Exploring Miscommunication and Collaborative Behaviour in Human-Robot Interaction

Theodora Koulouri
Department of Information Systems and Computing
Brunel University
Middlesex UB8 3PH
theodora.koulouri@brunel.ac.uk

Stanislao Lauria
Department of Information Systems and Computing
Brunel University
Middlesex UB8 3PH
stasha.lauria@brunel.ac.uk

Abstract

This paper presents the first step in designing a speech-enabled robot that is capable of natural management of miscommunication. It describes the methods and results of two WOz studies, in which dyads of naïve participants interacted in a collaborative task. The first WOz study explored human miscommunication management. The second study investigated how shared visual space and monitoring shape the processes of feedback and communication in task-oriented interactions. The results provide insights for the development of human-inspired and robust natural language interfaces in robots.

1 Introduction

Robots are now escaping laboratory and industrial environments and moving into our homes and offices. Research activities have focused on offering richer and more intuitive interfaces, leading to the development of several practical systems with Natural Language Interfaces (NLIs). However, there are numerous open challenges arising from the nature of the medium itself as well as the unique characteristics of Human-Robot Interaction (HRI).

1.1 Miscommunication in Human-Robot Interaction

HRI involves embodied interaction, in which humans and robots coordinate their actions sharing time and space. As most speech-enabled robots remain in the labs, people are generally unaware of what robots can understand and do resulting in utterances that are out of the functional and linguistic domain of the robot. Physical copresence will lead people to make strong but misplaced assumptions of mutual knowledge (Clark, 1996), increasing the use of underspecified referents and deictic expressions. Robots operate in and manipulate the same environment as humans, so failure to prevent and rectify errors has potentially severe consequences. Finally, these issues are aggravated by unresolved challenges with automatic speech recognition (ASR) technologies. In conclusion, miscommunication in HRI grows in scope, frequency and costs, impelling researchers to acknowledge the necessity to integrate miscommunication in the design process of speech-enabled robots.

1.2 Aims of study

The goal of this study is two-fold; first, to incorporate "natural" and robust miscommunication management mechanisms (namely, prevention and repair) into a mobile personal robot, which is capable of learning by means of natural language instruction (Lauria et al., 2001). Secondly, it aims to offer some insights that are relevant for the development of NLIs in HRI in general. This research is largely motivated by models of human communication. It is situated within the language-as-action tradition and its approach is to explore and build upon how humans manage miscommunication.

2 Method

We designed and performed two rounds of Wizard of Oz (WOz) simulations. Given that the general aim of the study is to determine how robots should initiate repair and provide feedback in collaborative tasks, the simulations departed from the typical WOz methodology in that the wizards were also naive participants. The domain of the task is navigation. In particular, the user

Proceedings of SIGDIAL 2009: the 10th Annual Meeting of the Special Interest Group in Discourse and Dialogue, pages 111–119,
Queen Mary University of London, September 2009.

guided the robot to six designated locations in a simulated town. The user had full access to the map whereas the wizard could only see the surrounding area of the robot. Thus, the wizard relied on the user's instructions on how to reach the destination. In this section we outline the aim and approach of each WOz study, the materials used and the experimental procedure. Sections 4 and 5 focus on each study individually and their results.

2.1 The first WOz study

This study is a continuation of previous work by the authors (Koulouri and Lauria, 2009). In that study, the communicative resources of the wizard were incrementally restricted, from "normal" dialogue capabilities towards the capabilities of a dialogue system, in three experimental conditions:

- The wizard simulates a super-intelligent robot capable of using unconstrained, natural language with the user (henceforth, Unconstrained Condition).
- The wizard can select from a list of default responses but can also ask for clarification or provide task-related information (henceforth, Semi-Constrained condition).
- The wizard is restricted to choose from a limited set of canned responses similar to a typical spoken dialogue system (SDS).

The current study investigates the first two conditions and presents new findings.

2.2 The second WOz study

The second round of WOz experiments explored the effects of monitoring and shared visual information on the dialogue.

2.3 Set-up

A custom Java-based system was developed and was designed to simulate the existing prototype (the mobile robot). The system consisted of two applications which sent and received coordinates and dialogue and were connected using the TCP/IP protocol over a LAN. The system kept a log of the interaction and the robot's coordinates.

The user's interface displayed the full map of the town (Figure 1). The dialogue box was below the map. Similar to an instant messaging application, the user could type his/her messages and see the robot's responses appearing on the lower part of the box. In the first WOz study, the user's interface included a small "monitor" on the upper right corner of the screen that displayed the current surrounding area of the robot, but not the robot itself. Then, for the purposes of the second study, this feature was removed (see Figure 1 in Appendix A).

Figure 1. The user's interface.

The wizard's interface was modified according to the two experimental conditions. For both conditions, the wizard could only see a fraction of the map- the area around the robot's current position. The robot was operated by the wizard using the arrow keys on the keyboard. The dialogue box of the wizard displayed the most recent messages of both participants as well as a history of the user's messages. The buttons on the right side of the screen simulated the actual robot's ability to remember previous routes: the wizard clicked on the button that corresponded to a known route and the robot automatically executed. In the interface for the Unconstrained condition, the wizard could freely type and send messages (Figure 2).

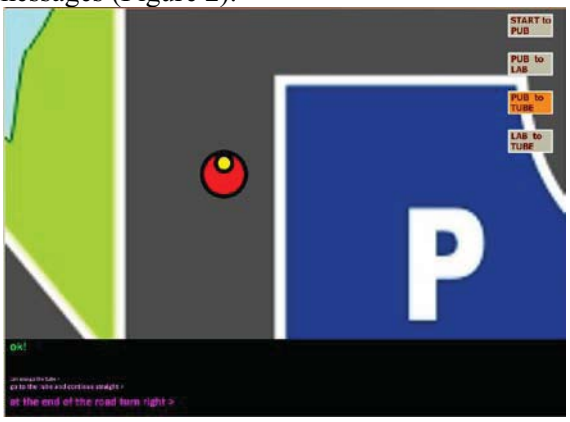

Figure 2. The wizard's interface in the Unconstrained condition.

In the version for the Semi-Constrained condition, the wizard could interact with the user in two ways: first, they could click on the buttons, situated on the upper part of the dialogue box, to automatically send the canned responses, "Hel-

lo", "Goodbye", "Yes", "No", "Ok" and the problem-signalling responses, "What?", "I don't understand" and "I cannot do that". The second way was to click on the "Robot Asks Question" and "Robot Gives Info" buttons which allowed the wizard to type his/her own responses (see Figure 2 in Appendix A).

2.4 Procedure

A total of 32 participants were recruited, 16 users and 16 wizards. The participants were randomly assigned to the studies, experimental conditions and to the roles of wizard or user. The pairs were seated in different rooms equipped with a desktop PC. The wizards were given a short demonstration and a trial period to familiarise with the operation of the system and were also informed about whether the users would be able to monitor them. The users were told that they would interact with a robot via a computer interface; this robot was very fluent in understanding spatial language and could give appropriate responses, it could learn routes but had limited vision. The users were asked to begin each task whenever they felt ready by clicking on the links on their computer screen, start the interaction with "Hello", which opened the wizard's application, and end it with "Goodbye" which closed both applications. The participants received verbal and written descriptions of the experiment. They were not given any specific guidelines on how to interact or what routes to take. However, the users were asked not to use directions such as "north", "south", "up", "down" etc. The wizards were also advised not to try to "sound like a robot", but respond normally.

3 Data analysis

The dialogues collected were annotated on the basis of dialogue acts (DAs), miscommunication, and task status. An example of an annotated dialogue is shown in Table 1.

ID	(x,y@TIME) MESSAGE	T.S.	MISC	HCRC TAG	
U1	(899,445@10:44:34) go straight on.			instruct	
R2	(880,286@10:44:59) There is a wall straight on.		Rej	explain	Imp
U3	(884,286@10:45:28) ok turn left here then take the third right			acknowledge/ instruct	
R4	(352,216@10:45:58) Before the bridge?		Non	check	Imp
U5	(351,68@10:46:17) this is the wrong place	WE	Corr	explain	
R6	(351,68@10:46:39) Wrong place. Should Robot go back?			acknowledge/ query-yn	
U7	(351,68@10:46:55) go forward and at the crossroads keep going forward and the tube is at the end of the road			instruct	
R8	(351,0@10:47:14) Out of bounds.	WE		explain	
R9	(351,608@10:47:47) Where to go?			query-w	
U10	(364,608@10:48:12) the tube is in front of you			instruct	
R11	(402,547@10:48:23) Is it this one?	BOT		query-yn	
U12	(402,547@10:49:7) yes it is.	SUC		reply-y	

Table 1. Example of an annotated dialogue. ID denotes the speaker (User or Robot), T.S. stands for task status and MISC for miscommunication.

3.1 Annotation of dialogue acts

The DAs in the corpus were annotated following the HCRC coding scheme (Carletta et al., 1996). Motivated by Skantze (2005), the last column in Table 1 contains information on the explicitness of the response. This feature was only relevant for repair initiations by the wizards. For instance, responses like "What?" and the the ones in Table 3 were considered to be explicit (EX) signals of miscommunication, whereas lines 2 and 4 in the dialogue above were labelled as implicit (IMP).

3.2 Annotation of task execution status

The coordinates (x,y) of the robot's position recorded for every exchanged message were placed on the map of the town (of dimensions 1024x600 pixels) allowing the analysts to retrace the movements of the robot. Wrong executions (WE) were determined by juxtaposing the user's instruction with the robot's execution, as indicated by the coordinates. Back-on-Track (BOT) was tagged when the first user instruction after a wrong execution was executed correctly. Finally, task success (SUC) was labelled when the robot reached the destination and it was confirmed by the user.

3.3 Annotation of miscommunication

The annotation of instances of miscommunication in the dialogues is based on the definitions given by Hirst et al. (1994). Miscommunication includes three categories of problems: misunderstandings, non-understandings and misconceptions. First, misunderstandings occur when the hearer obtains an interpretation which is not aligned to what the speaker intended him/her to obtain. In this study, without attempting to unveil the intention of the user, misunderstandings were

tagged when the user (who was monitoring the understanding) signalled a wrong execution (see line 5 in Table 1). These correction tags (Corr) did not always coincide with wrong execution tags, but were used when the user became aware of the error (after receiving visual or verbal information). Following the same definition, misunderstandings were also tagged as rejections (tag: Rej) when the wizard expressed inability to execute the instruction (for instance, given the robot's current location, as shown in line 2 in the dialogue), although he/she was able to interpret it. Secondly, non-understandings (tag: Non, line 4) occurred when the wizards obtained no interpretation at all or too many. Non-understandings also included cases in which wizards were uncertain about their interpretation (as suggested by Gabsdil, 2003). Lastly, misconceptions happen when the beliefs of the interlocutors clash, and are outside the scope of this study.

4 First WOz study

Skantze (2005) and Williams and Young (2004) performed variations of WOz studies to explore how humans handle ASR errors, using a real or simulated speech recogniser. They discovered that even after highly inaccurate recognition output, the participants rarely signalled non-understanding explicitly. Accordingly, the experimental hypothesis of the present study is that wizards in both conditions will not choose explicit responses to signal miscommunication (such as "I don't understand" or "What?") but responses that contribute with information.

ASR is a major source of errors in SDS. But as miscommunication is ubiquitous in interaction, there are many other sources of ambiguity that give rise to problematic understanding. Thus, for the current purposes of this work, it was decided that ASR would have an overwhelming effect on the interaction that might prevent the observation of other interesting dialogue phenomena.

This section describes further work on the Unconstrained and Semi-Constrained conditions (see Section 2.1). Twenty participants were recruited and randomly allocated to each condition.

4.1 Results

Analysis of the dialogues of the Unconstrained condition reinforced previous findings and confirmed the experimental hypothesis. In particular, wizards never used explicit repairs, but preferred to describe their location, request clarification and further instructions. Integrating finer classification of clarification requests (CRs) and the original dialogue act tagging, the DAs used by the wizards to signal non-understandings and rejections were categorised as shown in Table 2.

Dialogue Act	Explanation
Explain	The wizard gives description of robot's location. E.g., "I crossed the bridge.", "I am at a crossroad".
Check	This category covers CRs. The corpus contained two types of CRs: first, task-level reformulations (as in line 4 in Table 1), which reformulate the utterance on the basis of its effects on the task, showing the wizard's subjective understanding (Gabsdil, 2003). Second, alternative CRs which occur when the wizard gives two alternative interpretations, trying to resolve referential ambiguity. For instance, "back to the bridge or to the factory", to resolve "go *back* to last location".
Query-w	The wizard asks for further instructions. E.g., "Please give me further instructions."
Explain+Query-w	A combo of actions; the wizard provides information on location and asks for further instructions. E.g., "crossroads, now where?"

Table 2.Wizard DAs after miscommunication.

Figure 3 illustrates the distribution of these responses to signal non-understandings and rejections (columns labelled "Uncons-NON" and "Uncons-REJ", respectively). Evidently, there is a much greater variety of CRs than the two CR types reported here, as described in the work of Purver (2006) and Schlangen (2004). However, for a navigation task and having excluded ASR errors, problems occurred mainly in the meaning recognition level (explained below) and aimed for reference resolution.

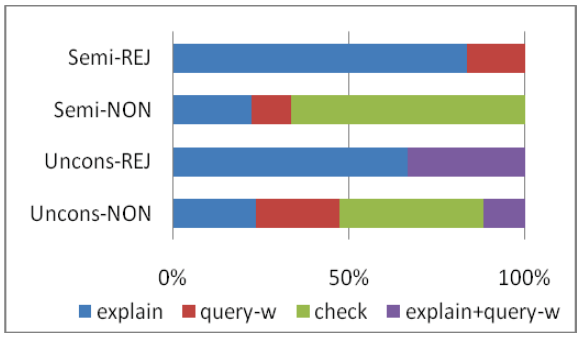

Figure 3. Use of strategies to signal non-understandings or rejections, for either condition.

In conclusion, wizards in the Unconstrained condition did not directly signal problems in understanding but, instead, they attempted to advance the dialogue by providing task-related information in either the form of CRs or simple statements. The study contributes to the findings presented in Skantze (2005) and Williams and Young (2004) in that it demonstrates the use of similar strategies to deal with different sources of problems.

In the Semi-Constrained condition, a degree of restrain and control over the error handling capacity of the wizards was introduced. In particular, the wizards could explicitly signal communication problems in the utterance, meaning and action level using three predefined responses. This is inspired by the models of Clark (1996) and Allwood (1995), according to which, miscommunication can occur in any of these levels and people select repair initiations that point to the source of the problem. The model (adapted from Mills and Healey, 2006) and the responses are schematically shown in Table 3 below.

Levels of Communication		Wizard Responses
Level 1	Securing Attention	-
Level 2	Utterance Recognition	"What?"
Level 3	Meaning Recognition	"Sorry, I don't understand."
Level 4	Action Recognition	"I cannot do that."

Table 3. Levels of communication.

Moreover, based on the classification of the wizard's error handling strategies in the Unconstrained condition (Table 2), we collapsed the observed strategies in two categories of responses which resulted in adding two more error handling buttons; namely, the button denoted as "Robot Asks Question" corresponded to the "Check" and "Query-w" strategies. The "Robot Gives Info" was associated with "Explain". This clear labelling of error handling actions presented to the wizards of the Semi-Constrained condition aimed to "coerce" them to use the strategies in a more transparent way. This could allow us a glimpse to the mechanisms and processes underlying human miscommunication management.

Analysis of the dialogues revealed that in the Semi-Constrained condition wizards employed both explicit and implicit strategies. Figure 4 shows the distribution of explicit and implicit responses to signal non-understandings and rejections. Figure 3 shows the frequency of each implicit strategy to signal non-understandings (Semi-NON) and rejections (Semi-REJ).

The initial prediction was that wizards will not use explicit signals of problems in the dialogue. This was contradicted by the results. It can be argued that the physical presence of the buttons and the less effort required account for this phenomenon. On the other hand, it is also plausible to assume that these strategies matched what the wizards wanted to say. Finally, there were no significant differences between conditions in terms of user experience, task success and time on task (as reported in Koulouri and Lauria, 2009).

Figure 4. Occurrence of implicit and explicit miscommunication signals (Semi-Constrained).

4.2 Discussion and future work

The findings of this study could be extrapolated to HRI. Classification of the responses of the wizards resulted in a limited set of error signalling strategies. Therefore, in the presence of miscommunication the robot could use the static, explicit strategies. But these strategies alone are inadequate (as shown by Koulouri and Lauria, 2009). They need to be supplemented, but not entirely replaced, with dynamic error handling strategies; namely, posing relevant questions and providing descriptions of location. Yet this entails several challenges. Gabsdil (2003) identifies the complexity of adding clarification requests to systems with deep semantic processing. With regard to alternative clarifications, systems would need to generate two alternative interpretations for one referent. Task-level reformulations would also require the system to have the capability to identify the effects of all possible executions of the instruction. As a next step, we will focus on issues concerning the implementation of such functionality.

Schlangen (2004) suggests that "general-purpose" repair initiations, such as "What?", which request repetition of the whole utterance, are more severe for the dialogue compared to reprise fragments (e.g., "Turn where?") that accept part of the utterance. Mills and Healey (2006) also found that "What's" were more disruptive to the dialogue than reprise fragments. Guided by these insights, our current work looks at how each error strategy affects the subsequent unfolding of the dialogue.

5 The second WOz study

Research in human communication has shown that in task-oriented interactions visual information has a great impact on dialogue patterns and improves performance in the task. In particular, Gergle at al. (2004), Clark and Krych (2004) and

Brennan (2005) explored different communication tasks and compared a condition, in which visual and verbal information was available, with a speech-only condition. In their experiments, a person gave instructions to another participant on how to complete a task. Their findings seem to resonate. In terms of time for task completion and number of words per turn, the interactions in the visual information condition were more efficient. The physical actions of the person following the instructions functioned as confirmations and substituted for verbal grounding. Regarding errors, no significant differences were observed between visual and speech-only conditions. Motivated by these findings in human-human interaction, the second study aims to identify the differences in the processes of communication depending on whether the user can or cannot monitor the actions of the robot.

5.1 Experimental design

The study followed a between-subjects factorial design. Experiments were performed for four different conditions, as illustrated in Table 4. The conditions "Monitor, Unconstrained" and "Monitor, Semi-Constrained" were the same as in the first study. Five pairs of participants were recruited to each of the Monitor Conditions and three pairs to each of the No Monitor Conditions.

	Unconstrained	Semi-Constrained
Monitor	Monitor, Unconstrained	Monitor, Semi-Constrained
No Monitor	No Monitor, Unconstrained	No Monitor, Semi-Constrained

Table 4. The design of the 2nd study.

5.2 Results

The data collection resulted in 96 dialogues, 93 of which were used in the analysis. The data were analysed using a two-way ANOVA. All effects that were found to be significant were verified by T-tests. The efficiency of interaction was determined using the following measures: time per task, number of turns, words, miscommunication-tagged turns, wrong executions and task success.

Time per task: The second column of Table 5 displays the average completion time per task in the four conditions. As expected, a main effect of the Monitor factor was found ($F=4.879$, $df=1,11$, $p<0.05$). Namely, when the user could monitor the robot's area the routes were completed faster. The interaction effect between factors was also marginally significant ($F=4.225$, $df=1,11$, $p<0.1$); pairs in the No Monitor, Semi-Constrained condition could not compensate for the lack of visual information and took longer for each task.

Number of turns and words: The aforementioned studies correlate task efficiency with number of turns and words. In terms of the mean number of turns per interaction, no significant differences were found across the groups. Nevertheless, we measured the number of words used per task and in accordance with previous research, we observed that pairs in the No Monitor conditions used more words ($F=4.602$, $df=1,11$, $p=0.05$). However, it was the wizards under the No Monitor conditions that had to be more "talkative" and descriptive ($F=10.324$, $df=1,11$, $p<0.01$). Figure 5 shows the "word-possession" rates attributed to wizards in the four conditions. Moreover, there seems to be a difference ($F=4.397$, $df=1,11$, $p=0.05$) in the mean number of words per turn. In particular, when the wizards' actions were visible to the users, the wizards required fewer words per turn. There is also an interaction effect showing more significant differences between the Monitor, Semi-Constrained condition and the No Monitor, Semi-Constrained condition ($F=5.970$, $df=1,11$, $p<0.05$); in the former, wizards managed with less than 2 words per utterance, taking full advantage of the luxury of the buttons and the fact that they were supervised. In the latter, wizards used more than 6 words per turn.

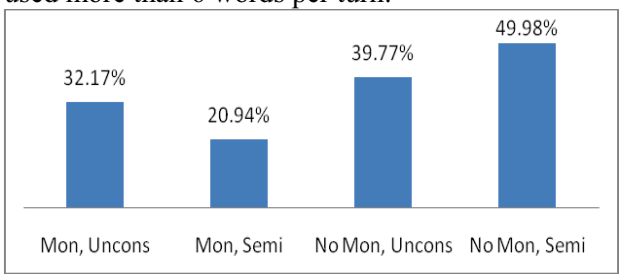

Figure 5. Words used by wizards over total.

Frequency of miscommunication: We measured the number of turns that were tagged as containing miscommunication. Surprisingly, miscommunication rates were much lower in the No Monitor conditions ($F=13.316$, $df=1,11$, $p<0.01$) and not in the conditions in which the user could check at all times the actions and understanding of the robot. The same pattern was found for user-initiated and robot-initiated miscommunication. The rates of miscommunication are included in the third column of Table 5.

Wrong executions: Analysis of number of wrong executions per task reveals a similar effect; wrong executions occurred much less frequently when the wizards were not supervised by

the users (F=6.046, df=1,11, p<0.05). They made on average 1 mistake per task, whereas the average number of wrong executions for the pairs in the Monitor conditions was 5 (fourth column in Table 5).

Task success rates: There were no differences in the number of interrupted or aborted tasks.

Condition	Time per Task (min)	Miscommunication Turns/Total Turns	#Wrong Executions per Task
Mon, Uncons	4.57	8.21%	4.2
Mon, Semi	4.63	8.82%	5.8
No Mon, Uncons	5.67	2.55%	1.0
No Mon, Semi	7.41	1.71%	0.7

Table 5. Summary of results (mean values).

5.3 Discussion and future work

These results are consistent with previous research. The conditions in which the user could see exactly what the robot saw and did resulted in faster task completion and shorter dialogues. However, a finding emerged which was not expected based on the aforementioned studies: in the conditions in which users could not monitor the robot's actions, the wizards were more accurate, leading to low occurrence of wrong executions and miscommunication (see column 3 and 4 in Table 5). The "least collaborative effort" is balanced and compromised against the need to ensure understanding. Thus, wizards provided rich and timely feedback to the users in order to compensate for the lack of visual information. This feedback acted in a proactive way and prevented miscommunication and wrong executions. In the Monitor conditions, asymmetries in perceived responsibility and knowledge between the participants could have encouraged wizards to be less cautious to act. In other words, as the user had access to the full map and the location of the wizard, the wizard felt less "obliged" to contribute to the interaction. However, due to the complex nature of the task, unless the wizard could sufficiently communicate the relevant position of the robot, the directions of the user would more likely be incorrect. It could also be assumed that since visual feedback is instant, the users were also more inclined to issue commands in a "trial and error" process. Irrespectively to the underlying motives, these findings show that despite higher costs in time and word count, linguistic resources were adequate for completing complex tasks successfully. The findings also resonate with the collaborative view of communication. The wizards adapted their behaviour in response to variations in the knowledge state of their partners and made up for the lack of visual information with rich verbal descriptions of their locations.

We are currently performing more experiments to balance the data sets of the study and validate the initial results. Moreover, a fine-grained analysis of the dialogues is under way and focuses on the linguistic content of the interactions. The aim is identical to the first WOz study, that is, to identify the strategies of the wizards in the presence and absence of visual information.

These results have important implications for HRI. As in human collaborative interaction, the robot's communicative actions have direct impact on the actions of the users. In real-world settings, there will be situations in which the users cannot monitor the robot's activities or their information and knowledge are either constrained or outdated. Robots that can dynamically determine and provide appropriate feedback could help the users avoid serious errors. Nevertheless, this is not a straightforward process; providing excessive, untimely feedback compromises the "naturalness" and efficiency of the interaction. The amount and placement of feedback should be decided upon several knowledge sources, combined in a single criterion that is adaptive within and between interactions. These issues are the object of our future work and implementation.

6 Concluding remarks

One of the most valuable but complex processes in the design of a NLI for a robot is enacting a HRI scenario to obtain naturally-occurring data which is yet generalisable and relevant for the future implementation of the system. The present study recreated a navigation scenario in which non-experienced users interacted with and taught a mobile robot. It also simulated two different setups which corresponded to the realistic situations of supervised and unsupervised interaction.
The current trend in the fields of linguistics and robotics is the unified investigation of spatial language and dialogue (Coventry et al., 2009). Exploring dialogue-based navigation of a robot, our study aimed to contribute to this body of research. It can be argued that there were limitations in the simulation as compared to the experimental testing of a real system and, thus, the study was primarily explorative. However, it yielded natural dialogues given that naive "confederates" and no dialogue script were used. The data analysis was more qualitative than quantita-

tive and followed established methods from previous research. Finally, the results of the study matched and extended these findings and provided useful information for the next version of the system as well as some insight into the processes of conversation and social psychology.

The next step in our research is to develop the dialogue manager of the robot to incorporate the feedback and miscommunication management strategies, as observed in the collected data. This holds the promise for a robust NLI that can handle uncertainties arising from language and the environment. However, miscommunication in HRI reaches beyond preventing and repairing recognition errors. Mills and Healey (2008) demonstrate that miscommunication does not inhibit but, on the contrary, it facilitates semantic coordination. Martinovsky and Traum (2003) suggest that through miscommunication, people gain awareness of the state and capabilities of each other. Miscommunication, thus, is seen as an opportunity for communication. Under this light, natural miscommunication management is not only the end, but also the means to shape and advance HRI, so that robots are not tools but partners that play a positive, practical and long-lasting role in human life.

References

Bilyana Martinovsky and David Traum. 2003.The Error Is the Clue: Breakdown in Human-Machine Interaction. *In Proceedings of the ISCA Workshop on Error Handling in Dialogue Systems.*

Dan Bohus and Alexander I. Rudnicky. 2005. Sorry, I Didn't Catch That! – An Investigation of Non-understanding Errors and Recovery Strategies. In *Proceedings of SIGdial2005*. Lisbon, Portugal.

Darren Gergle, Robert E. Kraut and Susan E. Fussell. 2004. Language Efficiency and Visual Technology: Minimizing Collaborative Effort with Visual Information. *Journal of Language and Social Psychology,* 23(4):491-517. Sage Publications, CA.

David Schlangen. 2004. Causes and Strategies for Requesting Clarification in dialogue. In *Proceedings of the 5th Workshop of the ACL SIG on Discourse and Dialogue (SIGdial04)*, Boston, USA.

Gabriel Skantze. 2005. Exploring Human Error Recovery Strategies: Implications for Spoken Dialogue Systems. *Speech Communication,* 45(3):207-359.

Graeme Hirst, Susan McRoy, Peter Heeman, Philip Edmonds, Diane Horton.1994. Repairing Conversational Misunderstandings and Nonunderstandings. Speech Communication 15:213–230.

Gregory Mills and Patrick G. T. Healey. 2008. Negotiation in Dialogue: Mechanisms of Alignment. In *Proceedings of the 8th SIGdial workshop on Discourse and Dialogue, Columbus*, OH, USA.

Gregory Mills and Patrick G. T. Healey. 2006. Clarifying Spatial Descriptions: Local and Global Effects on Semantic Co-ordination. In *Procs. of the 10th Workshop on the Semantics and Pragmatics of Dialogue.*

Herbert H. Clark. 1996. *Using Language*. Cambridge University Press, Cambridge, UK.

Herbert H. Clark and Meredyth A. Krych. 2004. Speaking While Monitoring Addressees for Understanding. *Journal of Memory and Language*, 50:62-81.

Jason D. Williams and Steve Young. 2004. Characterizing Task-Oriented Dialog Using a Simulated ASR Channel. *ICSLP.* Jeju, South Korea.

Jean Carletta, Amy Isard, Stephen Isard, Jacqueline Kowtko, Gwyneth Doherty-Sneddon and Anne H. Anderson.1996. HCRC Dialogue Structure Coding Manual(HCRC/TR-82). Human Communication Research Centre, University of Edinburgh.

Jens Allwood. 1995. An Activity based Approach to Pragmatics. Gothenburg Papers in Theoretical Linguistics, 76, Göteborg University, Sweden.

Kenny Coventry, Thora Tenbrink and John Bateman, 2009. Spatial Language and Dialogue: Navigating the Domain. In K. Coventry, T. Tenbrink, and J. Bateman (Eds.) Spatial Language and Dialogue. 1-8. Oxford University Press. Oxford, UK.

Malte Gabsdil. 2003. Clarification in Spoken Dialogue Systems. In: *Proceedings of 2003 AAAI Spring Symposium on Natural Language Generation in Spoken and Written Dialogue*, Stanford, USA.

Matthew Purver. 2006. CLARIE: Handling Clarification Requests in a Dialogue System. *Research on Language and Computation* . 4(2-3):259-288.

Stanislao Lauria, Guido Bugmann, Theocharis Kyriacou, Johan Bos and Ewan Klein. 2001. Training Personal Robots Using Natural Language Instruction. *IEEE Intelligent Systems.* 38–45.

Susan E. Brennan . 2005. How Conversation is Shaped by Visual and Spoken Evidence. In J. Trueswell & M. Tanenhaus (Eds.) *Approaches to Studying World-situated Language Use: Bridging the Language-as-product and Language-action Traditions*. 95-129. MIT Press, Cambridge, MA.

Theodora Koulouri and Stanislao Lauria. 2009. A WOz Framework for Exploring Miscommunication in HRI, In *Procs. of the AISB Symposium on New Frontiers in Human-Robot Interaction.* Edinburgh, UK.

Appendix A. Screenshot images of the interface

Figure 1. The interface of the user without the monitor (as used in the second WOz study).

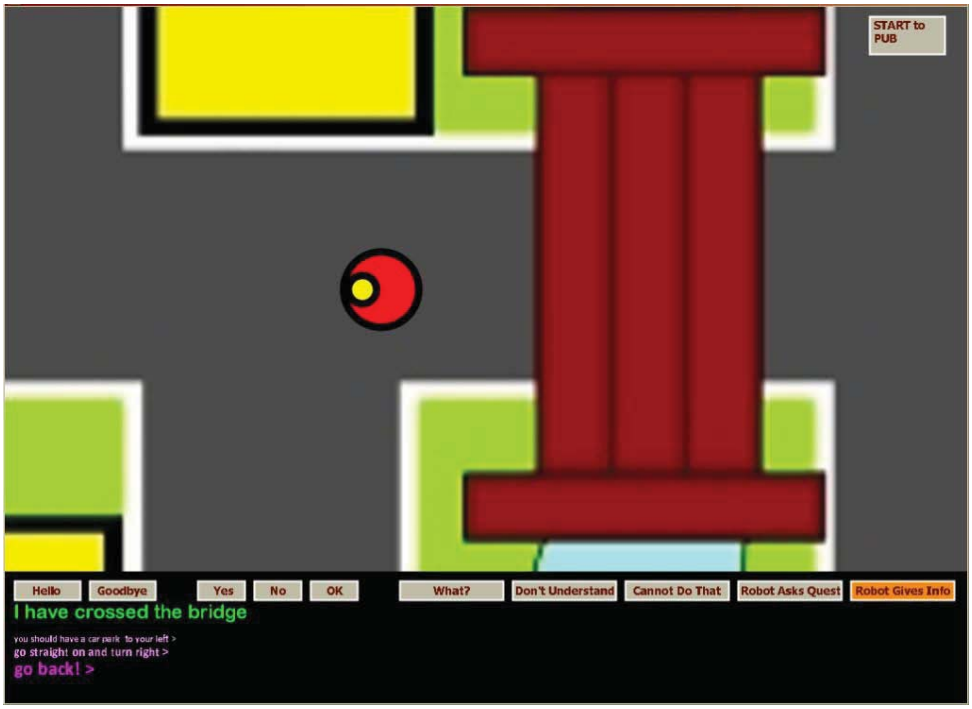

Figure 2. The interface of the wizard in the Semi-Constrained condition.