# Acquiring a Shared Environment Representation

E.A. Topp*, H. Hüttenrauch†, H.I. Christensen* and K. Severinson-Eklundh†

* Centre for Autonomous Systems † Interaction and Presentation Laboratory
Royal Institute of Technology
10044 Stockholm, Sweden
{topp,hehu,hic,kse}@nada.kth.se

## ABSTRACT

Interacting with a domestic service robot implies the existence of a joint environment model for user and robot. To enable robot navigation within such a setting requires further to embed a user's mental environment model in the corresponding robotic model. Robots typically use metric models for navigation, therefore those metric models need to be integrated with models provided by users. This paper presents a pilot study that investigates, how humans present a familiar environment to a mobile robot. Results from this pilot study are used to evaluate a proposed generic environment model for a service robot.

## Keywords

User study; Environment representation; Human Augmented Mapping; Personalisation

## 1. INTRODUCTION

Service robots are – often mobile – platforms that provide assistance to humans. Thus, a basic competence for a mobile robotic system is the ability to move from one location to another which requires navigation and localisation functionalities. Also, the robot has to share the environment with its potential users, which means that it has to move around and reason about its whereabouts in a way that is comprehensible. Mobile robots can navigate on the basis of metric, often feature based, maps, and they can build those maps autonomously while exploring an environment for the first time. Methods in robotics research are dealing with this issue of Simultaneous Localization and Mapping (SLAM) [3; 4; 6, among others]. Humans have a topological, (partially) hierarchical, view on their environment [12, among others]. To enable a service robot to perform tasks for users in arbitrary environments (well known to the user, initially unknown to the robot), a spatial representation that is understandable for both, the robot and the user, is needed. Assuming an indoor environment such as a home or office building, we mean by a commonly understandable map representation, that the robot's notion of the environment appears to be the same as the one the user might refer to. In other words, we need to build a "shared mental model". Such a model is likely to depend on a very personal view a potential user has on the environment. Thus, a service robot, provided with some general world knowledge can be – and ought to be – personalised. We assume a scenario of a "guided tour" to be an appropriate way to "teach" the robot its environment. The user can guide the robot around and name important regions and specific locations. At the same time, the robot can build a (metric) map of the environment. This map is augmented by the user's information which allows to integrate the robot's metric, feature-based map with the topological map representation of the user. Figure 1 gives an idea on how a scene of such a guided tour could look like.

**Figure 1: Illustration of a user showing the kitchen to her robot**

An open issue is the question, what strategies to present an environment would be used by different users, and how the given information can be incorporated into an environment model, to actually satisfy the requirements for a shared representation. In earlier work we already introduced the concept of *Human Augmented Mapping* (HAM,[16]), which allows us to subsume different aspects of Human-Robot Interaction (HRI) and robotic mapping. In a previous study [7] aspects of interaction as well as posture and positioning of subjects in relation to a robot were studied. In this case the scenario was also a "guiding the robot around" scenario, but the environment was limited to one room and the robot used in the study was controlled remotely.

With the present paper we describe a user study, in which

*Human Robot Interaction* '06 Salt Lake City, Utah, USA

subjects guide around an autonomous mobile robot in a complete floor of an office building that they are familiar with. The study investigates how different users present a well known environment to a robot. We suggest a generic robotic environment model and demonstrate with results from an initial pilot study, how this model can be personalised to different individual representations of a given environment.

## 1.1 Outline of the paper

The rest of this paper is organised as follows. We give an overview of related work and refer to hierarchical environment representations motivated from results in Cognitive Science and Psychology to propose a general robotic environment representation in section 2. Section 4 explains the design of our study, in section 5 we present the results from initial experiments, and in section 6 we draw conclusions on this pilot study and its results.

# 2. REPRESENTATION OF SPACE

In "The intelligent use of space" [9] Kirsh stated, that in order to understand complex (human) models of an environment, we have to observe the interaction of the (human) agent with this environment. Based on those observations, corresponding *robotic* models can be obtained. Transferring this to the interaction of two agents in and about a certain environment, observations from human-human interaction could be the base for a general robotic environment model. Such a model can be used to incorporate information from the interaction with a user, to personalise the robotic system. Personalisation along the taxonomy of Blom [2] means in this context to *accommodate work goals* (to "customise" the robotic system for certain tasks) and to *accommodate individual differences* (of different users in the explicitly stated representation). We propose to observe a human and a robot interacting in an environment (instead of two interacting humans), to learn, what robotic model can be used to build a "shared mental model" that both the user and the robot can refer to later.

In a study that uses a miniature robot on a table top street "map", Kyriakou *et al.* investigate, how computer vision can be used to follow verbal guiding directions [11], by having subjects guide the robot with commands like "follow this road to the station, then turn left". This is another form of "guiding a robot" without actually being part of a collaboratively working duo in the same working space. One condition for such a setup is the availability of a map that contains all items a potential user considers important at the respective location. Since this is what we wanted to learn about (what do users present in a given environment and how) we do not consider such a setup an option for our study.

Kuipers *et al.* presented a mapping approach that represents the environment as a combination of global topological and local metric maps [10]. The main aspect of this work however is the handling of large scale maps, that can be achieved by representing the environment as local metric maps that are linked in a global, topological (and as such hierarchical) representation. Also in other approaches the segmentation of metric maps and/or organisation of them into hierarchies has been studied as part of SLAM, but primarily as a way to limit computational complexity [3; 14; 15, among others].

Approaches to interactive robotic mapping have been reported by Diosi *et al.* [4] as well as by Althaus and Christensen [1]. Diosi *et al.* obtain a purely metric spatial representation of an office environment by guiding a robot around and defining labelled regions. Althaus and Christensen model the environment rather as a topological graph, but do not consider different levels of granularity in their representation. We believe that not only rooms (or regions) are needed, but also a lower level of complexity has to be integrated in a topological model. This allows to integrate places into the specified regions.

## 2.1 Motivation for an environment model

A number of different theories on how spatial relations are acquired and represented have been proposed throughout the years. According to McNamara [12] those theories can be grouped different the dimensions of a) format (analog vs. propositional), b) functionality (spatial configuration vs. semantic or logical knowledge), c) structure (flat vs. strongly hierarchical), and d) contents (encoded information vs. procedural knowledge to compute information).

McNamara used this categorisation to design a psychological study on spatial representations that concentrated only on the two latter characteristics (structure and contents). Subjects were given recall and distance estimation tasks on items that were spread out in physically separated regions on a "map". The results indicated, that distance between two items matters as well as co-coexistence in one region. In other words, if two items were close to each other, but in different regions, it was still possible for the subjects to recall and estimate their spatial relation. If the distance was large, this recall and estimation worked better within the same region. Thus McNamara came to the conclusion, that a *partially hierarchical model* supported his findings most appropriately.

Following these findings, we assume, that users would not necessarily follow a hierarchical order when explaining the environment to the robot, e.g. explain a certain place first and then give information about the room or present certain places only, that are located in different rooms. Transferring this to our guided tour implies, that the assumed robotic environment model has to be able to handle spatial information given in arbitrary order. Thus we propose a hierarchical structure, that incorporates the required flexibility with generic entries on each level, in which places can be represented. We express this assumption as well in a number of working hypotheses for the pilot study in section 4.3.

To incorporate other dimensions, particularly the *functionality*, the hierarchy needs to be extended. Galindo *et al.* [5] propose *Multi-Hierarchical Representations* to incorporate semantic information into their environment model used for mobile robotics. Two hierarchies, one conceptual, the other spatial, are linked with anchoring to enable reasoning. Their spatial hierarchy is build from local map representations obtained from sensory data, that are interpreted as open spaces (rooms, corridors) connected in a topological structure. The conceptual hierarchy incorporates concepts such as workspace, room, object and instances of those categories. A semantic model is given a priori, that links objects conceptually to rooms. For example it is assumed that an object "bathtub" is to be found in a room called "bathroom". By observing objects, the conceptual hierarchy is used to assign a specific concept ("bathroom") to a local map representation in the spatial hierarchy. As stated above we assume a hierarchical representation of the environment, but do not incorporate the semantics so far.

Along with the functionality one issue is in fact the personalisation [2] of a particular environment representation. From intuition one would expect, that individuals have different preferences and ideas how to interpret and use their surroundings. We consider the fact that different users might give different information to the very same robot as an issue of future work. Our environment model though is flexible enough to model those individual differences within the same framework.

# 3. HUMAN AUGMENTED MAPPING

With our concept of *Human Augmented Mapping* we can establish the link between a robotic map that enables the robot to navigate and the environment representation of a user (also referred to as "cognitive map"). We use a graph based model of the environment, described in section 3.1 to incorporate the information that is given interactively. Our assumption is that a "guided tour" is an appropriate way to give the user the possibility to personalise the robot's general environment representation. An "off-line" personalisation that could be achieved by pointing out items, places and regions on a metric map representation that was *autonomously* created by the robot does not seem useful, since the user would have to remember *exact* spatial relationships between items. When the robot and its user share the same workspace in an interactive setting, it is presumably easier to determine for the user, if the robot "understood" a piece of information correctly. From the robotic point of view we see the advantage of an interactively controlled mapping process in the disambiguation possibilities that arise from the interaction. We further do not assume a full initial environment model, that allows the robotic system to instantiate content entities by autonomous exploration, but consider a more general, structural model that can be filled with personalised information and that can be revised if necessary.

## 3.1 A hierarchical graph structure

We model the environment by using a hierarchy of graphs. The main concepts we incorporate so far are *locations* (or places) and *regions*, as depicted in figure 2.

We define a *location* as
**Specific positions/areas that can represent the position of large objects that are considered static.**
Such locations can for example be a closet, a refrigerator or a sofa.
A *region* is then
**Any portion of space that is large enough to allow for different locations in it, or at least large enough to navigate in it.**
Typically this would be rooms or corridors or parts of those.

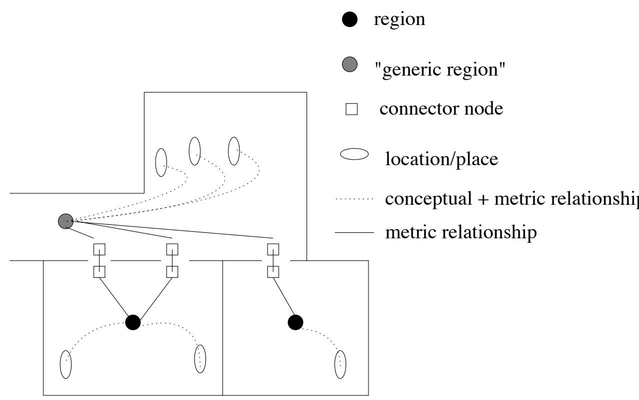


**Figure 2: Our graph structure visualised in 2D**

A natural extension to a higher level would be *floor* or *building*, but this was not considered for this work. On a lower level smaller objects that can (hypothetically) be manipulated and change their position frequently could be integrated, such as milk bottles in the fridge or brooms in a closet.

Regions are represented by local (metric) maps that can be used for navigation. The local maps are linked metrically by pairs of internal *connector nodes* that have an absolute position with respect to the local map they are in. Since those maps are built at the same time as the graph structure is filled with the information from the user, an initial internal hypothesis needs to be introduced. To maintain the hierarchical structure but allow for partially hierarchical representations as well, we assume a "generic region" in which we start the mapping process. With the "generic region" we can guarantee, that all mapped areas are represented as a region on the respective level of the hierarchy. As soon as a region is assigned a name by the user, it is stored together with the corresponding local metric map, that might already contain information on specific locations. When a specified region is left, and the adjacent area was not explored before, this "new" region becomes the "generic region". Note that the "generic region" can consist of several, topologically delimited regions (in the sense of an autonomous mapping approach). Only specified regions are entities in the hierarchy that form a new branch from the respective level downward. This makes it possible to define a specific location in a region, that is not (yet) relevant to name (e.g. a corridor).

# 4. THE PILOT STUDY

We conducted a pilot study to test our proposed robotic environment model against the information on a specific environment given explicitly by a human user to a mobile robot. Additionally the pilot study serves as a proof of concept for a more comprehensive user study currently prepared. The pilot study comprised experiments with five subjects of about 45 minutes duration. Within this time period the subjects spend about 20 minutes interacting with the robot. All of them received a cinema ticket voucher as compensation for their participation.

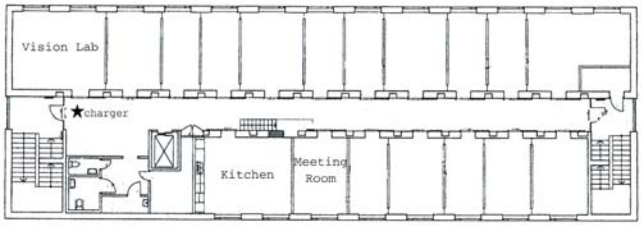


**Figure 3: The floor plan of our office environment on which the experiments took place. The star marks the starting point, where subjects encountered the robot**

## 4.1 Scenario

The scenario of the pilot study was a "guided tour" through a portion of an office building. Figure 3 shows the floor plan with offices (not marked), the kitchen, the meeting room and the computer vision laboratory of our office building. Subjects were instructed to show the robot around in the environment so that it later could perform service tasks and in order to do this needed to have "seen" the respective locations (a more detailed description of the instructions and the technical realisation is given in sections 4.2.2 and 4.2.3).

## 4.2 Method

In the following section we explain our selection of subjects, the instructions given to them, and the methods used for data collection.

### 4.2.1 Subjects

As important precondition to our pilot study we assumed subjects to know the environment they would guide the robot around in. This precondition is important and based on the idea that potential users will "add" service robots to their (to them already well known) homes and offices. Subjects were therefore recruited from the laboratory environment the experiments took place in. We make this a requirement also for our future study as it will ensure that subjects are not primed or biased by information that would have to be given to participants unfamiliar with the given environment and spatial settings. To require familiarity with the robot's operation area is a design choice that differs from other human-robot interaction studies, where subjects often are invited into an unfamiliar or "simulated" environment. The choice comes at a price however: in our own office environment some subjects of the pilot study were expected to be familiar with the internals of robotic systems. As a consequence we plan on (also) using a different environment for our future user study, to make sure that the familiarity with our robotic system is counterbalanced by subjects without experience in robotics research.

To assure at least some variety in familiarity with robotic systems we selected our five subjects actively among the members of the Computer Vision and Active Perception Laboratory[1] that hosts a part of the Centre for Autonomous Systems[2] on our campus. The group of pilot subjects included one secretary (familiar with robots from films, presentations and frequent encounters in the office environment, but not familiar with the internals), three computer vision researchers, one of them somewhat familiar with the internals of robotic systems, and one robotics researcher from the field of robotic mapping. All of them had been working in this particular office environment for about two years.

[1]http://www.nada.kth.se/cvap
[2]http://www.nada.kth.se/cas

### 4.2.2 Instructions

Our subjects were given an instruction sheet that explained the task and the functionalities and abilities of the robot. The task was to use a number of commands (`follow me, go to <target>, stop, turn left, turn right`) and explanations (`this is <item>`) to make the robot follow and to point out everything that the subject considered important for the robot to know on the floor the experiments took place on. The time frame given to the subjects for the completion of their task was about 20 minutes (15 minutes for the guided tour and five minutes to test the robots "memory"). In the instruction none of the words *region, location, position* or *place* was named. We referred to *"everything, that you think the robot needs to know", "whatever you pointed out before", etc.*, so that subjects were completely free to decide, what they would present to the system and how they would name it. Neither did we give any example (e.g. "You can name for example the coffee maker"), so that we would not include any items that a particular subject would not have considered important in the first place. Nevertheless subjects were informed, that we were not interested in small objects, since the robot had no object recognition abilities, it just would need to know "where" to go. The instruction sheet included a drawing that showed, how the field of view of the robot looked like, and that it would use a laser range finder to detect the subject for following and in order to "look around". This information was important, since the laser range finder only offers a forward field of view of 180°, with a range of 8 meters (for the detection of users we reduced it further to 3 meters), and the subjects could thus understand, how the robot perceived its environment. Particularly they were instructed to move a few steps in front of the robot so that it would detect and classify them as user. To make the robot start to move when following they should gain a distance to it of at least one meter, to give it the space to actually move.

The subjects were also informed that the robot was moving autonomously, when it was following a subject according to the task scenario, but that commands were interpreted by an experiment leader and fed manually into the system. Since we did not incorporate any object recognition, we stated that a service task (`go to <target>`) would be successfully completed, when the robot could find its way to the location where the task would have been performed. Also for the actual presentation of an item, the robot was assumed to "see", when it was "facing" the item. The instruction sheet was very open about the robot's abilities: we clearly stated which of the functionalities of the robot were in fact simulated or remotely controlled (see 4.2.3 for details) by an experiment leader that followed the (subject and robot) pair. We also explained, what the subjects should not try to do, as for example to send the robot around alone to explore the environment on its own, use the elevator or try to send it somewhere unknown. Subjects were offered to ask for help before and during the actual experiment, and knew that they could abort the experiment at any time.

#### 4.2.3 Technical realisation

The study was performed with a commercially available Performance PeopleBot by ActivMedia[3]. In a previous study this robot was used in a Wizard-of-Oz-setting [7], where the robot's functionalities were remotely controlled or simulated by two experiment leaders. For the technical realisation of our pilot study scenario we used a laser range data based tracking and following approach [16], which has been extended to incorporate a metric laser range data feature based SLAM method [6] and an input option to label regions or locations with name tags. Basic platform control and access to the sensors and text-to-speech system (Festival[4]) are provided by the Player/Stage[5] software library.

The system represents labelled locations in a simple graph structure that distinguishes between specifically labelled positions ("defined place") and internal navigation nodes. The internal nodes are used to build a navigation graph on which the system can perform a graph search to plan a path[6] to a previously named position. Note that this system does not implement the hierarchical model we proposed above, but enables a user to act and interact with the robot according to our scenario to test the validity of our proposed model.

The verbal interaction of the user with the robot was still controlled by the experiment leader, i.e. utterances from the subject were interpreted by the experiment leader and labels of locations or regions and commands were fed into the system via a graphical user interface (GUI) running on a laptop. This allowed us to avoid having problems due to miscommunication (as studied by Green *et al.* [8]) interfere with the actual task. For verbal feedback though we used the text-to-speech system with precoded utterances, so that the robot could refer to its own state and the task given to it (e.g. "I will follow you", "Stopped following", "I think I have lost you", "Stored <item>").

As the experiment took place on an entire floor of the building, one experiment leader (the robot's supervisor) had to follow the subjects to observe the experiment including all utterances and in general to assure the subject's safety at any time. The implemented system allowed switching from autonomous following based on the mentioned tracking approach to full remote control immediately by invoking a soft joystick implementation. Thus tracking failures and other inconvenient situations could be solved by the experiment leader as in a Wizard-of-Oz setting without having consequences for the mapping process and the labelling.

We provided the robot with two different behavioural strategies for the labelling of either a location or a region. If a location (including a "link" to a region, e.g. a doorway) was presented, the robot did not move and stated immediately, that it stored the given information. If on the other hand a region was presented, the robot stated, that it needed to have a look around and performed a 360° turn before confirming the information. The decision, which behaviour to choose, was made by the experiment leader according to our generic environment model and the respective definitions of regions and locations.

#### 4.2.4 Observation methods and data collection

By storing the data provided by the sensory systems we used for the technical realisation we could get a full "real time" (graphical) representation of each of the experiments. Additionally we recorded the experiments with two digital video cameras each. One video was recorded from the point of view of the robot, by mounting the video camera on its upper platform. The other camera recorded an external point of view by accompanying the user and the robot. After their experiments our pilot subjects were asked to answer a number of questions on the experiment in a short interview. This interview was roughly scripted with a list of questions on the motivation of the subject for naming or not naming certain locations or regions and for the way to handle the tour scenario. We were particularly interested in whether the subjects had perceived the behaviour of the robot differing depending on what was pointed out (a location of a region) and what they thought about it.

### 4.3 Hypotheses

We wanted to study, how different individuals present a known environment to a mobile robot and relate the resulting information to an environment model we consider appropriate in the context of Human Augmented Mapping. We assumed that humans do not necessarily follow a hierarchical structure, when they present a known environment to a robot (see section 2). Thus, we started out with a number of working hypotheses about the way subjects would present the regions and locations they considered relevant, as well as about the entities that would be named: "users do not name all regions in the environment" (`WH1`), "users point out locations in regions they did not name before" (`WH2`), and "users point out regions without entering them" (`WH3`). We use these hypotheses to show, in how far the outcome of the pilot study can be related to our environment model. We did not formulate a specific hypothesis for the dependency "familiarity with robotic systems vs. way of explaining the environment to a robot" to explore this issue. Nevertheless we expected robotic researchers familiar particularly with map representations to be more explicit than subjects not familiar with robotic environment representations. Further we assumed, along the argumentation of Sidner *et al.* [13], that the difference in the robot's behaviour would allow the subjects to "understand" the robot's internal processes, when storing either a region or a location.

## 5. RESULTS FROM THE EXPERIMENTS

In this section we present the results from our pilot study. We are aware that the data set is small and consequently not entirely representative. However, it is possible to analyse the outcome of the experiments in terms of *occurrence* of different phenomena. Additionally, our observations and the subjective answers we obtained in the short interviews allow us to investigate how subjects reasoned about their strategy to show regions and locations and to improve the

---

[3]http://www.activmedia.com

[4]http://www.cstr.ed.ac.uk/projects/festival/

[5]http://playerstage.sourceforge.net

[6]implementation part of the CURE library (©2005 Patric Jensfelt and John Folkesson, Centre for Autonomous Systems, Royal Institute of Technology, Stockholm, Sweden)

system for further studies. In general we can state, that the pilot study verified the validity of our approach to get information on different ways to handle an interactive process to build a map representation. Furthermore we believe that the soundness of our environment model can be demonstrated by its ability to handle the different situations we observed. In table 1 we summarise the quantifiable results to give an overview over our observations and statements from the interview.

## 5.1 Observations

All subjects but one used the full time frame to present the environment to the robot. The "tour" started for each experiment at one end of the corridor (see Figure 3), where the robot awaited its user. An initial location (the "charger") was generated automatically directly after the system was initialised to enable the robot to go back to this starting point. As a consequence we do not count this location as relevant to our results. All subjects took the robot into the kitchen, probably because this is a central room in our office environment, both from a topological, a functional, and a social point of view. However, the observed diversity in strategies to introduce the kitchen to the robot was quite large, ranging from the pure introduction of *the kitchen* over some combination of *specific locations in the kitchen and the kitchen itself* to *specific locations only*. Already from our small sample of data we can thus conclude that the variety of explicitly stated information that a robotic system in an interactive mapping process would have to cope with is large and needs to be handled by the robot's environment representation. More specifically, these differences in naming observed for the kitchen and its locations correspond to our expectations expressed in hypotheses `WH1` and `WH2`.

We also noted that none of the subjects named the corridor or hallway itself as a region, but all of them pointed out specific locations in it, which gives us further evidence for our hypotheses. One frequently presented location in the corridor was the "elevator" (or "lift") (named by four of the five subjects), which was however only shown by positioning the robot in front of it and pointing to the *doors*. Also rooms were indicated only by pointing to the respective door, confirming our expectation expressed in hypothesis `WH3`.

Most of the subjects stated in the interview that they had pointed out those locations or rooms, that they personally considered important, and left out others on purpose. In other cases the time constraints kept the subjects from presenting more to the robot. We see this as a sign for a strategy to personalise the robot's environment representation to personal needs and preferences.

We asked all subjects that had presented a mixture of rooms and locations (four out of five), if they had perceived the difference in reaction of the robot (turning by 360° for a region vs. not turning for a location). Three out of those four answered, that they had observed the difference in behaviour. All three stated that this behaviour seemed *appropriate* and/or made the robot *look smart*, since it obviously wanted "to understand its surroundings". One subject did not notice the difference in behaviour, possibly because only two rooms were presented, and the subject stated to have been busy figuring out, "why the robot sometimes needed a long time to understand me, and sometimes not". Note that this was stated despite the fact, that written information had been given to all subjects, that all dialogue would be simulated by the experiment leader.

**Table 1: Quantifiable results from the pilot study**

| Observation Subject | VR | VR | VR | SE | RR |
|---|---|---|---|---|---|
| Interaction time | 22 min | 19 min | 11 min | 25 min | 24 min |
| # regions | 4 | 2 | – | 2 | 2 |
| # locations$^{I}$ | 4 | 4 | 5 | $4^{II}$ | $8^{III}$ |
| # regions w o loc. | 3 | 2 | – | 1 | 1 |
| # loc. w o region | 3 | 4 | 5 | 2 | $3^{IV}$ |
| # regions w o entering | 1 | 2 | 1 | 1 | – |
| Behaviour noticed | Yes | Yes | – | No | Yes |
| – appropriate | Yes | Yes | – | – | Yes |
| – appears smart | Yes | No | – | – | Yes |

VR: Vision researcher, SE: Secretary,
RR: Robotics researcher

I: including regions that were only pointed to
II: including one small object (salt)
III: including one person and two doorways to respective rooms
IV: excluding doorways

Despite some technical problems (see section 5.4 for details) and the above mentioned timing problem all subjects expressed their satisfaction with the flow of interaction and communication as well as the robot's performance.

## 5.2 Particular situations

Even with the limited number of subjects we were able to observe some interesting strategies for the presentation of the environment. We relate the observations to statements from the short interviews where possible. We attempted to order them with increasing relevance to the environment model.

**Pointing out persons** In two cases subjects also tried to point out a person. In one case the person was sitting at her desk and the robot was made to store the respective location by the experiment leader. In the other case the subject reacted spontaneously to someone walking out of the elevator right in front of the robot. Here the robot was not made to store the information from the introduction. Nevertheless these situations show, that the system would have had to handle introductions of persons as well, since the introductory phrase "this is $< name >$" was exactly the same as for the kitchen[7].

**Possessive pronouns and relations** One of the subjects presented "my office" to the robot. In such a case a dialogue system would actually have to analyse "my" and relate it to the subject's name, but this was beyond the scope of our pilot study. In the experiment, the robot was thus constantly referring to "my office".

**Extreme personal point of view** In one experiment ses-

[7] In this particular experiment the subject left out all articles when presenting items

sion we observed that two rooms were pointed out, but no locations in them. In the corridor, none of the service points (pigeon holes, printer, etc.) was named, but the two exits to either side of the building and the elevator, as well as the experiment leader's office (only the door was pointed to). When asked why no other locations in e.g. the kitchen were named, the subject stated that the exits were considerably important, as well as the kitchen as a room, in case that guests had to be served. The coffee machine and the refrigerator were not important since the subject does not drink coffee at all nor uses the refrigerator. The observation holds both evidence for our hypothesis H2 and an extreme personal point of view on the environment.

**Explaining no rooms at all** One of our subjects concentrated only on locations (e.g. pigeon holes, coffee machine, refrigerator) and did not name any room (or other region). On the question, why not for example the kitchen as a whole was named, the answer was, that the robot should rather know about the whereabouts of the places, where it should *do* something. Just sending it to *the kitchen* would by no means help to get a coffee, the subject stated. We see this as a strong evidence for hypothesis H2.

**Explaining doorways** We expected our subject with robotic research and mapping experience to be more precise and explicit than other subjects. This expectation could be confirmed by the fact, that the doors to showed rooms (two in this case) were pointed out explicitly *when the robot was standing exactly in the door opening*. We could also observe, that both named rooms were actually entered. Since only two rooms were presented during this experiment, we can of course not generalise, but we consider at least our expectations for the robotics researcher's strategy confirmed.

## 5.3 Relation to our environment model

Our observations show, that even with a small, rather homogeneous group of subjects different ways to show and explain the environment are to be expected and dealt with, depending on the individual view and use of particular items and rooms. We see these differences as a proof of concept for our proposed environment model (as introduced in section 3.1) which we consider usable for a robotic map representation.

A general assumption is that a given robotic system has the ability to perceive regions that are delimited from other regions autonomously. This could for example be achieved by door detection or a method like the watershed algorithm [4, as an example of use]. We also assume, that we have a general knowledge model that distinguishes between regions and locations and a dialogue model that uses this knowledge base. From the experiments we got some evidence already on the strategy of the users to point out a region by only showing the respective door. In all observed cases, subjects positioned the robot with the help of "turn commands" so that it was facing the particular "link" (doorway or elevator doors), before naming the region. If these subjects on the other hand presented the region they were currently in they just stated that this was "the $<$ *name* $>$" without positioning the robot with "turn commands". The detection of such differences in the user's behaviour could give a signal on the actual intention of the user. We hope to find further evidence for such a differing behaviour in our future study.

Departing from our observations we can postulate some key situations, that need to be handled by our robotic environment model and suggest possible solutions.

**Presenting persons** Given an appropriate dialogue model, it would be possible to ask, if actually the region/room the person is in should be named accordingly (e.g. "Elin's office", in case "Elin" was introduced to the robot).

**Locations in an unnamed region** If a location is named before the region it is in, or the region is not named at all, this location would end up in the branch of the "generic region" in our hierarchy. If later the information about the region is given, the region needs to be delimited and separated from the generic region. All locations within the observed delimiters (e.g. walls, doorways) are now associated to this new branch in the hierarchy.

**Links to regions/rooms** With the "connector nodes" of our representation links to rooms (pointed out doorways) can be handled. In the current region (which might be the "generic region") a connector node with a virtual directed edge to the named region is created. Thus the system knows, that it can find the way *to* a certain region, without knowing anything about its appearance.

**Pointing out doorways explicitly** The environment model could cope with explicitly pointed out doorways by generating a location with the respective name. There are several possibilities to represent it in the hierarchy though. One option is to decide which region it belongs to, based on the name of the respective region (e.g. as observed "this is the door to the kitchen"). The second option is to keep the location in both regions, with a relative position to the respective local map that relates to the same absolute position (if possible). A third option would be to generate an entry of the generic region, that would allow to state that the robot is "in between two regions". However, since we could observe the respective strategy only with the robotics researcher, we assume this to be rarely observable with a differently structured sample.

Summarising we believe that our model holds at least for the variety of strategies to present a known environment to a robot observed in our pilot study.

## 5.4 Technical issues

During the pilot experiments we observed several issues of the technical realisation that had consequences for the actual interaction between subjects and the robot.

Despite the instruction to give the robot space when it was about to follow, subjects waited standing still for the robot to move. The robot's verbal indication to follow ("I will follow you") was obviously not enough, to indicate that it would actually follow. From carefully studying the recorded interaction on video we concluded that the robot actually needs to indicate with a body (movement) gesture like turning toward the user that it has seen the user and is ready to follow. A similar problem occurred, when subjects made the robot face something to "look at it" and wanted to continue the tour afterwards. We plan to make the robot turn back toward the user to indicate, that it is ready to continue after storing a presented item.

# 6. CONCLUSION AND FUTURE WORK

In this paper we presented two important aspects of our concept of Human Augmented Mapping, namely the environment representation of the robot and the interactive context that allows to build a shared mental model of an environment. We explained our approach to a robotic map representation, and showed, to what extend this representation holds in different situations within an interactive mapping process. A pilot study was conducted to investigate strategies of users to present a for them well known environment to a robot.

Despite the small number of subjects we were able to observe a rather large variety of strategies to present a known environment to the robot in a "guided tour". Parts of this diversity might be due to differing knowledge in robotics or the individual interest in the robot that our subjects had. However, we can state that all the different situations or strategies characterised in our hypotheses occurred at least once.

We got mostly positive feedback on the behaviour of the robot, especially on the "region observation strategy" we implemented to enable subjects to understand the internal processes of the robot to some degree. This assures us to keep such a behavioural strategy for further studies, to allow subjects to understand more of the internal procedures of the robot. The variety in presentation strategies we observed and the self reflecting comments on them showed us, that there is a need for quite flexible representations, when one robotic system should be used and guided around by different users. The results from the pilot study encourage us to use the proposed setup in a more comprehensive user study.

# 7. ACKNOWLEDGMENTS

The work described in this paper was conducted within the EU Integrated Project COGNIRON ('The Cognitive Companion' - www.cogniron.org) and was funded by the European Commission Division FP6-IST Future and Emerging Technologies under Contract FP6-002020.

We also thank Patric Jensfelt and John Folkesson for a their technical help with the implementation of the system, especially the SLAM method used.

# 8. REFERENCES

[1] P. Althaus and H.I. Christensen. Automatic map acquisition for navigation in domestic environments. *in Proc. of the International Conference on Robotics and Automation,* September 2003, Taipei, Taiwan.

[2] J. Blom. Personalization: a taxonomy. *CHI'00 extended abstracts on Human factors in computing systems,* 313-314, ACM Press.

[3] M. Bosse, P. Newman, J. Leonard, S. Teller. SLAM in Large-scale Cyclic Environments using the *Atlas* Framework. *International Journal of Robotics Research,* 23(12):1113-1140, December 2004.

[4] A. Diosi, G. Taylor and L. Kleeman. Interactive SLAM using Laser and Advanced Sonar. *in Proc. of the IEEE International Conference on Robotics and Automation,* April 2005, Barcelona, Spain.

[5] C. Galindo, A. Saffiotti, S. Coradeschi, P. Buschka, J.A. Fernández-Madrigal, and J. González. Multi-Hierarchical Semantic Maps for Mobile Robotics. *in Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems,* August 2005, Edmonton, AB, Canada.

[6] J. Folkesson, P. Jensfelt, and H.I. Christensen. Vision SLAM in the Measurement Subspace *in Proc. of the IEEE International Conference on Robotics and Automation,* April 2005, Barcelona, Spain.

[7] A. Green, H. Hüttenrauch, K. Severinson-Eklundh. Applying the Wizard-of-Oz Framework to Cooperative Service Discovery and Configuration. *in Proc. of the 13th IEEE International Workshop on Robot and HumanInteractive Communication,* September 2004, Kurashiki, Okayama, Japan.

[8] A. Green, H. Hüttenrauch, K. Severinson-Eklundh, B. Wrede, S. Li. Integrating Miscommunication Analysis in Natural Language Interface Design for a Service Robot. *submitted*

[9] D. Kirsh. The intelligent use of space *Artificial Intelligence,* 73:31–68, 1995.

[10] B. Kuipers, J. Modayil, P. Beeson, M. MacMahon, and F. Savelli. Local Metrical and Global Topological Maps in the Hybrid Spatial Semantic Hierarchy. *in Proc. of the IEEE International Conference on Robotics and Automation,* April 2004, New Orleans, LA, USA.

[11] T. Kyriakou, G. Bugmann, and S. Lauria. Vision-based urban navigation procedures for verbally instructed robots. *Robotics and Autonomous Systems,* 51:69–80, January 2005.

[12] T.P. McNamara. Mental Representations of Spatial Relations. *Cognitive Psychology,* 18:87–121, 1986.

[13] C.L. Sidner, C. Lee, C.D. Kidd, N. Leash, C. Rich, Explorations in engagement for humans and robots. *Artificial Intelligence,* 166(1):140-164, 2005

[14] N. Tomatis, I. Nourbakhsh, and R. Siegwart. Hybrid Simultaneous Localization and Map Building: a Natural Integration of Topological and Metric. *Robotics and Autonomous Systems,* 44:3–14, 2003

[15] A. Tapus, G. Ramel, L. Dobler, and R. Siegwart. Topology Learning and Place Recognition using Bayesian Programming for Mobile Robot Navigation. *in Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems,* September 2004, Sendai, Japan.

[16] E.A. Topp and H.I. Christensen. Tracking for following and passing persons. *in Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems,* August 2005, Edmonton, AB, Canada.