Spatial Cognition 2004:
Poster Presentations

Frauenchiemsee, Germany, October 2004

Thomas Barkowsky, Christian Freksa, Markus Knauff,
Bernd Krieg-Brückner, Bernhard Nebel (Eds.)

SFB/TR 8 Report No. 002-10/2004
Report Series of the Transregional Collaborative Research Center SFB/TR 8 Spatial Cognition
Universität Bremen / Universität Freiburg

Contact Address:

Dr. Thomas Barkowsky
SFB/TR 8
Universität Bremen
P.O.Box 330 440
28334 Bremen, Germany

Tel +49-421-218-8625
Fax +49-421-218-8620
barkowsky@sfbtr8.uni-bremen.de
www.sfbtr8.uni-bremen.de

Table of Contents

Topological operators for ontological distinctions: disambiguating the geographic concepts of *place*, *region* and *neighbourhood*

Pragya Agarwal

School of Computing, University of Leeds, Leeds, UK.

1.1 Objectives

Problems related to the semantic ambiguity between geographic concepts for modelling of geographic, especially cognitive conceptualisations, has been outlined in a related work (Agarwal in press). Experiments conducted to disambiguate such relations were also presented in this publication. The aim of the work presented here is to assess whether it is possible to specify basic topological operators for spatial concepts from cognitive conceptualisations (such as for *region*, *neighbourhood* and *place* used in this experiment). The experiment was also designed to assess whether techniques, like those employed for formalising spatial and linguistic expressions, can be applied to enhance understanding of ontological distinctions between spatial concepts based on topological properties. The primary hypothesis for the experimental work presented here is that cognitive schema can be used for externalising the properties and relations of concepts, both representational and topological.

1.2 Methods

Empirical tests using map-based experiments are developed to test shape and scale independent prototypical relations between cognitive conceptualisations of spatial entities. The principles of topological operators and prototypical relations are used to define the core metric and topological associations between conceptualisations of place, region and neighbourhood. In doing so, the basis for formal models of these concepts is defined. In addition, such methods are tested to define ontological distinctions between these concepts, and hierarchical relationships between spatial entities that are primarily of the mesoscopic kind. Experimental work was based in real-world setting and methods were tested and validated against a sample size of N=50. Random assignment was carried out and experimental framework was determined so as to minimise bias. This experiment was designed to externalise geographical knowledge within an egocentric framework, where individual estimates and conceptual structures were stimulated and externalised with the help of real world information. The design of the experiment was map-based using a map of Nottingham, UK as the primary cue. The context of 'live in' was used to ground the individual conceptualisations within a standardised context. This context developed from an egocentric, home-based, personal conceptualisation of place that emerged from theoretical literature and models for place. The respondents were not constrained to any specific shape associations in the mapping process. Although it is possible that the scale and design of the base map provided as the primary cue to the respondents and the context of 'live in' imposed constraints on the shape and relations that were conceptualised, this experiment was developed to indicate a possible set of relations that could then be tested for applicability and extended by applying similar methods to different contexts. This experimental design was also developed to provide a test-bed for the kind of empirical methods that can be developed for use with human subjects, for operationalising spatial relations from the cognitive schema, and for providing indications of how these methods can be incorporated in developing formal models and theories for spatial concepts.

The respondents were not constrained to any specific shape association. The drawings were scanned and incorporated in a desktop GIS. Using automated procedures (such as dot-density function), the different sets of drawings were merged for each of the concepts (*region*, *place* and *neighbourhood*), and then overlaid for each subject to determine the topological relationship for each. Each set of relationships were then overlaid to determine the generalised prototypical relations that emerged by correlating edges and centres. The relative size of the individual entities, and the angle and orientation of the relations were not primary considerations in this conceptualisation, because the emphasis was on understanding the 'containment', 'coincidence' and 'adjacency' operators conceptualised between these concepts when they are operationalised as spatial entities in the real world.

1.3 Results

Prototypical categories were seen in the set of relations defined by the respondents. Despite the specific nature of the relations and prototypical operators that resulted, the analysis shows that a set of prototypical relations forming the 'core' associated with the semantics of the concepts can be extracted from cognitive schema through human subject testing, thereby supporting the hypothesis proposed. It is also suggested from the results that *place* and *neighbourhood* are conceptualised as lower-order concepts than *region*, and are linked by 'intersects' and 'containment' relations. *Place* and *region* are also shown to have 'containment' and 'adjacency' relations, and indicated that *neighbourhood* is conceptualised as a lower-order concept to *place*. The results provide a basic set of relations that give indications of the hierarchical organisation of *place*, *region* and *neighbourhood*, semantically as well as topologically. This basic set of relations can form the basis for further testing to develop any formal set-theoretic relations and models for *place*, rather than applying models that have been developed for geographic *regions* to concepts such as *place* and *neighbourhood* that have a higher cognitive dimensionality. It was shown that cognitive schemas also contain indications to the geometrical and topological relationships specified on geographic concepts in commonsense notions. From a map-based experimental design, it was demonstrated that the primary theory for concepts constitutes, not only the semantic associations, but also a core set of topological relations that can be defined by operationalising the geographic concepts as spatial entities. For an exhaustive set of relations, further dimensionalities of concepts and complex arguments for contexts must be considered. It is proposed that this work be carried out, to validate and extend the set of prototype relations obtained from this experiment, using methods applied here in combination with other experimental protocols based on descriptive and selective design principles. The results from the experiment described in this paper have significant implications for the way a formal model of *place* can be developed and how set-theoretic arguments and intersection models can be applied to the representation of *place*.

The Landmark Spider: Weaving the Landmark Web

David Caduff and Sabine Timpf

Department of Geography, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland
{caduff,timpf}@geo.unizh.ch
http://www.geo.unizh.ch/

Introduction

Landmarks play an important role when humans navigate through foreign environments [1,2]. For example, trying to find the way is much easier if the navigator can rely on a description of the route based on well-recognizable objects in the environment, instead of navigating solely on the basis of street names and metric directions [3]. Therefore, collecting and incorporating landmarks along a route is a crucial task of pedestrian navigation systems that aim at providing efficient and reliable route instructions. Several proposals have been made on how to automatically extract landmarks from data sets and how such landmarks could be used to enhance wayfinding instructions for pedestrians [4-6]. So far, however, the question what landmarks to consider and how to integrate them in the route generation process is only poorly understood. We propose a *landmark spider* that assesses the relevance of landmarks along a route, includes these landmarks in the route generation process, and weaves a web of landmarks along that route.

Background

Cognitive research has indicated that the complexity of route instructions may be as important in pedestrian navigation as the overall length of the route. Streeter and co-authors found that human navigators were prepared to take suboptimal routes in terms of travel time, if these routes were potentially easier to describe and to follow [7,8]. Landmarks are important elements in route instructions as they ensure efficient and reliable navigation [9]. So far, the typical approach to incorporate landmarks in route descriptions has been to enrich navigation instructions with information about landmarks [4,6,10]. Our approach is different in that it uses a subset of the available landmarks, which are most prominent and easy to find, to determine the optimal route. The introduction of landmarks in the route generation process is advantageous, since it enables us to avoid areas of low landmark density and ensures consistent access to landmark information along the complete route.

The Landmark Spider

The landmark spider consists of a route network and a set of landmarks, including their spatial attributes and non-spatial properties (salience, visibility, etc.). The network is represented as a connected, simple, directed graph. The landmark spider uses a revised version of Dijkstra's shortest weighted path algorithm to find the optimal path between two nodes. The weighting function reflects the presence of landmarks at each point along the route where navigators might need assistance. The model framing the weighting function is based on findings by Michon and Denis [11]. According to them, the three most important reasons why landmarks are required during navigation are: 1) signaling where an action should be executed, 2) creating the link to the next section of the route in terms of assisting in locating the next landmark, and 3) reassuring navigators that they are still on track. These reasons are applied at specific locations along the way, and hence, can be mapped onto any network graph representing routes pedestrians may take.

Signaling where an action should be taken occurs at places where three or more edges meet, typically at nodes, with the exception of turning around, which may happen along any given edge. However, for the purpose of this study we neglect this case and assume that any decision taken by the navigator is correct and no turning around is required. Hence, each node needs to have at least one landmark associated with it. The ideal case would be a set of landmarks that are visible from both ends of the edge connecting the nodes and which are prominent enough as to provide guidance from one decision point to the next. As the distance between nodes grows, however, one landmark eventually gets out of sight and the need arises to find the next reference point. Hence, landmarks along edges may be required. In this case, we consider the vertices of the route network, which we use as positional reference in assisting navigators to find the next landmark as well as affirming that they are still on track.

Defining the Weighting Function

Now that we know where landmarks are required along the way, we can define the weighting function. Basically, the weight is the sum of the factors that define the binary relation between points on the route and potential landmarks. These factors are 1) the distance between a point and the landmark, 2) the orientation of the user with respect to the landmark, and 3) the salience of the landmark itself. The distance is important since close landmarks make better reference points than distant landmarks. Orientation is included due to the fact that following route instructions is a directed process. Hence, the orientation of a landmark with respect to the navigator is important when referring to landmarks. For instance, landmarks located in the line of view are more likely to serve as navigational references than objects located at the navigator's back. Finally, the saliency of the landmark indicates its quality, which is in essence a measure of the 'visibility' of the landmark. As a result, the salience is an essential part of the relation between points on the route and the landmark. This configuration ensures that each node and each vertex along the route is associated with a weight indicating the relevance of potential landmarks. Introducing these weights in the weighted shortest path algorithms and summing them up results in the optimal route in terms of landmark coverage.

Summary

In the best case, the optimal route computed by the landmark spider is identical to the shortest path between starting point and destination. The worst-case scenario occurs if the density of available landmarks is too low, in which case parameters of the weighting function may have to be adjusted appropriately. Questions on results analysis and comparison, as well as computational issues and performance are subject of ongoing research. Nevertheless, the landmark spider provides a first approach to integrating landmarks in the route generation process, in contrast to linking landmarks and route description after route generation.

References

1. Lynch K: *The Image of the City*. Boston: The M.I.T. Press; 1960.
2. May AJ, Ross T, Bayer SH, Tarkiainen MJ: Pedestrian Navigation Aids: Information Requirements and Design Principles. *Personal Ubiquitous Computing* 2003, 7:331-338.
3. Tom A, Denis M: Referring to Landmark or Street Information in Route Directions: What Difference Does It Make? In *International Conference on Spatial Information Theory, COSIT*; *Kartause Ittingen, Switzerland*,Edited by Kuhn W, Worboys M, Timpf S: Springer-Verlag: 2003:362-374.
4. Nothegger C: Automatic Selection of Landmarks [Diplomarbeit]. Vienna: University of Technology: 2003.
5. Elias B: Extracting Landmarks with Data Mining Methods. In *International Conference on Spatial Information Theory, COSIT 2003*; *Kartause Ittingen, Switzerland*,Edited by Kuhn W, Worboys M, Timpf S: Springer-Verlag: 2003:375-389.
6. Raubal M, Winter S: Enriching Wayfinding Instructions with Local Landmarks. In *Second International Conference on Geographic Information Science September 25 - 28, 2002*: Springer-Verlag: 2002:243-259.
7. Streeter LA, Vitello D, Wonsiewicz SA: How to Tell People Where to Go: Comparing Navigational Aids. *International Journal of Man-Machine Studies* 1985, 22:549-562.
8. Streeter LA, Vitello D: A profile of driver's map-reading abilities. *Human Factors* 1986, 28:223-239.
9. Denis M, Pazzaglia F, Cornoldi C, Bertolo L: Spatial Discourse and Navigation: An Analysis of Route Directions in the City of Venice. *Applied Cognitive Science* 1999, 13:145-174.
10. Redish AD, Touretzky DS: Navigating with Landmarks: Computing Goal Locations from Place Codes. *Symbolic Visual Learning* 1995.
11. Michon P-E, Denis M: When and Why Are Visual Landmarks Used in Giving Directions. In *International Conference on Spatial Information Theory, COSIT 2001*; *Morro Bay, CA, USA*,Edited by Montello DR: Springer-Verlag: 2001:292-305.

Addressee-dependent Blending in Object Localisation Tasks*

Kerstin Fischer

University of Bremen – kerstinf@uni-bremen.de

1 Blending in Spatial Reference

In this study, cognitive operations will be investigated for their use in spatial object localisation tasks involving different addressees. The cognitive operations considered are conceptual metaphor (1), metonymy, and blending (2). Blending concerns the integration of two independent representations into a new structure and is particularly relevant in our scenario.

The first pervasive finding in our data is that as soon as three similar objects are close to each other, speakers refer to them as a group (3; 4). For instance, reference to an object of this group is achieved by referring to it as the left, middle, or right object, irrespective of whether they are right, middle or left from the participants. Thus, the *right object* may well be left of speaker or hearer. We can understand this as a blending operation of the concept of a group with the close spatial proximity of the objects in question, in comparison with other objects. The group is then taken as the relatum for spatial reference.

Furthermore, if the placement of the objects 'resembles' the shape of an object, speakers refer to the groups in terms of these objects, for instance, *the square, the triangle, the top row, the bottom line*. That is, objects were conceptualised as points marking the corners of imaginary objects. Yet, they are not *treated* as 'making up' these corners, but as parts of real objects, as can be seen in the following utterance: "the brick is *located* at the top corner of the triangle."

Blending of different conceptual spaces occurs at several levels. Especially the up/down dimension (even though our settings are two-dimensional, that is, objects are placed on the ground in front of the participants) can be found pervasively, for instance, "top row of group of three" or "bottom of square". According to (5), the relation between *up* as referring to a path on the vertical dimension and *up* as a goal-directed path is metonymical: the reaching of its full extension in the vertical dimension is taken to stand for the reaching of a goal in general, as in *drive up the street*. In a next step, the metonymic extension of *up/down* as a path is extended to a whole region, such that the blended geometrical figures have a top and a bottom. Thus, we find the metonymic treatment of *up* and *down* for *far* and *close* regions, as well as a blending of object locations with geometrical figures. The latter is in line with reference to, for instance, Tangram figures in conversation (6).

* This research has been supported by the German Research Foundation in the framework of the SFB/TR8 *Spatial Cognition* at the University of Bremen.

2 Addressee-dependent Cognitive Operations

The cognitive operations for spatial reference just described are quite complex. The question I would like to address now is whether speakers can be shown to be aware of this complexity and whether they adapt their instructions to communication partners who may be suspected to have problems with such conceptualisations. The example investigated here is human-to-robot communication (HRI).

2.1 Data

The set of human-robot dialogues analysed here were elicited in a scenario in which the users' task was to instruct a robot to measure the distance between two objects out of a set of seven. These objects differed only in their spatial position. The users typed instructions into a notebook, having an overview of the objects to be referred to and the robot (a joint attention scenario). The relevant objects were pointed at by the instructor of the experiments. There were 21 participants from all kinds of professions and with different experience with artificial systems. The robot's output was either "error" (or a natural language variant of it) or a distance in centimeters. Since by reformulating their utterances the users display their hypotheses about the functioning of the system (see 7), error messages were frequent. The user utterances are typed and thus transcription was not necessary; typos were not corrected.

The human-to-human dialogues used were elicited with the same object configuration. The task was to describe one of the objects so that the addressee could identify it by pointing at it. To match the fact that in the HRI dialogues users typed their utterances, participants in this experiment were instructed to write their instructions on a piece of paper, which was then handed to the addressee. Participants were placed in the same positions as in the human-robot scenario. The participants were not allowed to talk or to communicate non-verbally.

2.2 Blending in Human-Robot Interaction

Many behaviours of speakers in human-computer and human-robot interaction show that users believe their artificial communication partners to be restricted in particular ways, especially regarding natural language processing, perception, and human abilities such as identifying the intended meaning of a sentence in spite of spelling errors. Thus, we may expect that users, in case they suspect non-literal language use and blending operations to be complex, will also adapt to their artificial interlocutor by using simpler or different types of descriptions (7).

What can be found is that speakers indeed employ these cognitively complex spatial descriptions, but not to same amount as in the human-to-human setting. Furthermore, as soon as there are problems, speakers change to another way of referring to the objects in question. As the repair in the example below shows, users are uncertain whether the notion of a group, here 'collection', can be processed by an artificial system, and thus in case of problems change to some other reference system:

(1) usr15-6:messe distanz zwischen den inneren beiden Tassen der rechten Tassenansammlung [**measure distance between the two inner cups of the right collection of cups**]
sys:ERROR 652-a: input is invalid.
usr15-7:messe distanz zwischen der tasse in der Mitte und der Tasse schräg rechts dahinter [**measure distance between the cup in the middle and the cup diagonally right behind**]

The same holds for blended geometrical figures. In the next example, the blend is marked as an imagined representation, not as a real object, and even though the system signals understanding by providing a distance, the speaker employs another kind of description which does not make use of cognitively complex concepts, such as an imaginary rectangle:

(2) usr17-9:miß den Abstand zwischen den beiden Tassen, die die längere Seite eines Rechtecks darstellen würden [**measure the distance between the two cups that would represent the longer side of a rectangle**]
sys:31,1 cm
usr17-10:miß den Abstand zwischen Roboter und der Tasse, die am zweitnächsten steht [**measure the distance between robot and the cup that is second next**]

Moreover, even the metonymic *up/down* dimension which can be seen as highly conventionalised is hardly used at all. In contrast to the pervasive usage in the human-to-human situation, only five out of twenty-one users in the HRI distance measurement task employed *up* and *down* at all. These users are furthermore likely to change to another way of instructing, as in the example below:

(3) usr13-30:miss die entfernung zwischen dem objekt links oben und dem objekt rechts oben [**measure the distance between the object up left and the object up right**]
sys:Die Instruktion konnte nicht erkannt werden. Bitte formulieren Sie neu. [**The instruction could not be processed. Please reformulate.**]
usr13-31:miss die entfernung zwischen dem objekt links hinten und dem objekt rechts hinten [**measure the distance between the object back left and the object back right**]

3 Conclusion

We can conclude that the speakers' ideas about the cognitive complexity of spatial descriptions influences the strategies they take in the communication with their artificial communication partners. Speakers are aware of potential problems in the use of cognitively complex ways of talking about space by means of blended spatial expressions, such as *the top row of the square*, and they adapt the complexity of their instructions to the (supposed) capabilities of their communication partners.

Bibliography

[1] Lakoff, G., Johnson, M.: Metaphors We Live by. The University of Chicago Press (1980)

[2] Fauconnier, G., Turner, M.: The Way we Think. Basic Books (2002)

[3] Moratz, R., Fischer, K., Tenbrink, T.: Cognitive modelling of spatial reference for human-robot interaction. International Journal on Artificial Intelligence Tools **10** (2001) 589–611

[4] Tenbrink, T., Moratz, R.: Group-based spatial reference in linguistic human-robot interaction. In: Proceedings of EuroCogSci 2003: The European Cognitive Science Conference, September 10-13, Osnabrück, Germany. (2003) 325–330

[5] Lindner, S.J.: A Lexico-Semantic Analysis of English Verb-Particle Constructions with *out* and *up*. PhD thesis, University of California (1983)

[6] Clark, H.H., Wilkes-Gibbs, D.: Referring as a collaborative process. Cognition **22** (1986) 1–39

[7] Fischer, K.: Linguistic methods for investigating concepts in use. In Stolz, T., Kolbe, K., eds.: Methodologie in der Linguistik. Frankfurt a.M.: Peter Lang (2003)

Study of Eye Movements in Landmark Recognition: An Experiment in Virtual Reality

Yu Jin, Sabine Gillner and Hanspeter Mallot

Department of Cognitive Neuroscience, Universität Tübingen
Auf der Morgenstelle 28, D-72076 Tübingen, Germany
yu.jin@uni-tuebingen.de
http://www.uni-tuebingen.de/cog

Abstract. Previous research on landmark has focus on landmark function in navigation. However, it is still not clear how a landmark is identified in a visual scene. Studies on active vision show that eye movements are proactive, anticipating actions. Our eyes actively gather task-relevant information from the visual scene. Here we are interested in how eye movements involve in landmark recognition in navigation. We analyzed eye movements of human subjects with respect to landmarks and distractors when they performed a navigation task in a virtual reality environment. Landmarks are fixated longer and more frequently than other objects in the scene.

1 Introduction

In everyday life, people have to deal with different navigation tasks such as way-finding or route-planning. Successful navigation is facilitated by using landmark information in the visual scene. Studies on animals and humans ([1], [2], [3]) have shown that landmark plays a dominant role in visual navigation. How does an agent define an object as a landmark in a scene and use it in a navigation task? Landmark is often noticed or memorized because it is capable of attracting attention. Landmark saliency may accrue by a unique visual feature of an object. In computer vision, a landmark can be distinguished from the background image by comparing visual properties. However, the Eiffel Tower can be easily identified as a landmark and one's own house may increase saliency to the individual at the same level. So landmark saliency is not equal to visual saliency of an object, but depends on the functional content of this object in a certain navigation task.

Our eyes have limited spatial resolution in the periphery, which requires the oculomotor system to direct the eyes towards objects in order to bring them into the fovea area [4]. Yarbus's study [5] has shown that the perception of a complex scene involves a complicated pattern of fixations. The generation of eye-gaze patterns is a processing of selecting task-relevant information. In recent years, researchers have increasingly employed eye movements to study covert cognitive process in different tasks ([6], [7], [8], [9]). Those studies have shown that eye-gaze patterns are highly specialized for different tasks. Here we studied landmark recogntion in a navigation task by examining eye movements. We were interested in whether there are landmark-associated eye movment patterns.

2 Methods

The experiment was run on a desktop computer. 10 subjects participated in the experiment. All subjects had normal visual acuity.

Eye movements were recorded using a mobile eye-tracking system (ASL 501) with an accuracy of about 1^o. Video data were analyzed offline with Matlab. Fixations were defined as any period in which a subject's gaze remained stationary with respect to an object in the field.

The virtual reality environment was programmed with OpenGL performer. It was adapted from a Hexatown model in which subjects can not infer the current location from the geometrical layout. Landmarks and distractors were introduced in the environment and had matched visual propertis. Unlike landmarks, distractors were duplicated at each place and identical in the whole environment.

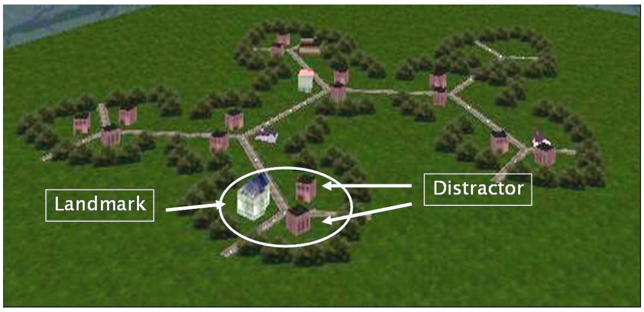


Fig. 1. Bird-eye view of the virtual reality environment with landmark and distractors

The experiment had a training phase and a test phase. In the training phase, subjects were driven passively along a route through the environment. They were instructed to memorize the route. The route was designed as a loop route so that subjects could not memorize it by using sequential memory. In the test phase, subjects were placed randomly at some place along the trained route and asked to reproduce the route by making next-step movement decision: move straight, turn left or turn right. The movement decisions were performed by pressing the corresponding cursor buttons (i.e. forward arrow ↑ move forward). Feedback was given after subjects made the movement decision. The test phase consisted of 12 trials. The training phase restarted automatically if subjects made any error in the test phase. There were two experiments (A & B). An object served as a landmark in one experiment and as a distractor in the other experiment.

3 Results

In order to perform the experiment, subjects need to identify landmarks in the scene and associate them with the current location. We examined eye movements data from experiment A (inital trial), experiment A (last trial, subjects learned route and performed tests without errors) and experiment B (last trial). Fixation number and fixation duration on landmarks and distractors were analyzed. Figure 2(a) illustrates relative fixation number on landmarks (fixations number on distractors is set to 1) in experiment A (first trial & last trial) and experiment B (last trial). The mean fixation duration is shown in figure 2(b). Landmarks are fixated longer than distractors (432 ms on landmarks and 323 ms on distractors).

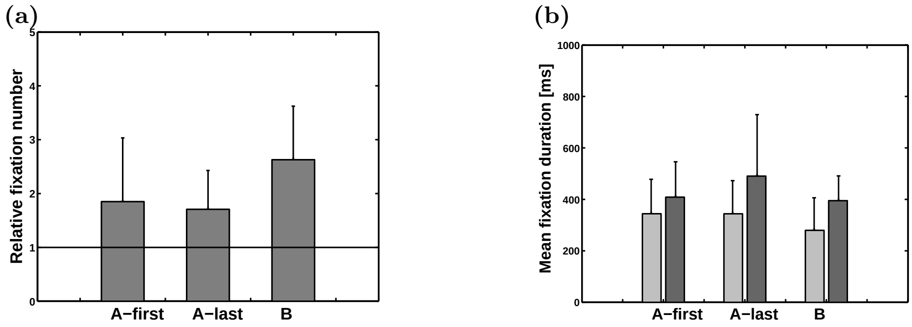


Fig. 2. **(a)** Relative fixation number on landmarks. THICK LINE: fixation number on distractors is set to 1. **(b)** Mean fixation duration on distractors (light grey) and landmarks (dark grey) in the experiment A and B.

We compared fixation duration on the same object in experiment A and B (figure 3). The fixation duration is longer when the object acted as a landmark than as a distractor in the experiment.

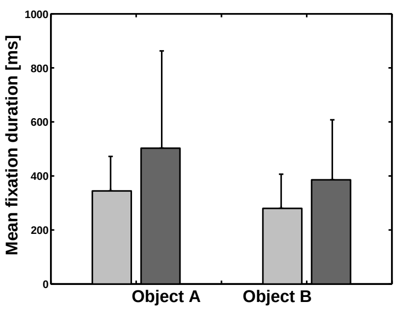


Fig. 3. Mean fixation duration on objects (distractor in light grey & landmark in dark grey)

4 Discussions

Here we examined eye movements towards landmarks in a navigation task. Virtual reality technique enables the possibility of manipulating the environment parameters (environment layout, objects). Landmarks in the experiment are unique for each place and contain local spatial information. Subjects can use them to infer the current location in the environment. Unlike landmarks, distractors contain no spatial information. Landmarks and distractors have matched visual properties. So they have similar visual saliency in the experiment. The difference between landmarks and distractors is their functional content. Distractors were duplicated at each place in the environment so they were presented more often in the scene than landmarks. However fixation number data show that subjects more often have their eyes gazed on landmarks than distractors. Also landmarks gain longer fixation duration than distractors. It indicates subjects selectively focus on the important information (landmarks) in order to perform the navigation task, rather than view the visual scene randomly. Interestingly, an object gains longer fixation duration when it acted as a landmark than as a distractor in the environment. Different eye movment patterns could be explained by landmark saliency but not by visual saliency. We have demonstrated studying eye movement patterns in navigation (active vision) provides a potential window for covert cognitive processing. It gives a reliable, objective measurement to investigate landmark recogntion in navigation: how subjects identify landmarks in the environment and use them to perform a navigation task.

References

1. Cohen R, Schuepfer T: The representation of landmarks and routes. Child Development, Vol. 51: (1980) 1065–1071.
2. Cartwright BA, Collet TS: Landmark learning in Bees. Journal of Comparative Physiology A, Vol. 151: (1983) 521–543.
3. Steck S, Mallot H: The role of global and local landmarks in virtual environment navigation. Teleoperators and Virtual Environments, Vol. 9: (2000) 69–83.
4. Carpenter RHS: Movements of the Eye. Pion Press, London. (1988)
5. Yarbus A.L: Eye movements and vision. New York: Plenum Press.(1967)
6. Land M.F: Predictable eye-head coordination during driving. Nature, Vol. 359: (1982) 318–320.
7. Land M.F, Lee D: Where we look when we steer. Nature, Vol. 369: (1994) 742–743.
8. Land M.F, Hayhoe A: What ways do eye movements contribute to everyday activities? Vision Research, Vol. 41: (2001) 559–565.
9. Shinoda H, Hayhoe M, Shrivastava A: What control attention in natural environments? Vision Research, Vol. 41: (2001) 3535–3545.

A Framework Modelling Connotations of Spatial Objects

Manfred Loidold and Christine Giger

GeoInformation Technologies Group
Institute of Geodesy and Photogrammetry
ETH Zurich
{loidold,giger}@geod.baug.ethz.ch

1. Lacking Abilities of Conventional GIS

Geographic Information Systems (GIS) are made to support men in dealing with spatial tasks and decision-makings. But the difference between human perception of space and the way, GIS handle spatial information limits the applicability and spread of GIS. Differences between human and technical perception of space are:

- GIS tend to represent spatial objects in a static way ignoring the dynamics of change.
- They do not take into account human connotations about spatial objects (and relations).
- GIS ignore the functional aspect of a spatial object.
- They simplify interdependencies to topological relations.

These deficiencies define the requirements for a GIS representing spatial objects and relations in a more sophisticated way: the capability

- to take into account the function of an object
- to acquire and represent dynamic processes
- to model and formalise the different perceptions, views and estimations of individuals towards spatial objects

2. Functional Aspects of Spatial Objects

The purpose of this approach is to contribute to the spatial cognition research by investigating the functional aspects of spatial object. The framework is based on the assumption that the functions of an object determine its perception. The value of an object for satisfying individual needs defines the individual evaluation of a spatial object.

The functional aspects of spatial objects can be divided into two different classes. On the one hand, geographic objects contain universally valid aspects, which are independent of the perception of a specific viewer. On the other hand, spatial objects are characterised by specific aspects depending on the particular user.

In general, the purpose of a street is to support the demand for mobility (universally valid worth). But the perception and appreciation of a specific street will vary, whether someone drives on it (=enabling aspect) or tries to cross the street by walking. Depending on the circumstances (e.g. traffic volume) the latter perceives this street as an obstacle (disabling aspect). As soon as this persons walks along the street on the sidewalk, the street (including the sidewalk) will turn into an enabling object. The changed viewpoint of perception modifies the actual appreciation.

This example illustrates the necessity to distinguish the validity of functional aspects associated to a spatial object:
a) Universally valid independent of a particular viewer

- Geometry
- Topology
- Attribute data
- The potential function denotes all possible functions, which an object can fulfil – regardless of its actual function.
- The potential modification includes possible changes on the ontology of an object taking into account the dynamic aspect.

b) Specific aspects depending on the viewer

- Actual modification denotes any change on the object or the view point of an observer, which actually takes place.
- Actual function means a task a person assigns to a spatial object.
- The individual benefit of an object is the worth for satisfying one's basic needs.
- The relational aspect denotes perceived interdependencies of objects.

=> Results are the effects of spatial objects (aims, consequences).
=> These effects shape connotations and consequently the individual's evaluations of objects.
=> They reflect human perception of space.

3. Functional Relations

In the modelling of the functional aspects (f) our approach aims to rely on applied methods (CSL) and to figure out interdependencies:

- The ontology (f_0) of a spatial object contains geometry, topology and attributive data, which is formalised by a Conceptual Schema Language (CSL) like INTERLIS (cf. figure 1 at the end of this list).
- The potential function depends on the ontology and the respective class of objects: $f_1(f_0$, class of objects) $\in F_1$, where F_1 denotes the set of all possible potential functions .
- The potential modification depends on the potential function and – if not included in it – the respective class of objects: $f_2(f_1$, class of objects).
- The actual function is a subset of the potential functions of an instance of a class of objects: $f_3(Obj_i) \in F_1$.

- The actual modification is a subset of the potential modifications: $f_4(Obj_i) \in F_2$.
- The individual benefit (f_5) is determined by the ontology, the actual function and the actual modification. The individual benefit of an object for satisfying needs is reduced, if the function can be satisfied by another facility (Substitution S): $f_5(f_0(Obj_i), f_3(Obj_i), f_4(Obj_i) - S)$.
- The relational aspect is determined by the ontology and the functional aspect of an object. As we cannot exclude further implications, we include a factor x (external factors) in this framework: $f_6(f_0(Obj_{i...n}), f_3(Obj_{i...n}), x)$.
- The functional aspect result in certain effects (f_7) determined by the ontology, the actual function, the potential modification and the individually perceived benefit: $f_7(f_0, f_2(Obj_i), f_3(Obj_i), f_5)$.

```
!! Version 2000-03-26

INTERLIS 2.0; LANGUAGE = en; !! 2-letter code (ISO 639)

DATA MODEL Roads2Example =

  DOMAIN
    Point2D = COORD 0.000 .. 200.000, !! Min_East  Max_East
                    0.000 .. 200.000, !! Min_North Max_North
                    ROTATION 2 -> 1;
    Orientation = 0.0 .. 360.0;

    TOPIC Roads =

    CLASS LandCover =
      Type: MANDATORY (building, street, water, other);
      Geometry: MANDATORY SURFACE WITH (STRAIGHTS) VERTEX Point2D
          WITHOUT OVERLAPS > 0.100;
    END LandCover;

    CLASS Street =
      Name: MANDATORY TEXT*32;
    END Street;

    CLASS StreetAxis =
      Street: MANDATORY -> Street;
      Geometry: MANDATORY POLYLINE WITH (STRAIGHTS) VERTEX Point2D;
      Precision: MANDATORY (precise, unprecise);
    END StreetAxis;

    CLASS StreetNamePosition =
      Street: MANDATORY -> Street;
      NamPos: MANDATORY Point2D;
      NamOri: MANDATORY Orientation;
    END StreetNamePosition;

    CLASS PointObject =
      Type: MANDATORY (tree, geodetic_point, other);
      Position: MANDATORY Point2D;
    END PointObject;

  END Roads; !! of TOPIC
```

Fig.1: INTERLIS (Conceptual Schema Language) formalises geometry, topology and attribute data.

We argue that all functional aspects interfere with the effects of a spatial object directly or in an indirect way. Together with external factors these effects have a bearing on the connotations of a person concerning a spatial object. They are represented in figure 2 as cloud, because these external factors are numerous and the list is not complete.

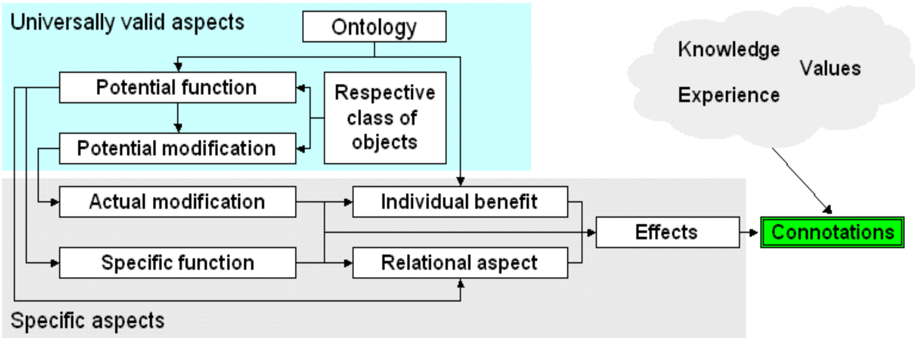


Figure 2: interdependencies of the universally valid and the specific aspects resulting in certain connotations concerning a spatial object.

Based on observations in a decision making process we figured out three typical distributions of connotations in a participatory planning group:

1. Homogenous distribution occurs if all participants share connotations of a spatial object or a class of objects. The homogenous evaluation can either be positive (e.g. wood), neutral (cable car), or negative (waste deposit).
2. A distribution is heterogeneous if no peak can be stated and the connotations cover a wide range. All evaluations (very high – very low) are assigned to a spatial object. No generally valid connotation could be identified.
3. A bipolar distribution is characterised by peaks at both ends of the scale. Street is a typical example of polarising object.

Although these observations were made in the specific scope of a participatory planning process, we suppose that these classifications of connotations are universally valid. Our further research aims at developing and confirming our framework in order to verify if this approach supports the understanding of spatial cognition and evaluation processes.

Route Description System: Architecture for Managing Geo-referenced Content

Sebastian Matyas, Christoph Schlieder [1]

[1]Chair for Computing in the Cultural Sciences,
Laboratory for Semantic Information Processing, Otto-Friedrich-University Bamberg,
D-96049 Bamberg, Germany

Abstract. Also Satellite positioning Systems such as GPS have drastically simplified the task of geo-referencing physical objects; geo-referencing semantic objects such as multimedia documents still are a challenging field in computer science. In the tourism area the precision of a GPS handheld with an average error smaller than 5m is good enough for geo-referencing physical objects like train station or sightseeing spots. We propose to indirectly geo-reference the corresponding Semantic objects through these physical objects they refer to. Our paper describes the architecture of a content management system which provides support for both direct and indirect geo-references and which have been implemented in the Route Description System authoring tool. For example it is now possible for a provider of a web-based information service for bikers to easily generate the description of a particular tour for a biker according to his personal preferences regarding content and media.

Geo-referencing semantic objects

Satellite positioning systems such as GPS have drastically simplified the task of geo-referencing, i.e. the task of determining the geographic coordinates of physical objects. However, geo-referencing semantic objects such as multimedia documents still constitute a major challenge for research in computer science. Semantic objects are geo-referenced indirectly through the physical objects they refer to. We study the problem of indirectly geo-referencing documents with touristic content. Our paper describes the architecture of a content management system which provides support for both direct and indirect geo-references and which have been implemented in the Route Description System authoring tool.

In several application areas, the geographic coordinates of physical objects can be determined with sufficient precision by anyone in possession of a GPS-handheld device produced for the consumer market. This is certainly the case for touristic applications in which points of interest (POI) need not be localized with an average error smaller than 5m. As a consequence, an increasing number of tourists make use of GPS-handhelds to support them in activities like biking or hiking. On the Web, they exchange GPS-data recorded in the field, e.g. route data and POIs of last summer's biking tour.

Today's content management systems (CMS) are designed to store, retrieve, and edit content for the Web and do not provide specific support for geo-referenced content. This makes it extremely difficult to reuse the content according to spatial criteria. Consider the following example. The provider of a web-based information service for bikers would like to manage all touristic content relating to a regional network of biking trails. It should be possible to easily generate the description of a particular tour for a biker according to his personal preferences regarding content and media. The result could be a text giving navigation instructions and describing a personalized selection of POIs or, equally well, a GPS track and an audio file with background information about history and geography of the region. What impedes current CMS reusing descriptions for parts of a route is the lack of support for organizing content in a way that reflects the spatial organization of the route network.

Route Description Mark-up Language

The foundation of the proposed CMS architecture is a representation language in which *n*-dimensional geo-features can be described and associated with non-spatial content. Zero-dimensional geo-features represent POIs, one-dimensional geo-features route segments and two-dimensional geo-features regions of interest. Image 1 and 2 illustrate these geo-features on the basis of two drawings generated in Autodesk Map. One requirement for the *route description mark-up language* (RDml) is that it also serves as an exchange format between various applications over the web. Designing RDml as a XML language was therefore a straightforward choice.

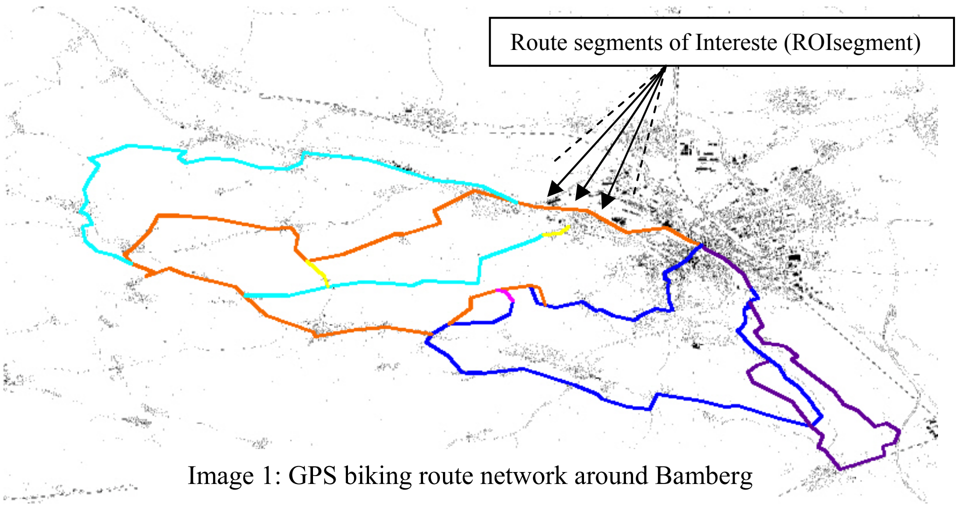


Image 1: GPS biking route network around Bamberg

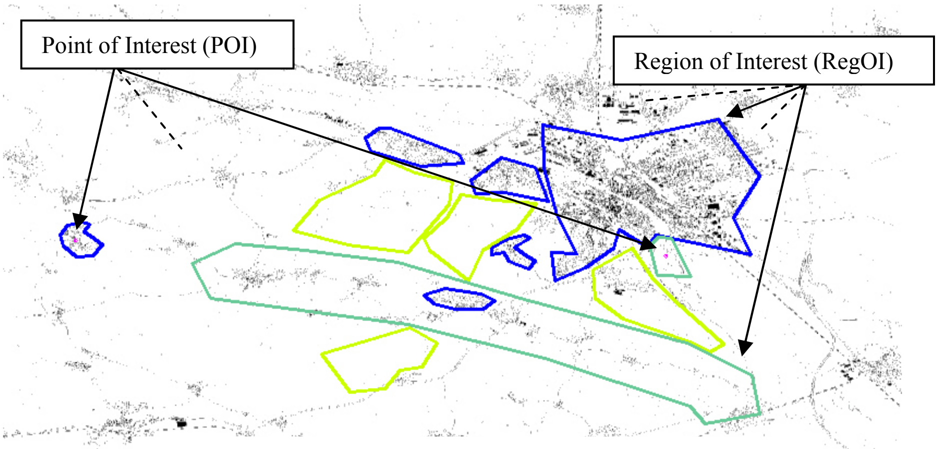


Image 2: Sample regions and POIs around Bamberg

A number of XML-based spatial languages have been proposed, mostly for the purpose of exchanging GPS data. A good example is the NaVigation Mark-up language (NVML) described by Takayama et al. (2003), Korkea-aho et al. (2001). Although this language is suitable for the representation of GPS raw data it is limited in its expressiveness. For example, the representation of two-dimensional geo-features or the association of video content is not possible.

For this reason, we decided to design RDml, as a special-purpose language for directly and indirectly geo-referenced content. Basic primitives of RDml are route segments, POIs and regions. These zero- to two-dimensional geo-features can be associated with multimedia content objects. Multiple associations of multimedia content from different sources are also possible.

The Route Description System Architecture

The *route description system* (RDS) architecture consists of four layers of abstraction. The lowest is the base layer which implements the topology of the route network, the POIs and regions in terms of standard GIS functionality. The topology of the route network is also used as an index for the multimedia content associated with the geo-features.

Next comes the thematic layer where the geo-features are associated with thematic information. On this layer, they appear as routes, POIs and regions.

On the following semantic layer all content is represented, i.e. the multimedia content as well as GPS data.

The top layer represents metadata for the content. Beside document-related metadata such as “data format” also judgments about the content (e.g. trust in data quality) are included. Image 3 shows the RDS layer architecture.

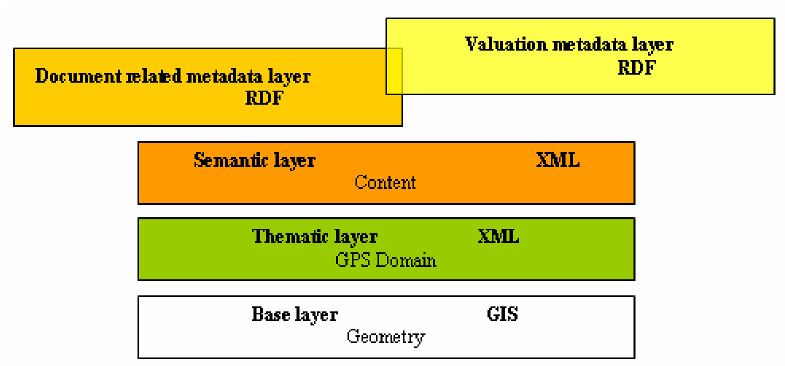


Image 3: RDS architecture

Conclusion and Outlook

Within this paper we presented architecture and a representation language with which it is possible for a CMS provider to organize content in a way that reflects the spatial organization of the route network. The indirectly geo-referenced content can be reused and put together based on personal preferences. The individual route of interest composed out of route segments is associated with thematically selected POIs and regions, through which the route leads. The multimedia content is chosen based on document related metadata.

Future research will have to make the "valuation" metadata usable for the selection of the multimedia content for the route description. This will be one of the next steps to build a recommender system for GPS biking tours.

Furthermore the implementation of both metadata aspects in the representation language RDml, e.g. in form of a RDF/XML-serialization, will be necessary for the recommender system and useful for exchanging this "hybrid" documents, GPS data and multimedia data, between two RDS applications.

References

Mari Korkea-aho and Haitao Tang
Experience of Expressing Location Information for Applications in the Internet
Department of Computer Science and Engineering, Helsinki University of Technology, Finland
Nokia Research Center, FIN-00045 Nokia Group, Finland
2001

Kuniharu Takayama, Yoshiharu Maeda, Hirohisa Naito
NVML: NaVigation Markup Language
Fujitsu Laboratories
June 2003

Graphical vs. Verbal Route Descriptions: Looking for Supra-modal Invariants

Edyta Przytula-Machrouh[1], Gérard Ligozat[1], Michel Denis[1]

[1] Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur LIMSI-CNRS, Université Paris-Sud
Bâtiment 508
91403 Orsay, France
{edyta, ligozat, denis}@limsi.fr

As the saying goes, "a picture is worth one thousand words". Is that true for route directions? Thanks to its universal character, a drawing can be understood independently of the language used by the reader. Furthermore, the inspection time of a sketch is shorter than that of a text, because the first allows the reader to access directly the elements he/she needs while neglecting secondary pieces of information [1], [2]. In the domain of spatial descriptions, the pictorial mode of expression is expected to be privileged. Indeed, it allows representing in a synthetic way spatial relations which would be long to describe by verbal means. Unsurprisingly, maps are frequently used for pedestrian or car navigation. However, in spite of the advantage of maps for representing global, immediately accessible information, and for making spatial relations explicit, they are still not the favourite aids of users for navigation tasks, and they tend to be judged as unclear and difficult of use [3], [4].

This work examines the use of graphical descriptions of routes, and its relationship to verbal descriptions. A central issue is that of the type of information that people are able to extract from them. To such purpose, we conducted an experimental study which consisted for participants having no prior knowledge about the route in re-expressing the content of a graphical description into a verbal description. The analysis of verbal descriptions obtained in that way, as translations of the graphical descriptions, provided us with interesting clues about the way individuals use this mode of description and about the type of information they are able to extract from a sketch. The initial material for the study consisted in a set of sketches produced previously by persons having a good knowledge of the route. As opposed to a map, such sketches are pre-processed artefacts in the sense that, ideally, they will contain only necessary and sufficient information about the route. Good and poor graphical descriptions were selected by judges having a good knowledge of the route (Fig 1). The verbal descriptions produced as the result of the interpretation of a sketch reflect the cognitive processes involved in the use of a graphical description. Since verbal descriptions of routes possess properties similar to those of the actual process of navigation, they tend to appear as linear, composed of sequences representing local information presented in the "correct" temporal-spatial order. They also provide indications about the processes which are activated when "reading" a sketch.

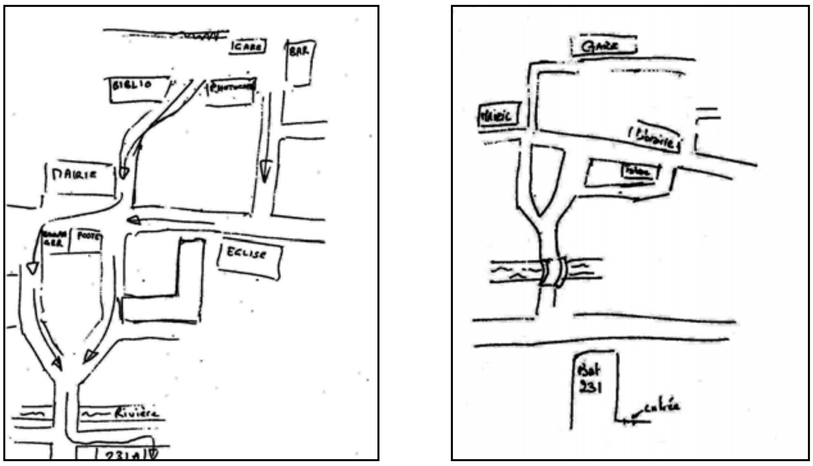


Fig. 1. Selected sketches: a good graphical description (left) and a poor one (right)

Our results show that a user first locates his/her initial position, which is usually made clear as a comment. This initial information ensures that the person will begin his/her progression in the correct direction. This initial step may involve the choice of a particular route when several possibilities appear. We also noted in verbal descriptions a systematic positioning of the subject, which results in the use of corresponding lexical terms (*on your right*). This positioning fixes the origin of the reference frame used in the description for positioning the landmarks (*Take the street on your left*). Moreover, landmarks and actions are often localized by the user.

The user then fragments the global information into several local pieces of information. He/she then proceeds to the identification and to the grouping of the elements presented in the sketch. The translation of a sketch into a verbal description imposes a division of the sketch into minimal units of information which we call "graphical schemas". These "graphical schemas" stem from the grouping together of elements according to specific rules; for instance, the grouping is made between elements situated in a relation of proximity on a sketch, and between elements representing an action and elements representing landmarks. Figure 2 shows an example of the grouping. During that phase, "bounded" landmarks are first taken into account. In the absence of such landmarks, either the grouping is made using the unbounded landmarks, or only the action itself is considered.

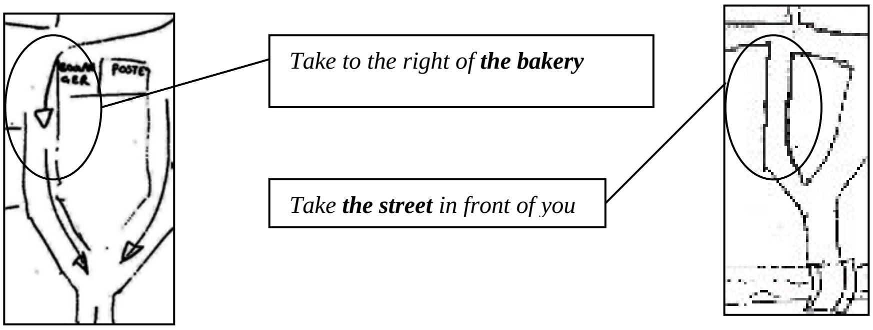


Fig. 2. Examples of the grouping together of elements in the fragments of sketches representing the same part of the environment: on the left, a fragment from a good graphical description, on the right, the fragment from a poor one.

In our experiment, the participants segmented the sketches according to a small subset of types of graphical schemas. This suggests that graphical schemas can be considered as primitive route elements, or "wayfinding choremes"[5]. Based on our study, we defined a repertoire of basic graphical schemas used in graphical route directions, which were shared by all participants. The basic units differ from the traditional maps, being simpler and more schematic. We suggest that such units can contribute to the development of "a language for space", as proposed by Brunet [6], [7].

References

1. Ganier, F., Gombert, J.-E., Fayol, M.: Effet du format de présentation des instructions sur l'apprentissage de procédure à l'aide de documents techniques. *Le Travail Humain*, 63 (2000) 121–152.
2. Ganier, F., Gombert, J.-E., Fayol, M.: Discours procédural et activités mentales: de la comprehension d'instructions complexes à la planification de l'action. *Langages*, 141 (2001) 47–63.
3. Petchenik, B.B.: Value and values in cartography. *Cartographica*, 22 (1985) 1-59.
4. Streeter, L.A., Vitello, D., Wonsiewicz, S.A.: How to tell people where to go: Comparing navigation aids. *International Journal of Man-Machine Studies*, 22 (1985) 549-562.
5. Klippel, A.: Wayfinding choremes. In: Kuhn, W., Worboys, M., Timpf, S. (eds.): *Spatial Information Theory. Foundations of Geographic Information Science*. Springer-Verlag, Berlin (2003) 349-361.
6. Brunet, R.: *La carte, mode d'emploi*. Fayard-Reclus, Paris (1987)
7. Brunet, R.: Building models for spatial analysis. *Two decades of l'Espace Geographique*. Gip-Reclus, Montpellier (1993) 109-123.

Eye movements during a locomotion task in virtual and real environments

Dagmar Schoch, Sabine Gillner & Hanspeter A. Mallot

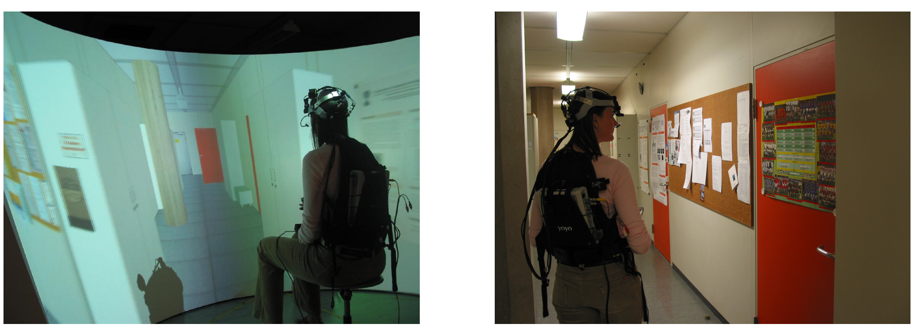
University of Tübingen, Institute for Cognitive Neuroscience, Morgenstelle 28
72076 Tübingen, Germany
dagmar.schoch@uni-tuebingen.de

Abstract. We investigated the eye movements of human subjects, walking through a real and a virtual hallway. Subjects in both types of experiments prefer to look on a few objects along the travelled path, the number of fixations to the same objects in the real and the virtual environment show a positive correlation (r = 0.88). From this result we can conclude that virtual reality is a reliable tool for eye movement studies.

1 Introduction

In recent years eye movement studies in natural environments have become a focus of research. At the same time, virtual reality has become an accepted tool in psychophysical studies. Here we investigate the eye movements of human subjects while walking through our lab and compare them with eye movements of subjects driving throught the "same" hallway in a virtual environment.

2 Methods



(a) **(b)** s

Fig. 1. (a) In the experiments in virtual reality subjects were seated in front of a round, tilted projected screen with a large field–of–view in horizontal and vertical direction. They could move with a joystick. **(b)** In the experiments in the real environment subjects walked a route through the hallway of our institute.

11 subjects took place in the experiment. Their task was to walk along a route of about 50 m through our lab, including three left–turns. For the experiments in virtual reality we build a model of the floor in a highly detailed manner. In these kind of experiments subject were seated in front of a round, tilted projection screen (horizontal field–of–view: 163^o, vertical field–of–view: 70^o) and could move through the virtual environment with a joystick. In both types of experiments eye movements were recorded (*ASL 501 mobil*) with an accuracy of about 1^o. Instructions about the route were given previously. Subjects performed three trials: either vr–real–real or real–vr–vr. One experiment lasted about one hour, including a calibration procedure of the eye tracker.

3 Results

Subjects in both types of experiments prefer to look on a few objects along the travelled path. Interestingly the number of fixations to the same objects in the real and the virtual environment show a positive correlation. A linear fit (see fig. 2) has a slope of almost one ($f(x) = 1.2x - 0.003$), the correlation coefficient is 0.88. The majority of fixations are pointed straight ahead in the direction of progression(around 40 % of the fixations). In virtual reality there is a stronger tendency to look on the floor (11 % of the fixations in VR vs. 6 % of the fixations in the real environment). The mean duration of the fixations in the two types of experiments showed no significant difference (401 ms in the real environment vs. 398 ms in the virtual environment), whereas the total number of fixations in the real environment was much higher as in the virtual environment (1393 in the real environment vs. 972 in the virtual environment).

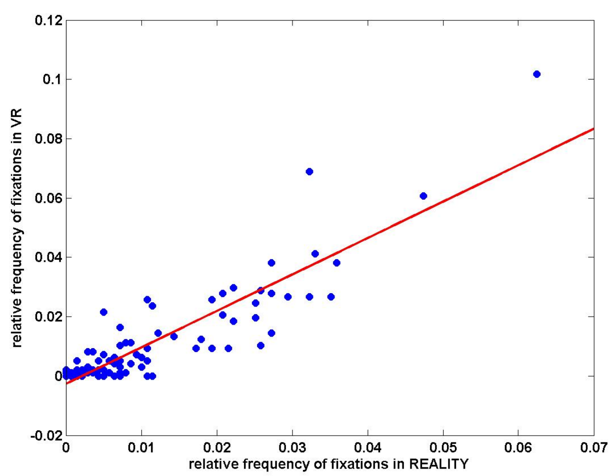


Fig. 2. Correlation plot of number of fixations to selected objects in the real and virtual conditions (linear fit: $f(x) = 1.2x - 0.003$, correlation r = 0.88).

4 Discussion

In virtual reality, subjects fixate the same objects along a travelled path as in a similar real environment. In both environments most of the fixations are directed towards the direction of progression, which was also found in other studies [1], but in a much higher extent (67 % up to 92 %). Also the fixation durations in both environments were comparable. The largest difference concerns the number of fixations to the ground floor. This could be due to the different locomotion modalities in the two experiments: In the real hallway we investigate the eye movements of pedestrians, whereas in the virtual environment the movement with the joystick resembles rather driving a car. From driving studies it is known [2] that eye movements are directed in particular towards the street. In summary we can conclude from our results, that virtual reality is a reliable tool for eye movement studies.

References

1. Hollands MA, Patla AE & Vickers JN (2002) "Look where you're going !": gaze behavior associated with maintaining and changing the direction of locomotion. *Expt. Brain Res.* 143:221–230
2. Land MF (1998) The visual control of steering. In *Vision and Action* (Eds. Harris, LR & Jenkins M) 163-180 (Cambrigdge Univ. Press, Cambrigde UK)

Associative Learning of Spatio-Temporal Sequences in a Reaction Time Experiment

Rainer Spiegel[1, 2, 3]

[1] Department of Computing, Goldsmiths College, University of London
New Cross, London, SE14 6NW, United Kingdom
r.spiegel@gold.ac.uk
http://www.goldsmiths.ac.uk/departments/computing/staff/RSP.html
[2] Department of Experimental Psychology, University of Cambridge
Downing Site, Cambridge, CB3 9BB, United Kingdom
r.spiegel@psychol.cam.ac.uk
[3] Wolfson College, University of Cambridge
Barton Road, Cambridge, CB3 9BB, United Kingdom
rs272@cam.ac.uk

Abstract. Recent data from experiments on sensory-motor learning of spatio-temporal sequences will be summarized along with simulations from a recurrent neural network. The network predicts human learning as well as generalization to novel sequences. The close match between the results of the network and the human data suggest that human learning in this task can be explained by a rather general, associative learning algorithm. It has to be kept in mind, though, that this is no proof that humans generally learn spatio-temporal tasks associatively. In addition, alternative explanations will be considered as well.

1 Introduction

In psychology experiments, recurrent neural networks have become popular models of human learning. One example is the simple recurrent network, abbreviated SRN [1], which was the most cited paper in Psycholinguistics and the 11th most cited paper in Psychology from 1990 to 1994 [2]. One reason for its high impact might be that its learning principles have a long tradition, not only in psychology, but also in related disciplines investigating sensory-motor learning [3]. These principles are mechanisms of associative learning, where connections between units are formed through simple associations, much as learning takes place in the majority of neural networks. Moreover, there is evidence that animal learning can often be explained by associative learning mechanisms, [4]-[7], and at least some human learning can be explained by those mechanisms as well, e.g. [8]-[13]. SRNs were also successful in serial reaction time experiments, where participants respond to sequences of signals on a computer screen by pressing designated keys on their keyboard. What all serial reaction time experiments in this summary have in common is the screen display depicted in Figure 1. The signals consisted of a circle filling with white color until the participant pressed a key. Subsequently, the three outlines momentarily re-appeared until the next circle filled with white color.

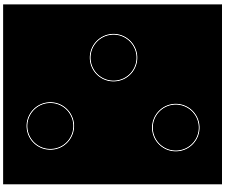


Fig. 1. The screen display in my Serial Reaction Time Experiment

1.1. The Sequence Learning Task

To clarify the sequential order of circle flashes, the letters A, B and C will be used to illustrate circle locations, but participants never saw any letters in the course of the experiment. They only saw the 3 circles. The Experimental group received the following consistent sequences

AB(varying numbers of C elements)BA or
ABB(varying numbers of C elements)BBA

In other words, the number of B elements before the Cs determines how many Bs appear after the Cs. The Control group was trained on the same sequences as the Experimental group (=consistent sequences) in 50 percent of the cases, and on the following inconsistent sequences in the other 50 percent of the cases:

AB(varying numbers of C elements)BB
ABB(varying numbers of C elements)BAA

In the Control group, the number of B elements before and after the C elements was not always the same during training. Following the training phase there was a testing phase where both Experimental and Control group were exposed to both consistent and inconsistent sequences. If the Experimental group was able to learn the relationship that the number of Bs is the same before and after the Cs, they should anticipate an A as the final location and a B as the penultimate location at the location where Experimental and Control group diverged. Faster reaction times and less errors on the respective locations would indicate anticipation. In my simulations with the SRN, anticipation is expressed by means of higher activities on the units representing the respective locations. The network is trained on the same sequences as the participants and the results are analyzed with the same methods. More details on testing the participants and running the network simulations can be found in [3], [14], [15].

2. Summary of Results

Participants in the Experimental group not only performed better than the Control group on the respective sequence locations (see 1.1). Rather, their performance was also determined by the varying numbers of C elements, which were themselves irrelevant with regards to the relationship between the Bs. If participants had been trained on an odd number of C elements (such as 1, 3, 5 Cs), participants would not only learn this relationship for the trained sequences. Rather, they would also generalize to novel sequences with odd numbers (such as 7 Cs). If participants had been trained on both odd and even numbers of C elements (e.g. 1, 3, 4 and 6 Cs), they would generalize to novel sequences with both odd and even numbers, such as 2 Cs and 5 Cs. However, they would not generalize to novel sequences outside the training range (such as 7 and 8 Cs). The remarkable finding is that this pattern of results was exactly predicted by the SRN. I replicated this finding in further experiments with different numbers of C elements. Analyses and results are beyond page limit, but details are provided in [14] and further experimental results are in preparation [15].

3. Discussion

The close match between the results of the SRN and the human data suggest that human sequence learning, in so far as this task is concerned, can be explained by an associative learning algorithm. The successful model for this experiment [1] belongs to a group of connectionist models that make use of associative learning algorithms, of which error correcting algorithms are prominent examples [16]. It has to be kept in mind, though, that these particular algorithms learn the statistical regularities of the training set, so that statistical models would also be able to account for these results, e.g. [14], [17], [18]. A recent statistical approach to sensory-motor learning [19] could also be extended to model this sequence learning task. On the other hand, associative learning algorithms and statistical approaches (e.g. Bayesian learning) are linked, so that convergence to similar results comes by no surprise, e.g. [14] and [20]. However, statistical models are usually not based on the type of auditory feedback that our participants received (a beep for the wrong key press and no beep for the right key press). This type of feedback is reminiscent of the feedback given in the SRN's learning algorithm. Moreover, it has been shown in another paper [3] that participants reach the point where their performance breaks down, i.e. their learning stops. So far, this finding was only predicted by the associative SRN.

What would be difficult, though, is to explain these findings with a rule-based mechanism. If participants figured out the underlying rule that the number of Bs is the same before and after the Cs, why would they generalize to some novel numbers of C and not to others. Even if they formed a rule for the odd-even distinction, why would they generalize outside the training range in one experiment and not in the other, and why can all these results including the performance break-down be predicted by a purely associative model? Nevertheless, this type of task provided evidence for rule-based performance on shorter and less numerous sequences [14].

References

1. Elman, J.L.: Finding Structure in Time. Cognitive Science 14 (1990) 179-211.
2. Marcus, G.F.: The Algebraic Mind. Integrating Connectionism and Cognitive Science. MIT-Press, Cambridge MA (2001) [page 25]
3. Spiegel, R., McLaren, I.P.L.: Computational Modeling of Human Performance in a Sequence Learning Experiment. Proceedings of the International Joint Conference on Neural Networks, Vol. 1. IEEE, Piscataway N.J. (2003) 212-217.
4. Waelti, P., Dickinson, A., Schultz, W.: Dopamine Responses Comply with Basic Assumptions of Formal Learning Theory. Nature 412 (2001) 43-48.
5. Rescorla, R.A., Wagner, A.R.: A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Non-reinforcement. In: Black, A.H., Prokasy, W.F. (eds.): Classical Conditioning II: Current Research and Theory. Appleton-Century-Crofts, New York (1972) 64-99.
6. Mackintosh, N.J.: A Theory of Attention: Variations in the Associability of Stimuli with Reinforcement. Psychological Review 82 (1975) 276-298.
7. Dickinson, A.: Contemporary Animal Learning Theory. Cambridge University Press, Cambridge (1980).
8. Christiansen, M.H., Chater, N.: Toward a Connectionist Model of Recursion in Human Linguistic Performance. Cognitive Science 23 (1999) 157-205.
9. Cleeremans, A.: Mechanisms of Implicit Learning. Connectionist Models of Sequence Processing. MIT-Press, Cambridge MA. (1993).
10. Dienes, Z., Altmann, G.T., Gao, S.J.: Mapping across Domains without Feedback: A Neural Network Model of Transfer of Implicit Knowledge. Cognitive Science 23 (1999) 53-82.
11. Spiegel, R., McLaren, I.P.L.: Abstract and associatively-based representations in human sequence learning. Phil. Trans. R. Soc. London B 358 (2003) 1277-1283.
12. Spiegel, R., McLaren, I.P.L.: Recurrent Neural Networks and Symbol Grounding. Proceedings of the International Joint Conference on Neural Networks, Vol. 1. IEEE, Piscataway N.J. (2001) 320-325.
13. Spiegel, R., Suret, M., Le Pelley, M.E., McLaren, I.P.L.: Analyzing State Dynamics in a Recurrent Neural Network. Proceedings of the IEEE World Congress on Computational Intelligence. Conference: International Joint Conference on Neural Networks, IEEE, Piscataway N.J. (2002) 834-839.
14. Spiegel, R.: Human and Machine Learning of Spatio-Temporal Sequences: An Experimental and Computational Investigation. PhD Thesis, University of Cambridge UK (2002).
15. Spiegel, R., McLaren, I.P.L.: Associative Sequence Learning in Humans. Paper under review by the Journal of Experimental Psychology: Animal Behavior Processes. (under review).
16. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning Representations by Backpropagating Errors. Nature 323 (1986) 533-536.
17. Spiegel, R.: Relating Bayesian Learning to Training in Recurrent Networks. Proceedings of the International Joint Conference on Neural Networks, Vol. 2. IEEE, Piscataway N.J. (2003) 908-913.
18. Russell, J.: What is Language Development? Oxford University Press, Oxford (2004).
19. Körding, K.P., Wolpert, D.: Bayesian Integration in Sensorimotor Learning. Nature 427 (2003) 244-247.
20. Mitchell, T.M.: Machine Learning. McGraw-Hill, Singapore (1997). [Chapter 6].

Intention Detection by Motion Track Analysis on Spatial Partonomies

Klaus Stein and Christoph Schlieder

Laboratory for Semantic Information Processing, Otto-Friedrich-Universität Bamberg, D-96049 Bamberg, Germany,
{klaus.stein,christoph.schlieder}@wiai.uni-bamberg.de

1 Introduction

Information services designed for use on handset devices such as smart cellular phones or GPS-equipped PDAs exploit knowledge about the user's location to decide what type of information to present. Typical examples are location-aware information systems for museums (see [1, 2]) or computer-mediated outdoor games (see [3, 4]). To be able to provide helpful context-sensitive information not only the location of the user ("*where* is the user?") but also his intention ("*why* is he there?", "what does he *want* there?") has to be taken into account.

2 Intention Detection

While the location of an user can be measured (more or less accurate), his intentions are less obvious but may be inferred from his behaviour. Although the number of user intentions is virtually unlimited, generally, only few types of intentions need to be discriminated in order to determine the user's information needs (see [5]): specific locations entail specific sets of possible intentions (at a bus stop you may wait for the bus, enter the bus or simply pass by on a walk, and even if there are many more possible intentions only these are important to distinguish for the information system), enabling the mobile system to provide useful information (Fig. 1).

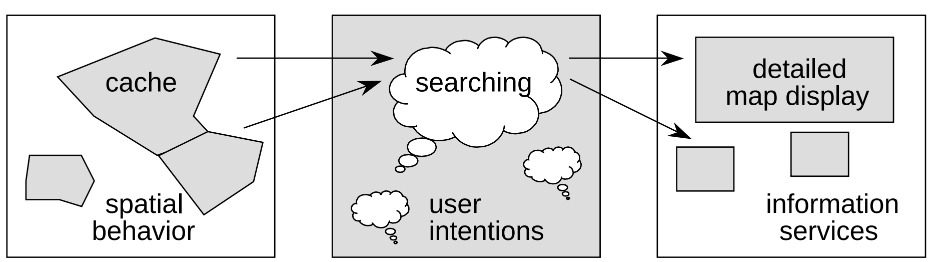


Fig. 1. Intentions as an intermediate level in modeling behavior-service mappings

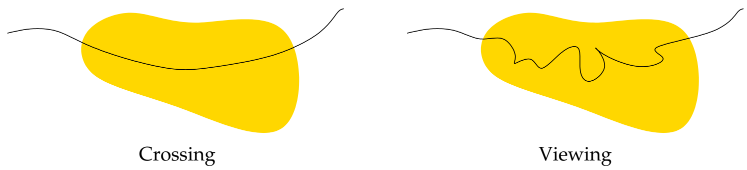


Fig. 2. Examples of motion track patterns.

One simple approach of behavior identification is motion track analysis. By tracking the user's location over time, his current speed, acceleration, motion shape etc. are known and reveal hints on his intention, as shown in fig. 2. To the left, the user's motion through a given region is fast and smooth, to the right the motion is much more winding. By using the information that the given region is an exhibition we can deduce the users intention (crossing vs. viewing the exhibition) from the fast and smooth vs. slow and winding motion track structure.

For motion track analysis the micro structure of the track is singled out by calculating the ε-histogram (see [6]) and the velocity curve. These data (supported by other sources, e. g. the main shape of the motion track) is matched to a given ruleset determining the current behavior.

3 Using Spatial Partonomies

As stated before, the mapping from a certain behavior to an intention depends on the current location of the user. Therefore any spatial region can have an own specific ruleset mapping behavior to intention. The motion track is cut at region borders and each part is handled separately.

By using partonomic structured regions ("regions within regions") a hierarchical intention detection can be applied. Fig. 3 shows the motion track of a person moving through a partonomic structure $\mathbf{P}$ incrementally entering and leaving regions $A, B, C, \ldots \in \mathbf{P}$: the person enters F, leaves F, enters G, leaves G, enters D and so on, giving the sequence

$$-F - G - D - C(A - B-) - G - .$$

For each region R a set of possible intentions $\mathbf{I}_R$ is predefined and a set of rules/prototypes is given to identify the present intention. This is done by selecting the part p_R of p inside R (by splitting p at the point, where the motion track enters and where it leaves R).

Regarding region F the matching is simple: the measures (ε-histogram, velocity curve, main shape, ...) of the part p_F of p inside F are calculated and the rules/prototypes specifying the different intentions I_1, I_2, ... are applied to them. If one rule set matches, the corresponding intention I_k is selected. If nothing matches the intention within F is unknown. The same procedure is valid

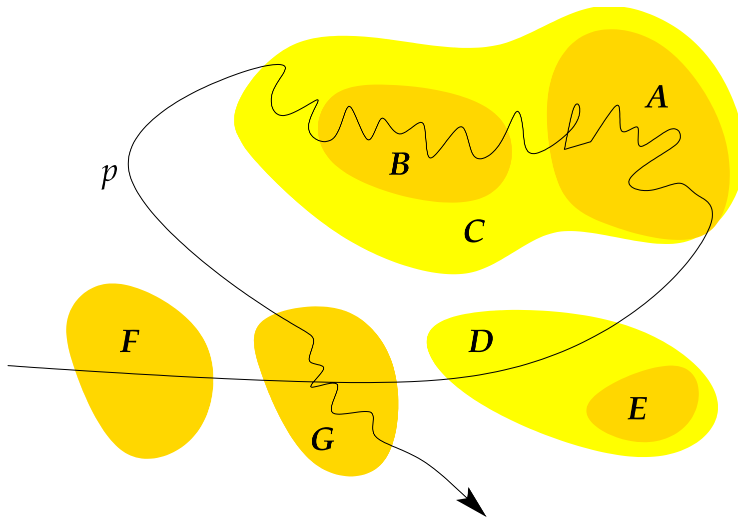


Fig. 3. Moving through a partonomic structured space

for the regions D and G. Even though there is a region $E \subset D$, E is not crossed by p and therefore savely ignored.

Within region C things become more sophisticated. Here p first enters C and within C crosses A and B before leaving C again and now the hierachical structure of $\mathbf{P}$ is used. In our example the part p_C of p inside C is cut in four segments p_A (part of p inside A), p_{AB} (part between A and B), p_B (part inside B) and p_{B0} (part between B and the border of C), we get

$$p_C = p_A \circ p_{AB} \circ p_B \circ p_{B0}.$$

On the first step the behavior within A and B is processed, revealing the intentions I_A and I_B, we get the mixed representation

$$r_C = I_A \circ p_{AB} \circ I_B \circ p_{B0}.$$

(If no intention I_B is found, because no rule set matches, the intention set $\mathbf{I}_B$ is ignored and we get the sequence

$$r_C = I_A \circ p_{A0} \quad \text{with} \quad p_{A0} = p_{AB} \circ p_B \circ p_{B0}\,,$$

likewise for region A). For processing r_C the rule set for $\mathbf{I}_C$ is enhanced. The parts p_{AB} and p_{B0} are handled by the given rulesets, but additionally the knowledge about the intentions I_A and I_B is incorporated.

Figure 4 gives an example. Here we have a museum M with two exibition rooms E_1 and E_2 and a cafeteria C. The motion track p gives the sequence

$$-M(-E_1 - E_2C-)- .$$

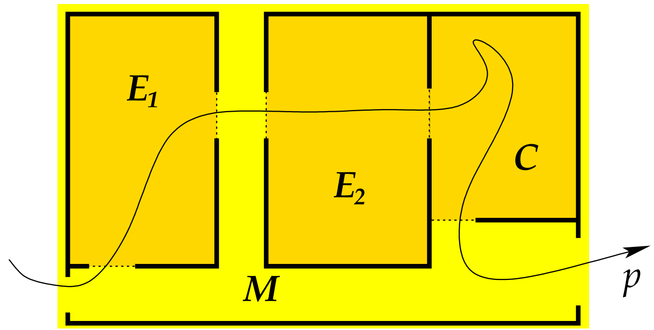


Fig. 4. Visiting a Museum?

To analyze the intention of the observed person within M the intentions found by analyzing the track within E_1, E_2 and C can be used: if I_{E_1} and I_{E_2} are "crossing" and I_C is "resting" we can conclude that I_M is "visiting the cafeteria" and not "visiting the museum".

Certainly the partonomic hierarchy can be deeper, the museum is part of a town and so on, and on each step (each region) the intentions found in the enclosed regions are used.

A more complete technical description on motion analysis and intention mapping is given in [7].

References

1. Abowd, G., Atkeson, C., Hong, J., Long, S., Kooper, R., Pinkerton, M.: Cyberguide: A mobile context-aware tour guide. Wireless Networks **3** (1997) 421–433
2. Baber, C., Bristow, H., Cheng, S., Hedley, A., Kuriyama, Y., Lien, M., Pollard, J., Sorrell, P.: Augmenting museums and art galleries. In: Human-Computer Interaction, INTERACT '01. (2001) 439–447
3. Flintham, M., Anastasi, R., Benford, S.D., Hemmings, T., Crabtree, A., Greenhalgh, C.M., Rodden, T.A., Tandavanitj, N., Adams, M., Row-Farr, J.: Where on-line meets on-the-streets: experiences with mobile mixed reality games. In: CHI 2003: Conference on Human Factors in Computing Systems, ACM Press Florida (2003)
4. Boll, S., Krösche, J., Wegener, C.: Paper chase revisited: a real world game meets hypermedia. In: Proc. of the 14th ACM conference on Hypertext and hypermedia. (2003) 126–127
5. Schlieder, C., Werner, A.: Interpretation of intentional behavior in spatial partonomies. In Freksa, C., Habel, C., Wender, K., eds.: Spatial Cognition. Lecture Notes in Artificial Intelligence. (Springer) 401–414
6. Stein, K.: Qualitative Repräsentation und Generalisierung von Bewegungsverläufen. doctoral dissertation, TU München (2003)
(*Qualitative Representation and Line Generalization of Motion Tracks*).
7. Stein, K., Schlieder, C.: Recognition of intentional behaviors in spatial partonomies. In Guesgen, H.W., Anger, F.D., Freksa, C., Ligozat, G., eds.: ECAI-2004-Workshop "Spatial and Temporal Reasoning", Valéncia (2004)

A Qualitative Description of Traffic Maneuvers

H. Joe Steinhauer

Department of Computer Science, Linköping University, S-58 183 Linköping, Sweden
joest@ida.liu.se

Abstract. In the poster we state our preliminary research results in developing a hierarchical qualitative mechanism of traffic recognition for communication. Traffic maneuvers can be classified according to certain qualitative aspects. Here a simple and intuitive model is used to capture the different states of traffic maneuvers. A hierarchical structure makes it possible to view the maneuvers on different abstraction levels. Natural language communication is done according to the different levels more or less detailed.

1 Motivation

Imagine an autonomous helicopter like the one developed in the WITAS project [1]. It is communicating the actions of a vehicle on the ground in natural language to a human operator who himself doesn't see the situation. According to the required amount of detail it could say: "The red car is driving behind a green truck, …, now the car is sheering out from behind the truck, …, coming alongside the green truck, …, passing it and sheering in in front of it." Or if less information will do "The red car is overtaking the green truck."

Now consider a driver support system that can provide you with the information what the cars around you are up to. Knowing which maneuver another driver intends to do, while you can not predict this yourself since you might not be familiar with the road system, might be of great help for you to drive smoothly in the traffic and to prevent accidents.

2 The Research Approach

We aim to develop a qualitative approach whereby traffic maneuvers can be represented in an artificially intelligent system. This representation will be sufficient to classify traffic maneuvers, give the background for communicating them in natural language and provide the basis for detecting dangerous traffic situations. For the natural language communication we want to describe every elementary movement of an object in focus as well as to give a description of composite maneuvers, which contain several sub maneuvers.

If we want to describe where an object is situated we normally do this in relation to other objects, like "The green car is overtaking the truck" where the truck is the reference object. If we say "The yellow bus is turning right at the crossing" the crossing is the reference object.

In our work we begin with the directed moving object model (DMO model) shown in Fig. 1. Several approaches exist that use this type of model for qualitative reasoning about directed and/or moving objects.

Freksa [4] describes the double cross calculus for oriented objects. He uses a neighborhood-oriented representation to reason about spatial direction information. His approach is one-dimensional. For our purpose we need a two-dimensional model but as well orientation and direction information. The terminology in the DMO model is thus the same as it is by Freksa.

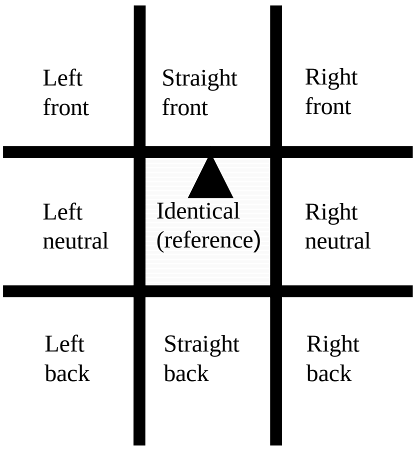
Mukerjee and Joe [6] use a two-dimensional model for calculating position relations of objects. They use an enclosing box around the object that first has to have an identified front direction. They then divide the plane around this box by extending the boxes' lines in eight two-dimensional regions that are named 1 to 8.

The direction relation matrix used by Goyal and Egenhofer [5] to calculate distances in similarity between spatial scenes forms a grid around the target object that divides the plane into nine quadrants where the

eight surrounding quadrants are named north, northeast, east, southeast, south, southwest, west, and northwest.
One approach for tracking traffic, including learning traffic maneuvers from the observed information, from a video input taken by a stationary camera has been done by Fernyhough et. al. [2]. The relative position of objects that are close to the reference object are given as well as a grid around the reference object. The areas around it are named: Ahead, Ahead Right, Right, Behind Right, Behind, Behind Left, Left, and Ahead Left.

2.1 The Directed Moving Object Model

Although this model is very simple in itself, it can be used as the basis for defining more detailed position descriptions and transitions; it can also be used as a framework for relating different aggregation levels.
In the model, shown in Fig. 1, the plane in which the two-dimensional directed object (the reference object) is situated is divided into nine two-dimensional regions. The arrow indicates the direction of the reference object relative to which the resulting regions are named. These are the relative positions where another object, the object in focus, can be in relation to the reference object. The square in the middle is the place where the reference object itself is situated.



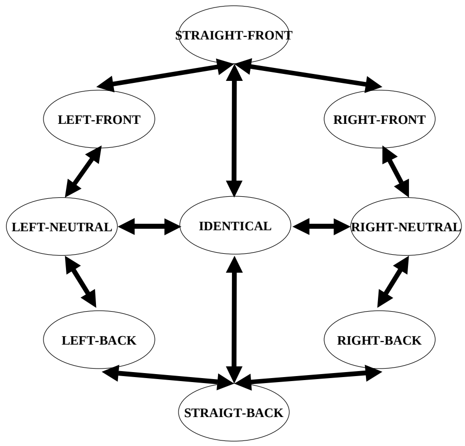


Fig. 1. The directed moving object model

Fig. 2. The conceptual neighborhood graph

If an object is in one of those regions this can be stated in natural language. It is sufficient to only state it once without mentioning any changes in position as long as these are within the same qualitative area. The second thing we would like to express are changes between the areas. There are 24 primitive transitions that describe a change from one area to a conceptual neighbor area if we assume a four-neighborhood of the qualitative areas.
The nine positions where an object can be located with regard to the reference object and the 24 possible transitions between those positions can be organized in the conceptual neighborhood graph shown in Fig. 2. For the traffic domain where movement is normally continuous it is sufficient to only concentrate on continuous processes. That means that all movement can be described along the transitions of the conceptual neighborhood graph. A real maneuver can then be seen as an actual path through that graph.

2.2 Usage of the DMO Model

The input in the system can be a sequence of snapshots of the scene of interest wherein all objects have been identified or a sequence of messages about state transitions. As an example we give the description of an overtaking maneuver: We concentrate on one object (the vehicle in focus) that is the white car in Fig. 3, and want to describe its actions. First we find a reference object that will be the black car in front of it.

Around the reference object the DMO areas are established which makes it possible to state that the white car is in the straight back area as shown snapshot a) in Fig.3. While formulating the natural language expression we will take into account what kind of objects participate in the maneuver and adjust the terminology in the way humans would do. In this case the first snapshot would lead to the sentence: “The white car is driving behind the black car.” For the following snapshots the statements “The white car is sheering out from behind the black car”, “It is driving left behind the black car”, “It is catching up to the black car, driving beside it, passing it, driving left in front of it and is sheering in in front of it ” and finally “The white car is driving in front of the black car”. It is not always appropriate to describe every step of a maneuver. On a higher level of communication the whole chain of elementary actions can be grouped together and stated as one overtaking maneuver, as well as e.g. several overtaking maneuvers in a row can be grouped to an overtaking of a queue.

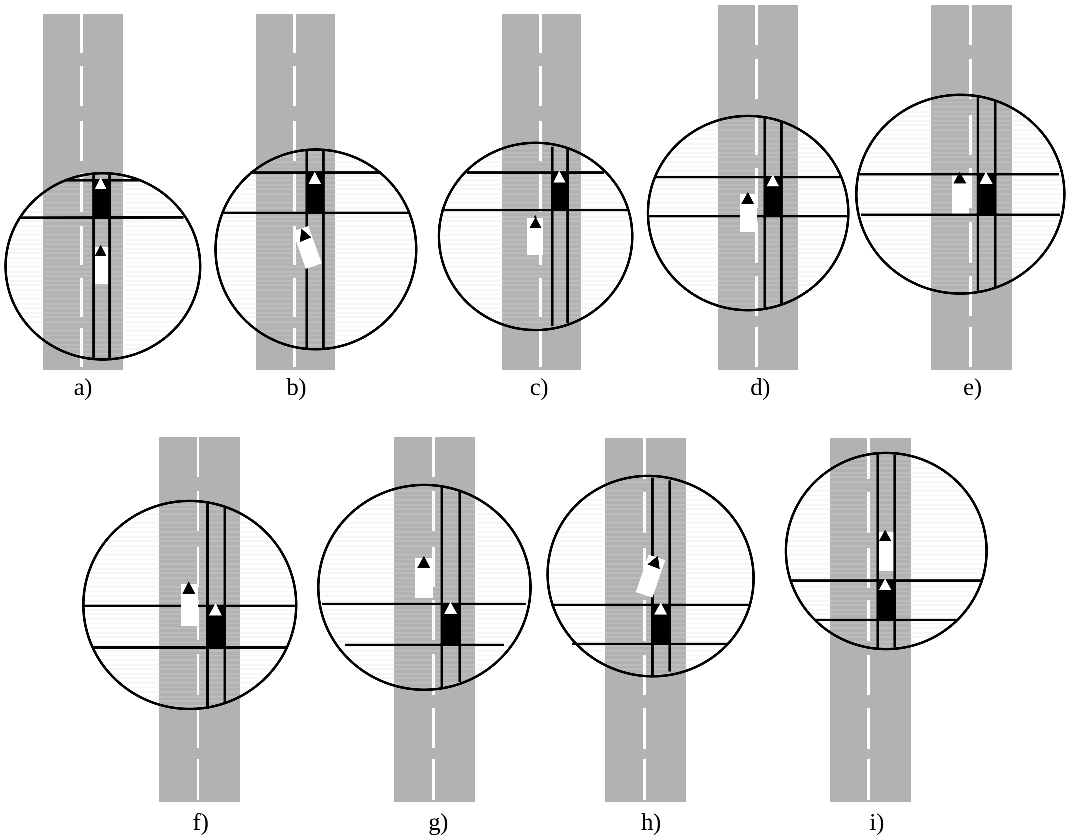


Fig. 3. Snapshots of the overtaking maneuver. The black car is the reference object; the white car is the vehicle in focus which actions are described according to its position relative to the reference object. a) driving behind b) sheering out from behind c) driving left behind d) catching up e) driving beside f) passing g) driving left in front h) sheering in in front i) driving in front.

3 Future Work

For the future of our project we plan to collect a suitable number of traffic maneuver descriptions that include maneuvers with several reference objects, combined maneuvers, and complex traffic situations and combine them with the dialog system that the helicopter uses to communicate with the operator. Prediction of traffic maneuvers will be part of this work. Furthermore we will decide on a modeling language for our approach.

References

1. P. Doherty, G. Granlund, K. Kuchcinski, E. Sandewall, K. Nordberg, E. Skarman, and J. Wiklund: The WITAS Unmanned Aerial Vehicle Project. In Proceedings of the 14th European Conference on Artificial Intelligence. IOS Press Amsterdam (2000) 747-755

2. J. Fernyhough, A. G. Cohn and D. Hogg: Constructing qualitative event models automatically from video input. In Image and Vision Computing, 18 (2000) 81-103)

3. Kenneth D. Forbus: Qualitative Reasoning. In Allen B. Tucker (Ed.): The Computer Science and Engineering Handbook. CRC Press (1997) 715-733

4. C. Freksa: Using orientation information for qualitative spatial reasoning. In A. U. Frank, I. Campari, U. Formentini (Eds.): Theories and methods of spatio-temporal reasoning in geographic space. Springer, Berlin (1992) 162-178

5. Roop, K. Goyal, Max, J. Egenhofer: Similarity of Cardinal Directions. In C.S. Jensen et al. (Eds.): Advances in Spatial and Temporal Databases. SSTD 2001, LNCS 2121. Berlin Heidelberg: Springer-Verlag (2001) 36-55

6. Amitabha Mukerjee, Gene Joe: A qualitative model for space. In Proceedings of the AAAI-90. The MIT Press Boston (1990) 721-727

7. Erik Sandewall, Patrick Doherty, Oliver Lemon, and Stanley Peters: Words at the Right Time: Real-Time Dialogues with the WITAS Unmanned Aerial Vehicle. In: Andreas Guenter et al (eds): KI 2003: Proceedings of the 26th Annual German Conference on AI.
Springer (2003)

8. WITAS homepage
http://www.ida.liu.se/ext/witas/

View-Dependent Spatial Knowledge Representation by Snapshots of Ordering Information

Thomas Wagner[1] and Otthein Herzog[2]

[1] Center for Computing Technologies (TZI), Universität Bremen, D-28359 Bremen
`twagner@tzi.de`
[2] Center for Computing Technologies (TZI), Universität Bremen, D-28359 Bremen
`herzog@tzi.de`

Classical studies in cognitive science indicate that the spatial representation used for navigation and localization tasks in cognitive systems is based on a static and allocentric *cognitive map* (see, e.g., [9], [5]). Nevertheless, recent studies propose a different view. First evidence against the concept of a static allocentric cognitive map was raised by [1]. They showed that a viewpoint-independent representation, as claimed by the cognitive map thesis, can only be generated under specific, i.e. unnatural conditions. Wang and Spelke are the first who argue clearly against the cognitive map approach [12] (see also the acknowledgement of Wang and Spelke in [12]) and claim that most of the observations and results from spatial cognition can be explained in two different ways, either by updating the egocentric movement in an allocentric map using movement information (path integration), or in contrary, that humans build a complete egocentric representation and are able to map between *different views*. Similar strong evidence has been provided by the research group of McNamara (in particular, [4], [3], [7], [6]). They showed in a series of experiments that a viewpoint-independent representation can only be achieved when the observed configuration of objects provide certain geometric properties, i.e., that the configuration of objects can be aligned with the surrounding environment (e.g., a room) and when the object configuration shows geometric properties like parallelism (*intrinsic frame of reference* in ([2], [4])).

On the poster we show how the snapshot-based approach to ordering information (see, e.g., [11], [10]) accounts for several important observations of the experiments from [4], [3], [7] and [8].

Literatur

1. R.D. Easton and M.J. Sholl. Object-array structure, frames of reference, and retrieval of spatial knowledge. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21(2):483–500, 1995.
2. W. Mou and T.P. McNamara. Intristic frames of reference in spatial memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 28(1):162–170, 2002.
3. W. Mou, T.P. McNamara, C.M. Valiquette, and B. Rump. Allocentric and egocentric updating of spatial memories. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 30(1):142–157, 2004.
4. W. Mou, K. Zhang, and T.P. McNamara. Frames of reference in spatial memories aquired from language. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 30(1):171–180, 2004.
5. J. OKeefe and A. Speakman. *The Hippocampus as A Cognitive Map.* Oxford, Clarendon Press, 1978.
6. A.L. Sheldon and T.P. McNamara. Systems of spatial reference in human memory. *Cognitive Psychology*, 43:274–310, 2001.
7. A.L. Sheldon and T.P. McNamara. Orientation and perspective dependence in route and survey learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 30(1):158–170, 2004.

8. D.J. Simons and R.F. Wang. Percieving real-world viewpoint changes. *Psychological Science*, 9(4):315–321, 1998.
9. E.C. Tolman. Cognitive maps in rats and men. *Psycholoical Review*, 55:189–208, 1948.
10. T. Wagner and K. Huebner. An egocentric qualitative spatial knowledge representation based on ordering information for physical robot navigation. In *Proceedings of Paper the RoboCup VII Symposium, to appear*, 2004.
11. T. Wagner, U. Visser, and O. Herzog. Egocentric qualitative knowledge representation for physical robots. *Journal for Robotics and Autonomous Systems, to appear*, 2005.
12. R.F. Wang and E.S. Spelke. Human spatial representation: Insights from animals. *Trends in Cognitive Science*, 6(9):176–182, 2002.

Spatial Representations for Graphical Reasoning

Lara Webber & Aidan Feeney

University of Durham
{l.j.webber; aidan.feeney}@durham.ac.uk

Abstract. We describe an experiment designed to investigate whether people represent information presented graphically by analogy to space. Participants were presented with pairs of line graphs and bar graphs that specified the relationships between pairs of terms (A, B; B, C). The order of these terms was manipulated so that the end terms were either separated by the repeated term (e.g. ABBC) or were adjacent (e.g. BACB). Participants were asked to reproduce one of the premise graphs or to draw a graph relating the A and C terms. The order of the terms in these conclusions suggests that participants reordered the terms in the premise graphs so as to produce an integrated linear model of the information in the graph. Spatial ability was found to be associated with the tendency to reorder. These results suggest that people create spatial mental models for graphical reasoning.

1. Introduction

Whilst there is universal agreement that it is the spatial nature of graphs and diagrams that underlies their benefits for communication [see 1; 2], there has been very little work that has investigated the representations that people build when they interact with graphs. Thus, graphs and diagrams are held to be efficient because their spatial nature facilitates search, makes certain information explicit, obviates the need for sophisticated inferencing, and makes them isomorphic to the phenomena that they represent, yet we do not know whether people's representations of graphs and diagrams possess spatial characteristics. To be sure, many researchers assume that there is a correspondence between external diagrammatic representations and our internal representations of what they convey [for a discussion see 3] but, until recently, there has been little or no direct evidence that the mental representation of graphs and diagrams can be spatial in nature and models of graph comprehension tend to assume that people represent graphical information using language-like representations [see 4] or are neutral with respect to mental representation [see 5; 6].

We have been carrying out work for a number of years attempting to explore the characteristics of people's mental representations of information presented graphically [7; 8; 9]. In this work we have adapted tasks that have primarily been used to study mental representations for reasoning. For example, Webber & Feeney [9] gave participants a series of three term reasoning problems ($A > B$, $B > C$ $\therefore A > C$) where the premises were presented graphically. Participants saw pairs of premise graphs (see Figure 1a) and then had to verify whether a conclusion graph (see Figure 1b) followed logically from the premises. For our present purposes the two key manipulations in this experiment were of the order of the terms in the premises (their figure) and the order of the terms in the conclusion. We found that participants spent longer encoding premises where the repeated term was separated by the end terms (e.g. BACB) than those where the end terms were separated by the repeated term (e.g. ABBC). In addition, we found that for ABBC problems participants took longer, and made more errors, in verifying CA conclusions than AC conclusions. However, for BACB problems these findings were reversed.

These results are very similar to effects of figure that have been reported across a variety of logical forms in the literature on deduction. They suggest that people are reordering the referents in the graph during the course of building up an integrated mental representation. This reordering occurs because people’s mental representations of the graphical premises take the form of a linear ordering of the referents. Such representations make information about relationships explicit, thus facilitating relational inferences, and are isomorphic in structure to the phenomena that they represent. These findings suggest that people build spatial representations for graphical reasoning.

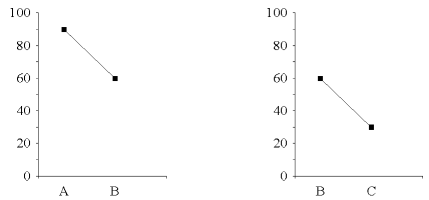


Fig. 1a. Two simple line graphs with the end terms separated by the repeated term.

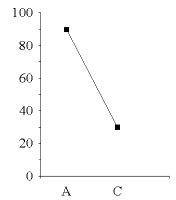


Fig. 1b. A valid conclusion graph in which term order is consistent with term order in Figure 1a.

In the experiment to be described here we were interested in whether we could obtain evidence of spatial reordering when people were asked to produce conclusions relating the end terms in sets of graphical premises. We also wished to examine whether our indices of spatial representation are sensitive to individual differences in spatial ability. That is, are people who are high in spatial ability more or less likely to reorder the terms in premises presented graphically?

2. Experiment

Method: We presented a series of 16 graphical reasoning problems to 52 participants via a computer screen. Once participants had indicated that they had looked at each set of premises (see Figure 2a) for long enough, on the next screen was a prompt requiring them to reproduce one of the premise graphs or to produce a conclusion relating the end terms. In half of the premise graphs the end terms were separated by the middle term (i.e. ABBC) and in the other half the end terms were adjacent (BACB). We coded the conclusions produced according to whether the order of the terms in the conclusion was consistent (see Figure 2b) or inconsistent (see Figure 2c) with their order in the premise graphs (see Figure 2a). In addition, as a measure of spatial ability we asked participants to complete a standard mental rotation task.

Predictions: As they do not require reordering, we predicted that most conclusions produced for ABBC graphs would be consistent. In the case of BACB graphs, one likely strategy that leads to an integrated linear representation is to reorder the premises so that term order becomes CBBA. Use of such a strategy would result in conclusion graphs that were consistent with the order of the terms as they appeared on the screen becoming inconsistent with their order in the mental representation. Accordingly we predict fewer consistent conclusions for BACB premise graphs. Because of the reordering process necessary to produce an integrated spatial representation of BACB graphs, we predicted a significantly more positive association between spatial ability and the tendency to produce consistent conclusions in response to ABBC graphs than between ability and the tendency to produce consistent conclusions in response to BACB graphs.

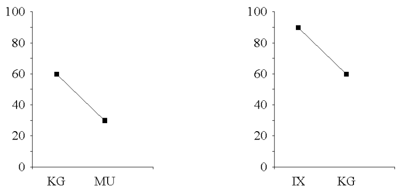


Fig. 2a. Premise graphs for ABBC problem

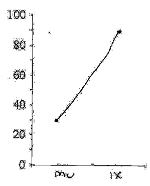


Fig. 2b. A consistent conclusion

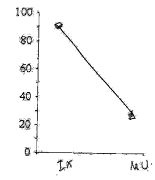


Fig. 2c. An inconsistent conclusion

Results: Five participants appeared to misunderstand the instructions so their data was excluded from all subsequent analysis. For the remaining 47 participants spatial ability was positively correlated with rate of correct responding on experimental trials (r = .40, p < .01). Thus, overall there was an association between spatial ability and performance on the task.

To test specific predictions about the effects of term order on how people mentally represented the graphical information we categorised participants as high or low in spatial ability and carried out a 2x2 mixed design ANOVA with repeated measures on the term order factor. As predicted, a much higher percentage of the conclusions produced for ABBC graphs were consistent (90%) than was the case for BACB graphs (33%), $F(1, 43) = 96.75$, $MSE = 696.134$, p <. 001. The analysis also revealed a significant interaction between term order and ability, $F(1, 43) = 4.08$, $MSE = 696.134$, p <. 05. Tests for simple effects revealed that high ability participants produced significantly more consistent conclusions in response to ABBC graphs than did low ability participants. Although this pattern was reversed for BACB graphs, the difference was not statistically significant.

Finally, individual differences analyses revealed that the correlation between spatial ability and the production of consistent conclusions in response to ABBC graphs (r = .36) was significantly (p < .02) more positive than the correlation between ability and consistent conclusions when the order of the terms was BACB (r = -.09).

3. Conclusions

We have found evidence for reordering in people's mental representations of information presented graphically. The tendency to reorder is associated with spatial ability, suggesting that people draw on general spatial resources in order to manipulate their mental representations of graphical information. Some people represent graphical information by analogy to space and the representational format they appear to adopt possesses several of the spatial characteristics held to make graphs and diagrams cognitively efficient.

References

1. Larkin, J. H. & Simon, H. A. Why a diagram is (sometimes) worth 10,000 words. Cognitive Science Vol. 11 (1987) 65-99
2. Shimojima, A. The graphic-linguistic distinction. Artificial Intelligence Review Vol. 13 (1999) 313- 336
3. Scaife, M. & Rogers, Y. External cognition: How do graphical representations work? International Journal of Human-Computer Studies Vol. 45 (1996) 185-231
4. Pinker, S. A theory of graph comprehension. In R. Freedle (Ed.), Artificial Intelligence and the Future of Testing. Hillsdale: Lawrence Erlbaum Associates (1990) 73-126
5. Lohse, G. L. A cognitive model for understanding graphical perception. Human-Computer Interaction Vol. 8 (1993) 353-388
6. Carpenter, P. A. & Shah, P. A model of the perceptual and conceptual processes in graph Comprehension. Journal of Experimental Psychology: Applied Vol. 4 (1998) 75-100
7. Feeney, A., Hola, A. K. W., Liversedge, S. P., Findlay, J. M., & Metcalf, R. How people extract information from graphs: Evidence from a sentence-graph verification paradigm. Lecture Notes in Artificial Intelligence Vol. 1889 (2000) 149-161
8. Feeney, A. & Webber, L. J. Analogical representation and graph comprehension. Lecture Notes in Computer Science Vol. 2733 (2003) 212-221
9. Webber, L. J. & Feeney, A. How people represent and reason from graphs. In Proceedings of the 25th Annual Conference of the Cognitive Science Society. Mahwah, NJ: LEA (2003) 1206-1211

Dissociable retrosplenial and hippocampal contributions to successful formation of allocentric memory representations

Thomas Wolbers & Christian Büchel

NeuroImage Nord, Department of Neurology, Universitätsklinikum Hamburg-Eppendorf, Germany
wolbers@uke.uni-hamburg.de

1 Introduction

Even though hierarchical models have conceptualised spatial knowledge derived from ground-level navigation as a succession of landmark, route and survey knowledge, there is growing evidence for the potential to develop route or survey knowledge from the very beginning of a learning experience. In this context, the importance of the applied learning strategy for the resultant representation has been emphasised repeatedly.

Whereas many neuroimaging and lesion studies on spatial navigation have successfully investigated the retrieval of previously learned environments [1,3,4], the neural foundations of learning complex spatial layouts have received less attention [2,6]. Even though these studies have identified a network of areas including frontal, posterior parietal, retrosplenial and medial temporal regions, the precise role of these structures remains controversial, particularly with respect to parahippocampal and hippocampal function. One reason might be the lack of a behavioural measure allowing for a clear distinction between different representations, since many navigational tasks can be performed based on either route or survey representations.

In the current study we aimed at identifying the neural foundations of survey learning by instructing subjects explicitly to infer the spatial layout of a complex virtual environment and the locations of twelve landmarks. To obtain a sensitive behavioural parameter allowing for a reliable identification of survey knowledge, we employed a specific environment that provides accuracy and reaction time measures precisely indicating the emergence of route vs. survey knowledge [8].

2 Methods

Eleven participants were included in the study that comprised six learning sessions. During encoding, subjects were moved throughout the environment, thereby encountering twelve distinct landmarks. Encoding was immediately followed by a retrieval condition requiring subjects to assess spatial relationships between pairs of buildings. Whereas the Euclidean distance between each of two buildings was held constant across pairs, differences in egocentric distance travelled between buildings were introduced. Buildings constituting *direct* pairs were separated by only one path segment during navigation, *close* and *remote* pairs contained buildings that were separated by largely varying numbers of up to eleven path segments.

This discrepancy between constant spatial distance and varying temporal delays provided an objective behavioural measure reflecting the acquisition of route vs. survey representations. If subjects were able to infer a map-like representation, performance improvements should emerge for all three categories of retrieval pairs, thus being independent of the varying temporal delays. Furthermore, we included additional sessions to control for perceptual input across the learning process.

3 Results

Participants were classified as allocentric learners based on their behavioural performance. Using the state-space smoothing algorithm developed by Smith and colleagues [7], we identified learning trials for direct, close and remote pairs in all eleven participants. The behavioural performance was characterised by increasing accuracy and decreasing response times across sessions for all types of retrieval pairs.

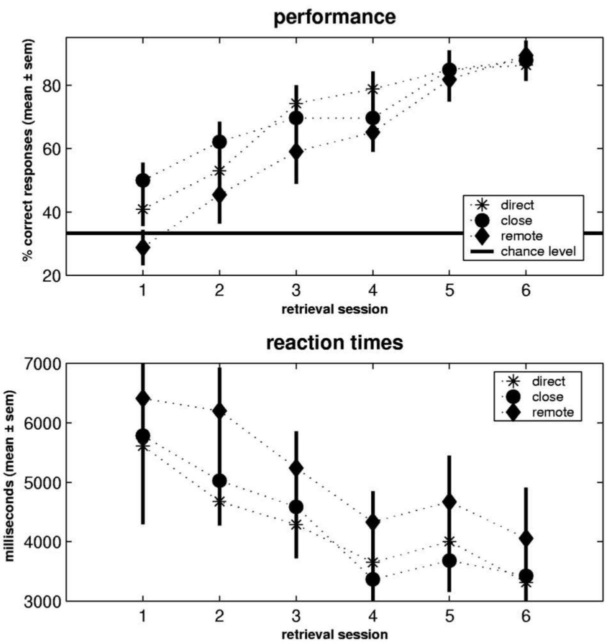


Fig. 1. Performance and reaction time data during retrieval (n=11). Significant changes across sessions were obtained for direct, close and remote pairs, thereby indicating the emergence of allocentric representations.

To identify the neural bases of survey learning we first contrasted activation during encoding with the control conditions. Neuroimaging results revealed a main effect of learning the spatial layout of the environment in frontal and posterior parietal cortices. Moreover, we observed increasing activation across learning sessions in the retrosplenial cortex, thus paralleling behavioural measures of survey expertise. In contrast, hippocampal activation did not follow the estimated learning curves, but reflected the amount of knowledge acquired in a given session. In other words, hippocampal activation was most prominent during the initial learning phase and decayed after performance had approached ceiling level.

4 Discussion

In conclusion, successful transformation of a complex environment experienced from ground-level perspective into an observer-independent, allocentric representation requires the dynamic interplay of medial temporal, posterior parietal, retrosplenial and frontal regions. Whereas the recruitment of frontal areas in our paradigm presumably reflects working memory processes necessary for successful long-term encoding of the spatial layout, parietal regions are known to process spatial positions and relations in egocentric coordinates. The subsequent transformation into an allocentric reference system crucially depends on retrosplenial cortex as reflected by the performance dependent activation. This region is thought to act as a transition zone between egocentric and allocentric reference frames [5], allowing for the conversion of stimulus information into the appropriate reference frame for the task at hand. In contrast, the hippocampal formation appears to support the incorporation of new spatial information into an existing memory representation. With increasing accuracy of this representation the hippocampal contribution becomes less important, thus leading to the observed activation decrease during the late phase of learning.

References

1. G.K. Aguirre, M. D'Esposito, Environmental knowledge is subserved by separable dorsal/ventral neural areas, J Neurosci 17 (1997) 2512-2518.
2. J. Barrash, H. Damasio, R. Adolphs, D. Tranel, The neuroanatomical correlates of route learning impairment, Neuropsychologia 38 (2000) 820-836.
3. T. Hartley, E.A. Maguire, H.J. Spiers, N. Burgess, The well-worn route and the path less traveled. Distinct neural bases of route following and wayfinding in humans, Neuron 37 (2003) 877-888.
4. G. Iaria, M. Petrides, A. Dagher, B. Pike, V.D. Bohbot, Cognitive strategies dependent on the hippocampus and caudate nucleus in human navigation: variability and change with practice, J Neurosci 23 (2003) 5945-5952.
5. E.A. Maguire, The retrosplenial contribution to human navigation: a review of lesion and neuroimaging findings, Scand J Psychol 42 (2001) 225-238.

6. E.A. Maguire, C.D. Frith, N. Burgess, J.G. Donnett, J. O'Keefe, Knowing where things are parahippocampal involvement in encoding object locations in virtual large-scale space, J Cogn Neurosci 10 (1998) 61-76.
7. A.C. Smith, L.M. Frank, S. Wirth, M. Yanike, D. Hu, Y. Kubota, A.M. Graybiel, W.A. Suzuki, E.N. Brown, Dynamic analysis of learning in behavioral experiments, J Neurosci 24 (2004) 447-461.
8. T. Wolbers, C. Weiller, C. Büchel, Neural foundations of emerging route knowledge in complex spatial environments, Cog Brain Res (in press).