands-free robot and two types of teleoperated robots.
- Results demonstrating that participants performed the task faster, could recall their surroundings better, and reported lower workload during HHO.

II. BACKGROUND

In this section we discuss related work and studies that use person following, gestures, and other modes of teleoperation for human-machine interaction.

Person Following: An extensive body of work has focused on the development of techniques for person following (e.g., [3,4,5]). Young and colleagues showed how an operator can physically instruct a robot to follow a person's walking path by having it learn from an operator that physically pushed it along the prescribed path [6]. Gockley and colleagues found that observers perceived a robot that followed a person's general direction as more natural than when it imitated the person's walking path [7]. Our work advances the general understanding of person following in HRI by directly comparing person following operation in a mobile task to traditional teleoperation. In particular, we seek to understand how such an interaction impacts the operator.

Gesture Recognition and Hands-Free Robot Interfaces: Although we manually simulated gesture recognition (using WoZ methodology), recent research has demonstrated success in automated gesture recognition. There is a great deal of work on gestures primarily designed for human-robot interaction (e.g., [8,9,10,11]), suggesting that gesture recognition technology is close to becoming robust enough for open-world interactions.

A number of promising works have demonstrated the feasibility of gestures in human-robot interfaces. Otero and colleagues collected gestures that came naturally to people in teaching scenarios involving a robot, and found that people accommodate a gesture recognition system's limitations [12]. This study suggests that people are willing to use gestures for robot control, even if they may need some basic training. Weiss

and others found that onlookers willingly helped a robot with directions by communicating with it through gestures [13]. The most similar study to ours is Guo and Sharlin's comparison of Wiimote-based gesture control of a robot to keyboard-based control [14]. They found that people performed a robot navigation task faster and in general preferred using the gesture controls. Our work seeks to extend this investigation by comparing gesture operation to joystick-style control, which is heavily used in UGV operation [15].

Other hands-free approaches to robot control exist. Coyote and TeamTalk are mobile robot systems capable of interpreting verbal navigation commands [16,17]. Some robotic wheelchairs can interpret an operator's gaze and head poses as steering controls [18,19]. There is even a mobile robot controllable by brainwaves [20]. Although these systems demonstrated hands-free robot operation, to the best of the authors' knowledge, no such interface has been evaluated through a user study or comparison study to teleoperation.

Teleoperation Control in HRI: The work in this paper follows methodologies from several comparison studies that observe operator behavior using a range of approaches to robot teleoperation. Comparative studies are important because there are a variety of teleoperation interfaces. Traditional interfaces include keyboards, joystick controllers, and multimodal displays [15]. In navigation tasks, the design and presentation of live video feeds and mapping can also impact a robot's effectiveness [21]. One such comparison study, by Nielsen and Goodrich, explored methods of presenting video and map updates to operators, concluding that operators performed a navigation task better with an integrated view than side-by-side views [22]. Adams and Kaymaz-Keskinpala compared PDA robot interfaces by varying the presence of camera and sensor information [23]. When operators could directly view the robot's environment, the additional camera feed from the robot added to the operator's workload. Our work differs from these by comparing teleoperation to an alternative mode of human-robot communication, HHO.

Similarly, the robot's level of autonomy can impact operator performance; several groups have examined the role of autonomy in teleoperation. Yanco, Drury, and Scholtz developed a set of general design guidelines for HRI after comparing operator performance on teleoperation interfaces with various modes of input (e.g., joystick, keyboard, touch screen) and levels of autonomy (e.g., collision avoidance, autonomous waypoint navigation) [24]. Luck and colleagues found that fully autonomous waypoint navigation yielded the least impact from latency problems and best performance on a navigation task when compared to less autonomous teleoperation interfaces [25]. Finally, Bruemmer et al.'s findings provide further support to this result; they showed that novices familiar with the task domain but unfamiliar with robot control benefited more than expert operators when using autonomous waypoint navigation [26]. Although these studies found an increase in autonomy to be beneficial, the operator was not walking around test areas with the robot. Our interest is in comparing HHO to teleoperation when the operator and robot are on foot and copresent.

III. System

Our heads-up, hands-free system consists of the following components:

- A stock iRobot 510 PackBot robot chassis, which is a tracked ground robot with flippers.
- A computational payload housing an Intel 1.2 GHz Core 2 Duo for on-board perception processing, and a Hokuyo LIDAR for obstacle avoidance.
- A Tyzx G2 stereo head and inertial measurement unit (IMU) mounted to a 3-degree of freedom neck.
- An implementation of a person following behavior, described in more detail below.

The heads-up, hands-free system uses a person following behavior [27] which integrates three software sub-components: (a) an image-based pedestrian detector, (b) a particle filter to track a pedestrian, and (c) a following component to steer the robot chassis and command the neck pan axis. We base the pedestrian detection component on a Histograms of Oriented Gradients (HOG) features approach, which is described in [28]. The HOG-based pedestrian detector is robust to changes in the person's pose and lighting conditions. We filter raw pedestrian detections using a particle filter and estimate range to the pedestrian using the depth data measured by the stereo camera. The particle filter produces a continuous track and eliminates false positives. Given the filtered position of the person, the robot tries to maintain a fixed distance from the person, while avoiding collisions. It does not attempt to follow the path of the person. The camera neck simultaneously pans in an attempt to keep the person in the center of the camera's field of view.

The teleoperated system consists of a robot of the same model as the heads-up, hands-free system, but the operator controls the robot using a ruggedized tablet controller and stylus (see Fig. 3, left). A backpack radio facilitates wireless teleoperation, and the operator drives the robot with an on-screen joystick and a stylus. The tablet's screen also shows a live video feed of the robot's forward-facing camera. The video feed displays a projected path to show the robot's heading as it moved. Projected path has been shown to be an improvement over standard teleoperation [29]. The drive speed of this robot is similar to that of the heads-up, hands-free robot. The interface is designed for and deployed in real-world field operations.

IV. Method

In the study, participants evaluated three different modes of operating a UGV in a room-by-room clearance scenario. The participant's task was to search hallways and rooms along a predefined route for markers ("stimuli") on the wall; the robot's task was to monitor the halls while the participant entered the rooms. The task of identifying and remembering these stimuli is designed to place a cognitive load on the operator (other than driving the robot). In this room-by-room tactic, these stimuli represent objects or unknown persons to investigate. We chose this variant of building clearance since it is a common operation encountered in combat situations.

Each participant engaged in five elements (listed below) of the experiment involving two different robots (a teleoperated robot and a controller-free robot, see Fig. 1). Elements 3, 4, and

Figure 3. Left: This is how participants operated the robot in the teleoperation condition. Teleoperation required a tablet OCU and backpack radio. Right: This is the 'T' gesture participants used to operate the heads-up, hands-free robot.

5 were uniformly rearranged across participants to mitigate any ordering effects.

1. The participant demonstrated several gestures as part of an ongoing data collection process. While standing in front of the controller-free robot, the participant was asked to show gestures they would use to initiate several different commands. (The data from this collection is not discussed in this paper, but will be used in future analysis.)

2. The participants received 5-10 minutes of training with each of the two robots immediately before a new robot was used. Training included specific exercises and free-form practice time.

3. *Teleoperation with video (TV):* The participant operated the robot using a tablet and an on-screen, virtual joystick (this is similar to the interfaces currently in the field, as holding a laptop while viewing the laptop's video feed and holding a controller while moving would be impractical). While searching the rooms, the participant also had to monitor the robot's video feed, which could be viewed on the tablet. Participants watched the video feed for any passing people (represented by a marker held in front of the robot's camera). This interface permits the operator to move the robot and view its camera feed simultaneously while mobile.

4. *Teleoperation with speech (TS):* The participant performed the same building clearing scenario task as in (3). While monitoring the hall, however, the robot reported any person detections via verbal notification. Thus, the participant was not tasked with monitoring the robot's video feed for the stimuli.

5. *Heads-up, hands-free operation (HHO):* The participant performed the same building clearing scenario task as in (4), but with the controller-free robot. The robot again reported any person detections via verbal notification. The participant did not carry the tablet and used an arm gesture when commanding the robot to remain in the hall. Motivated by military handbooks for tactical operations [30], we chose a "T" gesture (see Fig. 3, right). When the participant placed both of his arms out to form a "T", the robot toggled between being in follow mode and stop mode (see section IV.B). The position of the robot's flippers indicated its mode, i.e., when the flippers were

TABLE I. BREAKDOWN OF STIMULI ACROSS THE THREE TRIALS IN THE BUILDING CLEARANCE SCENARIO. CASES TV, TS, AND HHO WERE REORDERED TO PREVENT ANY ORDER BIAS.

Trial #	# Persons	# Objects	# Persons Walking By
1	5	8	3
2	6	7	4
3	8	3	5

out, the robot was stopped; when the flippers were retracted, the robot was in follow mode.

Because there is no operator control unit (OCU) in the HHO case, the user is informed of stimuli detections via verbal notification. Both the method of control (teleoperation vs. hands-free) and notifications (watching the video feed vs. verbal notification) differed between TV and HHO modes, preventing a clear comparison. However, the TS condition allows us to distinguish the effects of each variable. In TS mode, the UGV is teleoperated, but reports stimuli detections via speech, as in HHO. In this way, we can remove the impact of verbal notifications in our analysis.

A. Experiment Design

In this study, we gave participants a minimally labeled, single floor map of the nearby area (see Fig. 4). We tasked participants with exploring the hallway and numbered rooms on the map, both color-coded on the map for them (distance was ~45m, three corners to turn). The studies were conducted in one of two relatively vacant building areas, and participants went through all rooms in numeric order. Participants had fifteen minutes to complete the task. We told participants that their primary objective was to locate "persons and objects of interest" in the hallway and rooms. Persons and objects of interest were represented as transparencies that were taped onto the walls at floor level, eye level, or halfway in-between.

Each of the three trials (cases TV, TS, and HHO) had 16-17 potential stimuli to observe and remember (we did not inform participants of the number of markers to remember). The exact breakdown is specified in Table I. There was a mix of stimuli in hallways and in rooms. In any given trial, 1-2 rooms had no markers. We determined the positions of the markers, counts of each type of marker, along with the locations themselves, using randomly generated number sequences. We also selected the location and number of rooms where the robot reports a person detection to the operator (either by verbal notification or via the video feed) using the same approach.

After each trial, participants labeled the map and answered a survey about their interaction with the robot. The survey asked about their perceptions of the robot and their perceived cognitive load. After each of the first two trials, the experimenter also administered a confederate task, a spatial skills exercise. Following the first trial and post-surveys, participants took a mental rotation test [31]. After the second trial participants took a similar spatial skills test involving cross-section analysis [32] after the second trial and post-surveys. The purpose of the spatial tasks was to distract the participant and keep them from further focusing on the course while the experimenter rearranged markers in the space for the next trial. When the participant completed all three trials, they filled out a post-experiment survey on demographic information.
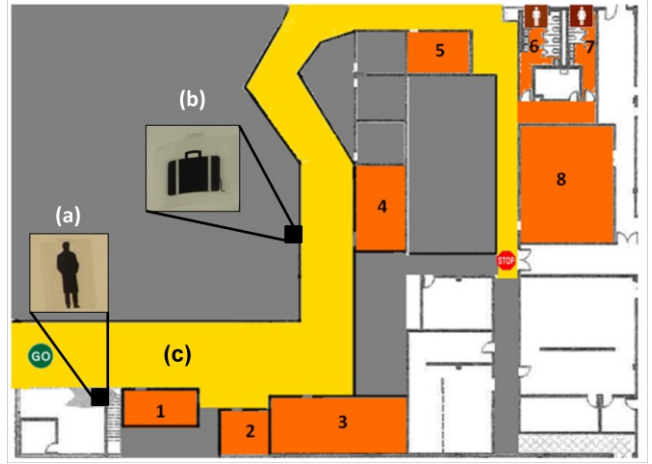

Figure 4. An example map where participants searched for two types of stimuli: (a) persons of interest and (b) objects of interest, throughout the (c) hallway and (numbered) rooms.

The entire procedure took an average of two hours to complete per participant. Two experimenters ran the study: (1) the primary experimenter, who was responsible for running the scenario and was the participant's point of contact; and (2) the data collector, who recorded video of the participant and provided WoZ gesture control for the HHO case. The primary experimenter memorized a scenario script to conduct the study and held to that script for every participant. The experimenters avoided interacting with the participant by following behind both the robot and participant. External participants received $100 for their participation.

B. Wizard-of-Oz Components

This experiment used Wizard of Oz (WoZ) methodology [33] for gesture recognition and stimuli detection. In other words, participants perceived the robot as performing these functions autonomously, when in reality one of the experimenters was manually triggering events when appropriate. Such studies are common in the human-robot interaction design process [34,35]. The goal of WoZ evaluation studies is to focus on the effects of a technology before it has been built, in order to determine whether and where to focus investment to further mature technologies.

When the robot was in follow mode during the HHO case, it was autonomously tracking and following the participant with its camera, maintaining a distance of three meters as the participant moved throughout the scenario. A "T" gesture stopped the robot and commanded it to remain stationary and monitor the hall (it was reactivated with another "T" gesture). The "T" gestures were recognized by the experimenter and manually triggered in WoZ style (the participants were not informed about the manual gesture recognition).

In the HHO and TS cases, when the participant entered a room, the experimenter simulated the detection and notification of stimuli by the robot. The experimenter pressed a button on his controller, which then synthesized a voice that said, *"A person just walked by"*. In the TV condition, the experimenter held one of the "person of interest" markers in front of the robot's camera for two seconds during which time the participant had the opportunity to notice the marker.

C. Hypotheses

Under the conditions described above, our goal was to evaluate four hypotheses on how HHO compares to teleoperation (both TV and TS):

- *Hypothesis 1:* Participants perform the building clearance task faster with HHO than with teleoperation.

- *Hypothesis 2:* Under a heads-up, hands-free operating condition participants are better able to discover and recall the locations of markers in their environments compared with a teleoperation condition.

- *Hypothesis 3:* Participants perceive that under a heads-up, hands-free operating condition their cognitive load is lower than under a teleoperation condition, as measured by the NASA Task Load Index (TLX) survey that measures operator workload [36].

- *Hypothesis 4:* Participants perceive the heads-up, hands-free operating condition more positively than the teleoperation condition.

D. Measures

The key independent variable in the building clearance scenario was the "control type" the participant used in the current trial (i.e., TV, TS, or HHO). The dependent variables involved observational and questionnaire measurements.

Observational measures – We measured participants' time on task in each trial from start to finish. We also measured participants' recall ability and environment awareness using a map labeling task at the end of each trial. Participants labeled where they saw persons and objects of interest, and where the robot saw persons of interest on the map they were given at the start of each trial.

Questionnaire measures – We measured participants' perceptions of cognitive task load after each trial using the NASA Task Load Index (TLX) survey [36]. We also measured participants' perceptions of each robot using the Knowledge, Dominance, and Humanlikeness scales described in [37]. The questionnaire also included a demographic survey where we asked participants about their past robot experience, video game experience, and military experience. All questions related to the perception of the robot and cognitive load were asked on seven-point Likert scales—i.e., scales that indicate their level of agreement with a statement. We also polled participants about their general robot interface preference.

E. Participation

We recruited 30 participants for this experiment. Of the 30 participants, 21 were on active duty in the military (1 US Army, 1 US Navy, 19 US Air Force). The remaining 9 participants were iRobot employees (3 had military experience), but were unfamiliar with the research and unaware of the goals of the study. The gender differential was 27 male to 3 female. Participants ranged in age from 19 to 49, (*mean:* 30.1 years, *standard deviation:* 7.2 years). Five participants claimed to have had significant experience (more than 400 hours) using robots.

V. RESULTS

In this section we present an analysis of the data collected during the experiment. We analyzed the observational

TABLE II. SCALES AND THEIR RELIABILITY.

Scale Name	Questions	Alpha[a]
Participant's Video Game Skills[b]	How long, how often, how good; adapted from [39]	0.84
Participant's Experience With Robots[b]	Hours of experience; experience operating a teleoperated robot; robot training; teleoperation training	0.27[d]
Robot's Competence[c]	competent; knowledgeable; intelligent; expert; reliable; useful; trustworthy; likable	0.98
Robot's Dominance[c]	strong personality; assertive; dominant; power	0.91
Robot's Humanlikeness[c]	natural; humanlike; like a human; lifelike; moves like a human; has a mind	0.92
Nasa Task Load Index (TLX)[c]	perceived mental load; physical load; time pressure; performance; stress; and perceived success	0.79

a. Cronbach's alpha is a measure of a scale's reliability as a whole. Alpha values range from 0.0 to 1.0, with 1.0 being perfect correlation among scale questions.
b. This scale was included as an independent variable effect in the regression model.
c. A dependent variable that we measured in the regression model.
d. This was a low alpha because people had different kinds of experience. However, the scale was predictive of people's performance and perceptions of the robot, so it was important to include it in the regression model.

measures using a mixed-effects analysis of variance model (a standard least squares regression using the reduced maximum likelihood (REML) method [38]). We first tested overall effects with the mixed-effects model. Next, for each dependent variable that had a main effect between the 3 control types, we tested all possible pairs of comparisons using the Tukey HSD test. The key independent variable was the control type used.

- To model a repeated measure for each of the 3 trials per participant, we included participant ID in the model as a random effect and trial order as a fixed effect. As expected, participants improved performance as they repeated the task, due to their learning about the task.

- A few participants were civilian employees, so we included a factor to account for interactions between control type and whether participants were civilian or military. This factor had no differentiating effect on most dependent variables except for time on task, which we report separately (see section V.A).

- For logistical reasons, we were forced to conduct the experiment in two different locations (same number of rooms and distance in each). We therefore included location as a factor in the model. Location had no measurable impact on the results.

- We also included a scale assessing participant robot experience and one assessing their video game skills (adapted from [39]). We analyzed the questionnaire measures related to task load and perceptions of the robot by first assessing how reliably each question set held together as a scale using Cronbach's α (see Table II). An alpha of 0.70 or greater traditionally indicates a reliable scale and we report these values in Table II. We included each scale in the mixed-effects model.

A. Observations

As discussed in section IV.D, we included two main observational measures, (1) the time that participants took to

complete the building clearance task, and (2) their performance on a map labeling task. Time on task assesses the participant's ability to navigate the robot through the course while also performing the primary search task. The map labeling task assesses the participant's awareness of markers in the environment and their ability to recall where they found markers later. We coded the participant maps such that each true marker location was worth 3 points: the participant received 1 point for labeling the correct area, 1 point for the correct corner of a room or hallway, and 1 point for the correct type (person of interest, object of interest, or person that walked by). The sum of all point values for each location represented the participant score, minus 3 points for each false positive. Lastly, the score was normalized by the total possible points. Eighty-seven maps were graded; three maps were incomplete due to logistical issues (two participants were affected). Their results provided additional testing data to evaluate our hypotheses, so we included them in our analysis.

Hypothesis 1 predicted that participants would finish the building clearance task significantly faster when interacting with HHO than teleoperation. The mixed-effect model, when observing only externally-recruited military personnel, supported this hypothesis. Control type (i.e., TV, TS, or HHO) had a significant main effect on time on task for military personnel ($F[2, 32.7] = 15.0$, $p < 0.0001$). HHO was significantly better than teleoperation with video ($p < 0.0001$; Tukey test) and significantly better than teleoperation with speech ($p = 0.01$; Tukey test) at the 95% confidence interval. Time on task was not significantly different between control types for civilian employees; this was an exception to our overall findings (see section VI for more details). Fig. 5 presents these results.

Hypothesis 2 predicted that participants would be better at discovering and recalling the locations of persons and objects of interest with HHO compared to teleoperation. This was measured by performance on the map labeling task. The mixed-effect model for map score performance marginally supported this hypothesis. Control type had a significant main effect on map score ($F[2, 49.8] = 5.5$, $p = 0.007$). HHO was significantly better than teleoperation with speech ($p < 0.006$; Tukey test) and marginally better than teleoperation with video ($p = 0.07$; Tukey test). Fig. 5 provides details on participant scores.

B. Questionnaire Responses

We analyzed the questionnaire measures by first performing a factor analysis for each questionnaire participants answered using Cronbach's α. These surveys assessed the task load and perceptions of the robot in each building clearance trial [36,37]. The NASA-TLX survey is a six question survey in which participants answer questions about perceived mental load, physical load, time pressure, performance, stress, and perceived success regarding the task. There were three surveys that assessed perceptions of the robot in the task: competency (8-question scale), dominance (4-question scale), and humanlikeness (6-question scale).

Hypothesis 3 predicted that participants would perceive heads-up, hands-free control to have lower cognitive load than either teleoperation condition. Results confirmed our prediction. Control type had a significant main effect on TLX-
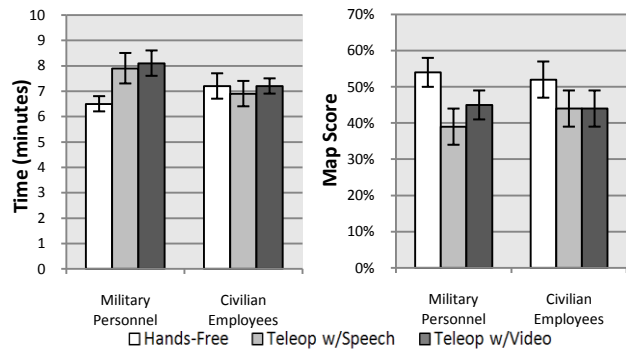

Figure 5. The time duration for the building clearance task (left, in minutes) and participant scores (right).
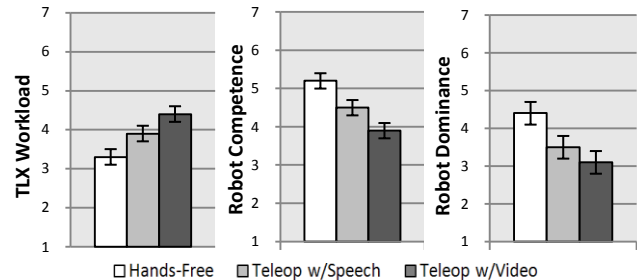

Figure 6. All participant scale responses related to workload (left, NASA-TLX), robot competence (center), and robot dominance (right). 1 = low, 7 = high. Lower is better for TLX, higher is better for competence. Results strongly favored the heads-up, hands-free robot.

scale scores ($F[2, 49.6] = 23.7$, $p < 0.0001$). HHO was significantly better than teleoperation with video ($p < 0.0001$; Tukey test), and significantly better than teleoperation with speech ($p = 0.001$; Tukey test). Fig. 6 shows these results.

Hypothesis 4 predicted that participants would perceive the heads-up, hands-free robot more positively than either teleoperation condition. Results showed a strong overall preference for the heads-up, hands-free robot (22 of the 30 participants preferred it to teleoperation) and more positive questionnaire ratings for the heads-up, hands-free robot. Control type had significant main effects for robot competency ($F[2, 50.0] = 17.9$, $p < 0.0001$), dominance ($F[2, 49.8] = 14.0$, $p < 0.0001$), and humanlikeness ($F[2, 49.6] = 32.3$, $p < 0.0001$). We summarize these results in Fig. 6. Participants perceived the heads-up, hands-free robot as significantly more competent than either teleoperation condition ($p < 0.0004$; Tukey test). Although participants perceived the heads-up, hands-free robot as significantly more dominant, this dominance was closer to neutral than the other two control types ($p < 0.0004$; Tukey test). While participants did not perceive any robot as exceptionally humanlike, the heads-up, hands-free robot had higher scores than the teleoperated robots.

VI. DISCUSSION

In general, the results of the study support the four hypotheses. In two cases the support was marginal. In particular, Hypothesis 1 was strongly supported by military personnel, but not by civilians; and Hypothesis 2 was marginally supported by the data. We discuss each of the hypotheses in order below.

In regards to Hypothesis 1, participants from the military took less time to complete the highly mobile building clearance task with the heads-up, hands-free robot than the teleoperated robots. The civilian employees in the study, though, did not perform the task any faster with the hands-free robot. Further analysis of the robot experience questionnaire revealed that 5 of the 9 civilian employees considered themselves "experts" or "near-experts" in robot operation, while only 6 of the 21 military personnel considered themselves experienced. Results from civilian employees did not impact any of the other measurements in the study; their results were consistent with the military personnel. As such, we included their results in all other measurement analyses. This result suggests, but does not confirm, that less experience is required to achieve the same competence when using HHO. On the other hand, the question remains for future work whether "expert-level" experience with HHO—which none of the participants had in this study—would reduce time on task.

Our results show that participants discovered more about their surroundings and recalled more of the markers in the environment when operating the heads-up, hands-free robot, supporting Hypothesis 2. This result was marginally significant, however, when comparing HHO to teleoperation with video. We speculate that this is because the faster pace of HHO may have reduced the thoroughness of the search task for a select number of participants. In the context of the robot's intended use, HHO serves as a less taxing alternative to teleoperation, since an operator need not devote full attention to robot operation. In military applications, this has the added benefit of freeing another soldier from devoting full focus to defending the operator.

Results from our analysis of NASA-TLX responses support Hypothesis 3 and demonstrate that participants had reduced physical and mental loads during HHO compared with teleoperation (which also required operating a tablet and wearing a backpack radio). This suggests that heads-up, hands-free robots are more desirable than teleoperation in situations that are mentally and physically demanding, such as when the operator must focus on the surroundings.

Finally, the data also supports Hypothesis 4. Nearly 75% of participants preferred heads-up, hands-free robot operation over teleoperation. Participants also found the heads-up, hands-free robot to be more competent than the teleoperated robots, and on average this measure was positive (above neutral). This suggests that participants perceived the heads-up, hands-free robot as more intelligent during the building clearance task. Participants also viewed the heads-up, hands-free robot as neutrally dominant, while viewing the teleoperated robots as negatively dominant.

A. Limitations

In this study we did not recruit participants from the general population; we recruited from a population that is more likely to use these robots "in the field"—military personnel. Most of these personnel had experience or training in building clearance scenarios. This group represents users who will most likely be directly impacted by these technologies in the future.

Although participants performed significantly better on the map labeling task when they interacted with the heads-up, hands-free robot compared to the teleoperated robots, we cannot independently claim that they had both increased situational awareness and improved recall ability from these results alone. Errors on the map labeling task could occur due to a missed marker in the environment or a forgotten one. However, we argue that participants were less distracted during HHO, determined by some combination of increased situational awareness and increased recall ability.

A significant majority of the participants in the study preferred HHO, but those that did not had one frequent reason for preferring teleoperation: access to live video. The current setup of HHO does not permit access to a visual camera, though the robot conveyed information verbally and through the position of its flippers. One direction for future work would be to observe the impacts on performance that camera access would have to HHO.

We acknowledge limitations with respect to the scope of the contributions presented in this work. First, this study does not result in general conclusions for gesture interaction with robots, though gesture operation did play a prominent role in HHO. The scenario design did not demand for more than one type of gesture to operate the robot, which simplified the task to focus on robot operation. We believe that current gesture technology can perform the full-body motion "T" gesture reasonably well. However, since gesture recognition was not the focus of this work, we opted to use Wizard of Oz methodology for this component. Second, comparing HHO and teleoperation modes combines several dimensions of interaction design. The differences include whether or not the operator is holding a tablet, using an input device, or can at any time look at a screen. Future work should independently compare each of these differences for their respective influence on robot operation. Finally, we do not address the issue of experience, either in teleoperating or heads-up, hands-free operating modes. The differences between civilian employees and external military personnel in time on task indicate that experience with robots does affect this measure. Future work should explore whether experience with HHO can further reduce time on task.

B. Design and Research Implications

Although the current study did not test heads-up, hands-free robots "in the field", this work has design and research implications for human-robot interaction, especially in military settings. Our measures provide strong support for HHO to be a compelling alternative to teleoperation in highly mobile scenarios. In the context of military applications, we believe that these robots might eventually be effective in scenarios such as building clearance and patrol, where the robot's role would most likely be to provide backup support to a team of personnel.

Analysis of post-experimental interviews with military participants revealed that their physical loads in the field are already high, given they carry heavy military equipment in standard battle uniform. HHO is a more desirable mode of operation, since it minimizes the equipment that they need to carry to control the robot.

VII. CONCLUSIONS

Traditional robot teleoperation commonly requires an operator's full attention to achieve task goals. In this paper, we

presented a heads-up and hands-free approach to controlling ground robots using person following and gesture control. We designed a study to compare this approach to robot teleoperation in a highly mobile task (room-by-room building clearance). In an experiment with 30 participants, including 21 military personnel, we showed that they had better recall of surroundings and a lower cognitive load with heads-up, hands free operation than when they teleoperated a robot at close distances. Military personnel also performed missions faster with HHO. We conclude that HHO holds potential as an effective alternative to teleoperation.

REFERENCES

[1] Wise County, Texas. Building Clearing / Tactical Raid. [Online]. http://www.co.wise.tx.us/constable/Downloads/Building%20Clearing,Tactical%20Raid.pdf

[2] A. Steinfeld et al., "Common metrics for human-robot interaction," in *Proc. of HRI*, 2006.

[3] M. Liem, A. Visser, and F. Groen, "A hybrid algorithm for tracking and following people using a robotic dog," in *Proc. of HRI*, 2008.

[4] E. A. Topp and H. I. Christensen, "Tracking for following and passing persons," in *Proc. of IROS*, 2005.

[5] T. Yoshimi, M. Nishiyama, T. Sonoura, H. Nakamoto, and S. Tokura, "Development of a person following robot with vision based target detection," in *Proc. of IROS*, 2006.

[6] J. E. Young, K. Ishii, T. Igarashi, and E. Sharlin, "Style-by-demonstration: Using broomsticks and tangibles to show robots how to follow people," in *Proc. of CHI*, 2009.

[7] R. Gockley, J. Forlizzi, and R. Simmons, "Natural person-following behavior for social robots," in *Proc. of HRI*, 2007.

[8] A. G. Brooks and C. Breazeal, "Working with robots and objects: Revisiting deictic reference for achieving spatial common ground," in *Proc. of HRI*, 2006.

[9] N. Miller, O. C. Jenkins, M. Kallmann, and M. Mataric, "Motion capture from inertial sensing for untethered humanoid teleoperation," in *Proc. of Int'l Conference on Humanoid Robotics*, 2004.

[10] R. Steifelhagen et al., "Natural human-robot interaction using speech, gaze and gestures," in *Proc. of IROS*, 2004.

[11] S. Waldherr, S. Thrun, and R. Romero, "A gesture-based interface for human-robot interaction," *Autonomous Robots*, vol. 9, no. 2, pp. 151-173, 2000.

[12] N. Otero et al., "Distribution and recognition of gestures in human-robot interaction," in *Proc. of RO-MAN*, 2006.

[13] A. Weiss et al., "Robots asking for directions: The willingness of passers-by to support robots," in *Proc. of HRI*, 2010.

[14] C. Guo and E. Sharlin, "Exploring the use of tangible user interfaces for human-robot interaction: A comparative study," in *Proc. of CHI*, 2008.

[15] T. Fong and C. Thorpe, "Vehicle teleoperation interfaces," *Autonomous Robots*, vol. 11, no. 1, 2001.

[16] M. Skubic et al., "Spatial language for human-robot dialogs," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 34, no. 2, 2004.

[17] T. Harris and A. Rudnicky, "TeamTalk: A platform for multi-human-robot dialog research in coherent real and virtual spaces," in *Proc. of AAAI*, 2007.

[18] Y. Matsumoto, T. Ino, and T. Ogsawara, "Development of intelligent wheelchair system with face and gaze based interface," in *Proc. of RO-MAN*, 2001.

[19] P. Jia, H. Hu, T. Lu, and K. Yuan, "Head gesture recognition for hands-free control of an intelligent wheelchair," *Industrial Robot: An International Journal*, vol. 34, no. 1, 2007.

[20] W. Amai, J. Fahrenholtz, and C. Leger, "Hands-free operation of a small mobile robot," *Autonomous Robots*, vol. 11, no. 1, 2001.

[21] T. Fong, C. Thorpe, and C. Baur, "Advanced interfaces for vehicle teleoperation: Collaborative control, sensor fusion displays, and remote driving tools," *Autonomous Robots*, vol. 11, no. 1, 2001.

[22] C. Nielsen and M. Goodrich, "Comparing the usefulness of video and map information in navigation tasks," in *Proc. of HRI*, 2006.

[23] J. Adams and H. Kaymaz-Keskinpala, "Analysis of perceived workload when using a PDA for mobile robot teleoperation," in *Proc. of ICRA*, 2004.

[24] H. Yanco, J. Drury, and J. Scholtz, "Beyond usability evaluation: Analysis of human-robot interaction at a major robotics competition," *Human-Computer Interaction*, no. 19, pp. 117-149, 2004.

[25] J. Luck, P. McDermott, L. Allender, and D. Russell, "An investigation of real world control of robotic assets under communication latency," in *Proc. of HRI*, 2006.

[26] D. Bruemmer, C. Nielsen, and D. Gertman, "How training and experience affect the benefits of autonomy in a dirty-bomb experiment," in *Proc. of HRI*, 2008.

[27] J. Brookshire, "Person following using histograms of oriented gradients," *Int'l Journal of Social Robotics*, vol. 2, no. 2, 2010.

[28] N. Dalal and B. Triggs, "Histogams of oriented gradients for human detection," in *Proc. of CVPR*, 2005.

[29] R. A. Pomranky, K. Cosenzo, A. Bodenhamer, and B. Pettijohn, "Experimental evaluation of computer-aided tele-operation (CATO) and computer-aided robotic manipulation (CARMAN) technology," Army Research Laboratory, Technical Report 2010.

[30] *Visual Signals, FM 21-60.*: Department of the Army, 1987.

[31] S. Vandenberg and A. Kuse, "Mental rotations, a group test of three-dimensional spatial visualization," *Perceptual and Motor Skills*, 1978.

[32] C. Cohen and M. Hegarty, "Sources of difficulty in imagining cross sections of 3D objects," in *Proc. of CogSci*, 2007.

[33] N. Dahlback, A. Jonsson, and L. Ahrenberg, "Wizard of Oz studies -- why and how," *Knowledge-Based Systems*, vol. 6, no. 4, pp. 258-266, 1993.

[34] D. Perzanowski et al., "Finding the FOO: A pilot study for a multimodal interface," in *Proc. of IEEE SMC*, 2003.

[35] H. Huettenrauch, K. S. Eklundh, A. Green, and E. A. Topp, "Investigating spatial relationships in human-robot interaction," in *Proc. of IROS*, 2006.

[36] S. Hart and L. Staveland, "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," *Human Mental Workload*, 1988.

[37] A. Powers and S. Kiesler, "The advisor robot: Tracing people's mental model from a robot's physical attributes," in *Proc. of HRI*, 2006.

[38] D. Harville, "Maximum likelihood approaches to variance component estimation and to related problems," *Journal of the American Statistical Association*, vol. 72, no. 358, pp. 320-338, 1977.

[39] M. S. Terlecki and N. S. Newcombe, "How important is the digital divide? The relation of computer and videogame usage to gender differences in mental rotation ability," *Sex Roles*, vol. 53, pp. 433-441, 2005.