Dialog Strategies for Handling Miscommunication in Task-Related HRI

Barbara Gonsior, Christian Landsiedel, Antonia Glaser, Dirk Wollherr, and Martin Buss

***Abstract*— As communication quality in public spaces often is impaired by noisy environment, it is difficult for a robot to retrieve missing task-information from humans. In this paper, different dialog strategies are modeled and evaluated with respect to user experience and error handling capabilities in HRI in order to cope with erroneous speech recognition. Since correct recognition of spoken language is a bottleneck for real-world dialog systems, special emphasis is placed on the issue of adapting dialog strategies to the conditions under which the dialog is held to thereby provide for adaptability of the dialog strategy to variable speech recognition performance. Experimental evaluations are conducted in a fully automated indoor setting, and in a Wizard-of-Oz outdoor setting. Results indicate that a critical point exists, up to which the use of requests for handling miscommunication improves the user experience of a dialog strategy.**

I. Introduction

In Human-Robot dialog, the extraction of knowledge from humans is vital in situations where a robot cannot fulfill its task based on the information it has or can acquire from its own sensor information. For robots in unknown and changing environments, this is a common situation. In order to have this information transfer work reliably, robust dialog systems are necessary, which are able to detect and handle errors and miscommunication while taking into account the task-relatedness of these dialogs. As discontinuation due to weak speech recognition performance prevents any information retrieval from humans in public spaces, proper handling of errors and miscommunication has to be ensured in order to provide validity of the acquired information on the one hand, and a satisfactory experience for the human interaction partners on the other hand.

Robots that ask humans for directions in order to extract information about environments are all still operating in very simple structured indoor environments. Coarse qualitative route descriptions can be given to a wheelchair robot [1] that navigates in an office floor. The office robot Jijo-2 [2] can learn the locations of offices and staff by moving around and asking humans for information. A robot asking for the way at a robotics conference is presented in [3]. A miniature robot that can find its way in a model town by asking for directions is described in [4]. These robots are able to interpret and follow simple route instructions, but cannot cope with the complexity and vagueness of natural language. Thus, careful design and robustness of the dialog is required, as well as proper environment modeling for situatedness of the dialog.

As there is no control over the environmental conditions, which may have great influence on speech recognition performance, misrecognitions may occur frequently and eventually mislead the dialog. Hence, miscommunication has to be handeled. Several Wizard-of-Oz studies successfully explored miscommunication and complexity arising from users giving verbal route instructions in a simulated dialogic real-time interaction with a robot executing the route instructions during the experiment [5], [6]. However, this paper adresses a robot that executes previously gained route instructions autonomously within real environment. Thus, complexity and the range of potential errors increases enormously, e.g. informational misalignment may be undetected during dialog but lead to errors during execution of the gained route knowledge. Thus, it is necessary to represent and evaluate the route description.

In this paper different dialog strategies are presented in order to fill a previously developed dialog framework for handling miscommunication with concrete configurations regarding the applicability in real world HRI. Four different dialog strategies are modeled and evaluated in two different experiments. Firstly, a fully automated (**FA**) indoor experiment evaluates each dialog strategy with respect to user experience based on a questionnaire. Secondly, the experiment is replicated within an outdoor Wizard-of-Oz (**WOz**) setting, and additionally, three different types of requests for handling miscommunication employed by the wizard are explored in combination with each dialog strategy.

The remainder of the paper is structured as follows: Section II provides an overview of the integrating dialog framework and corresponding dialog strategies applicable to real-world HRI settings together with requests for handling miscommunication; in Section III, four different dialog strategies and requests for handling miscommunication are evaluated; conclusions are given in Section IV.

II. Dialog Strategies

A. Dialog Framework

As introduced in [7], a dialog framework was further developed in order to provide an all-embracing structure for handling miscommunication in human-robot dialog, see Figure 1. The Theory of Perceptual Hypotheses (TPH), as originally formulated by Bruner & Postman [8] and extended

This work is part of the EU-STREP project IURO (Interactive Urban Robot), supported by the 7th Framework Programme of the European Union, ICT Challenge 2 Cognitive Systems and Robotics, contract number 248317, see also www.iuro-project.eu.

Support of the TUM Institute for Advanced Study (IAS), Technische Universität München, is hereby gratefully acknowledged, see also www.tum-ias.de.

The authors wish to thank Gabriel Skantze of the Department of Speech, Music and Hearing at the Royal Institute of Technology, Stockholm, for fruitful discussions and valuable comments.

Institute of Automatic Control Engineering (LSR), Technische Universität München, D-80290 Munich, Germany email: bg@tum.de, christian.landsiedel@tum.de, antonia.glaser@tum.de, dw@tum.de, mb@tum.de

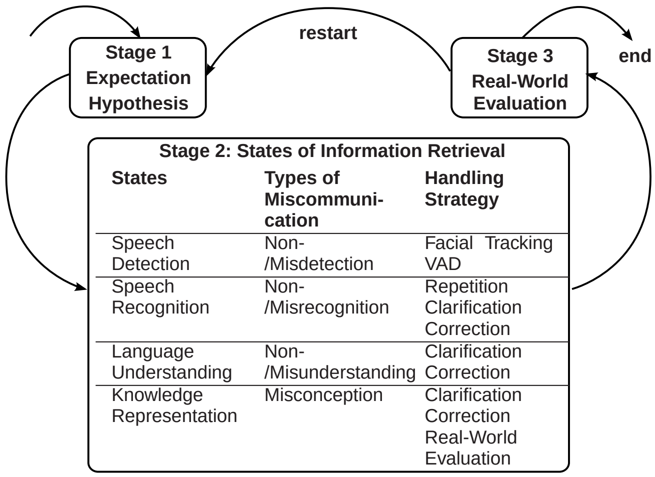

Fig. 1. Dialog framework for handling miscommunication in HRI

by Lilli & Frey [9] as *Hypotheses Theory of Social Perception*, provides the conceptual basis revealing that for humans any (mis-) interpretation results from a perceptual decision process evaluating an environmental stimulus input. This decision process can be seen as a loop controlled decision process composed of three stages: 1) provision of expectation hypothesis, 2) information input, and 3) confirmation or disproof of the hypothesis. In case of a disproved hypothesis, the cycle restarts as often as a hypothesis is confirmed. Transferred to spoken language dialog in HRI this means:

Stage 1) Expectation Hypothesis: Based on a context model the robot creates a hypothesis on what to expect from the human speech input, e.g. landmarks and directions. The robot should have a knowledge base containing both, predefined domain knowledge and previously stored task information gained from HRI.

Stage 2) States of Information Retrieval: In this stage, HRI takes place. A robot has to pass four states of information retrieval represented in a spoken dialog system in order to extract needed task-information from speech input: *Speech Detection*, *Speech Recognition*, *Language Understanding*, and *Knowledge Representation*. As Miscommunication has to be handled mainly in this stage, the robot is equipped with predefined associations between these states and different categories of miscommunication. For each of these, handling strategies deduced from human-human corpora related to each state, see Figure 1, can be assigned.

In order to become aware of putative miscommunication, each state calculates a confidence value that indicates the extent in which the extracted information meets a hypothesis. In case of low confidence, one of the related handling strategies is triggered, e.g. caused by ambiguity in natural language. By employing handling strategies, the robot reduces the number of possible hypotheses regarding interpretation of the speech input and thus raises the confidence values for other hypotheses until it is able to decide for one interpretation.

Stage 3) Real World-Evaluation: As miscommunication may be undetected during HRI, the robot has to evaluate the extracted information while performing its task. Therefore, the robot looks selectively for confirming or disproving information while performing its task within the real world by targeted questions to confirm the desired task-status If different and/ or conflicting hypotheses show a common denominator, the task will be performed until a critical point is reached and then the cycle of evaluation restarts.

After this overview of the functionalities embedded within the framework, the following subsection provides more detailed information on the design of **Stage 2)**, i.e. how HRI can be arranged and actually carried out by the robot.

B. Dialog Strategies

Since speech recognition in outdoor environments is highly subject to errors, resulting misrecognition may eventually mislead the dialog into wrong directions. Thus, it is essential to develop dialog strategies for a robot in order to control the dialog structure and thus triggering suiTable speech input on the one hand, and to provide a sense of naturalness and intuitiveness for the user on the other hand. Current approaches either use open requests, e.g. “how may I help you”, and then classify the recognized speech input by means of machine learning in a second step [10], [11], or the other way is to use a dialog strategy, where the systems asks rather closed questions and thus breaks the task down into several subtasks in order to get more and more required information with every question [12].

As outlined in [13], Wunderlich [14] analyzed asking-for-directions dialogs and found a common structure of four consecutive phases:

Introduction: The asker addresses a respondent and defines the task, i.e. giving directions to a specified goal location, possibly defining the mode of transportation or other individual requirements.

Giving directions: The respondent provides the necessary information by means of natural language and gestures, sometimes additionally with the help of a sketch.

Confirmation: Either of the two partners confirms the information. In this phase further inquiries can be made.

Conclusion: The asker thanks the respondent and they part.

This schematic structure is very flexible, i.e. some phases may be interchanged or recur. Nevertheless it is a well-proven guideline reflecting the intrinsic cognitive processes involved within human-human interaction and thus serves as a common ground to be transferred to HRI.

In the following subsections, different dialog strategies based on the above mentioned basic structure, but with variations regarding their flexibility and using open- and closed or even mixed promts, are presented in order to evaluate their impact on information retrieval in terms of user satisfaction and naturalness, exemplarily confined to the context of asking passers-by for directions but applicable to any task-oriented dialog.

1) Open Dialog: This strategy exactly applies the basic structure of asking-for-directions dialogs: The robot opens the dialog by introducing itself and asks for its way to a certain goal location. After the passer-by has given route instructions, the robot asks if it may repeat the route. If the reproduced route was incorrect, the robot asks the human to speak more clearly and describe the route again. After the human interactant confirms the reconstruction of the whole way, given by the robot during *confirmation* phase, the robot thanks and closes the dialog.

2) Divided Dialog: The divided dialog strategy coincides with the open dialog strategy with regards to the above mentioned basic structure. But, unlike open dialog this strategy differs within the phase of *Giving directions*: The robot asks directly for route segments beginning with ”please describe the first route segment” and reconstructs each route segment before it motivates the human to proceed with the next route segment. Compared to the Open Dialog this strategy shall reduce the time spent on repeating the whole instruction compared with correcting only one route segment.

3) Requesting Divided Dialog: This strategy coincides with the structure of the Divided Dialog but, unlike the latter, counts for each route segment if at least one landmark and one direction have been given. If no landmark was given, the robot requests the landmark by asking ”How far should I go in that direction or up to which point?”. Correspondingly, if no direction was given in a route segment, the robot asks ”In which direction shall I head?” and inserts this requested landmark or the gained direction into the reconstruction of the route description in order to be confirmed by the human after each route segment. There is no reconstruction of the complete route description at the end of the dialog in order to be confirmed.

4) Closed Dialog: The Closed Dialog-Strategy differs from the other strategies regarding its flow: The human can not give any free information input, but is asked to confirm or revise mainly closed questions. Again, the robot introduces itself, but directly after asking for its way, the robot continues with questions like ”Should I continue going in this direction?”, ”In which direction shall I head?” or ”In which direction shall I turn then?, followed by ”How far should I go in that direction or up to which point?”. The robot asks for directions and landmarks as long as it gets at least one of each for a route segment. Accordingly, the human interlocutor has very limited input-possibilities but speech recognition should be more robust due to the limited vocabulary. During *confirmation* phase at the end of the dialog the robot repeats the whole route description by combining all route segments, but unlike within the open dialog strategy, the robot asks for confirmation after the reconstruction of each route segment.

In order to improve information retrieval the closed dialog already contains requests to confine the vocabulary and to trigger the needed information input. Nevertheless, miscommunication may occur within all dialog strategies. Hence, there is an additional need for handling strategies, as presented in [7].

C. Handling of Miscommunication

The following subsection refines the selected requests for handling miscommunication related to the four states of information retrieval as presented on ***Stage 2*** in Figure 1, which were implemented and combined with the above mentioned dialog strategies in the Wizard-of-Oz experiment presented in Section III-B. As the experiment was conducted in German language it is important to note that the following requests are translated as far as possible into English, but in some cases meet the original meaning only approximately.

1) Repetition Requests: Assuming successful speech detection, miscommunication may initially occur in the state of speech recognition as ”non-recognition”, i.e. the robot could not gain any interpretation on what has been said by the human. In order to cope with non-recognition the following repetition requests were implemented:

- *”Could you repeat that, please?”*
- *”Excuse me, I didn’t get your answer. + [Repetition of the previous question]”*

2) Clarification Requests: As clarification requests are to confirm an interpretation [15] these kinds of requests are employed in case of ”mis-recognition”. They are used as well in every case of miscommunication within all following states of information retrieval given that speech recognition already released one possible interpretation of the speech input which can be taken as a hypothesis in order to be confirmed by the following clarification requests as explained more detailed in [7].

Reprise sluices mark the interpretation gap by emphasizing, e.g. ”wh”-alliterated words: ”*Sorry,...*

- *where?*”
- *when?*”
- *how far?*”

Wh-substituted reprises repeat the well-understood part and substitute interpretation gap: ”*Excuse me,...*

- *in which direction?*”
- *up to which landmark?*”
- *how far should I continue in this direction or up to where?*”
- *in which direction should I turn then?*”
- *how far should I go in this direction or up to where?*”

These requests are already integrated within *Closed Dialog* strategy as they are part of the closed questions in order to confine the needed vocabulary for speech recognition.

Reprise fragments are to emphasize an uncertain part of a gained interpretation: ”*Excuse me, did you mean...*

- *to the +[direction]?*”
- *at/up to/near +[router]?*”
- *I must pass by +[controller]?*”
- *en route, i will see +[controller] on the right?*”

Alternative clarification questions are to explicitly mention alternating interpretations in case of acoustic and referential ambiguity: ”*Excuse me, did you mean...*

- *sight or side?*”
- *turn to the right or turn to the side?*”
- *break or fake?*”

”*Excuse me, ...*

- *do I have to turn left or right at +[router]?*” in order to confirm the direction.

- *do I have to turn at + [router] or head straight on?*" in order to confirm if a certain landmark depicts a *controller*, and not a *router*.
- *do I have to pass +[controller] or turn there?*" in order to confirm if a certain landmark depicts a *router*, and not a *controller*.

Task-level reformulations are used to clarify more complex actions by reformulating the consequences of an utterance and thereby demonstrating subjective understanding. Thus these requests confirm the practical implication within an utterance: "*This would mean that...*

- *I have to turn back?*"
- *I should not turn until I have passed +[controller]?*"
- *at + [router] I have to turn + [direction]?*"

3) Correction Requests: If miscommunication is detected, e.g. by the user within the dialog phase of *confirmation* after the robot reconstructed the route description, correction requests are employed to revise an interpretation in order to meet the underlying intention of the speaker [15]: "*Excuse me, I think I got you wrong,...*

- *please tell me where I have to go instead.*"
- *please give the directions again, and a bit slower.*"

Within the wizard-of-Oz experiment these handling requests were combined with each dialog strategy within an experimental setting, which is described more detailed in the following subsection.

III. Experimental evaluation

The dialog strategies described and motivated in Section II were evaluated. For the experiments, the route description domain was chosen as it provides a valid and rather well-explored structure for human-machine interaction. Route descriptions given by subjects are stored and processed internally based on route graphs [16], which represent a route as a sequence of route segments. Each segment can consist of a *controller*, describing the traversal of the segment, a *router* describing the location at the end of the segment, and an *action* to take once this location has been reached, e.g. a change of direction. This representation is used to form clarification requests and to relate the described route to the user for possible correction.

The different dialog strategies were each modeled as state sequences according to their specifications given in Section II-B. Each state could either be a textual output node with speech output generated from templates that were filled with stored information given by the user if necessary, and input nodes where information given by the user was entered into the internal knowledge representation of the system. The transitions between the nodes, determining the course of the dialogue, were specified with conditions on the previous course of the dialog (e.g, requirements on the information given as response to question, such as posing a more refined question when not all relevant question had been provided).

A. Fully Automatic (indoor): Evaluation of Dialog Strategies

A first evaluation of user acceptance and user experience of the dialog strategies described in Section II was conducted in a user experiment. The dialog strategies, modelled as state sequences as described above, were used as templates for the DialogOS [1] tool, which provided text-to-speech synthesis and speech recognition for a fully automated (FA) natural language dialog. In this experiment, finite-state grammars for speech recognition were created for each input node.

During the experiment, a route displayed on a map was presented to the subjects for each of the dialog strategies separately. Then, subjects were asked to interact with the dialog system, which resulted in a description of the route. Following each dialog, participants filled out a questionnaire with questions describing their impression of the dialog. The experiment was held in a laboratory setting without embodiment of the dialog system.

B. Wizard-of-Oz (outdoor): Evaluation of Dialog Strategies and Handling Requests

The dialog strategies were evaluated with respect to user experience also in the style of a Wizard-of-Oz experiment (WOz), in which the task of entering user input into the knowledge representation of the system and the choice of system actions were performed by a human operator, which subjects were unaware of.

For this, the wizard was asked to perform in a way similar to a system restricted in its input by a predefined grammar for the user answers depending on the dialog state, similar to the system used for the experiment described in Section III-A. The transitions between the nodes were chosen depending on the dialog state by the wizard within the bounds of the respective dialog strategy, as defined by the corresponding state machine models.

In addition to the mere modeling of succession of dialog states, miscommunication handling strategies were implemented in this experimental setting. After each input node in which relevant, task-related information had been requested, the wizard had the possibility to choose from a number of handling requests as described in Section II-C.

In this experiment, for each dialog strategy, subjects engaged in dialogs with a dialog system embodied in a robot platform with human-like features. A map was not presented to the participants, but the users were asked to describe a way of their own choice to a well-known, nearby location.

In order to provide a realistic setting for the dialog, the experiment was conducted in an outdoor environment at Technische Universität München, and test persons faced the *IURO* robotic platform based on the Autonomous City Explorer (ACE) robot [17] as interaction partner. In order to enable natural language dialog and other modalities of interaction, the robotic platform is equipped with a number of sensors including cameras and microphones, a loudspeaker and a mechanical actuated head capable of displaying emotions and emulating speech movements [18].

C. Results

Results can be deduced from the initial fully automatic experiment including 16 subjects (11 male and 5 female,

[1]CLT Sprachtechnologie GmbH, www.clt-st.de

between 22 to 35 years with a mean of 27.0 years) described in Section III-A and the Wizard-of-Oz experiment including 29 subjects (21 male and 8 female, between 19 and 39 years with a mean of 22.9 years) described in Section III-B.

1) Quantitative Results: After each interaction following one of the four dialog strategies, which were presented in random order, subjects filled out a questionnaire including seven items regarding different aspects of the interaction. In detail, for experiments **FA** and **WOz** respectively, these items were (*translated into English*):
comprehension: "The system understood what I said/IURO understood me well."
duration: "The duration of the interaction was appropriate."
expectation: "I always knew which comments the system/IURO expected from me."
structure: "The structure of the dialog was sensible."
correction: "When there was a misunderstanding, the correction effort was appropriate."
request: "The system/the robot asked wisely for missing or uncertain information."
satisfaction: "Overall, I was satisfied with the dialog."
Every item was rated by participants on a Likert scale ranging from *1* = *"strongly disagree"* to *5* = *"strongly agree"*. Regarding reliability, coefficients of internal consistency for these items were good (Cronbach's $\alpha > .80$ overall).

Significance level for all performed tests was $\alpha = .05$ except for multiple testing where it had to be adjusted using the correction method of Bonferroni.

Fully Automatic (FA): According to the results of Kolmogorov-Smirnov tests, normal distribution could be accepted for every single item as well as for the total scores (calculated as means of all single item values per dialog strategy). Therefore parametric comparisons and correlations were performed.

An analysis of variance (ANOVA) with repeated measures revealed no significant difference between the total scores of the four dialog strategies. However, mean values show a trend towards a difference between the total rating of the *Open Dialog* and the *Closed Dialog*. Further repeated measure ANOVAs analyzing the ratings of single items provided significant differences between the dialog strategies for *comprehension* ($F = 4.16, p = .011$), *correction* ($F = 3.72, p = .023$), *request* ($F = 7.89, p = .001$) and *satisfaction* ($F = 4.11, p = .012$). Post-hoc tests revealed a significant difference between *Open Dialog* and *Closed Dialog* for *comprehension* ($t = -3.31, p = .005$). Similar deviations indicating that *Closed Dialog* was rated higher than *Open Dialog* were also obtained for *correction*, *request* and *satisfaction* as well, but failed to reach significance due to Bonferroni correction of significance level to $\alpha = .0083$. Means of single item ratings and total scores of the different dialog strategies are displayed in Table I.

Correlation analyses focused on the item *satisfaction* led to the finding of most meaningful relations in the *Open Dialog* condition: subjects' general satisfaction with the dialog significantly correlated with *comprehension* ($r = .83, p < .001$), *duration* ($r = .75, p = .001$), *structure* ($r = .79, p < .001$), *correction* ($r = .89, p < .001$) and *request* ($r = .98, p < .001$). The impression arises that more aspects of the interaction with the system had to be pleasing to satisfy the user in *Open Dialog* condition in comparison to the other strategies.

TABLE I
MEAN RATINGS WITH STANDARD DEVIATIONS (IN BRACKETS) OF SINGLE ITEMS AND TOTAL SCORES FOR EACH DIALOG STRATEGY DERIVED FROM THE FULLY AUTOMATIC (FA) EXPERIMENT (RATED ON LIKERT SCALES FROM 1 = STRONGLY DISAGREE TO 5 = STRONGLY AGREE).

	Dialog			
	Open	Divided	Requesting Divided	Closed
comprehension	2.9(1.3)	3.4(1.0)	3.2(1.1)	4.1(0.8)
duration	3.1(1.4)	3.3(1.0)	2.9(1.1)	3.8(0.9)
expectation	3.4(1.0)	4.1(0.9)	3.3(1.2)	3.7(1.2)
structure	3.6(1.4)	3.9(0.9)	4.0(0.9)	4.1(0.9)
correction	2.9(1.4)	3.5(1.2)	3.4(1.1)	4.1(1.0)
request	2.5(1.5)	3.6(1.2)	3.5(1.2)	3.9(1.1)
satisfaction	2.7(1.4)	3.3(0.8)	3.1(0.9)	3.9(1.0)
total score	3.1(1.1)	3.6(0.8)	3.3(0.7)	3.9(0.8)

Wizard-of-Oz (WOz): For the single items, normal distribution had to be rejected according to the results of Kolmogorov-Smirnov Test, but was accepted for the total scores (again calculated as means of all single item values per dialog strategy). Hence, comparisons and correlations regarding single items were performed non-parametrically and parametric methods were used for total scores.

Again, no significant difference between total scores could be derived from ANOVA with repeated measures, but means show a trend towards higher ratings of the *Open Dialog* and *Divided Dialog* compared to the remaining two strategies. On single item level, the ratings of *duration* ($\chi^2 = 9.00, p = .027$) and *satisfaction* ($\chi^2 = 8.62, p = .035$) significantly varied. According to post-hoc analyses the rating of *duration* considerably differed between *Open Dialog* and *Closed Dialog* ($Z = -2.68, p = .006$), whereas after adjusting the α-value no significance remained for *satisfaction*. Means of single item ratings and total scores of the different dialog strategies are displayed for all three conditions in Table II.

Correlations of *satisfaction* with other single items once again varied between dialog strategies. In condition *Closed Dialog*, there were clear relations to *comprehension* ($r = .57, p = .001$), *duration* ($r = .59, p = .001$), *structure* ($r = .59, p = .001$), *correction* ($r = .57, p = .002$) and *request* ($r = .56, p = .002$). Fewer significant correlations could be obtained for the other strategies and there was only one in *Open Dialog* condition (*comprehension*: $r = .51, p = .006$). Apparently, more aspects of the interaction were important to satisfy participants in the *Closed Dialog* condition compared to the other strategies.

In addition, to solve miscommunication problems, several handling requests could be used by the wizard during each dialog. The distribution of *repetition*, *clarification* and

TABLE II

MEAN RATINGS WITH STANDARD DEVIATIONS (IN BRACKETS) OF SINGLE ITEMS AND TOTAL SCORES (RATED ON LIKERT SCALES FROM 1 = STRONGLY DISAGREE TO 5 = STRONGLY AGREE) PLUS NUMBER OF HANDLING REQUESTS FOR EACH DIALOG STRATEGY DERIVED FROM THE WIZARD-OF-OZ (WOZ) EXPERIMENT.

Means of single items and total scores.

	Dialog			
	Open	Divided	Requesting Divided	Closed
comprehension	3.6(1.3)	3.9(1.1)	3.7(1.2)	4.0(1.0)
duration	4.2(1.0)	3.8(1.1)	3.5(1.3)	3.2(1.3)
expectation	4.0(1.0)	4.0(1.1)	4.0(1.0)	3.5(1.4)
structure	4.4(1.0)	4.3(1.2)	4.0(1.1)	3.9(1.2)
correction	3.9(1.2)	4.0(1.1)	3.7(1.2)	3.7(1.2)
request	3.8(1.4)	4.4(0.8)	4.2(0.9)	4.0(1.0)
satisfaction	4.3(0.8)	4.1(1.1)	3.9(1.2)	3.8(0.9)
total score	4.0(0.7)	4.1(0.8)	3.8(0.9)	3.7(0.8)

Number of handling requests per type and in total.

repetition	4	4	3	9
clarification	17	53	63	70
correction	5	1	3	7
total	26	58	69	86

correction requests was examined with Friedman tests. A significantly different use between dialog strategies was only obtained for the total amount ($\chi^2 = 45.90, p < .001$) and for *clarification* requests ($\chi^2 = 41.76, p < .001$). These results mainly arose from deviations between *Open Dialog* compared to the other strategies. The absolute number of applied handling requests was the highest for *Closed Dialog* for every type of request and in total. The number of included requests both for different types and for all requests in sum per dialog is displayed by Table II.

Comparison of ratings between both experiments: Total scores of strategy ratings were compared between the **FA** and **WOz** experiment with paired *t* tests resulting in a significant difference for *Open Dialog* ($t = -3.09, p = .005$) and marginally significant deviations for *Divided Dialog* ($t = -1.98, p = .054$) and *Requesting Divided Dialog* ($t = -2.00, p = .052$). The means indicate higher ratings for *Open Dialog*, *Divided Dialog* and *Requesting Divided Dialog* in the **WOz** compared to the **FA** experiment, whereas for *Closed Dialog* the relation is vice versa. Means and standard deviations of total scores for both experiments are displayed in Figure 2. In sum, quantitative results show varying ratings of the different dialog strategies between the two conducted experiments. In **FA** *Closed Dialog* was rated highest and *Open Dialog* lowest, whereas in **WOz** *Closed Dialog* was the most unpopular strategy. Surprisingly, in the second experiment most handling requests were used in *Closed Dialog* and least in *Open Dialog*.

Duration of interaction: Actual durations of the interactions per dialog strategy were distributed normal in each condition and experiment. Hence, comparisons were performed parametrically.

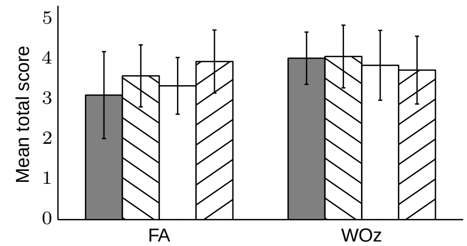

Fig. 2. Means and error bars ($\pm$ one standard deviation) for the total scores both of the Fully Automatic (FA) and the Wizard-of-Oz (WOz) scenario for all different dialog strategies. From left to right, the bars for each condition describe ratings obtained using *Open*, *Divided*, *Requesting Divided* (yellow) and *Closed Dialog* strategies, respectively.

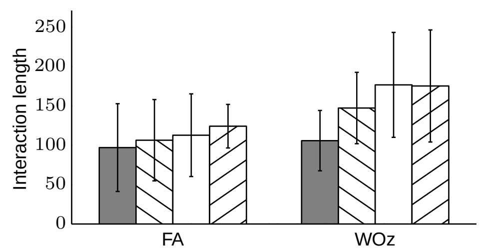

Fig. 3. Means and error bars ($\pm$ one standard deviation) for the duration of dialogs both of the Fully Automatic (FA) and the Wizard-of-Oz (WOz) scenario for all different dialog strategies. From left to right, the bars for each condition describe ratings obtained using *Open*, *Divided*, *Requesting Divided* and *Closed Dialog* strategies, respectively.

Concerning the interaction durations of conditions between both experiments, paired *t* tests revealed significant results for *Divided Dialog* ($t = -2.74, p = .009$), *Requesting Divided Dialog* ($t = -3.29, p = .002$) and *Closed Dialog* ($t = -2.75, p = .009$). As indicated by the means, interactions in these conditions were clearly longer in the **WOz** compared to the **FA** experiment (means and standard deviations of durations per dialog strategy for both experiments are displayed in Figure 3). Longer durations in these conditions probably resulted from the high amount of included handling requests as depicted in Table II.

One significant difference was obtained within the **WOz** experiment ($F = 11.13, p < .001$). According to post-hoc tests, only the actual duration of *Open Dialog* strongly deviated from all other conditions. This finding fits results derived from the questionnaire, in which the perceived duration of the *Open Dialog* strategy was rated best. For the initial **FA** experiment, no significant differences between the four strategy durations were found and again, this fact suits the duration ratings derived from the questionnaire.

IV. CONCLUSIONS

Four different dialog strategies were modeled and evaluated in two different experiments. Firstly, a fully automated (**FA**) indoor experiment was conducted, where each dialog strategy was evaluated with respect to user experience based on a questionnaire. Secondly, the experiment was replicated within an outdoor Wizard-of-Oz (**WOz**) setting, and additionally, three different types of requests for handling miscommunication employed by the wizard were evaluated in combination with each dialog strategy.

The fact that the total user experience scores within *Closed Dialog* condition decrease marginally in **WOz** compared to the previous **FA** experiment goes hand in hand with the usage of handling requests which is highest within *Closed Dialag* and lowest within *Open Dialog* condition. This finding might speak against the acceptance of the proposed handling requests. However, in all other conditions the total scores are raised compared to the **FA** experiment without handling requests, which indicates a slightly positive impact of those. Additionally, the above mentioned decreasing trend within *Closed Dialog* might be due to the fact that the most used *clarification* requests were already partly implemented in this condition, because of beeing part of the closed dialog strategy. This might have caused a feeling of over-usage of handling requests for the users compared to the other conditions. Furthermore, as the Wizard was only able to employ handling requests while being the initiative part within all dialogs, *Closed Dialog* provided more chances for usage than within *Open Dialog*, because there was least turn-taking due to the open prompt-strategy at the beginning of the dialog which did not allow the wizard to interrupt the user while giving route instructions. In contrast, the *Closed Dialog* strategy provokes frequent turn-taking and thus allows for more handling requests.

Summing all up, experimental results indicate that the application of handling requests raises user satisfaction as can be seen in the total scores for *Open*, *Divided*, and *Requesting Divided* Dialog. However, at a certain point, when there are too many handling requests employed, the effect changes into the opposite and decreases user ratings again. Another factor of influence is the duration of the interaction: Within the **WOz** experiment the duration of *Open Dialog* condition is significantly shorter than all other dialog strategies and accordingly rated as most convenient duration in the questionnaire. Due to the resulting significant increase of the total scores for *Open Dialog* compared between **FA** and **WOz** experiments it is suggested to employ the proposed handling strategies in a flexible way, but confined in a way to avoid a critical increase of dialog duration and a feeling of over-usage.

Future work will explore and refine this critical point by investigating if an error-rate of speech recognition can be calculated online during interaction and thereby provide a threshold for employing handling requests and/or switching to a different dialog strategy in order to prevent exceedance of another critical threshold regarding dialog duration. The goal is to improve natural error handling in terms of user experience within HRI.

REFERENCES

[1] R. Müller, T. Röfer, A. Lankenau, R. Musto, K. Stein, and A. Eisenkolb, "Coarse qualitative descriptions in robot navigation," in *Spatial Cognition II*, 2000, pp. 265–276.

[2] H. Asoh, Y. Motomura, F. Asano, I. Hara, S. Hayamizu, N. Vlassis, and B. Kröse, "Jijo-2: An office robot that communicates and learns," *Intelligent Systems*, vol. 19, no. 5, pp. 46–55, 2001.

[3] M. Michalowski, S. Sabanovic, C. DiSalvo, D. B. Font, L. Hiatt, N. Melchior, and R. Simmons, "Socially Distributed Perception: GRACE plays Social Tag at AAAI 2005," *Autonomous Robots*, vol. 22, no. 4, pp. 385–397, 2007.

[4] S. Lauria, G. Bugmann, T. Kyriacou, and E. Klein, "Instruction Based Learning: How to Instruct a Personal Robot to Find HAL." in *European Workshop on Learning Robots*, 2001.

[5] G. Skantze, "Exploring human error recovery strategies: Implications for spoken dialogue systems," *Speech Communication*, vol. 45, no. 3, pp. 325 – 341, 2005, Special Issue on Error Handling in Spoken Dialogue Systems.

[6] T. Koulouri and S. Lauria, "Exploring Miscommunication and Collaborative Behaviour in Human-Robot Interaction," in *Proc. SIGDIAL 2009*. Morristown, NJ, USA: Association for Computational Linguistics, 2009, pp. 111–119.

[7] B. Gonsior, D. Wollherr, and M. Buss, "Towards a Dialog Strategy for Handling Miscommunication in Human-Robot Dialog," in *IEEE Int. Symp. on Robot and Human Interactive Communication*, 2010, pp. 284–289.

[8] J. Bruner and L. Postman, "On the perception of incongruity: a paradigm." *Journal of Personality*, vol. 18, pp. 206–223, 1949.

[9] W. Lilli and D. Frey, *Theorien der Sozialpsychologie [Theories of Social Psychology]*, 2nd ed. Bern: Hans Huber Verlag, 1993, vol. 1: Cognitive Theories, ch. Die Hypothesentheorie der sozialen Wahrnehmung [The Hypotheses Theory of Social Perception], pp. 49–78.

[10] A. L. Gorin, G. Riccardi, and J. Wright, "How May I Help You?" *Speech Communication*, vol. 23, pp. 113–127, 1997.

[11] J. Chu-Carroll and B. Carpenter, "Vector-Based Natural Language Call Routing," *Computational Linguistics*, vol. 25, no. 3, pp. 361–388, September 1999.

[12] J. Boye and M. Wirén, "Multi-slot Semantics for Natural-Language Call Routing Systems," in *Proc. of the Workshop on Bridging the Gap: Academic and Ind. Research in Dialog Technologies*, 2007, pp. 68–75.

[13] A. Bauer, B. Gonsior, D. Wollherr, and M. Buss, "Heuristic rules for human-robot interaction based on principles from linguistics - asking for directions," in *AISB Convention - Symposium on New Frontiers in Human-Robot Interaction*, 2009, pp. 24–30.

[14] D. Wunderlich, "Wie analysiert man Gespräche? Beispiel: Wegauskünfte," *Linguistische Berichte*, vol. 53, pp. 41–76, 1978.

[15] M. Gabsdil, "Clarification in Spoken Dialogue Systems," in *Proc. of 2003 AAAI Spring Symp. Workshop on Natural Language Generation in Spoken and Written Dialogue*, Stanford, USA, 2003.

[16] S. Werner, B. Krieg-Brückner, and T. Herrmann, "Modelling Navigational Knowledge by Route Graphs," *Spatial cognition II*, pp. 295–316, 2000.

[17] A. Bauer, K. Klasing, G. Lidoris, Q. Mühlbauer, F. Rohrmüller, S. Sosnowski, T. Xu, K. Kühnlenz, D. Wollherr, and M. Buss, "The Autonomous City Explorer: Towards Natural Human-Robot Interaction in Urban Environments," *International Journal of Social Robotics*, vol. 1, no. 2, pp. 127–140, 2009.

[18] C. Mayer, S. Sosnowski, K. Kühnlenz, and B. Radig, "Towards Robotic Facial Mimicry: System Development and Evaluation," in *IEEE Int. Symp. on Robot and Human Interactive Communication*, 2010, pp. 198–203.