Measuring Up as an Intelligent Robot – On the Use of High-Fidelity Simulations for Human-Robot Interaction Research

Anders Green
KTH School of Computer Science and Communication
Royal Institute of Technology
100 44 Stockholm, Sweden
green@csc.kth.se

Helge Hüttenrauch
KTH School of Computer Science and Communication
Royal Institute of Technology
100 44 Stockholm, Sweden
hehu@csc.kth.se

Elin A. Topp
KTH School of Computer Science and Communication
Royal Institute of Technology
100 44 Stockholm, Sweden
topp@csc.kth.se

***Abstract*— In this article we describe and discuss how Hi-Fi simulation methods, or Wizard-of-Oz [25], can be employed in a user-centered approach to develop natural language user interfaces for robots with cognitive capabilities. By actively shaping the system both in terms of technology and in interface design the interest of both users and designers are involved in the process as stake-holders. We have analyzed two different simulation studies performed sequentially with different foci, but within the same overall scenario. The data obtained from Hi-Fi simulation studies is to a large extent qualitative, but data can also be used as a resource to be used for component development, e.g. as training data for speech recognizers. The collected data may also be used for quantitative evaluation depending on the experiment design.**

I. Introduction

The creation of an interface for an autonomous robot is to some extent a very different undertaking than developing a graphical user interface for software application for the standard desktop computer.

In the European Cogniron project [1] robots with cognitive capabilities are investigated from different perspectives. The overall aim of the project is not to produce a single robot, instead several scenarios that explore and demonstrate different types of capabilities are investigated.

In this article we describe and discuss how Hi-Fi simulation methods, or Wizard-of-Oz [25], can be used in the research on human-robot interaction, where the specific aim is to explore and investigate interaction models for robots with cognitive capabilities. We have chosen to work with Hi-Fi simulations because we are dealing with technology for which users have very vague ideas, i.e, there are no cognitive service robots available on the market. Instead users' conceptions of service robots with cognitive capabilities are formed from how robots are presented in popular culture. Not even designers, i.e, roboticists, interface designers and usability experts etc, can fully predict in what manner the *robotic artifact* they are creating will be used and accepted. Maulsby et al [27] noted that "playing the role of an agent" enabled the designer to get an in-depth understanding of what type of actions the system needed to provide. This is also appears to be the case for human-robot interaction, thus in order to understand how to design the interactive behavior of a cognitive companion robot we need to understand what this behavior should be like, i.e., formulated as a research challenge:

- *What are the requirements for the interactive behavior of a cognitive companion?*

This research challenge is not as straight-forward as it may seem. First of all the term "cognitive companion" is not well defined, but if we use the scenarios defined in the Cogniron project as a definition we end up with a system that should have the ability to use vision and audio to perceive and understand its environment, to engage in a continuous process of acquiring new knowledge and skills to learn tasks from users by interacting with them. The robot also needs to exhibit socially acceptable behaviors [1].

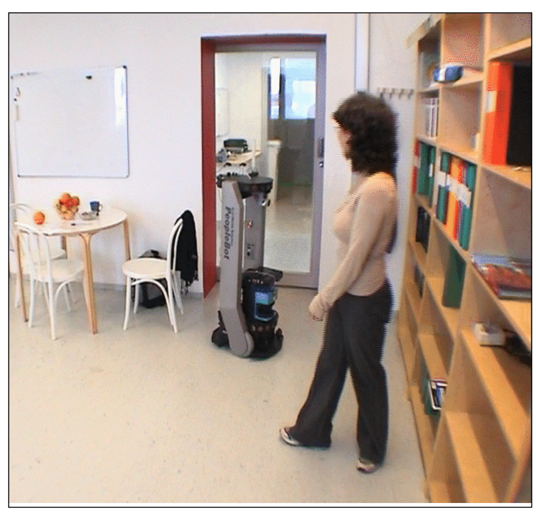

Fig. 1. The robot used in the two user studies in a scene from the first scenario.

The rest of this article is outlined as follows, first we describe the Wizard-of-Oz methodology and how this is can be used in user-centered approaches to design of human-robot interfaces. Then we describe, compare and discuss two different studies that explore the theme of the "Home tour" scenario.

A. *Wizard-of-Oz methodology*

High-fidelity simulation, or Wizard-of-Oz simulation, is a methodology used for simulation of high-level functions in an interactive system. The general idea is to simulate those parts of the system that require most effort in terms of development (like a natural language understanding module) or to assess the suitability of the chosen metaphor. One of the most common uses of Wizard-of-Oz is employed for finding out how users treat a system that uses natural language as an interface.

Wizard-of-Oz simulation methodology in its classical form, i.e., where one user interacts with one (desktop) computer in a lab environment, has been used since the 1970s. The term itself was first used by Kelley [25] two decades ago. Malhotra [26] used the method for simulation of natural language expert systems. The method has also been used for simulation of database question answering systems [7], [27] and uni-modal speech interfaces [3], but also for development of multi-modal interfaces [29], [30], [32].

The starting point for a Wizard-of-Oz study involves the construction of a prototype where some features of the system are for real and where some functions are simulated by one or more operators who control the system's actions and responses.

A classical setup is to put a user in front of a desktop computer in one room and an operator, a wizard, in another room. The user is given a scenario by a test leader; a set of tasks to solve using the novel system and the interactions between the user and the system are recorded. Since the user often is unacquainted with systems of that particular kind, e.g. speech interfaces, or the task, the characteristics of the setup are that of a kind of a role-play, where the user tries to engage and act within the given scenario. Once the experiment is started the user is allowed to interact with the system in the same way as if the system was for real.

The wizard acts as the system's high-level reasoning component responding to the users actions. During the experiment the test leader may intervene if the user gets into trouble related to the use of the (simulated) system. After the user has completed the scenario the test leader normally performs some kind of post session interview. At the end of this interview the test leader briefs the user that the system was in fact simulated and asks permission to use the data that has been collected during the experiment for research purposes. If the user does not give his/her permission the recordings are noted as unanswered and the data media is erased. If there is some kind of reward to be given to the user this will be handed over irrespectively of a negative or positive reply to the request to use data.

B. *User-centered design process*

Hi-fi simulation is to a large extent a method that provides qualitative data on human-robot interaction. Hence, we see it as primarily important in the very early stages of the design process. During the early phases of the system development methods that can be used to communicate the intended use of the product, e.g, mock-ups [9], [14] and scenarios with synthetic dialogues [17] are of great interest in order to inform design. When more mature prototypes have been implemented standard methods for usability evaluation [22] can be employed in order to ensure that the desired usability goals are met [8].

By now, the interaction model for a graphical user interface has become a de facto standard, and is thus known to virtually all users. The interaction model for a robot, however, is presented to the user as a *tabula rasa*, both in terms of lack of knowledge to predict the behavior of interface components and the type of tasks that the robot can perform. Normally a product that is released to the market has undergone several iterations of design and re-design. This process may be of lesser or greater complexity and the methods and work practices of may differ. However, given the fact that we in a sense are developing a "product", we may still employ methods from user-centered software engineering.

By employing a user-centered approach we implicitly accept that we are not only observing The process of developing the user interface, e.g., by taking on the role of evaluating system and human performance along the lines of [35]. Instead we are actively shaping the system both in terms of technology [36] and in interface design [14], [20], hence both users and designers have a role as stake-holders in this process. Thus we need to identify values that we aim to support with our systems [12], as well as to understand that we as researchers are part of the process [4] and not merely evaluating a system that has been created by others.

C. *Hi-Fi Simulation studies of Human-Robot Interaction*

In the Human-Robot Interaction domain there are some compelling reasons for performing simulations rather extracting data from human-human communication.

In many of the envisioned scenarios human-robot communication is focused on communication about topics that are already known to the participants during conversation, making it unnecessary to address these explicitly using speech during the execution of a task. First of all there is the matter of the task that the robot is performing. The tasks that we address the robot with are either simple or focused on domains where explicit verbal instructions are rare or only sparsely used by humans, but seem very important for robots. For instance, explicit negotiation of easily understood tasks (e.g. fetching an object from a known location) [17], and communication concerning detection of humans and description of routes [13]. Secondly, even in scenarios where robots are intended to replace people, in dirty, dangerous or distant environments, the work allocation between actual workers and robots may differ (see [6] for an example of extreme use). This means that the conversation between field workers, like divers or rescue workers would probably be of little direct use to when developing modules for spoken human-robot communication.

We may also perform simulation studies in order to investigate hypotheses concerning general aspects of human-robot interaction, such as social behaviors, e.g, studies of spatial positioning [40], [41] and collaboration [21].

As a method Hi-fi simulation supports a wide variety of activities where designers and users increase their understanding of a future system. According to [15] these activities can be categorized along the dimensions try-out, training and testing:

- *Try out and conceptualize design ideas*, i.e., to rapidly set-up and perform formative studies in order to conceptualize and try out new ideas for the interface that is being developed.
- *Training and education*, i.e., to train natural language components such as speech recognizers. It also provides an opportunity to educate interface designers by letting them to take on the role of the system [27], and thereby increase their understanding of the task.
- *Testing, evaluation and study of use*, using standard usability methods for the type of interface that is being evaluated, for instance using the "Common metrics" approach according to [35] or evaluation methods that are especially targeted for natural language user interfaces, e.g., PARADISE [39]. Last but not least, we include the study of systems in realistic usage contexts, i.e, observations that can be analyzed to increase our understanding of human-robot interaction in general.

II. Wizard simulation strategies

When employing Hi-Fi simulation techniques, such as Wizard-of-Oz, we need to establish what kind of data we are aiming to collect and how it should be analyzed. When we are simulating an autonomous robot with the ability to communicative in natural language we have identified a set of dimensions, in terms of behaviors and appearance, along which the system can vary:

- Degree of system realism: Are we going to simulate a "realistic" system? We must decide to what extent we are going to simulate system behavior so that they appear as realistic to the user. For instance, if we are simulating navigation capability we need to assure that the wizard does not "equip" the system with a navigation system that is able to plan ahead, seemingly inferring the intent of the user. Hence it is important to provide a set of constraints that bring some realism into the situation of use. This is what Maulsby [27] refers to being "true to the algorithm".
- Degree of exploratory freedom: Are we going to simulate according to an "algorithm"? Decisions that are made along this dimension concern whether or not we should allow a completely free and explorative interaction style, or if we should restrict the task
- Behavioral mimicry: how "natural-like" should the system be in terms of appearance and interactive behaviors, i.e, to what extent should the system imitate "nature", in terms of appearance (e.g, using close resemblance with nature [31]) or interactive capability, e.g, allowing for a conversational style of interaction that is close to human capability.

A. Validity of Wizard-of-Oz simulation

Sometimes the Wizard-of-Oz method is criticized for not providing necessary data for evaluating practical dialogue.Wizard-of-Oz methods typically fail to involve users that bring real tasks to the system [23]. This should be seen in contrast to what is generally believed, namely that the Wizard-of-Oz method is an open-ended method for collection of data on user behavior. Allwood and Haglund [2] have noted that the wizard operator acting a scenario is involved in roles on different levels. Thus the researcher role involves acting as a system (wizard operator) and during sessions the wizard can take on different communicative roles like the sender role, the receiver role etc. Bell [5] notes that in a task scenario, like a travel agency dialogue, the wizard not only acts in a system role but in the role of a travel agent. In reality the behavior of people acting in a real situation may be quite different from what people do in a simulated scenario even if the user believes that she is interacting with a real system.

When simulating a multi-modal system with many different ways for the user of providing input the cognitive load of the operators increases. There is sometimes a mismatch between what needs to be simulated and what suits the cognitive and perceptual abilities of the wizards. This phenomenon was been noted by Fitts [10] who proposed a list describing what people do better than machines and what machines do better than people[1]. In a simulation scenario, the problem of function allocation is twofold. First of all we need to consider the characteristics of the system we are simulating and secondly we need to think about the function allocation of the system we are using for the simulation. Following Sheridan [33] we see may see Fitts' List as a set of accepted statements making up the foundation for how to reason when designing function allocation. According to Fitts people are better than machines at detecting small changes in the environment, perceiving patterns; improvising, memorizing, making judgments and reasoning inductively. Machines are better at responding quickly to signals, applying great force, storing and erasing information and reasoning deductively [33]. It is therefore important that we consider the strengths and weaknesses of machines and humans when designing and planning Wizard-of-Oz setups.

Ethical considerations

The use of the Wizard-of-Oz method provides a possible ethical dilemma for the researcher, since in the classical setup, the user is led to believe that the system is a real machine. When the truth is revealed afterwards, the test-leader normally asks for the permission to use the data. There are a number of studies that use deception; in fact it seems that the use of

[1]The so-called MABA-MABA list.

the term Wizard-of-Oz is always connected to the practice of deceiving the user. Fraser and Gilbert [11] have argued against deceiving users on ethical grounds. This ethical dilemma might be solved by telling the user that the system is simulated, making the setup part of the role-play. The validity of the results from a study that is performed in this overt fashion may be questioned but this is dependent on the purpose with the study. Dahlbäck et al [7] argue that for some aspects of dialogue, like the type and frequency of anaphoric references, it is important that users are deceived rather than engaging in a role-play type of scenario. We should note that the work by Dahlbäck [7] and others in the late eighties was mostly done with systems that used text input. For spoken scenarios role-playing has been considered useful in multi-user scenarios [24] and elicitation of error handling strategies [34]. Human natural language performance is automated to a great extent and therefore we may assume that phenomena on the low-level like syntax, and vocabulary used probably are little affected by the fact that the user knows that the system is simulated or not.

However, using the Wizard-of-Oz method in the classical, deceptive manner, ensures the illusion of speaking with a computer. In a non-deceptive Wizard-of-Oz study, this illusion might be broken, something which may have an adverse effect on the result. However, it seems strange to use made up scenarios and role-play while maintaining the standpoint that we should deceive the user to the extent that the system is in fact real. In our view is that it is still an open question if deception is really necessary?

III. Studying the "Home Tour scenario"

The two studies that we are discussing in this paper both stem from work that is performed in the context of the European project Cogniron [1] where a set of scenarios has been defined. In the scenario we have worked with primarily, the user gives the robot a "tour" of the environment:

> *In this experiment, a robot discovers a home-like environment and builds up an understanding of it and of artifacts in it as taught by humans. This process is open-ended, i.e., it has no completion: the robot continues to learn as it faces new situations. A human shows and names specific locations, objects and artifacts, to the robot. The robot can engage in a dialogue in case of missing or ambiguous information.*

This scenario can be characterized as kind of Co-operative Service Discovery and Configuration, stressing the way the user and robot is intended to engage in a joint effort to inform each other of relevant knowledge about the environment. The central themes that the scenario aims to address concerns two main types of information to be jointly discovered by the user and the robot:

(i) the artifacts present in the environment (e.g. objects and locations) and,

(ii) the actions that the robot can perform related to these artifacts.

The two studies that we discuss in this paper explore two key concepts. We have already introduced *Cooperative-Service Discovery and Configuration* which can be viewed as an abstraction of the scenario described above. In the second study the notion of Human-Augmented Mapping (HAM, [36]) was evaluated both in terms of interaction and from a technical point of view. In the following we will compare the set-up and the result of these studies. An overview is given in the Tables I and II.

The general purpose of the first study was to evaluate an extended dialogue model based on [19], [38] in order to provide a rich set of data for further analysis of human-robot communication. The first study was primarily aimed at exploration and not formal hypothesis testing. We did however put down a list of things that we believed that users would do in different situations during the study in order figure out what the wizards should do with the aim to provide a consistent behavior throughout the session.

In the second study, which can be characterized as a pilot study, the general aim was slightly different. We still have an explorative purpose based on the assumption that individuals have different preferences and strategies for how to present their everyday environment. The second study can also be seen as a technical test of a way of represent the environment. Using the notion of MacNamara [28]) we assume that individual strategies are based upon a common psychological representation the meaning and conventionalized use of the terminology that describe users environment. Thus, during the sessions, we use the communicated concepts to annotate the environment representation (i.e., nodes with the types location or region). This means that the role of the wizard is to decide whether the communicated concepts, e.g. phrases like "kitchen" and "my office", can be categorized as locations and regions according to the environment representation. Study two thus aims to answer to questions:

- What strategies are used by individual users to *present* the environment?
- How can the information given by users be incorporated in a representation that can be used as a shared representation.

Both studies can be characterized as Hi-Fi simulations, the first study was performed as pure wizard study, i.e, users were not informed that the system was controlled by a wizard operators. The wizard operators were not completely hidden, instead the users were told that they were "technical support staff" that "monitored the experiment".

When there are two wizard operators it is necessary to carefully consider the task allocation between wizards and to support their collaboration. In the first study there were two wizards, one controlling the dialogue (communicator) and one controlling the movements of the robot (navigator). The wizards stood close together during the session to support contextual awareness, meaning that they easily could get a glimpse of what the other wizard was doing while concentrating on the task. The setup can be seen in Figure 2.

In the second study the users were informed that the speech system was controlled by the accompanying operator. Whereas the first study was performed with 22 invited users (undergraduates from our campus) and thus did not know the environment, the 5 users in the second study were all staff, familiar with the lab environment. As familiarity with the environment was a requirement depending on the research questions posed in the second study lab staff was chosen. In future studies we aim to move the robot to an environment outside the lab so that users that have little experience with robotics can be studied.

The first study was performed in a single room in our lab (Figure 2), furnished like a living room to provide a home-like character. The second study was performed in a significantly larger environment, i.e., a whole floor of our lab comprising about 20 rooms (offices and common areas like kitchen, meeting room and printer area).

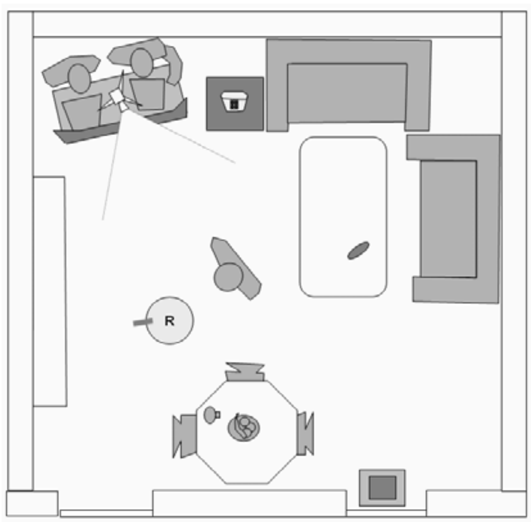

Fig. 2. Map showing the single room used in Study 1

a) Robot appearance, behavior and instructions to users: In both studies we used an ActivMedia PeopleBot equipped with an on-board pan-tilt-camera (see Figure 1. In the first study the robot also was equipped with visible stereo microphones.

In the first study, in terms of behavior the movements (navigation and camera movements) of the robot was controlled by one wizard and the speech output was controlled by another. Since we were aware that a real robot navigation component would make the robot behave in a specific way we tried to mimic this behavior as far as it was possible. In the second study a real follow behavior was used to solve all tasks, except for error recovery, which was performed overtly as teleoperation.

In terms of interactive behavior the first study was more focused on providing a coherent and consistent dialogue behavior, i.e., the dialogue wizard tried to follow an "algorithm" using the terminology by [27]. Hence the subjects received instructions with clear examples of how they could use the system. Our goal was not to test the system "in the wild" without informing the users about how the system was intended to work. Instead we very much instructed the users so that they would be able to use the system. Hence, the instructions were intentionally clear with exact phrasing. The navigation wizard, who also acted as the test leader, also showed the user how the robot worked by giving a small demonstration.

The robot had a dialogue model that accommodated for the actions:

- Greeting, e.g., phrases like "hello" and "my name is john".
- Follow, e.g., phrases like "follow me"
- Demonstration, e.g., referential phrases and gestures like "this is an orange" + pointing gesture.
- Validation, e.g., phrases like "find the orange!"

For the second study the robot used two different ways of giving feedback on users' input, dependent if the user provided information for labeling a *location* or a *region*. Thus if a location, e.g., "coffee machine" or "door to" a *region*, was presented by the user the robot would not move and immediately state that it had stored the location. If a region, e.g., "kitchen", "my office", was presented, the robot would state that it needed to "look around" and make a full 360 degree turn, before confirming that a *region* had been stored. The role of the wizard in was to select which behavior to use based on the wizards understanding if whether the concept used by the user could be interpreted as referencing a *location* or a *region*.

Users were told to use commands, i.e., follow me, *go to* ⟨*target*⟩, *stop, turn* ⟨*left/right*⟩, to control the robot and referential expressions like *this is* ⟨*term*⟩ together with deictic gestures. No terminology that could be directly related to the concepts tested was mentioned to the users, e.g., *location, region, place* etc. Instead phrases like *"everything, that you think the robot needs to know"* was used to guide the user. Since a real follow behavior was used special instructions was needed in order that the system would be able to track and follow the user (i.e., the user had to move about 1 meter to make the robot start following).

b) Data collection setup : In the first study we could collect data in a fairly controlled manner using a video camera placed in the location of the users. We also had web-cameras placed in each corner of the room. Mounted on-board the robot an audio recorder captured high quality stereo sound. Data from the laser range finder was also stored.

During the second study we used a hand-held camera and on-board camera to capture the "scene" of the user and the robot acting together in the environment. Since the environment comprised over twenty offices and common areas (e.g, kitchen, printer area) we could not cover the whole area with web cams or sensors.

In both studies we administered a questionnaire and interviewed the users after the end of the trial. In the first study we asked the users about who was controlling the robot before informing them about that they had been involved in a wizard study and asked again about their willingness

TABLE I

GENERAL PURPOSE AND SETUP OF USER STUDIES 1 AND 2

STUDY TYPE	Study 1: Single room	Study 2: Multiple rooms
Purpose	Data collection and exploration of interaction strategies related to a specific dialogue model [19], [38]	Investigate interaction strategies used by individual users to *present* the environment
Type, size of study	Explorative, 22 subjects (20+20 min)	Pilot-study with 5 subjects (20+25 min)
Key concept	Co-operative Service Discovery and Configuration	Human-Augmented Mapping (cf. [36])
Type of subjects	Undergraduate students	Staff that knew their environment
Environment	Single "living room"	20 rooms + kitchen

TABLE II

INTERACTION DIMENSIONS OF USER STUDIES 1 AND 2

INTERACTION	Study 1: Evaluation of Dialogue model	Study 2: Presentation strategy
Robot behavior and appearance	Dialogue model accomodating tasks: greet, follow, show (object, location), validate, close	Two behaviors: show (acknowledge) and show (look-around)
Instructions	User were provided with a written instruction with examples. Testleader gave a demonstration.	Users were told to use a small set of commands and gestures
Wizard-status	Users were informed about the role of robot operator	Pure wizard – user were told that wizard was an operator
Number and role of wizards	Two wizards (navigator and communicator) behind a screen (cf. [15])	One wizard accompanied user/robot
Data collection setup	• Video camera (overview). Onboard stereo audio (robot perspective). • Network web cameras in each corner. • Questionnaire • Post-trial interview	• Hand-held camera (overview) / onboard camera (robot perspective) • Questionnaire • Post-trial interview (scripted) • Sketch session
Type of data	• Single audio/video track (Mini DV) • Web-cam images ($\sim$ 1 fps) • Data from laser range finder • Stereo audio (22 KHz) • Notes from interviews • Questionnaire data	• Two video tracks with audio (Mini DV) • Notes from interviews • Data from laser range finder • Environment representation • Questionnaire data • Video from sketch session (Mini DV)

to participate given this new information. No user had any objections once we had revealed the truth and explained the reasons for our conduct. In the second study the subjects were informed that the system was partially simulated. The post-session interviews in the second experiment also included a sketch session where the user was instructed to draw a map on a white-board and explain how they had instructed the robot.

A. Findings from the studies

Several analyses have been performed using the data collected. An analysis of the communicative acts have been made and reported in [16]. We have also studied the role of posture and positioning [20] and active spatial influence on the part of the robot [18].

We performed an analysis of the miscommunication using the video recordings from the first 12 user sessions. These were transcribed and synchronized on the utterance level. We then printed out all dialogues and analyzed them by marking utterances that could be interpreted as symptoms of miscommunication.

We found miscommunication that could be attributed to the users' erroneous inferences about the system's capability. For instance, users tried to hold up objects in front of the camera. This was considered to be an error according to the task model and an a repair was issued by the communicator wizard.

We also noted several types of problems related to feedback in our data. Providing relevant and timely feedback essential to maintaining an orderly and well managed dialogue. We have identified problems related to timing, i.e., feedback is *ill-timed*, something which may render it incoherent. Another problem that occurred in the material was lack of feedback, i.e., the robot does not respond to the user's contribution before the user decides to make another contribution. Overlapping speech was also a problem in the corpus, i.e, the robot should have stopped the synthesized speech when the robot started to speak, something which lead to misunderstandings.

Another problem that can be described as an effect of the modality is miscommunication related to reference. In the manner the system was simulated we allowed for a ”robust” object recognition system, meaning that the system would recognize any object given that it was small enough and placed on a flat surface. The information can be said to have been negotiated, but since there is no pointing capability apart from the general direction indicated by the front robot and the on-board camera, there is no way of indicating precisely which object has been detected.

Results from the second study have been reported in terms of a technical evaluation of the environment representation [37]. The strategies employed by the users were very different and no convergence in behavior could be established since there were only five subjects. The observed behaviors still need to be taken into account when constructing future systems. Some of the phenomena are provided below to give a sense of what went on during the sessions:

- Users' personal view of the environment affects the way it is introduced to the robot, e.g., a person not drinking coffee might not mention the coffee maker.
- Persons were pointed out to the robot as offices (with occupants in them) were passed. Bystanders that passed the robot were also pointed out by one user.
- Locations pointed out reflected an action oriented view of the environment, i.e., a coffee machine is mentioned as a place where the robot should get coffee, but the kitchen is not mentioned.
- Doorways were pointed out by subjects with knowledge about robot navigation
- After giving the follow-command, the users stood still and waited for the robot to move, something which was not anticipated. Hence it needs to be equipped with an active way of prompting the user to move (i.e., perform an act of *spatial prompting* [18]).

IV. Discussion

When looking at the two different studies in a bird's eye perspective, the generic aspects of both studies can be contributed the context of use provided by the overall scenario, the Home Tour:

- The use of natural language using speech and gestures
- The initiative during the interaction lies with the user
- The behavior of the robot influences the user both in terms of communicative behavior and spatial adaptation.

In some respect the different foci of the studies affect the way studies have been setup. When mobility of both users and systems become a topic for investigation, the complexity of setting up the scenario increases, something which have an effect the way data can be collected.

Clearly the wizard-of-oz framework can be employed successfully for simulating a full fledged interactive robot system. There are several dimensions that have been added since the early work on screen-based applications [7], [25]–[27]. The dimension of collaborative behavior has always been in focus of wizard-of-oz studies for natural language user interfaces. For human-robot interaction the situated use context is in focus, i.e, the human and robot share the same environment and partially share the same perceptual context.

Since a robot system allow several degrees of freedom the manner and ability for the wizards to control the robot system almost becomes research topic in itself. It is clear that to some extent we need to assure that we do not spend more resources on building a simulation environment than on building the real system. The use of a real robots also makes safety issues come in focus, hence we need to assure that we can maintain safety during the sessions. We also need to check that the more formal aspects of user studies, i.e, make certain that legal and ethical issues are compatible with the type of studies we will perform (see Walters et al [40] for a specific example from the UK).

V. Conclusions

In this article we have described the way in which Hi-Fi simulation studies can be used in the process of developing interactive service robots. When carefully designed, simulation studies will provide data about different aspects of human-robot interaction that would otherwise be unattainable until large effort had been spent on the creation of a working prototype. Thus, the data that we get from a Hi-Fi simulation study is to a large extent qualitative, but we may also collect data as a resource to be used for component development, e.g. as training data for speech recognizers. We may also use the collected data to perform quantitative evaluations using different performance metrics, e.g, time-to-completion for a defined task [35] or user satisfaction [39] with respect to interactivity.

The type of data we can get a wizard-of-oz type simulation study ranges from collection of data that can be analyzed in different ways.

- Data on language use and task strategies, especially spatial language (including gesture and speech) can be analyzed with methods from psychology, linguistics etc.
- Visualization of the whole interaction to enable the designer to conceptualize what the system could or should do in different realistic situations.
- Assessment of users' attitudes towards a future system or towards robots in general.

In the future our aim is to evaluate systems where tasks that now are handled by wizard operators gradually are replaced with real components. To succeed with this we need investigate how we can provide a system which *can* be managed by wizards, while providing the necessary exploratory freedom for users together with requirement of an appropriate degree of system realism.

VI. Acknowledgment

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

REFERENCES

[1] COGNIRON, Annex 1 - Description of Work, 2003. EU Sixth Framework Program FP6-IST-002020, http://www.cogniron.org.

[2] J. Allwood and B. Haglund. Communicative Activity Analysis of a Wizard of Oz Experiment. Technical report, Department of Linguistics, Göteborg University, 1992.

[3] G. Antoniol, R. Cattoni, M. Cettolo, and M. Federico. Robust Speech Understanding for Robot Telecontrol. In *Proceedings of the 6th International Conference on Advanced robotics*, pages 205–209, Tokyo, Japan, 1993.

[4] L. J. Bannon and S. Bœdker. *Designing Interaction: Psychology at the human-computer interface*, chapter Beyond the Interface: Encountering Artifacts in Use. Cambridge University Press, New York, 1991.

[5] L. Bell. *Linguistic Adaptations in Spoken Human-Computer Dialogues: Empirical studies of User Behavior*. PhD Thesis, KTH Royal Institute of Technology, 2003. TRITA-TMH 2003:11.

[6] J. Casper and R. R. Murphy. Human-robot interactions during the robot-assisted urban search and rescue response at the World Trade Center. *IEEE Transactions on Systems, Man, and Cybernetics Part B*, 33(3):367–385, 2003.

[7] N. Dahlbäck, A. Jönsson, and L. Ahrenberg. Wizard of Oz studies - why and how. *Knowledge-Based Systems*, 6(4):258–256, 1993.

[8] P. Dario, E. Guglielmelli, and C. Laschi. Humanoids and personal robots: Design and experiments. *Journal of Robotic Systems*, 18(12):673–690, December 2001.

[9] P. Ehn and M. Kyng. *Cardboard computers: mocking-it-up or hands-on the future*, pages 169–196. Lawrence Erlbaum Associates, Inc., 1992.

[10] P. M. Fitts. *Human engineering for an effective air navigation and traffic control system*. National Research Council Committee on Aviation Psychology, Washington, DC, 1951.

[11] N. M. Fraser and G. N. Gilbert. Simulating Speech Systems. *Computer Speech & Language*, 5(1):81–99, 1991.

[12] B. Friedman and P. Kahn. The human-computer interaction handbook. chapter Human values, ethics, and design. Lawrence Erlbaum Associates, Mahwah, NJ, 2003. Eds: J. A. Jacko and A Sears.

[13] J. Fry, H. Asoh, and T. Matsui. Natural Dialogue with the JIJO-2 Office Robot. *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2:1278–1283, 1998.

[14] A. Green, H. Hüttenrauch, M. Norman, L. Oestreicher, and K. Severinson Eklundh. User-Centered Design for Intelligent Service Robots. In *Proceedings of 9th IEEE International Workshop on Robot and Human Interactive Communication*, Osaka, Japan, 2000.

[15] A. Green, H. Hüttenrauch, and K. Severinson Eklundh. Applying the Wizard-of-Oz framework to Cooperative Service Discovery and Configuration. In *13th IEEE International Workshop on Robot and Human Interactive Communication RO-MAN 2004*, pages 575–580, 20-22 Sept 2004.

[16] A. Green, H. Hüttenrauch, E. A. Topp, and K. S. Eklundh. Developing a Contextualized Multimodal Corpus for Human-Robot Interaction. In *Proceedings of the Fifth international conference on Language Resources and Evaluation LREC2006*, 2006.

[17] A. Green and K. Severinson Eklundh. Task-oriented Dialogue for CERO: a User-centered Approach. In *Proceedings of 10th IEEE International Workshop on Robot and Human Interactive Communication*, Bordeaux/Paris, September 2001.

[18] A. Green, B. Wrede, K. S. Eklund, and S. Li. Integrating Miscommunication Analysis in the Natural Language Interface Design for a Service Robot. submitted to IROS2006, 2006.

[19] A. Haasch, S. Hohenner, S. Huewel, M. Kleinehagenbrock, S. Lang, I. Toptsis, G. A. Fink, J. Fritsch, B. Wrede, , and G. Sagerer. BIRON – The Bielefeld Robot Companion. In *Proceedings of ASER 2004 - 2nd International Workshop on Advances in Service Robots*, Stuttgart, Germany, May 21 2004.

[20] H. Hüttenrach, A. Green, M. Norman, L. Oestreicher, and K. Eklundh Severinson. Involving Users in the Design of a Mobile Office Robot. *Systems, Man and Cybernetics, Part C: Applications and reviews*, 34(2):113–124, 2004.

[21] H. Hüttenrauch and K. Severinson Eklundh. To Help or Not to Help a Service Robot. In *Proceedings of the 12th IEEE International Workshop on Robot and Human Interactive Communication RO-MAN'2003*, Millbrae CA, USA, 2003. IEEE.

[22] ISO. International Standard ISO/IEC 9126. Information technology - Software product evaluation - Quality characteristics and guidelines for their use, International Organization for Standardization. 1991.

[23] A. Jönsson and N. Dahlbäck. Distilling dialogues - A method using natural dialogue corpora for dialogue systems development. In *Proceedings of 6th Applied Natural Language Processing Conference*, pages 44–51, 2000.

[24] K. Kanto, M. Cheadle, B. Gambäck, P. Hansen, H. K. Kristiina Jokinen, and J. Rissanen. Multi-Session Group Scenarios for Speech Interface Design. In C. Stephanidis and J. Jacko, editors, *Human-Computer Interaction: Theory and Practice (Part II)*, volume 2, pages 676–680. Lawrence Erlbaum Associates, Mahwah, New Jersey, 2003.

[25] J. F. Kelley. An iterative design methodology for user-friendly natural language office information applications. *ACM Transactions on Information Systems (TOIS)*, 2(1):26–41, 1984.

[26] A. Malhotra. Design criteria for a knowledge-based English language system for management: an experimental analysis. Technical Report MAC TR-146, MIT, 1975.

[27] D. Maulsby, S. Greenberg, and R. Mander. Prototyping an Intelligent Agent through Wizard of Oz. In *INTERCHI'93*, pages 277 – 282. ACM, April 1993.

[28] T. McNamara. Mental representations of spatial relations. *Cognitive Psychology*, 18:86–121, 1986.

[29] S. Oviatt. User-centered Modeling for Spoken Language and Multimodal Interfaces. *IEEE Multimedia,*, 3(4):26–35, 1996.

[30] D. Perzanowski, A. Schultz, W. Adams, E. Marsh, and M. Bugajska. Building a multimodal human-robot interface. *Intelligent Systems, IEEE [see also IEEE Expert]*, 16(1):16–21, 2001.

[31] T. Saito, T. Shibata, K. Wada, and K. Tanie. Relationship between interaction with the mental commit robot and change of stress reaction of the elderly. In *Proceedings of IEEE International Symposium on Computational Intelligence in Robotics and Automation*, volume 1, pages 119–124, 16-20 July 2003.

[32] D. Salber and J. Coutaz. Applying the Wizard of Oz Technique to the Study of Multimodal Systems. In *EWHCI*, pages 219–230. 1993.

[33] T. B. Sheridan. Function allocation: algorithm, alchemy or apostasy? *International Journal of Human-Computer Studies*, 52(2):203–216, 2000.

[34] G. Skantze. Exploring Human Error Handling Strategies: Implications for Spoken Dialogue Systems. In *Proceedings of ISCA Tutorial and Research Workshop on Error Handling in Spoken Dialogue Systems*, pages 71–76, 2003.

[35] A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, and M. Goodrich. Common Metrics for Human-Robot Interaction. In *In Proceedings of HRI2006 1st annual conference on Human-Robot Interaction*, Salt Lake City, UT, USA, March 2-3 2006. ACM.

[36] E. A. Topp and H. I. Christensen. Tracking for Following and Passing Persons. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2005)*, Edmonton, Alberta, August 2005.

[37] E. A. Topp, H. Hüttenrauch, and H. I. C. K. Severinson Eklundh. Acquiring a shared environment representation. In *Proceedings of the 1st ACM Human Robot Interaction Conference, HRI2006*, Salt Lake City, Utah, USA, March 2006. (extended abstract).

[38] I. Toptsis, S. Li, B. Wrede, and G. A. Fink. A Multi-modal Dialog System for a Mobile Robot. In *INTERSPEECH: 8th International Conference on Spoken Language Processing (ICSLP)*, volume 1, pages 273–276, Jeju, Korea, 2004.

[39] M. A. Walker, D. J. Litman, C. A. Kamm, and A. Abella. PARADISE: A Framework for Evaluating Spoken Dialogue Agents. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL-97)*, pages 271–280, Madrid, Spain, July 1997.

[40] M. Walters, S. Woods, K. Koay, and K. Dautenhahn. Practical and methodological challenges in designing and conducting interaction studies with human subjects. In *Proceeding of AISB Symposium on Robot Companions Hard Problems and Open Challenges in Human-Robot Interaction*, pages 110–120, Hertfordshire UK, April 2005 2005.

[41] M. L. Walters, K. Dautenhahn, K. L. Koay, C. Kaouri, R. te Boekhorst, C. L. Nehaniv, I. Werry, and D. Lee. Close encounters: Spatial distances between people and a robot of mechanistic appearance. In *Proceedings of IEEE-RAS International Conference on Humanoid Robots (Humanoids2005)*, pages 450–455, Tsukuba, Japan, December 5-7 2005.