

Human-Driven Spatial Language for Human-Robot Interaction

Marjorie Skubic and Zhiyu Huo
Electrical and Computer Engineering Dept.
University of Missouri

Laura Carlson, Xiao Ou Li and Jared Miller
Dept. of Psychology
University of Notre Dame

Abstract

This extended abstract outlines a new study that investigates spatial language for use in human-robot communication. The scenario studied is a home setting in which the elderly resident has misplaced an object, such as eyeglasses, and the robot will help the resident find the object. We present preliminary results from the initial study in which we investigate spatial language generated to a human addressee or a robot addressee in a virtual environment.

Introduction

When people communicate with each other about spatially oriented tasks, they more often use qualitative spatial references rather than precise quantitative terms, for example, “*Your eyeglasses are behind the lamp on the table to the left of the bed in the bedroom*”, (Carlson and Hill, 2009). Although natural for people, such qualitative references are problematic for robots that “think” in terms of mathematical expressions and numbers. Yet, providing robots with the ability to understand and communicate with these spatial references has great potential for creating a more natural interface mechanism for robot users. This would allow users to interact with a robot much as they would with another human, and is especially critical if robots are to provide assistive capabilities in unstructured environments occupied by people, for example in an eldercare scenario.

In a new collaborative project between the University of Missouri and the University of Notre Dame, we are investigating spatial language for robots within the context of helping older adults find lost items in the home, such as a book, a coffee mug, or eyeglasses. In the scenario, the older adult has misplaced an item. The robot knows the locations of important objects and can provide spatial descriptions. The elderly resident then moves through the house to retrieve the objects. The scenario is motivated by the observation that elders have increased difficulty with memory and eyesight, leading to difficulty in finding objects whose locations typically change (e.g., eyeglasses).

In addition, elders have increased difficulty with locomotion, but physical movement and exercise are known to facilitate cognitive function (Kramer et al., 2005). Thus, the robot can be assistive by conveying the locations of the objects so that the elderly resident can navigate an efficient path to retrieve them. This requires algorithms so that the robot can produce appropriate spatial descriptions. We are also studying a related scenario in which the elderly resident does not retrieve the objects, but provides descriptions to a robot that navigates a path to retrieve the objects. This requires algorithms for the robot to comprehend spatial descriptions.

As a first step, a series of human subject experiments is being conducted to study the preferences for spatial language in human to human communication, as well as in human to robot communication. Here, we present an overview of the experiments and report preliminary results. Currently, 21 out of 128 college-age participants have completed the study. This summer, we will begin a study with older adults. Based on the results of these human subject experiments, algorithms will be developed for robot spatial language, similar to prior work (Skubic et al., 2003, 2004), both in the language being generated and in the ability to understand spatial language terminology. The human-driven spatial language algorithms will ultimately be tested in future human subject experiments using both virtual and physical robots. Thus, at the end of the project, we will have captured spatial language results comparing college age students vs. older adults, human partners vs. robot partners, and virtual partners vs. physical partners.

Human Subject Experiments

The initial human subject experiments with college students are being conducted in a virtual setting. We are first investigating the type of spatial language that participants use intuitively in addressing either a human avatar or a robot avatar. Figure 1 shows part of the environment, as well as the robot avatar used in the study. The use of a virtual environment provides a better controlled setting and is easier for capturing potentially subtle metrics between different test conditions. In pilot work, the use of a virtual environment was shown to have



Fig. 1. The virtual scene used for the human subject experiments, showing the robot avatar in the hallway with the living room on the left and the bedroom on the right. The robot and virtual environment are modeled on the physical robot and environment in which future experiments will be conducted. The Microsoft Kinect will be used for perception on the robot.

sufficient sensitivity to detect differences and was also shown to replicate key findings from work done in physical environments, such as (Schober, 1995).

In this first set of experiments, each participant begins in the hallway (see Fig. 1) and is provided some time to explore the environment and note the room layout and the locations of various objects in the scene; at this point, candidate reference objects are shown but no target objects are included. The participant then returns to the hallway and is given a target object to find in the scene. The participant again explores the scene, this time looking for the location of the specified target object, which has now been placed in the environment. Eight target objects are used in the study, including a book, cell phone, eyeglasses case, keys, letter, mug, notepad, and wallet; each participant has eight trials, one for each target object. After looking for the target object, the participant once again returns to the hallway and is asked to give instructions to the avatar as to the location of the target object. Test conditions include the type of addressee (human or robot), alignment with respect to the addressee (either aligned at 0 degrees or face to face, i.e., 180 degrees), and a request as to the type of communications (either tell the addressee WHERE to find it or HOW to find it).

The environment has typical furniture found in living room and bedroom settings, as well as a number of other objects that could serve as potential reference objects within a spatial language description, including flowers, a vase, alarm clock, picture frame, tray, purse, tissue box,

lamp, fan, candle, statue, computer monitor, laptop, hat, game box, and plant. The goal is to provide appropriate furniture and interesting objects within each room so the participants can use a comfortable form of spatial language.

The spatial language used by the participants is coded as follows:

- Perspective taken (self or addressee)
- Type of description: hierarchical (e.g., in the bedroom on the table behind the lamp), beacon (by the plant), or mixed
- Number of spatial phrases
- Sequence of the types of spatial phrases
- Type of sequence order (Ascending, small to big, e.g., on the table in the bedroom; or Descending, big to small, e.g., in the bedroom on the table)
- Reference object selected
- Type of spatial term
- Use of spatial hedges (e.g., kind of near the middle of the room)

Preliminary Results

Here, we report preliminary results on 21 test subjects; 7 participants were given a robot addressee and 14 were given a human addressee. Because this is a small sampling of the final participant count and unbalanced as to the type

of addressee, it is not clear whether these data will hold. We will not claim significant differences until the complete results have been studied.

Tables 1-3 show the preliminary results. In each case, we report the results of the 0 degree and 180 degree alignment for participants interacting with the human addressee and the robot addressee. Table 1 shows the averages of using a self perspective as opposed to using the addressee’s perspective. For example, when addressing the human face to face (180 degree alignment), a self perspective was taken about 18% of the time, whereas the addressee’s perspective was taken about 82% of the time. When facing the robot, the self perspective was taken about 39% of the time, and the robot’s perspective was taken about 61% of the time. Although there are differences in the various test conditions, the addressee’s perspective was taken in the majority of the spatial language instances. This is consistent with other related work in HRI studies (Tenbrink, Fischer & Moratz, 2002). Note that this coding ignores ambiguous statements in which the perspective is unclear, which is common in the 0 degree alignment cases. Under this condition, it is possible that an assumed perspective is taken.

Table 1. Self Perspective (average over all trials)

Addressee	0 deg align	180 deg align
Human	42.9	18.1
Robot	31.4	38.8

Table 2 shows the results of descending vs. ascending descriptions. In most but not all of the cases addressing the human, the descriptions were descending, i.e., big to small, such as in the bedroom on the table next to the lamp. However, in the case of the robot addressee, all of the descriptions were descending. This includes trials asking WHERE to find the target object as well as HOW to find the object. This is contrary to the expected results. One would expect that the how question would result in descending descriptions modeling spatial directions or navigation commands (e.g., go to the bedroom, find the table, look next to the lamp on the table). However, we expected the where question to result in ascending spatial descriptions of small to big, such as the object is next to the lamp on the table in the bedroom.

Table 2. Percentage of Descending Descriptions (average over subject’s percentages)

Addressee	0 deg align	180 deg align
Human	85.7	92.9
Robot	100.0	100.0

The use of candidate reference objects is shown in Table 3. The interesting result here is the much higher use

of reference objects when the participant and addressee are aligned, compared to the face to face speaking. We will be watching whether this holds after the study is completed.

Table 3. Use of Candidate Reference Objects (sum over all trials)

Addressee	0 deg align	180 deg align
Human	14	1
Robot	15	4

Conclusions

Preliminary results were presented on a new collaborative study investigating preferences for spatial language in human-robot communication. We are particularly interested in identifying differences between human to human language vs. human to robot language. In the initial study, college age students are used as participants; in the later months, older adults will be recruited for the study. We will also look for differences and similarities between the two target age groups.

Acknowledgements

This work is funded by the National Science Foundation, grant# IIS-1017097.

References

- Carlson, L. A. & Hill, P. L. (2009). Formulating spatial descriptions across various dialogue contexts. In K. Coventry, T. Tenbrink, & J. Bateman (Eds.), *Spatial Language and Dialogue* (pp. 88-103). New York, NY: Oxford University Press Inc.
- Kramer AF, Colcombe SJ, McAuley E, Scalf PE, Erickson KI. (2005). Fitness, aging and neurocognitive function. *Neurobiol Aging*. Dec 26 Suppl 1:124-7.
- Skubic, M., P. Matsakis, G. Chronis, and J. Keller. (2003). Generating multilevel linguistic spatial descriptions from range sensor readings using the histogram of forces, *Autonomous Robots*, vol. 14, no. 1, 2003.
- Skubic, M., D. Perzanowski, S. Blisard, A. Schultz, W. Adams, M. Bugajska, and D. Brock. (2004). Spatial language for human-robot dialogs, *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, vol. Special Issue on Human-Robot Interaction, pp. 154-167, 2004.
- Schober, M.F. (1995). Speakers, addressees and frames of reference: Whose effort is minimized in conversations about locations? *Discourse Processes*, 20, 219-247.
- Tenbrink, T., K. Fischer, and R. Moratz, (2002). Spatial strategies of human-robot communication, KI, no. 4, 2002. [Online]. Available: iteseer.ist.psu.edu/tenbrink02spacial.html