Novel Mechanisms for Natural Human-Robot Interactions in the DIARC Architecture

Matthias Scheutz[1], **Gordon Briggs**[1], **Rehj Cantrell**[2], **Evan Krause**[1], **Tom Williams**[1] & **Richard Veale**[2]
Human-Robot Interaction Laboratory
Tufts University[1] and Indiana University[2], USA
{matthias.scheutz,gordon.briggs,evan.krause,thomas_e.williams}@tufts.edu, {rcantrel,rveale}@indiana.edu

Abstract

Natural human-like human-robot interactions require many functional capabilities from a robot that have to be reflected in architectural components in the robotic control architecture. In particular, various mechanisms for producing *social behaviors*, *goal-oriented cognition*, and *robust intelligence* are required. In this paper, we present an overview of the most recent version of our DIARC architecture and show how several novel algorithms attempt to address these three areas, leading to more natural interactions with humans, while also extending the overall capability of the integrated system.

Introduction

The 2005 AAAI robot competition featured an "Open Interaction Event" where robots were supposed to freely and naturally interact with humans. In preparation for the event, we described an envisioned waiter scenario (reminiscent of the 1999 AAAI "Hors d'Oeuvres Anyone?" robot competition) and introduced a novel hybrid robotic architecture called DIARC (short for "Distributed Integrated Affect, Reflection, and Cognition" architecture), which had been under development in our lab for several years and was employed on our robot in the competition. DIARC was specifically intended to make the robot's interactions more "natural" (Scheutz et al., 2005), integrating novel algorithms for affective computing and incremental natural language processing.

Subsequently, we introduced the theme of "natural human-like human-robot interaction" or "natural HRI", for short, to lay out a framework and research program that would allow for the development of intelligent robots that could interact with humans *in natural ways* (Scheutz et al., 2007). In the context of HRI, we defined "natural" to roughly mean that "any restrictions on possible interactions are due to human capacities (i.e., the limitations of human perceptual, motor, or cognitive system), and not the a priori functionality of the robot.[..] including interactions that can occur in any typical human setting, such as the ability to use language freely in any way, shape, or form; to make reference to personal, social, and cultural knowledge; or to involve all aspects of human perception and motor capabilities." (Scheutz et al., 2007). As a corollary, robots will have to respect human timings of actions such as the timing of eye movements and eye gaze, gestures and other bodily expressions, natural language understanding and production, automatic situation- and knowledge-based inferences, and many others. For example, it is critical that backchannel feedback during natural language interactions be fast enough and occur at the appropriate places during a speaker's utterance. Similarly, it is critical that eye gaze patterns underlying joint attention processes exhibit the appropriate coupling and decoupling among interactants. Failure to do so will in the best case result in unnatural interactions, but in the worst case in complete interaction break-down and frustration on the side of the human. Critically, failing to respect the human timing might cause significant changes in the human interactant's cognitive processes (e.g., we discovered in an HRI eye-tracking study that the human's allocation of attention was significantly altered throughout the experiment when the robot failed to establish eye contact at the appropriate time (Yu, Scheutz, and Schermerhorn, 2010).

Among the many areas required for natural HRI, we previously highlighted three areas: (C1) *social behaviors*, (C2) *goal-oriented cognition*, and (C3) *robust intelligence*. The first class was intended to include natural language capabilities, but also affective computing as well as non-linguistic interactions (e.g., such as gestures or joint attention). The second class was focussed on intentional behavior, including both the robot's ability to explicitly express and pursue goals, as well as any mechanisms that would allow humans to perceive the robot as an intentional agent. And the third class was intended to include various monitoring mechanisms that would allow the robot to detect all kinds of faults and recover from them (e.g, crashed component or miscommunications with interlocutors).

While significant progress had been made in intelligent robots by 2005 (e.g., compared to the 1990s, see the discussion in Scheutz et al. 2007) which was in part demonstrated at the competition, it was also clear that "none of the systems that competed in 2005 (including ours) has demonstrated the second requirement for natural interaction with humans in real-world environments: the ability to demonstrate and recognize intent." (Scheutz et al., 2007). Unfortunately, the development of integrated mechanisms for inferring human intent and tracking interlocutors' mental states in the context of human-robot interaction has remained a challenge to

this day, and few current architectures for intelligent robots attempt to address this problem. One of them is the current version of our DIARC architecture, which has seen significant developments over the last decade in all three of the above mentioned areas. In addition to addressing the second class by way of introducing explicit mechanisms for building and maintaining mental models of interlocutors and handling indirect speech acts that require the recognition of intent (Briggs and Scheutz, 2011, 2012b, 2013), the first class has also been addressed by developing novel mechanisms for robust task-based dialogue interactions (Cantrell et al., 2010; Scheutz, Cantrell, and Schermerhorn, 2011), including a tight integration between vision and natural language processing (Cantrell et al., 2012a; Krause et al., 2013), and the third class has been addressed by developing fault detection and recovery mechanisms (Kramer and Scheutz, 2007a) as well as novel notification mechanisms for multi-level introspection (Krause, Schermerhorn, and Scheutz, 2012).

The goal of the present paper then is to present an overview of the architectural changes and novel capabilities of the integrated DIARC architecture in all three areas. We start with a brief overview of DIARC and the robotic middleware ADE (Scheutz, 2006) in which it is implemented. Then we focus on our developments for all three classes – social behaviors, goal-oriented cognition, and robust intelligence – briefly describing for each class the new functional capabilities. We also briefly summarize the various diverse application domains in which DIARC has been successfully employed.

The DIARC Architecture for Natural Human-Robot Interactions

The DIARC architecture for natural human-robot interaction has been under development in our lab for more than a decade and uniquely integrates typical (lower-level) robotic capabilities (for visual perception, laser-based mapping and localization, navigation, and others) with (higher-level) cognitive capabilities such as robust incremental natural language understanding (Brick and Scheutz, 2007; Dzifcak et al., 2009; Cantrell et al., 2010), task-based dialogue interactions (Scheutz, Cantrell, and Schermerhorn, 2011; Briggs and Scheutz, 2013), task-based planning (Talamadupula et al., 2010), one-shot learning of actions and plan operators from natural language dialogues (Cantrell, Schermerhorn, and Scheutz, 2011; Cantrell et al., 2012b), mental modeling and belief inference (Briggs and Scheutz, 2012b, 2011), and others. DIARC also deeply integrates various *affect mechanisms* that bias goal prioritization, action selection, behavior arbitration and general deliberative processing (Scheutz and Schermerhorn, 2009; Schermerhorn and Scheutz, 2009b; Scheutz et al., 2006), in addition to modifying speech and facial expressions of the robot.

Analogous to other intelligent robotic architectures, in particular, cognitive architectures, DIARC makes several theoretical commitments, which we will briefly discuss:

- All processing in architectural components occurs asynchronously to other components (e.g., as in the subsumption architecture), and no assumptions can be made about messages passed between components (e.g., about their timely arrival). This is viewed as a feature, rather than a constraint, that allows for the distribution of architectural components over multiple computational hosts compared to classical cognitive architectures that are monolithic.
- Each component operates on a "cognitive cycle" (the "loop time") and may run multiple threads of control within itself (e.g., perception, natural language processing, and action execution are examples of highly parallelized components). This is in contrast to common cognitive architectures (e.g., ACT-R and SOAR) where processing occurs in one overall cognitive cycle.
- There is no centralized controller ("homunculus" as the philosophers call it) such as the central rule interpreter commonly used in production systems and thus classical cognitive architectures. Rather, control is distributed and any synchronized activity must be accomplished by way of multiple components working together in a fault-tolerant, robust fashion.
- Goals are explicitly represented in terms of pre-, operating-, and post-conditions, have a priority that is computed based on the goal's urgency, expected utilities and overall affective state (which, in turn, is computed for each component based on its operation). Goals are attached to skills that accomplish them, which can be retrieved based on their post-condition and executed.
- Action selection is distributed and priority-based (using goal priorities for resource allocation and Behavior arbitration) and uses locking mechanisms for mutual exclusion of resources (e.g., robot effectors).
- No single architectural learning mechanism is prescribed; rather, different forms of learning can occur in different components (e.g., statistical learning in components close to perception and action, symbolic learning in higher-level components).
- No single knowledge representation is prescribed; rather, knowledge representations may take different forms within components depending on the nature of the process operating on them (e.g., saliency maps inside the vision processing component, dependency graphs in the parser, clauses in the reasoner,...).
- Logical expressions are used as a common currency and data representation format across components wherever possible (e.g., between the natural language and vision processing subsystems) and are used as part of introspective access to system features and capabilities.
- Architectural components can tightly interact with the underlying implementation platform, the ADE middleware, using the ADE notifications mechanisms which allow for introspection, monitoring, and discovery of system features and failures (to our knowledge, this is different from any other current architecture for intelligent robotic systems).

DIARC is implemented in the "Agent Development Environment" ADE (Scheutz, 2006) which was conceptualized and developed as an implementation environment for future

complex distributed robotic architectures addressing various challenges posed by natural real-time human-robot interactions. Analogous to other current robotic infrastructures (such as JAUS, Player/Stage, Yarp, ROS, and others)[1], ADE provides the basic communication and computational infrastructure for parallel distributed processes which together implement various functional components of an agent architecture (e.g., the interfaces to a robot's sensors and actuators, the basic navigation and manipulation behaviors, path and task planning, perceptual processing, natural language parsing, reasoning and problem solving, etc.). ADE also provides interfaces to all widely-used robotic environments (Kramer and Scheutz, 2007b), allowing for the seamless integration and re-use of their components. In a running ADE system, all participating ADE components start up independently and connect to each other as prescribed in the architecture diagram – the blue-print of the system – to allow for information flow among architectural components. Different from all other robotic infrastructures, ADE was *designed to be as secure, fault-tolerant and scalable as possible*.

Being implemented in ADE, DIARC can run in a parallel distributed fashion with its components distributed over multiple computing hosts (based on host availability) to address the critical *real-time* and *robustness constraints* required of intelligent robots in HRI domains. Different from other infrastructures and architectures, we have been able to repeatedly demonstrate the robustness of ADE in light of various system faults during task performance in human-robot interaction experiments (Kramer and Scheutz, 2007a, 2006). Moreover, different from many other intelligent agent architectures, DIARC can be easily extended by new architectural components that are implemented via ADE components "wrapping" existing software. These ADE components can then be added to an existing ADE system (Scheutz, 2006) and become available as services to existing components in DIARC, addressing important questions concerning the *extendability* and *scalability* of intelligent agent architectures.

Novel Architectural Extensions for Natural Human-Robot Interactions

Natural human-robot interactions require robots to be capable of a great variety of social behaviors, most critically spoken natural language dialogues. We will thus first and foremost address our developments on situated natural language processing, which in the contexts of joint tasks, for example, requires robots to handle the typical disfluencies and infelicities of spontaneous speech exhibited by humans. Moreover, dialogue interactions (and, a fortiori, task-based interactions) require robots to maintain and update a mental model of their interlocutors. Hence, we will describe our work on developing a pragmatic framework that integrates natural language understanding with mental belief modeling. And finally, sustained interactions require mechanisms to cope with various types of errors and failures. We thus also briefly point to our work on introspection, fault detection, and fault recovery.

[1]A detailed comparison of robotic infrastructures up to 2006 can be in found in Kramer and Scheutz (2007b).

Robust Natural Language Interactions

To facilitate robust social interactions, (1) the natural language understanding (NLU) system must handle a wide range of inputs, using failures (such as unknown words and miscommunications) as learning opportunities; and (2) spoken inputs must be grounded to actions, locations, and agents the robot knows what to do with. The first requirement is handled by a robust trainable dependency parser with online learning capabilities, while the second is addressed by systematic integration with the robot's perceptual and cognitive capabilities.

To facilitate robust speech recognition, we have developed a neural any-time speech recognizer using a liquid-state machine (LSM) back-end and successfully applied it to the recognition of both short and longer phrases (Veale and Scheutz, 2012). The practical advantages of our approach for real-world applications are (1) the ability to access results at any time and (2) the robustness of the recognition in somewhat noisy environments. Intermediate predictions can be accessed at any time during an utterance, enabling early actions based on the recognizers predictions, or allowing for the biasing of other cognitive components (such as parsers, visual search systems, etc.) based on the current best-guess. Conversely, the recognizer can be biased in parallel in real-time based on other information available from other perceptual modalities or top-down information (Veale, Briggs, and Scheutz, 2013). The system is also robust in certain situations because, unlike traditional Markov-model based speech recognizers, the neural circuit may find highly non-linear relations between very different parts of phrases, which will be used to separate the phrases at the holistic level (in contrast to having to break down sound into phonemes and using n-gram windows).

Robust parsing of spoken inputs requires (1) the ability to handle verbal disfluencies such as interjections and repetitions (Cantrell et al., 2010) and (2) the ability to learn new vocabulary items on the fly. The NLU Component (NLUC) of ADE uses an incremental dependency parser trainable from annotated corpora. Because it does not require fully-connected parses, it is not strongly impeded by spoken disfluencies such as interjections and repetitions. As the NLUC identifies syntactic dependencies, it uses a dictionary of known concepts and their semantic valancies to produce semantic representations. If the identified structure does not match the system's expectation based on the dictionary — or if the system has no expectations because the word is unknown — the robot can request missing information, and can learn either a new word or a new valancy for a previously-known word.

However, a semantic representation is useless unless it is grounded in the robot's physical and cognitive environment. For example, the name of an action is not useful unless the robot understands (can perform, recognize or plan with) the action. Thus the robot also learns new grounded concepts from spoken interaction, for example it can ask questions to elicit a procedural definitions for verbs (Cantrell, Schermerhorn, and Scheutz, 2011) and information for use in planning (Cantrell et al., 2012b). Similarly, mentioned entities must be grounded. The NLUC incrementally grounds refer-

ences by collecting and storing information about the entity to which each noun phrase refers. If the noun phrase ultimately appears to refer to a perceived feature of the environment or to co-refer with a previously-known entity, information from both structures is merged.

In order to resolve referents in the environment, the NLUC is integrated with perceptual components. When an interlocutor begins a noun phrase, the NLUC requests that components such as vision begin to search the environment. Each search is incrementally modified as more information about that noun phrase is collected (e.g., as adjectives and nouns are heard). All descriptors (adjectives and nouns) are sent to each perceptual component (unless specifically restricted); unknown descriptors are ignored. When the end of the noun phrase is identified, the NLUC notifies perception to end the search and receives the final list of matching objects. The NLUC then compares the number of matching entities it sees with the expected number (for example, "the red block" should have a single referent, "the red blocks" should have multiple referents, and "all blocks" or "any blocks" may have any number of referents including zero). If there is a discrepancy (e.g., multiple referents were returned for "the red block"), the NLUC can mention this problem: "There is more than one red block".

Referents that are known but not immediately present in the environment (e.g., previously mentioned objects or locations) are also identified. For example, the NLUC is integrated with the SPatial EXpert (SPEX) (Williams et al., 2013), which reconciles place descriptions whose semantics it receives from the NLUC with its representations of places previously mentioned in dialogue or perceived by the robot's sensors. SPEX maintains a map of the robot's environment which can be used to perform spatial reference resolution for the NLUC. For example, given the semantics for the phrase "the third room on the right," SPEX returns a reference to a known room matching that description if one already exists in SPEX's world model. Other components that can be queried in this manner include belief, which maintains a list of previously-discussed or encountered entities.

Finally, if these methods fail, the NLUC can resolve references to hypothesized or inferred entities. Some mentioned objects imply other objects; for example, when a door is mentioned, the NLUC assumes it is attached to a room. Components such as SPEX enlarge their world models based on these new objects. For example, if a location is described that does not match any on the map, SPEX enlarges the map with a hypothesized unknown location and returns a reference to it. Thus the robot's world model is updated not only from exploration but also from dialogue, a novel ability compared to previous systems (e.g., Matuszek, Fox, and Koscher 2010), which have been unable to update their world model after the system's initial training phase or tour. This allows for more natural interaction between the robot and its interlocutor, as navigation, discovery and dialogue can occur concurrently.

Mental Modeling and Indirect Speech Acts

For goal-oriented cognition and social cognition the ability to model the mental states of an interaction partner is vital to successful interaction. This ability requires not only a means of representing such a mental model, but also the ability to make inferences about an agent's mental state based on observed communicative acts (e.g. speech acts, gestures) in addition to non-communicative acts (e.g. physical task-relevant actions). Also, not only must one be able to model the mental states of others, but one must also pro-actively communicate one's own beliefs and intentions to one's interaction partners using similar linguistic and non-linguistic mechanisms. Below we will describe the progress made in DIARC to construct mechanisms that assist in mental state inferences in cooperative contexts– as well mechanisms that allow for natural and human-like communication of one's own beliefs and intentions.

One mechanism found in human-human interaction is the use of certain linguistic cues to communicate beliefs about the mental state of the interactant, specifically certain adverbial modifiers such as "yet", "still" and "now." For example, one would not say "are you at the store yet?" if he or she did not believe his or her interlocutor had a goal to or anticipated being at the store. In Briggs and Scheutz (2011) we provided the first formal pragmatics for sentences containing adverbial modifiers that link pragmatic representations to mental states of interlocutors (e.g., expected goals or currently held beliefs). In addition, not only did we develop pragmatic rules to infer the beliefs of an agent's interlocutor based on use of adverbial cues, we devised an utterance selection algorithm for natural language generation that would appropriately select utterances with the correct adverbial modifier based the agent's own belief (Briggs and Scheutz, 2011).

Another phenomenon in human-human interaction are indirect speech acts (ISAs). These include indirect requests such as "Could you get me a coffee?" or "I would like a coffee" (as opposed to a direct request, "get me a coffee!"). Not only did we develop rules to understand indirect requests, but we also extended our utterance selection method and natural language generation system to select socially appropriate request forms (which may or may not be indirect) based on formalization of social roles, obligations, and context. Additionally, we developed plan-reasoning mechanisms that could assist with making sense of indirect answers – that is, responses to questions that do not answer the immediate question, but convey understanding of the broader goal and assist with the completion of that goal. For instance, if one were to ask, "Do you know where the meeting room is?", an example of an indirect answer would be, "Follow me!" (Briggs and Scheutz, 2013). This extends previous work in understanding indirect requests (Wilske and Kruijff, 2006; Perrault and Allen, 1980) and generating indirect requests (Gupta, Walker, and Romano, 2007).

Underlying the abilities to handle these are general principles and rules for updating mental models, which we will outline here. We will use τ to denote the interlocutor, ρ to denote the robot, and $[[..]]_c$ to denote the "pragmatic meaning" of an expression (e.g., a natural language utterance) in some context c, which often includes task and goal information, as well as beliefs (about the interlocutor, perceptions, objects, etc.) and discourse aspects (about the previous in-

teractions with the interlocutor). Overall, updates to robot's mental model of the interlocutor will be triggered by various events, mediated through the robot's perceptual system. For example, the robot might perceive a new task-relevant object. Assuming that all agents store such perceptions, we can formulate a general principle that if an agent α perceives an object o at location l at time t, then α will believe (B) that it perceived o at l at t:

$$Perceives(\alpha, o, l, t) \Rightarrow B(\alpha, Perceives(\alpha, o, l, t))$$

Another example would include natural language utterance from the interlocutor, such as "can you get me a coffee?", which in certain social contexts c should be treated as a request. In addition to updating its own beliefs, the robot ρ needs to model the interlocutor's τ beliefs in response to its utterances, (i.e., ρ has to derive its mental model $\{\psi | B(\tau, \psi) \in Bel_\rho\}$) and update it by using them same rules it applies to its own beliefs. The same is true when ρ notices that τ has certain perceptions or performs certain actions. The above principles (and a few related ones) have already been successfully integrated into a special belief modeling component part of DIARC and evaluated in simple human-robot interactions (Briggs and Scheutz, 2012b, 2011).

Introspection and Architectural Adaptation

To support long-term sustained interactions, ADE provides advanced architectural mechanisms for system-wide simulation and introspection on component services that can be used to detect faults and failures (Kramer and Scheutz, 2007a), but also missing competencies, and it allows the robot to discover new capabilities at run-time. Specifically, ADE provides support for fine-grained architectural simulations by *duplicating* all architectural components and running the duplicated architecture with a simulated robotic body in a simulated environment which is initiated based on the state of the current environment. As a result, it is possible for agents to discover that they are missing low-level capabilities to complete a task such as certain perceptual algorithms or motor primitives that are not explicitly represented in any part of their system (e.g., in a self-model). Moreover, ADE supports the discovery of dynamic changes to the architecture, including the addition of new capabilities at run-time.[2]

ADE also integrates introspection mechanisms at all levels of the architecture: agent-level, infrastructure-level, and component-level. Typically, self-adjusting agent architectures employ either component-specific introspection mechanisms or attempt to integrate all levels of self-reasoning and self-adjustment into a single system-wide mechanism (Morris, 2007; Haidarian et al., 2010; Sykes et al., 2008; Georgas and Taylor, 2008). Our approach combines the benefits of system-wide and component-specific mechanisms and allows for self-observation, self-analysis, and self-adjustment at all three levels of the architecture. These mechanisms not only provide the typical failure detection and recovery and system reconfiguration, but have also been show to provide performance improvements. In a concrete implementation on a robot, we demonstrate how the high-level goal of the agent is used to automatically reconfigure the vision system to minimize resource consumption while improving overall task performance (Krause, Schermerhorn, and Scheutz, 2012).

Applications of DIARC

Robotic architectures for human-robot interaction can be employed in two important ways (in addition to being deployed on robots in application domains): as *experimental tools* and as *computational models*. In the first case, the architecture is used to study human social cognitive processes that unfold in real-time and where interaction responses depend on past behaviors ("contingent experimental design"), including the evaluation of human responses to and interactions with future robots. In the second case, the architecture is used to develop and implement computational models of "situated embodied cognition" which focuses on the role of body situated in an environment for understanding cognition, including the evaluation of different human-machine interfaces and their efficacy.

As an experimental tool, DIARC has been run on various robotic platforms to study

- Joint attention processes (e.g., establishing and maintaining joint attention, or breaking joint attention through "abnormal attention") (e.g., see Yu, Scheutz, and Schermerhorn, 2010; Yu, Schermerhorn, and Scheutz, 2012)
- Human attitudes about robots (e.g., social facilitation and social inhibition to probe agency, or investigations of the effects of robotic voices, social presence, etc.) (e.g., see Crowell et al., 2009; Schermerhorn, Scheutz, and Crowell, 2008)
- Human reactions to autonomous robots in cooperative tasks (e.g., to robot affect, robot autonomy, to local/remote HRI) (e.g., see Schermerhorn and Scheutz, 2011; Scheutz et al., 2006)
- Robot ethics (e.g., whether humans will accept robots that ignore commands in the interest of team goals or that point out unethical aspects of commands) (e.g., see Briggs and Scheutz, 2012a; Schermerhorn and Scheutz, 2009a)
- Philosophical and conceptual inquiry (e.g., what it is like to be an agent/have a red experience, or the effects of "ethical robots" on human decision-making)

As a model, DIARC has been used to demonstrate various advanced capabilities:[3]

- Spoken natural language and dialogue interactions (e.g., instructing and tasking in natural language, dialogue-based mixed initiative, robust NL interactions under time

[2] See, for example, the demonstration video at http://www.youtube.com/watch?v=9KLwELMatcg which shows a robot that simulates itself performing a task to discover that it has no ability to perform a visual search; but when it is given the missing camera, it immediately notices that it is now capable of completing the task and consequently resumes the task right away.

[3] For videos of DIARC in operation, see http://www.youtube.com/user/HRILaboratory/.

pressure) (e.g., see Cantrell et al., 2010; Scheutz, Cantrell, and Schermerhorn, 2011)

- Planning, reasoning, and problem solving in open worlds (e.g., planning and reasoning with incomplete knowledge, determining optimal policies in open worlds) (e.g., see Talamadupula et al., 2010; Joshi et al., 2012)
- Knowledge-based learning (e.g., one-shot learning of new actions, new plan operators, and new perceivable objects) (e.g., see Cantrell, Schermerhorn, and Scheutz, 2011; Cantrell et al., 2012b)
- Mental models, simulation, and counterfactual reasoning (e.g., adverbial cues for inferring false beliefs, automatic inference from mental models, simulations of actions) (e.g., see Briggs and Scheutz, 2011, 2013)
- Introspection and self-awareness (e.g., detecting faults and failures, detecting capabilities, automatic adaptation of architectural components for improved autonomy) (e.g., see Kramer and Scheutz, 2007a; Krause, Schermerhorn, and Scheutz, 2012)

Conclusion

We provided an overview of the current version of the DIARC architecture for human-robot interaction and described some of its novel extensions that are intended to address critical shortcomings of prior versions of the architecture in an effort to allow for more natural human-robot interactions. Specifically, we reviewed developments on the situated natural language processing side, the mental modeling and intent inference algorithms, and the mechanisms for multi-level introspection and fault tolerance mechanisms. We also provided a brief summary of the diverse application domains in which DIARC has been employed. In the architectural development phase, DIARC's component model will be extended to improve introspection capabilities that will increase the "self-awareness" of the architecture and thus its ability to predict intended and unintended future states. These predictions will then be used to adapt the robot's behavior in a way that will make its interactions more reliable and robust, and improve its overall performance. Most implemented DIARC components implemented in ADE are freely available for download from http://ade.sourceforge.net/.

Acknowledgments

This work was in part funded by NSF grant 111323 and ONR grants #N00014-07-1-1049 and #N00014-11-1-0493.

References

Brick, T., and Scheutz, M. 2007. Incremental natural language processing for hri. In *Proceedings of the Second ACM IEEE International Conference on Human-Robot Interaction*, 263–270.

Briggs, G., and Scheutz, M. 2011. Facilitating mental modeling in collaborative human-robot interaction through adverbial cues. In *Proceedings of the SIGDIAL 2011 Conference*, 239–247.

Briggs, G., and Scheutz, M. 2012a. Investigating the effects of robotic displays of protest and distress. In *Proceedings of the 2012 Conference on Social Robotics*, LNCS, 238–247. Springer.

Briggs, G., and Scheutz, M. 2012b. Multi-modal belief updates in multi-robot human-robot dialogue interactions. In *Proceedings of AISB 2012*.

Briggs, G., and Scheutz, M. 2013. A hybrid architectural approach to understanding and appropriately generating indirect speech acts. In *Proceedings of AAAI*.

Cantrell, R.; Scheutz, M.; Schermerhorn, P.; and Wu, X. 2010. Robust spoken instruction understanding for HRI. In *Proceedings of the 2010 Human-Robot Interaction Conference*.

Cantrell, R.; Krause, E.; Scheutz, M.; Zillich, M.; and Potapova, E. 2012a. Incremental referent grounding with nlp-biased visual search. In *Proceedings of AAAI 2012 Workshop on Grounding Language for Physical Systems*.

Cantrell, R.; Talamadupula, K.; Schermerhorn, P.; Benton, J.; Kambhampati, S.; and Scheutz, M. 2012b. Tell me when and why to do it!: Run-time planner model updates via natural language instruction. In *Proceedings of the 2012 Human-Robot Interaction Conference*.

Cantrell, R.; Schermerhorn, P.; and Scheutz, M. 2011. Learning actions from human-robot dialogues. In *Proceedings of the 2011 IEEE Symposium on Robot and Human Interactive Communication*.

Crowell, C.; Scheutz, M.; Schermerhorn, P.; and Villano, M. 2009. Gendered voice and robot entities: Perceptions and reactions of male and female subjects. In *IROS 2009*.

Dzifcak, J.; Scheutz, M.; Baral, C.; and Schermerhorn, P. 2009. 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 *Proceedings of ICRA*.

Georgas, J. C., and Taylor, R. N. 2008. Policy-based self-adaptive architectures: a feasibility study in the robotics domain. In *Proceedings of the 2008 international workshop on Software engineering for adaptive and self-managing systems*, SEAMS '08, 105–112. New York, NY, USA: ACM.

Gupta, S.; Walker, M. A.; and Romano, D. M. 2007. How rude are you?: Evaluating politeness and affect in interaction. In *Affective Computing and Intelligent Interaction*. Springer. 203–217.

Haidarian, H.; Dinalankara, W.; Fults, S.; Wilson, S.; Perlis, D.; Schmill, M.; Oates, T.; Josyula, D.; and Anderson, M. 2010. The metacognitive loop: An architecture for building robust intelligent systems. In *PAAAI Fall Symposium on Commonsense Knowledge (AAAI/CSK'10)*.

Joshi, S.; Schermerhorn, P.; Khardon, R.; and Scheutz, M. 2012. Abstract planning for reactive robots. In *Proceedings of the 2012 IEEE International Conference on Robotics and Automation*.

Kramer, J., and Scheutz, M. 2006. ADE: A framework for robust complex robotic architectures. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 4576–4581.

Kramer, J., and Scheutz, M. 2007a. Reflection and reasoning mechanisms for failure detection and recovery in a distributed robotic architecture for complex robots. In *Proceedings of the 2007 IEEE International Conference on Robotics and Automation*, 3699–3704.

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

Krause, E.; Cantrell, R.; Potapova, E.; Zillich, M.; and Scheutz, M. 2013. Incrementally biasing visual search using natural language input. In *Proceedings of AAMAS*.

Krause, E.; Schermerhorn, P.; and Scheutz, M. 2012. Crossing boundaries: Multi-level introspection in a complex robotic architecture for automatic performance improvements. In *Proceedings of AAAI*.

Matuszek, C.; Fox, D.; and Koscher, K. 2010. Following directions using statistical machine translation. In *Proceeding of the 5th ACM/IEEE international conference on Human-robot interaction*, HRI '10, 251–258. New York, NY, USA: ACM.

Morris, A. C. 2007. *Robotic Introspection for Exploration and Mapping of Subterranean Environments*. Ph.D. Dissertation, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA.

Perrault, C. R., and Allen, J. F. 1980. A plan-based analysis of indirect speech acts. *Computational Linguistics* 6(3-4):167–182.

Schermerhorn, P., and Scheutz, M. 2009a. Dynamic robot autonomy: Investigating the effects of robot decision-making in a human-robot team task. In *IEEE ICMI-MLMI Conference*.

Schermerhorn, P., and Scheutz, M. 2009b. The utility of affect in the selection of actions and goals under real-world constraints. In *Proceedings of the 2009 International Conference on Artificial Intelligence*.

Schermerhorn, P., and Scheutz, M. 2011. Disentangling the effects of robot affect, embodiment, and autonomy on human team members in a mixed-initiative task. In *Proceedings of the 2011 International Conference on Advances in Computer-Human Interactions*, 236–241.

Schermerhorn, P.; Scheutz, M.; and Crowell, C. 2008. Robot social presence and gender: Do females view robots differently than males? In *HRI 2008*, 263–270.

Scheutz, M., and Schermerhorn, P. 2009. Affective goal and task selection for social robots. In Vallverdú, J., and Casacuberta, D., eds., *Handbook of Research on Synthetic Emotions and Sociable Robotics: New Applications in Affective Computing and Artificial Intelligence*. Idea Group Inc. 74–87.

Scheutz, M.; Schermerhorn, P.; Middendorff, C.; Kramer, J.; Anderson, D.; and Dingler, A. 2005. Toward affective cognitive robots for human-robot interaction. In *Proceedings of AAAI 2005 Robot Workshop*. AAAI Press.

Scheutz, M.; Schermerhorn, P.; Kramer, J.; and Middendorff, C. 2006. 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.

Scheutz, M.; Schermerhorn, P.; Kramer, J.; and Anderson, D. 2007. First steps toward natural human-like HRI. *Autonomous Robots* 22(4):411–423.

Scheutz, M.; Cantrell, R.; and Schermerhorn, P. 2011. Toward humanlike task-based dialogue processing for human robot interaction. *AI Magazine* 32(4):77–84.

Scheutz, M. 2006. ADE—steps towards a distributed development and runtime environment for complex robotic agent architectures. *Applied Artificial Intelligence* 20(4-5):275–304.

Sykes, D.; Heaven, W.; Magee, J.; and Kramer, J. 2008. From goals to components: a combined approach to self-management. In *Proceedings of the 2008 international workshop on Software engineering for adaptive and self-managing systems*, SEAMS '08, 1–8. New York, NY, USA: ACM.

Talamadupula, K.; Benton, J.; Kambhampati, S.; Schermerhorn, P.; and Scheutz, M. 2010. Planning for human-robot teaming in open worlds. *ACM Transactions on Intelligent Systems and Technology* 1(2):14:1–14:24.

Veale, R., and Scheutz, M. 2012. Neural circuits for any-time phrase recognition. In Miyake, N.; Peebles, D.; and Cooper, R. P., eds., *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, 1072–1077. Austin, TX: Cognitive Science Society.

Veale, R.; Briggs, G.; and Scheutz, M. 2013. Linking cognitive tokens to biological signals: Dialogue context improves neural speech recognizer performance. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, in press. Austin, TX: Cognitive Science Society.

Williams, T.; Cantrell, R.; Briggs, G.; Schermerhorn, P.; and Scheutz, M. 2013. Grounding natural language references to unvisited and hypothetical locations. In *Proceedings of AAAI*.

Wilske, S., and Kruijff, G.-J. 2006. Service robots dealing with indirect speech acts. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, 4698–4703. IEEE.

Yu, C.; Schermerhorn, P.; and Scheutz, M. 2012. Adaptive eye gaze patterns in interactions with human and artificial agents. *ACM Transactions on Interactive Intelligent Systems* 1(2):13.

Yu, C.; Scheutz, M.; and Schermerhorn, P. 2010. Investigating multimodal real-time patterns of joint attention in an hri word learning task. In *Proceedings of the 2010 Human-Robot Interaction Conference*.