Which way to turn? Guide orientation in virtual way finding

Mark Evers
Technical & Professional Communication
University of Twente
The Netherlands
M.Evers@alumnus.utwente.nl

Mariët Theune
Human Media Interaction
University of Twente
The Netherlands
M.Theune@utwente.nl

Joyce Karreman
Technical & Professional Communication
University of Twente
The Netherlands
J.Karreman@utwente.nl

Abstract

In this paper we describe an experiment aimed at determining the most effective and natural orientation of a virtual guide that gives route directions in a 3D virtual environment. We hypothesized that, due to the presence of mirrored gestures, having the route provider directly face the route seeker would result in a less effective and less natural route description than having the route provider adapt his orientation to that of the route seeker. To compare the effectiveness of the different orientations, after having received a route description the participants in our experiment had to 'virtually' traverse the route using prerecorded route segments. The results showed no difference in effectiveness between the two orientations, but suggested that the orientation where the speaker directly faces the route seeker is more natural.

1 Introduction

When someone approaches us and asks which way to go, we naturally turn – if necessary – so we face the direction to take (which makes it also easier for ourselves to imagine traversing the route). Generally, the route seeker then also turns to adapt his or her orientation to match ours, and we end up sharing the same perspective on the route to take.[1] Presumably, this matching of physical orientation is meant to reduce the mental effort that is involved in matching another person's perspective on a spatial scene for both speaker and hearer (Shelton and McNamara, 2004). However, someone who faces an embodied virtual agent presenting a route description in a virtual environment (projected on a computer screen) cannot turn to match his or her perspective with that of the agent, as turning away from the screen would result in losing sight of both the agent and the virtual environment. In this situation, the only way to bring the perspectives of route provider (agent) and route seeker (user) closer together is for the agent to adapt its orientation to match that of the user. In this paper, we describe an experiment carried out to determine if such a change in orientation by the route provider helps the route seeker with virtual way finding. Although the experiment was aimed at determining the most effective and natural orientation of a Virtual Guide, we used prerecorded route descriptions presented by a human route provider. The Virtual Guide that we have developed (see next section) was still being implemented at the time.

2 The Virtual Guide

We have developed an embodied Virtual Guide[2] that can give route directions in a 3D environment, which is a virtual reality replica of a public building in our home town. When navigating through this virtual environment, shown on the computer screen from a first person perspective, the user can approach the Virtual Guide to ask for directions. Currently the

[1] This observation is based on personal experience. We also observed this behaviour in a small corpus of route description video's.

[2] See http://wwwhome.cs.utwente.nl/˜hofs/dialogue for an online demo.

Proceedings of the Workshop on Embodied Language Processing, pages 25–32,
Prague, Czech Republic, June 28, 2007.

Guide is behind the reception desk (see Figure 1), but she can be situated anywhere in the building.

The first part of the interaction between the Virtual Guide and the user consists of a natural language dialogue in which the Guide tries to find out the user's intended destination. This may involve subdialogues, in which either the Guide or the user asks the other for clarification, and the resolution of anaphoric expressions (e.g., *How do I get there?*). Input and output modalities include text, speech and pointing. For an in-depth description of the dialogue module of the Virtual Guide, see Hofs et al. (2003).

When the user's destination has been established, the Virtual Guide gives a natural language route description, in the form of a monologue that cannot be interrupted. This is somewhat unnatural since in real direction giving, the route seeker tends to give feedback and, if necessary, ask for clarification while the route is being described. However, since in our system dialogue management and the generation of route descriptions are handled by separate, specialised modules this is currently not possible.

The route is presented as a sequence of segments, which are mostly expressed as "point+direction" combinations (Dale et al., 2005). That is, they consist of a turn direction combined with the location where this turn is to be made, specified in terms of a landmark. For example, *You go left at the information sign*. The route description is generated as follows. First, the shortest path between starting point and destination is computed based on predefined paths in the virtual environment. Turn directions are derived from the relative angles of subsequent path segments, and landmarks are selected based on their relative salience (e.g., in terms of size or colour) and proximity to a turning point. The sequence of turn directions and associated landmarks is then given as input to the natural language generation component, which is based on Exemplars (White and Caldwell, 1998). After a first version of the route description has been generated using a collection of standard sentence structures, this initial description is revised by randomly aggregating some sentences and adding cue phrases such as *and then, after that* etc. to achieve some variation in the generated text.

To generate appropriate gestures to accompany the verbal route description, the generated text is extended with tags associating the words in the route description with different types of gestures. Currently this is done using a simple keyword approach. Direction words (*left, right*) are associated with pointing gestures in the corresponding directions, and references to landmarks are associated with deictic gestures pointing to either the absolute or the relative location of these objects (see Section 3). Some iconic gestures (i.e., gestures that have a resemblance in shape to what they depict) are also available, for example a horizontal tube-like gesture that can be used in references to corridors and tunnels. Unlike the pointing gestures, which are generated "on the fly", the iconic gestures of the Virtual Guide are generated by using canned animations. For a more sophisticated approach to the generation of iconic gestures, see the work by Kopp et al. (in press) who describe the dynamic planning of novel iconic gestures by NUMACK, an embodied conversational agent that functions as a virtual guide for the Northwestern University campus.

The last stage of the route description process in our Virtual Guide is to send the marked-up text to the animation planner, which actually generates the required animations in synchronization with text-to-speech output. The animation planner is based on the work by Welbergen et al. (2006).

Figure 1: The Virtual Guide.

3 The Guide's gestures and orientation

During the route description, the Virtual Guide can make pointing gestures from either an 'objective' viewpoint, i.e., pointing at the absolute locations of objects, or from a 'character' viewpoint, i.e., pointing at locations relative to the position of a person who is walking the route. An objective viewpoint makes most sense when pointing at objects that are (in principle) visible to both the agent and the user, which is only the case for objects that are located at the start of the route. So, most of the time the Guide will be using the character viewpoint, pointing left and right relative to its own body to indicate landmarks and directions from the perspective of someone who is walking along the route being described.

The typical orientation of information presenting agents is facing the user. However, it is not a priori clear that this would be the best option for the Virtual Guide. When facing the user, all pointing gestures made by the guide from a character viewpoint would mirrored in the eyes of the user, so the latter would have to perform a mental 180° re-orientation of the gestures. This would demand extra cognitive effort on top of processing and storing the verbally presented route information, and might negatively influence the user's ability to reproduce the route directions during actual traversal of the route.

In actual direction giving situations, people often tend to minimize the difference in orientation between them. Therefore we wondered if reducing the difference in orientation between the agent and the user would help the user to find his way during traversal. If the agent would turn to face almost the same direction as the user, its gestures could be expressed as close to the route seeker's perspective as possible, thus reducing the cognitive load for the user in processing them. Also, we wondered if this configuration would yield a more natural effect than having the agent directly face the user during the route description. We investigated these questions in an experiment where participants had to virtually follow a route, presented to them in one of two versions that differed in the orientation of the route provider. Because the Virtual Guide was still being implemented at the time, we used route descriptions by a human route provider. The experimental setup and its results are presented below, followed by some conclusions and future research directions.

4 The orientation experiment

The goal of the experiment was to investigate the effect of speaker orientation on the effectiveness and naturalness of a route description. For our experiment, we opted to use prerecorded route descriptions, as this matched the capabilities of our Virtual Guide (which can only present the route as a monologue with no interaction) and also ensured an unlimited number of reproductions of constant quality and content. We recorded two separate route descriptions that differed in speaker orientation with respect to the route seeker, but were otherwise (largely) the same:

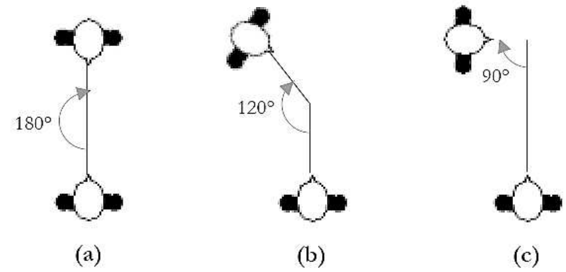


Figure 2: Angle between route provider and route seeker (camera)

180° version The route provider is oriented at a 180° angle with respect to the route seeker, i.e., he directly faces the camera lens, creating mirrored gestures (his left is seen as right by the viewer and vice versa). See Figures 2(a) and 3(a).

120° version The route provider is oriented at a 120° angle toward the route seeker, as if to adapt his orientation to that of the route seeker. See Figures 2(b) and 3(b).

We chose an orientation of 120° for the route seeker-oriented version, so as to maintain visibility of non-verbal signals. If the route provider were to assume an orientation of 90° or less, as illustrated in Figure 2(c), not all gestures would be visible and maintaining eye contact could make his posture unnatural.

The 120° and the 180° condition only differed in bodily orientation while eye contact remained unchanged and facial expressions remained visible. Also, although wording slightly varied, the presented information was the same in both conditions. The route descriptions were recorded on location in a small town with short streets and plenty

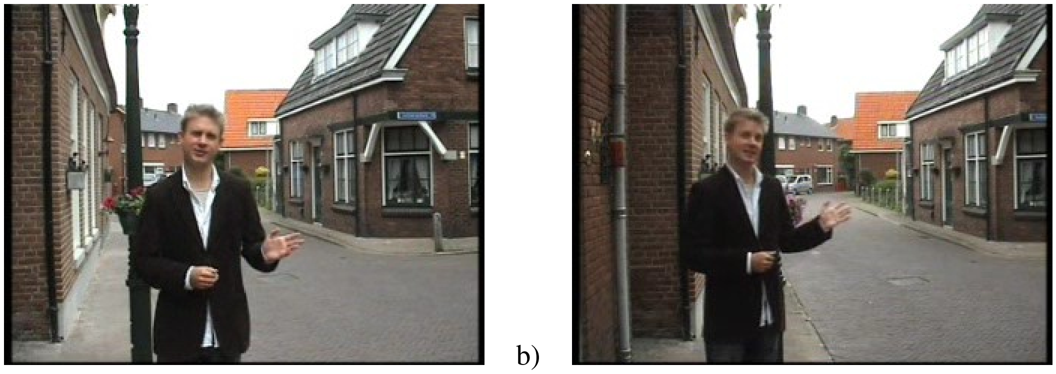
a)  b)

Figure 3: “Turn left at the white building” (a: 180°, b: 120°)

of landmarks. The route being described led from the recording location to the town hotel. The verbal description was similar in structure to those generated by the Virtual Guide. It mentioned five decision points, each connected with one or two characteristic landmarks. For example, *At the men’s fashion shop, you turn right*. During the route description, the route provider made beat gestures and pointing gestures from a character viewpoint, taking his own body orientation as a reference for left and right. Apart from a few slight variations, the gestures used in both versions of the route description were the same; see Figure 3. At the start of the route description, both route provider and route seeker were exactly (180° version) or almost (120° version) perpendicular to the starting direction of the route.

After viewing one of the two versions of the route description, the participants in the experiment had to ‘virtually traverse’ the route (to measure effectiveness of the route description) and were asked how natural they found the route description. The most realistic way to measure effectiveness of the route description would have been to have the participants walk the route in reality after having received the description, as was done by Fujii et al. (2000) and Michon and Denis (2001). However, conducting such an experiment is a very time consuming activity. As a more practical alternative we developed a reconstructive method allowing participants to traverse the route on the computer, instead of in a real (live) environment. In this set-up, participants ‘traversed’ the route by viewing prerecorded route segments, showing a moving scene from a first person perspective as if they walked through the streets themselves, accompanied by street sounds. Apart from practical considerations, an additional advantage of this set-up is that it yields full control with respect to repeatability and the participation setting because of its playback nature.

Our hypotheses were as follows:

1. The 120° version is more effective, i.e., yields a more successful traversal than its 180° counterpart.

2. The 120° version yields a more natural route description than its 180° counterpart.

4.1 Participants

A total of 49 participants were involved in the experiment, aged 20 to 64 years (with an average of 33 years). Since no participants were younger than 12 or post 70, no specific effect of age on their spatial skills was expected (Hunt and Waller, 1999). Since gender is an influential factor in orientation and way finding (Hunt and Waller, 1999; Lawton, 1994), we used a 50% male - 50% female test population. The 120° version of the route description was shown to 13 male and 12 female participants; the 180° version to 11 male and 13 female participants.

4.2 Procedure

The experiment consisted of the following steps.

Introduction - After reading an introductory text explaining the experiment, the participant filled in a pre-questionnaire asking for age, gender, and educational level. We also asked how familiar the participant was with the route location, indicated on a

5-point scale ranging from not at all familiar (1) to very familiar (5). If the participant indicated being moderately or more familiar with the location, his or her results were discarded. The questionnaire was followed by an example question to familiarize the participant with the controls and with the set-up of the traversal part of the experiment.

Route description - First, the participant was shown a video impression of the location where he or she, being lost in an unfamiliar town, supposedly approached someone to ask the way to the hotel. Then the participant watched one of the two prerecorded route descriptions. To compensate for the fact that, unlike a real-life situation, there was no opportunity to verify understanding or ask for clarifications, the participants were allowed to play the route description video twice.

Traversal - After having received the route description, the participant had to virtually traverse the route by watching six prerecorded traversal segments in succession, appearing in a pop-up window. The first segment began at the starting point of the route and ended at the first decision point (intersection). Each following segment started where the previous one ended, with the final segment ending at the destination of the route. At the end of each route segment, an overview of the next intersection was provided by moving the camera viewpoint gradually so the entire intersection was shown. The average length of each traversal segment was around 1.5 minutes.

After watching each segment, the participant had to select which direction to take next from a limited set of options: left, straight ahead or right (if applicable). Each option was accompanied with a photo of the corresponding view from the crossing. After answering the question, the participant was informed which direction was correct. Then the participant proceeded with the route traversal from the correct turn, regardless whether the correct direction had been chosen or not.[3]

[3]This differs from the effectiveness measure of Fujii et al. (2000), who used a movement failure rate defined as Out/N, with Out being the number of times a participant lost the way and was unable to return to the route, and N being the number of trials. We found this method too complicated in design and too confusing for the participants to be used in this experiment. In our set-up, the participant was only allowed one trial per decision point and always traveled along the correct route.

	120°	180°	Total
Male	3.46 (0.88)	3.27 (1.19)	3.38 (1.01)
Female	4.00 (1.04)	3.62 (0.77)	3.80 (0.91)
Total	3.72 (0.98)	3.46 (0.98)	3.59 (0.98)

Table 1: Number of correct decisions as a function of gender and version (results are presented as Means with Std. Deviations in brackets).

Post-questionnaire - After route traversal, the participants answered several questions about the route description. Here we only focus on one of the questions, i.e., “Do you think the route provider described the route in a natural way?”, to be answered on a 5-point scale ranging from very natural (1) to very artificial (5). The participants were also offered the opportunity to comment on their answer.

5 Results and discussion

Here we present and discuss the main findings from our experiment.

5.1 Effectiveness of the route description

Hypothesis 1 concerned the influence of speaker orientation on the effectiveness of the route description. We measured this by counting the number of correct turns taken by the participants during route traversal. The route contained five decision points (intersections), so participants’ scores ranged from 0 to 5 correct turns. Gender has been proved to strongly influence way finding ability (Hunt and Waller, 1999; Lawton, 1994), so gender was accounted for as a fixed factor in our analysis.

The results are summarized in Table 1, which shows that participants performed slightly better in the 120° version than in the 180° version, and that women performed slightly better than men. However, these differences were not significant; neither for version nor gender. Thus, our first hypothesis is not supported.

This lack of effect might be taken as evidence that gestures hardly play a role in conveying information, so that a difference in their orientation would not affect the route seeker’s mental processing of the route description. It has been argued that the main function of gestures in conversation is not to transfer information to the interlocutor, but to facilitate the cognitive process of speaking

(Rimé and Schiaratura, 1991; Morsella and Krauss, 2004). Still, though most spontaneous gestures may not be produced for the interlocutor's benefit, it has been shown experimentally that people do make use of the information conveyed by gestures (Kendon, 1994; Cassell et al., 1999; Kelly et al., 1999). The communicative power of gestures does seem to depend on the task and the type of gesture, however (Bangerter and Chevalley, 2007). In fact, in our experiment the gestures were not essential for understanding the route description. All pointing gestures were accompanied by explicit verbal descriptions of the corresponding landmarks and/or directions; in other words, the gestures were redundant with respect to speech. So, regarded from a purely informational point of view, these gestures were superfluous and the participants may have paid only limited attention to them or even consciously ignored them. This explanation is supported by the comments of various participants who said they tried to focus on the verbal instructions because the description was extensive and they found the gestures distracting.

We consciously limited the number of decision points in the experiment to five, well within the 7 ± 2 range of short term memory, but for each decision point the route provider not only mentioned the direction to take, but also one or two landmarks. Furthermore, he gave some auxiliary hints of what to do in-between turns (*Walk straight ahead until you see a traffic sign; there you keep walking straight ahead*) and some more details. In their comments, several participants mentioned being distracted by too much detail in the description, and said they found the directions hard to remember. As a consequence, some participants tended to ignore the gestures or look away from the computer screen altogether. Obviously, doing so would clearly impair the effect of speaker orientation to be demonstrated by the experiment. On the other hand, not all participants ignored the gestures (at least not initially) as in the 180° version, some participants declared that they found the mirrored gestures annoying.

5.2 Naturalness of the route description

In Table 2, test results on the naturalness of the route description are shown for speaker orientation and gender. Orientation had an almost-significant effect on participants' judgement of naturalness (two-way ANOVA; $F(1,45)=3.35$, $p=0.07$ two-tailed).[4] The effect would have been significant if it had been the other way around. The effect of gender was not significant, and neither was the interaction of version and gender.

Contrary to our hypothesis, the participants judged the 180° version as being more natural than the 120° version. This was contrary to what was expected, because 'in the real world' route providers and seekers tend to minimize the difference in their orientation. In fact, as mentioned above, several participants reported being annoyed by the mirrored gestures in the 180° version. These contradictory findings suggest that it was not the route provider's gestures or their orientation that were crucial for the judgement on naturalness, but only whether the route provider's body was fully turned toward his audience – directly addressing them – or not. This may be the result of many previous confrontations with presenters (human or other) displayed on television or computer screens, explaining things to an audience. Perhaps the natural tendency to make orientations as similar as possible when explaining a route to someone does not transfer to a situation where the route is presented by somebody on a screen: a form of presentation in which we expect someone to be facing us.

Furthermore, the fixed position of the camera during the route description may also have interfered with its naturalness. If the route provider points into some direction, we tend to turn our heads to that direction, maybe in the assumption he will point at some landmark that can help us orientate or navigate. The fixed position of the camera, in contrast with the adaptive orientation of the route provider, may have yielded an unnatural combination in the case of the 120° version of the route description.

5.3 Gender effects

For both versions of the route description, women performed better than men. Although not significant, the difference in performance is sufficiently remarkable to merit some discussion. We believe the difference may be explained by the fact that women and men employ different strategies for way find-

[4] A two-tailed test was performed in spite of our one-sided hypothesis 2, because the effect was contrary to what was expected.

	120°	180°	Total
Male	2.62 (1.26)	1.73 (0.91)	2.21 (1.18)
Female	2.75 (1.14)	2.46 (1.13)	2.60 (1.12)
Total	2.68 (1.18)	2.13 (1.08)	2.41 (1.15)

Table 2: Naturalness as a function of gender and version (results are presented as Means with Std. Deviations in brackets).

ing (Hunt and Waller, 1999): women's strategies are most suited for tracking and piloting, whereas men use strategies appropriate for navigation. Tracking is a point-to-point way finding strategy that relies on information limited to environmental characteristics along the route. Piloting combines these environmental characteristics with self-centered orientation and direction (e.g., "When you're facing the main entrance, turn to the right"). Navigation, on the other hand, uses configurational information: routes are derived from knowledge of the surroundings of the destination or its global position. Thus, men tend to pay attention to bearings while women often rely on descriptions of control points and cues to the route such as landmarks (Lawton, 1994).

Looking at the set-up of our experiment, we see that it seems to favour a strategy of point-to-point decision making instead of relying on a more general and global sense of direction, as in navigation. First, the route description consisted entirely of landmarks to identify decision points and turns to be made when encountering them, fitting a tracking and piloting approach to way finding. Second, both the route description and the traversal segments were shown on a screen, with a restricted and forced field of vision. This may have impeded the estimation of global position, direction and distance, i.e., the kind of spatial knowledge men rely on for orientation and way finding. So, the way finding strategy that women already tend to employ in everyday life may have been most suited to this experiment and hence their higher score.

6 Conclusions and future work

The goal of this study was to find out which orientation of the Virtual Guide would be most effective and natural for providing route descriptions in a virtual environment. To test effectiveness, we devised a method that allowed participants to 'virtually' traverse a route by watching pre-recorded route segments and making turn decisions at intersections. We hypothesized that a speaker orientation of 120° with respect to the route seeker would result in a more effective and natural route description than a 180° orientation, because it would take the route seeker less effort to match the speaker's gestures with his or her own perspective. However, we found no effect of speaker orientation on task performance. A possible explanation lies in the complexity of our route description, which caused some participants to focus only on the verbal part of the description. Contrary to our expectation, the 180° orientation was judged to be more natural, in spite of the fact that some participants found the mirrored gestures annoying. The reason for this may be that people expect a speaker to be directly facing them when presenting information on a screen.

Based on these results, we decided to stick to the standard 180° orientation for our Virtual Guide. However, some reservations are in order when applying the results of our study to the Virtual Guide. For one thing, the route descriptions used in the experiment were not given by an agent but by a real human, albeit pre-recorded. This is still far from the situation in which an embodied agent is communicating with a user by means of an interface. A second difference with the Virtual Guide lies in the participant's navigational control. In the context of the Virtual Guide, the user can actively navigate through, and look around in, the environment to be traversed. In our experiment, the participants' view was restricted and forced by that of the camera which severely restricted their possibilities for orientation and navigation.

An obvious line of future research is therefore to repeat our experiment with the Virtual Guide, and have participants actually traverse the route by navigating through the 3D virtual environment, with total freedom of movement. This will make the traversal part more realistic and also more suitable for male way finding strategies, thus providing a better and more neutral measure for the effectiveness of the route description. In addition, we expect that the participants will be less inclined to see the guide as a kind of TV presenter and more as a real presence, because they will (virtually) share the same 3D environment with it. This may lead the participants to be less biased toward a 180° orientation of the route provider. Finally, all information not strictly necessary for way finding will be left out of the route description. This includes landmarks located along traversal segments rather than at intersections, and instructions to go 'straight ahead' (which several participants found confusing in the current experiment). With a less complex description, participants may refrain from ignoring the gestures made by the route provider and thereby be more susceptible to manipulation of speaker orientation.

Acknowledgements

The authors would like to thank Mark Tempelman and Job van den Wildenberg for their help with the experiment. The Virtual Guide was implemented by Dennis Hofs, Rieks op den Akker, Marco van Kessel, Richard Korthuis and Martin Bouman. The research reported here was carried out within the context of the project ANGELICA (A Natural-language Generator for Embodied, Lifelike Conversational Agents) sponsored by the Netherlands Organisation for Scientific Research, NWO (grant number 532.001.301).

References

A. Bangerter and E. Chevalley. 2007. Pointing and describing in referential communication: When are pointing gestures used to communicate? In *Proceedings of the Workshop on Multimodal Output Generation (MOG 2007)*, pages 17–28.

J. Cassell, D. McNeill, and K.E. McCullough. 1999. Speech-gesture mismatches: Evidence for one underlying representation of linguistic and non-linguistic information. *Pragmatics and Cognition*, 7(1):1–33.

R. Dale, S. Geldof, and J. Prost. 2005. Using natural language generation in automatic route description. *Journal of Research and Practice in Information Technology*, 37(1):89–105.

K. Fujii, S. Nagai, Y. Miyazaki, and K. Sugiyama. 2000. Navigation support in a real city using city metaphors. In T. Ishida and K. Isbister, editors, *Digital Cities*, Lecture Notes in Computer Science 1765, pages 338–349. Springer-Verlag, Berlin Heidelberg.

D. Hofs, R. op den Akker, and A. Nijholt. 2003. A generic architecture and dialogue model for multimodal interaction. In P. Paggio, K. Jokinen, and A. Jnsson, editors, *Proceedings of the 1st Nordic Symposium on Multimodal Communication*, volume 1, pages 79–91, Copenhagen. CST Publication, Center for Sprogteknologi.

E. Hunt and D. Waller. 1999. Orientation and wayfinding: A review. ONR technical report N00014-96-0380, Office of Naval Research, Arlington, VA.

S. D. Kelly, D. Barr, R.B. Church, and K. Lynch. 1999. Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory and Language*, 40:577–592.

A. Kendon. 1994. Do gestures communicate? a review. *Research on Language and Social Interaction*, 27(3):175–200.

S. Kopp, P. Tepper, K. Striegnitz, and J. Cassell. in press. Trading spaces: How humans and humanoids use speech and gesture to give directions. In T. Nishida, editor, *Engineering Approaches to Conversational Informatics*. John Wiley and Sons.

C.A. Lawton. 1994. Gender differences in wayfinding strategies: Relationship to spatial ability and spatial anxiety. *Sex Roles*, 30(11-12):765–779.

P. Michon and M. Denis. 2001. When and why are visual landmarks used in giving directions? In D.R. Montello, editor, *Spatial Information Theory. Foundations of Geographic Information Science: International Conference, COSIT 2001*, Lecture Notes in Computer Science 2205, pages 292–305. Springer-Verlag, Berlin Heidelberg.

E. Morsella and R. Krauss. 2004. The role of gestures in spatial working memory and speech. *American Journal of Psychology*, 117(3):251–270.

B. Rimé and L. Schiaratura. 1991. Gesture and speech. In R. Feldman and B. Rimé, editors, *Fundamentals of Nonverbal Behavior*, pages 239–281. Cambridge University Press, Cambridge.

A.L. Shelton and T.P. McNamara. 2004. Spatial memory and perspective taking. *Memory and Cognition*, 32(3):416–426.

H. van Welbergen, A. Nijholt, D. Reidsma, and J. Zwiers. 2006. Presenting in virtual worlds: Towards an architecture for a 3d presenter explaining 2d-presented information. *IEEE Intelligent Systems*, 21(5):47–53.

M. White and T. Caldwell. 1998. EXEMPLARS: A practical, extensible framework for dynamic text generation. In *Proceedings of the Ninth International Workshop on Natural Language Generation*, pages 266–275.