Knowledge Representation for Generating Locating Gestures in Route Directions

Kristina Striegnitz, Paul Tepper, Andrew Lovett, Justine Cassell

ArticuLab, Northwestern University

{kris,ptepper,andrew-lovett,justine}@northwestern.edu

Abstract

When humans give route directions, they use gestures to indicate the location of landmarks. The form of these gestures reflects that speakers take one of several perspectives when producing them. They may locate the landmark with respect to the speaker, with respect to the person following the route, or with respect to other landmarks. A corpus study shows that which perspective is chosen is partly determined by the function of the discourse segment these gestures occur in. Since locating gestures are so prevalent in direction-giving, in this paper we address the kinds of dialogue information and knowledge representation that is needed to generate them automatically.

1. Introduction

Route directions commonly consist of descriptions of paths, followed by key landmarks, and then a suggested re-orientation of the body, followed by the next path description (Denis, 1997). When giving route directions, humans use gestures for a variety of purposes, such as indicating turns and movement direction, to describe the location of landmarks, and to depict their shape. In previous work (Kopp, Tepper & Cassell, 2004), we have studied how gestures are used to describe the *shape* of landmarks and how such gestures can be generated in an embodied conversational agent (ECA). In this paper, we look at the way humans use gesture to indicate the *location* of landmarks. Emmorey, Tversky, and Taylor (Emmorey, Tversky & Taylor, 2001, Taylor & Tversky, 1996) have found that people change between different perspectives when giving directions. We examine the use of these different perspectives in our data (Section 2). Next, we formulate requirements on knowledge representation for generating such gestures in an ECA (Section 3), and propose a way of implementing these requirements (Section 4). We then sketch how this information is used in a direction giving ECA (Section 5). Finally, Section 6 relates our results to previous work before we conclude in Section 7.

2. Gestures in Direction Giving Dialogues

2.1 Data

The observations described in this paper are based on 28 videos of people giving directions across Northwestern University's campus to another person who (they believe) is unfamiliar with the campus. In addition to transcribing the speech, we have identified and annotated gestures referring to landmarks, annotated them with their referents (a basic name for what they seem to depict) and information about the perspective used (as described below). Utterances have, furthermore, been marked for the dialogue moves that they accomplish.

Most importantly for this paper, utterance units have been marked for the instruction type that they carry out and for their dialogue function. Tables 1 and 2 list the instruction types and dialogue functions that we use. Some utterance units do not pertain directly to the task because they are incomplete, uninterpretable or meta-communication about the task or conversation. All other units are annotated with at most one instruction type and as many dialogue functions as apply. We coded 753 utterance units by the person giving the directions and 234 utterance units by the person receiving the directions in this way.

move/move+lm	instruction to proceed/to proceed wrt. a landmark
reorient/reorient+lm	instruction to reorient/to reorient at a landmark
lm	mentioning of a landmark (with no turning or proceeding instructions)
dir	„pure“ directional information (no instruction, no landmark mentioned)

Table 1: Instruction type dialogue acts.

clarif	clarification questions
a-clarif	answer to clarification question
repeat	marks all utterance units in redescriptions of a segment of the route which are not part of an answer to a clarification question
elab	marks all utterance units which are part of an elaboration on a landmark or action
ack	backchannel feedback
req-feedback	requests for feedback
ask	questions which are not clarification questions
answer	answers to ask moves

Table 2: Dialogue functions.

2.2 Perspective of Locating Gestures in Direction Giving Dialogues

The literature on route description discusses two perspectives that people use for describing space along the route (Taylor & Tversky, 1996). In *route* perspective, landmarks are described in the frame of reference of a person walking the route. The *survey* perspective is like a birds-eye view. Buildings are described relative to each other or to an absolute frame of reference (e.g. cardinal directions). These two different perspectives are also reflected in the gestures that accompany speech (Emmorey, Taylor & Tversky, 2001), and we find examples of both perspectives in our data.

Figure 1: “on your left once you hit this parking lot [is the Allen Center]”

Figure 2: “and [it’s really big]”

In our data, we also find gestures that do not fall into these two categories. First, we find gestures that seem to be purely shape depicting, and do not refer to the location of the referent landmark at all. Second, we find gestures which locate the object with respect to the speaker’s actual position and orientation.

Figure 1 shows an example of a gesture where the speaker takes on the perspective of the person following the route (the route perspective). He speaks and gestures as if he has the position and orientation that an imaginary direction-follower would have at this point along the route. Therefore, the location of his gesture (to the left of his body) corresponds to the location of the landmark relative to the location and orientation of the imaginary direction-follower. This perspective is by far the most common in our data (54.1% of all gestures referring to landmarks).

Another way in which people use their hands and the space around their bodies is to lay out virtual maps, as shown in Figure 3. Map gestures are unique in that after one gesture is made, the hand is held in place, while the next location is depicted relative to the first, by placing the other hand relative to the position of the first. As Figure 3 illustrates, the right hand representing University Hall is the anchor, held in exactly the same position throughout the three-gesture sequence, while the locations of Kresge and Harris Hall are shown relative to it. The virtual map is oriented such that it matches up with the direction a person walking the route would be facing. 16.3% of the landmark depicting gestures in our data are map gestures.

It is important to note that gestures referring to landmarks do not necessarily have a locating function. For

Figure 3: ‘[University Hall] is on your right, [on the left is Kresge], and [then straight ahead is Harris]” (The order of the pictures from left to right corresponds to the order of the three gestures.)

example, after having located the Allen Center to the left of the direction-follower, the speaker in Figure 1 continues by saying *and it's really big*. He accompanies this elaboration with the gesture shown in Figure 2, which refers to the landmark's shape by indicating its horizontal extent. This gesture does not locate the landmark to the left, which would be its position with respect to the point of view assumed for the previous utterance, but places it in front of the speaker's body. In our data, 15.8% of the gestures referring to landmarks are of this non-locating kind.

However, often gestures are neither purely locating nor purely shape depicting. For instance, the gesture used in Figure 1 seems to indicate the wall of the building being described, as the shape of the hand is flat and vertically oriented. It thus has a shape depicting component in addition to its locating function. In this paper, we are concerned with the locating function of gesture and will not address the issue of how to determine which shape features to depict and how to depict them (but see Kopp et al., 2004 for more on these questions).

Finally, gestures may be used to locate objects with respect to the speaker. That is, the speaker simply points to a real object. This type of gesture is extremely rare in our data (only 1.9% of all gestures referring to landmarks fall in this class). Table 3 shows the distribution of perspective among gestures referring to landmarks in our set of direction giving dialogues.

perspective	#	%
route perspective	199	54%
survey perspective	60	16%
non-locating	58	16%
locating wrt. speaker	7	2%
unclear/ambiguous	44	12%
	368	100%

Table 3: Distribution of perspective among gestures referring to landmarks.

2.3 Perspective and Dialogue Structure

In order to generate locating gestures with different perspectives, we must address the following question: “When are the different perspectives used?” As the following results show, the use of these perspectives seems to be determined in part by the dialogue move that the speaker is trying to perform. Tables 4-6 show the distribution of route perspective gestures, survey perspective gestures, and non-locating gestures with respect to dialogue moves.

The route perspective seems to be the default perspective when gesturing about landmarks. Survey perspective gestures are used most frequently in answers to clarification questions and in redescriptions of route segments. This fits findings of a previous study on direction-giving, which differed from our own in that the subjects could use a physical map (Cassell et al., 2002). Here, subjects only refered to the map if their purely verbally given directions were not understood.

dialogue act	#	%
lm (not a-clarif, elab, repeat)	76	38%
a-clarif (incl. 26 UUs marked as lm)	31	16%
repeat (incl. 45 UUs marked as lm)	12	6%
elab (incl. 10 UUs marked as lm)	46	23%
other	34	17%
	199	100%

Table 4: Distribution of **route perspective gestures** wrt. dialogue acts.

dialogue act	#	%
a-clarif	32	53%
repeat	16	27%
elab	7	12%
other	5	8%
	60	100%

Table 5: Distribution of **survey perspective gestures** wrt. dialogue acts.

dialogue act	#	%
elab	51	88%
other	7	12%
	58	100%

Table 6: Distribution of **non-locating gestures** wrt. dialogue acts.

Non-locating gestures occur mainly in elaborations. That is, if after having introduced a landmark, speakers give further information about the shape of this landmark, as opposed to location or location combined with shape. This is reflected in their gestures, in which the locating component may be absent or deemphasized.

3. Requirements on Knowledge Representation

To generate any kind of route description, a map of the relevant area is needed. Minimally, the map must include the paths that can be taken, so that the system can calculate the route. Unlike direction-giving systems like MapQuest, our system gives directions using landmarks to indicate reorientation points and other important points along the path. Therefore, we also need a representation of the landmarks located along these paths.

As demonstrated in Section 2, generating gestures referring to these landmarks requires information about position and orientation relative to the map of both the imaginary person following the route and the speaker, as well as mechanisms for inferring spatial relations between the entities in the map representation. First, to generate route perspective and survey perspective gestures (Figures 1 and 3), we need to keep track of the position and orientation that a person following the route would have at each point of the description. Second, we need the position and orientation of the (virtual) person giving the directions. This is needed to generate gestures which locate landmarks relative to the speaker. Additionally, the system requires mechanisms for inferring spatial relations between the entities in the representation. For example, the system needs to be able to infer the location of landmarks relative to paths, other landmarks, the speaker, and the direction-follower. It also needs to know how to calculate the orientation of these entities with respect to landmarks or cardinal directions. This is necessary in order to locate gestures referring to landmarks correctly in the gesture space. It is also necessary for the decision which landmarks to mention in the route description; landmarks that are mentioned at a specific point in the description should be visible to the direction-follower when he/she reaches the corresponding point of the route.

Generating virtual-map gestures, as shown in Figure 3, additionally requires the following information. First, it requires a way of mapping any part of the map representation onto the gesture space in front of the speaker. This allows us to associate landmarks with hand locations in such a way that the relative location of the hands reflects the relative location of the landmarks in the world. Second, the discourse history has to contain information about the current location of the hands and which landmark they stand for such that multimodal anaphoric expressions can refer back to these landmarks in later utterances.

Finally, landmarks and paths must be associated with semantic information. For instance, a description of a landmark could draw upon information about its name, type (building, lake, monument, etc.), size, color, and shape. For paths, we may specify what type of path it is, a street, parking lot, courtyard, etc. This information is necessary for generating descriptions of landmarks together with gestures depicting their shape and/or size. In the next section, we propose a way of implementing the knowledge requirements formulated above in an ECA.

4. Locating Landmarks in Space

The basis for generating locating gestures is a map representation consisting of two interlinked components: (i) a graph, where edges represent the paths that can be walked and nodes (path points) represent points on the map where the direction-follower may have to change his direction, and (ii) a set of landmarks. Landmarks are associated with areas and path points are associated with points in a common coordinate system (see Figure 4). In addition, path points can be linked to landmarks by qualitative relations specifying whether a path point is the entrance of a building or whether it is next to a landmark. Finally, landmarks and path points are associated with semantic information as described above (type of landmark, size, color, shape, etc.).

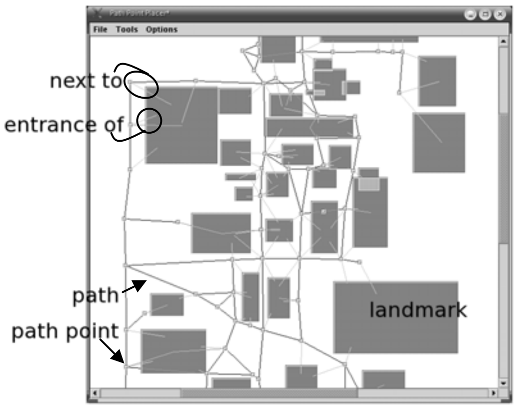


Figure 4: The map representation.

4.1. Locating Landmarks with respect to the Direction-follower’s and the Speaker’s Perspective

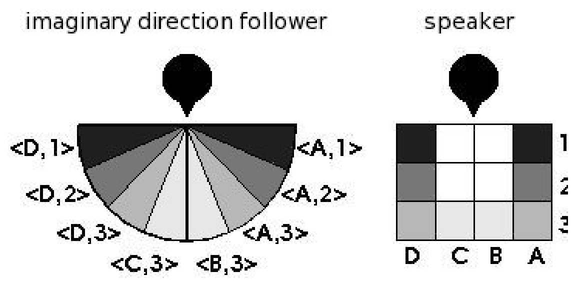
When gestures are used to locate landmarks with respect to the *direction-follower’s* point of view, they depict the landmark at a location in the gesture space. This location corresponds to the location of the landmark relative to the position and orientation that the direction-follower would have in the world at that moment if he/she were walking the route. This holds whether it is a simple pointing gesture or a gesture that depicts some aspect of the landmark’s shape, as in Figure 1. In order to generate such gestures, we need to keep track of the position and orientation of the direction-follower in the map representation. These values change continually over the course of the dialogue, as the description (and the imaginary direction-follower) progresses along the route.

Given a route between two points on the map graph, we can derive the direction-follower’s orientation for each point along this route, based on the location of the previous point on that route To facilitate this calculation, we represent both the position and the orientation of the speaker in terms of the coordinate system used in our map representation for placing path points and landmarks. This allows us to calculate the angle at which landmarks are located with respect to the direction-follower’s orientation, which can then be mapped to different positions in the speaker’s gesture space.

Gestures that locate objects with respect to the *speaker* can be generated using the same mechanisms, given that the location and orientation of the speaker within the map representation are recorded. Note that in our current application the speaker is our ECA, which is part of a stationary information kiosk. The agent is displayed on a fixed screen, so its position and orientation remain the same over the course of an interaction.

4.2. Generating Map Gestures

In their simplest form, map gestures resemble the act of placing objects in the horizontal, tabletop plane in frt of the speaker. While they can get more complicated than this, e.g., by also depicting information about the shape of the objects, here we will just consider this basic case of positioning objects. For the spatial representation, we use a simple grid of regions called a Gesture Space schema (see Figure 5b) which represents the tabletop plane in front of



(a) (b)

Figure 5: Mapping locations relative to the position and orientation of the direction follower in the world (a) to the gesture space in front of the speaker (b).

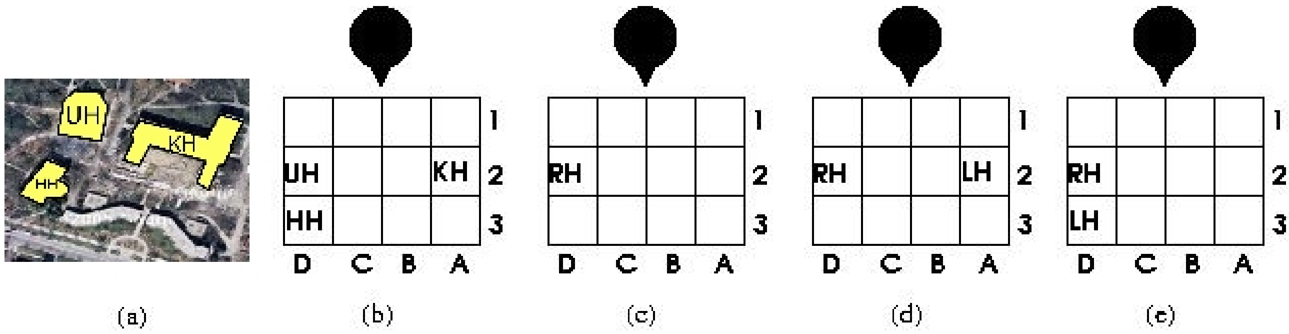


Figure 6: Spatial schemata for map gestures. (a) Map of the world, where UH is University Hall, KH is Kresge Hall and HH is Harris Hall. (b) Map translated into the schema w/r/t the speaker. (b-d) ECA's right and left hands (RH & LH) as they move

the speaker. Each map gesture depicts a limited section of the map of the world. When a map gesture is planned, the locations of the landmarks in the desired section are mapped onto the grid. This mapping associates objects in the map representation — each located at a particular angle with respect to the direction-follower's direction of view — with particular fields in the grid, as illustrated in Figure 5. For example, all landmarks that are directly in front of the direction-follower are mapped to fields *<B,3>* (if they are slightly to the left of the center) or *<C,3>* (if they are slightly to the right).

Each landmark is then positioned by one of the hands, in an order specified by the content planner. Figure 6a shows the area of campus talked about in the example of Figure 3, and Figure 6b shows how it is mapped onto the grid of the Gesture Space schema (assuming that the imaginary direction-follower is at the top of the map and is looking downwards). Figures 6c-e show the sequence of states accompanying the three gestures in Figure 3. In each of these states, the current positions of the agent's hands are indicated by RH for right hand and LH for left. In this example, the right hand is the anchor, i.e., the hand which stays stationary across the three states. Numbering the grid regions using matrix notation, the information in (c), for example, can be read as *RH remains in <2,1>, LH moves to <2,3>*.

Next, we use a data structure called Gesture Representation Structure (GRS) to link the agent's hands to both their location and the entities that they represent. Each hand is represented by a pair comprising the current location and the referent it is representing. Location is itself a pair of coordinates, recording a location in the Gesture Space schema. Discourse referents are logical terms denoting the entity that the gesture refers to. The GRS is a pair of these pairs, RH and LH, one for each hand. Figure 7 shows this data as a feature structure.

The Gesture Representation Structure is stored in the Information State, as part of the dialogue context. It

$$\text{GRS:}\begin{bmatrix} \text{LH:} \begin{bmatrix} \text{Loc:}\langle D, 2\rangle \\ \text{Referent: universityHall} \end{bmatrix} \\ \text{RH:} \begin{bmatrix} \text{Loc:}\langle A, 2\rangle \\ \text{Referent: kresgeHall} \end{bmatrix} \end{bmatrix}$$

Figure 7: Gesture Representation Structure, part of the NUMACK discourse context.

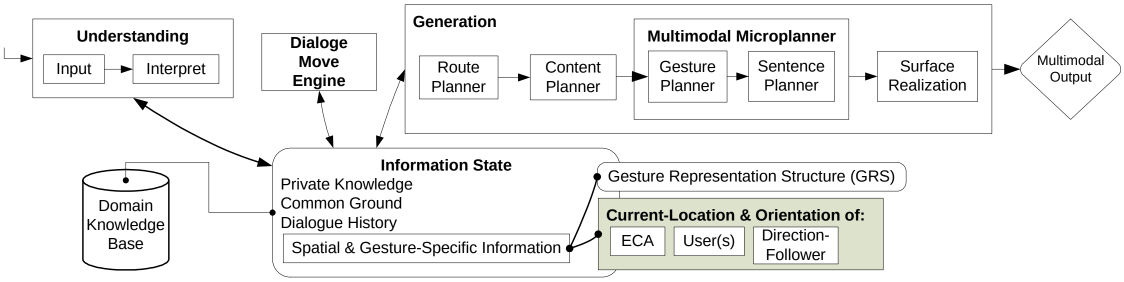


Figure 8: Architecture of a direction giving ECA.

gets updated as the relations between hands, locations, and landmarks change. This allows later utterances to make use of the information, e.g., in order to generate appropriate multimodal anaphoric references to landmarks, where the ECA continues using the same hand and location to refer to the same landmark as long as the direction-follower's position and orientation remains stable.

5. Architecture of a Direction Giving ECA

Now, we move on to describing the architecture of our ECA, as illustrated in Figure 8. First, we discuss the dialogue management module, or Information State. Next, we describe the content planner, which includes a route planner that works from a map representation specialized for gesture and natural language generation (see Section 4). The content planner also determines the perspective used in each gesture. Lastly, we give a brief description of the multimodal microplanner and surface realization components.

At the center of the system is the Information State (Traum & Larsson, 2003). This is a data structure that keeps track of the dialogue history, the private knowledge of the system as well as the shared knowledge of user and system, and the current state of the system. In addition to this kind of information, which is commonly found in any Information State, we also use the Information State to store the output of the content planner, and to keep track of the point in the route the description has reached. We are currently working on integrating the Gesture Representation Structure into the Information State.

The Dialogue Move Engine determines how to integrate user dialogue moves into the Information State and chooses the moves of the system. We use the MiDiKi system (Burke et al., 2003) for maintaining the Information State and for specifying the rule system of the Dialogue Move Engine.

Once the system has determined where the user wants to go and where he wants to leave from, the route planner calculates the shortest path between these two points. The map representation that the route planner currently works with has been coded by hand. Ultimately, we would like to automatically derive the necessary information from existing sources of geographic information. The output of the route planner is a sequence of path points and the task of the next step, i.e., content planning, is to map this to a sequence of preverbal messages, which can then be

(a) (b) (c)

Figure 9: NUMACK, our ECA, producing (a) a route perspective gesture, (b) a non-locating gesture, (c) a survey perspective gesture.

turned into multimodal utterances by the multimodal microplanner. More specifically, the content planner (i) chooses which path points to mention, (ii) decides which instruction types[1] to use for describing each step in the route, (iii) selects landmarks that can be used to identify path points to the user, and then (iv) determines the content of the expressions referring to those landmarks. In step (iv), the content planner chooses the properties of the landmark that need to be expressed either in the language or in gesture to distinguish the landmark from its surroundings. It also determines the perspective that should be used with respect to gesture.

So, it is in these last two steps that the data structures described in the previous sections come to bear. By default, the system assumes the route perspective. Figure 9a shows an example of a route perspective gesture, which accompanies the words "*Pass the Allen Center on your left.*" Non-locating gestures are only used in elaborations on landmarks that do not mention the location of that landmark (e.g., Figure 9b: "*Dearborn Observatory is the building with the dome*"). Currently, we only generate survey perspective gestures that lay out a virtual map in front of the speaker. That is, these gestures depict the relative locations of landmarks, but not other features like shape or path trajectories relative to landmarks. Figure 9c shows an example. The accompanying speech is "*Annenberg Hall is here and the Seminary is here*" where the first occurrence of *here* refers to the position of the right hand and the second one to the left hand.

The output of the content planner is a plan of the route description which specifies the structure of the text and the semantic content that needs to be expressed by each utterance. It is stored in the information state. Based on user feedback, the dialogue manager chooses when to send the next utterance specification to the microplanning and realization modules. The multimodal microplanner determines the form of the utterance, including the actual words as well as the form of the gestures and the coordination between language and gesture (Kopp et al., 2004). Finally, the

[1] Reorientation, reorientation with respect to a landmark, move, move with respect to a landmark etc.; see the description of dialogue moves in Section 2.

surface realization component turns the utterance specification produced by the microplanner into speech and movements of the animated character on the screen (Kopp & Wachsmuth, 2004).

6. Related Work

Most literature on deictic gestures in multimodal interfaces is on the *interpretation* of such gestures (see, e.g., Bolt, 1980, Johnston & Bangalore, 2000). There are systems which *generate* deictic gestures, such as the COMIC system (Foster, 2004), DFKI's PPP Persona (André, Rist & Müller, 1998), but these systems only handle pointing gestures that point to objects presented on the screen. They are, hence, what we have called gestures that locate objects with respect to the speaker. To the best of our knowledge, there is no system which assumes different perspectives to generate locating gestures referring to objects not immediately available in the environment of the virtual agent.

Another body of research that is relevant to our application is the existing work on generating natural language route descriptions. For example, Dale, Geldof & Prost (2005) generate driving directions from GIS data. Look et.al. (2005) produce walking directions, but concentrate on the representation of the information necessary for planning the route rather than the planning and realization of the natural language output. Habel (2003) concentrates on the architecture of a generation system for route directions, arguing for an incremental processing model. None of these systems model face-to-face dialogue and, hence, none of them look at generating the gestures that humans use when giving route directions.

7. Conclusions and Future Work

Previous work on human face-to-face dialogue has shown that speakers assume different perspectives when giving route directions (Taylor & Tversky, 1996). In particular, they use the route perspective, which refers to landmarks with respect to an imaginary direction-follower’s point of view, and the survey perspective which locates landmarks using a birds-eye view. Our data supports this finding and also shows that, in addition to route perspective and survey perspective gestures, people use non-locating gestures and gestures that locate landmarks with respect to the speaker’s point of view. The distribution of these gestures is partly determined by the dialogue move of the utterance they occur in. Our goal is to model the different uses of locating gestures in a direction giving ECA in order to produce route descriptions which are more natural and easier to understand. To the best of our knowledge the issue of perspective in locating gestures has never been addressed with the aim of generating such gestures in a virtual agent.

In this paper, we have discussed the knowledge necessary for generating such gestures and we have proposed a way of representing this knowledge in an implemented system. More specifically, we have argued that we need a suitable map representation (representing not only the paths that can be walked on but also landmarks in relation to these paths as well as additional semantic information about properties of paths and landmarks) and that we have to be able to keep track of the position and orientation of entities in this map (i.e., landmarks as well as the direction-follower and the speaker). This information is necessary for generating route perspective and survey perspective gestures as well as gestures that locate a landmark with respect to the speaker's point of view. In the case of map gestures, the position of the speaker's hands needs to be recorded, linked to landmarks, and this information needs to be appropriately updated as the discourse proceeds.

The proposal made in this paper is implemented in a direction giving ECA. Our next step will be to use this system to evaluate our proposal and to inform its further development.

Acknowledgements: We would like to thank Magan Sethi and Laura Cristina Stoia for their help with implementation. This research was supported by a fellowship from the Postdoc-Program of the German Academic Exchange Service (DAAD).

References

André, E., Rist, T. & Müller, J. (1998). WebPersona: A Life-Like Presentation Agent for the World-Wide Web. *Knowledge-Based Systems*, 11(1), 25-36.

Bolt, R. (1980). “Put-that-there”: Voice and gesture at the graphics interface. In *Proceedings of the 7th annual conference on Computer graphics and interactive techniques*, (pp. 262-270).

Carl Burke, Christy Doran, Abigail Gertner, Andy Gregorowicz, Lisa Harper, Joel Korb, Dan Loehr. Dialogue complexity with portability? Research directions for the Information State approach. In *the Research Directions in Dialogue Processing Workshop at the 2003 HLT-NAACL/NSF Human Language Technology Conference.*

Cassell, J., Stocky, T., Bickmore, T., Gao, Y., Nakano, Y., Ryokai, K., Tversky, D., Vaucelle, C., Vilhjálmsson, H. (2002). MACK: Media lab Autonomous Conversational Kiosk. *Proceedings of Imagina02*. February 12-15, Monte Carlo.

Dale, R., Geldof, S., & Prost, J.-P. (2005). Using Natural Language Generation in Automatic Route Description. *Journal of Research and Practice in Information Technology,* 37(1), 89-105.

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

Emmorey, K., Tversky, B., & Taylor, H. A. (2001). Using space to describe space: Perspective in speech, sign, and gesture. *Spatial Cognition and Computation, 00*, 1-24.

Foster, M. E. (2004). Corpus-based Planning of Deictic Gestures in COMIC. In Belz, A.; Evans, R. & Piwek, P. (Eds.), *Proceedings of the Third International Conference on Natural Language Generation*, (pp. 198-204), Springer, Lecture Notes in Computer Science, Vol. 3123.

Habel, C. (2003). Incremental Generation of Multimodal Route Instructions. In Proceedings of the AAAI Spring Symposium on Natural Language Generation in Spoken and Written Dialogue, (pp. 44-51).

Johnston, M. & Bangalore, S. (2000). Finite state multimodal parsing and understanding. In *Proceedings of the International Conference on Computational Linguistics (Coling)*, (pp. 369 – 375).

Kopp, S., Tepper, P., & Cassell, J. (2004). Towards Integrated Microplanning of Language and Iconic Gesture for Multimodal Output. In *Proceedings of the International Conference on Multimodal Interfaces*.

Kopp, S. & Wachsmuth, I. Synthesizing Multimodal Utterances for Conversational Agents. *The Journal Computer Animation and Virtual Worlds.* 15(1): 39-52, 2004.

Larsson, S. (2002). *Issue-Based Dialogue Management.* PhD thesis, Goteborg University.

Look, G., Kottahachchi, B., Laddaga, R. & Shrobe, H. (2005). A Location Representation for Generating Descriptive Walking Directions. In *IUI '05: Proceedings of the 10th International Conference on Intelligent User Interfaces*, (pp. 122-129).

Taylor, H. A., & Tversky, B. (1996). Perspective in spatial descriptions. *Journal of Memory and Language, 35*, 371-391.

Traum, D. and Larsson S. (2003). The Information State Approach to Dialogue Management. In Smith & Kuppevelt (Eds.), *Current and New Directions in Discourse and Dialogue*, (pp. 325-353). Kluwer.