Verbal Assistance in Tactile-Map Explorations: A Case for Visual Representations and Reasoning

Christopher Habel, Matthias Kerzel, Kris Lohmann

Department of Informatics, University of Hamburg
Vogt-Kölln-Str. 30; 22527 Hamburg, Germany
{habel, kerzel, lohmann}@informatik.uni-hamburg.de

Abstract

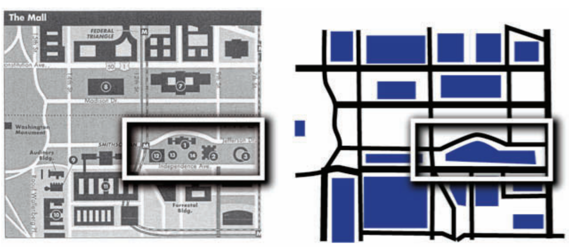
Tactile maps offer access to spatial-analog information for visually impaired people. In contrast to visual maps, a tactile map has a lower resolution and can only be inspected in a sequential way, complicating the extraction of spatial relations among distant map entities. Verbal assistance can help to overcome these difficulties by substituting textual labels with verbal descriptions and offering propositional knowledge about spatial relations. Like visual maps, tactile maps are based on visual, spatial-geometric representations that need to be reasoned about in order to generate verbal assistance. We present an approach towards a verbally assisting virtual-environment tactile map (VAVETaM) realized on a computer system utilizing a haptic force-feedback device. In particular, we discuss the tasks of understanding the user's map exploration procedures (MEPs), of exploiting the spatial-analog map to anticipate the user's informational needs, of reasoning about optimal assistance by taking assumed prior knowledge of the user into account, and of generating appropriate verbal instructions and descriptions to augment the map.

1 Introduction

Tactile maps provide blind and visually impaired people with useful means to acquire knowledge of their environment. As such, they can be used as substitutes for visual maps, see Ungar, Blades, and Spencer (1993). As Espinosa et al. (1998) point out, tactile maps can potentially increase the independence and autonomy of blind and visually impaired people, since they make navigation in a complex urban environment possible without the assistance from a sighted guide. Compared to visual maps one drawback of tactile maps is the restriction of the haptic sense regarding the possibility of simultaneous perception of information, for an overview see Loomis, Klatzky and Lederman (1991). The additional effort in haptic perception leads in the case of map exploration to specific limitations for building up cognitive maps, such as sparse density of information and disadvantage of survey knowledge compared to route knowledge. Due to the restriction of the haptic sense in simultaneous perception of information, additional information given in another modality, e.g., speech, can be very useful (Wang, Li, Hedgpeth, and Haven 2009). The increasing availability of haptic interfaces for human-computer interaction (HCI) offers a large variety of prospects for training and assisting blind people. In particular, by the means of such devices (such as the PHANToM® desktop) it is possible to realize map-like representations of physical environments that are HCI-counterparts to traditional tactile maps (Kostopoulos et al. 2007; Lahav and Mioduser 2008). Virtual-environment (VE) tactile maps offer the option to generate verbal descriptions. Thus, both representational modalities, maps and language, can be used for providing spatial information. In particular, the sequential nature of verbal descriptions supports incremental construction and updating of spatial knowledge.

Although some proposals for and approaches to verbally augmented tactile-map systems exist, these systems do not take the generation of complex natural language descriptions in interaction with the user's exploratory movements into account (Wang, Li, Hedgpeth, and Haven 2009; Jacobson 1998; De Felice et al. 2007; Parente and Bishop 2003). In contrast to these approaches, we propose knowledge-based generation of verbal instructions and descriptions that are elicited by the user's tactile explorations of the VE tactile map. With an abstract semantic categorization of the users' movements (see Section *Semantic Representation and Segmentation of the User's Movements*), knowledge about what they explore can be used to compute verbal assistance in scenarios where verbally augmenting the haptic representation provides additional hints either for further exploration procedures or for the efficiency of building up spatial knowledge of the environment (re-)presented in the map. As Wang and colleagues (2009) describe, besides the description of labels in visual maps, users demand information about locations of 'auditory landmarks' like audio-enabled traffic lights and further information about the relations of complex entities, such as long roads.

The usefulness of improving tactile maps with verbal descriptions is exemplified by maps depicting the *National Mall at Washington* (see Figure 1)[1]: Even though the visual map (Figure 1.a) does not contain much details, elaborated information about the shape of buildings is not suitable to be represented in a tactile map (Figure 1.b). Instead, verbal descriptions should be used to communicate further knowledge about a given entity. This can be information about the name of buildings, other landmarks (e.g., *'This is the Washington Monument'*), or even information about a block of buildings too complex to be represented in the haptic modality like the one marked in Figure 1. An appropriate verbal description could be: '*The landmark you are exploring consists of four large buildings. In the west is Freer Gallery of Art. In the east is Hishborn Museum and Sculpture Garden. In between there is the Smithonian Institution Building and the Arts and Industries Building.'* This example considers an important class of problems and phenomena: The level of granularity used in representations either in verbal descriptions or in maps. During route planning humans use different levels of granularity, as shown by Timpf and Kuhn (2003) for the highway domain. It is likely that granularity transformations also happen during route planning by visually impaired or blind people. Using the example of the four buildings described above: When planning a route simply passing by the museum buildings, it may be sufficient to inform the user that there are buildings, whereas when planning a route to the Hishborn Museum, more detailed information about the location and the spatial relations among the buildings should be given verbally.



(a) Visual Map (b) Abstraction for Tactile Map

Figure 1: Example of an abstraction for a tactile map of the National Mall of Washington

External as well as internal representations play a major role in human cognition and especially in human problem solving. Additionally, another dichotomy plays a major role in Cognitive Science and Artificial Intelligence: *Propositional* representations—often characterized as *linguistic* or *language oriented*, which contain language as well as number symbols—in opposition to *analog* representations, also called *visual, diagrammatical,* or *image-like,* e.g., maps, diagrams, etc. The benefits of analog representations[2]—whether they are internal or external, whether they are called pictorial, spatial or diagrammatic—are based on their property of sharing relevant inherent constraints with the domain they represent. In contrast, propositional representations must code such constraints, which are implicitly embodied in analog representations, explicitly (see Palmer 1978; Haugeland 1987; Habel 2003a). Additionally, the conception of *visual representations* and *visual reasoning* foregrounds the usefulness of principles of perceptual organization for reasoning and problem solving: Such—in reasoning about maps—neighborhood of regions or connectedness of lines *pop out* by perception and have not to be deduced via chains of inferences.

Maps, in particular *visual maps*, have mostly the character of *hybrid representation systems* (see Myers and Konolige 1995; Habel 2003a). On the one hand, they usually rely on linguistic labels and symbols, on the other hand, they provide visual-spatial representations 'analog' to the geographical world. Furthermore, there exist conventions about map semantics, i.e., the systematic semantic links among graphical entities constituting maps (*map concepts, MCs)*, their counterparts on the layers of concepts represented propositionally and entities in the external world, e.g., that streets are depicted as lines or water is depicted as a blue area.

2 Verbally Assisting Virtual-Environment Tactile Maps

Looking from the user's perspective the VAVETaM system (*verbally assisting virtual-environment tactile map* system) we outline in this paper, consists of a VE tactile map provided by a haptic device that is augmented by verbal assistance. Thus, it offers a multimodal interaction to the user. The perceivable VE tactile map is based on a virtual three-dimensional haptic space, which can be explored by manually moving a *virtual interface point (IP)* with a three-dimensional pointing device and perceiving 3D-graphics-like virtual objects by force feedback. The virtual tactile map can be displayed to a sighted user for control and assistance purposes. Figure 3 shows a visualization of a VE tactile map, in which streets and landmarks are rendered as depressed areas on a map plane.

Figure 2 shows the architecture of the VAVETaM. The *Virtual-Environment Tactile Map* (VETM) represents a hybrid map and thus consists of two layers: The *Spatial-Geometric Layer* provides a model for haptic perception mediated by the *Haptic Device* and is also used to facilitate categorization of the user's movements in the haptic space in relation to the explored map. The *Propositional Layer* serves as basis for communication by providing the propositions used to generate verbal assistance.

[1] The depiction of the left map is derived from: Fodor's Washington, D.C. 2001, page 29. © Fodor's Travel Publications; Random House: New York.

[2] Strictly speaking, analog representations are analog with respect to specific aspects of the domain to be represented. The assumed structural correspondence between representation and domain holds with respect to some prominent properties and relations (see Palmer 1978).

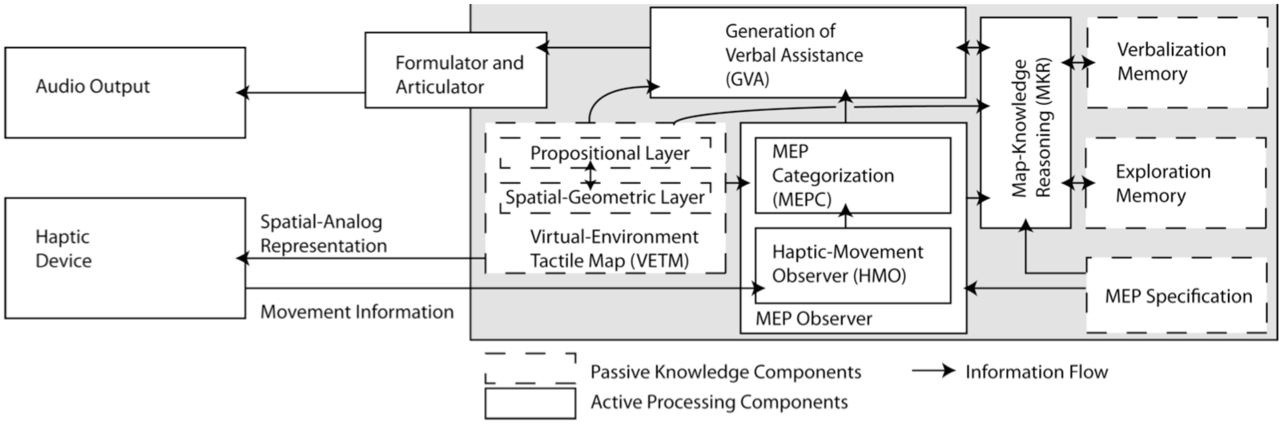


Figure 2: Structual model of VAVETaM

Exploration of VE tactile Maps. In haptic exploration users instantiate *map exploration procedures (MEPs)*, which are realized as movements. The Haptic Device provides position and hence movement information. These map movements are related to the Spatial-Geometric Layer of VETM, and are represented abstract and semantically. To facilitate this, the movements are categorized in the *MEP Observer* component according to the taxonomy of MEPs provided by the *MEP Specification* component. The MEP Observer consists of the *MEP Categorization (MEPC)* and the *Haptic-Movement Observer (HMO)*, which are further discussed in the next section. For the generation of verbal descriptions, it is essential to analyze the user's exploration, in particular, to build assumptions about what parts of the map are known and what parts are still unexplored. This reasoning is done by the *Map-Knowledge Reasoning* component[3] *(MKR)*, which is linked to both a memory of verbalized assistance (*Verbalization Memory*) and the users map exploration procedures (*Exploration Memory*).

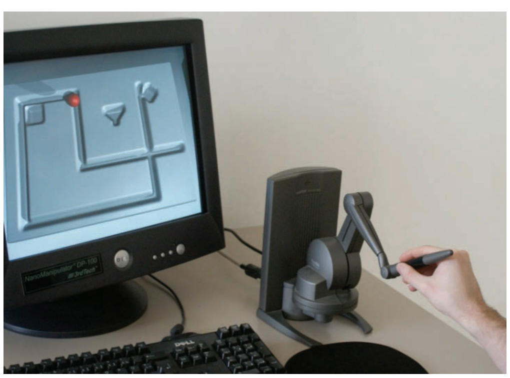


Figure 3: PHANToM® desktop and visualized VETM

The output of the MEP Observer and the assumptions about the user's proposed knowledge are used within the *Generation of Verbal Assistance (GVA)* component to generate propositional, pre-verbal messages (Levelt 1989) in order to satisfy *informational needs* (Pirolli and Card 1999), see Section 4. Once such a pre-verbal description is generated, it is stored in the *Verbalization Memory*.

The VETM Component. In the *Introduction* the need for a model of the spatial representation within the VETM was described. In order to construct such a model, the formats for representing spatial knowledge within maps were discussed.

Generally, it is plausible to assume that tactile maps work in the same way as visual maps, even though map concepts vary due to the representational possibilities of the tactile-map setup, that is, the resolution and complexity that can be perceived haptically and understood sequentially is reduced. As in the visual scenario, within the VETM component one layer is spatial-geometric. In addition to this spatial-geometric representation, a propositional layer includes conventional information about MCs and specific labels linked to the analog representation (compare Maaß (1994) for a similar approach). This allows computing the information to be verbalized in the GVA component (see Section 4).

3 Semantic Representation and Segmentation of the User's Movements

MEPs as Semantic Concepts for Language Generation. The goal of the VAVETaM project is to generate verbal assistance for VE tactile map exploration, both to communicate labeling information like street and building names and to assist the user's exploration. To do so, the natural language generation process has to be based on the user's interaction with the VE tactile map. In this respect, the generation process differs from generating in-advance route instructions or the description of complete static networks (Habel 2003b, Levelt 1989). The user interacts with the IP of the haptic device with the VAVETaM. The map entities like streets and landmarks are explored in a sequential way. As during exploring a hardcopy tactile map, the user has to start at one point of the map and sequentially gather information about spatial relations

[3] In our current design the Map-Knowledge Reasoning does not try to recognize the user's plan, though this might be a focus for future research (we thank an anonymous reviewer for this remark).

among map entities resulting in a movement of the IP. Even though the movements are not generated with the purpose of controlling the generation of verbal assistance, they are the input to VAVETaM, determining the verbal output. The VE tactile map of VAVETaM is realized as a plane with depressed areas indicating the existence of landmarks and tracks (see Section 4). The human-map interaction is realized within this plane area and can thus be described two-dimensionally. Hence, the interpretation of the movements of the IP can follow similar reasoning procedures as the interpretation of visual sketch gestures (Adler and Davis 2009, Habel and Tappe 1999).

The generation of verbal assistance is based on a semantic representation of the movements, i.e., the continuous movements have to be categorized. An example can clarify the need for categorization: To be able to generate a verbal output like *'You are exploring Independence Avenue, which is parallel to Constitution Avenue,'* it is necessary to understand that the user is exploring the track segments t_2 and t_6 denoting Independence Avenue in Figure 4. This reasoning is made in the MEP Observer component, which has access to the movement information of the haptic device and to the MEPs Specification component.

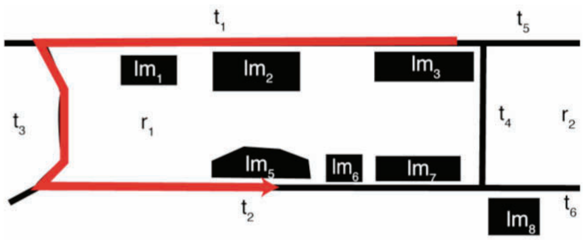


Figure 4: A part of a visualized VAVETaM exploration of Washington with labels (*t* denoting track segments, *lm* landmarks and *r* regions)

The classification made to categorize the fluidly sequential movement into semantic categories called map exploration procedures (MEPs) is cognitively motivated. Basically, it follows Lederman's and Klatzky's seminal investigation that haptic exploratory procedures (EPs) are systematically related to kinds of desired knowledge about the object (Lederman and Klatzky 1987). Exploring an object haptically includes several types of contact, e.g., static contact and exploration over time. While static contact is suitable to gain knowledge about the temperature of an object such as a bottle of water, exploration over time, such as contour following, is suitable to gain knowledge about the shape of objects.

A discussion of MEPs necessarily has to start with an investigation of the knowledge categories involved in navigational map reading. This knowledge is strictly domain-dependent. An investigation of the cognitive categorization of geographic objects is described in Mark, Smith, and Tversky (1999).

In a user study, which was accomplished on a hardcopy tactile map allowing only one-finger exploration and later video analysis, we focused on the investigation of exploration of tactile maps of urban domains. The resulting preliminary set of MEPs has yet to be tested for its reliability for VAVETaM. The main knowledge categories for navigation are (1) *tracks* (the term track is used as a general term for street-like structures involved in route planning), (2) *landmarks*, and (3) *regions*.

Besides this set of three knowledge categories, knowledge about the (4) *frame* of the tactile map is necessary to integrate the percepts to form knowledge about global spatial relations, especially as the frame of a rectangular tactile map is used as reference for the investigation of the orientation of linear objects, such as streets. In conclusion, our hypothesis is that the four knowledge categories result in four MEPs. Those are: (a) *track-MEP*, (b) *landmark-MEP*, (c) *region-MEP*, and (d) *frame-MEP*.

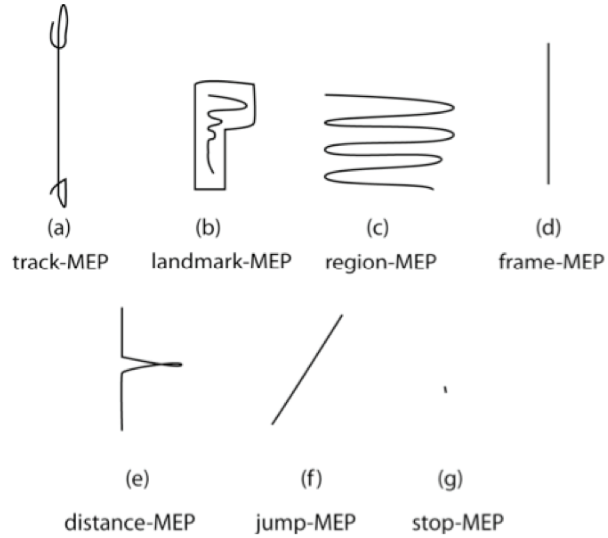


Figure 5: Movement patterns (preliminary set of MEPs)

The user study revealed the regular occurrence of three further MEPs, which seem to fulfill functions of checking and correction, a phenomenon described for tactile exploration by Zinchenko and Lomov (1960). These three MEPs are (e) *distance-MEP*, (f) *jump-MEP*, and (g) *stop-MEP*. Figure 5 shows schematic depictions of typical MEP movement patterns. The *distance-MEP* is a small movement to evaluate the distance between two previously explored entities of the map shown in Figure 5.e for the distance estimation between a track segment and another map entity. During the *jump-MEP*, a jump from a map entity to another map entity is executed, lifting the finger of the map. This MEP is frequently followed by the region-MEP if the map entity that was the goal is not met straightaway. The *stop-MEP* indicates periods without movements.

With an MEP categorization, map explorations can be semantically classified and hence, a propositional representation enables the reasoning for the generation of verbal assistance. An example of a sequence of MEPs is: *track-MEP-1*(t_1), *track-MEP-2*(t_3), *track-MEP-3*(t_2), and so on (compare Figure 4).

The MEP Observer Component. The recognition of MEPs is performed in the MEP Observer component. The stream of input data of the haptic device is segmented into semantic MEPs in the MEP Observer component. Hence, the MEP Observer is a core component for the reasoning for language generation. In this component, spatially organized gestures are categorized in a semantic way. The stream of data is on a first level segmented into perceptual units with a fixed temporal and spatial resolution. On higher levels, these perceptual units are combined to aggregates of conceptual representations in a hierarchical process resulting in assumptions about the MEPs executed by the user. As it is the aim of VAVETaM to provide helpful verbal assistance accompanying the haptic exploration of the VE tactile map, it is important that the MEP Observer recognizes MEPs as early as possible, even if the MEPs are not yet finished. For example, if the user is tracing a track, the system might recognize this *track-MEP* and provide verbal assistance while the user is still exploring the track. To accomplish this, an approach similar to the sketch recognition process outlined in Adler & Davis (2009) is used: Basic level units are recognized and grouped to more abstract aggregates, the *micro-level movements (MLMs)*. These MLMs form the next level categories called *MEP Parts (MEPPs)*. The MEPPs are aggregated to MEPs on a semantic level. Compare Figure 6 for an illustration of the hierarchical combination process. To be able to proceed as described, the Haptic Movement Observer (HMO) uses both, procedures similar to those used for gesture recognition, and pre-segmentation depending on the position of the IP in relation to the different objects in the VE tactile map. The HMO outputs MLMs, which represent basic user movements in relation to objects of the VE tactile map. When the IP touches the surface of an object representing a track or a landmark, an MLM is recognized, namely the *touch-track-MLM*. This MLM can be followed by a *trace-track-MLM*. Once an MLM is recognized subsequent perceptual units may still be associated with the same MLM, i.e., a *trace-track-MLM* might be recognized while the user is still tracing the track. A further tracing of the same track will not result in recognition of another *trace-track-MLM*. The MEP Categorization (MEPC) aggregates MLMs to the higher-level categories in the hierarchy, namely MEPPs and MEPs. For example, an MEPP describing a single track segment being explored would be constituted by the MLMs of touching, tracing, and finally leaving the track in question. MEPPs are further combined in a hierarchical manner to form MEPs (kindred to Guhe, Habel and Tappe 2000). For example, the MEPP denoting an exploration of the track segment t_1 and the subsequent MEPP for exploring a track segment t_2, with track segment t_2 being connected to track t_1, may constitute the *track-MEP*(t_1,t_2) as shown in Figure 6.a. Like MLMs both MEPPs and MEPs need not be complete in order to be recognized. In Figure 6.a the depicted *track-MEPP*(t_2) and the *track-MEP*(t_1, t_2) may be recognized although the final *cease-touch-track-MLM(t2)*, which is depicted in grey, might still be missing and more perceptual units might be associated with the *trace-track-MLM*(t_2). Figure 6.b shows a visualization of the described process. As a track segment as linear map object is referred to by a line-like object with a relatively small width, the movement pattern of the *trace-track-MLM* can be described quite well in relation to the map objects in the VE tactile map, which already includes the information that the *trace-track-MLM* follows the movement pattern of the corresponding track segment. Other MEPs that are related to larger map objects or more than one object have to rely on gesture-recognition-like processes to enable movement categorization and by this, MEP recognition. The *region-MEP* serves as a good example for this process: It generally begins with a *touch-region-MLM*. This MLM is followed by a *linear-region-expl-MLM*, then a *direction-change-MLM* is performed, followed by another *linear-region-expl-MLM*, and so on, until the final *cease-touch-region-MLM* is recognized. For an identification of these MLMs, the observed movements of the IP have to be transformed into a qualitative movement representation to enable the recognition of linear movements, curved movements with different radii and direction changes.

The recognition of MEPs enables the generation of verbal assistance. A problem to be solved is the possibility of changing qualitative relations of the IP to the map object. During the exploration of t_2 in Figure 4, t_4 might be in front concerning the direction of the exploratory movement, but when the user proceeds to t_6, it is in the back. This means, some of the qualitative directions for utterances have to be calculated online at the time of the utterance. The MEP Observer provides the actual IP position to the GVA component to enable this online spatial reasoning.

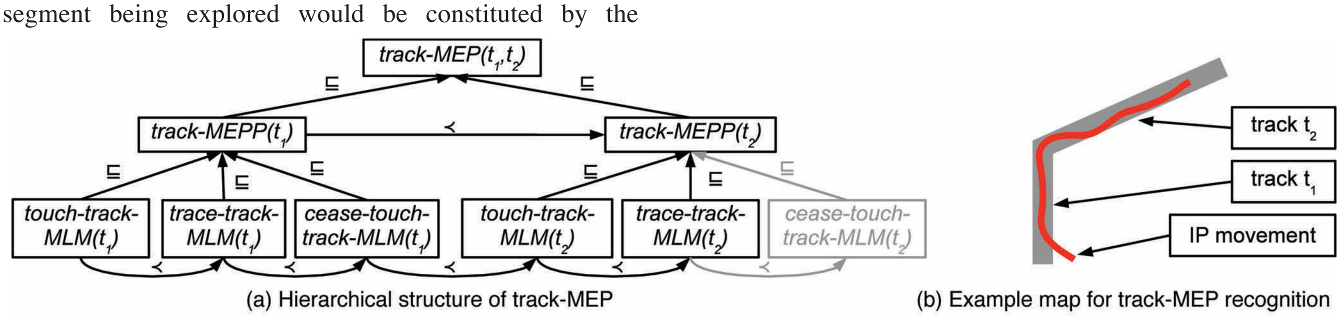


Figure 6: (a) Hierarchical structure of a *track-MEP* (≺ indicates part of relation, ⊑ indicates subsequent MLMs or MEPPs)
(b) Visualization of a *track-MEP*(t_1,t_2) recognition process

4 Conceptualizing Qualitative Spatial Relations for Verbalization

In order to generate helpful verbal assistance for VETM exploration the GVA has to relate the user's MEPs to both the propositional and the spatial-geometric layer of the VETM in a semantic way, a necessity caused by the propositional nature of verbal language and the reasoning processes about the user's informational needs outlined below. To give a short example in advance: The user is exploring a track and the VAVETaM should announce that an easy to hear fountain is ahead. The fact that a certain region of the spatial-geometric layer of the VETM is called a 'fountain' is stored in the propositional layer, while the reasoning that the MEP is heading for this area is done in the spatial-geometric layer.

Lovett and Forbus (2009) suggest *visual routines* for the interpretation of visual media representing spatial knowledge in a set of propositions describing spatial relations. In the context of VAVETaM a set of propositions describing spatial relations between different map entities or the IP and map entities needs to be generated as part of the conceptualization process for verbal assistance generation. The difficulty of this task results from the large and inhomogeneous set of possible qualitative spatial relations used in both natural language and computational spatial reasoning. Describing all possible relations between all entities of the map would produce very large sets of propositions and thus lead to computational problems and more important to problems of cognitive adequacy. Thus, the set of propositions that is true for even the small visual map depicted in Figure 7 is quite large.

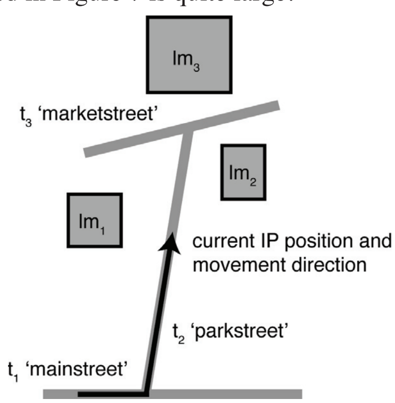


Figure 7: Example of map exploration in progress. The bold arrow symbolizes the IP movement. The currently detected MEP is *track-MEP(t_2)*

The challenge of conceptualizing spatial relations in the context of VAVETaM is focused on the selection of relevant spatial relations by taking into account the user's exploratory movements. By basing this selection on both currently observed MEPs and the memory component, we exploit the users actions for this selection, gaining the following advantages: (a) the extracted information is relevant to the current state of exploration, this gives the user the possibility to explore in a more active way, (b) the assumptions about the user's knowledge are taken into into account, both, to avoid redundant information and to relate new information to map entities already known by the user, (c) references to the user's spatial actions (e.g., moving the IP) can be used within the verbal assistance instead of just giving static descriptions, and (d) the amount of propositions to be communicated is limited, which fosters easy communication and encoding by the user.

Different MEPs are associated with different sets of *Informational Needs* that range from labeling information to distal spatial relations of yet unknown map entities and hints for further exploration. For a *track-MEP* this set of informational needs might include answers to the questions: How is the explored track labeled? Where is the exploratory movement headed in general? What is the start and end point of the explored track? What salient landmarks are near to the current position or passed along the upcoming way? Assuming no prior exploration by the user, a useful verbal assistance in the scenario depicted in Figure 7 might be: *'You move along Parkstreet in the direction of lm3. Currently lm1 is on your left side and you will pass lm2 on the right side before you meet Marketstreet in a T-crossing.'*

informational-needs hypothesis 1	
MEP:	track-MEP(t_2)
label(t_2):	'parkstreet'
direction:	landmark(lm_3)
landmarks right	-
landmarks left	lm_1
landmarks ahead right	lm_2
landmarks ahead left	-
track startpoint	t-crossing(t_2,t_1)
track endpoint	t-crossing(t_2,t_3)

Figure 8: Informational-needs hypothesis for the situation depicted in Figure 7

Once an MEP is recognized by the MEP Observer, the GVA conceptualizes an *Informational-Needs Hypothesis* that consists of a set of propositions that answer the specific questions depending on the type of MEP, see Figure 8 for an example of an informational-needs hypothesis in a frame like pseudo code describing the map depicted in Figure 7. These answers are derived by applying visual routines to the spatial-geometric layer of the VETM in combination with the current position and movement data of the IP. I.e., for a landmark to count as being left or right of the IP it has to intersect a line through the current IP position orthogonal to the explored track and be within a specified distance to the track, as shown in Figure 9. As Lovett and Forbus point out, simple visual routines can be combined to form more complex visual routines. For example, (1) visual scanning in the direction of the currently explored track can detect the endpoint of this track and can determine characterizations of it, such as 'being a T-junction with t_2', (2) visual-geometric routines considering lines orthogonal to the track allow the construction of regions that contain the landmarks left and right ahead on the currently explored track, respectively. These two visual routines are depicted in Figure 9.

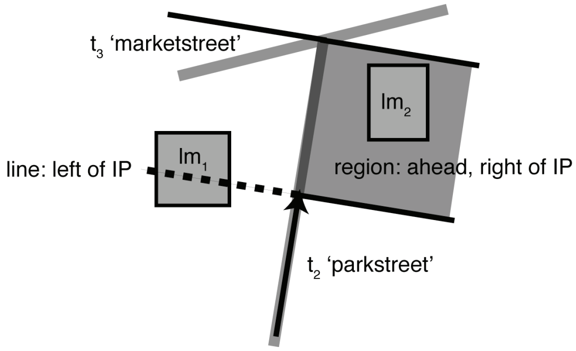


Figure 9: Visual routines for 'landmarks left' and 'landmarks ahead, right': All landmarks intersecting the dotted line are considered left of the current IP position. All landmarks intersecting the grey area are considered ahead, right of the current IP position

Not only the observation of a single MEP, but also the observation of two consecutive MEPs can be used to generate verbal assistance. In Figure 7 the current MEP is *track-MEP*(t_2), which was preceded by *track-MEP*(t_1). This setup could result in a verbal assistance like: *'You left mainstreet and entered parkstreet,'* making sure the users are aware they no longer follow the same street.

All propositions stored within the informational-needs hypothesis are possible candidates for verbalization to assist the user. Due to the fact that the user is probably not interested in redundant information, the Map-Knowledge Reasoning (MKR) is consulted. Any proposition the user is already assumed to know is deleted from the set of propositions in the informational-needs hypothesis.

Producing verbal utterances takes time, so the amount of verbal assistance for an ongoing exploration is limited by the time the user is willing to wait until performing the next MEP. Therefore, propositions have to be chosen from the informational-needs hypothesis according to the criteria: (a) importance, (b) domain constraints, (c) temporal constraints, and (d) conversational maxims. I.e., (a) an audio-enabled traffic light might be more important to be mentioned than a streetlight, (b) if several landmarks are passed, they should be mentioned in the right order, (c) a landmark that is currently direct left of the IP should be mentioned preferred to one that is still ahead, because the user might proceed along the track, and (d) mentioning *'three trees'* is more economic than mentioning *'a tree'* three times.

Only one verbal assistance is generated a time and is then both stored in the Verbalization Memory and sent to the Formulator and Articulator component. Once the verbal assistance is uttered, the informational-needs hypothesis is computed again, taking into account the updated position and movement direction of the IP. If the current informational-needs hypothesis differs from the former hypothesis, the selection process starts over again. This makes sure the generated messages are in accordance with the current IP position. Figure 10 shows an example of the emergent behavior of this system: Instead of generating a static description like: *'There is lm2 on the left side of the track,'* which does not take the user's exploratory movements into account, it might generate an assistance like *'You will pass lm2 on the right side,'* which offers a more specific localization to the user.

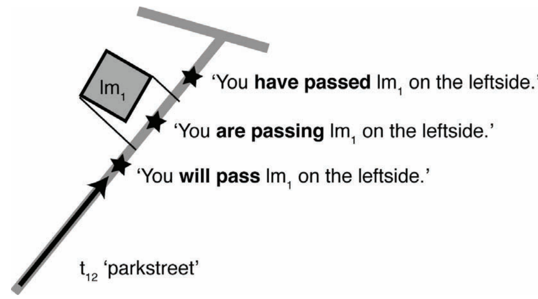


Figure 10: Verbal assistance depending on current IP position, symbolized by the star

5 Conclusion

This paper illustrates how the interaction of haptic perception and verbal communication can be realized in the VAVETaM system to generate verbal assistance. This assistance includes information about labels, such as names of map entities, exploration hints, and information about spatial relations. The accessibility of verbal assistance eases tactile-map exploration, resulting in a higher speed and accuracy.

In comparison to existing verbal assistance systems, VAVETaM offers customized assistance that depends on both recognition of users' motions and position in the VE tactile map and assumptions about their proposed knowledge. This kind of assistance fosters an active exploration. Hybrid representations combining visual representations and propositional representations constitute the knowledge base used for analyzing the user's exploration processes. The same representation is the basis for the conceptualization of suitable verbal instructions and descriptions. To allow the reasoning for the generation of verbal assistance, the user's movements are categorized into map exploration procedures (MEPs). Visual reasoning is used to facilitate understanding of the user's actions in a semantic way. Thus, the provided assistance can specifically address the user's current informational needs, which usually surpass simple labeling information in complexity. Visual reasoning offers a way to generate this assistance without the need for manual pre-encoding and enables the generation of verbal assistance in relation to the IP's actual position, which is a common case for assistance. Visual routines are, in the context of the VAVETaM project, not only used for reasoning in the visual modality. Instead, they provide a way of generating multimodal representations of spatial domains.

Acknowledgment

The research reported in this paper has been partially supported by DFG (German Science Foundation) in ITRG 1247 'Cross-modal Interaction in Natural and Artificial Cognitive Systems' (CINACS).
We thank the anonymous reviewers for their highly useful commentaries.

References

Aaron, A., and Davis, R. 2009. Symmetric Multimodal Interaction in a Dynamic Dialogue. In *Intelligent User Interfaces Workshop on Sketch Recognition*.

De Felice, F.; Renna, F.; Attolico, G.; and Distante, A. 2007. A Haptic/Acoustic Application to Allow Blind the Access to Spatial Information. In *World Haptics Conference*, 310-315.

Espinosa, M.A.; Ungar, S.; Ochaita, E.; Blades, M.; and Spencer, C. 1998. Comparing Methods for Introducing Blind and Visually Impaired People to Unfamiliar Urban Environments. *Journal of Environmental Psychology* 18(3): 277-287.

Guhe, M.; Habel, C.; and Tappe, H. 2000. Incremental Event Conceptualization and Natural Language Generation in Monitoring Environments. In *Proceedings of the first international conference on Natural Language Generation,* 85-92.

Habel, C. 2003a. Representational Commitment in Maps. In Duckham, M. Goodchild, M., & Worboys, M., eds. *Foundations of Geographic Information Science*, 69-93. London: Taylor & Francis.

Habel, C. 2003b. Incremental Generation of Multimodal Route Instructions. In *Natural Language Generation in Spoken and Written Dialogue. Papers from the 2003 AAAI Spring Symposium,* 44–51.

Habel, Ch., and Tappe, H. 1999. Processes of Segmentation and Linearization in Describing Events. In Klabunde, R., and v. Stutterheim, Ch. eds., *Representations and Processes in Language Production*, 117–152. Wiesbaden: DUV.

Haugeland, J. 1987. An Overview of the Frame Problem. In Pylyshyn, Z. ed. *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*. 77–93. Norwood, NJ: Able.

Jacobson, R.D. 1998. Navigating Maps with Little or no Sight: An Audio-tactile Approach. In *Proceedings of the Workshop on Content Visualization and Intermedia Representations,* 95-102.

Kostopoulos, K.; Moustakas, K.; Tzovaras, D.; Nikolakis, G.; Thillou, C.; and Gosselin, B. 2007. Haptic Access to Conventional 2D Maps for the Visually Impaired. *Journal on Multimodal User Interfaces* 1: 13-19.

Lahav, O., and Mioduser, D. 2008. Haptic-feedback Support for Cognitive Mapping of Unknown Spaces by People who are Blind. *International Journal of Human-Computer Studies* 66(1): 23-35.

Lederman, S.J., and Klatzky, R.L. 1987. Hand Movements: A Window into Haptic Object Recognition. *Cognitive Psychology* 19(3): 342–368.

Levelt, W.J.M. 1989. *Speaking: From Intention to Articulation*. The MIT Press, Cambridge, MA.

Loomis, J.M.; Klatzky, R.L.; and Lederman, S.J. 1991. Similarity of Tactual and Visual Picture Recognition with Limited Field of View. *Perception* 20(2): 167-177.

Lovett, A., and Forbus, K. 2009. Using a Visual Routine to Model the Computational of Positional Relations. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society*.

Maaß, W. 1994. From Vision to Multimodal Communication: Incremental Route Descriptions. *Artificial Intelligence Review* 8(2): 159-174.

Mark, D.M.; Smith, B.; and Tversky, B. 1999. Ontology and Geographic Objects: An Empirical Study of Cognitive Categorizaton. In Freksa, C., and Mark, D.M. eds. *Spatial Information Theory,* 283–298. Berlin: Springer.

Myers, K., and Konolige, K. 1995. Reasoning with Analogical Representations. In Glasgow, J.; Narayanan, H.; and Chandrasekaran, B. eds. *Diagrammatic Reasoning: Cognitive and Computational Perspectives*, 273–301. Cambridge, MA: MIT-Press.

Palmer, S. 1978. Fundamental Aspects of Cognitive Representation. In Rosch, E., and Lloyd, B. eds. *Cognition and Categorization,* 259-303. Lawrence Erlbaum: Hillsdale, NJ.

Parente, P., and Bishop, G. 2003. BATS: the Blind Audio Tactile Mapping System. In *Proceedings of the ACM Southeast Regional Conference*.

Pirolli, P., and Card, S. 1999. Information Foraging. *Psychological Review* 106: 643–675.

Timpf, S., and Kuhn, W. 2003. Granularity Transformations in Wayfinding. In Freksa, C.; Brauer, W.; Habel, C.; and Wender, K.F. eds. *Spatial Cognition III*. 77-88, Berlin: Springer.

Ungar, S.; Blades, M.; and Spencer, C. 1993. The Role of Tactile Maps in Mobility Training. *British Journal of Visual Impairment* 11(2): 59-61.

Wang, Z.; Li, B.; Hedgpeth, T.; and Haven, T. 2009. Instant Tactile-audio Map: Enabling Access to Digital Maps for People with Visual Impairment. In *Proceeding of the eleventh international ACM SIGACCESS conference on computers and accessibility*, 43–50.

Zinchenko, V. P., and Lomov, B. F. 1960. The Functions of Hand and Eye Movements in the Process of Perception. *Problems of Psychology* 1: 12-25.