Communications of the ACM
(to appear)

More Than Just Another Pretty Face: Embodied Conversational Interface Agents

Justine Cassell
MIT Media Laboratory
E15-315, 20 Ames St
Cambridge, Massachusetts
+1 617 253 4899
justine@media.mit.edu

Introduction[1]

Both animals and humans manifest social qualities. Dogs recognize dominance and submission, stand corrected by their superiors, demonstrate consistent personalities, and so forth. On the other hand, only humans communicate using language, and carry on conversations with one another. And the skills of conversation have developed in humans in such a way as to exploit all of the unique affordances of the human body. We make complex representational gestures with our prehensile hands, gaze away and towards one another out of the corners of our centrally-set eyes, and use the pitch and melody of our flexible voices to emphasize and clarify what we are saying. Perhaps because conversation is so defining of humanness and human interaction, the metaphor of face-to-face conversation has been applied to human-computer interface design for quite some time. One of the early arguments for the utility of this metaphor gave a list of features of face-to-face conversation that could be fruitfully applied to human-computer interaction, including mixed initiative, non-verbal communication, sense of presence, rules for transfer of control ([9]). However, although these features have gained widespread recognition, human – computer conversation has only very recently become more than a metaphor. That is, it is only just recently that designers have taken the metaphor seriously enough to attempt to design a computer that could hold up its end of the conversation.

In this article I describe some of the features of human-human conversation that are being implemented in this new genre of *embodied conversational agents*. Then I describe an embodied conversational agent that is based on these features. I argue that, because conversation is such a primary skill for humans, and such an early-learned skill (practiced, in fact, between infants and mothers who take turns cooing and burbling at one another), and because the body is so well-equipped to support conversation, embodied conversational agents may turn out to be powerful ways for humans to interact with their computers. However, I claim that in order for embodied conversational agents to live up to their promise, their implementations must be based on actual study of human – human conversation, and their architectures must reflect some of the intrinsic properties found there.

Embodied conversational interfaces are not just computer interfaces represented by way of human or animal bodies. And they are not just interfaces where those human or animal bodies are *lifelike* or *believable* in their actions and their reactions to human users. Embodied conversational interfaces are specifically *conversational* in their behaviors, and specifically human-like in the way they use their bodies in conversation. That is, embodied conversational agents may be defined as those that have the same properties as humans in face-to-face conversation, including:

- The ability to recognize and respond to verbal and non-verbal input.
- The ability to generate verbal and non-verbal output.
- The ability to deal with conversational functions such as turn taking, feedback, and repair mechanisms.
- The ability to give signals that indicate the state of the conversation, as well as to contribute new propositions to the discourse.

[1] Research leading to the preparation of this article was supported by the National Science Foundation (award IIS-9618939), AT&T, Deutsche Telekom, and the other generous sponsors of the MIT Media Lab. Heartfelt thanks to Tim Bickmore, Lee Campbell, Kenny Chang, Joey Chang, Sola Grantham, Erin Panttaja, Jennifer Smith, Scott Prevost, Kris Thorisson, Obed Torres, Hannes Vilhjalmsson, Hao Yan, and the other talented and dedicated students and former students who have worked with me on the Embodied Conversational Agents project. Many thanks also to colleague Matthew Stone for his continued invaluable contribution to this work. Finally, thanks to James Lester, Jeff Rickel and Ben Shneiderman for generous comments that improved the paper.

Embodied conversational agents bring a new slant to the argument about whether it is wise to anthropomorphize the interface. It has been shown that humans respond to computers as if they were social entities. Even experienced computer users interact with their computers according to social rules of politeness and gender stereotypes, accept that some computers are more authoritative than others, and that some computers are experts while others are generalists, and in many, many other ways react to the computer as if it were another human ([10]). Nevertheless, the fact that we do react to computers in this way begs the question of whether interface designers should accede to these illogical tendencies by building computers that look like humans. Critics (such as Shneiderman [12]) have asked what function this would serve. He points out that anthropomorphized interfaces have never been successful in the past, and in fact they may even lead to slower response times or confusion on the part of the user. Our response might be to say that only conversational embodiment – giving the interface the appearance *and the function* of the human body in conversation – will allow us to evaluate the function of embodiment in the interface. Simply building anthropomorphized interfaces that talk (but that do not use their talk in human-like ways) will not shed any light on the debate about embodiment. It is my belief that well-designed embodied conversational interface agents will address particular needs that are not met in current interfaces. For example, ways to make dialogue systems robust in the face of imperfect speech recognition, to increase bandwidth at low cost, and to support efficient collaboration between human and machines, and between humans mediated by machines. This is exactly what bodies bring to conversation.

Embodied conversational agents also bring a new dimension to discussions about the relationship between emulation and simulation, and the role of foundational principles that are true to real-world phenomena in the implementation of interfaces that sell. The first wave of interface agents with bodies and autonomous embodied characteristics – often called autonomous synthetic characters -- were not focused on conversation but on more general interactional social skills. Researchers developing these characters discovered, sometimes to their surprise, that believability and lifelikeness may not be best derived from modeling life. Instead, these researchers have found themselves turning to insights from Disney animators and others about caricaturization and exaggeration as a way of getting users to suspend disbelief and attribute reality to an interactive character. Thus, for example, the OZ project at CMU brought artists and actors in early in the development process of their interactive characters to help them convey features of personality in a compelling way [2]).

This design approach cannot be said to have carried as much weight in the development of embodied *conversational* agents. Here, much like the scientists who first began to build dialogue systems to allow computers to understand human language, researchers are finding themselves forced to turn to theories of human-human interaction, and to investigate the nitty-gritty details of conversation as a way of ensuring that their interfaces share the conversational skills of human users. Thus, for example, Lester's COSMO character ([8]) refers to the objects in his environment using pronouns, descriptions and/or pointing gestures, according to a complex algorithm based on the linguistic theory of referential ambiguity. André and Rist ([1]) associate particular gestures to aspects of planning. They generate pointing gestures as a sub-action of the rhetorical action of labelling, in turn a sub-action of the action of elaborating. Similarly, Rickel and Johnson ([11]) have their pedagogical agent move to objects in the virtual world and then generate a pointing gesture at the beginning of an explanation about that object.
And so, rather than engaging in debates about whether anthropomorphization is good or evil, an emphasis on implementing precisely-described and well-motivated characteristics of human conversation into the interface allows researchers who develop embodied conversational agents to turn to the issue of the adequacy of their theories of human behavior for the task of implementing effective interfaces.

So, what are the conversational skills that embodied conversational agents display? Let's start with an actual scenario between a human user and an embodied conversational agent, and then turn to the behaviors that characterize it.

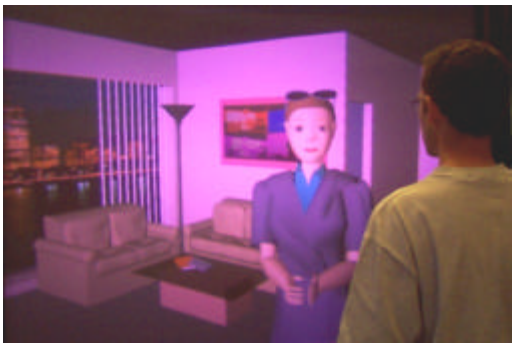

Figure1:User Interacting with Rea

A Sample Interaction

Rea is an embodied conversational agent whose domain of expertise is real estate and she has access to a database of available condominiums and houses for sale in Boston. She can display pictures of those properties and of their various rooms, and point out and discuss their salient features. The following is an excerpt from an actual interaction:

Mike approaches the projection screen. Rea is currently turned side on and is gazing idly about. As Mike moves within range of the two cameras mounted above the screen, Rea turns to face him and says:
REA> Hello. How can I help you?
MIKE> I'm looking to buy a place near MIT.
Rea nods, indicating that she is following.
REA> I have a house to show you.
A picture of a house appears on-screen behind Rea who blinks and looks at the house and then at Mike.
REA> It's in Somerville.
MIKE> Tell me about it.
Rea looks up and away while she plans what to say
REA> It's big.
Rea makes an expansive gesture with her hands.
Mike brings his hands up as if he is about to speak and so Rea does not continue; instead waiting for him to speak.
MIKE> Tell me more about it
REA> Sure thing. It has a nice garden.
Rea sketches a curved gesture with her hands indicating that the garden extends along two sides of the house
MIKE> How far is it?
REA> It is five minutes to the Porter Square T station.
*Rea makes it clear that it is five minutes **on foot** from the T station by making a walking gesture with her fingers*
MIKE> How big is the house?
REA> It has four bedrooms, three bathrooms. . .
Mike interrupts Rea who stops speaking immediately.
MIKE> Wait. Tell me, where is the master bedroom?
REA> I'm sorry, I didn't catch that. What did you ask me?
MIKE> Where is the master bedroom?
REA> It's upstairs.
Rea points up
MIKE> Where is the master bathroom?
REA> It's next to the bedroom.
Rea brings her hands together to indicate the relationship between the bedroom and the bathroom.
And the house tour continues.

Conversational Models

Why is this scenario so exciting or difficult? Because Rea is engaging in some very subtle human-like conversational patterns, and because a set of five properties of human conversation had to be modeled in order for the system to be able to demonstrate those patterns. Those five properties are (a) function rather than behavior, (b) synchronization (c) the division between interactional and propositional contributions, (d) multi-threadedness, and (e) entrainment.

Function rather than Behavior

Even though conversation looks orderly, governed by rules, no two conversations are exactly the same and the set of behaviors exhibited differs from person to person and from conversation to conversation. Therefore to successfully build a model of how conversation works, one cannot refer to surface features, or *conversational behaviors* alone. Instead, the emphasis has to be on identifying the high level structural elements that make up a conversation. These elements are then described in terms of their role or *function* in the exchange. Typical discourse functions include *conversation invitation, turn taking, providing feedback, contrast and emphasis*, and *breaking away*.

This is especially important because particular behaviors, such as the raising of the eyebrows, can be employed in a variety of circumstances to produce different communicative effects, and the same communicative function may be realized through different sets of behaviors. The form we give to a particular discourse function depends on, among other things, current availability of modalities such as the face and the hands, type of conversation, cultural patterns and personal style. Thus, in the dialogue above, in order to indicate that she is listening (that is, as a way of providing feedback), Rea nods. She might instead have said "uh huh" or "I see". Note that in a different context these behaviors may carry a different meaning; for example a head nod can indicate emphasis or a salutation rather than feedback.

Communicative Functions	**Communicative Behavior**
Initiation and termination:	
React to new person	Short glance at other
Break away from conversation	Glance around
Farewell	Look at other, head nod, wave
Turn-Taking	
Give Turn	Look, raise eyebrows (followed by silence)
Want Turn	Raise hands into gesture space
Take Turn	Glance away, start talking
Feedback	
Request Feedback	Look at other, raise eyebrows
Give Feedback	Look at other, nod head

Table 1. Some examples of conversational functions and their behavior realization (taken from [6])

Synchronization

Behaviors that fill the same function, or achieve the same communicative goals, occur in synchrony. This property leads to humans assuming that synchronized phenomena co-carry meaning. That is, the meaning of a nod is determined by where it occurs in an utterance, to the 200 msec scale (consider the difference between "you did a [great job]" (square brackets indicate the temporal extent of the nod) and "you did a [. . .] great job"). Thus, in the dialogue above, Rea *says* "it has a nice garden" at exactly the same time as she sketches the outlines of the garden (in fact, the most effortful part of the gesture, known as the stroke, co-occurs with the noun phrase "nice garden"). The same gesture could mean something quite different if it occurred with different speech, or could simply indicate Rea's desire to take the turn if it occurred during the human user's speech.

Division between Propositional and Interactional Contributions

Contributions to the conversation can be divided into *propositional information* and *interactional information*. Propositional information corresponds to the content of the conversation. This includes meaningful speech as well as hand gestures and intonation used to complement or elaborate upon the speech content (gestures that indicate size in the sentence "it was *this* big" or rising intonation that indicates a question with the sentence "you went to the store"). Interactional information consists of cues that regulate the conversational process and includes a range of non-verbal behaviors (quick head nods to indicate that one is following, bringing one's hands to one's lap and turning to the listener to indicate that one is giving up the turn) as well as regulatory speech ("huh?", "do go on"). In short, the interactional discourse functions are responsible for creating and maintaining an open channel of communication between the participants, while propositional functions shape the actual content. Both functions

may be fulfilled by either verbal or non-verbal means. Thus, in the dialogue excerpted above, Rea's non-verbal behaviors sometimes contribute propositions to the discourse, such as the gesture that indicates that the house in question is five minutes *on foot* from the T stop, and sometimes regulate the interaction, such as the head-nod that indicates that Rea has understood Mike's utterance.

Multi-threadedness

Interactional behaviors tend to be shorter in duration than propositional. In fact, conversation among humans is striking for the variety of time scales involved. A 500 msec pause is long enough to signal to a participant in a conversation that she must indicate that she is following. At the same time, the other participant will continue to deliver his contribution to the conversation, which may go on for as long as several minutes. This *multi-threadedness* means that only some conversational behaviors – the longer ones, such as deciding what to say – are deliberate (or, planned) while others – the shorter ones, such as producing a feedback nod – are simply reactive (carried out unconsciously). Thus, in the dialogue above, only 200 msec into Mike's speech, Rea nods that she is following. Her later verbal response to that same message, however, takes more than 1 sec to plan and deliver.

Entrainment

One of the most striking aspects of human-human conversation is *entrainment*. Through gaze, eyebrow raises and head nods both speakers and listeners collaborate in the construction of synchronized turns, and smooth conversation. In fact, over the course of a conversation, participants increasingly synchronize their behaviors to one another. Entrainment ensures that conversation will proceed efficiently (one of the functions that Brennan & Hulteen ([3]) suggest are needed for more robust speech interfaces). Rea cannot yet entrain her non-verbal behaviors to those of the listener. Human users, however, very quickly entrain to her, and begin to nod and turn their heads in sychrony with her within one or two conversational turns.

REA: An Embodied Conversational Agent

Thus far we have talked about some of the essential properties of embodied human – human conversation, and we have sketched some of the benefits of incorporating these properties into human – computer interfaces. In this section we turn to the details of how that implementation can be accomplished. In order for embodied human – computer conversation to be successful, the insights set out above must be incorporated into every stage of the architecture of the Embodied Conversational Agent. To demonstrate, we turn back to Rea, an embodied conversational agent whose verbal and non-verbal behaviors are designed in terms of the conversational properties described above.

- Rea has a human-like body, and uses her body in human-like ways during the conversation. That is, she uses eye gaze, body posture, hand gestures, and facial displays to contribute to the conversation, and to organize and regulate the conversation. She also understands (some aspects of the use of) these same modalities when employed by her human interlocutor.
- Because of the property of multi-threadedness, the system allows Rea to watch for feedback and turn requests, while the human user can send these at any time through various modalities. The architecture must be flexible enough to track these different threads of communication in ways appropriate to each thread. Because different threads have different response time requirements, the architecture must allow different processes to concentrate on activities at different time scales.
- Dealing with propositional information requires building a model of the user's needs and knowledge. Thus the architecture includes both a static knowledge base that deals with the domain (here, real estate) and a dynamic discourse knowledge base (dealing with what has already been said). To generate propositional information the system must plan how to present multi-sentence output and manage the order of presentation of interdependent facts. To understand interactional information, on the other hand, the system builds a model of the current state of the conversation with respect to conversational process (who is the current speaker and who is the listener, has the listener understood the speaker's contribution, and so on).
- The core modules of the system operate exclusively on functions (rather than sentences, for example), while other modules at the edges of the system translate input into functions, and functions into outputs. This also produces a symmetric architecture because the same functions and modalities are present in both input and output. Such models have been described for other conversational systems: for example, by Brennan and

Hulteen ([3]). Our work extends this work by developing a conversational model that relies on the function of non-verbal behaviors as well as speech, and that makes explicit the interactional and propositional contribution of these conversational behaviors.

Architecture

Figure 2 shows the modules of the Rea architecture. The three key points for Embodied Conversational Agents are:

- 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 semantic representation frame 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 (the understanding, response planner, and generation modules), and moves to the periphery decisions made in terms of behaviors (the input manager and action scheduler).

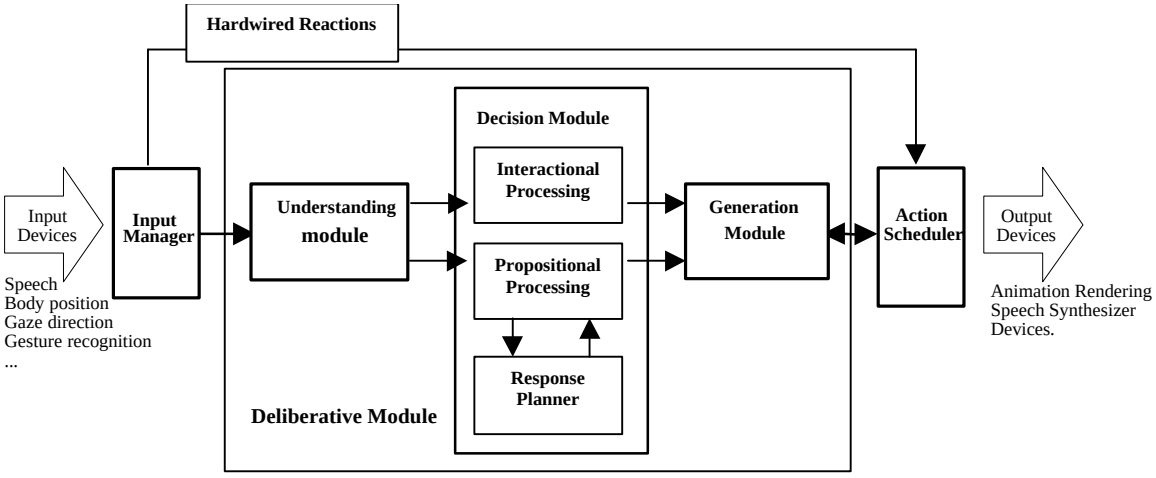

Figure 2: Rea Architecture (co-developed with the Fuji-Xerox Palo Alto Laboratory).

The Input Manager collects input from all modalities and decides whether the data requires instant reaction or deliberate discourse processing. Hardwired Reaction handles spontaneous reaction to stimuli such as the appearance of the user. These stimuli can then directly modify the agent's behavior without much delay. For example, the agent's gaze can seamlessly track the user's movement. The Deliberative Discourse Processing module handles all input that requires a discourse model for proper interpretation. This includes many of the interactional behaviors as well as all propositional behaviors. Lastly the Action Scheduler is responsible for scheduling motor events to be sent to the animated figure representing the agent. A crucial function of the scheduler is to prevent collisions between competing motor requests. The modules communicate with each other using KQML, a speech-act based inter-agent communication protocol, which serves to make the system modular and extensible.

Implementation

The system currently consists of a large projection screen on which Rea is displayed and in front of which the user stands. 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 (written in SGI OpenGL) and conversation engine (written in C++ and CLIPS), while several other computers manage the speech recognition (until recently IBM Via Voice; currently moving to SUMMIT) and generation (previously Microsoft Whisper; currently moving to BT Festival) and image processing (STIVE).

In the implementation of Rea we have attended to both propositional and interactional components of the conversational model. In terms of the propositional component, Rea's speech and gesture output is generated in real-time. The descriptions of the houses that she shows, along with the gestures that she uses to describe those houses are generated using the SPUD natural language generation engine, modified so as to also generate natural gesture ([5]). In this key aspect of Rea's implementation, speech and gesture are treated on a par, so that a gesture may be just as likely to be chosen to convey Rea's meaning as a word. The approach is motivated by literature in Psychology and Linguistics suggesting a similar process in humans (see citations in [4]). For example, in the dialogue above, Rea indicates the extent of the garden with her hands, while conveying the attractiveness of the garden in speech. Rea's other responses (greetings, off-hand comments) are generated from an Eliza-like engine.

In the interactional component, as demonstrated in the dialogue above, the following functions are possible:

- Acknowledgment of user's presence - by posture, turning to face the user;
- Feedback function - Rea gives feedback in several modalities: she may nod her head or emit a paraverbal (e.g. "mmhmm") or a short statement such as "okay" in response to short pauses in the user's speech; she raises her eyebrows to indicate partial understanding of a phrase or sentence.
- Turn-taking function – Rea tracks who has the speaking turn, and only speaks when she holds the turn. Currently Rea always allows verbal interruption, and yields the turn as soon as the user begins to speak. If the user gestures she will interpret this as an expression of a desire to speak, and therefore halt her remarks at the nearest sentence boundary. Finally, at the end of her speaking turn she turns to face the user.

These conversational functions are realized as conversational behaviors. For turn taking, for example, the specifics are as follows: Rea generates speech, gesture and facial expressions based on the current conversational state and the conversational function she is trying to convey. For example, when the user first approaches Rea ("User Present" state), she signals her openness to engage in conversation by looking at the user, smiling, and/or tossing her head. When conversational turn-taking begins, she orients her body to face the user at a 45 degree angle. When the user is speaking and Rea wants the turn she looks at the user. When Rea is finished speaking and ready to give the turn back to the user she looks at the user, drops her hands out of gesture space and raises her eyebrows in expectation. Once again, this approach comes directly from Social Science literature on human – human conversation [4]. Table 2 summarizes Rea's current interactional output behaviors.

State	Output Function	Behaviors
User Present	Open interaction	Look at user. Smile. Toss head.
	Attend	Face user.
	End of interaction	Turn away.
	Greet	Wave. Say "hello" .
Rea Speaking	Give turn	Relax hands. Look at user. Raise eyebrows
	Signoff	Wave. Say "bye"
User Speaking	Give feedback	Nod head, paraverbal ("hmm")
	Want turn.	Look at user. Raise hands.
	Take turn.	Look at user. Raise hands to begin gesturing. Speak.

Table 2. Output Functions

By modeling behavioral categories as discourse functions we have developed a natural and principled way of combining multiple modalities, in both input and output. Thus when REA decides to give feedback, for example, she can choose any of several modalities based on what is appropriate and available at the moment.

Conclusion

In this paper I have argued that embodied conversational agents are a logical and needed extension to the conversational metaphor of human – computer interaction, and to anthropomorphization of the interface. Following Nickerson ([9: 54]) I hasten to point out that "an assumption that is not made, however, is that in order to be maximally effective, systems must permit interactions between people and computers that resemble interperson conversations in all respects." Instead, I argue that, since conversation, anthropomorphization, and social interfaces in general are so popular in the interface community, attention needs to be paid to how they are implemented. That is, embodiment needs to be based on an understanding of conversational function, rather than an additive – and ad hoc -- model of the relationship between non-verbal modalities and verbal conversational behaviors.

The qualitative difference is that the human body enables the use of certain communication protocols in face-to-face conversation. The use of gaze, gesture, intonation, and body posture play an essential role in the proper execution of many conversational behaviors–such as conversation initiation and termination, turn-taking and interruption handling, and 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 message a speaker sends.

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. Thus, an embodied interface agent which exploits them has the potential to provide a higher bandwidth of communication than would otherwise be possible. However, the flip side is that these communications protocols must be executed correctly for the embodiment to bring benefit to the interface.

To date empirical investigations of any kinds of embodied interfaces have been few, and their results have been equivocal. As Shneiderman points out ([12]), there is ample historical evidence, in the form of a veritable junk pile of abandoned anthropomorphic systems, against using anthropomorphized designs in the interface. And Dehn and van Mulken ([7], specifically examining evaluations of recent animated interface agents, conclude that the benefits of these systems are arguable in terms of user performance, engagement with the system, or even attributions of intelligence. They point out, however, that virtually none of the systems evaluated exploited the affordances of the human bodies they inhabited: this design paradigm "can only be expected to improve human – computer interaction if it shows some behavior that is functional with regard to the system's aim." In other words, embodiment for the sake of the pretty graphics will probably not work.

But note that it is only very recently that embodied conversational agents have been implemented with anywhere near the range of conversational properties outlined above. For this reason, it is only now that we can start to carry out rigorous evaluations of the benefits of conversational embodiment. In my own lab we have been encouraged by the results of early comparisons of embodied conversational agents (a) to an embodied interface without conversational behaviors, and (b) to a menu-driven avatar system. Comparing one of Rea's ancestors (see [4] for further details and citations) to an identical body uttering identical words, but without non-verbal interactional behaviors, we found that users judged the version with interactional behaviors to be more collaborative, more cooperative, and to exhibit better natural language (even though both versions had identical natural language abilities). On the other hand, performance on the task was not significantly different between the groups. An evaluation of one of Rea's cousins – a 3D graphical world where anthropomorphic avatars autonomously generate the conversational behaviors described here – did show positive benefits on task performance. And users in this study preferred the autonomous version to a menu-driven version with all of the same behaviors [6].

One of the motivations for embodied conversational agents – as for dialogue systems before them – comes from increasing computational capacity in many objects and environments outside of the desktop computer – smart rooms and intelligent toys, in environments as diverse as a military battlefield or a children's museum – and for users as different from one another as we can imagine. It is in part for this reason that we continue to pursue the dream of computers without keyboards, that can accept natural untrained input. In situations such as these, we will need robustness in the face of noise, universality and intuitiveness, and a higher bandwidth than speech alone.

Such benefits may come from embodied conversational interface agents. We demonstrated our approach to this new paradigm with the Rea system. 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. Rea is an embodied conversational agent who is increasingly able to hold up her end of the conversation.

References

[1] André, E., T. Rist, & J. Mueller, "Employing AI Methods to Control the Behavior of Animated Interface Agents," *Applied Artificial Intelligence*, vol. 13, pp. 415-448, 1999.

[2] Bates, J., "The Role of Emotion in Believable Agents," *Communications of the ACM*, vol. 37, pp. 122-125, 1994.

[3] Brennan, S. E. & E. A. Hulteen, "Interaction and Feedback in a Spoken Language System: A Theoretical Framework," *Knowledge Based Systems*, vol. 8, pp. 143-151, 1995.

[4] Cassell, J., "Elements of Face-to-Face Conversation for Embodied Conversational Agents," in *Embodied Conversational Agents*, Cassell, J., J. Sullivan, S. Prevost, et al., Eds. Cambridge, MA: MIT Press, in press.

[5] Cassell, J. & M. Stone, "Living Hand to Mouth: Theories of Speech and Gesture in Interactive Systems," *Proceedings of AAAI Fall Symposium: Psychological Models of Communication in Collaborative Systems*, Cape Cod, MA, 1999.

[6] Cassell, J. & H. Vilhjalmsson, "Autonomy vs. Direct Control: Communicative Behaviors in Avatars," *Autonomous Agents and Multi-Agent Systems*, vol. 2, pp. 45-64, 1999.

[7] Dehn, D. & S. v. Mulken, "The Impact of Animated Interface Research: A Review of Empirical Research," *Human-Computer Studies*, in press.

[8] Lester, J., S. Towns, C. Calloway, & P. FitzGerald, "Deictic and Emotive Communication in Animated Pedagogical Agents," in *Embodied Conversational Agents*, Cassell, J., J. Sullivan, S. Prevost, et al., Eds. Boston: MIT Press, 2000.

[9] Nickerson, R. S., "On Conversational Interaction with Computers," *Proceedings of User Oriented Design of Interactive Graphics Systems: Proceedings of the ACM SIGGRAPH Workshop*, 1976.

[10] Reeves, B. & C. Nass, *The Media Equation: How People Treat Computers, Television and New Media Like Real People and Places*. Cambridge: Cambridge University Press, 1996.

[11] Rickel, J. & W. L. Johnson, "Task-Oriented Collaboration with Embodied Agents in Virtual Worlds," in *Embodied Conversational Agents*, Cassell, J., J. Sullivan, S. Prevost, et al., Eds. Boston: MIT Press, 2000.

[12] Shneiderman, B., *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, Third ed. Reading, MA: Addison-Wesley, 1998.