# Human and Automated Indoor Route Instruction Following

**Matt MacMahon (adastra@mail.utexas.edu)**
Department of Electrical and Computer Engineering;
The University of Texas at Austin

**Brian Stankiewicz (bstankie@psy.utexas.edu)**
Department of Psychology;
Center for Perceptual Systems
Consortium for Cognition and Computation
The University of Texas at Austin
1 University Station A8000
Austin, Texas 78712 USA

**Abstract**

Humans possess the remarkable ability to give and follow natural language route instructions through large-scale spaces. In this process, a *director* describes the actions and observations along the route, recalling the environment's topology, metrical layout, and visual features. A *follower* interprets these descriptions, navigating by applying the instructions to the possibly unfamiliar environment. Furthermore, followers must account for mistakes, ambiguities, and omissions in the route description. To study how instructions are written and followed, we collected 756 free-form route instructions from six participants for 126 routes in three virtual environments. A second group of participants and a computational model (MARCO) followed these instructions. Humans successfully reached the destination on 68% of the instructions and MARCO followed 61% of the instructions. MARCO's performance was a strong predictor of human performance and ratings of individual instructions.

**Keywords** Artificial Intelligence; Spatial Cognition; Natural Language Understanding; Cognitive architectures; Human experimentation; Symbolic computational modeling Knowledge representation

## Introduction

Imagine while walking across a campus, a stranger approaches you to ask how to get to another location. The destination is not within sight, so you cannot simply point at the goal. Instead, you must reference your memory of routes between you and the goal. Once a route is selected, you need to access your knowledge of specific landmarks and distances to provide references for the *follower*. You translate this knowledge into a verbal description of the route. Remarkably, a short verbal description is often sufficient to guide a follower through an unfamiliar large-scale space.

Despite *directors*' best efforts, not all instructions are perfectly clear and reliable for reaching the goal. Often instructions contain ambiguous information (e.g. which tree is "`the oak tree`"), qualitative mistakes within the instruction (e.g. "`turn right`" where no right turn is possible) or metrical mistakes (e.g. "`go forward 3 blocks`" when the distance is 4 blocks). Because of these failings, the follower must treat the instructions as guidance, not as strict commands.

With all the potential for miscommunication arising from the complexity of giving and understanding route instructions, the human ability to provide instructions that can typically be followed is remarkable. The current paper investigates how people give route instructions about indoor environments they have learned through navigation. We present a computational model of route instruction following called MARCO and investigate the following three questions:

1. How do quality and style vary in route instructions?
2. How well does MARCO follow route instructions and how closely does the model's behavior correlate with human behavior?
3. Can MARCO differentiate good versus bad instructions, by both human performance and ratings?

### Why are some route instructions reliable?

Though a large literature examines route instructions, there is no consensus about what differentiates good instructions from bad. Vanetti and Allen (1988) found spatial ability had a larger effect in the accuracy of subjects' described routes than verbal ability. Daniel et al. (2003) found "good", "poor", and "skeletal" instructions were differentiated by whether the proper action was associated with the proper landmark. Allen (2000) suggests descriptives clauses and action delimiters should be inserted at choice points and near the destination, instead of en-route. Lovelace et al. (1999) found good route instructions mentioned many landmarks along the paths, off the route, and at the choice points.

Some of these studies rate route instructions subjectively (Vanetti and Allen, 1988; Lovelace et al., 1999; Tversky and Lee, 1999), but do not test if navigation success is affected. Others have participants follow a small number of participant- and experimenter-written instructions (Daniel et al., 2003; Allen, 2000). Our work implements a study suggested by Lovelace et al. (1999), where participants follow route instructions from different routes and directors in virtual reality.

### Computational models of route instructions

Software systems that analyze or follow route instructions can be distinguished by how they represent space. Freundschuh and Egenhofer (1997) survey a variety of spatial representation models and define broad categories based on (1) if the objects in the space are *manipulable*, (2) if the space requires *locomotion* to experience, and (3) the size, or *scale*, of the space. This work focuses on non-manipulable, large-scale spaces that cannot be experienced from any one

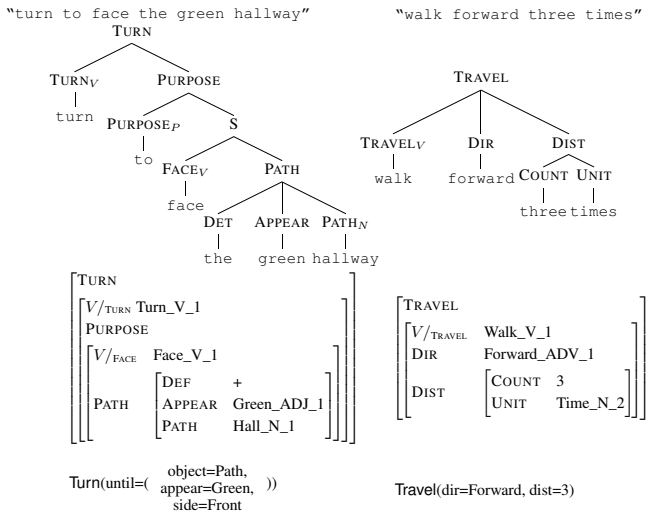


Figure 1: The *Text Interpreter* parses each sentence to get a syntactic tree, then labels the word senses and transforms the tree to a content frame. The *Instruction Modeler* interprets the frame as an underspecified command that the *Route Executor* acts to fulfill, guiding the follower's navigation through the environment.

perspective: the agent must turn (*panoramic space*) or move (*environmental space*) to see the space.

Skubic et al. (2004) developed software that can recognize and reason about spatial relations. It moves a robot within a room to achieve a spatial command. Bugmann et al. (2004) compared the performance for a robot navigating through a tabletop model environment a system following (1) programs translated by hand from speech, (2) software-generated models of the instructions, and (3) people controlling the motion. Bugmann's participants saw an outside, panoramic perspective of a small model of a town neighborhood the robot navigated. This paper builds on Simmons et al. (2003), a system that follows route instructions through large-scale environments.

## Marco Route Follower Model

MARCO is designed to follow route instructions in large-scale spaces, between places not mutually visible and separated by travel. We tested MARCO on instructions written from memory by people who learned the environments from a first-person perspective while navigating. MARCO follows these route instructions without any *a priori* environmental model by reasoning about actions, views, and topology. These are the Causal and Topological representation levels of the Spatial Semantic Hierarchy (Kuipers, 2000).

MARCO has modules for interpreting and following written, natural language route instructions (MacMahon et al., 2006). MARCO consists of three primary modules (see Figure 1 for a trace of the linguistic modeling): A *Text Interpreter*, an *Instruction Modeler*, and a *Route Executor*. For the sake of brevity, we describe only the fundamental properties of these modules. For details, see MacMahon et al. (2006).

### Text Interpreter

The Text Interpreter models the surface structure of an utterance and the surface meaning of an utterance. The Syntactic Parser parses raw text into grammatical structures. Our grammar directly models verb-argument structure, instead of part-of-speech syntax (see the parse tree in Figure 1). Next, the Text Interpreter translates the surface structure of an utterance to a model of the surface meaning, that drops arbitrary word ordering and marks words with meaning sense, abstracting over changes in morphology, spelling, and synonyms.

### Instruction Modeler

The Instruction Modeler translates the surface meaning of what was said into an imperative model of what to do, when. The Instruction Modeler combines information across phrases and sentences to generate either *imperatives* (specific action instructions; e.g. "go until") or *declaratives* (i.e. information about the environment; e.g. Path(appearance:Blue, length:Long). The representation captures the underspecified commands in the route instructions, modeling the route as a sequence of simple actions (Turn, Travel, Verify) to be taken under certain perceptual (e.g. seeing a view) or cognitive conditions (e.g. estimating a distance). This step is similar to the "minimal units of information" Denis (1997) derived manually. Figure 1 shows the transformation from text to the imperative model.

### Route Executor

To navigate, MARCO interprets the imperative model in the context of the environment. The Route Executor picks actions given the context from perceiving the environment and tracking the state within the route instructions. It checks symbolic view descriptions against sensory observations and spatial models. MARCO performs *symbol anchoring* (Coradeschi and Saffiotti, 2003) by tying each concept to its experience. It verifies if the described attributes of the environment are consistent with the observation stream and acts to reach the described states along the route.

## Route Instruction Experiment

To understand how humans give and follow instructions, we developed a collection of novel virtual environments. We collected route instructions from six participants who learned to navigate efficiently in the environments. Thirty-six human participants and two variants of MARCO followed these instructions. The human participants subjectively rated the instructions. We are interested in how well MARCO can follow these instructions and how MARCO's performance correlates with human navigatation and ratings of route instructions.

### Methods

**Apparatus** These experiments used desktop virtual reality, with human participants with a first-person perspective moving through a computer-animated three-dimensional world. Figure 2 provides an overhead map of

one of environments (Top) and the first-person perspective of the participants navigating through it (Bottom). The experiments are controlled by Python scripts using the WorldViz Vizard software (Vizard, 2003). Subjects navigate in discrete motions using the keyboard: '8' key moves forward one hallway, '4' turns left, and '6' turns right.

**Stimuli** These environments and the experiment control software build on top of previous studies on spatial navigation (Stankiewicz et al., 2001; Kuipers et al., 2003). To provide useful cues for the directors, we placed 11 objects of 6 different types within each environment. Furthermore, each environment was divided into three separate regions, designated by distinct pictures on the walls (see figure 2). Finally, 7 long hallways within each environment had a visually distinct texture mapped onto the floor. Figure 2 (Top) shows the layout for one of the environments and Figure 2 (Bottom) shows the view of an easel on a black stone hallway in the fish region.

The three testing enviroments varied in the density of the layout, as measured by the shortest travel routes betweeen the named positions. The most compact had a mean shortest path length of 4.2 (median 4), the most spread-out enviroment, mean 6.0 (median 6). The shortest route was one travel action, the longest was 13.

**Procedure** Two sets of participants were used in this study: *Directors* and *Followers*. Directors learned each environment until they could navigate efficiently, then wrote text route instructions, navigated the routes, and rated themselves. Followers followed the Directors' instructions without any previous knowledge of the layout, rated the quality of the instruction, and rated their certainty of reaching the described destination.

**Directors** Directors were instructed to learn the environment and the location of the seven *target* locations. The name of each target was announced by a computer-generated voice (e.g. "Position 3") when the participant entered a location. The directors were told that later they would give instructions to travel between these target locations. Directors had free exploration sessions of 120 travel actions to learn the spatial layout.

After each free exploration period, we evaluated how well the director knew the environment by giving a *navigation efficiency test*. In this test, the computer started the participant at one of the target locations and instructed the participant to travel to a target location (e.g. "Go to position 2."). The participant was instructed to travel to that location by the shortest route. After reaching the designated target location, the computer compared the number of travel actions used to the shortest route. When director reached all goals within 150% of the shortest path for each route for seven consecutive routes, the navigation efficiency test ended. After three routes over this threshold, the director returned to the free exploration phase.

After passing the navigation efficiency test, the director entered the *route instruction* phase of the study. In this phase, the director gave instructions for routes between the target locations. For each route, the director

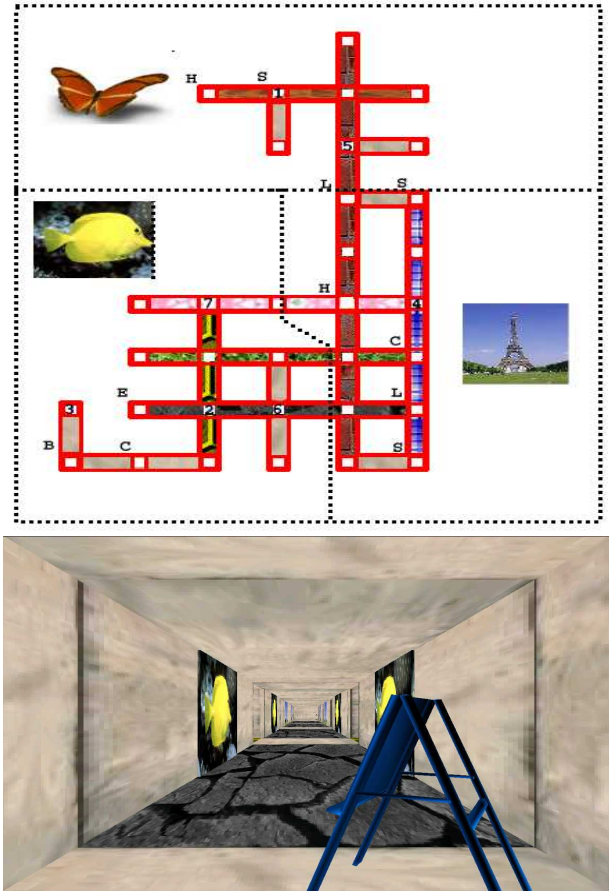

Figure 2: **Top**: Map of one of three virtual environments (not seen by participants). Three regions share a wall hanging of a fish, butterfly, or Eiffel Tower. Each long hallway has a unique flooring. Letters above mark objects (e.g. 'C' is a chair), numbers indicate named positions. **Bottom**: Participants' first-person view from the easel ('E') at the end of the black hall in the map.

experienced the following sequence of events:

1. The director was placed at the starting location, facing a random direction.
2. The position name was announced (e.g. "Position 7") and the director was allowed to turn freely to orient, but not to move forward. Once ready, the participant pressed a button.
3. A text-entry window then appeared on the screen. The director typed their instructions without a time limit. The director could also move the cursor to correct previously typed text in this step. The director clicked a button when the instructions were finished.
4. After giving instructions, the director navigated from her current location to the specified target location. Upon reaching the target location, the director pressed the space bar indicating that she was at the goal.
5. On a six-point scale, the director rated:
   (a) how certain she was she reached the destination
   (b) and the quality of her own instructions.

For the seven locations within an environment there are 42 possible pairs (7 choose 2). Each participant gave instructions between all 42 ordered pairs of named positions in a random order. Each director repeated this procedure for all three environments, on separate days.

**Followers** Participants in the *follower* group were told that they would be given a set of instructions written previously by other participants. The followers were instructed that they should follow the instructions to the best of their ability. The route instructions were presented on the computer screen as text as typed by the director. However, any reference to a target position by name (e.g. "`This is Position 1`") was replaced with an anonymizing 'X' or 'Y.' The followers recognized destinations from the descriptions, not a name.

Each follower followed and rated 126 instructions balanced across routes, directors and environments. The procedure interleaved instructions from all three environments to discourage the followers learning the environments. No follower experienced exactly the same route twice, so none repeated a route with new route instructions.

Each follower experienced the following sequence of events for each route instruction:

1. The computer presented a text box containing the route instruction text. The follower was allowed to read the instructions without a time limit, selecting an 'OK' button when finished.
2. The follower was placed at the starting position facing a random direction.
3. The follower navigated through the environment.
4. At any time, the follower could review the instructions by pressing 'd' on the keyboard. The instruction display fully obscured the view of the environment.
5. When the follower believed that he had reached the destination described in the instructions (or finished trying), he pressed the space bar.
6. The follower rated how confident he was that he had reached the goal and the quality of the instructions, both on a six-point scale.

**Participants** Forty-two participants were used in the study. Six participants were directors (3 females) and thirty-six were followers (15 females). The directors were paid $10.00/hour for an average of 7 hours. The followers participated for one or two hours to help satisfy course credit in an undergraduate psychology course.

### Human Instruction Experiment Results

Table 1 provides a sample of the instructions given by the six directors. The instructors' styles varied from very sparse instructions providing specific move sequences (EDA) to very rich and elaborate instructions (KLS).

### Marco Performance Study

Over runs through the 756 route instructions (42 routes in 3 environments for 6 directors), we measured how often MARCO successfully reaches and recognizes the destination. For analysis, we focus on the 682 of the route instructions where director typed at least one sentence. We also used MARCO and the test corpus to examine why people prefer some instructions over others, even when both lead to the goal.

Table 1: Example instructions for routes from same start and end. Instructions include errors (e.g. "`halllway`").

**EDA:** `turn to face the green halllway, walk three times forward, turn left, walk forward six times, turn left, walk forward once`

**EMW:** `Follow the grassy hall three segments to the blue-tiled hall. Turn left. [...] Turn left. Go one segment forward to the corner. This is Position 5.`

**KLS:** `take the green path to the red brick intersection. go left towards the lamp to the very end of the hall. at the chair, take a right. [...] at the end of this hall at the corner, you are at position 5`

**KXP:** `head all the way toward the butterfly hallway, keep going down it until you reach a dead end square area. pos 5 is in the corner to the left as you enter the block.`

**TJS:** `go all the way down the grassy hall, take a left, go all the way down the blue hall until you see a coat rack, take another immediate left.`

**WLH:** `from four face the grass carpet and move to the hat rack, [...] move into the corner such that the lamp is behind you and to your right you see a gray carpeted alley`

We ran MARCO with and without an error recovery strategy. The recovery strategy is a *Find* behavior, which simply performs a "drunkard's walk" when the follower does not see a view that matches a necessary description in the instructions. The follower randomly picks one of the paths at the current location to travel along, then checks whether the sought-after view is now visible. There is also a small chance of giving up, which gradually increases with each forward move.

## Marco and Human Comparison

Table 2: MARCO model's success predicts people's with 84% precision, 75% recall, and 79% F-measure.

| | Human Success | Human Failure |
|---|---|---|
| MARCO Success | 1809 | 338 |
| MARCO Failure | 618 | 758 |

Human followers successfully followed the instructions on 68% of the route instruction runs. Tables 2 and 3 and Figure 3 summarize the success rates of MARCO and people and how subjects rated the instructions.[1]

[1] Comparisons are with MARCO as of April 21, 2006.

Table 3: Correlations of success rates and human subjective ratings. Spearman rank-order correlation coefficients ($R_S$) are all significant at $p < 0.001$.

| | Route Success Rate | Human Subjective Rating $R_S$ | Human Success Rate $R_S$ |
|---|---|---|---|
| Human | 68% | 0.578 | *1.000* |
| Full MARCO | 61% | 0.544 | 0.607 |
| MARCO w/o Find | 53% | 0.610 | 0.585 |

As expected, there is a strong correlation between the instruction rating and the human rate of stoping at the goal ($R_S$=0.578). The mean human subjective rating has a slightly lower correlation to the success rate of the full MARCO model ($R_S$=0.544), but a higher correlation to the success rate of MARCO without Find ($R_S$=0.610).

Comparing the success rates, a different pattern emerges. Human success rate per route instruction correlates more highly with the success rate of the full MARCO ($R_S$=0.607) than with the success rate of MARCO without Find ($R_S$=0.585). The full MARCO model is a better predictor of the objective human success rate for a route instruction, while MARCO without Find is better predicts human subjective rating.

## Analysis of Follower Performance

Figure 3 shows the navigation success rates for people, the full MARCO and MARCO without the Find behavior. Each plotted point is the arithmetic mean over the set instructions with post-hoc human subjective rating of $n \pm 0.125$, with poor instructions (from 1.0) on the left to and excellent on the right.

The top line (○) shows the success rate by human followers increasing with instruction rating. The default MARCO (□) system approximates human performance for highly-rated instructions ($> 4.0$), while succeeding less often on poorly-rated instructions ($< 4.0$). On highly-rated instructions (right side), using the Find error recovery method (▷) does not greatly affect performance. However, for poor instructions, Find actions become as crucial, as can be seen as the performance of the system without Find (▷) drops on the poor instructions on the left.

We coded the primary reason for failure on 118 instructions where Marco's success rate is 50 percentage points less than people's. There were 8 errors in modeling word meaning, 28 errors in modeling phrases, 8 errors in combining phrases within an utterance, 50 errors in combining information from separate utterances, 5 perceptual errors, 6 anaphora errors, and 8 errors of over-relying on part of the description.

Based on this discrepancy analysis, the most effective ways for MARCO to improve are in better and more comprehensive modeling of phrases and in modeling discourse context. Some phrases are not yet interpreted, such as some fictive motion phrases, while others are misinterpreted in some contexts. Some

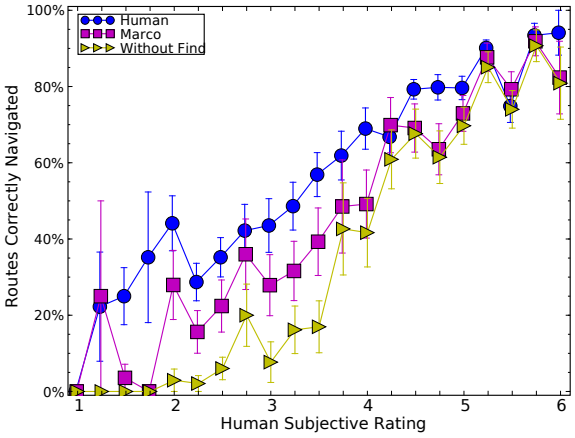


Figure 3: Route instruction following success for runs by human followers and MARCO with and without Find.

modeling of discourse context will allow interpretation when one utterance depending on the prior utterance. One example is "`Go down two intersections. At the third, turn right,`" which was also a difficulty for Bugmann et al. (2004).

## Analysis of Director Performance

Figure 4 shows the mean success rate on instructions from each director for people, MARCO and MARCO without Find. Each bar represents the mean percentage of routes successfully followed over multiple runs through the route instructions by a director (all directors for 'All'). Each group of bars displays the performance for runs using the instructions from one director.

MARCO best approximates human performance on two of the directors who give the more reliable and highest rated instructions (EDA and EMWC). These directors also show the least drop off without the Find behavior. MARCO does not perform quite as well on these instructions from the next two most reliable directors, KLS and WLH, but still does not require the Find behavior often. These directors all tend to give instructions explicitly covering entire route.

On instructions from the two poorest performing directors, KXP and TJS, MARCO does not perform as well as people do. Both of these directors wrote instructions that require frequent use of the Find behavior, as can be seen by the decrease in performance without it. The routes from these directors are often fragmentary and error-laden.

# Conclusions

This study examined what separates highly- and poorly-rated verbal route instructions. One human study collected a large corpus of text route instructions describing three complex large-scale virtual environments. A second human study followed and rated these route instructions. A software system that can parse, model, and reactively enact route instructions is presented. The

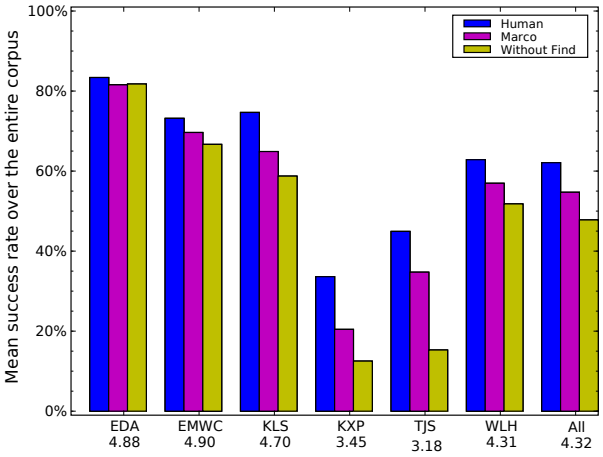


Figure 4: Route instruction following success and mean human instructions rating for each of the 6 directors by people and MARCO with and without Find.

system, MARCO, approximates human performance, as measured by whether the follower successfully navigates from a starting place to the destination and correctly declares reaching the goal.

We find that the base MARCO system is a strong predictor of which instructions people can follow, while the MARCO system without an error recovery behavior is a strong predictor of how people will rate the instructions. If MARCO successfully reaches the goal, people most likely will also. If MARCO must use its error recovery Find behavior to reach the destination, people will be able to reach the goal less often and will rate the instructions lower. Further work will use MARCO to investigate how people think and communicate about large-scale spaces and provide a practical interface for assistive technologies such as smart wheelchairs.

## Acknowledgments

This work was supported by AFOSR grants FA9550-04-1-0236, FA9550-05-1-0321 and NIH grant EY016089 to B.J. Stankiewicz, by NSF grant IIS-0413257 to Benjamin J. Kuipers, and under support for Matt MacMahon through ONR work order N0001405WX30001 for the NRL Research Option, Coordinated Teams of Autonomous Systems. We gained insights from discussions with Benjamin Kuipers, the members of the UT CS Intelligent Robotics Laboratory, and the UT Psychology Space and Shape Laboratory.

## References

Allen, G. L. (2000). Principles and practices for communicating route knowledge. *Applied Cognitive Psychology*, 14(4):333–359.

Bugmann, G., Klein, E., Lauria, S., and Kyriacou, T. (2004). Corpus-based robotics : A route instruction example. In *Proc. of the Intelligent Autonomous System*, pages 96–103, Amsterdam, The Netherlands.

Coradeschi, S. and Saffiotti, A. (2003). An introduction to the anchoring problem. *Robotics & Autonomous Systems*, 43(2-3):85–96.

Daniel, M.-P., Tom, A., Manghi, E., and Denis, M. (2003). Testing the value of route directions through navigational performance. *Spatial Cognition and Computation*, 3(4):269–289.

Denis, M. (1997). The description of routes : A cognitive approach to the production of spatial discourse. *Current Psychology of Cognition*, 16(4):409–458.

Freksa, C. and Mark, D. M., editors (1999). *Spatial Information Theory: Cognitive and Computational Foundations of Geographic Information Science (COSIT '99)*, volume 1661 of *Lecture Notes in Computer Science*, Stade, Germany. Springer.

Freundschuh, S. and Egenhofer, M. (1997). Human conceptions of spaces: Implications for GIS. *Trans. on Geographic Information Science*, 2(4):361–375.

Kuipers, B. J. (2000). The Spatial Semantic Hierarchy. *Artificial Intelligence*, 119:191–233.

Kuipers, B. J., Tecuci, D. G., and Stankiewicz, B. J. (2003). The skeleton in the cognitive map : A computational and empirical exploration. *Environment &Behavior*, 35(1):80–106.

Lovelace, K. L., Hegarty, M., and Montello, D. R. (1999). Elements of good route directions in familiar and unfamiliar environments. In Freksa and Mark (1999), pages 56–82.

MacMahon, M., Stankiewicz, B., and Kuipers, B. (2006). Walk the talk: Connecting language, knowledge, action in route instructions. In *Proc. of the 21st National Conf. on Artificial Intelligence (AAAI-2006)*, Boston, MA.

Simmons, R., Goldberg, D., Goode, A., Montemerlo, M., Roy, N., Sellner, B., Urmson, C., Schultz, A., Abramson, M., Adams, W., Atrash, A., Bugajska, M., Coblenz, M., MacMahon, M., Perzanowski, D., Horswill, I., Zubek, R., Kortenkamp, D., Wolfe, B., Milam, T., and Maxwell, B. (2003). GRACE: An autonomous robot for the AAAI Robot Challenge. *AI Magazine*, 24(2):51–72.

Skubic, M., Perzanowski, D., Blisard, S., Schultz, A., Adams, W., Bugajska, M., and Brock, D. (2004). Spatial language for human-robot dialogs. *IEEE Trans. on Systems, Man&Cybernetics – Part C*, 34(2):154–167.

Stankiewicz, B. J., Legge, G. E., and Schlicht, E. (2001). The effect of layout complexity on human and ideal navigation performance. *Journal of Vision*, 1(3).

Tversky, B. and Lee, P. U. (1999). Pictorial and verbal tools for conveying routes. In Freksa and Mark (1999), pages 51–64.

Vanetti, E. J. and Allen, G. L. (1988). Communicating environmental knowledge : The impact of verbal and spatial abilities on the production and comprehension of route directions. *Environment &Behavior*, 20:667–682.

Vizard (2003). WorldViz Vizard virtual reality software. http://www.worldviz.com/vizard.htm.