**19th IEEE International Symposium on Robot and**
**Human Interactive Communication**
**Principe di Piemonte - Viareggio, Italy, Sept. 12-15, 2010**

# Towards a Dialog Strategy for Handling Miscommunication in Human-Robot Dialog

Barbara Gonsior and Dirk Wollherr and Martin Buss

*Abstract*— **This paper presents a first theoretical framework for a dialog strategy handling miscommunication in natural language Human-Robot Interaction (HRI). On the one hand the dialog strategy is deduced from findings about human-human communication patterns and coping strategies for miscommunication. On the other hand, relevant cognitive theories concerning human perception serve as a conceptual basis for the dialog strategy. The novel approach is firstly to combine these communication patterns with coping strategies and cognitive theories from human-human interaction (HHI) and secondly transfer them to HRI as a general dialog strategy for handling miscommunication. The presented approach is applicable to any task-oriented dialog. In a first step the conversational context is confined to route descriptions, given that asking for directions is an restricted but nevertheless challenging example for task-oriented dialog between humans and a robot.**

## I. INTRODUCTION

With increasing demand for autonomous robots that are to assist humans as personal assistants with more and more complex tasks [1], the research field of HRI has to meet new and challenging requirements. Since it is inevitable that robots will not always posses all required task information, future assistive robots must be able to retrieve missing information through human-robot interaction. As natural language is the most convenient way for humans to communicate information, there is a strong need for dialog systems with natural language interfaces (NLI) to meet the requirements of reliable information extraction (IE). Considering that ”vagueness is one of the most salient, but also one of the most effective features of natural language” [2], dialog systems of robots have to cope with possible ambiguities and interpret information extracted from dialogs with humans correctly. Given that dialogs with robotic systems have to be natural and intuitive for non-expert users, a strong need for error detection and correction mechanics arises for spoken dialog systems in order to enable them to control dialogs in a way that provides optimal information retrieval (IR) on the one hand, and intuitive naturality for a human user without any prior instructions on the other hand. For this purpose findings from human-human communication need to be involved into the spoken dialog system of robots.

Interacting with a user to get route directions can be regarded as a problem-solving application, in which one of the dialog partners has a problem he cannot solve (the novice) and the other partner has the required domain knowledge to solve the problem (the expert) [3]. Most common settings postulate the dialog system to be the expert and the user to be the novice, e.g. TRAINS [4], and Steve [5]. The few exceptions include [6], [7], where robots ask for directions. However, both projects operate in extremely structured indoor environments and none of these robots reason about the human input and missing information. However, information assessment and interpretation is essential in natural environment communication to detect and recover communication errors.

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 [8], [9]. 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 the goal is to develop a dialog strategy enabling robots to interact with humans in a natural and intuitive way within real environment in order to autonomously achieve a given interaction goal. More precisely, natural error handling should be achieved by selective rising and evaluating informational contents through well-directed requests at such a rate that miscommunication can be handeled. In order to cope with miscommunication in HRI a first framework is presented based on findings from human-human communication patterns. The conceptual basis is given through the transfer of relevant cognitive theories concerning human perception that are subsequently combined to a novel approach for handling miscommunication in task-oriented dialog. In [10] a robot dialog system has been developed that incorporates heuristic rules derived from linguistic principles regarding human-human communication. The dialog system enables robots to extract route information from dialogs with passers-by. Evaluation results confirm the importance of correction mechanisms as an appropriate reaction to miscommunication.

The remainder of the paper is organized as follows: Section II defines three different categories of miscommunication based on findings from human-human communication. Section III provides an overview of human strategies in handling miscommunication related to the corresponding

---

Support of the TUM Institute for Advanced Study (IAS) is hereby gratefully acknowledged. This work is supported in part within the DFG excellence initiative research cluster *Cognition for Technical Systems – CoTeSys*, see also `www.cotesys.org`, and within the European Commission as part of the *IURO project*, see also `www.iuro-project.eu`.

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

category and subtype of miscommunication. In Section IV relevant theories and models are introduced and subsequently transferred to HRI application concerning the developed dialog strategy. Section V presents an recapitulatory overview of the deduced dialog strategy for HRI assigned to a robot asking for directions. The strategy combines human handling strategies and relevant theories in a cycle of three consecutive stages the robot has to pass in order to evaluate the retrieved information and handle errors due to informational misalignment arisen from miscommunication in human-robot dialog. Conclusions are given in Section VI.

## II. CATEGORIES OF MISCOMMUNICATION

According to Hirst et al. [11] miscommunication can be devided into three different categories: *Non-understanding*, *Misunderstanding* and *Misconception*. These categories are described below more precisely regarding their symptomatic definition and awareness of the dialog partners.

*Non-understanding* — The hearer is aware of not obtaining any clear interpretation of the speaker‘s utterance. That means, he/she either finds no interpretation at all or is unable to choose between alternating interpretations. If there is only one available interpretation but the hearer is uncertain about getting it right, Gabsdil [12] introduces the subtype of ”uncertain understanding”. Accordingly, three subtypes of non-understanding are distinguished: *(i) total non-understanding* when the listener gets no interpretation at all, *(ii) uncertain non-understanding* when the hearer is not sure about his interpretation and *(iii) ambiguous non-understanding* when the hearer has to choose between more than one possible interpretation.

*Misunderstanding* — The hearer obtains an interpretation of the speaker’s utterance but is not aware of misinterpreting it. Thus, misunderstandings are not detected until one of the interlocutors gets aware of an informational misalignment.

*Misconception* — Misconception goes back to deeper differences in the (conceptional) beliefs of the dialog partners. In contrast to non-understanding and misunderstanding, misconception occurs when a hearer obtains the right interpretation of an utterance but unexpectedly does not align with its content or meaning. The hearer is aware of the informational misalignment.

Based on this classification the following section provides an overview on handling strategies from human-human interaction (HHI) related to each category of miscommunication.

## III. HANDLING OF MISCOMMUNICATION IN HHI

In order to develop a natural and intuitive dialog strategy for handling miscommunication in human-robot dialog, it is essential to explore in which way human users are accustomed to handle miscommunication. Linguistically motivated research has revealed several strategies for signalling and repairing informational misalignment due to miscommunication within HHI. This section points out which type of *clarification request* or *correction* is used for handling each (sub-)type of miscommunication.

### *A. Non-Understanding: Clarification Requests*

In a need for confirmation the hearer will usually employ different kinds of handling strategies in order to signal and repair non-understanding depending on the particular subtype. Gabsdil [12] defines *clarification requests* as backward looking questions, intuitively asked by a dialog partner ”only when he did not (fully) understand or is uncertain about what the previous speaker said or meant with an utterance”. In general, humans tend to clarify only parts of a dialog instead of the whole utterance. Thus, besides using a conventional ”sorry?” that is ambiguous and underspecified regarding the context of a dialogue, humans employ *Partial Clarification Questions*, and alternatively *Task-level Reformulations* in order to raise informational contents.

*1) Partial Clarification Questions (PCQ):* Specific repetition or replacement of only the unclear parts of an utterance in a fragmentary elliptical way. These kinds of questions can in turn be differentiated between

*(i)* **Wh-questions:** insertion of ”Wh”-alliterated words.
*Reprise sluices*: mark the interpretation gap in utterances like ”Look, there is Peter” by a wh-word, e.g. ”Who”?
*Wh-substituted reprises*: repeat the well-understood part of the utterance completed with a wh-word, e.g. ”He’s where”?
*(ii)* **Reprise fragments**: PCQ repeating the uncertain part of an utterance, generating a single polar question by emphasizing the uncertain part via specific intonation of the supposed interpretation, e.g. ”five hundred *miles*”?
*(iii)* **Alternative clarification questions**: closed PCQ that explicit mention the alternating interpretations.
*acoustic ambiguity*, e.g. ”Gablestreet or Babelstreet”?
*referential ambiguity*, e.g. ”next street left or right”?

*2) Task-Level Reformulations (TLR):* are used for clarification of more complex actions, e.g. within task-oriented dialog by reformulating the consequences of an utterance and thereby demonstrating subjective understanding. TLR are confirming the practical implication, e.g. ”You mean back to the main street?” within the utterance ”take a u-turn and go straight on to the next intersection”.

Table I provides an overview of all types of clarification requests for handling each type of non-understanding.

### *B. Misunderstanding: Correction*

If misunderstandings are detected in the course of a dialog they may be repaired by the dialog partners through correcting each other. While the above mentioned clarification requests should confirm the hearer’s interpretation, *corrections* are to revise the hearer‘s interpretation in order to meet the speaker’s underlying intention [12].

### *C. Misconception: Clarification Requests/ Correction*

The interpreter will employ again *clarification requests* and polite forms of *correction* in order to signal the need of repairing mutual alignment.

The human ability to choose an appropriate type of *correction* or *clarification request* goes back to several cognitive functions, e.g. to represent and evaluate any verbal input. Thus, it is not enough to consider explicit handling strategies

TABLE I

Overview of Clarification Requests for Handling Non-Understanding

| Degrees of Non-Understanding | Clarification Requests | | | | |
|---|---|---|---|---|---|
| | Partial Clarification Requests | | | | Task-Level Ref. |
| | Wh-questions | | Reprise fragments | Alternative clar. quest. | |
| total | Reprise sluices<br>*Wh*o? | Wh-substituted reprises<br>He‘s *wh*ere? | | | |
| uncertain | | | Five hundred *miles*? | | That means *back*? |
| ambiguous | | | | *Ga*bel- or *Ba*belstreet? (acc.)<br>*left* or *right* building? (ref.) | |

from HHI. In contrast, these strategies must be embedded into an all-embracing cognitive framework deduced from relevant cognitive theories and models. The next section introduces suitable backgrounds from linguistics and social psychology.

## IV. RELEVANT THEORIES AND MODELS TRANSFERRED TO HRI

With regard to a robot reliably obtaining missing task-knowledge from humans by handling miscommunication it is indispensable to figure out some selected theories and models describing cognitive processes that allow for successful communication in HHI. These models and theories regarding understanding, representing and evaluating any spoken input are outlined in this section. Subsequent to every subsection each of them is transferred to HRI.

### *A. States of Understanding*

Empirical HHI-studies revealed that, within verbal communication, successful understanding evolves from the listeners ability to pass through four consecutive phases, called the ”states of understanding” by Clark & Schaefer [13]: *state 0* where the listener does not notice that the speaker uttered anything, *state 1* when the listener noticed an utterance, but did not hear it correctly, *state 2* when the utterance was correctly heard, and *state 3* when the listener understood what the speaker ment by an utterance. Miscommunication can affect each state, and parts of the same utterance may be spread over different states of understanding. In case the interpreter supposes to be in a more advanced state than he/she really is, the communicative goal is not achieved until a mutual belief about being in final *state 3* is established for both interlocutors.

*1) Transfer to HRI: States of Information Retrieval (IR):* The states of understanding are the minimum basis for each successful communication act as passing them successfully assures that the listener understood what the speaker meant by an utterance. Thus, passing these states is indispensable for a robot talking to a human. [1]

For HRI these states provide a guideline for IR starting with speech recognition. As miscommunication can occur on each state it is important to assign each category of miscommunication and related handling strategy to one of these states of understanding. Endowing a robot with these consecutive states as a guideline for IR from human-robot dialog will enable it to choose the right handling strategy for each kind of miscommunication depending from the state in which it occurs. Table II schows the proposed transfer of the states of understanding to states of IR with correlations to miscommunication and corresponding handling strategy. Thereby, state 1: speech recognition, is unique for each human-robot dialog. For the special case of a robot asking for directions, state 2: extraction of relevant information, is to extract landmarks and directions. Accordingly, state 3: organisation/ representation of extracted RI corresponds to route graphs representing the retrieved route belief.

### *B. The Theory of Perceptual Hypotheses (TPH)*

Cognitive theories from social psychology are deduced from empirical proven data concerning human behaviour and problem solving. Thus they provide useful guidelines to be considered within HRI.

TPH as originally formulated by Bruner & Postman [14] and extended by Lilli & Frey [15] as *Hypotheses Theory of Social Perception* is based on regarding perception as a cognitive interaction between an organism and its environment. The process of perception is generally seen as reception and interpretation of stimuli managed by available hypotheses about the environment.

The basic assumption is that any process of perception starts with an expectation hypothesis, even before recognition of any environmental stimulus input. Such hypotheses originate from prior experiences of perception and can be seen as a set of cognitive predispositions. Accordingly, the chosen hypothesis affects perception to a certain degree by defining the kind of information to look for. Hence, the perceived objects can be seen as a selection out of diverse environmental stimuli organized by emphasizing some aspects of stimuli more than others. In other words, every perception can be seen as a result of former perceptions, successfully approved in prior similar situations. So the process of perception can be divided into three consecutive stages: *stage 1* - provision of expectation hypothesis, *stage 2* - input of information about the object of perception and *stage 3* - confirmation (end) or disproof (restart) of the hyposesis. Hence, each process of perception ends either with confirmation of an expectation hypothesis through perceived information or if the received input data disproves the expectation hypothesis the perception cycle restarts. If this cycle restarts several

---

[1] As recognition of speech input is the first step for a robot to retrieve any information from dialogs with humans, stage 0, where the listener did not notice the speaker's utterance, is ignored.

TABLE II

CONSECUTIVE STATES OF INFORMATION RETRIEVAL (IR) ADAPTED FROM THE STATES OF UNDERSTANDING [13] TO HRI

| STATE | INFORMATION RETRIEVAL (IR) | MISCOMMUNICATION | HANDLING |
|---|---|---|---|
| State 1 | Speech recognition | Non-understanding | Clarification |
| State 2 | Extraction of relevant information (landmarks, directions) | Misunderstanding | Correction |
| State 3 | Organization and representation of extracted relevant inf. (route graph) | Misconception | Clarification/Correction |

times the underlying strategy may be falsification of several hypotheses as perception can be seen as decision process.

*1) The strength of hypotheses:* As there can be more than one hypothesis at the same time the actual extent of influencing perception performance depends on the strength of a hypothesis. If an underlying hypothesis is very strong it primarily determines the perception process, i.e. concept-driven or top down information processing. In contrast, if the built hypothesis is rather weak it leads to data-driven information processing, i.e. bottom-up [16].

There is a continuous relation between the strength of an expectation hypothesis and available stimulus information input. Lilli & Frey [15] deduced three basic assumptions: *(i)* The stronger a hypothesis, the more likely it is activated, i.e. *priming*. *(ii)* The stronger a hypothesis, the smaller the amount of needed information to confirm it. *(iii)* The stronger a hypothesis, the larger the amount of needed conflicting information to disprove it.

The strength of a hypothesis itself depends on five determinates: (a) Frequency of former confirmation: the higher the frequency the higher is the subjective confidence. (b) Number of alternative hypotheses: the higher the number, the lower is the chance for each to take effect. (c) Motivational impacts: motivation triggers selection of hypothesis-supporting stimulus information and avoiding hypothesis-contradicting information. (d) Cognitive impacts: The more it is fixed within cognition, the more it is dominant and modification-resistant, forming a guiding hypothesis e.g. daily routines. (e) Social impacts: In the absence of suitable stimulus information the accordance of social group members can serve as hypothesis confirmation.

*2) Transfer to HRI: Hypothesis-Testing:* TPH provides a framework for evaluating the gained knowledge via hypothesis-testing and thus detecting and handling miscommunication for HRI, suitable for every task-oriented dialog. Since this work is specified on a robot asking for directions the following explanations refer to this context.

Given that the robot possesses no predefined knowledge and no prior experiences before asking the first passer-by on its way to a predefined goal it will only be endowed with its ability to extract route information from the human input on the one hand and to store the extracted route belief as a route graph. Thus, its initial expectation hypothesis $H_0$ will only be to receive directions and landmarks in order to create a first representation of the extracted route knowledge. But after the first dialog with any pedestrian crossing its way, it will retrieve the first route graph that forms its first real expectation hypothesis $H_1$. As $H_1$ is the only expectation hypothesis of the robot it is automatically primed as conversational knowledge-basis with maximum strength. That means, on it's way, the robot selectively looks for information to confirm it, e.g. by short and closed questions like asking someone if this is the right direction to its predefined goal. In the case of conflicting information caused by miscommunication it will build an alternative hypothesis $H_2$. If targeted questions rather confirm this new route graph it will assign $H_2$ to a higher strength value as $H_1$. Otherwise it will rather specify $H_1$ by means of detailed questions in order to facilitate its evaluation and grounding by retrieving additional information, e.g. "Can you tell me the name of this street?". In this way, the robot turns given relative information, e.g. second street, into absolute information, e.g. second street, called Gabelstreet. According to Gabsdil [12] these kind of refining questions are related to "other check questions", that differ from clarification requests and corrections as they cannot be assigned to the "backward looking function". They are used when the listener correctly understood an utterance but is aware of needing additional information in order to prevent miscommunication. In this way $H_1$ is executed until again conflicting information input occurs that triggers the evaluation cycle to start again.

In other words, if the expectation hypothesis $H_1$ and one or even more alternative hypothesis $H_2 + n$ are rivaling against each other, the robot will test and refine them by targeted questions as long as one of them dominates with a higher strength value. If miscommunication occurs it will lead to conflicting information that in turn creates alternative hypotheses. In this way even misunderstandings that have been undetected during the dialog will be detected sooner or later. In the case that one or more alternative hypotheses $H_2 + n$ are consistently on par with $H_1$, according to TPH's motivational impacts on the strength of a hypothesis, the robot may ask which route is easier to find or more convenient than another. Finally, in virtue of cognitive impacts, the more a hypothesis or parts of it could be confirmed through hypothesis-testing the more it is dominant and eventually turns into a guiding hypothesis that is (nearly) modification-resistant. As even undetected misunderstanding leads to conflicting information, this approach helps detecting it.

Fig. 1 shows four examples of route graphs extracted from dialog from the experiment that conflict from a certain point as alternative hypotheses. Nevertheless they are equal in their beginnings, namely the directions right and then two times straight on, building their least common denominator. According to TPH their common beginnings constitute a guiding hypotheses after successfully reaching this point. Afterwards the robot must start hypotheses-testing in order to decide which route to take.

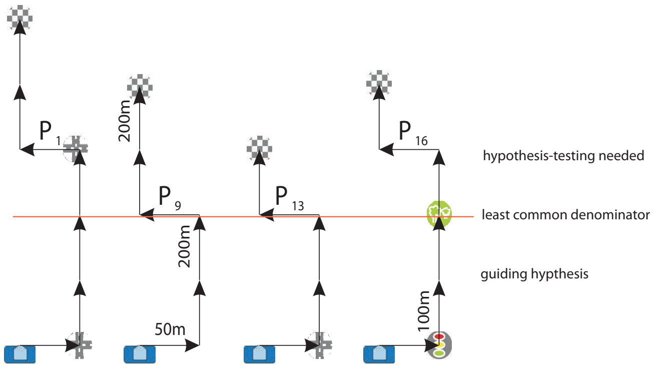

Fig. 1. Four examples of route graphs extracted from dialogs from the experiment (participants $P_1$, $P_7$, $P_{11}$, $P_{14}$). The red line marks the least common denominator as a critical point: below this line the route graph is confirmed and no conflicting information exists. Above the line hypothesis-testing is needed because of more conflicting hypotheses.

### *C. Deixis Theory*

In natural language communication referring is generally managed by deictics [17], e.g. ”here” and ”there”, which point to subspaces and are the verbal equivalents to pointing gestures. In order to refer to these subspaces both speaker and listener have to share a common ”deictic space”. At the beginning of a face-to-face route description this deictic space is given through the range of the visual perception. In the course of the conversation the route description usually leaves the shared visual perception range and a new deictic space is built by the geographical knowledge of both partners. Thus, orientation within these geographical spaces depends on particular contextual factors such as the position of the speaker and the direction of gaze [18], and can be summarized as the speaker's origo.

*1) The Origo:* Bühler introduces the term ”origo” [17] which is conceptually conceived as the origin of a ”coordinate system of subjective orientation”. It is derived from the need of a ”basic reference point” [18] in a given deictic space, which includes three deictic dimensions [19]: the *personal dimension*, the *spatial dimension*, and the *temporal dimension*. Thus the origo is defined by the personal mark 'I', the spatial mark 'here' and the temporal mark 'now'. Accordingly the origo is the basic reference point for referring to other elements within these three deictic dimensions.

*2) Transfer to HRI: Origo-Evaluation:* As the origo is the basic reference point for orientation within space it also constitutes the basis for evaluating expectation hypotheses. Given that the robot retrieved absolute information like street-names, it should evaluate its origo at each critical point on its way, e.g. shifting directions, by asking if it actually reached that certain point and thus successfully conducted the underlying route-hypothesis. In the case of conflicting hypotheses this constitutes the starting point for further hypothesis-testing. Moreover, through origo-evaluation, the previous part of one or more hypothesis-graphs can be approved and subsequently turned into a guiding hypothesis, e.g. the mutual initial route of the depicted graphs in Fig. 1.

In the following Section finally all deduced and transferred findings from Sections II, III and IV are combined and modeled towards a general dialog strategy serving as framework for prevention, detection and handling of erroneous information retrieval due to miscommunication in human-robot dialog.

## V. DIALOG STRATEGY FOR HANDLING MISCOMMUNICATION IN HRI

According to TPH the process of perception builds a cycle that is divided into three consecutive stages described in Section IV.B. This cycle can be transferred as a general loop-controlled dialog strategy including the above introduced states of IR in Table II, integrated on *stage 2: IR from human information input*. This dialog strategy is suitable for task oriented human-robot dialog by providing the robot with the ability to detect and handle miscommunication by evaluating recently obtained knowledge and comparing it to new retrieved information within dialog in a natural way. In the case of inconsistency or misalignment it will be able to initiate well-targeted questions without turning the dialog redundant as outlined in Section IV.

**Stage 1: Provision of expectation hypotheses**

On its way to a predefined goal location the robot looks for humans in order to approach them. As outlined in Section IV.B the robot's initial expectation hypothesis $H_0$ is to extract and organize relevant route information from human speech input including landmarks and directions.

**Stage 2: IR from human information input**

The robot addresses a passer-by and opens the dialog by asking for directions. Thereby, as outlined in Section IV.A the robot passes three consecutive *states of IR*. While passing these states miscommunication can affect each state of IR (see Table II).

*State 1 of IR: speech recognition* – If miscommunication occurs in this state, it can be categorized as *non-understanding*. As a consequence, the robot will not be able to extract relevant task information from the human speech input in state 2, e.g. landmarks and directions. In order to handle non-understanding, the robot must choose targeted questions related to the corresponding subtype of non-understanding (see Section II and Table I) until it is able to recognize the spoken input and extract relevant route information.

*State 2 of IR: extraction of relevant information* – In this state miscommunication is assigned to the subtype of *misunderstanding* and means that the robot extracts information in another way than inteded by the human speaker. If miscommunication is detected within dialog it can be corrected by the interlocutors. If not, the robot builds a route graph representing the extracted information including mistakes and either grounds it as expetation hypothesis $H_1$ for the first dialog partner or as alternative hypothesis $H_2 + n$ for another dialog partner in order to be evaluated.

*State 3 of IR: representation of extracted information* – If miscommunication affects this state, the corresponding subtype is *misconception*. That means, the robot already extracted all relevant information but his representation of

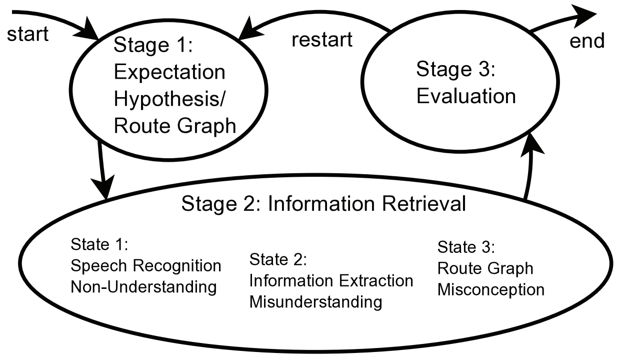

Fig. 2. Cycle of the dialog strategy as a framework for preventing, detecting and handling miscommunication in human-robot dialog.

the gained knowledge, the route graph, either doesn‘t meet the intention of the human dialog partner or it conflicts with its own expectation hypothesis $H_1$ gained from previous dialogs. In both cases, informational misalignment arises. If the human interlocutor disagrees with the robot route graph, it can be corrected during the dialog. However, if the retrieved information conflicts with the hypothesis-graph $H_1$ it builds an alternative $H_2$ for evaluation.

**Stage 3: Confirmation/ Disproof of the Hypothesis**

The first step is to execute the hypothesis with the highest current strength value until a critical point, e.g. shifting directions or a point of discrepancy with an alternative hypothesis, and evaluate the origo by well-directed questions about the current location as explained in section IV.C. If origo-evaluation is successful the route graph representing the way to this point turns into a guiding hypothesis and the robot may proceed with the evaluation of one or more on-going hypotheses via hypothesis-testing as outlined in section IV.B. If origo-evaluation is not successful the related hypothesis is rejected and the strength value of alternative hypotheses increases respectively.

The cycle as illustrated in Fig. 2 starts again with every rejected hypothesis and every extension of the guiding hypotheses through IR by asking human passers-by for directions in order to test a new expectation hypothesis.

## VI. CONCLUSIONS

As reasoned in [20] dialogues should be interpreted in terms of informational alignment rather than information transfer. Thus, handling miscommunication is essential for a spoken dialog system. As miscommunication can be divided into three different categories occurring on three different states of understanding, these states have now been transferred to states of information retrieval (see Table II). These states in turn, have been embedded into the second stage of a dialog strategy, consisting of three consecutive stages (see Fig. 2), derived from linguistsic models and cognitive theories. Thus, an integrated framework is built in order to render an autonomous speech-enabled robot capable of evaluation and repairing informational alignment with humans. As a first step, this is a theoretical framework for handling miscommunication in HRI and needs to be evaluated. Future work will reveal the practicability of the developed strategy implemented into a spoken dialog system, structuring the dialog as depicted in Fig. 2 within real-setting applications.

## VII. ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of the TUM Institute for Advanced Study (IAS) and all reviewers' comments. This work is supported in part within the DFG excellence initiative research cluster *Cognition for Technical Systems – CoTeSys*, see also `www.cotesys.org`, and within the European Commission as part of the *IURO project*, see also `www.iuro-project.eu`.

## REFERENCES

[1] S. Thrun, “Toward a framework for human-robot interaction,” *Human-Computer Interaction*, vol. 19, no. 1, pp. 1–24, 2004.

[2] W. Klein, *Spatial Orientation Theory: Research, and Application.* New York: Plenum, 1983, ch. Spatial orientation in route directions, pp. 283–311.

[3] J. Boye and M. Wirn, “”multi-slot semantics for natural-language call routing systems,” in *Proc. Workshop on Bridging the Gap: Academic and Industrial Research in Dialog Technologies*, 2007, p. 6875.

[4] J. Allen, L. Schubert, G. Ferguson, C. Hwang, T. Kato, M. Light, B. Miller, M. Poesio, and D. Traum, “The trains project: a case study in building a conversational planning agent,” *Journal of Ecperimental and Theoretical Artificial Intelligence*, vol. 7, pp. 7–48, 1995.

[5] J. Rickel and W. Johnson, “Steve: A pedagogical agent for virtual reality (video),” in *Proc. Int. Conf. on Autonomous Agents.* Minneapolis: ACM PRess, 1998.

[6] 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.

[7] 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.

[8] G. Skantze, “Exploring human error recovery strategies: Implications for spoken dialogue systems,” *Speech Communication*, vol. 45, no. 3, pp. 325 – 341, 2005.

[9] T. Koulouri and S. Lauria, “Exploring miscommunication and collaborative behaviour in human-robot interaction,” in *Proc. SIGDIAL Conference.* NJ: Ass. for Comp. Linguistics, 2009, pp. 111–119.

[10] A. Bauer, B. Gonsior, D. Wollherr, and M. Buss, “Dialog system for human-robot interaction for asking for directions, submitted to *Computational Linguistics*, 2010.

[11] G. Hirst, S. McRoy, P. Heeman, P. Edmonds, and D. Horton, “Repairing conversational misunderstandings and non-understandings,” *Speech Communication*, vol. 15, no. 3-4, pp. 213–229, 1994.

[12] M. Gabsdil, “Clarification in spoken dialogue systems:,” in *Proc. of 2003 AAAI Spring Symposium Workshop on Natural Language Generation in Spoken and Written Dialogue*, Stanford, USA, 2003.

[13] H. H. Clark and E. F. Schaefer, “Contributing to discourse,” *Cognitive Science*, vol. 13, pp. 259–294, 1989.

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

[15] W. Lilli and D. Frey, *Theorien der Sozialpsychologie [Theories of Social Psychology]*, 2nd ed. Bern: Hans Huber Verlag, 1993, vol. 1: Cognitive Theories, pp. 49–78.

[16] W. Wirth, T. Hartmann, S. Bcking, P. Vorderer, C. Klimmt, and H. e. a. Schramm, “A process model of the formation of spatial presence experiences,” *Journal of Media Psychology*, vol. 9, p. 493525, 2007.

[17] K. Bühler, *Theory of language. The representational function of language.* Benjamins, 1934/1990.

[18] W. Klein, *Speech, Place, and Action: Studies in Deixis and Related Topics.* New York: Wiley, 1982, ch. Local Deixis in Route Directions, pp. 161–182.

[19] S. Anderson and E. Keenan, *Language typology and syntactic description, Vol. 3: Grammatical categories and the lexicon.* Cambridge University Press, 1985, ch. Deixis, pp. 259–308.

[20] M. J. Pickering and S. Garrod, “Alignment as the basis for successful communication,” *Research on Language and Computation*, vol. 4, no. 2-3, pp. 203 – 228, October 2006.