A WOz Framework for Exploring Miscommunication in HRI

Theodora Koulouri[1] **and Stasha Lauria**[1]

Abstract. The aim of this paper is to investigate management of miscommunication in spontaneous interaction between a human and a speech-enabled robot. The paper describes initial results of an exploratory WOz study, in which pairs of naive participants interacted and collaborated in a navigation task. The study is motivated by how humans achieve mutual understanding as well as previous research conducted in the area of spoken dialogue systems. The dialogue and error handling capacity of the wizard is incrementally impaired towards the capabilities of a system in three experimental conditions. Preliminary analysis of the data reveals the necessity to endow the robot with richer error management resources and point to the efficacy of less explicit error handling strategies. Analysis of the data and further experimentation are currently performed and expected soon to shed light to some of the intricacies and unique characteristics of HRI.

1 INTRODUCTION

1.1 Personal Robotics

Recent progress in robotic technologies brings closer the vision of ubiquitous and commercial use of robotic systems. This poses new challenges for the vibrant and young field of Human Robot Interaction (HRI). Personal robots possess the highest expected rate of growth [1] and as non-expert user groups are targetted, more natural and rich interfaces are on demand. Thus, there are numerous prototypes equipped with natural language interfaces (NLI) and promising results have been generally reported [1]. Nevertheless, there remain many challenges which pertain to the development of a NLI between a human and a robot.

1.2 Natural (mis)Communication in HRI

Natural language is infinitely novel and ambiguous. Moreover, language in use, i.e., language that emerges in dialogic settings, is never perfect. In fact, error-free communication is now held to be an ideal rather what happens in everyday human interaction. Yet, human interlocutors manage miscommunication so efficiently that it rarely becomes an explicit focus in the dialogue. Additionally, the performance of state-of-the-art NLP technologies still leaves much to be desired. As a consequence, problems in understanding, uncertainty and out-of-grammar words will always occur in human-machine interaction. Most current spoken dialogue systems (SDS) produce prompts such as "Please repeat" and "I do not understand" to signal and resolve problems. Such strategies are, nevertheless, insufficient to handle all kinds of miscommunication, which has ultimately a great impact on performance and user experience. This issue has first been identified in research in the area of spoken dialogue systems, which has mostly dealt with information-seeking, telephone-based applications. But it is exacerbated in the area of personal robotics in which users are naive not only about the linguistic but also the functional competence of the robot [2]. It is evident that errors in understanding and execution give rise to many safety concerns which is not usually the case in other domains. Moreover, it has been argued [1] that language-endowed robots face higher expectations by people, who attribute human-like linguistic capabilities and intelligence to them. Thus, as the occurrence, scope and forms of miscommunication increase, the need for an approach that enriches the robot with a greater repertoire for dealing with problematic understanding also increases.

The aim of our research is to develop a natural framework for handling miscommunication in HRI. As people are extremely apt in preventing and repairing problematic understanding, the approach in this study is to explore and build on the principles of human error handling. This paper describes the first steps in identifying consistent linguistic behaviour when human users perceive communication failure within the context of HRI.

2 BACKGROUND AND MOTIVATION

2.1 Past Work

The current work extends previous research by the universities of Plymouth and Edinburgh [3]. The project explored Instruction Based Learning (IBL) and aimed to enable naïve human users to instruct a personal robot to perform a navigation task by means of natural language. The robot is equipped with a built-in knowledge of some basic navigation actions. On encountering a novel route description, the robot engages in a dialogue with the user who explains and decomposes the novel route into known actions. Subsequently, the route is incorporated to the robot's knowledge base for future use. This enables the robot to learn and execute increasingly complex tasks. This work demonstrated that IBL is a viable architecture for developing personal robots with capabilities of learning. However, evaluation of the system revealed the complexity of adding even simple interactive mechanisms to the robot [4].

2.2 Models of Communication

[1]Dept. of Information Systems and Computing, Univ. of Brunel, UB8 3PH, UK. Email: theodora.koulouri@brunel.ac.uk

Empirical studies of human-human interaction (HHI) describe communication as a collaborative process in which participants are continuously establishing that what has been said has also been understood. During interaction, interlocutors can be in any of these states [5]:

Table 1. States of Understanding. B, A and u stand for hearer, speaker and utterance, respectively.

State 0	B didn't notice that A uttered any u
State 1	B noticed that A uttered some u (but wasn't in state 2)
State 2	B correctly heard u (but wasn't in state 3)
State 3	B understood what A meant by u

Communication is achieved if interlocutors are in state 3. According to the model, listeners choose repair initiations that assert their current state of understanding. For example, "What?" is an assertion of being in state 1 and "Which building?", a presupposition of state 2.

Along the same lines, [6] and [7] propose a four-level model of communication which has been unified by [8] and is shown below. Miscommunication can occur in any of these levels:

Table 2. Levels of Communication (adapted from [8]).

Level 1	Securing Attention
Level 2	Utterance Recognition
Level 3	Meaning Recognition
Level 4	Action Recognition

2.3 Research in Miscommunication

2.3.1 Sources of miscommunication

Miscommunication covers three categories of problems in interaction; misunderstandings, non-understandings and misconceptions [9]. Misunderstandings occur when the hearer obtains an interpretation which is not aligned to what the speaker intended him/her to obtain and may not be readily detected. This category generally leads to direct corrections by the speaker. Non-understandings occur when the hearer obtains no interpretation at all or too many. This category could also include cases in which the hearer is uncertain about the interpretation which he/she obtained. Non-understandings are of special interest to our study as this type of problem triggers repair initiations (e.g., clarification requests) by the hearer. Last, misconceptions occur when the interlocutors' beliefs of the world clash.

2.3.2 Clarification requests

Clarification requests can be broadly defined as the dialog acts employed by the hearer to signal a problem in understanding. Most taxonomies of clarification requests [10, 11, 12] are motivated by the models of HHI discussed in Section 2.2. Namely, it is maintained that the choice of clarification request indicates the highest level of understanding currently available to the listener and signals the information required for understanding to eventually occur. Thus, there is the expectation that speakers modify their original utterance as a response to a clarification request.

Furthermore, "generic" repair utterances such as "What?", "Pardon?" and "Please repeat" (that is, the default error handling strategies of a SDS) are less informative as regards to the source of the problem compared to reprises (such as "Go where?"). According to [12], the latter type of clarification requests accepts a part of the problematic element and is less severe for the dialogue. A study in HHI [8] compared "What's" and reprise fragments and found that the former caused more dialogue disruption and impaired the coordination between participants.

2.3.3 Error handling strategies in SDS

In the area of SDS, there have been considerable efforts to build systems with more intelligent handling of errors. In [13], the problematic utterance was classified in one of twelve classes of errors and targetted help was provided by the system. In [14], the system gave feedback with the recognition hypothesis, a diagnosis of the problem and similar in-coverage examples of what to say. Both studies report more successful interactions. In [15], an extensive research in error recovery strategies from non-understandings is presented. They defined a set of error handling strategies and discovered a relation between them and recovery from errors. They provided evidence for benefit in performance by incorporating a smart policy of selecting strategies.

In previous studies [16, 17, 18], variations of Wizard of Oz (WOz) simulations were performed in order to discover how humans handle automatic speech recognition (ASR) errors. Zollo [16] lists a number of negative and positive feedback strategies such as reprise sluices and fragments, confirmation requests, simple acknowledgements etc. Moreover, Skantze [17] found that explicitly signalling non-understanding is not the most common strategy used by humans to handle errors, but providing task-related information results in more successful interactions. Findings from [15] and [18] for different domains seem to resonate with these conclusions. The aforementioned studies suggest that dialogue systems that make miscommunication another contribution circle and not a focus of the dialogue will enjoy higher success rates in terms of performance and user experience.

There have been several successful implementations of prototype SDS that incorporate clarification requests [11, 12] and more sophisticated error handling. However, it remains an open question how these features would operate in a robot which is perceptually grounded to its environment.

2.4 Challenges in HRI

Research in error handling focuses on uncertainties arising from poor speech recognition, which is justified by the fact that it is the source of the large majority of errors. However, a significant number of errors are attributed to out-of-grammar expressions and requests beyond the functionality of the system [15].

In the interaction with a personal robot acting in the same dynamic environment as its user, one should expect a different distribution of errors and possibly new sources of errors. To the knowledge of the authors there has not been an analysis of sources of errors in speech-enabled robots and how each of these affects the global performance of the robot.

In this section the unique characteristics of HRI that make NLI an even more challenging endeavour are outlined. First, the domain of interaction is wider than of SDS. Moreover, as HRI is a new field, users are more naive about what the robot can understand and do. It has also been suggested that they would expect human-like abilities [1]. As a consequence, misunderstandings and non-understandings are more likely to occur. Expanding the grammar by collecting larger corpora is a difficult and resource-demanding task with disproportionate outcome. Therefore, the robot should be capable of informing users of its competences in a natural manner [2]. In the face of problematic understanding, the robot should be able to act in a way that ensures the safety of the user but also the smoothness of the interaction.

On the other hand, models of HHI predict that co-presence will increase the perception of mutual knowledge [19] and users are more likely to refer to objects and entities in a way that a robot without human-like vision capacities is unable to resolve. Moreover, real physical environments are dynamic, even more so when the robot is mobile. Therefore, referential resolution obtains an elevated status in the dialogue manager of the robot, which is now required to be able to communicate and negotiate changes promptly and effectively.

2.5 Methods of Data Collection

It could be argued that since natural interaction is the end, data from human dialogues should be the starting point. However, it is well established among linguists that speakers adapt their linguistic behaviour according to the perceived characteristics of the hearer. For instance, we do not talk to a five-year old child in the same way we talk to a colleague. This might explain the differences in interaction patterns observed in HHI and HCI [20]. Moreover, communication in conversational settings comes with an abundance of shared assumptions and knowledge which are neither transparent nor relevant to the researcher. Another solution could be to use data from interactions with current spoken dialogue systems. This, however, seems to be of little utility as our aim is to build the systems of the future.

In order to collect data from a variety of interactions and phenomena as well as test functionality not yet implemented, NLI developers set up WOz experiments. In these experiments, a human operator ("the wizard") emulates the system (or parts of it) and interacts with a user who is under the impression that he/she is talking to a machine [20].

3 METHOD

This study aims to identify the strategies that humans use when their communicative ability and the information available to them are restricted in a way similar to artificial systems. To serve our purposes, we devised a series of WOz experiments which deviates in certain respects from the typical WOz methodology. Previous research that employed this method [15, 16, 17, 18] served as the motivation and basis for our design. WOz simulations have also been conducted in the area of HRI [3, 21]. However, to the authors' knowledge, none of them explored miscommunication and error handling.

In a WOz experiment the wizard is the trained experimenter. However, this seems to introduce a bias as the experimenter controls the interaction. As the object of this study is the wizard's dialogue actions, the experiments involved *two* naive subjects, that is to say, both the user and wizard are naive. A justification of this choice comes from [17] who maintains that "this experimental setting lacks the control that the consistent behaviour of a trained operator would give. Still, this method may be good for explorative studies, which aim at finding new ideas on dialogue behaviour, and especially on how error situations could be handled." The study explored recovery strategies from ASR errors. However, both participants were fully informed, hence, there was no wizard involved.

The experimental design described here is also largely motivated by the WOz paradigm for dialogue systems proposed by Levin and Passonneau [22] which draws on two methods used in AI, namely, ablation and comparison. In their design, the communicative resources of the wizard are incrementally restricted. The different conditions are compared in order to isolate the properties of the dialogue system that most affect the overall performance. Their paper also exemplifies the application of the method for exploring error handling strategies.

Our study involves three experimental conditions for data collection. In particular,

- The wizard simulates a super-intelligent robot and interacts with the user using unconstrained natural language (henceforth, referred to as Condition 1).
- The wizard selects from a list of utterances that point to the source of the problem (in the utterance, meaning or action level, see Section 2.2) but can also type in a clarification question or provide task-related information (Condition 2).
- The wizard is fully restricted to use the same limited set of utterances as a typical dialogue system. In case of problematic understanding, the wizard signals that there is some kind of problem (Condition 3).

3.1 Domain of the Experiment

The domain of the experiment is navigation on a miniature town which is similar to the Map Task [23]. As Brown [23] points out, in the Map Task, each subject has two overt sources of information: what the other speaker says and what is on his/her map. Thus, the participants are given the opportunity to interact with each other in a relatively natural manner, while controlling the information available to them at any given point in the dialogue. The analyst has access to the records of what each participant does and says, which allows for a degree of understanding of how people are tackling the task or the problems that arise, and where these arise. Yet, even in task-oriented interaction, repair instances are not frequent.

In a small-scale study, like the one described in this paper, the scarcity of such data would have prevented us from making any reliable inferences. Thus, the interface between the participants had to be further degraded (as explained below). In our study, the user guided the robot to six destinations. The user had full access to the map whereas the wizard could only see a small fraction of the map of the area surrounding the robot, so he/she had to rely on the user's instructions on how to go to a location. In addition, the user could not see the robot itself, but only its surroundings. Therefore, both participants needed to collaborate and exchange information in order to complete the task. In this sense, the task falls under the category of problem-solving tasks,

as opposed to information-seeking tasks commonly employed in the development of SDS. It is interesting to discover to what extent knowledge gained from error handling in the domain of information-seeking dialogue systems can be applied to a different domain of interaction with a robot.

3.2 Set-up

A custom java–based system[2] was developed for the experiments consisting of the map and the messaging box. The wizard's interface was modified according to each experimental condition but the user's interface remained the same. The two applications were linked together using the TCP/IP protocol, sending and receiving coordinates and messages over a LAN. However, there are no real constraints of location as the computers can also be connected via the Internet. Moreover, the system kept a log of the interaction but also, for every message exchanged, the coordinates of the robot at that given moment were recorded. This enabled us to monitor the wizard's level of understanding by comparing each instruction with how it was actually executed.

For all conditions, the wizard's map contained only a small fraction of the map with the area that surrounded the current position of the robot. All buildings were shown as yellow squares. The robot was displayed as a red circle with a yellow "face" and was operated by the wizard by pressing the arrow keys on the keyboard. The dialogue box was on the lower part of the screen. The messages from the wizard were shown on the upper part of the box (in green) and the user's messages on the bottom as well as a history of the user's previous messages. As explained in 2.1 above, the actual robot has the ability to learn and remember previous routes. To simulate this ability, once a route was completed, in the next task, a new button appeared on the right side of the wizard's screen that represented the newly learnt route. When the user requested to take a known route, the wizard clicked on the corresponding button and the robot automatically executed the route.

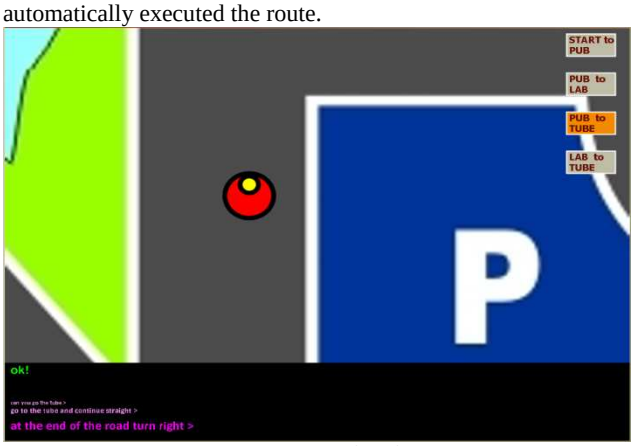

Figure 1. The Wizard's Interface for Condition 1.

In the interface version for Condition 1, the wizard could freely type messages and send them to the user like in a typical messaging application. Figure 1 displays a screenshot of the interface. Note that that the robot has already completed and "learnt" four routes.

[2] http://processing.org

In the version for Condition 3 (Figure 2), the wizards were deprived of typing their own messages. There was a set of buttons on their messaging box which the wizard clicked and the corresponding canned response was sent to the user. These responses were "hello", "goodbye" and "ok". In addition, there was a button denoted with a "?". The wizard was instructed to use this button to inform the user that there was some kind of problem in understanding or executing. This button randomly generated any of the following responses: "Can you please repeat that?", "Sorry, I don't understand", "What?" as well as a response that contained a fragment of the user's previous message (e.g., "Turn where?"). The wizard had no control over or any a priori knowledge of which response would be sent.

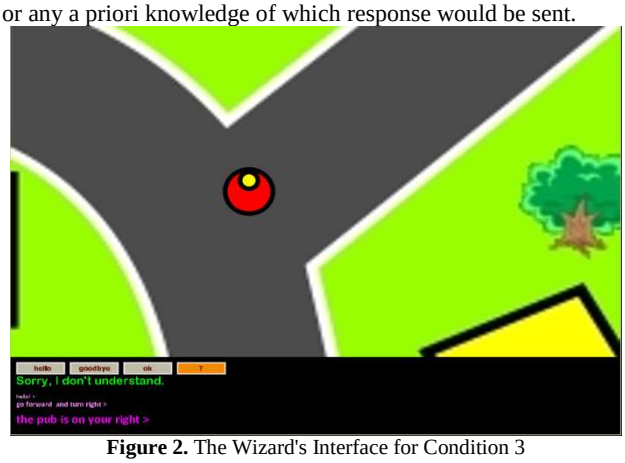

Figure 2. The Wizard's Interface for Condition 3

The interface version for Condition 2 (Figure 3) allowed for two ways of interaction with the user. First, the canned response-based interaction; in particular, the wizard could click on "Hello", "Goodbye", "Yes", "No", "Ok" and the problem-signalling buttons "What?", "I don't understand" and "I cannot do that"[3] to automatically send these messages to the user. The second way was to click on the "Robot Asks Question" and "Robot Gives Info" buttons so that the wizard could type in his/her own messages; the wizard was instructed to use the former to request clarification and the latter to provide the user with information.

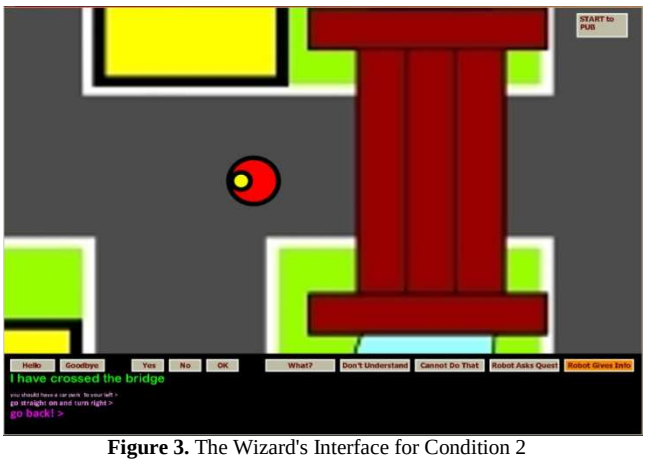

Figure 3. The Wizard's Interface for Condition 2

[3] These responses signal problems in the utterance, meaning and action level, respectively. See section 2.2, Table 2.

The user had access to the full map of the town (Figure 4), in which the destination was shown in red and the locations that had been reached in previous tasks in blue. On the upper right corner of the screen, the user could see a small fraction of the map that showed the surrounding area of the current position of the robot. Thus, the user could see the surrounding area changing while the robot was moving, but not the robot itself. The dialogue box was on the lower part of the screen. The user could type in messages and send them to the wizard in a manner similar to a messaging application. The messages from the robot were displayed at the bottom of the textbox (in purple). The user's interface did not show the history of the dialogue in order to simulate the transience of spoken language.

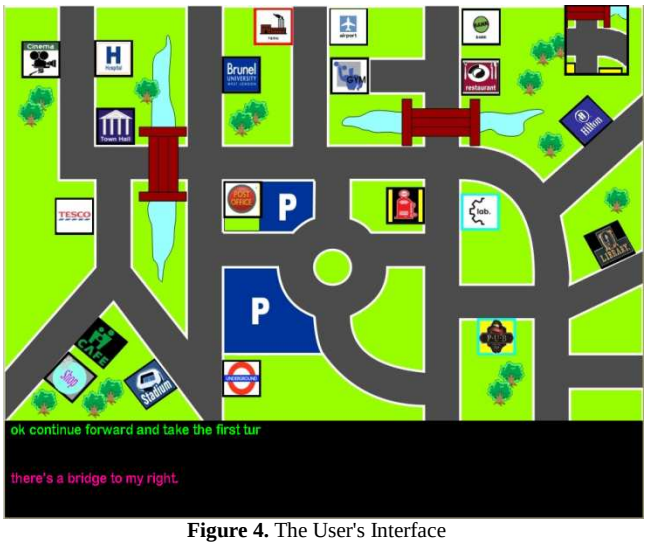

Figure 4. The User's Interface

3. 4 Procedure

Eighteen paid participants were recruited, nine users and nine wizards. Computer expertise was not required. For each experimental condition, three pairs were used. Participant pairs were randomly assigned to one of the three conditions, either as wizards or users. The pairs were seated in separate rooms in front of a desktop PC. The user was sitting alone during the experiment whereas the experimenter was in the same room as the wizard in case they needed technical assistance. The wizards received instructions, a short demonstration and had a trial period to familiarise with the functionality of the system.

The users were briefed about the experiment and the interface. Unlike the wizards, who were aware about the simulation, the users were told that they would interact with a real robot through a computer interface. They were asked to start each task whenever they felt ready by clicking on the links on their computer screen. They were also advised to start each interaction with "Hello", which automatically opened the wizard's application, and end the interaction with "Goodbye" which closed both user and wizard's windows. The user could terminate the interaction at any point by typing "abort task". After each task (completed or aborted), they had to fill in a short questionnaire. The questionnaire consisted of five Likert-scale statements for which the users stated their level of agreement.

Each pair attempted six tasks; specifically, the user guided the robot from a starting point to six designated locations. The destinations were selected in an effort to increase the number or complexity of instructions and the possibility of reference to known routes. Dialogues were allowed to run until the user ended them or up to 10-11 minutes (this cut-off time was decided on the basis of pilot studies).

Both participants received a written description of the experiment which they read before the beginning of the experiment and could consult throughout it. They were not given any instructions on what to say. However, they were explicitly instructed not to use directions such as "north", "south", "up" and "down", but use relative directions from the robot's point of view and common landmarks such as the bridges, the car parks and known locations. Additionally, the users were told that the robot was very fluent in understanding natural language and producing appropriate responses. They were also informed about the route-learning ability of the robot. Last, for the conditions in which the wizards were able to type their own messages, they were asked not to try to "sound like a robot" but talk naturally.

In designing and executing the experiment, a set of parameters as described in [24] were considered: first, the interface which was built for the experiment fully simulated the existing system and allowed for a mixed initiative interaction. Thus, the insights gained from the experiments are relevant to this domain and application while remaining generalisable. Second, the task was open and the focus was not its fast completion, that is, the user could plan or later modify the route in any way. This allowed for dialogue variation. Finally, pilot studies were conducted to ensure that the simulation environment operated by the naive wizard was usable.

3. 5 Experimental Hypotheses

The WOz simulations will be completed in two phases. Phase 1 is described in this paper. Phase 1 aims to generate a refined hypothesis and guide the experiments of the second phase.

The existing system, the robot, (see section 2.1, [3] and [4]), as with the majority of SDS, deals with all sorts of miscommunication in a formulaic way. On the opposite end, human interlocutors have a powerful repertoire of error management skills. Thus, we maintain that a system with more sophisticated error handling capabilities could improve performance. Moreover, previous research has shown that explicitly signalling non-understanding (e.g., "I don't understand" and "Please repeat") is not the default strategy used by people, but implicit strategies (such as clarification requests and providing task-related information) are typically opted for.

The experimental hypotheses are the following:

1. Condition 3, in which the wizard handles miscommunication in an "uninformed", prescribed way, will result in lower success rates and user satisfaction than the "informed" wizard condition (Condition 2).
2. In experimental Conditions 1 and 2, wizards will not resort to explicit signalling of non-understanding.

We also aim to observe several interesting phenomena that are specific to HRI and are not found in HHI and interaction with SDS. These insights, still sparse in literature, will help us develop a system that meets the needs and desires of its user.

4 RESULTS

This section presents an elementary analysis of the data obtained. A total of 54 dialogues were collected and analysed in terms of task success, task completion time, miscommunication, wrong executions and user perceptions.

4.1 Task Success

In Condition 3, one user aborted a task and another pair of participants exceeded the 10-minute time limit and the task was interrupted. In Condition 2, all tasks were completed within ten minutes (see third column of Table 3). This could suggest that in the restricted wizard condition, the participants were unable to recover from errors as effectively as participants in the other conditions. Nevertheless, given the size of the sample, further experiments are needed to validate this claim.

Table 3. Summary of Results from All Three Conditions.

Condit-ion	Ave-rage Time per Task (min)	Task Complet-ion Rate	Miscom-munication Turns/ Total Turns	Total Number of Wrong Executio ns	% Resolved Wrong Executions	Total Number of No Executio ns
1	4.58	94.44%	11.29%	10	72.22%	0
2	4.25	100.00%	11.13%	32	77.40%	0
3	5.52	88.89%	12.89%	10	68.75%	56

4.2 Task Completion Time

As shown in the second column of Table 3, Condition 2 resulted in faster interactions compared to the other conditions. The difference between conditions 2 and 3 is greater than the difference between conditions 1 and 2. This might indicate that simplistic signalling of non-understanding leads to longer interactions. Aborted and interrupted tasks were not considered in the analysis.

4.3 Miscommunication and Wrong Executions

Based on the definitions given in Section 2.3, dialogue turns that expressed non-understanding or contained user corrections were labelled as miscommunication. The data in column 4 of Table 3 reveal a high occurrence of such problems across all conditions. In Condition 3, the number of miscommunication turns is marginally larger.

As explained in Section 3.2, the path that the robot followed could be reproduced and each instruction and the corresponding robot action were juxtaposed. If the execution did not match the instruction, it was tagged as wrong execution. As shown in the fifth column of Table 3, the number of wrong executions was significantly higher in Condition 2. However, recovery rates from wrong executions were similar across the first two conditions, but slightly lower in Condition 3. This could imply that misunderstandings that led to wrong executions were more easily resolved in Condition 2 than in Condition 3. It should be also noted that although wrong executions in Conditions 1 and 3 were equally frequent, in Condition 3, 56 "no executions" were tagged (compared to none in Conditions 1 and 2). "No execution" tags were placed in turns in which, although the instructions had been received, there was no reaction (movement) whatsoever by the wizard. It could be speculated that the number of wrong executions could have been higher in Condition 3. But more significantly, this observation supports the idea that the wizards perceived the inadequacy of their expressive means to proactively or reactively deal with problematic understanding. On the other hand, wizards in Condition 1 and 2 felt more confident that they could prevent or resolve wrong executions by signalling non-understanding, requesting clarifications etc. and, thus, were more willing to act based on their assumptions.

4.4 User Perceived Task Success and Overall Satisfaction

After each interaction, the users completed a six-point Likert-scale questionnaire in which they rated their agreement with five statements. These statements covered ease of use, accuracy and helpfulness of the system, perceived task completion ("I think I did well in completing the task") and overall satisfaction ("I am generally satisfied with this interaction"). The responses were mapped to integer values ranging from 1 to 6 (with 6 representing the optimal score). The average scores for perceived task success and user satisfaction were calculated and plotted against each condition (see Figure 6). It seems that, in Condition 2, despite the similar frequency of miscommunication and higher wrong execution rate, the users experienced the dialogue as smoother and more successful.

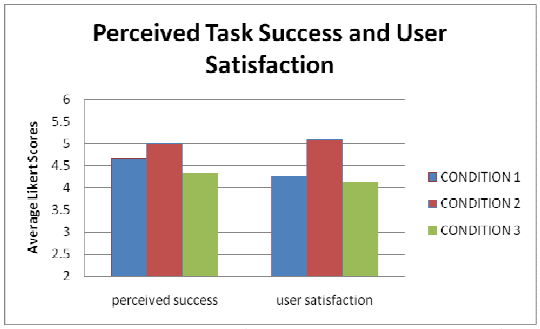

Figure 5. User's Perception of Task Success and User Satisfaction

4.5 Further Observations

Initial turn-based analysis of the dialogues produced several interesting observations. First, in Condition 1, in which wizards were free to formulate their own responses, there was not a single occurrence of explicit problem-signalling responses such as "I do not understand" or "What do you mean?". Even less "talkative" wizards preferred to request further instructions than to state that there was a problem. Explicit strategies should have been the easiest "way out", especially in Condition 2 in which the wizards could simply press a button. However, the "least effort principle" is always balanced and compromised against the need to ensure understanding [5]. Thus, wizards felt that in certain circumstances of interaction, explicit strategies were inadequate. Second, an interesting phenomenon was observed which was not anticipated; wizards in Condition 2 used a

balanced combination of explicit and implicit strategies. A sample of a dialogue under Condition 2 is shown in Table 4. In this example, the wizard explicitly signals non-understanding, using the button "I don't understand" (line 2), but then she types a clarification request termed task-level reformulation[4] (line 4).

It can be argued that the physical presence of the buttons in the wizards' interface biased them to use the predefined responses. However, it is equally plausible to assume that these strategies matched exactly what they intended to say. Within the context of the results discussed in Sections 4.1-4.4, this has important implications for the application of the findings of this study to a real-world system. In particular, unconstrained natural communication with a robot, as represented by Condition 1, is infeasible to implement with current technology. However, a system with a smart policy for selecting from a predefined set of explicit and implicit responses, which yields similar benefits in performance and user experience, is a solution within our reach.

Table 4. An Excerpt of a Dialogue in Condition 2.

1.	user said- turn right after the bridge then
2.	robot said- Sorry, I don't understand.
3.	user said- turn right now
4.	robot said - at a car park?
5.	user said- before the car park

Table 5. An Excerpt of a Dialogue in Condition 1.

1.	user said- please turn right and turn right again
2.	robot said- Crossroads, now where?
3.	user said- turn right
4.	user said- keep going right
5.	robot said- I cannot go down that road any more
6.	user said- go forward
7.	robot said- There is a fork in the road
8.	user said- ok turn left here then take the third right
9.	robot said- Before the bridge?
10.	user said- sorry this is the wrong place
11.	robot said- Wrong place? Should Robot go back?
12.	user said- go forward and at the crossroads keep going forward and the tube is at the end of the road
13.	robot said- Out of bounds

Table 5 contains part of a dialogue in Condition 1. In any given situation, the wizard did not explicitly signal the problem but either provided information (lines 2, 5, 7, 13) or used clarification requests (lines 9, 11). In this example, we have the opportunity to observe the balance in the interaction and roles of the interlocutors as they try to cope with and compensate for insufficient sources of information. Management of miscommunication, thus, is a joint activity and not the responsibility of one speaker. The issue of repair rarely becomes the focus of the interaction; the goal of the interaction, that is, obtaining adequate information to complete the task, was not sidetracked but also became a means of repair itself.

Another interesting phenomenon was encountered in Condition 3; as the dialogue progressed, the users tended to pay less attention to the wizard's messages. Instead, they focused on the movement of the robot as they could see it from the little window on their interfaces, and continued issuing commands in a "trial and error" process. This could suggest that the users felt that the robot was no longer helpful or reliable in terms of interaction. This observation is also supported by the post-task questionnaires.

Table 6. An Excerpt of a Dialogue in Condition 3.

1.	user said- go to the lab
2.	user said- do u know where the lab is?
3.	user said- go straight
4.	user said- go left
5.	user said- stop
6.	user said- turn right
7.	robot said- Can you please repeat that?
8.	user said- go straight
9.	user said- go right
10.	user said- move
11.	user said- go straight
12.	user said- go right
13.	robot said- Sorry, I don't understand.
14.	user said- right

Table 6 shows an excerpt of a dialogue under Condition 3. In line 2, the user asks a question but does not wait for a reply. Similarly, the problem-signalling messages of the robot (lines 7 and 13) seem to go unnoticed. The dialogue looks disconnected and the user gives commands and waits to see the action on the screen. Evidently, in real deployment environments such conduct has different consequences.

In conclusion, the numerical results presented in Sections 4.1-4.4 seem to provide support for the first experimental hypothesis. The data discussed in 4.5 only partially confirmed the second hypothesis but revealed a path worth pursuing. Quantitative analysis of the turn by turn unfolding of the dialogue is currently being carried out, using a dialog act tagging scheme. Results are expected soon to illuminate patterns of behaviour regarding wizard strategies and subsequent user responses to address what they perceived to be problematic.

5 DISCUSSION AND FUTURE WORK

The preliminary results reported in this paper are consistent with previous research (see Section 2.3.3) and point towards fascinating research directions. As part of our current and future work, we are running more experiments in order to ascertain that these initial findings are statistically sound and not opportunistic. Moreover, we are performing a fine-grained analysis of the dialogues which involves turn-based annotation of the data with dialogue act tags. The object of the analysis is the actions employed by wizards and users to maintain and restore understanding. These findings will feed the second cycle of simulations and are anticipated to offer additional insights in miscommunication management in HRI. Further, we will look at the relation between wizard strategies, user responses, error recovery, overall dialogue performance, task success and user perceptions.

In [17] and [18], the wizard had no direct access to the user's speech, but could listen to or read the output of a simulated or real speech recogniser. In our simulations we allowed the wizard to have full access to the messages sent by the user without a mid-component. The experimental setup was designed to provide a complex environment that would give rise to many instances of miscommunication (see Table 3) and

[4] In [10], Gabsdil defines task-level reformulations as clarification requests that reformulate the previous speaker's utterance in terms of its effects and the task rather than its surface form.

difficulties in coordination through high referential ambiguity and deictic discrepancies. At this stage in the research, it was decided that ASR would unnecessarily add to the task complexity and negatively affect the consistency and validity of the results; first, the role of the wizard which was assumed by a non-expert participant was already demanding as they had to make decisions and respond fast and accurately with limited and ambiguous sources of information while operating the robot. Dealing with the often incomprehensible output of the speech recogniser would have rendered their task impossible. Secondly, ASR performance is a major source of problems with detrimental effects on the overall system performance. This could confound the effect of the other sources of problems and the error handling strategies, which form the aim of this study, and obscure the observation of other interesting phenomena. The proposed framework will be ultimately implemented and tested using a fully operating system. However, time and resources permitting, we aim to design another experimental condition in which the same set of error handling strategies, as defined by the second round of experiments, is used by the wizard whose abilities are further constrained by ASR.

A valid argument against the experimental setup could be that the results are an artifact of or only relevant to text-based interaction. However, in [11] it is argued that text-based synchronous interaction is now a commonplace means of communication between people (e.g., instant messaging and chat rooms). Moreover, spoken dialogue contains many other sources of information that play a role in participants' understanding. These are extra-linguistic features such as variations in voice amplitude, pitch and speed which function as cues for the listener. When speech is transcribed for analysis these are very hard to represent and are often ignored. On the other hand, text-based interaction constrains users to convey meaning using only linguistic means, which minimises the analyst's manipulation of the data and provides clearer indications of how understanding is achieved. Nevertheless, experiments using real-time speech and audio are planned after these exploratory studies are terminated.

The Map Task offers a rich domain for task-oriented interaction [23]. In the Map Task design, there is one A-role speaker who holds all information necessary for the completion of the task. In the setup of our study, we attempted to raise the status of the wizard; namely, the wizard needed to successfully communicate the surroundings and the exact position of the robot, otherwise the route instructions from the user that were mainly formed with deictic expressions were meaningless. Similar to normal HHI and human cooperative behaviour, in Conditions 1 and 2, participants collaborated and helpfully shared relevant knowledge. Breazeal et al. [25] state that the goal of the field of human-robot interaction is broader than interaction; rather it should be pursued as human-robot collaboration. Thus, the insights and findings that are beginning to emerge from these experiments could contribute in closing the gap between HHI and HRI, so that robots are not tools but partners that play a positive, practical and lasting role in human life.

ACKNOWLEDGEMENTS

The authors would like to thank the anonymous reviewers for their valuable comments and insights.

REFERENCES

[1] S. Thrun, *Toward a Framework for Human-Robot Interaction, Human-Computer Interaction*, 19: 9-24 (2004).
[2] G. Bugmann, Effective Spoken Interfaces to Service Robots: Open Problems, In: *Procs. AISB 2005*, Hatfield, UK, pp. 18–22 (2005).
[3] S. Lauria, G. Bugmann, T. Kyriacou, J. Bos and E. Klein, Training Personal Robots Using Natural Language Instruction. *IEEE Intelligent Systems*, pp. 38–45 (2001).
[4] G.Bugmann, Final Report to the EPSRC http://www.tech.plym.ac.uk/soc/staff/guidbugm/ibl/IGR-IBL.pdf (2003).
[5] H.H. Clark and E. Schaefer, E. Contributing to Discourse. *Cognitive Science*, 13:259–294 (1989).
[6] H.H. Clark, Using Language. Cambridge University Press, Cambridge (1996).
[7] J. Allwood, An Activity Based Approach to Pragmatics. In: *Abduction, Belief and Context in Dialogue, Studies in Computational Pragmatics*, Amsterdam. John Benjamins (1995).
[8] G. Mills and P. G. T. Healey, Clarifying Spatial Descriptions: Local and Global Effects on Semantic Co-ordination. In: *Procs. 10thWorkshop on the Semantics and Pragmatics of Dialogue* (2006).
[9] G. Hirst, S. McRoy, P. Heeman, P. Edmonds, and D. Horton, Repairing Conversational Misunderstandings and Non-Understandings. *Speech Communication*, 15:213–230 (1994).
[10] M. Gabsdil, Clarification in Spoken Dialogue Systems. In: *Procs. 2003 AAAI Spring Symposium Workshop on Natural Language Generation in Spoken and Written Dialogue*, Stanford, USA. (2003).
[11] M. Purver, P. G. T. Healey, J. King and G. Mills, Answering Clarification Questions. In: *Procs. SIGDIAL*, Sapporo, Japan (2003).
[12] D. Schlangen, D. Causes and Strategies for Requesting Clarification in Dialogue. In: *Procs. SIGDIAL*, Boston, MA (2004).
[13] G. Gorrell, I. Lewin, and M. Rayner. Adding intelligent help to mixed-initiative spoken dialogue systems. In: *Procs. 7th ICSLP*, Denver, CO (2002).
[14] B. A. Hockey, G. Aist, J. Dowding, and J. Hieronymus. Targeted help and dialogue about plans. In: *Procs. 40th Annual Meeting of the ACL*, Philadelphia, PA (2002).
[15] D. Bohus and A.I. Rudnicky, Sorry, I Didn't Catch That! - An Investigation of Non-understanding Errors and Recovery Strategies, In: *Procs. SIGDIAL*. Lisbon, Portugal, (2005).
[16] T. Zollo, A Study of Human Repair Initiation Strategies in the Presence of Speech Recognition Errors, In: *Working Notes of the AAAI Fall Symposium on Psychological Models of Communication in Collaborative Systems*, (1999).
[17] G. Skantze. Exploring Human Error Recovery Strategies: Implications for Spoken Dialogue Systems. *Speech Communication*, 45(3):207-359 (2005).
[18] J. D. Williams and S. Young. Characterizing Task-Oriented Dialog Using a simulated ASR Channel. *ICSLP*. Jeju, South Korea (2004).
[19] H.H. Clark and C. Marshall, Definite Reference and Mutual Knowledge. In: *Elements of Discourse Processing*. B.L. Webber, A.K. Joshi, and I.A. Sag (Eds.) Cambridge University Press, New York (1995).
[20] N. M. Fraser and G.N. Gilbert, Simulating Speech Systems. *Computer Speech and Language*, 5: 81–99 (1991).
[21] S. A. Green, S. M. Richardson, R. J. Stiles, Billinghurst, and J. G. Chase, Multimodal Metric Study for Human-Robot Collaboration. In: *Procs. ACHI 2008.* Washington, DC (2008).
[22] E. Levin and R. Passonneau, A WOz Variant with Contrastive Conditions, In: *Procs. Dialog-on-Dialog Workshop, ICSLP*. Pittsburg, PA (2006).
[23] G. Brown, Speakers, Listeners and Communication. Cambridge: Cambridge University Press, (1995).
[24] N. Dahlback, A. Jonsson and L. Ahrenberg, Wizard of Oz Studies-Why and How, In: *Readings in Intelligent User Interfaces*, M. Maybury and W.Wahlster (Eds.), Morgan Kaufmann (1998).
[25] C. Breazeal, A. Brooks, D. Chilongo, J. Gray,G. Hoffman, C. Kidd, et al. Working Collaboratively with Humanoid Robots. In: *Procs. International Conference on Humanoid Robots*, Los Angeles (2004).