2013 IEEE RO-MAN: The 22nd IEEE International Symposium on
Robot and Human Interactive Communication
Gyeongju, Korea, August 26-29, 2013

Testing an Assistive Fetch Robot with Spatial Language from Older and Younger Adults

Marjorie Skubic, Zhiyu Huo, Tatiana Alexenko, Laura Carlson, and Jared Miller

Abstract—**Methods and experimental results are presented for interpreting 3D spatial language descriptions used for human to robot communication in a fetch task. The work is based on human subject experiments in which spatial language descriptions were logged from younger and older adult participants. A spatial language model is proposed, and methods are presented for translating natural spatial language descriptions into robot commands that allow the robot to find the requested object. Robot command representation and robot behavior are also discussed. Experimental results compare path metrics of the robot system and human subjects in a common simulation environment. The overall success rate of the robot trials is 85%.**

I. INTRODUCTION

A growing elderly population and shortages of healthcare staff have created a need for autonomous agents capable of performing assistive tasks. Recent studies [1] have shown that seniors want assistive robots capable of performing household tasks, such as the "fetch" task. Older adults also prefer a natural language interface [2] when interacting with robots. In this paper, we present methods and experimental results for interpreting 3D spatial language descriptions used in a fetch task.

Robot understanding of spatial language has been explored previously. Much of the work has focused on 2D navigation. For example, Gribble et al [3] proposed using the Spatial Semantic Hierarchy [4] to represent and reason about space, using commands such as “go there” and “turn right”. There is a body of work on understanding 2D route instructions for guiding an agent or robot through an environment [5][6][7][8]. Tellex et al. also consider manipulative commands that move beyond the 2D ground plane, e.g., “*put the pallet on the truck*” [8]. This work has informed our project; however, it is focused on the more general natural language processing (NLP) problem and is limited in addressing the perceptual and cognitive challenges of the fetch task.

What separates our work from much of the related research is: (1) the integrated perceptual capabilities of our robot, (2) our support of 3D spatial relationships vs. the 2D representations others use, and (3) our human-centric approach. The perceptual capabilities of our robot, which include furniture and object recognition, bring it closer to being able to interpret language in a "human" way. The complex language with 3D spatial relationships warrants NLP and inference processes that model those of a human and are rooted in existing cognitive science research.

Our goal is to interpret language used in the context of human-robot interaction. Natural, unconstrained language is notorious for a lack of punctuation, numerous stop-words, unnecessary repetitions and other elements that create noise and hinder interpretation. In this paper, we soften the noise by first exploring the use of templates created from transcribed spoken spatial language logged in studies with older and younger adults.

We discuss the basis for these templates in Section II. Methods for interpreting the spatial language are presented in Section III. The human language model and our robot methods for interpreting the language have been designed for real spatial language as well as the template descriptions tested here; our proposed approach will be tested on "noisy" natural spatial language in the next phase of the project.

In Section IV, an experiment is presented that compares the performance of humans with a robot in following spatial descriptions to fetch a target object. Various path metrics are introduced for comparing the performance. Results and discussion are included. We analyze some of the failures and discuss strategies that will be addressed in future work.

II. SPATIAL LANGUAGE MODEL

A. Templates Derived from a Spatial Language Corpus

In our previous work, we created a corpus of 1024 spatial language descriptions collected in a human subject experiment [15][16]. In the experiment, which was conducted in a virtual environment, 64 younger adult participants and 64 older adult participants were asked to tell a human or robot avatar *where* a given object was or *how* to get to it. The experiment was based on the idea of a user explaining to an assistant where or how to find an object for the purpose of fetching the item. Test manipulations included how vs. where instructions, robot vs. human addressee, and older adult vs. younger adult subjects.

The resulting spatial language was analyzed, and templates were created to capture language structure that was common to the spatial language descriptions logged for each test manipulation. Then, using these templates, representative spatial descriptions were generated and used for the experiment presented in this paper. The templates also played a major part in the creation of our human spatial language model.

The general structure of the templates was determined by examining syntactic differences across age (older/younger adults), instruction (how/where) and addressee (human/robot) manipulations. A significant difference that emerged was a function of instruction [18]. *How* descriptions were overwhelmingly dynamic, following a sequential, direction-like structure such as [Move] +

[Direction] + [Move] + [Direction] + [Goal]. For example, "Go forward, turn left, go straight, and you'll find the target. *Where* descriptions were more split, with a significant number of static descriptions, following a structure such as [Target] + [In] + [Room] + [Room Reference]. For example, "The book was in the living room, against the back wall." Result details can be found in [16].

To capture linguistic differences between age and addressee groups, word counts for spatial terms, house items and furniture items were calculated. Differences in the occurrence of given words for a given category (e.g. younger vs. older) were considered of importance when they occurred a minimum of ten times, and twice as often as a competing category. For example, older adults were more than twice as likely to use "door" and "take" in their descriptions than were younger adults. These word count differences were used to modify the template structure defined by the instruction manipulation described above. For example, the younger adult template in the How-Robot condition contained the following description for the location of the glasses case, "Walk straight and take a right. Go forward and turn right and you'll find the glasses case". Older adult templates in the How-Robot condition contained the following description for the glasses case, "Take a right through the door. Go forward and turn right and you'll find the glasses case." Thus, the templates capture the differences observed in the different test manipulations, although some differences are subtle.

The templates were also generated for different landmark conditions. The *No-Landmark* templates were unaltered. *Goal-Landmark* templates included a description of the table where the target object was located. For example, the Older-How-Robot description for the glasses case would read, "Take a right through the door. Go forward and turn right and you'll find the glasses case on the table." *Path-landmark* templates included a description of a furniture item in the environment in addition to the table where the target was located. For example, the Older-How-Robot description for the glasses case would read, "Take a right through the door. Go forward and turn right and you'll find the glasses case on the table behind the couch." In all, across all categories, there were 149 unique templates. There was some repetition across categories because of a lack of a meaningful difference between word counts. Because of this repetition, we examine the templates subsequently simply as a function of how/where and as a function of landmark type (none, goal, path). The other differences that emerged from the older vs. younger and robot vs. human addressee manipulations in the original study were not expected to be reflected in the path metrics used for comparing robot and human performance, due to the very subtle differences.

B. Semantic Chunks

The spatial language templates, derived from the spatial language descriptions observed in the fetch task, drove the development of our human spatial language model. The templates are essentially containers for certain components commonly found in spatial language. By identifying these components, we created a method for segmentation (or chunking) of spatial language. The lengthy and often complex spatial instructions can be broken up into smaller, meaningful parts (or chunks). Furthermore, these chunks can be nested to capture relations between them.

Table I shows the chunks which constitute the proposed human spatial language model. Figure 1 shows an example of one of the spatial descriptions generated from a template that has been chunked using the model. Table I shows chunk types that include perspectives of both outside and inside the target room. The parts of each description are first differentiated based on whether or not the addressee is outside or inside the room at the time of execution. The first two chunk types are meant to provide information outside the room and determine room choice. The rest of the chunk types are meant to be interpreted within the room. This separation has roots in the works of Radvansky [17][18], who showed that humans appear to unload old and load new cognitive maps whenever they pass through a boundary between two enclosures, such as doorways.

Next is the separation between targets and references. In the previous section, these were referred to as goal and path types. Each of the chunk types is either a target or a reference type. Targets indicate the goal states, which, in this case, are in the correct room near the furniture item with the target object. References are states that help the executor of the description *achieve* the goal state, for example, references to furniture items that are near the goal or descriptions of the path to the goal.

Finally we further separated within-room instructions into three categories: room regions, furniture items, and small objects. This separation was motivated by both the human tendency to create hierarchies for objects and spaces based on their size, as well as the perceptual capabilities of our robot. Our robot can recognize furniture items and small objects; however, this recognition becomes more accurate when the robot knows whether it is looking for a small object or a furniture item. For parts of the description that refer to fixed walls or other structures in the room, the robot relies on a map of the environment rather than recognition capabilities. Therefore it becomes necessary, when interpreting spatial descriptions, for the robot to know when it needs to switch between different recognition modes. Also, larger furniture items such as a bed are less mobile and are more likely to stay in place, whereas small objects can be moved within and between rooms; over time, the robot could learn how to take advantage of such tendencies to improve the efficiency of the fetch.

C. Static and Dynamic Differences

The proposed human spatial language model is suitable for both static and dynamic spatial instructions, unlike related work [10][11][12], which concentrated on sequential dynamic instructions. Our spatial language corpus shows that static spatial language is often not sequential and many of the static spatial language instructions we collected started with the Object Target Phrase (OBTP) rather than information about the room in which the object is located. This was particularly prevalent in language used by older adults [16]. Because we use semantic chunk labels, we can

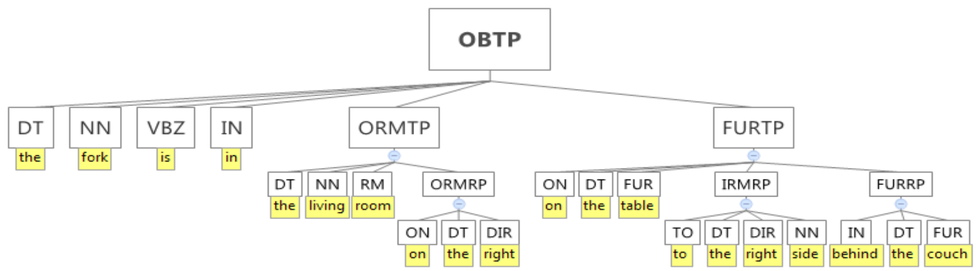

Figure 1. An instance of a chunked spatial description. Chunk types are shown in Table 1.

rearrange the static descriptions into sequential order for ease in translation into robot commands. Methods proposed in [10], [11] and [12] which focused on inherently sequential directions would not be suitable for static descriptions where order is often arbitrary or reversed.

TABLE I. CHUNK TYPES

Chunk Type	Explanation
ORMTP	Outside Room Target Phrase
ORMRP	Outside Room Reference Phrase
FURTP	Furniture Target Phrase
FURRP	Furniture Reference Phrase
OBTP	Object Target Phrase
OBRP	Object Reference Phrase
IRMRP	Inside Room Reference Phrase

III. Interpreting Spatial Language

A. Modeling Spatial Relationships

When people communicate with each other about spatially oriented tasks, they typically choose relative spatial references rather than precise quantitative terms, e.g., *the eyeglasses are in the living room on the table in front of the couch* [20]. Although natural for people, it is not easy for a robot to follow such a description. Providing robots with the ability to understand and communicate with these spatial references has great potential for creating a more natural approach for human-robot interaction [13]. In previous work [21], we used the histogram of forces (HoF) [22] to model spatial relationships and, thus, provide a method for interpreting spatial language references in human-robot interaction. The HoF can quantize the spatial relationship between two crisp or fuzzy objects by providing weights of different directions [22]. By providing a quantitative model of these relationships, the HoF can be used to translate qualitative spatial relationships into robot instructions.

B. Modeling the Fetch Task

The environment of the fetch task investigated here is a two-room home with a hallway between the rooms. The robot stands at the end of the hallway to wait for instruction before starting the task. To simplify the fetch task, we divide the process into three sub-tasks: (1) determine the target room and move to enter the room through the doorway, (2) move within the room to the place where the target object is located by following the spatial description, (3) search for the object around the goal location as specified in the spatial description. In the fetch task, the target objects are assumed to be on the surface of furniture items so that the robot does not need to search inside the furniture. The robot uses its local perception for navigation in this task.

C. Reference-Direction-Target Model

Because the robot has no prior information about the furniture and object placement inside the room, it needs to use the information provided by the spatial language description. Therefore, we propose a Reference-Direction-Target (RDT) model to translate the spatial description into navigation information that can be used directly as a navigation command for the robot. The RDT model includes the three parts of Reference, Direction and Target, which together comprise a RDT node. The three RDT components represent all types of navigation instructions a robot may need in an indoor environment. We assume that the robot has a map of the environment structure, although we do not assume the robot knows where the furniture items are located within the rooms of the structure.

<u>Reference</u> refers to the object or structure that is used in a relation. Several types of references are used in the fetch task, as described below.

NONE – No reference object is mentioned in the instruction, i.e., the robot action is not dependent on the objects around the robot. For example, "*turn right*" or "*go forward*" are instructions with no reference.

ROOM – The room is used as a reference for navigation, e.g., *"move halfway in*" or "*to the left of the room*". The Direction component then determines which part of the room is the destination. Using a sense of direction, e.g. from a compass, and prior knowledge of the room structure, the robot can move to the target area and search for the target object.

WALL – A wall is used as the reference, e.g., "*to the back wall*". The robot navigates close to a wall and may search for the target object.

ROBOT – The robot itself is used as the reference. The reference object does not directly appear in the description, but rather ego-centric references are used, e.g., "*to the left*" or "*in front of you*". These mean "to the left of the robot" or "in front of the robot" which uses the robot's local reference frame.

FURNITURE – A furniture item is used as the reference object. The reference frame that defines the direction differs for different types of furniture. These have been defined based on the results of our spatial language experiments. For example, "in front of the couch" is typically defined using the intrinsic frame of the couch. The front side refers to the seating side of the couch independent of viewing angle.

Direction represents the position relationship between objects and tells the robot where it should move to search for the target. For the different references described above, the meaning of *direction* is different. For NONE, the direction tells the robot the angle for motion with respect to the robot's local reference frame. For other reference types, direction shows where the robot should move, relative to the specified reference. For different types of navigation instructions, the reference frame for direction may be defined differently [14]. The directions used in robot fetch commands include: front, left, right, back side, and between. Table II shows the combinations of references and their corresponding directions.

Target indicates the target furniture in the instruction or the reference of the target object. If there is not a target in a RDT node, the target is defaulted to be a table type furniture item. This is a natural assumption, in the context of the fetch task.

TABLE II. REFERENCES AND CORRESPONDING DIRECTIONS

Reference	Category	Corresponding Direction
NONE	Dynamic	Front, Left, Right
ROOM	Dynamic	Left, Right, Back
WALL	Dynamic	Left, Right, Back, Side
ROBOT	Static	Front, Left, Right, Back
FURNITURE	Static	Front, Left, Right, Back, Between

D. Translating Chunks into Navigation Commands

For the fetch task, we manually built a dictionary of spatial language phrases for translating the words and phrases in the chunks to navigation commands that can be understood by the robot. The knowledge to build this dictionary is based on our human-robot spatial language experiments [15][18]. From the 3 sub-tasks described above, the information also has three parts: (1) target room, (2) inside-room navigation command, and (3) target object. They can be extracted by searching the words, phrases and their corresponding tags in the chunks from the dictionary of spatial language phrases. In the fetch task, the target room is extracted directly from the ORMTP and ORMRP chunks, and target object is extracted from the OBTP chunk. FURRP chunks, FURTP chunks and IRMRP chunks provide navigation instructions within rooms.

The translation is a traversal process along the leaves of the parse tree. For the example shown in Figure 1, we convert the parse tree to a robot behavior model by 3 steps.

1. Preorder traverse the parse tree. List the phrases of the corresponding chunks sequentially. The phrases are: (1) OBTP: "*the fork is in*", (2) ORMTP: "*the living room*", (3) ORMRP: "*on the right*", (4) FURTP: "*on the table*", (5) IRMRP: "*to the right side*". (6) FURRP: "*behind the couch*".

2. Extract room information and target object information from ORMTP, ORMRP and OBTP using the dictionary. The room is *living room* and the target object is the *fork*.

3. Generate navigation instructions by building the RDT nodes. The result is "robot-right-table". In a complex command, there may be more than one phrase that can be translated to a RDT node. Connect them sequentially to build a RDT chain (Fig. 2).

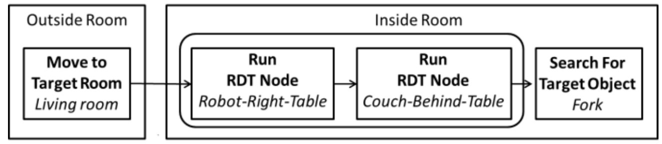

Figure 2. RDT Chain Model for the spatial description in Fig. 1.

E. Robot Behavior Model

After translating the spatial descriptions into robot commands, the robot behavior model can be instantiated, and the robot is then ready to execute the command. The robot behavior model has a two-tier structure. The higher tier is a global view of the whole task which is the 3-subtask model. The lower tier drives the robot actions as given by the RDT nodes. In the RDT model, the reference also provides a label that tells the robot what kind of behavior it should perform. The behavior can either be a basic action like spinning and moving forward or a complex action like searching or following a path. The dynamic instructions and static instructions have different strategies which can be represented by state machines. The dynamic model is not as dependent on perception and recognition abilities but rather relies on sequential movement commands. However, the static command strategy requires the robot to search and recognize the reference and target items.

IV. EXPERIMENT

The methods described above have been evaluated experimentally by executing robot spatial descriptions in a simulation environment and comparing the results to human performance using the same descriptions. The human performance provides some context in interpreting the robot results.

A. Simulation Environment and Experiment Design

Microsoft Robotics Studio is used for the robot simulation experiment environment. The virtual environment is a two-room home with a hallway between rooms, as shown in Fig. 3. The robot starts at the back of the hallway. The robot used in this experiment is a differential drive Pioneer 3DX mobile robot with a Kinect mounted at a height of 1m. For the physical robot, RGB and depth images are used to recognize the furniture and small objects inside the room [14]. For the simulation experiment, the robot uses the Kinect viewing cone and distance to determine when perception is likely to succeed. That is, if a furniture item or small object is in the viewing cone and at a close enough distance, the robot assumes that perception is successful. This method is used to approximate the robot's performance in a physical setting, which will be tested in future work. It also serves to test the spatial language methods independent of the perceptual challenges.

There are 6 scenarios in the experiment. Each has a unique target object, which are fork, glasses case, laptop, monitor, statue, and mug. In each scenario, the furniture positions are fixed while the object placement is different. Fig. 3 shows the furniture and object locations in the scene. There are 149 template spatial language descriptions for the 6 robot fetch scenarios. The descriptions are converted to tree structures and translated to robot commands as described above. In this experiment, the descriptions have been manually chunked so that they are reliable as ground truth for future NLP work.

An experiment was also conducted with human subjects; 48 undergraduates were asked to navigate through the same virtual environment using a mouse and keyboard interface to arrive at a target specified in a spatial description. This was designed to test the effectiveness of the descriptions for finding the specified target. Several cases were intentionally designed to include an ambiguous phrase, in an effort to observe how the human subjects would handle such situations. For example, the region "*in front of the couch*" might refer to the seating side of the couch if the couch's intrinsic frame is used, or it might refer to the opposite side depending on the robot position and a different reference frame being used. Each participant performed 12 trials, each with a template description; 576 trials were tested in total which were taken from the 149 unique spatial descriptions. Target objects were specified in the spatial descriptions, and subjects navigated until they reached the target location.

For the robot, the same 149 descriptions were used; however, the target object was not included in the descriptions to test how well the robot could determine the target based on the description structure and content. Each robot trial ended when the robot arrived at the position of the target furniture (as determined through the robot's reasoning processes) and turned its viewing cone on the target furniture item, i.e., the furniture that held the small target object.

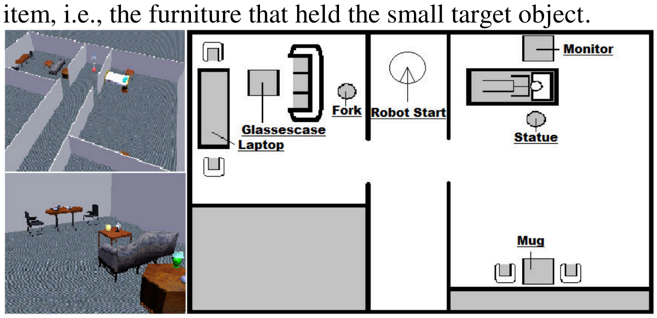

Figure 3. Simulation Experiment Environment

B. Results

We recorded robot state in each frame for each trial and took snapshots of the robot's sensor at the end of the trials. To analyze the results of the robot experiment, we tested several metrics. Here, we present results for path length, percent spin time, percent stop time, and success rate. Path metrics are generated from the robot state record and compared to the human performance using the same metrics. The success rate is analyzed for the robot only, as all paths in the human subject data ended with the specified target object. To determine whether the trial was successful, we checked whether the target object was in view in the sensor snapshot taken at the end of the trial.

Tables III through VII show the results of the experiment based on an items analysis using the 149 unique template descriptions. Mean values and standard deviations are included for each path metric. To better compare robot and human path metrics, we include only robot trials that were successful in determining the correct target. There are 123 successful robot trials out of the total 149 unique descriptions tested. The robot success rates are then analyzed for the how/where and different landmark test conditions. The overall success rate for the robot was 85%.

TABLE III. PATH LENGTH FOR HUMAN VS. ROBOT (METERS)

	Landmark	Mean	SD
Human	Goal	9.71	1.95
	None	9.82	2.26
	Path	9.22	2.08
	Total	9.54	2.10
Robot	Goal	8.66	2.30
	None	8.89	2.39
	Path	7.58	2.03
	Total	8.28	2.28

TABLE IV. PATH LENGTH FOR HOW VS. WHERE (METERS)

	Mean	SD
How	9.30	0.21
Where	8.42	0.28

TABLE V. PERCENT SPIN TIME FOR HUMAN VS. ROBOT (%)

	Landmark	Mean	SD
Human	Goal	17.82	6.78
	None	17.97	4.03
	Path	18.12	6.79
	Total	17.98	6.08
Robot	Goal	6.68	7.70
	None	6.20	7.54
	Path	28.24	25.69
	Total	15.31	20.36

TABLE VI. PERCENT STOP TIME FOR HUMAN VS. ROBOT (%)

	Type	Mean	SD
Human	How	10.25	4.95
	Where	7.43	4.41
Robot	How	0.22	0.42
	Where	1.15	7.14

TABLE VII. SUCCESSFUL RATE RESULT (%)

Types and Landmarks	How vs. Where		Goal vs. Path vs. None		
	How	Where	Goal	Path	None
Successful Rate	89.4	73.4	89.5	40.0	98.0

V. DISCUSSION

Several observations can be made from the experimental results. From the path length metric, we find that the robot has a shorter path than the human subjects in all command types and all landmark types. We also observe that the

"Where" type command results in a shorter path length than the "How" type command across all robot and human trials.

Considering percent spin time, the robot takes less spin time in the Goal and None landmark cases than the humans but considerably more spin time than humans in the Path landmark cases. This demonstrates that giving the robot more information may not necessarily help.

The percent stop time results show that the robot spends much less stop time compared to the human trials in all command types and landmark cases, perhaps because the robot is not using perception here.

Looking at the success rate results for the robot, we observe that "How" type commands have a higher success rate than "Where" type commands. Also, the commands with "Path" information show a much lower success rate when compared to other landmark cases. Several of the "Path" landmark cases intentionally included an ambiguous phrase, such as "*in front of the couch*" when the seating side was on the opposite side from the robot. In many of these ambiguous cases, the robot assumed an intrinsic reference frame by default, and got it wrong, because it was constrained from using any perceptual abilities to confirm the location as a person would. In spite of these ambiguities, the overall success rate was 85%, which indicates that performance is likely to improve if additional perceptual and reasoning capabilities are included.

We will continue to work on improving the system to robustly handle both static and dynamic commands. Future plans include a modified experiment in which the human will not be given the target object; this will provide a better comparison with the robot runs. Moreover, we will test the system on a larger set of spatial descriptions, using the actual descriptions logged in human subject experiments rather than the template descriptions. We are working on NLP methods to automate the chunking process. Our ultimate goal is to evaluate the robot in the physical environment and test the perceptual capabilities along with the spatial language methods.

ACKNOWLEDGEMENT

This work was supported in part by the U.S. National Science Foundation under grant IIS-1017097.

REFERENCE

[1] J. Beer, C. A. Smarr, T. L. Chen., A. Prakash, T. L. Mitzner, C. C. Kemp, and W. A. Rogers, "The Domesticated Robot: Design Guidelines for Assisting Older Adults to Age in Place," in *Proc., ACM/IEEE Intl. Conf. on Human Robot Interaction*, Boston, MA, 2012, pp. 335-342.

[2] M, Scopelliti, V. Giuliani, and F. Fornara, "Robots in a Domestic Setting: A Psychological Approach," *Universal Access in the Information Society*, vol. 4, pp. 146-155, 2005.

[3] W. Gribble, R. Browning, M. Hewett, E. Remolina, and B. Kuipers, "Integrating Vision and Spatial Reasoning for Assistive Navigation", in *Assistive Technology and Artificial Intelligence. Lecture Notes in Computer Science*, V. Mittal, H. Yanco, J. Aronis and R. Simpson (Eds.), Springer-Verlag, Berlin, pp. 179-193, 1999.

[4] B. Kuipers, "A Hierarchy of Qualitative Representations for Space," in *Spatial Cognition. Lecture Notes in Artificial Intelligence* 1404, C. Freksa, C. Habel, and K. Wender (Ed.), Berlin: Springer-Verlag, pp. 337-350, 1998.

[5] T. Kollar, S. Tellex, D. Roy, and N. Roy, "Toward Understanding Natural Language Directions," in *Proc., 5th ACM/IEEE Intl. Conf. on Human-Robot Interaction,* 2010, pp. 259.

[6] M. MacMahon, B. Stankiewicz, and B. Kuipers, "Walk the Talk: Connecting Language, Knowledge, and Action," *Route Instructions*, pp 1475-1482, 2006.

[7] A. Vogel and D. Jurafsky, "Learning to Follow Navigational Directions," in *Proc., 48th Annual Meeting of the Association for Computational Linguistics*, 2010, pp. 806-814.

[8] Y. Hato, S. Satake, T. Kanda, M. Imai, and N. Hagita, "Pointing to space: modeling of deictic interaction referring to regions," in *Proc. of the 5th ACM/IEEE Intl.Conf. on Human-Robot Interaction*, 2010, pp. 301-308.

[9] S. Tellex, T. Kollar, S. Dickerson, M. Walter, A. Banerjee, S. Teller and N. Roy, "Understanding Natural Language Commands for Robotic Navigation and Mobile Manipulation," *Proc., Conf. on Artificial Intelligence* (AAAI), 2011.

[10] S. Tellex, T. Kollar, S. Dickerson, M. R. Walter, A. G. Banerjee, S. Teller, and N. Roy, "Approaching the Symbol Grounding Problem with Probabilistic Graphical Models", *AI Magazine*. vol. 32, no. 4, pp. 64-76, 2011.

[11] T. Kollar, S. Tellex, D. Roy, and N. Roy, "Toward Understanding Natural Language Directions". In *Proc. ACM/IEEE Int'l Conf. on Human-Robot Interaction*, 2010, pp. 259–266.

[12] Matuszek, Cynthia, Evan Herbst, Luke Zettlemoyer, and Dieter Fox, "Learning to parse natural language commands to a robot control system." In *Proc. of the 13th Int'l Symposium on Experimental Robotics,* 2012.

[13] M. Skubic, Z. Huo, L. Carlson, X. Li, J. Miller, "Human-Driven Spatial Language for Human-Robot Interaction." *Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence*. 2011.

[14] M. Skubic, T. Alexenko, Z. Huo, L. Carlson, J. Miller, "Investigating Spatial Language for Robot Fetch Commands." *Workshops at the Twenty-Sixth AAAI Conference on Artificial Intelligence*. 2012.

[15] M. Skubic, L. Carlson; X. Li, J. Miller, Z. Huo, "Spatial language experiments for a robot fetch task." in *Proc., ACM/IEEE Intl. Conf. on Human-Robot Interaction*, Boston, MA, 2012.

[16] L. A. Carlson, M. Skubic, J. Miller, Z. Huo, and T. Alexenko. "Strategies for human-driven robot comprehension of spatial descriptions by older adults in a robot fetch task". *Topics in Cognitive Science*, in press.

[17] Radvansky, A. Gabriel, Sabine A. Krawietz, and Andrea K. Tamplin. "Walking through doorways causes forgetting: Further explorations", *The Quarterly Journal of Experimental Psychology* vol. 64, no. 8, 2011, pp. 1632-1645.

[18] Radvansky, A. Gabriel, and David E. Copeland. "Walking through doorways causes forgetting: Situation models and experienced space", *Memory & cognition,* vol. 34, no.5, 2006, pp. 1150-1156.

[19] Arkin, C. Ronald, "Behavior-based robotics", MIT press, 1998.

[20] L. A. Carlson, and P. L. Hill. "Formulating spatial descriptions across various dialogue contexts." *Spatial Language and Dialogue*, vol. 1, no.9, 2009, pp. 89-104.

[21] M. Skubic, D. Perzanowski, S. Blisard, A. Schultz, W. Adams, Magda Bugajska, and D. Brock, "Spatial language for human-robot dialogs", *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 34, no.2, 2004, pp. 154-167.

[22] Matsakis, Pascal, and L. Wendling. "A new way to represent the relative position between areal objects." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no.7, 1999, pp. 634-643.