

More Than Just a Pretty Face: Affordances of Embodiment

J. Cassell, T. Bickmore, H. Vilhjálmsón, H. Yan

Gesture and Narrative Language Group

MIT Media Laboratory

E15-315

20 Ames St, Cambridge, Massachusetts

+1 617 253 4899

{justine, bickmore, hannes, yanhao}@media.mit.edu

ABSTRACT

Prior research into embodied interface agents has found that users like them and find them engaging. In this paper, we argue that embodiment can serve an even stronger function if system designers use actual human conversational protocols in the design of the interface. Communicative behaviors such as salutations and farewells, conversational turn-taking with interruptions, and referring to objects using pointing gestures are examples of protocols that all native speakers of a language already know how to perform and that can thus be leveraged in an intelligent interface. We discuss how these protocols are integrated into Rea, an embodied, multi-modal conversational interface agent who acts as a real-estate salesperson, and we show why embodiment is required for their successful implementation.

INTRODUCTION

There is a qualitative difference between face-to-face conversation and other forms of human-human communication [4]. Businesspeople and academics routinely travel long distances to conduct certain face-to-face interactions when electronic forms of communication would seemingly work just as well. When people have something really important to say, they say it in person.

The qualitative difference in these situations is not just that we enjoy looking at humans more than at computer screens but also that the human body enables the use of certain communication protocols in face-to-face conversation which provide for a more rich and robust channel of communication than is afforded by any other medium available today. The use of gaze, gesture, intonation, and body posture play an essential role in the proper execution

of many conversational functions—such as conversation initiation and termination, turn-taking, interruption handling, feedback and error correction—and these kinds of behaviors enable the exchange of multiple levels of information in real time. People are extremely adept at extracting meaning from subtle variations in the performance of these behaviors; for example slight variations in pause length, feedback nod timing or gaze behavior can significantly alter the interpretation of an utterance (consider "you did a great job" vs. "you did a . . . great job").

Of particular interest to interface designers is that these communication protocols come for "free" in that users do not need to be trained in their use; all native speakers of a given language have these skills and use them daily. An embodied interface agent which exploits these protocols has the potential to provide a higher bandwidth of communication than would otherwise be possible.

Of course, depictions of human bodies are also more decorative than menus on a screen and, like any new interface design, they are also currently quite in vogue and therefore attractive to many users. Unfortunately, many embodied interface agents developed to date don't go further than their ornamental or novelty value. Aside from the use of pointing gestures and two or three facial expressions, many animated interface agents provide little more than something amusing to look at while the same old system handles the mechanics of the interaction. It is little wonder that these systems have been found to be likable and engaging, but to provide no improvement in task performance over text or speech-only interfaces.

In this paper, we first review the embodied interface agents developed to date and summarize the results of evaluations performed on them. We then discuss several human communication protocols along with their interface utility and requirements for embodiment. Finally, we present Rea, an embodied interface agent which implements these

protocols and describe our ongoing research program to develop embodied interface agents that leverage knowledge of human communication skills.

RELATED WORK

Other researchers have built embodied interface agents, with varying degrees of conversational ability. The closest to our own research in this area is the work of Rickel and Johnson [20], Andre and Rist [1], and Lester et al. [17] whose agents do use both verbal and nonverbal conversational behaviors, move to objects in the interface and use pointing gestures in combination with speech or text output. In these systems, however, the association between verbal and nonverbal behaviors is always additive – the affordances of the body are not exploited for the kinds of tasks that it performs better than speech.

"Animated Conversation" [8] was a system that automatically generated context-appropriate gestures, facial movements and intonational patterns. In this case the domain was conversation between two artificial agents and the emphasis was on the production of non-verbal propositional behaviors that emphasized and reinforced the content of speech. However, the system was not designed to interact with a user, and did not run in real time.

The work of Thorisson provides a good first example of how an embodied interface agent inspired by studies of human psychosocial competencies might be developed [22]. The agent, Gandalf, recognized and displayed interactional information such as gaze, simple gesture and canned speech events. In this way he was able to perceive and generate turn-taking and back channel behaviors that lead to a very natural conversational interaction. However, Gandalf had limited ability to recognize and generate propositional information, and was also limited in his ability to provide correct intonation for speech emphasis on speech output, or co-occurring gestures with speech.

The conversational character system developed by Prevost, et al [19], uses the same architecture as the one presented in this paper (it was co-developed by our two research groups), but their application domain and many implementation details are different. In their system a conversational character assists a user with a complex A/V system by controlling equipment, answering questions or giving tutorials. To date, the conversational behaviors of their agent is limited to greeting and farewell rituals, gaze, pointing gestures and body positioning.

User Studies on Embodied Interface Agents

Koda and Maes [16], and Takeuchi and Naito [23], studied user responses to interfaces with static or animated faces, and found that users rated them to be more engaging and entertaining than functionally equivalent interfaces without a face. Kiesler and Sproull found that users were more likely to be cooperative with an interface agent when it had a human face (vs. a dog image or cartoon) [14].

Andre, Rist and Muller found that users rated their animated presentation agent ("PPP Persona") as more entertaining and helpful than an equivalent interface without the agent [1]. However, there was no difference in actual performance (comprehension and recall of presented material) in interfaces with the agent vs. interfaces without it.

In a user study of the Gandalf system mentioned above [9], users rated the smoothness of the interaction and the agent's language skills significantly higher under test conditions in which Gandalf utilized limited conversational behavior (gaze, turn-taking and limited gesture) than when these behaviors were disabled.

Most of these evaluations have tried to address whether embodiment of a system is useful at all, usually by keeping the interaction the same, and then including or not including an animated figure. The studies, then, are not testing how particular uses of embodiment may improve task or learning performance. Therefore, although the previous studies inspire us by showing that the mere presence of a character wins us points, we now need to focus on the contribution of embodiment in fully functional conversational interfaces, and in order to do that, we need to start with a better understanding of what embodiment contributes to human-human interaction.

HUMAN COMMUNICATION PROTOCOLS REQUIRING EMBODIMENT

Embodiment provides us with a wide range of behaviors that, when executed in tight synchronization with language, carry out a communicative function. It is important to understand that particular behaviors, such as the raising of the eyebrows, can be employed in a variety of circumstances to produce different communicative effects, and that the same communicative function may be realized through different sets of behaviors. It is therefore clear that any system dealing with conversational modeling has to handle function separately from surface-form or run the risk of being inflexible and insensitive to the natural phases of the conversation. Here we briefly describe some of the fundamental communication protocols and their functional elements along with examples of nonverbal behavior that contribute to their successful implementation. Table 1 shows examples of mappings from communicative function to particular behaviors and is based on previous research on typical North American nonverbal displays, mainly [11] and [15].

This mapping from form to function relies on a fundamental division of conversational goals: contributions to a conversation can be propositional and interactional. Propositional information corresponds to the content of the conversation. This includes meaningful speech as well as hand gestures used to complement or elaborate upon the speech content (gestures that indicate the size in the sentence "it was this big"). Interactional

information consists of the cues that regulate conversational process and includes a range of nonverbal behaviors (quick head nods to indicate that one is following) as well as regulatory speech ("huh?", "Uh-huh"). This theoretical stance allows us to examine the role of embodiment not just in task- but also process-related behaviors. From this standpoint, we note that most previous embodied interface agents do not deal with interactional and propositional information in an integrated manner, which prevents them from fully exploiting the affordances of the body.

Conversation Initiation and Termination

Humans partake in an elaborate ritual when engaging and disengaging in conversation [18]. For example, people will show their readiness to engage in a conversation by turning towards their potential interlocutor, gazing at the person and then exchanging signs of mutual recognition typically involving a smile, eyebrow movement and tossing the head or waving of the arm. Following this initial synchronization stage, or distance salutation, the two

people approach one other, sealing their commitment to the conversation through a close salutation such as a handshake accompanied by a ritualistic verbal exchange. The greeting phase ends when the two participants re-orient their bodies, moving away from a face-on orientation to stand at an angle. Terminating a conversation similarly moves through stages, starting with non-verbal cues, such as orientation shifts or glances away and cumulating in the verbal exchange of farewells and the breaking of mutual gaze.

Conversational Turn-Taking and Interruption

Interlocutors do not normally talk at the same time, thus imposing a turn-taking sequence on the conversation. The protocols involved in floor management -- determining whose turn it is and when the turn should be given to the listener -- involve many factors including gaze and intonation [10]. In addition, listeners can interrupt a speaker not only with voice, but by gesturing to indicate that they want the turn.

Communicative Functions	Communicative Behavior
<i>Initiation and termination:</i>	
Reacting	Short Glance
Inviting Contact	Sustained Glance, Smile
Distance Salutation	Looking, Head Toss/Nod, Raise Eyebrows, Wave, Smile
Close Salutation	Looking, Head Nod, Embrace or Handshake, Smile
Break Away	Glance Around
Farewell	Looking, Head Nod, Wave
<i>Turn-Taking</i>	
Give Turn	Looking, Raise Eyebrows (followed by silence)
Wanting Turn	Raise Hands into gesture space
Take Turn	Glance Away, Start talking
<i>Feedback</i>	
Request Feedback	Looking, Raise Eyebrows
Give Feedback	Looking, Head Nod

Table 1. Some examples of conversational functions and their behavior realization

Content Elaboration and Emphasis

Gestures can convey information about the content of the conversation in ways that the hands are uniquely suited to fulfill. For example, the two hands can better indicate simultaneity and spatial relationships than the voice or other channels.

Probably the most commonly thought of use of the body in conversation is the pointing (deictic) gesture, possibly accounting for the fact that it is also the most commonly implemented use for the bodies of animated interface agents. In fact, however, most conversations don't involve many deictic gestures [18] unless the interlocutors are discussing a shared task that is currently present. Other

conversational gestures also convey semantic and pragmatic information [5]. Beat gestures are small, rhythmic baton like movements of the hands that do not change in form with the content of the accompanying speech. They serve a pragmatic function, conveying information about what is "new" in the speaker's discourse. Iconic and metaphoric gestures convey some features of the action or event being described. They can be redundant or complementary relative to the speech channel, and thus can convey additional information or provide robustness or emphasis with respect to what is being said. Whereas iconics convey information about spatial relationships or concepts, metaphorics represent concepts which have no

physical form, such as a sweeping gesture accompanying "the property title is free and clear."

Feedback and Error Correction

During conversation, speakers can non-verbally request feedback from listeners through gaze and raised eyebrows and listeners can provide feedback through head nods and paraverbals ("uh-huh", "mmm", etc.) if the speaker is understood, or a confused facial expression or lack of positive feedback if not. The listener can also ask clarifying questions if they did not hear or understand something the speaker said.

REA: AN EMBODIED CONVERSATIONAL AGENT

The Rea project at the MIT Media Lab [6,7] has as its goal the construction of an embodied, multi-modal real-time conversational interface agent. Rea implements the conversational protocols described above in order to make interactions almost as natural as face-to-face conversation with another person. In the current task domain, Rea acts as a real estate salesperson, answering user questions about properties in her database and showing users around the virtual houses (Figure 1).



Figure 1. User interacting with Rea

Rea has a fully articulated graphical body, can sense the user passively through cameras and audio input, and is capable of speech with intonation, facial display, and gestural output. The system currently consists of a large projection screen on which Rea is displayed and which the user stands in front of. Two cameras mounted on top of the projection screen track the user's head and hand positions in space. Users wear a microphone for capturing speech

input. A single SGI Octane computer runs the graphics and conversation engine of Rea, while several other computers manage the speech recognition and generation and image processing

Rea is able to conduct a conversation describing the features of the task domain while also responding to the users' verbal and non-verbal input. When the user makes cues typically associated with turn taking behavior such as gesturing, Rea allows herself to be interrupted, and then takes the turn again when she is able. She is able to initiate conversational error correction when she misunderstands what the user says, and can generate combined voice, facial expression and gestural output. Rea's responses are generated by an incremental natural language generation engine based on [21] that has been extended to synthesize redundant and complementary gestures synchronized with speech output. A simple discourse model is used for determining which speech acts users are engaging in, and resolving and generating anaphoric references.

Architecture

Figure 2 shows the modules of the Rea architecture that is designed to meet the requirements of real-time face-to-face conversation [7]. In this design, input is accepted from as many modalities as there are input devices. However the different modalities are integrated into a single semantic representation that is passed from module to module. This representation is a KQML frame [13] which has slots for interactional and propositional information so that the regulatory and content-oriented contribution of every conversational act can be maintained throughout the system.

The categorization of behaviors in terms of their conversational functions is mirrored by the organization of the architecture which centralizes decisions made in terms of functions (in the Deliberative Module), and moves to the periphery decisions made in terms of behaviors (the Input Manager and Action Scheduler).

In addition the Input Manager and Action Scheduler can communicate through a hardwired reaction connection, to respond immediately (under 200 msec.) to user input or system commands. Tracking the user with gaze shifts as they move is an example of a reactive behavior. The other modules are more "deliberative" in nature and perform non-trivial inferencing actions that can take multiple real-time cycles to complete. Rea is implemented in C++ and CLIPS, a rule-based expert system language [12].

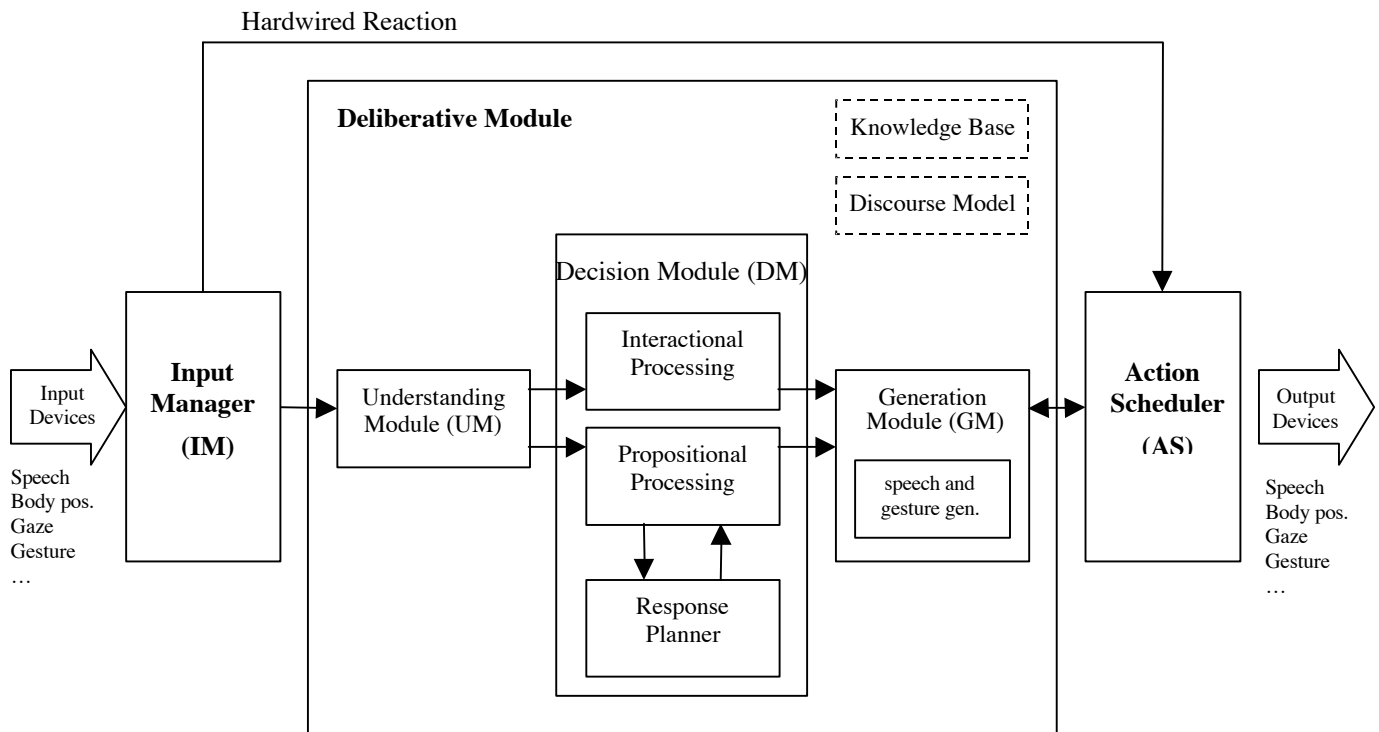


Figure 2. Rea's Software Architecture

Overview of Implemented Communication Protocols

Rea implements the human communication protocols previously described, as follows.

Conversation Initiation and Termination

Rea acknowledges the user's presence through posture, by turning to face the user, as detected by the vision system. She also exchanges greetings and farewells with the user using verbal and non-verbal (gestural) output, in response to the user's verbal greeting and farewell. Rea also recognizes when the user turns away during conversation (based on vision input) and suspends speech input processing until the user turns to face her again.

Conversational Turn-Taking and Interruption

Rea tracks who has the speaking turn (using a conversational state model), and only speaks when she holds the turn. Currently Rea always allows verbal interruption based on audio threshold detection, and yields the turn as soon as the user begins to speak. If the user gestures (as detected by the vision system) she will interpret this as expression of a desire to speak, and halt her remarks at the nearest sentence boundary. She exhibits the "look away" behavior while she is planning her response (which serves to hold the turn until she is ready to speak), and at the end of her speaking turn she turns to face the user to indicate a turn transition point.

Content Elaboration and Emphasis

Rea currently uses pointing gestures to refer to pictures of houses in her environment. Research into generation of pointing gestures to disambiguate objects during virtual house walk-throughs as well as recognition of user pointing gestures is currently being pursued.

Rea generates her natural language responses together with accompanying conversational gestures in a unified text Generation Module. This module distributes the information to be conveyed to the user across the voice and gesture channels based on semantic and pragmatic criteria, resulting in both redundant and complementary gestures. The gestures are composed as a function of hand starting and ending positions, trajectory, hand shape, and envelope size. Beat gestures are generated to mark new information when no semantic information is to be conveyed through the gesture channel. Thus, for example, when describing a new property, Rea says "It has a lovely garden" and demonstrates with her hands that the garden surrounds two sides of the house.

The development of a module to classify the user's conversational gestures, based on input from the vision system and statistical models, is currently underway.

Feedback and Error Correction

Rea provides non-verbal feedback during the user's turn by nodding her head at the end of user utterances (as detected through the audio threshold device) in which the user

keeps the turn. If Rea does not fully understand the user's input (typically due to errors reported by the speech recognition system) she attempts repair by asking a clarifying question.

Sample Interaction in Detail

In order to understand better how Rea processes user input, both propositional and interactional, and produces appropriate output behavior, it is helpful to look at a segment of interaction with a user and describe the messages sent between each of Rea's internal modules. The following paragraph records an actual interaction between a user and Rea:

Tim approaches Rea
 Rea notices and looks toward him and smiles
 Tim says "hello"
 Rea responds: "Hello, how can I help you", with a hand wave
 Tim says "I'm looking to buy a place near MIT"
 Rea glances up and away to keep the turn while "thinking"
 Rea says: "I have a house...", with a beat gesture to emphasis the new information "house"
 Tim interrupts by beginning to gesture
 Rea finishes the current utterance by saying "in Cambridge" and then she gives up the turn.
 Tim asks for details.

We will now focus on how the different modules of Rea's architecture contribute to carrying out this interaction.

All messages are packaged into a KQML tell-performative (Figure 3), where the sender and recipient fields contain the names of the modules communicating. For messages that have to do with describing the interaction between the user and Rea, including all messages in this example, the content field contains a frame of type *commact*. The sender and recipient fields of the commact denote where the communicative action originated and who is the intended recipient of it, the value being either *REA* or *USER*, depending on whether the commact is being interpreted or generated by Rea's Decision Module (DM).

The general processing sequence is as follows: Input Manager (IM) has some new information about the user's actions and creates a new commact with sender *USER* and recipient *REA*. In the input field it places a description of the behaviors detected. The Understanding Module (UM) receives the commact, interprets the behaviors and fills in the prop and intr fields accordingly, sending the commact on to the DM. In reaction to the incoming commact, the DM may construct a new commact, this time with sender *REA* and recipient *USER*. After filling in the prop and intr fields, the DM passes the frame on to the Generation Module (GM) whose job is to translate the propositional and interactional descriptions into a series of low level behaviors to be placed in the output field. Lastly the Action Scheduler (AS) receives the new commact and

using the output field, it coordinates verbal and non-verbal realization.

```
(tell :sender UM :recipient DM :content
  (commact :sender USER :recipient REA
    :input [(speaking :state TRUE)
      (gesturing :state TRUE) ]
    :prop NONE
    :intr [ (takingturn) ]
  )
)
```

Figure 3. A Sample Performative

As a user comes within a few feet of Rea, a stereoscopic vision system starts to track the user's head and hand movements [2]. Upon receiving this information from the IM, the UM sends the DM an interactional message saying that the user is now present. This makes the system transition into the *UserPresent* state (Figure 4), sending off to the GM an interactional request for generating an invitation to start a conversation. The GM maps the request to a sequence of behaviors, that includes a look towards the user and a smile, to be sent to the AS for execution. When the user responds to the invitation by saying "Hello", the IM reports the onset of voice to the UM that sends the DM an interactional message saying that the user has now taken the turn. The system transitions into the *UserTurn* state and stays there until the IM delivers the parsed speech content to the UM and the DM has received from the UM an interactional message saying the user has given up the turn, and a propositional message in the form of a speech act, in this case of the type *SA-RITUAL-GREET*. Inside DM this speech act generates an obligation to respond to the greeting. Since a similar *SA-RITUAL-GREET* speech act in return would fulfil that obligation, the DM sends such an act to the GM for execution. The GM breaks the speech act into a hand wave behavior and the spoken utterance "hello, how can I help you?" to be realized by the AS. The system is momentarily in a *ReaTurn* state while the speech act is performed, but returns back to an *OpenFloor* state when done.

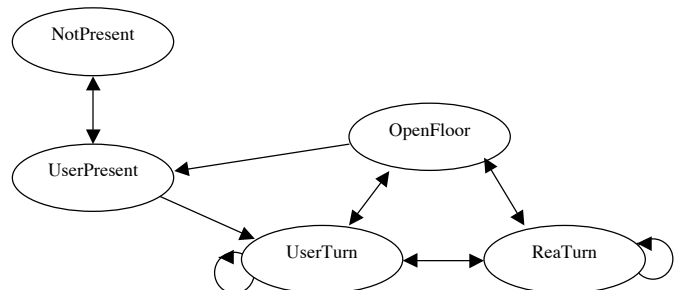


Figure 4. Rea's Conversational States

When the user starts speaking again, the GM produces an interactional message indicating that the user has taken the turn, shifting the system's state to *UserTurn*. As the user finishes asking "I'm looking to buy a place near MIT" the GM gives the DM the interactional message that the user has given up the turn along with the propositional message that the user performed a *SA-REQUEST-PLACE*. The GM also adds "NearMIT" as an attribute that the place has to have in order to be considered. The DM determines that *HOUSE1* meets the user's preferences so the *SA-REQUEST-PLACE* speech act generates an obligation to *OFFER-HOUSE1*. But since this is the first time *HOUSE1* is presented to the user, another obligation *DESCRIBE-HOUSE1* is also generated. Looking at the obligations one at a time, the DM sends off to the GM an *SA-OFFER-HOUSE1* to fulfil the first one. Along with this propositional message, an interactional message stating that Rea also needs to take the turn is sent to the GM. The GM consults a text and gesture generator¹ for generating an appropriate verbal and gestural expression of the proposition while instructing the AS to glance up and away in an effort to take and keep the turn. The user notices that Rea is planning to speak and does not grab the floor, allowing Rea to stay in a *ReaTurn* state. However, as Rea is delivering the utterance generated by SPUD, "I have a house", the user realizes that "near MIT" was perhaps too weak of a constraint and wants to add more detail and therefore spontaneously raises the hands in anticipation of further elaborating on the query. The vision notices the sudden hand movement and the UM sends a message to the DM saying that the user would like the turn. Gesture is treated as a low-priority interrupt, and Rea should finish her current utterance before giving the user the turn, so the DM removes the future obligation to *DESCRIBE-HOUSE1* but allows the GM and AS to continue executing the current utterance. (Had the user interrupted with speech overlapping Rea's, the DM would also have halted the GM and AS execution, causing Rea to give the user the turn immediately.) When Rea finishes her utterance ("...in Cambridge") she looks at the user in a *UserTurn* state and the user continues.

CONCLUSIONS

User-testing of Gandalf, capable of some of the conversational functions also described here, showed that users relied on the interactional competency of the system to negotiate turn-taking, and that they preferred such a system to another embodied character capable of only emotional expression. In fact, users became so comfortable with Gandalf that they began to overlap their speech with his, which overtaxed his limited speech recognition

¹ We are using the SPUD system developed by Matthew Stone, augmented to synthesize conversational gestures in addition to speech in real time [21].

capabilities [9]. Our next step is to test Rea to see whether the implementation of a larger set of conversational functions, including error correction and gesture synthesis, allows users to engage in more efficient and fluent interaction with the system.

In this paper we have argued that embodied interface agents can provide a qualitative advantage over non-embodied interfaces, if the bodies are used in ways that leverage knowledge of human communicative behavior. We demonstrated our approach with the Rea system. Increasingly capable of making an intelligent content-oriented – or propositional – contribution to the conversation, Rea is also sensitive to the regulatory – or interactional -- function of verbal and non-verbal conversational behaviors, and is capable of producing regulatory behaviors to improve the interaction by helping the user remain aware of the state of the conversation.

ACKNOWLEDGEMENTS

Thanks to the other members of the Rea team - Lee Campbell, David Mellis and Nina Yu - and to Jennifer Smith for their contribution to the work and comments on this paper. Thanks to Candy Sidner and several anonymous reviewers for helpful comments that improved the paper.

REFERENCES

1. Andre, E., Rist, T., Mueller, J. Integrating Reactive and Scripted Behaviors in a Life-Like Presentation Agent. *In Proceedings of Agents '98* (Minneapolis/St. Paul, May 1998), ACM Press.
2. Azarbayejani, A., Wren, C. and Pentland A. Real-time 3-D tracking of the human body. *In Proceedings of IMAGE'COM 96*, (Bordeaux, France, May 1996).
3. Ball, G., Ling, D., Kurlander, D., Miller, D., Pugh, D., Skelly, T., Stankosky, A., Thiel, D., Van Dantzich, M. and T. Wax. Lifelike computer characters: the persona project at Microsoft Research. *In Software Agents*, J. M. Bradshaw (ed.), MIT Press, Cambridge, MA, 1997.
4. Boyle, E., Anderson, A., and Newlands, A. The Effects of Visibility in a Cooperative Problem Solving Task. *Language and Speech 37(1)*, 1994. 1-20.
5. Cassell, J. "Embodied Conversation: Integrating Face and Gesture into Automatic Spoken Dialogue Systems." *In Luperfoy (ed.), Spoken Dialogue Systems*. (to appear) Cambridge, MA: MIT Press.
6. Cassell, J., Bickmore, J., Billinghurst, M., Campbell, L., Chang, K., Vilhjálmsón, H., Yan, H. "Embodiment in Conversational Interfaces: Rea", *ACM CHI 99 Conference Proceedings*, Pittsburgh, PA, 1999.
7. Cassell, J., Bickmore, J., Campbell, L., Vilhjálmsón, H., Yan, H. Human Conversation as a System Framework: Designing Embodied Conversational

- Agents, in Cassell, J. editor, *Embodied Conversational Agents*, MIT Press, Cambridge, MA, 1999
8. Cassell, J., Pelachaud, C., Badler, N.I., Steedman, M., Achorn, B., Beckett, T., Douville, B., Prevost, S. and Stone, M. Animated conversation: rule-based generation of facial display, gesture and spoken intonation for multiple conversational agents. *Computer Graphics (SIGGRAPH '94 Proceedings)*, 28(4): 413-420.
 9. Cassell, J. and Thórisson, K. The Power of a Nod and a Glance: Envelope vs. Emotional Feedback in Animated Conversational Agents. *Journal of Applied Artificial Intelligence*, in press.
 10. Cassell, J., Torres, O. and Prevost, S. Turn taking vs. Discourse Structure: how best to model multimodal conversation. In Wilks (ed.) *Machine Conversations*. Kluwer, The Hague, 1998.
 11. Chovil, N. Discourse-Oriented Facial Displays in Conversation. *Research on Language and Social Interaction*, 25, 163-194, 1992.
 12. CLIPS Reference Manual Version 6.0. *Technical Report*, Number JSC-25012, Software Technology Branch, Lyndon B. Johnson Space Center, Houston, TX, 1994.
 13. Finin, T., Fritson, R. KQML as an Agent Communication Language. In *Proceedings of the Third International Conference on Information and Knowledge Management (CIKM'94)*, November 1994), ACM Press.
 14. Keisler, S., and Sproull, L. "Social" Human-Computer Interaction. In B. Friedman (Ed) *Human Values and the Design of Computer Technology*. CSLI Publications, New York, 1997. 191-200.
 15. Kendon, A. *Conducting Interaction: Patterns of behavior in focused encounters*. Cambridge University Press. New York. 1990.
 16. Koda, T., and Maes, P. Agents with faces: The effect of personification. *Proceedings of the Fifth IEEE International Workshop on Robot and Human Communication (RO-MAN '96)*. 189-194.
 17. Lester, J., Converse, S., Kahler, S., Barlow, S., Stone, B., and Bhogal, R. The Persona Effect: Affective Impact of Animated Pedagogical Agents. In S. Pemberton (Ed.), *Human Factors in Computing Systems: CHI'97 Conference Proceedings*. 359-366. New York: ACM Press.
 18. McNeill, D. *Hand and Mind: What Gestures Reveal About Thought*. University of Chicago Press. 1992.
 19. Prevost, S., Hodgson, P., Cook, L., and Churchill, E. Face-to-Face Interfaces. In M. Altom and M. Williams (Eds.), *Human Factors in Computing Systems: CHI'99 Extended Abstracts*. 244-245. New York: ACM Press.
 20. Rickel, J., and Johnson, L. Task-Oriented Dialogs with Animated Agents in Virtual Reality. In *Proceedings of the Workshop on Embodied Conversational Characters* (Tahoe City, California, October 1998). 39-46.
 21. Stone, M. *Modality in Dialogue: Planning, Pragmatics, and Computation*. PhD Thesis, University of Pennsylvania, 1998.
 22. Thórisson, K. R. *Communicative Humanoids: A Computational Model of Psychosocial Dialogue Skills*. PhD Thesis, MIT Media Laboratory, 1996.
 23. Takeuchi, A., and Naito, T. Situated facial displays: Towards social interaction. In I. Katz, R. Mack, L. Marks, M. Rosson, and J. Nielsen (Eds.), *Human Factors in Computing Systems: CHI'95 Conference Proceedings*. 450-455. New York: ACM Press.