Towards Generic Spatial Object Model and Route Guidance Grammar for Speech-based Systems

Perttu Prusi, Anssi Kainulainen, Jaakko Hakulinen, Markku Turunen, Esa-Pekka Salonen and Leena Helin

Speech-based and Pervasive Interaction group, TAUCHI, Department of Computer Sciences
University of Tampere, Tampere, Finland
{Perttu.Prusi, Anssi.Kainulainen, Jaakko.Hakulinen, Markku.Turunen, Esa-Pekka.Salonen, Leena.Helin}@cs.uta.fi

Abstract

In-vehicle route guidance is an area where commercial applications have already come to market. However, indoor guidance is still under research. We present a grammar that can be used to generate speech-based route guidance descriptions based on the spatial object model also discussed in this paper. In speech-based route guidance, proper references to objects in the environment and appropriate segmentation of the guidance are important features.

1. Introduction

Previous studies about indoor route guidance have focused on the development of resource-adaptive systems capable of multimodal interaction, concerning especially user positioning and graphical presentations of route descriptions [1, 2, 3]. Route guidance systems for indoor and outdoor navigation are based on models describing objects, their attributes and the relations between the objects of the environment.

This paper introduces a route guidance grammar (RGG) for generating speech-based descriptions of different indoor environments. We also present a spatial object model (SOM) that describes the relevant attributes of any indoor object to be described using speech.

We started research concerning indoor route guidance back in the year 1999 by developing the Doorman, a speech based system including an indoor receptionist and a route guidance system [4]. Speech is an effective way to convey complex route guidance information [5, 6, 7], though our experience [4] in speech-based route guidance has shown that users have trouble in remembering long descriptions. The focus of our study is on developing an effective speech-based and multimodal user interface for route guidance. The solution presented is based on the route guidance implementation in the Doorman system and our experiences with it. While most of the features were already present in the Doorman implementation, the solution is more elaborated and systematic.

We construct route guidance out of segments. Dividing a route to a combination of segments brings two kinds of benefits [8]. Generic route descriptions that are suitable for different usage contexts and hardware setups are easier to produce from segments than a complete route. On the other hand, segmentation limits the load that route descriptions place for user's cognitive resources at one moment. The segmentation supports the human way to analyze information when creating an inner presentation, a *cognitive map* [9] used in navigating. Segmentation has been introduced in some form in the route guidance grammars in [5, 6, 10, 11]. RGG describes a segment that has dynamically changing elements depending on the part of the route that it is describing. The description for the whole route in indoor environment is formed, by generating these segments and chaining them together.

SOM presents a way to model the environment. Objects, their attributes and relations between the objects are described so that the model supports automated positioning, route finding and speech-based guidance. SOM is based on our experiences [4]. A collection SOM descriptions form a spatial model of the real indoor environment to which the guidance is generated. SOM is not restricted to describe objects of certain facilities or use context. It serves the purposes of different route finding and speech-based route guidance generation platforms. Furthermore, it is not bound into any specific interaction possibilities between the user and route guidance system. I/O-resources required for interaction can be embedded in the environment or they can be implemented in the user's personal mobile device like a PDA or a smart phone.

Next we introduce RGG, in detail with some examples describing its potential. After this we present SOM. We end the paper with discussion and conclusions.

2. RGG: Route Guidance Grammar

Studies concerning outdoor route guidance generation [5, 6, 7] present fragments of the grammars used in them. Although [7] presents a full grammar for generating speech-based tourist route descriptions, it needs to be adjusted for indoor environment and supplemented with descriptions concerning for example landmarks.

Elements and the structure of RGG will be discussed in detail, because they play a significant role in conveying complex route descriptions using speech. RGG is designed to be capable of generating flexible and compact speech-based route descriptions easily adoptable by humans because of their natural spoken language form.

The route descriptions consist of segments each guiding the user from one *decision point* [2, 3, 8, 13] to next. In the decision points, the direction of the progression is reconsidered. When segments from the start to the end of the route are chained together using aggregation and referring expressions, a fluent speech output is formed.

2.1. Indoor route description structure and vocabulary

RGG is presented in detail in Table 1. The left column presents the elements of the grammar. The right column describes the structure and vocabulary of each element. Next we go through the elements and their structure comparing them to previous studies.

2.2. Objects

In the direction phrase template for outdoor navigation presented in [11], proper names and traffic signs were used to describe objects. Also spoken route description generating system presented in [10] uses proper names for describing the objects in the environment. RGG describes *Object* using *Name* and *ObjectType* as shown in Table 1. Names can be proper names like: "Perttu's office" or more general names like "vending machine" or a "coat rack". ObjectType is used to classify objects by their shape and other attributes. We use ObjectType to clarify the structure of the spatial model needed for abstract level route finding. For example, when finding a suitable route for human to traverse, objects belonging to ObjectTypes like "corridors", "stairs" and "communication unit" [4] (ObjectTypes with speech-based interaction capabilities) are relevant to the search. During the surface generation objects belonging to ObjectTypes like "landmark" are important. If the object does not have a proper name, ObjectType is needed for describing it in descriptions.

Element	***Description***
SegmentDescription	[**Start**] **Instruction** [**End**]
Instruction	**Action** | **Confirmation**
Start	**Orientation**
End	**Orientation**
Action	[turn **Direction**] [and] [**Confirmation**] [**Distance**] [to the] [**Rank**] **Object**
Confirmation	{continue [through the **Area**]} | {continue past [the **Object**]} | {head towards the **Object**} | {follow **Area** | **Sign**}
DirectionCorrection	front | back
Direction	{in | to [your] [**DirectionCorrection**] right | left} | in front of [**Object**] | behind [the **Object**] | up | down
Distance	until next | {for {some time | a long time} | {soon | until} you arrive to}
Sign	follow the signs to **Area** | the **Object**
Rank	First | second | third | …
Object	**Description | Landmark | Area**
Landmark	[shape] [size] [color] | …
Orientation	{start near **Object Direction**} | {you see the **Object Direction**} | the **Object** you are looking for is **Direction** [near the **Object]**
Area	corridor | lobby | hall | ...
Description	Name | ObjectType

Table 1: RGG: a grammar for generating route description segments

In addition to single objects, *Area*s are elements that can be utilized especially in instructions that define the path of the segment. Areas in the pedestrian route description context are discussed in [12]. They point out that areas play a significant role in pedestrian route descriptions, since people can move more freely in the environment unlike vehicle drivers, whose moving is limited more predictably in roads. Indoor areas like corridors, halls and lobbies are significant objects for route descriptions because of their size. They are easily noticeable despite the effects of the observer movement. An area can be utilized in confirmation, e.g.: "…follow corridor…"

2.3. Action

Action in route descriptions presented in [12] is divided in orientation, starting, moving and destination. Action presented in [7] contains confirmation, distance, direction and orientation, e.g.: "…turn right and continue through the hall to the stairs..." or "…continue past the meeting room until next door…" We suggest that the structure of the action should be more flexible for generating different kind of route description segments for different environments.

The beginning of the action presented in Table 1 consists of expressions defining either the *Direction* optionally followed by *Confirmation* or with just confirmation advising the user to continue to the same direction as advised in previous segment. We define directions in Table 1 in quite a similar relative way as in [12]. Simple left and right instructions [6] are not exact enough to point out the wanted object indoors where similar appearing objects like offices next to each other might be easily mixed. Especially for describing the destination *DirectionCorrection* might be required.

The purpose of confirmations is to encourage the user to go through or alongside an area, head towards an object or continue past a landmark or other object e.g.: "…follow the corridor…", "…continue past the coffee machine…" or "…head towards the lobby…"

Directions of the progression on the path might be described in less precise way using objects like discussed above. These objects are used in the description as fixed points towards which the user is supposed to head [7]. These confirmation elements ensure the subject that he/she is on the right track as discussed also in [1].

The direction and/or confirmation elements are followed by elements defining the progression in the path. These elements are used to anchor the end of the path to a certain endpoint landmark, decision point or other object, e.g.: "…head towards the door..."

Descriptions of the decision points are crucial in route descriptions, because turning in a wrong direction may cause the user to get lost. User's trust towards the system increases, if he/she is able to verify the descriptions by constantly observing mentioned objects in the environment while traversing the route.

Studies have shown that humans prefer non-metrical distance expressions [1, 2]. This is probably due to the fact that we are poor at estimating exact distances [7]. *Distance* is described using relative expressions or references to other objects like structural objects of the environment that have a sequential character, e.g.: "…until you arrive to the third door…"

2.4. Orientation

Orientation in route description confirms user's position by describing objects in the environment. Orientation and reorientation elements in the route descriptions are discussed in [3, 12].

In the beginning of a route, orientation is used to form a common ground between the route guide and the receiver on the current position e.g.: "…start near couch to your left…" In the end of the route it is pointed out that the destination is reached e.g.:"…the laboratory you are looking for is behind the pillar…"

Orientation along the route can be given to the user by describing the start and/or end of the segment. This reorientation consists of definitions of the heading and starting point or ending point of the segment. This is done by using expressions defining heading relatively and describing objects in the path ahead e.g.: "…you see a big plant to your left…"

Orientation in the start and end of a route segment can be redundant and therefore meaningless burden to the cognitive resources of the receiver as pointed out in [5, 11]. Orientation is presented only when necessary. The need depends on the *continuity* and *forward progression* between two following segments [5]. According to continuity, if start point orientation is omitted, it is assumed to be the same as the previous end point. According to forward progression, the direction of the motion is assumed to be the same and therefore no orientation between the segments is required. Optional orientation descriptions bring flexibility and help to focus the speech-based route descriptions to convey only the important information.

3. SOM: Spatial Object Model

The basis of any route guidance system is in the model of the space where guidance is supported. The model has to cover the objects, their attributes and relations between them. SOM does not cover all the aspects of indoor objects, only the attributes relevant from speech-based route guidance generation perspective.

Areas form a middle level between the spatial model and SOM. Spatial model has to include also structurally unimportant and less static individual objects like bookshelves and vending machines that can be utilized in speech-based route descriptions as landmarks like discussed in [1, 3, 7, 8, 12].

Tag	***Subtag***	***Type***
ID		String
Name		String
	UsageType	Integer
Location		Coordinates
Center		Coordinates
ObjectType		Integer
Connection		Integer set
Pass		Binary
	Constraints	Time expression set
Sector		Coordinates
ShapeUniqueness		Integer
Color		Integer
	ColorUniqueness	Integer
Transparency		Integer
Rank		String
Sign		String set

Table 2: Object attributes in spatial object model (SOM)

The SOM presents an object as a set of attributes. The attributes can be found in Table 2. They form three groups; some attributes are common to all objects while others are specific to *decisions points* [3, 8] and some are used for landmarks only. An object can be a landmark and a decision point at the same time.

3.1. Attributes concerning all objects

A unique string of characters identifies each object and is used when queries to the spatial model are made.

To be able to generate a structural presentation of the environment, the *location* and *shape* need to be modeled from every object. A common way to present these is to use 2D or 3D coordinates to define the physical boundaries of objects [1, 7, 10]. We have added a *Center* attribute which gives a possibility to compare automatically the relative position between the objects. In speech-based descriptions this attribute enables expressions like: "…in front of the coat rack…" and "…behind the locker…" that define distance to objects relatively as emphasized in [1].

Objects can be references in various ways in speech-based descriptions. Each object must have a *Name* [1, 4, 7, 8, 10, 12] and *ObjectType* as discussed above. There can be several names for one object. For example, users speaking different languages and special groups, such as visually impaired users, require different names for the same object. We use *UsageType* paired with name that enables the dynamical selection of the name for each usage context.

3.2. Decision points and interaction possibilities

A series of decision points form the skeleton of the route connected by the paths. A decision point has *Connection(s)* to other objects. Connection between two objects means that one can be used to get to the other, e.g., a corridor has connection to an office room. When a suitable route is found, a description concerning this route is generated. Connections may also have restrictions for their usage. *Pass* attribute defines restrictions, if there are any, and its sub-attribute, *Constraints*, specify the restrictions. Constraints can be sets of temporal limitations concerning weekday, date and time. For example a door may be locked during the night time and the constraints may contain the information concerning the week days and times of day when the door is passable, hence the door is passable but with constraints.

A speech-based ubiquitous route guidance system like the Doorman [4] requires user interface components like microphones and loudspeakers embedded in the environment. These components need to be modeled in SOM. We use *Sector* attribute for components used in speech-based interaction to define the range and direction. Sector is a set of coordinates that define the area of audibility. Sector plays a central role in timing of speech-based interaction and can be utilized by the system when dividing the route descriptions into segments.

3.3. Landmark attributes and other features

Landmarks are memorable cues selected along the path. They are needed to communicate reorientations and/or path progression at decision points [3]. Separation from the background by some attribute or combinations of attributes defines an object's eligibility as a landmark. Objects, like stairs and elevators [2] and supporting structures of the building can act as reliable landmarks in route descriptions. Landmarks are discussed widely in route guidance literature [1, 2, 3, 7, 8, 12], but they are often mentioned as lists of objects suitable for landmarks.

In addition to certain objects and object groups known to have landmark value, we present attributes for *ShapeUniqueness*, *Color* and *ColorUniqueness*. They define the landmark eligibility of any object and can be utilized in speech-based route descriptions. The idea of using the landmark attributes instead of certain objects in route guidance is described in detail in [13]. We take a light weight approach in modeling the attributes. For example for confirmation in Table 1 expression like"…head towards the round window…" could be generated using SOM, even though windows as such are not necessary eligible as landmarks. Especially ShapeUniqueness and ColorUniqueness define the eligibility of a landmark when modeled numerically. High values of these attributes raise the possibility of an object to be selected in route description as a landmark. Objects behind transparent or translucent materials might be visible to the user and reliable landmarks though not reachable. *Transparency* attribute is utilized by the route description generation process when choosing suitable objects. *Rank* defines the attribute that can be used for descriptions based on serial numbering of objects. When using rank, a clear route description can be generated even for challenging environments without eligible landmarks, e.g., for describing an action element: "… continue through corridor to the third door…"

Explicit *Signs* and signboards are the conventional way to offer route guidance and used also in speech-based route guidance generation [7]. If properly implemented, they offer valuable additional information to the route descriptions. Sign attribute contains a set of objects that can be guided from current object using explicit signs.

4. Conclusions

We have introduced a SOM and RGG that can be used to generate speech-based route guidance for indoor environments. SOM describes the environment and RGG uses the SOM as a vocabulary for adaptable speech-based route guidance generation. Adaptable and extensive RGG allow it to be used in different indoor environments. SOM covers the needs of indoor route guidance system providing several ways to refer to objects using speech. In addition to route guidance, SOM can be used for other speech-based systems that need to make descriptions of the environment.

The next step in our study is to replace existing route description component of the Doorman [4] using RGG and modify its spatial model to respond to the demands of SOM. After these modifications to the system are done, we will be able to make setups in different environments to test their adaptability.

5. References

[1] Look, G., Kottahachchi, B., Laddaga, R. & Shrobe, H. "A Location Representation for Generating Descriptive Walking Directions", *Proc. of 10th International Conference on Intelligent User Interfaces (IUI '05)*, 122-129, 2005.

[2] Baus, J., Krüger, A. & Wahlster, W. "A resource-adaptive mobile navigation system", *Proc. of International Conference on Intelligent User Interfaces 2002 (IUI2002)*, 15-22, 2002.

[3] Butz, A., Baus, J., Krüger, A. & Lohse, M. "A hybrid indoor navigation system", *Proc. of the 6th international conference on Intelligent user interfaces 2001 (IUI2001)*, 25-32, 2001.

[4] Mäkelä, K., Salonen, E-P., Turunen, M., Hakulinen, J. & Raisamo, R. "Conducting a Wizard of Oz Experiment on a Ubiquitous Computing System Doorman", *Proc. of International Workshop on Information Presentation and Natural Multimodal Dialogue*, 115 – 119, 2001.

[5] Tversky, B. & Lee, P. "How Space Structures Language", *Spatial Cognition, An Interdisciplinary Approach to Representing and Processing Spatial Knowledge*, Springer-Verlag, 157-176, 1998.

[6] Streeter, L., Vitello, D. & Wonsiewicz, S. "How to tell people where to go: Comparing navigational aids", *International Journal of Man-Machine Studies, Vol. 22, 5*, 549-562, 1985.

[7] Höök, K. "An Approach to a Route Guidance Interface" *Stockholm University, Dept. of Computer and Systems Sciences, Licentiate Thesis* **ISSN 1101 8526**, 1991.

[8] Dale, R., Geldof, S. & Prost, J-P. "CORAL: using natural language generation for navigational assistance", *Proc. of twenty-sixth Australasian computer science conference on Conference in research and practice in information technology, Vol. 16*, 35 – 44, 2003.

[9] Kuipers, B. "The 'map in the head' metaphor", *Environment and Behaviour, Vol. 14, 2*, 202-220, 1982.

[10] Krahmer, E., Landsbergen, J. & Odijk, J. *A Guided Tour Through LGM. How to Generate Spoken Route Descriptions. IPO Report 1182*, 1998.

[11] Tversky, B. & Lee, P. "Pictorial and Verbal Tools for Conveying Routes", *Proc. of International Conference on Spatial Information Theory: Cognitive and Computational Foundations of Geographic Information Science*, 51-64, 1999.

[12] Corona, B. & Winter, S. *Guidance of Car Drivers and Pedestrians*, Technical Report, Institute for Geoinformation, Technical University, Vienna, Austria. 2001.

[13] Raubal, M. & Winter, S. "Enriching wayfinding instructions with local landmarks" in M. Egenhofer & D. Mark, (eds.), *Geographic Information Science, Lecture Notes in Computer Science 2478*, Springer, 243-259, 2002.