Robust Natural Language Dialogues for Instruction Tasks

Matthias Scheutz

Human-Robot Interaction Laboratory
Indiana University, 1910 E 19th Street, Bloomington, IN 47401, USA

ABSTRACT

Being able to understand and carry out spoken natural instructions even in limited domains is extremely challenging for current robots. The difficulties are multifarious, ranging from problems with speech recognizers to difficulties with parsing disfluent speech or resolving references based on perceptual or task-based knowledge. In this paper, we present our efforts at starting to address these problems with an integrated natural language understanding system implemented in our DIARC architecture on a robot that can handle fairly unconstrained spoken ungrammatical and incomplete instructions reliably in a limited domain.

Keywords: Spoken instruction understanding, human-robot interaction, integrated robotic architecture

1. INTRODUCTION

Natural language is the hallmark of human communicative exchanges and an important means of social coordination among humans working in teams. As such, it is a main goal of architecture and interaction design in human-robot interaction (HRI) research to develop robotic architectures that can effectively interact with humans in spoken natural language. However, spoken natural language understanding on robots is difficult for a variety of reasons, from problems with word error rates of speech recognizers, to difficulties connected to parsing disfluent speech or syntactically ambiguous sentences, to resolving references based on perceptual, task-based or other pragmatic knowledge. Moreover, language processing and action execution have to be very tightly integrated to allow robots to provide the kind of "back-channel feedback" that humans expect (e.g., verbal or gestural acknowledgments) and to enable robots to perform active perception (e.g., to resolve a reference while an utterance is still going on using other sensory modalities, such as vision). Ultimately, we would like spoken natural language systems on robots that are capable of dealing with incomplete, ungrammatical and disfluent utterances in real-time in a human-like fashion. While clearly a desirable goal, we are currently very far from even coming close to reaching it. Hence, we take achieving human-like natural language interactions as an important challenge in HRI that needs to be addressed, rather than a reason to postpone or even abandon work on natural language understanding on robots in HRI settings.

In this paper, we will provide a brief overview of our recent efforts in developing a robust natural language understanding system as part of an integrated robotic architecture. Most notably, the natural language understanding system is able to perform incremental understanding taking perceptual and task-based knowledge into account and can handle simple natural task-based dialogues. We start with several examples of task-based human-human dialogue interactions from a prior study that illustrate some of the challenges for spoken instruction understanding and execution on robots. Then we briefly introduce our DIARC architecture and demonstrate its capabilities with four example dialogues that showcase how DIARC can handle disfluencies, ambiguity, as well as task-based and natural dialog with verbal and multi-modal acknowledgments.

2. HUMAN PERFORMANCE IN A SEARCH TASK

It is difficult, if not impossible, to anticipate *a priori* the wide variety of spoken language interactions that will be initiated by humans during interactions with robots, even when the task is well-defined. To investigate the kinds of natural language interactions that might take place during human-robot interactions in a typical collaborative search task, we conducted several human-human studies.[1] In these studies, two humans must coordinate with each other via remote audio communication only (no visual links) to accomplish several tasks. In particular, one

Email: mscheutz@indiana.edu

person, the "director", uses a map to direct another person, the "member", through an unfamiliar environment to locate and perform various actions on target objects (colored boxes) scattered throughout the environment. Conversely, the member has to report the location of any green boxes encountered in the environment, which the director will mark on the map. In the following, we give four example dialogues that show different linguistic, perceptual, and cognitive challenges with which a robot performing the member function in the search task would have to deal.

```
Member:   okay um I see a green box with a one on it
Director: okay where is that located at?
Member:   um it's located right next to the doorway on the left on the floor
Director: on the left of the door?
Member:   yeah uh towards the . second room
Director: okay
Member:   yeah that doorway
Director: okay
Member:   um, okay
```

Most notably, this dialogue exhibits a recurrent aspect of most exchanges, namely repeated forms of acknowledgment and confirmation, including different versions of "okay", "yeah", "um okay", and others. These acknowledgments may differ in prosody, either to emphasize a particular referent as in "yeah that doorway", or to denote different functions in the dialogue (e.g., a brief short "okay" to mean "got it", or a stretched "okay" to indicate that the action finished as the utterance ended, or a rising "okay" indicating that the speaker needs more information). The above dialogue also shows syntactic ambiguity as in "right next to the doorway on the left on the floor" (where there are different prepositional attachments for "on"?) together with an ambiguous confirmation by the member ("towards the second room"), reflecting an implicit assumption on the part of both that the director has knowledge of the approximate location and perspective of the member.

```
Director: so where are you at now?
Member:   the - so I went back into the . bigger room . with the desk
Director: yes, right
Member:   um, on the bookshelf in the corner of the room
Director: yes
Member:   uh on the bottom shelf is the fifth box
Director: on the right or left hand side?
Member:   on the right hand side
Director: that's the fifth box?
Member:   yeah
Director: okay
```

In addition to various disfluencies like "uh" and "um" (which often indicate cognitive load), the above dialogue shows examples of interactions that violate typical assumptions such as a "where" question being answered with a description of a trajectory instead of a referent ("the - so I went ..."). And we see assumptions about the accuracy of the map, as in "on the bookshelf in the corner" (where it is assumed by the member that the director's map has a bookshelf marked on it and that there is only one).

```
Director: uh go back until you see the next pink box which should be: uh I guess
Member:   right before the corner
Director: your right hand side
Member:   yeah
Director: yeah, then go ahead and put one in there
Director: now that's all of them you have right?
Member:   yeah, all the yellow blocks?
```

```
Director: so
Member:   okay
Director: yes
```

The above dialogue shows several frequent features including task-based references such as "put one in" and "that's all of them", which assume knowledge about one of the task goals: to put yellow blocks in pink boxes. Note that without task knowledge it would not be possible to resolve the reference to "one" or "them" here. We also see contractions as such as "all of them you have", which is not about "having" them, but about *not* having them any more (i.e., "having put them in pink boxes"). And we see *hedged explains* (e.g., "the next pink box which should be right before the corner"), which are commands that use expressions such as "you should", "there ought to be", "it might", etc., to indicate that the member should check something and report back for verification. This request for verification usually requires the member to execute a sequence of actions (e.g., walking down the hallway to the corner, looking for the pink box, and then informing the director).

```
Director: okay wh-where are you at?
Member:   um I just entered the first room
Director: the first room-there should be like a wooden platform
Member:   yeah
Director: okay
Director: uh, I guess keep going straight unless you see anything
```

In addition to frequent additions of superfluous words that do not contribute meaning and must be omitted for parsing and interpretation, such as "like" (which could be a meaningful preposition in other contexts), we see an example of perspective-taking, as in "the first room" or "keep going straight", which again assumes the director's knowledge of the member's orientation or position in the environment. Moreover, we see additional sentences that cannot be understood properly without task-based knowledge, as in "keep going straight unless you see anything", where "anything" clearly does not literally mean "any thing" but rather means "green box" or "blue box" based on the task instructions.

In sum, we found

- *ungrammatical sentences*, including incomplete referential phrases, missing verbs, corrections, and others
- *wrong word substitutions for intended target words* such as "block" and "book" for "box"
- *underspecified directions, referents, and directives* which assume shared task-knowledge, knowledge of sub-goals, perspectives, etc.
- *frequent "ums", "uhs", and other disfluencies and pauses* indicating cognitive load
- *frequent coordinating confirmations and acknowledgments* as dialogue moves including prosodically different "okays", "yeahs", etc.

3. CHALLENGES FOR SPOKEN INSTRUCTION UNDERSTANDING AND EXECUTION

The above dialogues are typical examples for the kinds of coordinating natural language interactions humans exhibit in limited domains like the search task. The ten most frequent word types in the entire corpus (excluding acknowledgments) are `the`, `box`, `I`, `a`, `'s`, `there`, `you`, `on`, `and`, and `that`. The 151 most frequent words cover 60% of the tokens, 280 words reach 80% coverage, and 712 reach complete coverage. Yet, despite their seemingly very limited nature (based on vocabulary), the dialogues present a major challenge for HRI. It is clear that meanings are not constructed from sentences alone but from interactions that serve particular purposes and accomplish particular goals. Perception, action, and language processing in humans are obviously all intertwined, involving complex patterns of actions, utterances, and responses, where meaningful linguistic fragments result from their context together with prosodic, temporal, task and goal information, and *not* sentence boundaries. Consequently, we need to develop new models of interactive natural language processing and understanding for HRI that process language in very much the same interactive, situated, goal-oriented way as humans. In particular, we need to integrate the timing of utterances, back-channel feedback, perceivable context (such as objects, gestures, eye gaze of the participants, posture, etc.), as well as background and discourse knowledge, task and goal structures, etc., if we want to achieve human-level performance on robots. This includes both *functional challenges* and *architectural challenges.*

Functional challenges include (1) mechanisms for providing appropriate feedback that humans expect even while an utterance is still going on, using different kinds of acknowledgment based on dialogue moves; (2) new algorithms for anaphora and reference resolution using perceptual information as well as task and goal context; and (3) mechanisms for handling various kinds of disfluencies and incomplete and ungrammatical utterances, including robust speech recognition, parsing, and semantic analysis.

Architectural challenges include (1) real-time processing of all natural language interactions within an human-acceptable response time (e.g., typically acknowledgments have to occur within a few hundred milliseconds after a request); (2) integration of various natural language processing components (including speech recognition, parsing, semantic and pragmatic analyses, and dialogue moves) that allows for parallel execution and incremental multi-modal constraint integration; and (3) automatic tracking of dialogue states and goal progress to be able to provide meaningful feedback and generate appropriate goal-oriented dialogue moves.

While each of the above functional challenges is a research program in its own right and we are nowhere close to addressing any of them in a satisfactory manner, it is still possible to make progress in parallel on the architectural challenges, i.e., on defining appropriate functional components in the architecture with appropriate data structures and information flow among them to facilitate the integration of the algorithms that will meet the functional challenges. In the next section, we first provide a brief overview of our DIARC architecture which is our attempt at defining an architecture for natural HRI that can address the above architectural challenges, and then provide several examples of dialogue interactions that show the capabilities of DIARC.

4. EXAMPLES OF ROBUST NATURAL LANGUAGE DIALOGUES FOR COORDINATED MIXED INITIATIVE TASKS

The DIARC "distributed integrated affect cognition reflection" architecture is our attempt at providing an architecture for natural HRI.[2] It integrates cognitive processing, such as natural language understanding[3–5] and complex action planning and sequencing,[6–8] with lower-level activities, such as multi-modal perceptual processing using color and shape detection,[9–11] object detection and tracking using SIFT features,[12] face detection and person tracking,[13] gesture and pointing behavior detection, navigation, and overall behavior coordination.[14] DIARC has been showcased at AAAI robot competitions[15, 16] and used successfully in a variety of HRI experiments with human subjects.[3, 17–20]

DIARC is implemented in the distributed multi-agent robotic development infrastructure ADE,[21–23] which allows for the distribution of architectural components over multiple hosts and provides support for automatic error detection and recovery,[24] thus addressing critical *real-time* and *robustness* constraints required of robotic architectures for HRI. Moreover, ADE systems can be easily extended by "wrapping" existing software as "ADE agents" and adding them to an ADE system,[22] addressing questions concerning *extension* and *scalability.* Finally, ADE provides interfaces to all widely-used robotic environments,[21] allowing for the seamless integration and re-use of components available in these systems.[*]

We will now present four example dialogues to illustrate different aspects of typical human-like dialogues currently possible in DIARC. All dialogues have been performed on different types of robots.

[*] ADE and most DIARC components are freely available at `http://ade.sourceforge.net/`.

4.1 Handling disfluencies

The scenario here is taken from the search task[1] where the robot has to search its environment to report the location of green boxes to the human team leader. The search task here could be a generic version of a "search and rescue scenario" where the robot has to search a building for wounded humans and "green boxes" could serve as proxies for wounded humans. The robot has prior knowledge that green boxes might be hidden in rooms, so it will enter rooms whenever it detects doorways to see whether it can find green boxes. Once it finds a green box, it generates a report to inform the team leader of the location of the box. The following example shows how the robot might receive an instruction to go to a particular location in the environment (as seen in the human experiments) while having to pursue in parallel the goal of reporting the locations of green boxes. In particular, it gives an example of an instruction with disfluencies and corrections that consists of multiple actions that the robot needs to take.

```
Human: Ah, turn around and go to, ah, first, no, go to the end of the hallway.
Robot: Okay.
Human: I guess you have, ah, seven minutes.
Robot: Okay.
```

Here, the robot needs to filter disfluencies and extract the intended instruction–"turn around and go to the end of the hallway within the next seven minutes"–and execute it. After employing various disfluency filters (from finite-state machine and regular expression filters to eliminate lexical fillers and parsing filters to discard corrected elements, see[25]), the robot generates a logical action description `turn(180);goto(endof(currrent-hallway))` where `turn` and `goto` are action primitives, `endof` is a function and `current-hallway` denotes the current hallway (see[5] for details). It also generates a goal description with a timeout value of 7 minutes and associates the goal with the action description. It then instantiates an *action manager* that will schedule actions in pursuit of the goal based on its priority and utility values (the utility value here is inherited from the utility of the "follow-instruction" supergoal, for details see[2]). Note that the robot immediately acknowledges understanding in both cases using "okay" (with falling prosody to indicate the "got it" meaning of "okay" described in the previous section).

4.2 Handling semantic ambiguities

The next scenario shows how the robot can handle semantic ambiguities given by perceptions as well as singular and plural cases. In this case, the robot has just entered a room in which there are two yellow blocks (one in a blue and one in a green box) when it receives the command to get one.

```
Human: Get the yellow block.
Robot: Which one?
Human: Get any yellow block.
Robot: Okay.
```

The robot determines that the reference to "the yellow block" is not unique and consequently asks for clarification. Upon hearing "any yellow block" it picks one of the two yellow blocks and retrieves it. Contrast this with the command "get the yellow blocks". In this case, the robot needs to determine that it has to get both blocks and produce a plan to first get one and then the other. Note that the only syntactic difference between the two commands is in the final word "block". Semantically, however, the logical forms that are automatically generated are quite different. Specifically, the determiner "the" has two meanings associated with it depending on whether it denotes a singular or plural construction. In the singular, the associated meaning is $\lambda Y \exists x.Y(x) \wedge \forall y.Y(y) \rightarrow x = y$, while in the plural it is $\lambda Y\{x|Y(x)\}$. Note that in the first case the meaning is a formula, while in the second case it is a term. Assuming "yellow" is defined as $\lambda Z \lambda z.yellow(z) \wedge Z(z)$, the translation of "the yellow block" amounts to $\exists x.yellow(x) \wedge block(x) \wedge \forall y.(yellow(y) \wedge block(y) \rightarrow x = y)$, while the meaning of "the yellow blocks" is captured by $\{x|yellow(x) \wedge block(x)\}$. In the singular case, the robot finds the unique referent in its visual memory and replaces the formula with a new constant `yb1` denoting it, yielding the final action expression `get(yb1)`. In the plural case, the robot replaces the set expression by an enumeration over all yellow blocks in its visual memory `FORALL(`ϕ`(x),get(x))` where $\phi = block(x) \wedge yellow(x)$.

4.3 Handling task-based dialogues with acknowledgment

The next example also builds on the search task[1] and demonstrates the kinds of natural task-based dialogues found in the human experiments. In particular, the example shows how the robot can determine what objects to look out for and what facts about the environment to report based on its knowledge of the task. It also demonstrates how the robot can quickly react to a given instruction by changing its behavior based on a tight integration between natural language understanding and action execution[4] while providing the necessary feedback to the human commander.

In the following scenario, the robot has been going down one corridor and has stopped outside of a doorway.

```
Human: Is there a hallway?
Robot: I see a hallway.
Human: Okay, go down there.
Robot: Okay.
```

The robot drives down the hallway. As it is driving down it notices a doorway, which it reports to the team leader, also acknowledging its position.

```
Robot: Okay, I'm now in the hallway.  There is a doorway on the left.
Human: Good, go through that doorway.
Robot: Okay.
```

The robot enters the room through the doorway and notices several yellow blocks, some of which are in boxes. Since these are task-relevant, it reports them to the team leader.

```
Robot: I'm now inside the room.  There are yellow blocks in boxes.
Human: Get a yellow block from a blue box.
Robot: Okay.
```

The robot verifies that there is a yellow block in a blue box, approaches the blue box, and gets the yellow block from the blue box.

4.4 Handling multi-modal natural spoken dialogues

The final example shows interactions that take place in a setup in which an autonomous mobile robot with a humanoid torso is engaged by a human interlocutor in a simple dialogue that explores the robot's perceptual capabilities (such as its ability for perspective taking). In this case, because the robot has a humanoid torso, acknowledgments include both verbal and non-verbal components, reflecting the fact that humans also often use multiple channels to convey information (e.g., utterances in conjunction with head motions to indicate agreement or disagreement).

```
Human: Hello robot.
Robot: Hello.
Human: Come over here.
```

The robot uses sound source localization to determine the location of the human speaker to resolve the meaning of the indexical "here" as part of the action `move-to`(*here*) (i.e., the meaning of "here" in this context is the location of the speaker).

```
Human: Do you see a green, no, red box?
Robot: (shakes head) No, I don't see anything red.
```

The robot here manages to handle the correction “green, no, red” correctly and searches for a red box. Since it does not see a red box, and, in fact, sees nothing red, it generates a head motion together with an utterance stating that.

```
Human: Look to the left.
Robot: (looks to its left)
Human: Now do you see it?
Robot: (nods head) Yes, I see a red one.
```

The robot takes the instruction to look to the “left” to mean “left from its perspective” and turns its head in that direction. It also ignores the word “now” at the beginning of the following sentence and resolves the pronoun reference of “it” to refer to the “red box” from the previous sentence. Upon noting the red box now in its visual field, it nods its head and generates another anaphoric reference using “one” instead of “box”.

```
Human: Is the red box to the left of the blue box?
Robot: (nods head) Yes, it is.
Human: Well, ah, is it to the right of the blue box?
Robot: (shakes head) No, it is not.
```

Again, the robot handles anaphoric reference correctly and also handles fillers (“well”) and disfluencies (“ah”).

```
Human: Point to the blue box.
Robot: (points to the blue box) It’s over there.
Human: (moves out of robot’s visual field) Look at me.
Robot: (looks at the human)
Human: Thank you.
Robot: (smiles) You’re welcome.
```

The robot accompanies its pointing action with verbal feedback using the indexical “there”, which in this case indicates the location of the pointing action. Moreover, it resolves the indexical reference of “me” in “look at me” and uses sound localization again to resolve the destination of the “look” command (i.e., the location of the speaker, who has moved out of the robot’s field of vision – had the speaker remained within the robot’s visual field, the robot could have used vision to orient its head to face the speaker).

5. RELATED WORK

Spoken instruction understanding and execution on robots has attracted increasing attention by researchers from various communities both within and outside of HRI, including roboticists, AI researchers, and researchers working on dialogue systems. Given the importance of spoken language, is it unsurprising that several other research groups are pursuing spoken instruction understanding for HRI. Online NLP imposes a substantial burden for processing speed (because people expect acknowledgement very soon after completing an utterance, as noted above); other projects also take an incremental processing approach to address this problem, some integrated with robotic architectures,[26] others comprising standalone systems.[27] The interactive “give-and-take” nature of human spoken language evident from our human-human experiments is a focus of other projects, including systems that allow a human team member to teach the robot a complex task via spoken natural language.[28] And other groups are pursuing the typically multi-modal aspects of human communication, including instruction via a combination of speech and physical demonstration.[29] Different from our approach, however, these systems do not attempt to handle disfluencies, which (as the human-human corpus demonstrates) are very common in spoken instructions. In short, while important aspects of the spoken instruction problem are being attacked from various angles, there is currently no project that attempts to provide the kind of integrated robotic architecture that is necessary for robotic applications. The approach taken by DIARC combines mechanisms to ensure robustness (e.g., disfluency handling), speed (e.g., incremental processing), naturalness (e.g., establishing common ground by incorporating multi-modal functionality, such as vision processing), and others to come as close as currently possible to natural spoken instruction in the context of HRI.

6. CONCLUSION

In this paper, we demonstrated the challenges faced by a robotic natural language understanding system that is intended for natural human-like dialogue interactions with the robot in HRI contexts. And we briefly introduced our DIARC architecture which is starting to address some of these challenges. Specifically, we reported results of several simple natural dialogues that DIARC can handle together with a brief high-level description of the processes involved in handling them. Clearly, this is only a start and there is a great deal of work ahead of us before we can claim to have reached human-like *natural* spoken language interactions. However, by starting with limited domains and tasks (such as instruction tasks), it is possible to make progress right now toward a not-to-distant future point where the resultant architecture will be ready for transition into real-world application domains.

ACKNOWLEDGMENTS

This work was in part funded by ONR MURI grant #N00014-07-1-1049. The author would also like to acknowledge several people for their work on various parts of the robotic architecture, including Paul Schermerhorn, James Kramer, Tim Brick, Rehj Cantrell, Xuan Wu, and Juraj Dzifcak.

REFERENCES

[1] Scheutz, M. and Eberhard, K., "Towards a framework for integrated natural language processing architectures for social robots," in [*Proceedings of the 5th International Workshop on Natural Language Processing and Cognitive Science*], 165–174 (June 2008).

[2] Scheutz, M., Schermerhorn, P., Kramer, J., and Anderson, D., "First steps toward natural human-like HRI," *Autonomous Robots* **22**, 411–423 (May 2007).

[3] Brick, T. and Scheutz, M., "Incremental natural language processing for HRI," in [*HRI*], 263–270 (2007).

[4] Brick, T., Schermerhorn, P., and Scheutz, M., "Speech and action: Integration of action and language for mobile robots," in [*Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*], (October/November 2007).

[5] Dzifcak, J., Scheutz, M., and Baral, C., "What to do and how to do it: Translating natural language directives into temporal and dynamic logic representation for goal management and action execution," in [*Proc. IEEE International Conference on Robotics and Automation (ICRA'09)*], (May 2009).

[6] Scheutz, M., Eberhard, K., and Andronache, V., "A real-time robotic model of human reference resolution using visual constraints," *Connection Science Journal* **16**(3), 145–167 (2004).

[7] Schermerhorn, P., Benton, J., Scheutz, M., Talamadupula, K., and Kambhampati, R., "Finding and exploiting goal opportunities in real-time during plan execution," in [*IROS 2009*], (2009).

[8] Talamadupula, K., Benton, J., Schermerhorn, P., Kambhampati, R., and Scheutz, M., "Integrating a closed world planner with an open world robot: A case study," in [*ICAPS 2009 workshop on bridging task and action planning*], (2009).

[9] Andronach, V. and Scheutz, M., "Design and experimental validation of a minimal adaptive real-time visual motion tracking system for autonomous robots," in [*IC-AI 2005*], 663–669 (2005).

[10] Scheutz, M., "Real-time hierarchical swarms for rapid adaptive multi-level pattern detection and tracking," in [*Proceedings of the 2007 IEEE Swarm Intelligence Symposium*], 234–241 (2007).

[11] Middendorff, C. and Scheutz, M., "Real-time evolving swarms for rapid pattern detection and tracking," in [*Proceedings of Artificial Life X*], 419–425 (June 2006).

[12] Lowe, D. G., "Object recognition from local scale invariant features," in [*Proceedings of the Seventh International Conference on Computer Vision (ICCV '99)*], **1**, 1150–1157, IEEE (1999).

[13] Scheutz, M., McRaven, J., and Cserey, G., "Fast, reliable, adaptive, bimodal people tracking for indoor environments," in [*IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*], (2004).

[14] Scheutz, M. and Andronache, V., "Architectural mechanisms for dynamic changes of behavior selection strategies in behavior-based systems," *IEEE Transactions of System, Man, and Cybernetics Part B* **34**(6), 2377–2395 (2004).

[15] Scheutz, M., Schermerhorn, P., Middendorff, C., Kramer, J., Anderson, D., and Dingler, A., "Toward affective cognitive robots for human-robot interaction," in [*Proceedings of AAAI 2005 Robot Workshop*], AAAI Press (2005).

[16] Schermerhorn, P., Kramer, J., Brick, T., Anderson, D., Dingler, A., and Scheutz, M., "DIARC: A testbed for natural human-robot interactions," in [*Proceedings of AAAI 2006 Robot Workshop*], (2006).

[17] Crowell, C., Scheutz, M., Schermerhorn, P., and Villano, M., "Gendered voice and robot entities: Perceptions and reactions of male and female subjects," in [*IROS 2009*], (2009).

[18] Schermerhorn, P., Scheutz, M., and Crowell, C., "Robot social presence and gender: Do females view robots differently than males?," in [*HRI 2008*], 263–270 (2008).

[19] Scheutz, M., Schermerhorn, P., Kramer, J., and Middendorff, C., "The utility of affect expression in natural language interactions in joint human-robot tasks," in [*Proceedings of the 1st ACM International Conference on Human-Robot Interaction*], 226–233 (2006).

[20] Schermerhorn, P. and Scheutz, M., "Dynamic robot autonomy: Investigating the effects of robot decision-making in a human-robot team task," in [*IEEE ICMI-MLMI Conference*], (November 2009).

[21] Kramer, J. and Scheutz, M., "Robotic development environments for autonomous mobile robots: A survey," *Autonomous Robots* **22**(2), 101–132 (2007).

[22] Scheutz, M., "ADE - steps towards a distributed development and runtime environment for complex robotic agent architectures," *Applied Artificial Intelligence* **20**(4-5) (2006).

[23] Andronache, V. and Scheutz, M., "ADE - an architecture development environment for virtual and robotic agents," *International Journal of Artificial Intelligence Tools* **12**(2), 251–286 (2006).

[24] Kramer, J. and Scheutz, M., "ADE: A framework for robust complex robotic architectures," in [*IROS*], (2006).

[25] Canrell, R., Scheutz, M., Schermerhorn, P., and Wu, X., "Robust spoken instruction understanding for hri," in [*HRI 2010*], (forthcoming) (2010).

[26] Lison, P. and Kruijff, G.-J. M., "Efficient parsing of spoken inputs for human-robot interaction," in [*Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'09)*], (2009).

[27] Schuler, W., Wu, S., and Schwartz, L., "A framework for fast incremental interpretation during speech decoding," *Computational Linguistics* **35**(3) (2009).

[28] Rybski, P. E., Stolarz, J., Yoon, K., and Veloso, M., "Using dialog and human observations to dictate tasks to a learning robot assistant," *Journal of Intelligent Service Robots - To appear* **1**(2), 159–167 (2008).

[29] McGuire, P., Fritsch, J., Steil, J. J., Roẗhling, F., Fink, G. A., Wachsmuth, S., Sagerer, G., and Ritter, H., "Multi-modal human-machine communication for instructing robot grasping tasks," in [*IROS 2002*], (2002).