PALADYN Journal of Behavioral Robotics

VERSITA

Research Article · DOI: 10.2478/s13230-012-0100-6 · JBR · 3(2) · 2012 · 102-111

Improvements in robot navigation through operator speech preferences

Daniel N. Cassenti[1]*, Troy D. Kelley[1], Rosemarie E. Yagoda[2], Eric Avery[1]

1 U.S. Army Research Laboratory, RDRL-HRS-E APG, MD, USA 21005

2 U.S. Navy, NAVSEA, Carderock, Virginia Beach, VA, USA 23459

Received 012/7/31

Accepted 2012/12/10

Abstract

The capability to use voice commands to control robots offer an intriguing possibility to increase the efficiency with which robotic operators may give commands. This study consists of two experiments that investigate how robot operators prefer to speak to robots in a search-and-find task and to evaluate which mode of speaking generates the greatest performance. Experiment 1 revealed that operators used selective exocentric references when available to direct a confederate acting as a robot. Experiment 2 revealed that with the same exocentric references operators showed improved performance while directing an actual robot as compared with egocentric-only commands. Experiment 2 also revealed that performing a dual task was less detrimental to performance when using exocentric commands as compared to egocentric commands. Suggestions for improvements to a robot control system that follow from these results include developing recognition of structural properties of an indoor environment and improving map incorporation.

Keywords

Robot · mode of control · SS-RICS · performance · navigation

1. Introduction

With the advent of speech recognition tools, robotics developers have a new tool at their disposal to aid robot operators by incorporating the ability to use voice command in their robotic platforms (e.g., [1]). Using this capability to maximum potential requires understanding how people prefer to speak to robots when directing them to achieve certain goals. If certain types of speech improve task performance [2] then robotic platform developers should develop their platforms to conform to effective types of verbal command speech (see [3]).

Since instructions to robots are often directed towards navigation, our primary focus is navigation. Navigational language can be divided into two broad categories, egocentric and exocentric [4]. Egocentric language uses the subject as the referent. In the case of a task involving a robot, the robot is the egocentric reference. Navigational commands given egocentrically are focused on the action required of the robot with reference to its current position and orientation. In terms of robot operator commands, the following would qualify as egocentric: "turn left," "turn right," "go forward," "go backward," or "stop" with some additional commands or qualifiers depending on the situation.

Exocentric language uses the environment as a referent. Exocentric navigational commands are focused on objects in the environment. In terms of a robot, the operator could instruct the robot to "go to" any number of objects that the robot understood and could locate within the environment.

The navigational language preferences of robot operators have been previously studied [5–7]. In [5], participants were asked to type natural language commands to a robot to make it move next to an object. Only half of the participants referenced the object, even though the robot would have understood this command. Instead, participants chose to maneuver the robot using egocentric commands, and the amount of typing necessary was far greater than using a single exocentric command. The authors were surprised by the tendency to use egocentric commands and speculated that the participants determined exocentric commands would be too complicated for the robot to process. As a follow up to [5], [6] reaffirmed this speculation and in [7] a model was developed to represent egocentric and exocentric robotic commands.

The robotics control system of focus in the present study is the Symbolic and Sub-symbolic Robotic Intelligence Control System (SS-RICS; [8]). SS-RICS is based on production system modeling in which discrete step activity advances an agent from an initial state to a goal state. SS-RICS can be linked to a robotic platform in either a real environment or simulated environment. This system is particularly focused on navigational capabilities in an indoor environment. An operator can control a robot linked to SS-RICS by verbal commands or by using manual controls.

The results of [5] and [6] are particularly important for the development of SS-RICS. Performance of a task involving a robot linked to SS-RICS would be best if and when commands are conducive to an operator's natural language tendencies. If commands are written without consideration for natural tendencies, then this would force the operator to translate commands into an unnatural form before speaking them. The extra processing increases the time to produce the commands and the chance that an error (e.g., using a command that is not stored in SS-RICS' database) will occur.

Verbal commands are processed by speech recognition software and elicit an action for SS-RICS to search its database for a goal to match the command. SS-RICS will then take the selected goal and execute the programmed actions. Goals may also have other sub-goals contained within them. If so, each sub-goal is executed sequentially (e.g.,

*E-mail: daniel.n.cassenti.civ@mail.mil

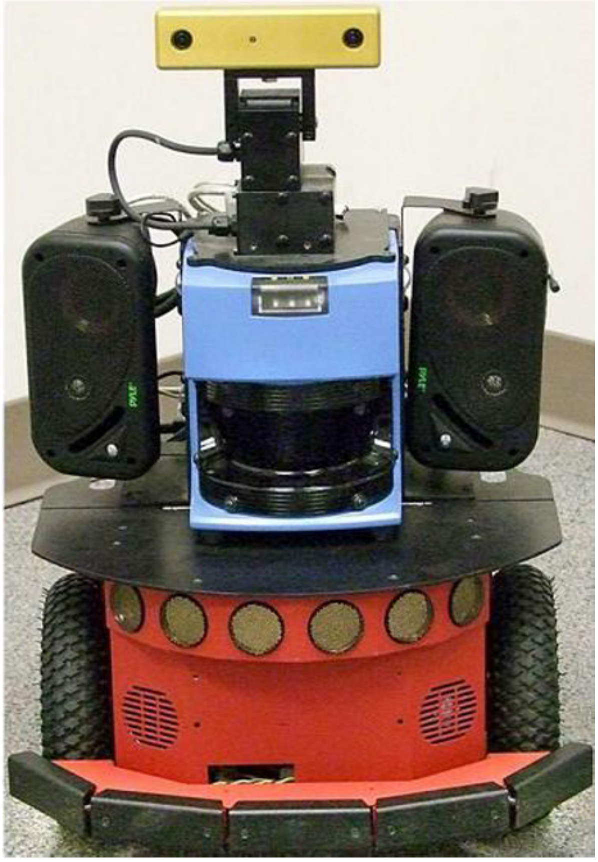


Figure 1. The robotic platform with the LIDAR system (blue and black section) and mounted camera (top).

a command to search a room may include goals to go to a series of waypoints).

Another important consideration is the development of visual object recognition capabilities. The results of [5] and [6] revealed that only half of participants were inclined to use exocentric commands. If exocentric commands do not improve performance, then object recognition would not be an important skill to include in SS-RICS. If performance is enhanced by using exocentric language, then visual recognition is an important skill to develop in SS-RICS.

The robotic platform (Figure 1) currently linked to SS-RICS is from the Pioneer series (Adept Technology, Inc.) and uses a Light Detection and Ranging (LIDAR) laser system. The LIDAR is a 180° laser, which projects and bounces off objects in the environment. The laser beams return distance information that supplies SS-RICS with a schematic map of half of its visible environment. SS-RICS may then take this information and generate a map of all areas that were visible to the LIDAR system using Simultaneous Localization and Mapping (SLAM) [9]. SS-RICS may also open a pre-built map to navigate and self-locate in an environment.

The following two experiments were designed in an iterative fashion where the second experiment was a response to build on knowledge from the first experiment, an experimental design similar to [10]. The experimental methods were intended to achieve two objectives. First, we aimed to understand how well robot operators performed in accomplishing a search-and-find task when using egocentric-only versus egocentric and exocentric commands. This represents a departure from the paradigm of [5], [6], and [7]. In the following experiments, the operators did not have exocentric command options available in some conditions; instead they were restricted to manual control, egocentric-only verbal control, or egocentric and exocentric control conditions. We hypothesized that exocentric commands would improve performance as operators were given the freedom to command the robot to go to an object rather than using a string of commands to precisely orient and position the robot.

The second objective was to analyze the most frequently used objects in the exocentric condition. The difficulty of programming object recognition skills from camera input [11] requires a conservative approach to picking objects to recognize. If several objects are found to be used frequently (e.g., doors), then SS-RICS developers would be advised to begin with the most frequently cited objects. Once SS-RICS consistently recognized the most frequently cited object, then the next most frequently cited object may be programmed. In this way the development of SS-RICS may be guided by empirical results rather than educated guesses.

Participants in Experiments 1 and 2 were tasked with controlling a robot to find a target object in a simulated environment (Experiment 1) or a real environment (Experiment 2). In each experiment, participants were instructed to use one of three conditions for controlling the robot, including manual control, egocentric-only commands, or egocentric and exocentric command options.

The paper is organized into four major sections including the introduction, Experiment 1, Experiment 2, and a general discussion. The experiment sections discuss the respective experiments at every level. These include an introduction to the experiment, the method, results, and discussion of those results. The general discussion will summarize the results of each experiment, discuss the relationship between the two sets of experimental results, discuss how to use the results of the studies in SS-RICS development, and finally draw conclusions.

2. Experiment 1

Experiment 1 was designed to provide empirical results to drive the development of SS-RICS. To allow operators freedom to use a variety of commands, the participants heard instructions on how they would control a robot and gave these commands to a confederate who adopted the role of the robot. However, subjects were not told a person was controlling the viewpoint and were guided to assume this was done autonomously by a robot. This "Wizard of Oz" paradigm [12] allowed the use of a variety of commands that an SS-RICS controlled robot would not be capable of understanding. The experimental paradigm is similar to [13] in which participants directed an actual robot that was controlled by an operator through a real environment.

In particular this arrangement allowed the participant freedom to use any of the objects in the scenes. This exocentric command freedom was also enabled by using object-rich environments through a first-person shooting game [14]. The participants were told that the robot operated within this environment could be controlled via manual keyboard and mouse controls or through verbal commands. The confederate was located in a separate room and heard the participant's commands through a speaker.

In addition to the control variable of manual, spoken egocentric, and spoken exocentric commands, participants received a map of the environment on half of the trials. The maps included the layout of the indoor structure and the objects contained within it. A representative

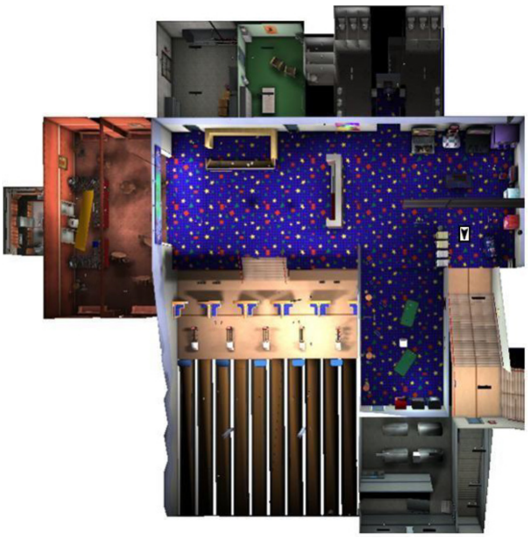


Figure 2. *An example map that details the accessible parts of one of the environments from Experiment 1 (i.e., a bowling alley). The icon with a black arrow on a white background in the right, middle portion of the map indicates the starting position and orientation of the robot. The robot was not allowed to climb stairs but could use ramps like the one illustrated in the bottom right corner of the figure.*

map is shown in Figure 2. The map was meant to simulate real world conditions in which the configuration of a to-be-explored room is already known.

The results of each trial included whether participants were able to find the target object within three minutes, the search time used if the target was found, the proportion correct of a Mental Rotation Test [15], the number of times the participant looked at a map, and voice recordings when in the exocentric command condition. The performance results were used to discover which condition or conditions offered the best results. A questionnaire was administered to see if there was a relationship between spatial ability, as measured by the test, and the ability to perform the task. The voice recordings were used to discover object label preferences from the exocentric condition.

2.1. Method

2.1.1. Participants

Nineteen volunteers participated in this experiment. All volunteers were civilian employees of U.S. Army Research Laboratory and received no compensation in exchange for their time. All participants were tested for visual acuity in both eyes and all had at least 20/40 vision. Three participants discontinued the study prior to completing an experimental trial after experiencing motion sickness during the practice trials. Of the 16 participants that completed the experiment, eight were male and eight were female.

2.1.2. Materials and Software

After signing an informed consent, participants received a Mental Rotation Test [15]. The test comprised twenty questions. Each question consisted of a blocked three-dimensional figure and was followed by four options listed A, B, C, and D. The participant was asked to put a check in a checkbox next to each option that represented a rotation of the sample figure in three-dimensional space. There were two correct options out of four presented for each of the items.

Figure 3. *The target object that participants were asked to find in all practice and experimental trials. The target was described as a "bomb."*

Instructional materials were presented next. One document had a picture of the target object (Figure 3), to familiarize the participant with how to the target looked in all experimental trials. The other document was a map of the bank setting where all three practice trials took place.

In twelve of the twenty-four experimental trials, participants also used a map. The trials that included a map were determined randomly per participant. Icons were added to each map to illustrate the starting position and orientation of the robot.

In the experimental trials, participants viewed a computer monitor, placed at eye-level on a desk in front of the participants. The software was Raven Shield: Rainbow Six [14] with freely available environments developed by Raven Shield users. Only indoor environments were selected and all adversaries were removed from the environment.

In a manual condition, participants used the arrow keys, on a standard keyboard, to move the first-person character forward or backward and a mouse to orient the field of view and direction of movement. The monitor showed the movement of the character in real time. In two verbal conditions the participant was asked to speak into the speaker phone that transmitted the commands to the confederate.

2.1.3. Procedure

Participants were told they had three minutes to navigate through an indoor environment and to find the target that was shown during instruction for each trial. Participants were informed the experimenter would remain in the room throughout the trials. If time expired prior to the participant declaring the target had been found, the experimenter would inform the participant to stop the trial. Participants were shown the map of the practice trials and were then informed that a map would be provided on half of the experimental trials.

Participants were then informed of the three robot control conditions. The first to be explained was the manual condition, and the experimenter explained how to use the arrow keys to move the character, the enter key to open doors, and the mouse to orient the character. Participants were then permitted to use these controls in the practice environment and to continue to use them until they felt comfortable with their ability to use the controls. There was a target in the practice, but the participants were not required to find it, however they were required to view the target in order to become familiarized with how the target appears in an environment.

After the manual practice, participants were asked to practice with "directions-only" commands (i.e., the egocentric condition). The experimenter explained that these commands were limited to "turn left," "turn right," "move forward," "move backward," "stop," "slow down," and "open door." The latter command was necessary though it did refer to a door because the environment editor in the gaming software did not have an obvious means of keeping doors open in the environment. Participants were restricted to saying "open door" when the character was situated next to a door. Also, the commands to turn left or right could be modified by a number of degrees (e.g., "turn right ten degrees"). If the participant did not modify a command to turn, the confederate turned the character 90° in the specified direction. Participants were told that the robot would not move if the participant used any other command or if their command was not spoken clearly enough. To avoid the latter, participants were encouraged to enunciate their words and to speak louder than normal. Like the manual practice, the egocentric practice was completed when the participant felt comfortable with the commands.

After the egocentric practice, the experimenter explained the exocentric condition as using all the commands from the other verbal condition as well as references to any visible objects. Participants were instructed that the robot would understand virtually any object label. All other instructions were maintained in the exocentric condition. Participants performed the practice trials until they felt comfortable with the new commands.

There were 24 experimental trials all with different environments. On half of trials, participants were given a map of the environment prior to the start of the trial. On the other half of trials, participants did not receive a map. The participants also had a three-level control variable with the aforementioned command options. Control and map variable combinations included four trials each. Variable levels were randomly assigned to the environments, and trial order was randomized.

Trials began when the experimenter loaded the environment in the gaming software. During this time the participant received a map to study for as long as they chose before starting the timer. The experimenter kept a tally during the experiment of the number of times the participant looked at the map including once if studied prior to the start of the timer. Once the time began, the participant was free to explore the indoor environment within the limits of the control condition. If the target was found before three minutes expired, they were asked to indicate that they found it. If the participant did not find the target before three minutes, the experimenter stopped the trial.

2.2. Results

The results were first analyzed by environment to determine if there were any particularly difficult environments. With one environment only two out of 16 participants found the target (.125), so the results of this environment were removed from all further analyses. All 15 of the remaining environments were analyzed.

2.2.1. Performance Results

Performance data includes Mental Rotation Test [15] score (i.e., the number of correctly identified options selected out of 40 correct options), proportion of targets found, number of looks at the map, and response time in the search-and-find task when the target was found. First, two correlation tests were run between Mental Rotation test score and search-and-find accuracy and response time. Of these two correlations, there was a significant positive two-tailed correlation of r (335) = .225 between mental rotation test score and proportion of targets found, $p < .001$.

Two within-subjects (map, no map) by three (manual, egocentric, exocentric control) repeated-measures analyses of variance (ANOVAs) were run to detect possible differences in proportion of targets found (.658) and in response time for successful trials only (9.7 s). First, having a map resulted in greater accuracy (.734) than not having a map

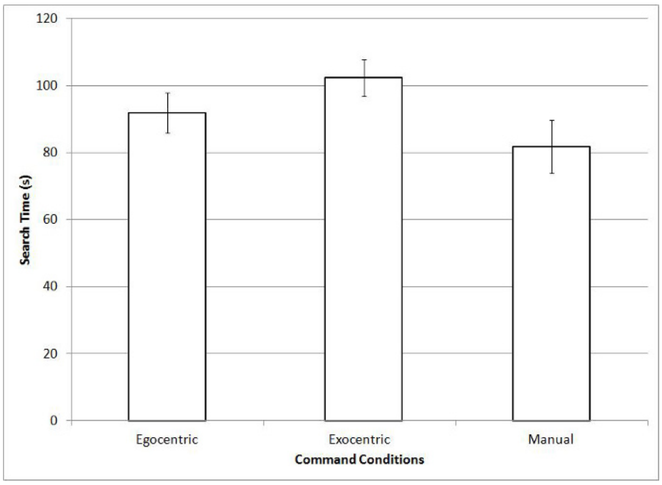


Figure 4. Search accuracy in Experiment 1.

(.620), $F(1, 15) = 8.81$, $p = .01$. Second, a potential difference of command conditions, $F(2, 24) = 2.96$, $p = .071$ was detected. T-tests were used to make two comparisons with a Bonferroni correction (i.e., traditional critical p-value of .05 becomes .025 after dividing by the number of comparisons). The tests revealed that manual control (81.7 s) was faster than the exocentric verbal control (106.7 s), $t(15) = 2.62$, $p = .019$. The manual and exocentric conditions were not different from the egocentric condition (91.8). Figure 4 presents search time per command condition.

Finally, two correlations were tested involving the number of times participants looked at the provided maps versus proportion of targets found and response time. A significant negative one-tailed correlation, r (188) = -.139 was found between number of map looks and search-and-find accuracy, p = .028.

2.2.2. Exocentric Command Results

The recorded utterances from the exocentric verbal control condition were analyzed to understand operator object label preferences. It is important to note that two participants did not speak loudly enough for their voice recordings to be analyzed (one male and one female participant) and were not included in the subsequent analyses.

As the first test, it was important to discover whether participants were or were not likely to use object labels in commands when give the chance to use them. A test was designed to discover whether object labels were used or if participants merely used egocentric commands. Utterances (i.e., a single command) using object names was divided by the total number of utterances to calculate a proportion of object names used (.144). A single-sample t-test showed that the proportion was greater than zero, $t(13) = 7.21$, $p < .001$, indicating that participants tended to use exocentric labels rather than merely egocentric commands.

The utterances with object labels were further categorized by label type. This analysis showed that "door" or "doorway" (hereafter referred to as "door") was the most frequently used object label. The proportion of door labels to all other object labels was calculated (.844). A single-sample t-test showed that of all object referents door was more likely to be used than not used, $t(13) = 18.79$, $p < .001$.

In addition to door, "hall" or "hallway" (hereafter referred to as "hall") and "room" were frequently used. Given how frequently the door label was used, the analyses of both hall and room was reduced to the proportional use of these words versus all object labels not including door or the other object label of interest (e.g., when analyzing "hall," neither "door" nor "room" were included in the total number of utterances). This reduced the remaining pool of participants by one for the following tests because one participant only used door in exocentric commands. One paired-samples t-test showed that of the remaining objects, "hall" (.248) was more likely to be used than not used, $t(12) = 2.71$, $p = .019$ and the other paired-samples t-test showed that of the remaining objects room (.229) "room" was more likely to be used than not, $t(12) = 2.68$, $p = .02$.

2.3. Discussion

The most interesting result is that response time degrades when participants are allowed to use exocentric commands; however, performance is not degraded when limited to egocentric commands. This result defies intuition since it seems reasonable that adding capabilities, especially when the participant could limit the number of commands used, would only improve performance.

The results from the verbal protocols in the exocentric condition may help to explain the finding. The results indicated that participants overwhelmingly used the terms "door," "hall," and "room" in their exocentric commands. This limited subset of words may suggest that there were too many object label options in the environment, and participants responded by limiting their options to a more constrained set. Sorting through more options for commands would necessarily take longer than sorting through few objects (e.g., [16]). The participants did not need to consider object labels at all in the egocentric condition. Though, they may have to speak more commands in the egocentric condition, it was faster to speak more than to think more.

The finding from [5] that shows that many participants were inclined to think that robots could not process exocentric commands may also be a factor in this finding. Participants generally reported the assumption that they were speaking to a confederate and not a robot, but many also said that they were trying to help out the experimenter by speaking to the confederate as if they were speaking to a robot. Though this is just an anecdotal finding, it illustrates that perhaps participants had to go through the extra cognitive step of reflecting on whether a robot would understand an object label or possibly to think of alternative labels that a robot would understand. This may suggest that the extra trial time required to find the target in the exocentric condition was consumed by unnecessary cognitive steps.

The map variable also has some interesting findings. Participants were more accurate in target finding when they had a map, but when the map was referenced during the trial, performance actually decreased. These results suggest that map reading prior to the start of the trial enhanced performance. Additionally, map reading during the trial was likely an indication of disorientation which would lead to decreased performance. Therefore, in situations where there are maps available for a place to be searched by a robot, robot operators should be encouraged to study the map prior to navigating the robot. This is not to say that the map should be withdrawn during search, only that a priori map reading is important.

Experiment 1 helps guide the development of SS-RICS. Although it is unclear whether the exocentric condition actually helps or hinders performance from this experiment, the verbal protocols indicate that object recognition capabilities may be limited to recognizing structural properties of the indoor environment. As of now, the robot already recognizes halls (or hallways) and doorways. Rooms and doors may only be recognized if a preloaded map has them explicitly marked by the mapmaker. The recommendation derived from Experiment 1 is that SS-RICS develop the capability of recognizing doors and rooms through either LIDAR or the camera.

Given the finding of greater response time in the exocentric control condition, a second experiment was necessary to see if exocentric commands enjoy any benefit relative to egocentric only. The two possible reasons for decreased performance of the exocentric condition may be moderated in a second experiment. First, participants may have had too many options for commands in the exocentric condition. We mitigated this problem in Experiment 2 by limiting the number of exocentric commands that the participant could use. The limited set of object labels included the most frequently used object labels from Experiment 1(i.e., room, hall, and door). Additional iterations were added on room including "go to room" and "search room."

The second possible reason for reduced exocentric performance was that participants thought that a robot would not be able to understand a wide range of objects and this required a greater number of cognitive steps before considering which object label to use. Though participants were instructed that the robot could understand virtually any object label, the participants may have engaged in confirmation bias to help the experimenter by speaking in a way that they would speak to an actual robot. Anecdotal evidence indicated that few of the participants were convinced that the robot was real and may have over-compensated to align with the story of the experiment. The potential reason was mitigated by using an actual robot in the study. Participants worked with a real robot over the SS-RICS interface. By seeing the robot and requiring use of object label commands in a practice trial, participants were less likely to doubt the robot's capabilities.

3. Experiment 2

Experiment 2 was designed to overcome the limitations of Experiment 1 that may have caused the exocentric condition to result in high response time. Experiment 2 was run using SS-RICS and the Pioneer robot platform. By using an actual robot that has a prescribed set of exocentric commands, participants can at once overcome the sense that the robot would not be able to understand object names and reduce the set of objects that the participants would need to consider.

We were also interested in the role of mental workload (see [17]) in operating the robot. In Section 2.3 above, an increase in cognitive processing in the exocentric condition may have contributed to worse exocentric performance. The increased information processing would have also increased mental workload, which is the amount of mental taxation required to perform a task or set of tasks. Increased workload tends to decrease performance [18], so increased workload in the exocentric control condition may have been responsible for decreased performance in Experiment 1. Though we did not consider mental workload when designing Experiment 1, Experiment 2 included mental workload as a variable.

The environment used was lab space that was more or less familiar to different participants. To equalize knowledge, the map variable from Experiment 1 was removed and all participants received a map with the robot's initial orientation and position displayed.

3.1. Method

3.1.1. Participants

Seventeen participants (12 male and 5 female) from the civilian employee population of U.S. Army Research Laboratory volunteered for

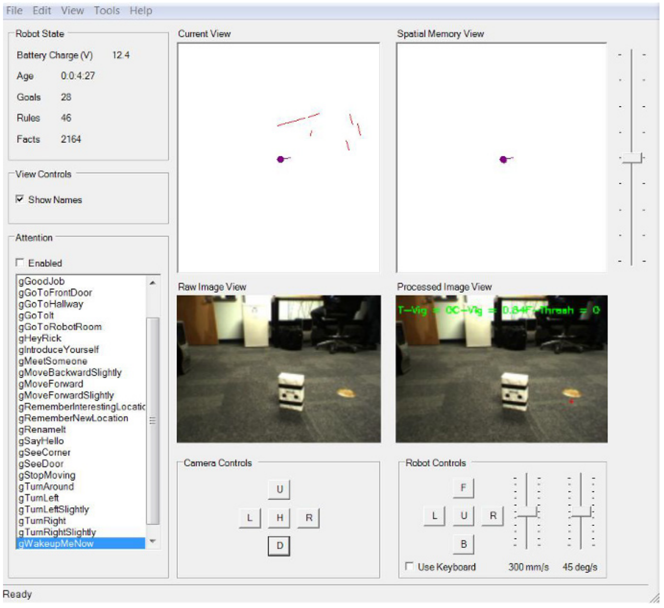


Figure 5. *An example interface of SS-RICS. Some of the features of SS-RICS include video capture in the center, a list of available goals in the bottom, left, camera directional controls in the middle, bottom, and robot directional controls in the right, bottom of the figure.*

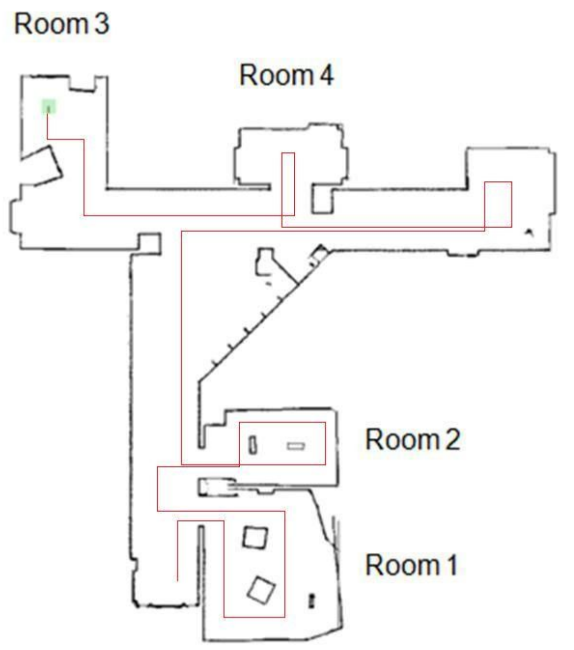


Figure 6. *An example map presented to the participants. The robot is represented by the box in the upper left corner with the line in the center marking the orientation of the robot. Room number labels were handwritten in the approximate locations represented here. The maroon lines indicate a path that the robot could have taken to account for all 177 of the possible distance.*

this study without compensation. None of the participants from Experiment 2 participated in Experiment 1.

3.1.2. Materials and Software

The materials in Experiment 2 were almost a complete departure from those used in Experiment 1. The participants were seated in front of a computer screen with a mouse for manual control and a microphone for verbal control conditions. The participants used the Human-Robot Interaction Workload Measure (HRI-WM; [19]) to document their subjective workload throughout the experiment.

Microsoft Voice for Vista [20] was used to train SS-RICS to the particulars of the participant's voice. The participant spoke prescribed sentences into the microphone and the software created a voice profile from this input, which was accessed as the participant spoke commands to SS-RICS.

The SS-RICS interface was used during the robot navigation portion of each trial. Although there are multiple ways to configure that interface, every participant in every trial had the same settings configuration. The components of the interface that may have been used by participants included a list of verbal command options, a window listing history of verbal commands, two directional keypads, and two streaming video windows from the perspective of the robotic platform, as shown in Figure 5. The directional keypad on left controlled the camera and allowed the participant to click to move the camera, up, down, left, right, or the straight-ahead home position. The keypad on the right directed the robot to turn left, turn right, move forward, or move backward.

The robot was a Pioneer P3-DX, with a mounted LIDAR system and color video camera. The robot was always located within a space in a laboratory facility. Practice sessions occurred in a single room in the laboratory facility.

3.1.3. Procedure

The participants began with voice training. Once the voice training was over, participants were given experimental instructions. Participants were instructed that they would need to control a robot to find a target object (i.e., a cardboard box with a black and white paper wrapping to make it resemble a bomb) that was shown to the participants.

Participants then completed three practice trials, one for each of three robot control conditions from Experiment 1. The practice trials were performed in the same room and the target was hidden in three different locations. Participants were able to take as much time as needed to feel comfortable with the control condition.

The participants received more instructions after completing the practice trials. On half of the trials, simple arithmetic (i.e., addition and subtraction) problems were performed with two single-digit addends and a single-digit solution. On math task trials participants verbally responded to the problems which were verbally presented every eight seconds. If the participant did not respond correctly prior to the start of the next arithmetic problem, the answer was marked as incorrect.

Each participant was also instructed that they would receive a map before each trial, which they could study while the next trial loaded. The map contained the starting position and orientation of the robot, structural features of the navigation space and objects placed in the space. Room labels on the map were added to each of the four rooms. For each participant, the starting location was different on every trial and equidistant from the starting location. An example map is represented in Figure 6.

The participants had eight minutes to find each target. If the participant did not announce finding the target before eight minutes expired, the trial ended. Overall, there were six trials with one trial for each combination of control and math task conditions. The order of conditions was randomized. Although multiple trials of each condition combination would have been ideal, technical limitations of the robot limited the experiment to three practice and six experimental trials.

In the manual condition, participants used a mouse to interact with the SS-RICS directional keypads to direct the robot and camera movement. In the egocentric verbal condition, participants spoke commands into a microphone. The commands were the same type used in Experiment 1 with the exception that the participants were not able to use degrees for a turn. A command to turn left or right was interpreted as a turn of 90 degrees. The participant could also request to turn left or right "slightly," which was interpreted to be a ten degree turn. Giving the command "turn around," made the robot turn 180 degrees.

The exocentric verbal condition included the same commands as the egocentric condition with some additional exocentric commands. First, the "go to door" command moved the robot to the nearest door. Next, the "go down hallway" command moved the robot down the nearest hallway in the direction that the robot was facing. Participants were able to instruct the robot to "go to" any given room (e.g., "go to room 2"). Lastly, the participant could instruct the robot to "search room" followed by a room number and the robot would search the requested room through a series of preprogrammed way points.

After each trial, participants responded to a modified version of the HRI-WM [19] using E-Prime [21]. The functionality of the measure remained the same expect for the fact no comments could be added by the participant.

The HRI-WM was divided into two parts. In the first part, participants assigned proportions of workload to five different factors of the robot navigation task. A one word label was chosen for each factor and participants were instructed on their meaning before the first time taking the HRI-WM. Configuration workload was described as difficulty "controlling the robot without help." Task workload was defined as finding the bomb and doing the math task, if that was required. Context workload was the environment and the interface. Team process workload was difficulty communicating with the robot. Finally, system workload was the properties of the robot itself (e.g., its height or speed). For the second part, the participants were asked to rate on a scale of one to five the amount of workload that they felt was required to perform the trial with higher values meaning a feeling of being very mentally taxed.

Participants entered this data through an interface that included a color-coded sliding bar scale for proportion of component contribution to overall workload. The bar had twenty slots filled with the background black color, which participants could fill by clicking to increase through button clicks on up arrows for each factor. The participant could adjust the proportion by decreasing any of the factors. The participants were instructed that the whole bar did not need to be filled as long as it reflected the desired proportions. Participants selected the overall workload by clicking to fill in a box above the desired number. The overall workload could be adjusted by clicking on a different box. An example interface is shown in Figure 7.

3.2. Results

In the course of all experimental trials, technical problems caused three participants (two women and one man) to have an uninterpretable trial each. The lack of replicates for condition combinations prevented analysis of the rest of the data from these participants and the data was removed from all subsequent analyses.

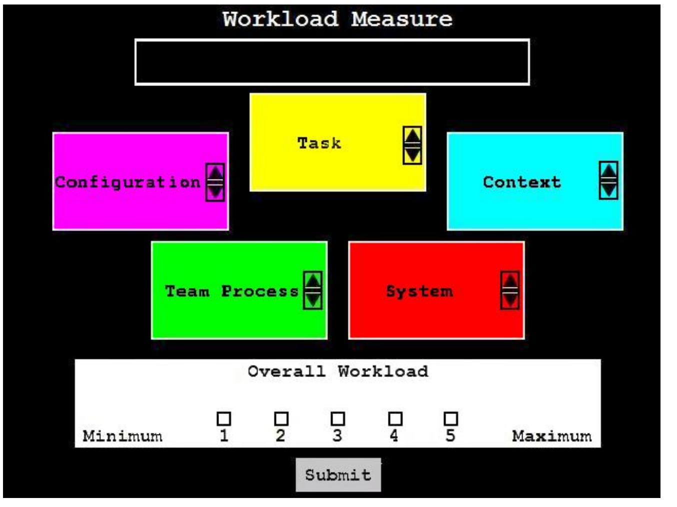


Figure 7. An example of the HRI-WM. Clicking the up or down arrow for one factor increased or decreased the amount of that color in the bar at the top of the page.

3.2.1. Performance Results

Performance variables included target search accuracy, distance navigated, and math task accuracy. Math task response time did not demonstrate any differences. Distance was a calculation indicating the amount of the environment investigated by a participant in each trial. None of these variables had a successful-trial-only filter like search time in Experiment 1.

A three by two repeated-measures ANOVA (Control: Manual, Egocentric, Exocentric / Math: Yes, No) was conducted on target search accuracy (overall mean = .560). A difference was detected among the levels of the control variable, $F(2, 26) = 4.48$, $MS_E = .162$, $p = .021$. Paired-samples t-tests with a Bonferroni correction revealed that Exocentric control (.714) had higher accuracy than Egocentric (.393), $t(13) = 2.86$, $SE_M = .113$, $p = .013$. No difference was detected with the Manual condition (.571).

Another three by two repeated-measures ANOVA was run on distance traveled by the robot across all trials (overall mean = 114 ft). The distance was calculated by adding the length of all hallways traversed and the perimeter of all rooms searched. This calculation was used to account for the distance that was put in view of the robot's camera rather than an absolute distance traversed by the robot. The absolute distance would have necessarily included jitter that would have increased the distance, but not aided the visual search process. A path that would have covered the maximum of 177 feet is included in Figure 6. A difference in the control variable, $F(2, 26) = 3.74$, $MS_E = 928$, $p = .037$, was detected. Follow-up paired-samples t-tests with a Bonferroni correction found that exocentric control (122.9 feet) resulted in more distance than egocentric control (101.6 feet), $t(13) = 2.86$, $SE_M = .113$, $p = .013$, which was an increase of almost 20% efficiency. There was no difference found with manual control (117.9 feet).

Finally, a one-way repeated-measures ANOVA isolated to the control variable was run on math task accuracy on the half of trials in which the math task was performed (overall mean = .910). A difference was detected within the control variable, $F(2, 26) = 5.93$, $MS_E = .0043$, $p = .008$. Subsequent paired-samples t-tests with a Bonferroni correction detected that egocentric control (.877) had lower accuracy than manual control (.959), $t(13) = 3.05$, $SE_M = .0266$, $p = .009$. No difference was found with exocentric control (.895), although a comparison with manual would have been significant, $t(13) = 2.16$, $SE_M = .0264$, $p = .050$ if not for the required Bonferroni correction for multiple t-test comparisons. Figure 8 presents all significant differences between performance variables.

3.2.2. Subjective Results

The HRI-WM results from two participants were compromised due to technical problems and were not analyzed in the remaining analyses. The HRI-WM measured overall workload ratings as well as proportional contribution of five components of workload.

First, a three by two repeated-measures ANOVA was conducted on overall workload ratings, which reflects how taxing a trial was in the participant's opinion on a five point scale. A main effect of math task found that participants felt trials with a math task were more taxing (3.06) than with no math task (2.53). Another possible difference was detected among the levels of the control variable, $F(2, 22) = 3.42$, $MS_E = 5.01$, $p = .051$. A single paired-samples t-test was run to compare the most theoretically interesting of the factors (i.e., exocentric and egocentric control). Egocentric control (3.21) had greater workload ratings than exocentric (2.58), $t(11) = 2.26$, $SE_M = .276$, $p = .045$.

The proportion of workload attributed to each component in the HRI-WM was also analyzed using three by two ANOVAs. Only statistically different findings are reported.

A significant interaction for the team configuration variable (i.e., number of operators to number of robots, in this case identified as "controlling the robot without any help") was found, $F(2, 22) = 5.39$, $MS_E = .012$, $p = .012$. Follow-up paired-samples t-tests with a Bonferroni correction found two statistical differences. First, workload ratings for exocentic control with the math task (.201) was greater than exocentric control without the math task (.131), $t(11) = 2.68$, $SE_M = .026$, $p = .021$. Second, the workload ratings for exocentric control with the math task was greater than egocentric control with the math task (.149), $t(11) = 2.83$, $SE_M = .018$, $p = .016$.

The task component had a significant main effect of math task, $F(1, 11) = 1.26$, $MS_E = .0048$, $p = .005$. Participants rated the task component with higher workload when there was a math task (.234) than when there was no math task (.182).

Finally, there was a possible difference within the control variable for the team process component (i.e., communicating with the robot), $F(2, 22) = 3.41$, $MS_E = .023$, $p = .051$. A single paired-samples t-test was run to compare the levels of the control variable of most theoretical interest (i.e., exocentric and egocentric control). The result showed a greater proportion of workload attributed to the team component for egocentric (.331) over exocentric control (.241), $t(11) = 2.35$, $SE_M = .039$, $p = .038$. Figure 9 presents the results of all significantly different workload findings.

3.3. Discussion

The target search accuracy and distance traveled results show a clear benefit for the ability to use limited exocentric commands in Experiment 2. The theory that increased workload affects the ability to control a robot was bolstered by two findings. First, there was lower math task accuracy for egocentric verbal control versus manual and second, a lack of a difference for exocentric verbal control. In Experiment 2, the math problems were answered verbally, which represents greater mental taxing on a concurrent verbal task (e.g., verbal control) than on a concurrent manual task (e.g., manual control) [17]. The higher math task accuracy for the manual control condition is not surprising. The lack of difference between manual and exocentric conditions and

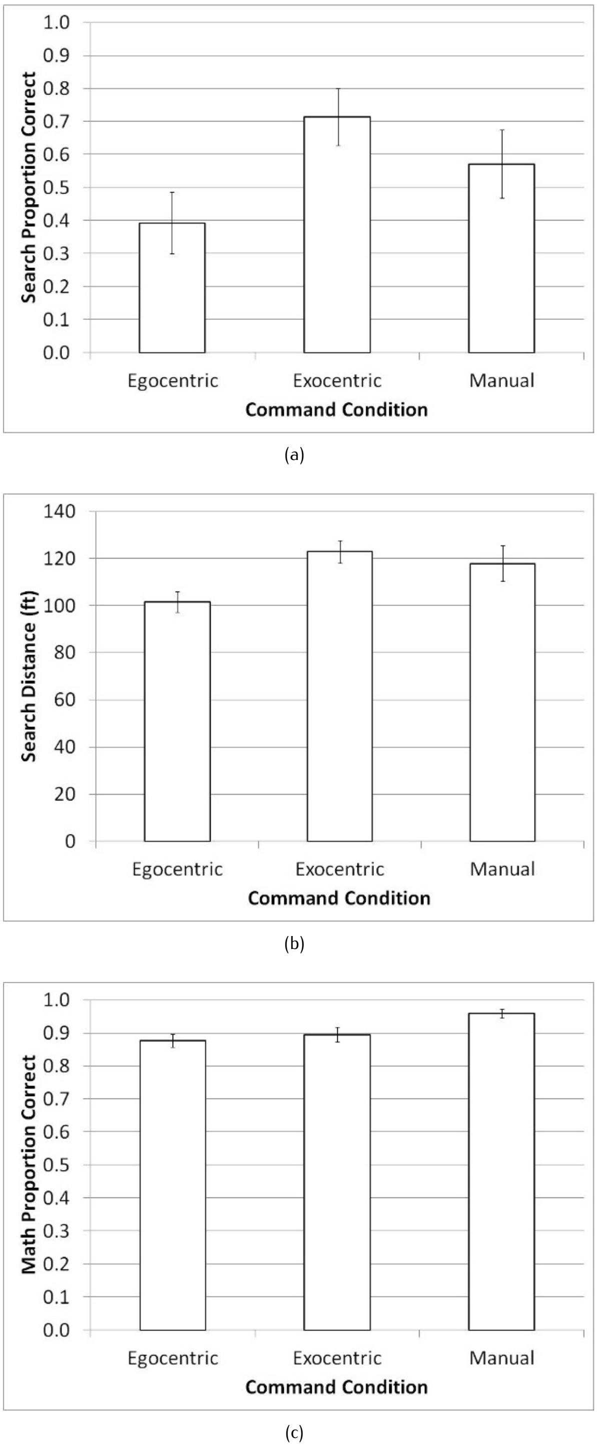
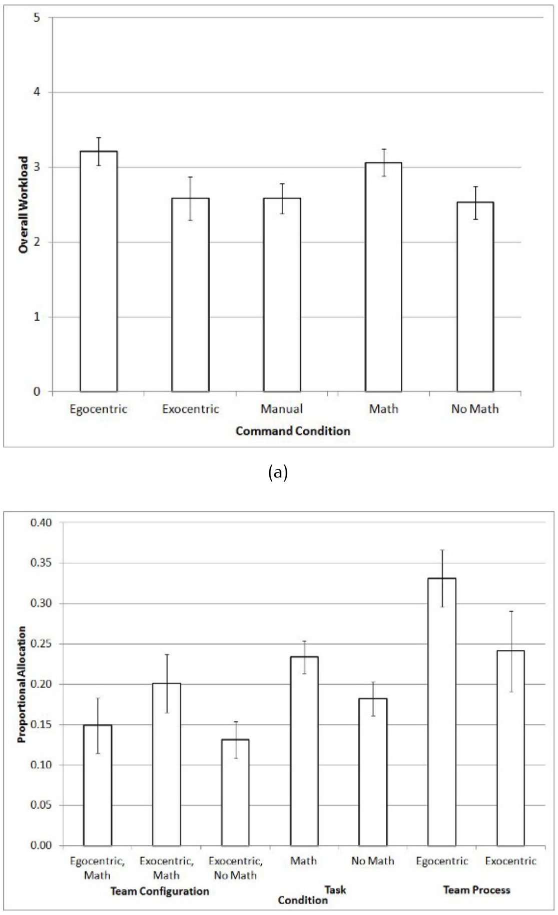


Figure 8. Performance variable in Experiment 2. Panel A represent search accuracy. Panel B represents distance searched. Panel C represents math task accuracy.



(b)

Figure 9. The workload variables of interest in Experiment 2. Overall workload is represented in Panel A. Workload component data including team configuration, task, and team process workload results are represented in Panel B.

a difference between manual and egocentric conditions draws the conclusion that the egocentric condition was more mentally taxing than the exocentric condition.

Two additional conclusions may be implied based on higher workload ratings when math was present and in the egocentric condition over the exocentric. First, the math task was an effective tool for increasing subjective workload. There was some concern that single digit arithmetic problems would not be taxing enough, yet the results reveal that the secondary task did create enough workload to be detected by participants. Second, greater subjective ratings of workload when in the egocentric condition indicate that the participants felt more taxed when location labels were not available. This reinforces the theory that participants are more mentally taxed when precise directional commands are required.

Another interesting set of results were from the proportional allocation of workload components with the HRI-WM. First, the task component had higher attributions of workload when there was a math task than when there was no math task. This again reinforces the notion that the math task represented a noticeable increase in mental workload.

The team configuration component also has interesting implications. When participants had exocentric command options while math was present, they felt that controlling the robot without help was more of a factor in determining workload than performing the math and search-and-find-tasks. This suggests that operators had some trouble in deciding which commands to select since the addition of exocentric options was the difference between the two conditions. This is in line with the hypothesis based on the Experiment 1 results that one of the difficulties of the exocentric condition in Experiment was selecting which commands to issue.

The last finding was that the team process component received more weight in the egocentric control conditions than with exocentric control. Team process workload was described as difficulty communicating with the robot and so the results indicate that the participants felt it was difficult to give commands to the robot. During the execution of the task, the robot had a high probability of being able to process a command issued by the participant. Given that more commands were likely given in the egocentric condition as the participant adjusted the robot's movements, this result is not surprising. This team process result may be attributed to a theoretical cause given that it may be a mechanical problem, but we posit that this is a legitimate concern nonetheless. If egocentric-only commands are the only options for a robot, then the operator will feel more workload as misunderstood commands grow. Exocentric commands options would improve performance over egocentric-only commands.

4. Conclusions and Future Work

4.1. Conclusions

The two experiments discussed above shed light on robot operator preferences for navigational language used for controlling a robot and how preferences translate into task performance (though it is important to consider that the experimental scenarios were narrowly constrained and had arguably low sample sizes). Taken at face value, the experimental results implicate a plan for how to proceed with the development of a cognitive-based architecture like SS-RICS in the most efficient manner that will also increase robot operator performance.

The Experiment 1 results indicated that exocentric command options may have slowed down target search. However, Experiment 2 showed the expected performance improvements in target search accuracy and distance when exocentric commands were allowed. Experiment 2 results also showed a decrease in secondary task performance for egocentric control relative to manual, yet there was no difference between exocentric control and manual. This suggests that there was less cognitive processing associated with exocentric than egocentric control as confirmed by lower overall workload ratings for exocentric than egocentric control.

The proportional allocation of the workload attributes is an additional point of interest. In the exocentric with math condition, participants tended to think that controlling the robot without help was more difficult than in any other condition. This seems to suggest that even with a limited set of exocentric label options, participants had difficultly selecting where and when to search. In the egocentric condition, a search strategy of serial location investigation based on starting location made sense; however, in the exocentric condition, participants could command the robot to go to any one of four rooms. This property of exocentric control apparently warranted increased workload ratings, thus

adding confirmatory evidence that plentiful object label options (as in Experiment 1) is a significant source of workload.
Another interesting finding is that in the egocentric condition participants found it more difficult to communicate with the robot than when using exocentric control. It was necessary to give more commands in the egocentric conditions than the exocentric condition; therefore, it would make sense that communication was more difficult if each command faces a certain small chance of not being correctly processed. Although this result seems to underlie a physical problem with the vocal utterances or a failure on the part of SS-RICS to correctly parse the misunderstood utterance, this still poses an important practical limitation to robot control systems that can receive only egocentric commands.

4.2. Future Work

A clear path for developing SS-RICS emerged from the results of both studies. First, developing exocentric command options is an important step to improve performance when operators control indoor robots. The results of Experiment 1 appeared to be in contradiction to this conclusion, but the changes in experimental method for Experiment 2 clarified the approach. Instead of open-ended exocentric commands, SS-RICS would benefit from commands limited to rooms, halls, and doors. These are not so much objects as they are locations inside a building that define the structure of the building. Not only were these the commands that led to improved performance in Experiment 2, but they are also the natural exocentric extensions from Experiment 1. Another approach proposed in [22] that recommends object recognition though functional identification of objects in the environment could appreciably expand SS-RICS vocabulary.
The robot executed the exocentric commands in Experiment 2 because pre-loaded maps included labels for doors, halls, and rooms. Preloaded maps are often impossible, so SS-RICS needs to be developed to recognize doors, halls, and rooms on its own. Currently, SS-RICS can recognize doorways and halls, but not doors or rooms. To ensure recognition of doors and rooms and to allow adaptability for hall and doorway recognition, SS-RICS will likely have to process additional payload inputs and integrate recognition algorithms.
A third development that would improve SS-RICS is the ability to upload maps (i.e., building floor plans) not generated by SS-RICS. The maps used in Experiment 2 were not just pre-loaded, but they were built by allowing SS-RICS to navigate the experimental environment. Based on the aforementioned results, the capability to upload maps in SS-RICS would most likely enhance mission performance, in general. In addition, SS-RICS would not have to previously explore the environment to generate a map.
Control systems should be designed to conform to behavioral patterns and preferences of robot operators. Keeping in mind the narrowness of the scenarios and the arguably low number of participants the experimental results from these two studies indicate that exocentric verbal control improves performance, but only if there is a limited number of commands. Integrating the changes described above into a robot control system would ultimately improve robot operator performance, and reduce workload overall.

References

[1] S. Grozny, Robust Adaptation to Non-Native Accents in Automatic Speech Recognition. Springer-Verlag, New York, 2003
[2] L. Talmy, How language structures space. In H. L. Pick, Jr., L. P. Acredolo (Eds.), Spatial Orientation: Theory, Research and Application, Plenum Press, New York, 1983, 225-282
[3] D. Sofge, M. Bugajska, J.G. Trafton, D. Perzanowski, S. Thomas, M. Skubic, S. Blisard, N. Cassimatis, D. Brock, W. Adams, Human-Robot perspective taking: Collaborating with Humanoid Robots in Space. International Journal of Humanoid Robotics, 2 (2005), 181-202
[4] G.K. Aguirre, M. D'Esposito, Topographic disorientation: A synthesis and taxonomy, Brain 122 (1999), 1613-1628
[5] R. Moratz, K. Fischer, T. Tenbrink, Cognitive modeling of spatial reference for human-robot interaction, International Journal of Artificial Intelligence Tools 10 (2001), 589-611
[6] T. Tenbrink, K. Fischer, R. Moratz, Spatial strategies in human-robot communication. Zeitschrift füur Knstliche Intelligenz, 16 (2002), 19-23
[7] R. Moratz, T. Tenbrink, J. Bateman, K. Fischer, Spatial knowledge representation for human-robot interaction, In C. Freksa, W. Brauer, C. Habel, K.F. Wender (Eds.), Spatial Cognition III: Routes and Navigation, Human Memory and Learning, Spatial Representation and Spatial Learning, Springer, New York, 2003, 263-286
[8] T.D. Kelley, International Journal of Advanced Robotic Systems, 3 (2006), 219-222
[9] S. Thrun, Simultaneous localization and mapping. In M. Jefferies, W.-K. Yeap, (Eds.), Robotics and Cognitive Approaches to Spatial Mapping, Springer, Berlin, 2008, 13-42
[10] R. Moratz, T. Tenbrink, Spatial Reference in Linguistic Human-Robot Interaction: Iterative, Empirically Supported Development of a Model of Projective Relations. Spatial Cognition and Computation, 6 (2006), 63-107
[11] N. Pinto, D.D. Cox, J.J. DiCarlo, Why is real-world visual object recognition hard? Computational Biology, 4 (2008), 0151-0156
[12] J.F. Kelley, An iterative design methodology for user-friendly natural language office information applications, ACM Transactions on Office Information Systems, 2 (1984), 26-41
[13] D. Perzanowski, D. Brock, W. Adams, M. Bugajska, A.C. Schultz, J.G. Trafton, Finding the FOO - A Pilot Study for a Multimodal Interface, Proceedings of the IEEE Systems, Man, and Cybernetics Conference (2003, Washington, DC), 3218-3223
[14] Red Storm, Raven Shield: Tom Clancy's Rainbow Six, Version 3, computer software, Ubi Soft, Montreal, Quebec 2003
[15] S. G.Vandenberg, A.R. Kuse, Mental rotations, a group test of three-dimensional spatial visualization, Perceptual and Motor Skills, 47 (1978), 599-604
[16] J.L. Zaichowsky, Measuring the involvement construct, Journal of Consumer Research, 12 (1985), 341-352
[17] C.D. Wickens, Multiple resources and performance prediction. Theoretical Issues in Ergonomic Science, 3 (2002), 159-177
[18] D.N. Cassenti, T.D. Kelley, Towards the shape of mental workload. Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting (2006, Santa Monica, CA), 1147-1152
[19] R.E. Yagoda, Development of the Human Robot Interaction Workload Measurement Tool (HRI-WM), Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting (2006, Santa Monica, CA) 2010, 304-308
[20] Microsoft, Micrsoft Voice for Vista, computer software, Microsoft, Seattle, WA 2006
[21] Psychology Software Tools, E-Prime version 2.0, computer software, Psychology Software Tools Inc., Pittsburgh, PA 2009
[22] R. Moratz, Intuitive linguistic Joint Object Reference in Human-Robot Interaction: Human Spatial Reference Systems and Function-Based Categorization for Symbol Grounding. Proceedings of the American Association of Artificial Intelligence 22nd Annual Meeting (2006, Menlo Park, CA), 1483-1488