Where You Point is Where the Robot is

Hokyoung Ryu
Massey University
Auckland, New Zealand
+64 (0)94140800 ext. 9140
h.ryu@massey.ac.nz

Woohun Lee
Korea Advanced Institute of Science and Technology Daejon, Republic of Korea
+82 (0)428694519
Woohun.lee@kaist.ac.kr

ABSTRACT
It is virtually envisioned that in the near future home-service robots will be assisting people in their daily lives. While a wide spectrum of utility of home-service robots has been proposed, i.e., cleaning, surveillance or go-and-fetch jobs, usability studies of the home-service robots have been less undertaken. This paper explores the usability issues, in particular, a map-based user interface for instructing home-service robots in the home environment. It focused on how the different map representation of the co-located environment would affect task performance of locating the home-service robots. The effectiveness of the map-based human-robot interface was thus analysed according to the dimensionality of the map, the location information of the elements in the co-located workspace. The experimental results showed that task performance was varied by the different map representation, providing a better understanding of what characteristics of the map representation were able to effectively support the human operator in instructing the home-service robots in the home environment.

Author Keywords
Map-based human-robot interaction, Design guidelines, Home-service robots, Human factors.

ACM Classification Keywords
H.1.2 User/Machine Systems, H.5.2. User Interfaces, I.3.6. Methodology and Techniques

INTRODUCTION

Many studies on Industrial Ergonomics (IE) or Human-Computer Interaction (HCI) have suggested a range of design implications for personal computing or technological environments, in an effort to ensure that new systems can accommodate intrinsic human capabilities. However, as the personal computing environment continues to advance we are led to revisit an underlying theme; whether our current understanding of the personal computing or technological environment is still valid when exploring new types of system, such as personal home-service robots.

In fact, the use of HCI techniques is somewhat unusual in Robotics studies [4], even though the premise of Human-Robot Interaction (HRI) is well fitted to HCI. Perhaps, it is because the main emphasis on Robotics studies is still placed on developing fully-autonomous systems, as a consequence, there have been few advantages for designers to consider effective human-robot interfaces. Second, current applications of HRI, e.g., NASA's Mars Rover, assume that all robot users are trained experts, and they will benefit less from well-designed interfaces. The two reasonable factors naturally have led few HCI researches and fewer Robotics researchers still who advocate the use of HCI techniques for creating effective human-robot interfaces, except Yanco et al. [15, 16] and Scholtz [11] who recently emphasized the understanding of collaborating work patterns between the human and the robot would be the most important design issue in HRI.

In this paper, we consider HRI under the HCI sphere. Following this, we apply the HCI techniques to the design of HRI and such an approach is expected to offer significant understandings of the human operator's cognitive capability in collaborating with robots. Naturally, it may provide the different accounts of HRI against the current understandings of HCI. In effect, this paper is expected to provide Robotics researchers with the issues involved in creating effective human-robot interfaces.

UNDERSTANDING HRI

One of the contrasting characteristics of HRI over HCI is that, generally, robots are mobile with supervision of human operators who may also be mobile. The autonomy of the robot, which enables to perform tasks in lieu of the human

Chinz'06, July 6–7, 2006, Christchurch, New Zealand.

operator, may allow us to describe HRI as Computer-Supported Cooperative Work (CSCW) [15]. Indeed, the canonical "time-space" framework in the CSCW domain categorizes HRI applications in a meaningful way. Robots, such as NASA's Mars Rover, that are designed for exploring space missions fall into the category of *asynchronous and non-co-located*, because they are highly autonomous and are remotely located from their collaborative team of human operators. By contrast, urban rescue robots operate primarily in a *synchronous and non-co-located* manner as they explore buildings or spaces too dangerous or too small for humans to work. Assistive robots, such as home-service robots (e.g., iRobot's Roomba™), operate in a *synchronous and co-located* fashion as they are intended to help a person to live better, sharing the environment with their human operators. Out of the wide spectrum of the HRI applications, synchronous and co-located applications, which are receiving wide attention as a place for the development of recent robotics technologies [5], require an effective interface for spatial interaction, ensuring monitoring of the robot and maintaining the safety of the human operator.

Within the scope of this paper, we narrow down the wide spectrum of the HRI applications into the synchronous and co-located HRI, in the sense that, firstly, the application is most likely to take up a future place of the robot development; and, secondly, it asks a higher level of collaborative interaction, mixing the human and the robot in an unstructured and uncontrolled environment.

Map design of the co-located workspace

Most robots working in the co-located environment populated with people have been instructed by the human operator, depending on *(i) figuring out where both the robot and the operator are; (ii) determining where the robot need to go, and (iii) moving the robot into the to-be-located place*. As a consequence, key design factors of the human-robot interface for the co-located workspace are to provide the relevant information and to implement an appropriate interaction style.

As to the former concern, the CSCW studies identified that all the levels of awareness, such as peripheral awareness [6] and workspace awareness [7], help collaborative participants to know who else are working in the shared workspace, what the others are doing and how they collaborate in trouble-free. *Peripheral awareness*, being aware of all the collaborative participants' existence (or their location), provides who else are working together. *Workspace awareness*, which is about who is working on what, allows the collaborative participants to have the up-to-the-second knowledge of the other participants' interactions with the elements in the environment.

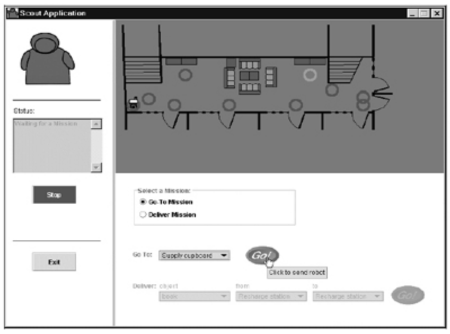


Figure 1. A map-based HRI interface, excerpted from [7]

In the same fashion, collaboration between the human operator and the robot asks the understanding that the human operator has of the location, and surroundings of the robot; and the constraints under which it must operate [3]. They help the human operator to complete human-robot collaborative activities in an effective and safe way.

There are a wide spectrum of collaborative activities that are currently being performed by human-robot systems, e.g., cleaning, surveillance, go-and-fetch jobs. Both cleaning and surveillance tasks appear to be well performed with little human operator's intervention. By contrast, go-and-fetch tasks that are envisioned to widely help elderly or disabled people require more human operator's instructions than the other tasks [7]. Carrying out go-and-fetch tasks depends on the three types of spatial understanding of the workspace: *landmark knowledge, route knowledge and survey knowledge* [13]. In particular, landmark knowledge that is characterized at an early stage of spatial understanding of an environment enables to build route knowledge that is essential to build a robot's traversed route, avoiding obstacles in the shared workspace. In this sense, map-based navigation *per se* quickly establishes these two types of knowledge should it be appropriately designed. See Figure 1. Huttenrauch and Norman [7] intended to provide these types of knowledge, by placing familiar objects, e.g., a coffee table and couches, on the map with the brief sketches of the objects. Therefore, the human operators were able to locate the robot into the position in which they want to locate, referring to the objects on the map. Yet, the map interface has some limitations. The implicit representation of the landmarks and the workspace may ask the human operator's local perception. Quite often, for the better workspace awareness and situation awareness, many vehicle navigation studies, e.g., Wickens [14], employed three dimensional maps for explicitly presenting the workspace and the objects.

Several virtual reality studies [e.g., 2] adopted a simple way to extend awareness of the landmarks on the map – the attachment of legends to the landmarks. Consider Figure 1

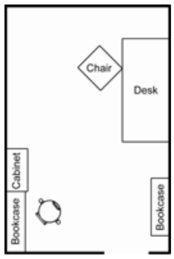


Figure 2. The map with the textual labels of the objects

again. Here, the perception of the landmarks would be reliant on the boundary drawings of the objects. By contrast, Figure 2 has the corresponding label of each object. The benefit from this labelled map is most likely to help human operators to readily comprehend the meaning of the objects without any local viewing. Further, many studies [e.g., 9] demonstrated that the locating tasks tended to be specified in locational reference terms such as "below the desk", rather than in the absolute location of the destination. It can thus be thought that the labelled maps quickly provide the to-be-located places with appropriate locational reference terms, even when the objects cannot be locally perceived due to some barriers between the landmarks and the human operator [9].

In conjunction with awareness of the workspace and the landmarks, the human-robot collaborative activities also ask where the collaborative participants are, in order to avoid possible collision among them. We can simply expect that the explicit representation with respect to the presence and location of both the collaborative participants would lead a better collaboration in the co-located situations.

Interaction design for HRI tasks

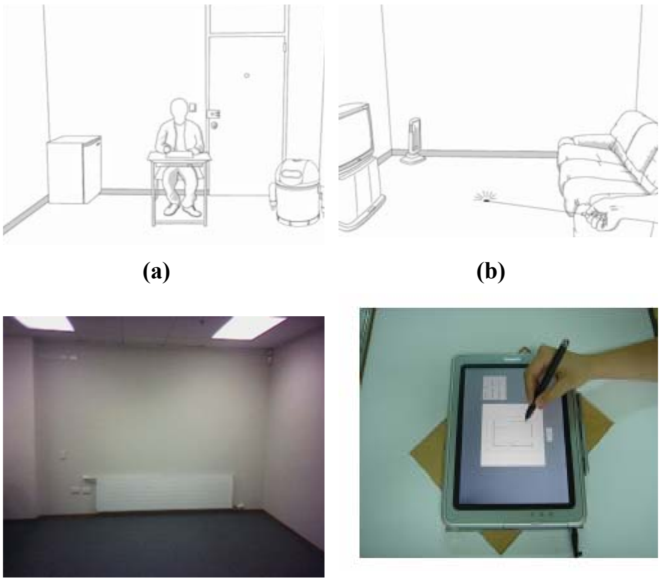
Apart from how to design the map for HRI tasks, an appropriate interaction style that exploits the maps should be considered. In this respect, perhaps, natural language interaction with the map [e.g., 8] might be the most attractive means, because the human operator does not need to remember how to use the map interface. However, the ambiguity of natural language is inevitable. By contrast, direct manipulation with the map has been considered as a most practical manner, replacing complex verbal commands with actions to manipulate directly the visible objects or points on the interfaces. For instance, many HRI studies, e.g., Huttenrauch et al. [7], Fong [4], and Skubic et al. [12], adopted direct manipulation interaction on their map-based interfaces, mostly with Personal Digital Assistants (PDAs). Yet, the PDA-based direct manipulation has several drawbacks. Firstly, it does not show a cursor that provides feedback of the current position of the possible point selection. Point selection with the stylus pen, therefore, results in no opportunity to figure out misjudgement of the point selection. Furthermore, technically, it cannot avoid optical distortion such as parallax error. Optical distortion in PDAs or touch-sensitive screen input generally is due to the fact that the image displayed in the touch-sensitive area is projected by a mirror located 2 millimetres below the touch screen. Therefore, even when the human operator precisely taps a point that is believed as being correct, the point hit generally is several millimetres away from the one that they want to select. In particular, this distance error depends on the angle of the stylus pen and the human operator's angle of vision. Finally, the technical specification of PDAs should also be considered, e.g., the size of the tip of the stylus pen and touch sense resolution. The tip size is varied, but generally more or less 0.5 mm. As a consequence, it cannot provide more precise pointing performance beyond this. Further, the tip size issue is closely related to the touch-sense resolution of the screen. This touch-sense resolution of PDA is determined by the number of sensors that detect the contacts. For instance, the commercial PDAs can provide around 0.2 mm touch-sense resolution. In Huttenrauch et al.'s study [7], for instance, a PDA – screen size 57 (width) × 76 (height) mm with 240 × 320 pixels screen resolution – was used to accommodate a large office area. It meant that roughly one pixel, i.e., around 0.24 mm, of the screen covers 10 inches more or less in the office area. Compared this setting with the tip size (more or less 0.5 mm) and the touch-sense resolution (0.2 mm) of the commercial PDAs, it is obvious that the Huttenrauch et al.'s study cannot avoid some distance-related errors as the human operator locates the robot into a particular place.

Nonetheless, this paper considers a handheld device with stylus-pen input, thanks to the portability of the handheld device, which effectively addresses human operator's mobile situations. To achieve a compromise of the problems discussed above, instead, the experiments in this paper employed a smaller space – 3.9 × 6 meters – and a Tablet PC that has higher touch-sense resolution than PDAs, so it is expected to partially lessen the touch-sense related problem of PDAs. Yet, the real map size that was used in the experiments is 39 × 60 mm, so it is mostly affordable in most PDAs.

Thus far, we discussed many design challenges that may be associated with map-based human-robot interfaces, ranging from psychological aspects of map representation to physical interaction issues. Yet, the aim of this paper is not empirically validating the design of a map-based human-robot interface in the home environment; that could only be done by studying its use when collaborated with the robot in a real use context. Rather, the subtle difference between HCI tasks, particularly vehicle navigation tasks, and HRI tasks in the home environment simply raised the questions of the following experiments and the results suggest a fragment of the design guidelines of map-based interfaces in synchronous and co-located human-robot interaction.

EXPERIMENTAL TASK

All the experiments in this paper were conducted in a 3.9 × 6 meters room with a Tablet PC as shown in Figure 3(d). The dimension of the map was 39×60 mm. It was presented on Compaq™ Tablet PC T1000 with 195 × 300 pixels. One pixel on the touch-sensitive screen was thus equivalent to 4 cm^2 in the real size of the room, which is much highly granulated than Huttenrauch et al.'s study [7]. The pointing tasks were performed with a stylus pen, of which the tip size is around 0.3 mm. To lessen optical distortion that partially comes from the angle of the stylus pen and the human operator's angle of vision, participants were asked to have the same posture to use the stylus pen and placing the PDA at the same position, as shown in Figure 3(a).



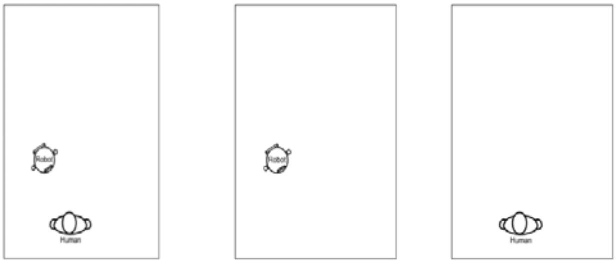
(a) (b) (c) (d)

Figure 3.

A typical experimental setting in this paper. (a) A participant was sat on the desk chair and tapped the to-be-located place of the robot on the map-based interface with a stylus pen, the actual robot was located in the left side of the participant, but it did not move to the to-be-located place as the participant tapped the point on the touch-sensitive screen; (b) the experimenter indicated the to-be-located places by using a laser pointer. They were predefined, tagged by a red sticker on the floor; (c) the room in which the experiments took place, and (d) the participant pinpointed the to-be-located places on Compaq™ Tablet PC T1000 with a stylus pen

Yet, the robot in the experimental tasks did not actually move to the places followed by the participant's instructions. This way of experimental settings has two advantages. Firstly, the primary concern of these experiments shows how the human operators would establish their spatial cognition via the map representation, so the best map would provide the better pointing performance. Experiment 1 for example, assessed whether the presentation of the human operator's position on the map would make effects on their spatial cognition to initially generate appropriate commands to the robot, and we thought that this could be achieved without any robot on the move. Second, this context of use is arguably more natural in the sense that this setting removes all distraction, which is not relevant to spatial cognition of the to-be-located places, and the whole procedure emphasizes to perceive the to-be-located places very simply and optimally.

There are of course, disadvantages to this experimental setting. Firstly, the behavioural record is limited to an error distance between the to-be-located places that were specified by the experimenter and the actual locations that were indicated by the participants. If the robot can follow the instructions, its error could be lessened by another instruction, guiding the robot into the correct position. However, it is notable that the tasks performed in this paper are not navigation guiding tasks that have been considered in the previous HRI studies [8, 12], but one-off locating tasks. Second, there might be included a certain level of uncertainty between the experimenter and the participants as for where the robot should be located. This drawback was avoided by: (a) defining the to-be-located places in advance, and (b) having a small-scale learning session so that random clicking was virtually minimized. Experiment 3 for instance, twelve to-be-located places were predefined. Each to-be-located point was tagged on the floor of the workspace with a red sticker at 2cm diameter. Therefore, if necessary, the participants were allowed to look at the to-be-located places while they were performing the tasks. The to-be-located places on the floor were not presented on the interface, only the participants were allowed to reason about the places on the map using both the map given and their own local perception. In addition, to avoid random clicking, all the participants had been asked to perform two trials before they carried out their main experiment. With this simple learning session, the participants were allowed to make brief experience how the to-be-located places would be indicated by the experimenter, and how they could pinpoint the to-be-located points on the map-based interface. These limitations of this experimental setting may underlie the accounts of the map-based interface which would be concluded by this paper. Yet the capital objective of the experiments in this paper was to focus on the map representation for controlling home-service robots and identifying the to-be-located places.



(a) (b) (c)

Figure 4. The three map representation in this experiment. (a) the current positions of the human operator and the robot are presented on the map, (b) only robot, (c) only human operator.

EXPERIMENT 1

The introduction of this paper discussed that typical HRI tasks may ask the locational information of the collaborative participants. Experiment 1 was intended to empirically investigate this issue. To emphasize the effects of the perception of the two collaborative participants, all possible distraction, such as surrounding objects, was removed. The form of representing the two collaborative participants on the map was the main manipulation (independent variable) in this experiment. The measure taken (dependent variable) follows from the practical implication of users having difficulty about the locating tasks, that is, they will lead larger error distance between the to-be-located places and the actual points on the Tablet PC. The three experimental groups were formed by the three types of representation of the two entities, as shown in Figure 4. If the absences of any entities in the map representation would have effects on task performance, it would lead to an unequal task performance against the map representing both the human operator and the robot.

Method

Participants

30 participants were all the undergraduate students at Massey University. Upon completing Experiment 1, two dollars were given to the participants.

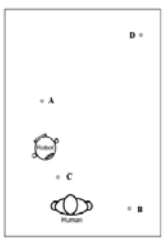


Figure 5. The four-to-be-located places

Design

The experimental design was a two-way mixed design. The different map presentation, as shown in Figure 4, was the between-subject independent variable. The four to-be-located places, as shown in Figure 5, served as the other within-subject independent variable. The sequence of the trials of the four to-be-located places was counterbalanced using a Latin square. The dependent variable, the Euclidian error distance between the to-be-located point that was asked by the experimenter and the actual point that the participants pinpointed on the Tablet PC, was used to assess the effect of the independent variables.

Apparatus

The room equipped with no objects (landmarks), except a desk and a desk chair on which participants sat and performed their pointing tasks (see Figure 3(a)). Four to-be-located places were predefined. Point A was considered as a relatively close place to the robot, but a relatively distant place from the human operator. By contrast, Point B was closely located to the human operator, but far from the robot's current position. Both the human operator and the robot are closely placed to Point C; and Point D was the farthest one from both the robot and the human operator.

Procedure

The participants were first provided with the instructions regarding the experiment. These gave information about the experiment, the purpose of the study, and the data protection policy. The participants were then randomly assigned into one of the three different map representations. Before the main experiment, the participants were allowed to be familiar with the experimental setting, performing two trials. As the experimenter indicated a point tagged by a red sticker on the floor of the room, the participants were first asked to look at the to-be-located point on the floor, and in turn, they were allowed to pinpoint the point on the Tablet PC. The to-be-located points in these trials were not the same ones used in the main experiment. The procedure in the main experiment was the same with that of the pre-test.

Results

Table 1. Task performance in Experiment 1. (unit: cm)

Entities on map	Mean error distance (s.d.)			
	A	B	C	D
Robot, Human	12.44 (3.48)	13.36 (3.95)	10.40 (2.04)	22.39 (8.56)
Robot	14.29 (6.09)	16.70 (4.24)	12.13 (3.71)	21.71 (5.64)
Human	15.23 (3.97)	13.41 (4.26)	12.27 (3.32)	19.81 (4.85)
Total	**13.99 (4.64)**	**14.49 (4.31)**	**11.60 (3.12)**	**21.31 (6.41)**

Table 1 gives the pointing accuracy between the to-be-located places and the actual pointing on the map at each point. Comparing the mean error distance, it appeared to be dramatically reduced as the points were near to either the robot (Point A) or the human operator (Point B) or both (Point C). Point D, which was the farthest point from the both entities, resulted in the greatest error distance (mean 21.31). This task performance was assessed by a two-way (maps ×points) analysis of variance with repeated measures, revealing that there was no significant effect of the different map representation ($F_{2, 8} = 0.71$, n.s); but the to-be-located points significantly affected this task performance ($F_{3, 7} = 35.96$, $p<.01$).

Summary

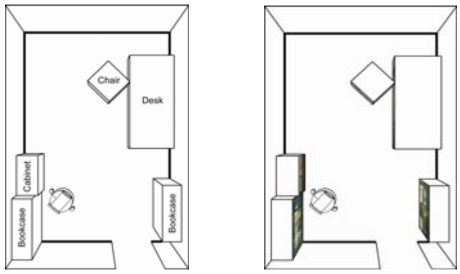
The main question concerned in Experiment 1 was whether all the collaborative participants in the limited space should be explicitly represented on the map. Two conclusions can be drawn from this experiment. Firstly, even in the situation where there was no information of the human operator position on the map the task performance was not deteriorated. This may be because the human operator seems to reason about their own location in terms of the visual angle and the relative distance from the robot that is locally perceived. Second, pointing accuracy relied heavily on the distance from the current position of the human operator or the robot. That is, if the destination of the robot was away from the robot or the human operator, the pointing task performance seemed to be very poor.

EXPERIMENT 2

Experiment 1 demonstrated, for example in the home context, the applicability of the map-based human-robot interface [e.g., 7, 8] in terms of task performance (max error distance 21.31 cm in 3.9×6 meters). In particular, it identified the location of the human operator was not necessary to be presented on the map.

In fact, the common home context has lots of objects, but Experiment 1 overlooked the existence of the objects in its setting. Indeed, it is known that navigational skills rely heavily on the identifications of the landmarks that are available at the time of interaction [13]. Experiment 2, therefore, intends to present these objects on the map, and investigates whether they have effects on task performance.

Here, as a way of presenting the objects in the environment, two characteristics were considered: *dimensionality*, and *legends*. The dimensionality issue of the map representation has been widely dealt with in vehicle navigation studies, as a way of extending situation awareness. For instance, Rate and Wickens [10] showed two dimension (2D) maps would be better to perform the lateral and vertical positioning over the three dimension (3D) counterpart, providing more accurate response. On the other hand, as being asked to report the current position on the map, most of users responded faster with 3D maps. The early studies on the vehicle navigation imply that the 2D map of the co-located environment would be of value for accurate lateral and vertical positioning of the robot. Yet, many robotics applications still prefer 3D modelling to the 2D model, thanks to the realism that it offers.



(a) (b)

Figure 6. The elevated 2D ($2_{1/2}$) maps (a) with the textual labels, (b) with pictorial images

To compromise these two approaches, the elevated 2D ($2_{1/2}$) model, which has been successful in Geological Information Systems (GISs), was considered. See Figure 6. The small-scale elevation of each object was determined by the relative size with the robot's stature (70 cm) in the map representation. The depth information that is presented on the elevated 2D map can convey information regarding the relative volume of each object.

The perception of the objects can also be extended by the explicit description of the object on the map. Of course, the objects would be locally perceived by the human operator in the co-located workspace. However, sometimes, in the situations in which the human operator cannot locally see the objects by some barriers, the explicit description of the object would be useful to specify locational reference terms for spatial interaction [9]. For this reason, two types of legends were considered: *textual label* and *pictorial image*. See Figure 6 again. The textual label of each object was added on the map as shown in Figure 6(a). The current vision technology can also manipulate photo images of each object, as shown in Figure 6(b). We assumed that the photo images were taken by an imaginary camera from the human operator's forward-field-of-view. In particular, the photo images were added on the face in which the human operator was supposed to see.

A within-subjects experimental design was developed, in which every participant serves under all combinations of both variables (dimensionality and legends). The six experimental treatments were formed by the three types of legends (none, text label, and picture) and the two types of dimensionality (2D and elevated 2D). Between the experimental treatments, the participants were asked to rate how much the map was easy to recognise the environment.

Method

Participants/Design

60 participants were recruited. The experimental design was a 2 (dimensionality) by 3 (legend) within-subjects design with three repeated measures. The participants experienced all the six combinations of both variables,

performing all the twelve to-be-located points that were predefined as shown in Figure 7. The performing sequence of the six experimental combinations was counterbalanced using a Latin square. The dependent variables, (i) the Euclidian error distance between the to-be-located point and the actual point; and (ii) a 7-point Likert scale rated by each participant as to how much each map representation was easy to recognise the environment where they were in, were used to assess the effects of the independent variables.

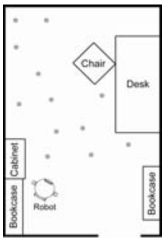


Figure 7. The twelve points used in Experiment 2

Apparatus/Procedure

The same apparatus with Experiment 1 was used, except the six different maps of the environment. The elevated 2D maps were designed by RHINOCEROS™. In addition, only the current position of the robot was presented on the map, given the results from Experiment 1. The procedure in Experiment 2 was almost the same with Experiment 1, except that the participants were asked to rate how much each map was easy to recognize the environment with a 7-point Likert scale between the experimental treatments.

Results

The first half of Table 2 gives the mean error distance in each experimental condition. Comparing the figures of each condition, it can be seen that there were no considerable difference in terms of legends. Whereas, a slight difference of the dimensionality was observed (20.13 in the 2D vs. 21.14 in the elevated 2D). The other half of Table 2 also gives the mean rates of map readability, showing that the different map representation would have extensive effects on the perception of the environment. Our participants reported that the elevated 2D (mean=5.02) would provide a better perception of the environment than the planar 2D (mean = 3.96), and also the textual label (mean=5.58) was

preferred to the others (mean = 3.64 in pictorial image, mean = 4.25 in no legend). A two-way analysis of variance with repeated measures was carried out for error distance. It was significantly decreased in the 2D maps over the elevated 2D ($F_{1, 59} = 5.28$, $p<.05$). However, it showed that our participants were little sensitive of the different legends of the objects on the map ($F_{2, 58} = 0.44$, n.s). The same analysis technique was applied to investigate the participant's readability ratings of each map representation, revealing that readability was significantly extended in the elevated 2D ($F_{1, 59} = 43.29$, $p<.01$), and text labels would best serve map readability ($F_{2, 58} = 82.39$, $p<.01$).

Summary

The main concern in Experiment 2 was whether the perception of the elements in the shared workspace could be extended through dimensionality and legends. Two conclusions can be drawn from this experiment. Firstly, in terms of map readability, the two features – dimensionality and legends – would have fair effects on the human operator's perception of the elements. Of course, most of the elements in the shared workspace might be locally recognized by the human operator. However, there are many situations in which the human operator cannot locally see some of the objects by environmental barriers such as interposition of the objects. In such cases, the explicit description and a more realistic model of the elements can help the human operator to recognize the environment. This result parallels many studies on vehicle navigation, which concluded that 3D maps could provide better perception of the elements. Interestingly, the pictorial image of the object was not preferred by our participants. It is probable that the image used in this experiment was not very high quality to identify the corresponding objects. Second, the dimensionality made considerable effects on task performance, concluding that 2D maps would provide more accurate response over the elevated 2D. This can be interpreted, as Rate and Wickens [10] claimed, by perceptual ambiguities of size and distance in the elevated 2D. In the 2D planar model, all the to-be-located places can be clearly recognizable in terms of the lateral and vertical position. On the other hand, the legends of the objects had little impact on task performance. This is partially because the objects used in this experiment could be locally viewed, so that the references to the objects were not necessary.

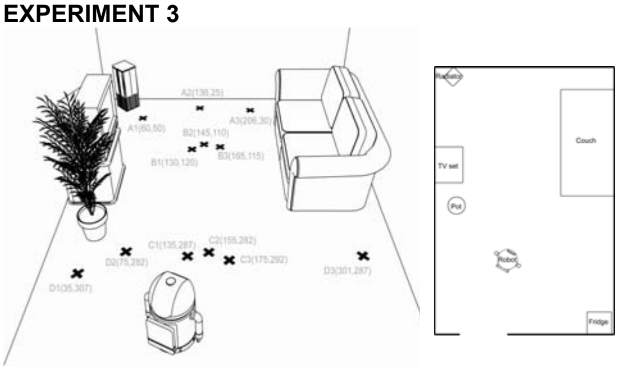
Table 2. Task performance and readability rating in Experiment 2

	Mean error distance (s.d) (unit: cm)				Mean rating (s.d) (1 – difficult, 7 – easy)			
	Text	Picture	None	**Total**	Text	Picture	None	**Total**
2D	20.15(10.86)	20.03(10.82)	20.21(10.35)	**20.13(10.66)**	5.13 (1.35)	3.22 (1.65)	3.52 (1.64)	**3.96 (1.76)**
Elevated 2D	21.18(13.67)	21.33(11.57)	20.91(11.44)	**21.14(12.25)**	6.02 (1.13)	4.07(1.57)	4.98 (1.42)	**5.02 (1.59)**
Total	**20.66(12.34)**	**20.68(11.21)**	**20.56(10.90)**	**20.64(11.49)**	**5.58 (1.24)**	**3.64 (1.61)**	**4.25 (1.53)**	**4.49 (1.68)**

This phenomenon was already identified by Colle et al. [1], coined it as *room effect*, maintaining that if the space is relatively small the users can easily afford local viewing in which objects are directly viewed irrespective of the representation of the space and the objects in the space.

Experiment 2 partially purported the current map-based interface for the co-located HRI, in the sense that most of the map-based interfaces [e.g., 7] employed the planar 2D map with the boundaries of the objects.Yet, Experiment 2 also indicated that textual labels would extend map readability, further its advantage would be obvious in which the robot should be located into a separate room or the human operator cannot locally see the objects. For this reason, a design implication – the 2D map with text label – was derived from the results of Experiment 2.

EXPERIMENT 3



(a) (b)

Figure 8. (a) the environment, (b) the map, for Experiment 3

Experiment 2 led a design implication – the 2D map with text labels. To explore its effectiveness, it is necessary to carry out an empirical study with respect to how much it can support the locating tasks. Experiment 3 directed to this issue. Consider Figure 8(a). The 12 to-be-located places were characterized by the distance from the landmarks. For instance, Point 'A1' was considered as a relatively closer point to an object, i.e., the radiator, and a relatively distant place from the current robot position. The criterion (2 meters) to specify the spatial relation from the robot was calculated by the fact that the robot was placed in 2 meters away from the bottom of the map, so that the remaining four meters out of 6 meters (room height) were divided into two categories, i.e., less than 2 meters is close and over 2 meters mean far. Likewise, the one meter criterion from the closest object was derived from the total 3.9 meters (room width), subtracting the width of the couch (1.2 meters) and the TV set (0.7 meters), and then dividing the remaining two meters into two sets, i.e., less than one meter is close and over 1 meter is far.

Method

Participants/Design

Twenty participants from Experiment 2 were invited to participate in Experiment 3. It was intended to form more homogeneous participants and reduce the experimental efforts without further training of the participants. The experimental design was a 2 by 2 within-subject design. Distance from the robot (close and far), and distance from the closest object (close and far) were served as independent variables. Twelve to-be-located points were predefined as shown in Figure 8(a). The sequence of the twelve trials by each participant was counterbalanced using a Latin square. The dependent variables, the Euclidian error distance and the completion time to locate a point on the Tablet PC, were used to assess the pointing task performance in each condition.

Apparatus/Procedure

The same apparatus with the previous experiments was also used here, except the different to-be-located points and map. The same procedure was used with the previous two experiments.

Results

The overall mean error distances and mean task completion time at each point were shown in Table 3. Two additional columns were added to help the reader to understand the spatial relation scheme used in this experiment. First, comparing the figures of error distance at every point, it can be seen that there was a consistent effect of the closeness from the objects. In both cases, i.e., Point As (mean 13.10) and Ds (mean 15.66), the mean error distances were almost half to the corresponding counterparts, i.e., Bs (mean 23.66) and Cs (mean 30.33), respectively.

Interestingly, the opposite effect of the distance from the robot was observed as our participants outperformed in the situations in which the points were distantly located from the robot (mean 13.10 in As, and mean 23.66 in Bs), rather than closely located to the robot, i.e., mean 15.66 in Ds, and mean 30.33 in Cs, respectively. These observations were further analyzed by a two-way analysis of variance with repeated measures of error distance. In terms of the spatial relation of the objects, error distance decreased significantly as the to-be-located places were closely placed to the objects ($F_{1, 59} = 6.75$, $p<.01$). Further, our participants provided significantly less error distance as the to-be-located places were distant from the robot ($F_{1, 59} = 95.97$, $p<.01$). It appeared to come from the fact that our participants were much careful to locate the robot into the places where they were far from the current robot position. The mean completion time of this task purported the interpretation given above. That is, as the to-be-located

Table 3. Task performance in Experiment 3

	Points	Mean Euclidian error distance (s.d) (unit: cm)	Mean completion time (s.d) (unit: secs)	Distance from the robot	Distance from the object
A	A1	9.30 (3.64)	9.85 (9.48)	Far	Close
	A2	16.14 (6.19)	11.26 (12.41)		
	A3	13.62 (5.97)	9.46 (9.31)		
	Total	**13.10 (3.84)**	**10.19 (10.40)**		
B	B1	27.39 (9.65)	10.11 (6.71)	Far	Far
	B2	19.89 (6.88)	11.49 (8.95)		
	B3	23.69 (8.42)	10.11 (7.02)		
	Total	**23.66 (8.31)**	**10.57 (7.56)**		
C	C1	30.74 (25.27)	5.32 (4.12)	Close	Far
	C2	29.44 (24.36)	9.37 (6.27)		
	C3	30.80 (25.07)	6.13 (4.31)		
	Total	**30.33 (24.90)**	**6.94 (4.90)**		
D	D1	15.01 (14.67)	5.73 (2.58)	Close	Close
	D2	18.05 (14.17)	5.91 (3.01)		
	D3	13.90 (6.35)	7.17 (4.43)		
	Total	**15.66 (11.73)**	**6.27 (3.34)**		

points were closer to the robot, i.e., Cs (mean = 6.94 secs.) and Ds (mean 6.27 secs.), the human operator could rapidly pinpoint the places than the other counterparts, i.e., Bs (mean = 10.57 secs.) and As (mean = 10.19 secs.), respectively. A two-way analysis of variance with repeated measures of task completion time revealed that the completion time were significantly affected by the distance of the to-be-located places from the robot ($F_{1, 59} = 12.80$, $p<.01$), not from the object ($F_{1, 59} = 0.58$, n.s).

SUMMARY

The conclusion to be drawn on from this experiment was spatial interaction would be highly affected by the fact that whether the to-be-located places could be referred by the surrounding elements on the map, as the human operators located the robot with the map-based interface. In fact, this result paralleled many psychological studies of spatial knowledge acquisition [13, 14]. In effect, the meaning of this experiment can be thought that it empirically supported that the map-based interface for HRI tasks should provide appropriate landmarks that are relevant to organize their HRI tasks, as that of vehicle navigation tasks.

SUMMARIES AND FUTURE WORKS

Taken together the three experiments demonstrated here that, first of all, the experiments empirically purported the applicability of the map-based human-robot interface, generating only around 5.4% error distance (0.21m/3.90m) with the setting used in the experiments. Secondly, the *room effect* was identified in the co-located HRI situation. That is, the human operator's local perception would marginally incorporate with the map interface. Finally, the distance from the surrounding elements was the most important factor to define task performance. None of these findings has been demonstrated empirically before.

The results demonstrated here raised three questions that could be pursued in the near future. Firstly, the results from these experiments lead us to consider how to improve the locating task performance for distantly located places from the elements, such as Point Bs and Cs in Experiment 3. Simply magnifying of the corresponding area of the map may overcome this limitation. Second, this paper considered the single room environment. However, the general home context has many rooms. It may lead to the different map design against what this paper had concluded. Finally, the situations in which the human operator is mobile are worth mentioning. If the human operator is mobile, the congruence issue between the human operator's local perception and the map representation is inevitable. Indeed, this congruence issue has been well investigated in vehicle navigation studies, but not in the map-based HRI sphere. These issues are currently investigating with a different experiment setting.

REFERENCES

[1] Colle, H.A. and G.B. Reid, Spatial orientation in 3D desktop displays: using rooms for organizing information. Human Factors, 45(3), 2003, 424-439.

[2] Darken, R. and H. Cevik. Map usage in virtual environment: Orientation issues. In Proceedings of the Virtual Reality, (VR1999) (Huston, Texas, 1999), IEEE Computer Society Press, 133-140.

[3] Drury, J.L., J. Scholtz, and H.A. Yanco. Awareness in Human-Robot Interactions. In Proceedings of IEEE Conference on SMC, 2003, 912-918.

[4] Fong, T., Collaborative control: a robot-centric model for vehicle teleoperation, Ph.D. Thesis, Carnegie Mellon University, 2001.

[5] Forlizzi, J., Robotic products to assist the aging population, ACM Interactions (March-April, 2005). 16-18.

[6] Gutwin, C., G. Stark, and S. Greenberg. Support for workspace awareness in educational groupware. In Proceedings of Computer Supported Collaborative Learning, 1995.

[7] Huttenrauch, H. and M. Norman. PocketCERO-mobile interfaces fro service robots. In Proceedings of the Mobile HCI, 2001.

[8] Perzanowski, D., et al., Building a multimodal Human-Robot interface. IEEE Intelligent Systems (Jan/Feb, 2001), 16-21.

[9] Poole, P.E. and C.D. Wickens, Frames of reference for electronic map displays: their effect on local guidance and global situation awareness during low altitude rotorcraft operations. University of Illinois Institute of Aviation, 1998

[10] Rate, C. and C.D. Wickens, Map dimensionality and frame of reference for terminal area navigation display: where do we go from here? Technical Report, University Illinois Institutive of Aviation, 1993

[11] Scholtz, J., Have robots, need interaction with humans! In ACM Interactions (March-April. 2005), 13-14.

[12] Skubic, M., et al. Hand-drawn maps for robot navigation. In Proceedings of AAAI 2002 Spring Symposium, 2002.

[13] Thorndyke, P. and B. Hayes-Roth, Differences in spatial knowledge obtained from maps and navigation. Cognitive Psychology, 14, 1982, 560-589.

[14] Wickens, C.D., Frames of reference for navigation, In Attention and Performance XVII, D. Gopher and A. Koriat, (Ed.)., MIT Press, MA, 1999, 112-144.

[15] Yanco, H.A. and J.L. Drury. Classifying Human-Robot interaction: An updated taxonomy. In Proceedings of IEEE Conference of SMC, 2004, 2841-2846.

[16] Yanco, H.A., J.L. Drury, and J. Scholtz, Beyond usability evaluation: Analysis of human-robot interaction at a major robotics competition. Human Computer Interaction, 19, 2004, 117-149.