

# Mixed-Initiative in Human Augmented Mapping

Julia Peltason, Frederic H.K. Siepman,  
Thorsten P. Spexard, Britta Wrede, and Marc Hanheide  
Applied Informatics Group  
Bielefeld University, Germany

{jpeltaso,fsiepman,tspexard,bwrede,mhanheid}@techfak.uni-bielefeld.de

Elin A. Topp  
School of Computer Science and Communication  
Royal Institute of Technology (KTH)  
Stockholm, Sweden

Elin-Anna.Topp@alumni.uni-karlsruhe.de

**Abstract**— In scenarios that require a close collaboration and knowledge transfer between inexperienced users and robots, the “learning by interacting” paradigm goes hand in hand with appropriate representations and learning methods. In this paper we discuss a mixed initiative strategy for robotic learning by interacting with a user in a joint map acquisition process. We propose the integration of an environment representation approach into our interactive learning framework. The environment representation and mapping system supports both user driven and data driven strategies for the acquisition of spatial information, so that a mixed initiative strategy for the learning process is realised. We evaluate our system with test runs according to the scenario of a guided tour, extending the area of operation from structured laboratory environment to less predictable domestic settings.

## I. INTRODUCTION

Traditionally, the ability of learning has been investigated in terms of information storage and representational formats. However, based on recent trends in psychology and different cognitive disciplines a new view on robotic learning is being established [1] termed “Learning by Interacting”. According to this view interaction is a necessary prerequisite not only for infants in order to learn but also for robots. For robotics this means that the interaction context has to be taken into account much more seriously than it has been up to now. In order to head towards life-long learning and adaptation abilities of today’s and future’s cognitive robots that interact with

humans, learning must not be separate, but part of the regular operation processes of a robot. In particular in scenarios that require a close collaboration and knowledge transfer between inexperienced users and robots, the “learning by interacting” paradigm goes hand in hand with appropriate representations and learning methods.

In the context of mobile domestic robots a central learning challenge is the acquisition of appropriate spatial models that bridge semantics of spaces and geometrical positions and metrics. In previous work the framework for *Human Augmented Mapping* (HAM) as a concept for the integration of robotic mapping and human environment representation was introduced [2]. Also the integration of a respective initial mapping subsystem into an interactive framework was discussed [3]. This paper presents the consequent extension of that work exploiting advanced interaction strategies leading to a novel spatial learning framework. The idea of “interactive mapping” or “semantic mapping” has in fact been discussed in a couple of works over the last few years. Earlier approaches to supervised learning of environment representations were reported by Althaus and Christensen [4] and Diosi *et al.* [5], in both cases however the user had to have the initiative and be very specific, particularly in the latter approach the complete environment had to be presented in one “tour”. An approach similar to our integrated system for interactive mapping is reported by Zender *et al.* [6]. Their system however focuses more on the conceptual/semantic level of the mapping process as far as the interaction and higher level functionalities are concerned. Our proposed mapping and learning subsystem is capable of handling both the user initiated specification of spatial entities and detecting significant changes in the spatial properties of the environment to generate a robot initiated clarification dialog as sketched in figure 1. Such a strategy for robot initiative was proposed also in other works [7], but requires more explicitly modeled knowledge, e.g., in terms of defined ontologies and a model of doors in order to detect transitions in pre-acquired maps. In contrast, we face up to the systemic challenge of closely interlinking a HAM-enabled learning approach with dialog management in order to realize a comprehensive social learning framework enabling mixed initiative and continuous learning in an interactive manner from scratch. The remainder of the paper is organised as follows. In section II we discuss the idea of *Learning by*

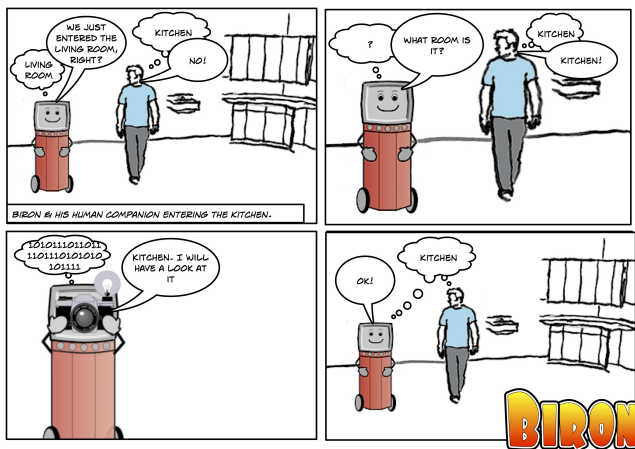


Fig. 1. The robot is taking the initiative and attempting to verify its (incorrect) hypothesis which induces the human to correct it.



Fig. 2. Hardware composition of BIRON: It is equipped with several sensors that allow perception of the current situation as a basis for interaction. Feedback is provided by speech synthesis and using a virtual character displayed on the touchscreen. The robot's on-board computational equipment in combination with a wireless LAN notebook is sufficient to achieve a system running responsive enough for fluent HRI.

*Interacting* and explain the requirements and scenario we assume for a respective system. We present the mapping and dialog (sub)systems that are central for our work in section III and IV. Our integration strategy for the complete system is presented in section V and we discuss the results we could achieve with our system in section VI. Conclusions and ideas for future works are presented in section VII.

## II. LEARNING BY INTERACTING

In contrast to general machine learning, learning by interaction is characterized by the close interrelation of learning and application phases and a mixed initiative kind of interaction. In the context of Human Augmented Mapping the central goal of learning is to successively acquire a spatial representation that is complete and consistent in order to enable the robot to provide services based on this knowledge. Strategies that are exploited to achieve completeness and correctness in interactive learning can be summarized as explicit tutoring, clarification turns, and correction turns. These strategies and the way they can be implemented are subject to discussion after the general learning scenario is presented.

### A. Scenario and Platform

The scenario, the mapping challenge is embedded in is the so-called *Home Tour*. The Home-Tour-Scenario for our robot envisions a newly purchased robot being introduced to its new working area – usually an apartment – by the human user. Due to the variety of applications in home environments only a smaller set of pre-programmed knowledge is useful. The main knowledge such as spatial layout or presence of objects in the new environment has to be obtained online during interaction with a human user. Thus, the Home-Tour-Scenario especially incorporates the requirement of a real-world environment with the additional constraint that the user has only minimal knowledge about the robot. With this minimal knowledge, e.g., taken from a single sheet of paper like a quick start guide for today's printers, the user must be able to interact with the robot. In our scenario this interaction consists of introducing the different rooms and objects of the apartment to the robot. This introduction is crucial for any further task like remembering the last position of the glasses

or fetching the user's favorite cup. The platform to study this scenario is the mobile robot *BIRON* (**B**ielefeld **R**obot **C**ompanion) as illustrated in Fig. 2.

### B. Facilitating Learning By Mixed Initiative Interaction

In the process of learning it is in particular important to ensure that the representation of the acquired knowledge is correct, consistent, and as complete as possible. As the scenario targets at inexperienced users that have no or only very limited knowledge about the learning system they will likely not provide the exact appropriate amount and type of information. Rather the model they intuitively teach by their own initiative is usually sparse and possibly erroneous. In a pilot study with a comparably small number of subjects already various strategies for the presentation of the same environment to a mobile robot could be observed [8]. Hence, it is a core principle of learning by interacting to continuously consider learning as an active, structuring process that allows to continuously monitor the already acquired knowledge to revise and extend it. There are several issues crucial for providing structure for interactive learning. In [9], we address the question how to *constrain the situation* by structuring the interaction in order to enable the robot to identify what needs to be learned. By contrast, in the work at hand we focus on the question of how to equip the robot with abilities that enable *active learning* and thus the improvement of the learned model within interaction. Having these abilities, the robot can take initiative and provoke situations that let it acquire new information which is more effective than simply waiting until the human eventually presents such situation to him. Therefore, we consider mixed initiative interaction style as key requirement for interactive learning.

As for the learning of a spatial representation, how can we exploit mixed initiative dialog in order to help human and robot to jointly build up a common understanding of their environment? One aspect of this kind of interaction is the human tutor taking the active role and showing the robot new rooms explicitly. Beyond providing the knowledge, the role of the tutor is also to check the learning success, for instance by asking monitoring questions like "Where are we right now?". As least as important, the other aspect is the robot taking the active role attempting to gain information, for instance by verifying the existing representation or by resolving uncertainty. Moreover, by actively communicating its hypotheses, the robot provides an insight into its model and thus provokes the user's feedback and gives her the opportunity to possibly correct the model.

In our scenario, providing such an interactive learning framework is the main task of the dialog system. We will describe it in section IV from the HRI point of view and in section V from the system integration point of view.

## III. THE MAPPING SUBSYSTEM

In the following we describe our mapping approach, that is originally a central part of a prototypical system for "Human Augmented Mapping" [2] and has been transferred to the more complex interactive framework on *BIRON*.

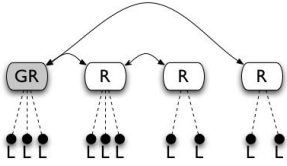


Fig. 3. The two-level model with “regions” (white in general (R), gray for the “generic region” GR) and “locations” (large black dots, L). The solid curved double arrows represent the links between the entries of one level (between “regions”), the dashed lines represent the relation of the “locations” to their surrounding “region”.

Our mapping subsystem uses a topological graph structure to implement a generic environment model. This model builds a hierarchy of spatial concepts, as is described in the following.

#### A. The graph structure

The topological graph of the Human Augmented Mapping framework as it is used for the work presented here implements two spatial concepts, termed *regions* and *locations*. A *region* is defined to be a delimited area of the environment, e.g., a room. A *region* can contain several distinct *locations*, corresponding to views or “snapshots” of the environment.

For the work presented here we focus on the segmentation of the environment into *region* nodes to build the topological graph structure in the assumed interactive process. In previous work we proposed to use a statistics based descriptor to segment *regions*, that can be used both for the representation and classification of specific *regions* as for the detection of transitions between them [10], [11]. This descriptor uses the ellipse generated by the two first Principal Components of a 360° laser range data set as one central feature. We assume that the system gathers information iteratively and not necessarily following the hierarchical model [8]. Thus, we use the concept of the “generic *region*”, which is topologically speaking a *region* without specification, i.e., a node in the graph that has no explicit spatial properties, but that subsumes (metric) positions in the environment, so that it can contain *locations* [2].

With this concept it is possible to start the mapping process “from scratch” without any *a priori* knowledge in the mapping subsystem about how many entities are to be presented or what their labels would be. Fig. 3 illustrates the hierarchy for the two conceptual levels.

The mapping subsystem assumes access to raw laser range finder data to generate *region* representations and to pose estimations. The latter are usually assumed to be corrected by a SLAM module to keep the *region* representations metrically consistent.

For the work presented here we focus on the segmentation and representation of *regions* that can be achieved in an interactive “guided tour”. We also use the term *room* instead of *region*, since all our examples refer to clearly distinguishable *rooms*. In the following we describe how we assume the environment to be segmented to construct the nodes of the underlying topological graph.

#### B. Segmenting regions / rooms

We consider two types of events that can trigger the system to segment a new room from the environment. One is to receive external input that annotates a certain spatial entity with a label (e.g., “... this is Elin’s office...”). The other type of event is the data driven detection of a “new area”, which can lead to a discourse with the user. The latter case can help the system to resolve ambiguities, thus, to reduce uncertainty about its whereabouts. Such a dialog with the user can be either a confirmation of the hypothesis the system generates or a correction of a wrong assumption. Both can then lead to the specification of a new room (and respective node), depending on the hypothesis generated by the system and the information given by the user.

In case a room is explicitly specified by the user, the mapping subsystem receives a respective request and triggers an “exploration turn” to gather a 360° laser range data set that is used for the computation of a concise representation of the surroundings. This “exploration turn” is used since the robot is only equipped with one laser range finder and proved actually useful for the interaction, as the robot is obviously “doing something” to obtain a representation of the surroundings. The computed representation is then used to generate a respective node in the graph structure and is additionally stored for comparison and classification purposes.

While traveling the mapping subsystem uses virtual scans computed from a local map to continuously generate hypothetical representations that are compared to the one assumed to represent the “current room”. In case a significant change is detected and / or the system’s pose estimation indicates that the delimiter of a known room has been passed, this is reflected in two internal flags, HYPOTHESIS\_UNCERTAIN (default value “false”) and HYPOTHESIS\_CHANGED (default value “false”).

Those flags together with the hypothesis for the “current room” can be extracted and used by the controlling programs to trigger a respective dialog with the user to resolve the ambiguity, as is described in section V. In case that the ambiguity is not resolved the mapping subsystem continues to work based on its current hypotheses. In this case the representation as such remains unchanged as do the flags, which means that a second request for clarification can be triggered.

An important decision was made regarding the initial assumption and initiative. We assume initially that the *user* has the initiative as long as the system is certain of being in the “generic *region*” and until at least one room is specified. Only after that, the mapping subsystem checks the internal environment representation continuously to detect transitions and maintain a consistent representation.

If the system receives at some point a confirmation of being in the “generic *region*” (i.e., having left a specified room without receiving more information) it would not notify any transition detection again until it hypothesizes to have come back into a known room. Thus, the actual initiative

arbitration can be handled by the higher level interaction supervision and dialog according to the interactive learning strategy.

#### IV. THE DIALOG SYSTEM

The multimodal dialog system presented in this section is based on the principle of grounding and was first introduced by Li et al. [12]. Here, we view it from the HRI point of view as a proxy for the overall robot system managing the interaction. Later on, we will discuss its communication with other components.

Both human’s and robot’s contributions are modeled as multimodal *Interaction Units* consisting of a *Verbal Unit* and *Non-verbal Unit* creating verbal and non-verbal output. For the robot’s dialog acts, we additionally model their underlying communication intention whereas we do not explicitly model the intention of the user. Her intention is implicitly computed by analyzing the semantic and pragmatic content of the utterance. However, this feature enables us to include an explicit user model about the user’s intentions in future work.

The Interaction Units are organized based on the principle of grounding [13]. Grounding is a well-known interaction concept assuming that both interaction partners aim to establish mutual understanding or *common ground* during their interaction. This results in utterances being arranged as pairs with the so-called *Presentation* initiating such a pair and the so-called *Acceptance* giving evidence of understanding. In case that the utterance has not been understood, a further Presentation will be created first in order to clarify the situation.

Altogether, figure 4 illustrates how mixed initiative interaction style is realized based on these concepts. In case of robot initiative, the intention of the robot’s Interaction Unit arises from an internal event. For instance, if the event occurs that the mapping subsystem recognizes a known room, the robot will have the intention to verify its hypothesis which will trigger the appropriate verbal and non-verbal behavior. Subsequently, it will wait for the human to create an appropriate Interaction Unit as Acceptance.

In case of human initiative, the intention of the robot’s Interaction Unit arises as reaction on the user utterance. For instance, if the user asks a question like “Where are we?”, the corresponding Interaction Unit will be created as Presentation which gives the robot the intention to answer the question and thus providing Acceptance.

Table I shows example subdialogs within the interactive location learning scenario. The robot’s behavior may vary depending on whether a representation already exists for a correction or on degree of uncertainty.

Note that if the human disagrees with the hypothesis communicated by the robot without providing the correct one, the robot will ask and keep waiting for a respective response until mutual understanding has been established. This is due to our strategy to have the robot ask for clarification; the mapping subsystem could in fact deal with a “pending

answer” and assume the “generic *region*” as the current one, as was mentioned already in section III-B.

In addition, it is of course an issue not to annoy the interaction partner with the robot’s self-initiated clarification behavior. Therefore, based on a person tracking and attention module [14], the dialog system triggers clarification questions only if the situation seems appropriate which depends on presence and behavior of the interaction partner. Accordingly, the robot does not ask in a situation where there is no interaction partner present or the interaction partner is engaged in conversation with another human.

TABLE I  
EXAMPLE SUBDIALOGS

Initiative	Subdialog purpose	Example dialog
Human	Room teaching	H: This is the kitchen! R: Kitchen. I will have a look at it. R: (looking) R: OK!
Human	Monitoring questions	H: Where are we? R: We are in the kitchen.
Robot	Verifying hypothesis	R: We just entered the hallway, right? a) <i>Human agrees</i> H: Yes. R: OK! b) <i>Human corrects</i> H: No, this is the living room. R: Living room. OK! c) <i>Human disagrees without providing correction</i> H: No! R: What room is it? H: The living room R: Living room. OK.
Robot	Resolving uncertainty	R: Are we still in the living room? a) <i>Human agrees</i> H: Yes. R: OK. b) <i>Human corrects</i> H: No, this is the kitchen. R: Kitchen. OK! c) <i>Human disagrees without providing correction</i> H: No! R: What room is it? H: The kitchen. R: Kitchen. OK.

#### V. SYSTEM INTERACTION STRATEGIES

Mixed initiative and grounding for learning are considered as guide lines for the composition and architecture of our system. Leaving the view of the dialog as proxy for the robot and considering it representing the user against the overall system presents the dialog system as one information source among others: Information concerning for instance the robot’s current location can be provided by the user via the dialog, by the presented localization component or any other source.

To provide the necessary flexibility for information coordination the architecture follows an information oriented approach employing an event-driven integration style. The system coordinates its actions triggered by the generation and modification of information entities (memory elements,

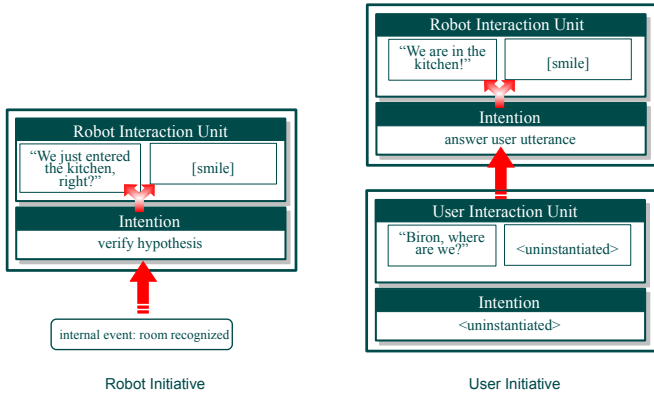


Fig. 4. Creation of Interaction Units. In case of robot initiative, the robot’s Interaction Unit is created based on an internal event. In case of human initiative, the robot’s Interaction Unit is created as respond on the humans’s Interaction Unit.

which are basically XML documents with binary attachments) in the so-called Active Memory [15], [14]. These information generation events are independent from system components as they only refer to a certain information within the Active Memory. This enables all system components to provide their information to the system and therefore they no longer depend on a certain system component but merely on a certain kind of information, independent of its source. In this paradigm, components are seen as information sources submitting to the central repository – the Active Memory.

#### A. Mixed initiative on system level

Mixed initiative in learning means that both user and robot may trigger the learning process. By means of the Active Memory concept, one basic requirement to implement mixed initiative is provided: A far-reaching equivalence of information sources. Components such as the mapping subsystem and the dialog system dynamically subscribe for particular events, e.g. the insertion of an information entity describing a room. In order to achieve mixed initiative, respective components always subscribe for the type of information they potentially generate themselves. Hence, they can react adequately, independent whether they generate the piece of information or any other component did. By these means a mixed-initiative-enabled component always takes both roles, as an information source and sink, simultaneously.

#### B. Grounding on system level

As described in the previous section, the equivalence of information sources provides a technical basis for mixed initiative. However, from the learning by interacting idea with several information sources existing arises the question how to make information consistent. Therefore, we elevate the concept of grounding which has up to now been mostly seen as the external communication protocol to a more systemic concept. Based on states describing the task progress, information is subject to a negotiation process related to the general idea of grounding. This negotiation process between components is triggered whenever an information

entity is inserted into the Active Memory. Its representation is augmented by a state encoding the negotiation progress, and by subscribing a component on information entities being in particular states, a sequencing of operation is achieved. Table II shows the different states with their respective semantics.

As the information is the result of such negotiation process, the information present in the Active Memory can to some extent be considered as common ground. From this perspective, state completed means that common ground has successfully been established. This view is supported by the fact that also the negation process is reflected by the Active Memory content: As shown in table III, the ROOM is set to generic region as soon as the robot realizes the incorrectness of its previous hypothesis. When the correct label has emerged from the negotiation, the ROOM is updated again.

The task completion states are more fine-grained than the grounding states which opens the possibility to provide more detailed feedback for the user. Since the robot’s feedback utterances serve mainly the purpose of information, it does not expect an answer from the user which results in “lonely” Presentations like Pre3 or Pre4 in table III.

In a real system, negotiation is not subject to only two components, but certain steps in this process might require involvement of others. For instance, a component might need to get access to a hardware component in order to assess current information which then involves arbitration. The Active Memory concept easily allows to account for this requirement by having the arbitration component subscribe for the particular state changes as will be illustrated in the following example.

TABLE II  
TASK PROGRESS STATES

State	Description
initiated	an information starting a task has just been inserted into the memory
accepted	it has been checked that no conflicts will prevent the task execution and all necessary preparation have been taken
rejected	the task is currently not executable and will not be started
completed	the task has successfully been executed
failed	though all preconditions were satisfied, an error occurred during task execution

TABLE III  
EXAMPLE DIALOG

Dialog act	Role wrt. grounding	State	Room
R: We just entered the living room, right?	Pre1	initiated	living room
H: No.	Acc1	completed	generic_room
R: What room is it?	Pre2	“	generic_room
H: This is the kitchen.	Acc2	initiated	kitchen
R: Kitchen.I will have a look at it.	Pre3	initiated	kitchen
R: OK.	Pre4	accepted	kitchen
		completed	kitchen

### C. Two Examples: Verification and room teaching

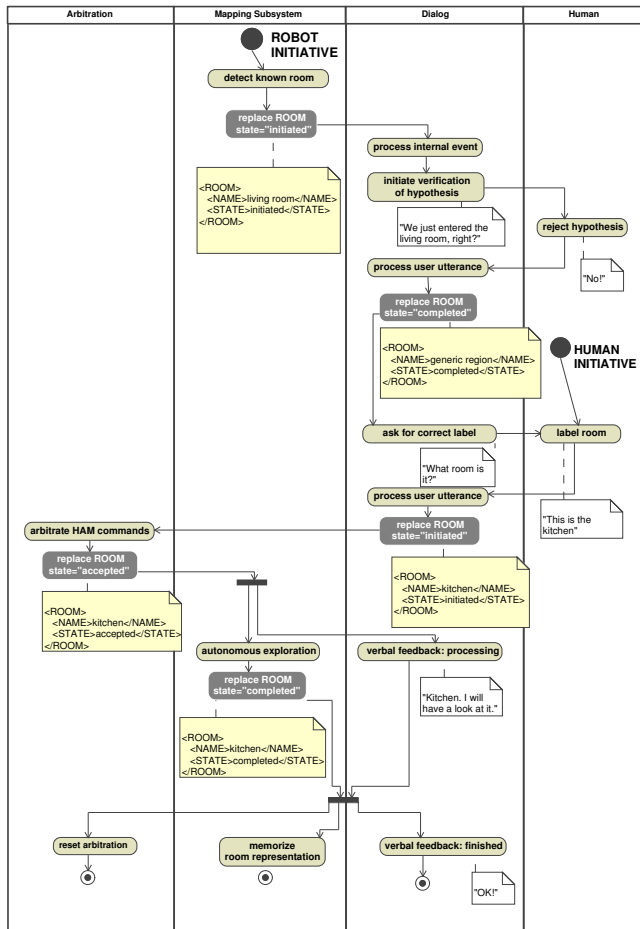


Fig. 5. Activities during robot-initiated clarification respectively human-initiated room teaching

Figure 5 illustrates two examples assuming a situation where the robot misclassifies a new room as the previously learned living room.

If in such situation the robot detects a known room, the task is initiated by the mapping subsystem by replacing the memory element `ROOM` with name `living room` and state `initiated` into the active memory. The dialog system is notified about this event and prompts a verification question (“We just entered the living room, right?”). Since the robot’s hypothesis is incorrect, the human will negate. Having processed the human’s answer, the dialog system finishes the learning task by setting name `generic region` and state `completed`. Driven by the intention to establish mutual understanding and thus enhance the environment model, the dialog system asks for the correct label (“What room is it?”).

What happens in the following is the same as for human initiated room teaching. Based on the human’s utterance, a new room learning task is initiated by the dialog system. The arbitration component now allows the mapping subsystem to take control of the robot’s driving motors by setting the state `accepted` which triggers the mapping subsystem to start the exploration by turning the robot around. At the same

time, the dialog is triggered to give verbal feedback about the processing state (“I will have a look at it.”). As soon as the exploration is finished, the mapping subsystem sets the state `completed`. In consequence, the hardware arbitration is reset, the robot memorizes the new representation and the dialog system gives another verbal confirmation.

Altogether, we realized a two-way negotiation protocol where both directions rely on the same interaction strategies. At the same time, this strategy is a very flexible way of coordination since states may be skipped. In fact, an alternative mapping and localization module has been integrated into the robot system that does not make use of the `accepted` state [16], [14]. For integration, the subscription of the Arbitration on the `accepted` state was removed. The interface to the dialog remained the same, though.

## VI. EVALUATION

In the following we discuss the results that we could achieve with our integrated system.

### A. Method

For the evaluation of the work presented in this paper we assume three important phases. The first *conceptual design and evaluation phase* consisted of a feasibility analysis and interface specification and resulted in a prototypical integration of an initial version of the mapping subsystem into *BIRON’s* interactive framework [3].

In the second *empirical evaluation phase* this initial system was extended and improved as described in this article and evaluated regarding its capabilities for interactive learning of a space representation with the help of test runs in a laboratory environment resembling a domestic setting. We consider this phase as crucial to understand whether the tested algorithms, in this case the mapping subsystem’s transition detection and *region* representation approach, do indeed solve the targeted problem of enabling the interactive acquisition of a consistent environment representation.

Only in the third *user evaluation phase* it is then possible to evaluate the system with naïve users in a real domestic setting as proposed in section II.

Here, we discuss our integrated system and the results we could achieve with respect to one example “tour”, i.e., one of the test runs of the second phase of evaluation.

### B. Result

The test run or “example tour” was conducted in a part of a laboratory at the University of Bielefeld that can be compared to a part of an apartment including living room, kitchen, and a part of a hallway, which were the labels chosen for the tour. The “kitchen” can be reached both from the “hallway” and the “living room” which integrates a loop in the environment, allowing to investigate the ability of the transition detection and room (*region*) representation to recognize a particular, already known room and react appropriately.

One interesting aspect of this experimental run was that no correction of the internal pose estimations coming from

the respective sensors on the robot platform was used. This deliberate decision was made to investigate the system’s capabilities in terms of the interaction it can support and be supported by when relying on the presumably erroneous original pose estimations delivered from the robot’s sensory system. Consequently, the results achieved in the presented experimental run are discussed before this background. As explained in section III, the mapping subsystem would only report transition detections after at least one room is specified. According to the strategy presented in section IV however, the system was assumed to take the initiative immediately after receiving a confirmation for “having left a specified room” and ask for the actual whereabouts, instead of accepting the “generic *region*” as “current room” and wait for further specifications of the user. Consequently, the discourse was after the first, user initiated specification, controlled mostly by the system’s clarification questions.

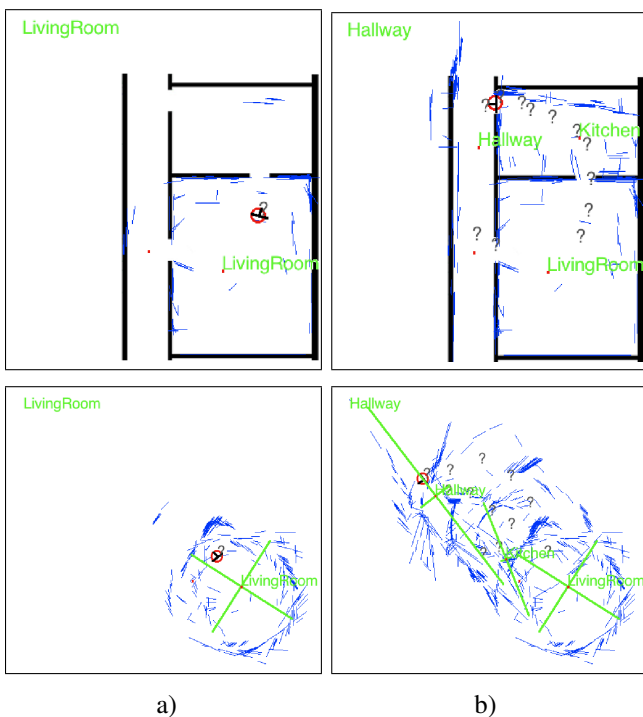


Fig. 6. The experiment with BIRON, visualized in a post-hoc run of the mapping subsystem. Question marks indicate positions where the robot asked the user for confirmation. (top) Reconstruction with the help of a pose estimation module with the room labels marked at the positions where they were given to the robot by the user. (bottom) Visualization of the original percepts with the labeled rooms (*regions*) depicted with the ellipse axes of the statistical descriptor at their computed centers. a) Starting in the “living room”, b) concluding the tour in the hallway after going through living room and kitchen twice. The respective hypotheses about the current region are shown in the upper left corner of each frame.

Figure 6 illustrates the guided tour with BIRON through the laboratory environment, conducted by a researcher acting as user. For better comprehension, the original odometry and scan data acquired by the robot during the trial were post-hoc processed with SLAM methods to be analyzed in consistent maps illustrated in the top row. The lower row illustrates the unprocessed laser scan segments used for the computation

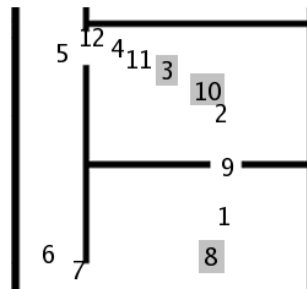


Fig. 7. All transition detections marked with their number in chronological order of occurrence. The spurious detections are highlighted with gray rectangles around them. One (no. 3) occurred during the first round through the environment, the other two (no. 8 and 10) occurred in the second round when the system had accumulated a rather high error in its pose estimation.

of the statistical descriptors of regions. From system logs the positions where the robot took the initiative and asked about a hypothesized transition to a new or known room are computed and indicated in the illustration by question marks.

The tour started in front of the “living room”, where initially nothing was specified. Overall 12 times the robot asked for a confirmation of a hypothesized transition. Five of these situations refer to actual transitions, i.e., door passages, while four more are plausible due to the robot being close to the respective door.

Fig. 7 illustrates the respective spots in the order of occurrence. The three somewhat questionable detections (number 3, 8 and 10) can be explained partially (8 and 10) with the uncertainty the system had gathered due to the error in the position estimation. For the spurious detection (3) a possible explanation might be that the user was blocking the robot’s “view” for a significant amount of time. Taking a look into the run itself it becomes obvious that the user and the robot spent a long time in the “kitchen”, being relatively static. Table IV summarizes the tour with respect to the situations in which the robot asked for clarification, as they are referred to also in fig. 7. The table shows the hypothesis the robot had at the respective time, the question it asked (“Are we still in the X?”, “We just left the X, right?”, “We just entered the X, right”, “What room is it”), the user’s answer (“Yes”, “No”, “This is the X”) and the system reaction (specifying a new room, confirming hypothesis, correcting hypothesis) with the abbreviations: GR = generic region, LR = living room, KI = kitchen, and HW = hallway.

Since the pose estimation error was obviously mostly depending on rotations of the robot platform (see the uncorrected illustration in fig. 6), the overall error was kept on a level that allowed to hypothesize the “hallway” correctly as “current room” when it was re-entered, since no significant turning movements “on the spot” had been made after its specification.

In general this tour shows the system’s ability to generate a basis for meaningful interaction with the user, by which it can update and correct its internal representation of the environment.

## VII. CONCLUSION AND FUTURE IDEAS

In this paper we presented our integrated system for mixed initiative Human Augmented Mapping. The challenge, namely to achieve a consistent and most complete spatial rep-

TABLE IV

THE SITUATIONS IN WHICH *BIRON* ASKED FOR CLARIFICATION

No.	Hyp.	Robot	User	Action
0	GR		"... LR."	new
1	LR	"Are we still ...?"	"Yes"	conf
2	GR	"We just left ...?"	"Yes"	conf
	GR	"What room is it?"	"... KI"	new
3	KI	"Are we still ...?"	"Yes"	conf
4	KI	"Are we still ...?"	"Yes"	conf
5	GR	"We just left ...?"	"Yes"	conf
	GR	"What room is it?"	"... HW"	new
6	HW	"Are we still ...?"	"Yes"	conf
7	GR	"We just left ...?"	"Yes"	conf
	GR	"What room is it?"	"... LR."	corr
8	GR	"We just left ...?"	"No"	corr
9	GR	"We just left ...?"	"Yes"	conf
	GR	"What room is it?"	"... KI"	corr
10	KI	"Are we still ...?"	"Yes"	conf
11	KI	"Are we still ...?"	"Yes"	conf
12	HW	"We just entered ...?"	"Yes"	conf

resentation, has been illuminated from different perspective: The design of the dialog as a proxy between user and system, the required characteristics of the mapping and recognition technique itself, and the interaction strategies between the subsystems in the system architecture. The interplay of the mapping subsystem and the dialog subsystem by means of grounding and mixed-initiative concepts elevated to a systemic level has turned out to be a powerful and generic approach in order to allow flexible learning by interacting. With "state sequencing" and the information-oriented architecture itself the foundation is laid for very flexible learning strategies. The exploitation of the central concepts has been proved by the spatial learning approach in this paper. In this context, we discussed our particular integration strategies that support both user and data (robot) driven initiative for the acquisition of (spatial) information and evaluated our system in the context of test runs and an example tour.

Of course, the cyclic evaluation of our robot must continue. Though our robot system basically has undergone "full-cycle" evaluation including extensive user studies [17], that evaluation release of the robot only featured a spatial learning limited to user's initiative. How naive users will now perceive the novel mixed-initiative learning, is still an open question. However, on the basis of system-level grounding and state sequencing in the information-oriented architecture different learning strategies can now be studied under one conceptual umbrella and compared in real world scenarios with naive users. An example for adaptive learning strategies following the same concept could be a robot that considers a personality models of a human user. So, it can adapt its own level of extrovertness, e.g., not being too very curious and asking for room labels every now and then, if a user obviously is irritated already. The system now also allows to extend the concept of human augmented mapping gradually towards more autonomy for the robot. It can for instance explore open spaces and take initiative whenever a person is in interaction distance, or accept a human's initiative occurring during the autonomous phase.

## ACKNOWLEDGEMENTS

This work has partially been funded by the German Research Foundation (DFG) under the contracts of the Collaborative Research Center 673 "Alignment in Communication", the Cluster of Excellence "Cognitive Interaction Technology" (CITEC), and the German Service Robotic Initiative (DE-SIRE).

## REFERENCES

- [1] B. Wrede, K. J. Rohlfing, M. Hanheide, and G. Sagerer, "Towards learning by interacting," B. Sendhoff, E. Koerner, O. Sporns, H. Ritter, and K. Doya, Eds. Berlin Heidelberg: Springer-Verlag, 2009, pp. 139–150.
- [2] E. Topp, "Human-robot interaction and mapping with a service robot: Human augmented mapping," Doctoral Thesis, KTH School of Computer Science and Communication (CSC), Stockholm, Sweden, Sept. 2008.
- [3] T. Spexard, S. Li, B. Wrede, M. Hanheide, E. Topp, and H. Hüttenrauch, "Interaction awareness for joint environment exploration," in *Proceedings of the Special Session on Situation Awareness in Social Robots at the IEEE International Symposium on Robot and Human Interactive Communication (Ro-Man)*, Jeju Island, Korea, 2007.
- [4] P. Althaus and H. Christensen, "Automatic map acquisition for navigation in domestic environments," in *Proceedings of the IEEE International Conference on Robotics and Automation, Taipei, Taiwan*, Sept. 2003.
- [5] A. Diosi, G. Taylor, and L. Kleeman, "Interactive slam using laser and advanced sonar," in *Proceedings of the IEEE International Conference on Robotics and Automation, Barcelona, Spain*, Apr. 2005.
- [6] H. Zender, P. Jensfelt, Ó. Mozos, G.-J. Kruijff, and W. Burgard, "An Integrated Robotic System for Spatial Understanding and Situated Interaction in Indoor Environments," in *Proceedings of the Conference on Artificial Intelligence (AAAI), Vancouver, BC, Canada*, July 2007.
- [7] G.-J. Kruijff, H. Zender, P. Jensfelt, and H. Christensen, "Clarification dialogues in human-augmented mapping," in *Proceedings of the ACM Conference on Human-Robot Interaction (HRI), Salt Lake City, UT, USA*, 2006.
- [8] E. Topp, H. Hüttenrauch, H. Christensen, and K. Severinson Eklundh, "Bringing together human and robotic environment representations – a pilot study," in *Proceedings of the IEEE/RSJ International Conference of Intelligent Robots and Systems, Beijing, China*, Oct. 2006.
- [9] I. Lütkebohle, J. Peltason, L. Schillingmann, C. Elbrechter, B. Wrede, S. Wachsmuth, and R. Haschke, "The curious robot – structuring interactive robot learning," in *Proceedings of the IEEE International Conference on Robotics and Automation, Kobe, Japan*, May 2009, to appear.
- [10] E. Topp and H. Christensen, "Topological modelling for human augmented mapping," in *Proceedings of the IEEE/RSJ International Conference of Intelligent Robots and Systems, Beijing, China*, Oct. 2006.
- [11] E. A. Topp and H. I. Christensen, "Detecting structural ambiguities and transitions during a guided tour," in *Proceedings of the IEEE International Conference on Robotics and Automation, Pasadena, CA, USA*, May 2008.
- [12] S. Li, B. Wrede, and G. Sagerer, "A computational model of multimodal grounding," ACL Press. ACL Press, 2006, inproceedings, pp. 153–160. [Online]. Available: files/papers/Li2006-ACM.pdf
- [13] H. H. Clark, *Arenas of Language Use*. University Of Chicago Press, May 1993.
- [14] M. Hanheide and G. Sagerer, "Active memory-based interaction strategies for learning-enabling behaviors," *International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 01/08/2008 2008.
- [15] S. Wrede, M. Hanheide, C. Bauckhage, and G. Sagerer, "An Active Memory as a Model for Information Fusion," *Proc. 7th Int. Conference on Information Fusion*, 2004.
- [16] O. Booij, B. Kröse, J. Peltason, T. P. Spexard, and M. Hanheide, "Moving from augmented to interactive mapping," Zurich, 2008.
- [17] M. Lohse, M. Hanheide, K. Rohlfing, and G. Sagerer, "Systemic interaction analysis in hri," 2009.