Spatial Language Experiments for a Robot Fetch Task

Marjorie Skubic[1], Laura Carlson[2], Jared Miller[2], Xiao Ou Li[2], & Zhiyu Huo[1]

[1]Electrical & Computer Engineering
University of Missouri
Columbia, MO 65211
1-573-882-7766

SkubicM@missouri.edu;
zhiyuhuo@mail.mizzou.edu

[2]Dept. of Psychology
University of Notre Dame
Notre Dame, IN 46556
1-574-631-8883

LCarlson@nd.edu;
jmille39@nd.edu; xiaoou_li@nd.edu

ABSTRACT
This paper 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 results from phase I of the study in which we investigate spatial language generated to a human addressee or a robot addressee in a virtual environment.

Categories and Subject Descriptors
H.5.2 [**Information Systems**]: User Interfaces – *Voice I/O.*

General Terms
Experimentation, Human Factors.

Keywords
Robots for elderly, Spatial language.

1. INTRODUCTION
Recent studies have shown that one of the top five tasks noted by seniors for assistive robots is help with fetching objects, for example, retrieving missing eyeglasses [1]. In addition, the most preferred domestic robot interface is natural language [2]. In this paper, we present the overview and initial results for a project designed to address the fetch task and study an appropriate language interface that allows the human users to communicate naturally and effectively with a robot.

When people communicate with each other about spatially oriented tasks, they typically use relative spatial references rather than precise quantitative terms, e.g., *the eyeglasses are in the livingroom on the table in front of the couch* [3]. Here, we explore the spatial language used in the context of the fetch task and investigate both language generation and understanding. A set of human subject experiments is planned, studying first college-age students and then adults over age 65 for comparison. Initial experiments are being conducted in a virtual environment (VE), which provides a controlled setting and is easier for capturing potentially subtle metrics between test conditions. In pilot work, the use of a VE was shown to have sufficient sensitivity to detect differences and also replicated key findings from work done in physical environments, e.g.,[4]. Later studies are planned with robots in the physical world.

HRI'12, March 5–8, 2012, Boston, Massachusetts, USA.
ACM 978-1-4503-1063-5/12/03.

2. EXPERIMENT PROTOCOL
The first round of human subject experiments has been completed with 128 college-age students. We investigated the type of spatial language used intuitively by participants when addressing either a human or a robot avatar. The VE included three rooms – a central hallway with a livingroom and a bedroom (Figure 1).

Each participant begins with a brief video illustrating the room layouts. At this point, candidate reference objects are shown but no target objects are included in the scene. The participant is then asked to explore the scene, to look for a specified target object which is now included in the VE. Eight target objects are used for the study: a book, cell phone, eyeglasses case, keys, letter, mug, notepad, and wallet; each participant has eight trials, one for each target object. After locating the target object, the participant is brought back to the hallway and is asked to give instructions to the avatar on the location of the target object. There are three test conditions: (1) the addressee, either human or robot, (2) the subject alignment with respect to the addressee, either aligned at 0 degrees or face to face, i.e., misaligned at 180 degrees, and (3) the instruction on the type of communications, either tell the addressee WHERE to find the target object or HOW to find it. The descriptions given by the participants were recorded, transcribed, and coded for analysis.

Figure 1. The virtual scene used for the experiments, showing the robot avatar in the hallway with the livingroom on the left and the bedroom on the right.

3. RESULTS
The more prominent results are summarized in Fig. 2-5. Participants had a preference for taking the addressee's perspective in all test conditions (Fig. 2), which agrees with prior HRI studies e.g., in [5]. Participants also had a strong preference for using descending descriptions (Fig. 3), that is, from outside to inside the room, e.g., in the bedroom on the table.

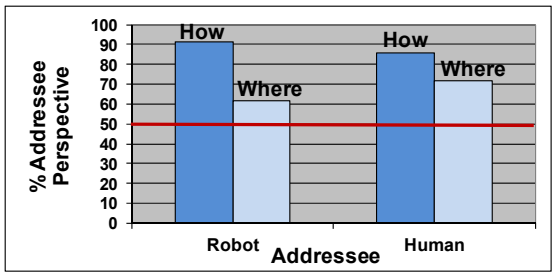

Figure 2. Percent of the time subjects used the addressee's perspective.

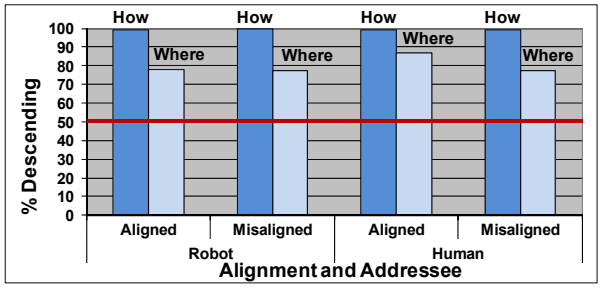

Figure 3. Percent of the time subjects used descending descriptions (outside to inside room).

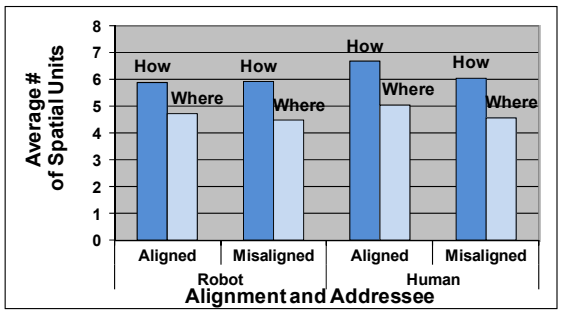

Figure 4. Average number of spatial units per description.

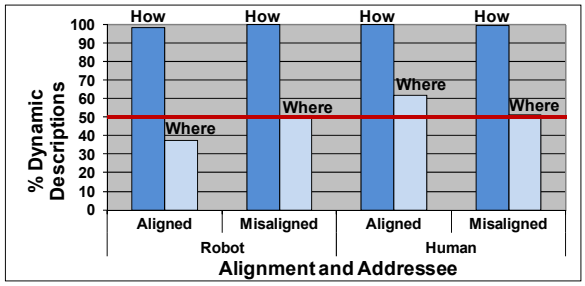

Figure 5. Percent of the descriptions that resemble step by step instructions to the target object, as opposed to static location descriptions.

We tested for differences between test conditions and found the only statistically significant differences were between the HOW and WHERE conditions. Fig. 2-5 all illustrate these differences. The HOW descriptions emphasized the preferences for addressee perspective and descending descriptions. In addition, participants used more spatial phrases on average for the HOW descriptions compared to the WHERE descriptions (Fig. 4). We also looked at whether subjects used static location descriptions (e.g., the book is in the bedroom on the table next to the bed) vs. dynamic descriptions that provide step by step instructions (e.g., go to the bedroom, move to the left side of the bed, look on the table and find the book). Not surprisingly, participants showed a very strong preference for these step by step instructions when asked to provide a HOW description (Fig. 5). It is perhaps more surprising that participants also used these step by step instructions about half the time for the WHERE descriptions.

There were no statistically significant differences between the robot and human addressees. The figures show some differences but none of these are statistically significant. It is possible that these college age subjects were viewing the human avatar in the VE as an anthropomorphic robot; we did not test for this. We anticipate that we may find differences with the older adults.

The most used spatial terms include on, to, right, and left. Top reference items include the furniture pieces (especially the table, bed, and desk) and house structure units (room, wall). Very few references were made to smaller objects lying on the table and desk next to the target objects.

3. CONCLUSIONS

The results of the experiments confirm the use of the addressee's perspective and show a preference for descending descriptions, which will be important in supporting a robot interface. They also show that the spatial descriptions depend on the task instructions given. The HOW descriptions convey step by step motion, and use many spatial units; the WHERE descriptions have no preference for conveying motion and use fewer spatial units. We are in the process of repeating the experiments with adults over 65 and will be especially looking for differences in the two age groups.

The results also highlight some of the challenges in implementing a robot fetch task in the physical world. We will need to recognize furniture items such as tables even when they are cluttered with objects, and support relative spatial terms with respect to the furniture. There are also significant challenges in speech recognition and natural language processing to translate the spatial descriptions into robot commands.

4. ACKNOWLEDGMENTS

This work is funded by the NSF under grant IIS-1017097.

5. REFERENCES

[1] Mitzner T., Smarr, C., Beer, J., Chen, T., Springman, J., Prakash, A., Kemp, C. and Rogers, W. 2011. Older Adults' Acceptance of Assistive Robots for the Home. Technical Report HFA-TR-1105. Georgia Tech Univ.

[2] Scopelliti, M., Giuliani, M., and Fornara, F. 2005. Robots in a domestic setting: a psychological approach. *Universal Access in the Information Society*, 4(2), 146-155.

[3] Carlson, L. A. and 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.

[4] Schober, M.F. 1995. Speakers, addressees and frames of reference: Whose effort is minimized in conversations about locations? *Discourse Processes*, 20, 219-247.

[5] Moratz, R., Fischer, K. and Tenbrink, T. 2001. Cognitive Modeling of Spatial Reference for Human-Robot Interaction, *Intl. J. on Artificial Intelligence Tools*, 10(4), 589-611.