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


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.

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.

<table>
<thead>
<tr>
<th>Segment/Description</th>
<th>Instruction</th>
<th>Start</th>
<th>End</th>
<th>Action</th>
<th>Confirmation</th>
<th>DirectionCorrection</th>
<th>Direction</th>
<th>Distance</th>
<th>Sign</th>
<th>Rank</th>
<th>Object Description</th>
<th>Landmark</th>
<th>Area</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Start] Instruction</td>
<td>Action</td>
<td>Orientation</td>
<td>Orientation</td>
<td>[turn Direction]</td>
<td>[continue through the Area]</td>
<td>[continue past the Object]</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>Name</td>
<td>ObjectType</td>
<td></td>
</tr>
</tbody>
</table>

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

In addition to single objects, Areas 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 end-point landmark, decision point or other object, e.g.: “…heheads 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 reorienta-
tion consists of definitions of the heading and starting point or
ending point of the segment. This is done by using expres-
sions 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 orienta-
tion 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 orienta-
tion between the segments is required. Optional orientation
descriptions bring flexibility and help to focus the speech-
based route descriptions to convey only the important infor-
mation.

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 at-
tributes 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 unimpor-
tant 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].

<table>
<thead>
<tr>
<th>Tag</th>
<th>Subtag</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>String</td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>String</td>
<td></td>
</tr>
<tr>
<td>UsageType</td>
<td>Integer</td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td>Coordinates</td>
<td></td>
</tr>
<tr>
<td>Center</td>
<td>Coordinates</td>
<td></td>
</tr>
<tr>
<td>ObjectType</td>
<td>Integer</td>
<td></td>
</tr>
<tr>
<td>Connection</td>
<td>Integer set</td>
<td></td>
</tr>
<tr>
<td>Pass</td>
<td>Binary</td>
<td></td>
</tr>
<tr>
<td>Constraints</td>
<td>Time expression set</td>
<td></td>
</tr>
<tr>
<td>Sector</td>
<td>Coordinates</td>
<td></td>
</tr>
<tr>
<td>ShapeUniqueness</td>
<td>Integer</td>
<td></td>
</tr>
<tr>
<td>Color</td>
<td>Integer</td>
<td></td>
</tr>
<tr>
<td>ColorUniqueness</td>
<td>Integer</td>
<td></td>
</tr>
<tr>
<td>Transparency</td>
<td>Integer</td>
<td></td>
</tr>
<tr>
<td>Rank</td>
<td>String</td>
<td></td>
</tr>
<tr>
<td>Sign</td>
<td>String set</td>
<td></td>
</tr>
</tbody>
</table>

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

The SOM presents an object as a set of attributes. The at-
tributes 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 en-
vironment, 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 attrib-
ute 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 im-
paired 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 descrip-
tion concerning this route is generated. Connections may also
have restrictions for their usage. Pass attribute defines restric-
tions, if there are any, and its sub-attribute, Constraints, spe-
cify the restrictions. Constraints can be sets of temporal limita-
tions 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 micro-
phones and loudspeakers embedded in the environment.
These components need to be modeled in SOM. We use Sec-
tor attribute for components used in speech-based interaction
to define the range and direction. Sector is a set of coordi-
nates 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 seg-
ments.

3.3. Landmark attributes and other features
Landmarks are memorable cues selected along the path. They
are needed to communicate reorientations and/or path pro-
gression at decision points [3]. Separation from the back-
ground 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 Sign 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 RGGo that can be used to generate speech-based route guidance for indoor environments. SOM describes the environment and RGGo uses the SOM as a vocabulary for adaptable speech-based route guidance generation. Adaptable and extensive RGGo 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 RGGo 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