Justine Cassell

Embodied Conversational Interface Agents

More than another friendly face, Rea knows how to have a conversation with living, breathing human users with a wink, a nod, and a sidelong glance.

Animals and humans all manifest social qualities and skills. Dogs recognize dominance and submission, stand corrected by their superiors, demonstrate consistent personalities, and so forth. On the other hand, only humans communicate through language and carry on conversations with one another. The skills involved in human conversation have developed in such a way as to exploit all the special characteristics of the human body. We make complex representational gestures with our prehensile hands, gaze away and toward 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 pointed to the application of the features of face-to-face conversation in human-computer interaction, including mixed initiative, nonverbal communication, sense of presence, and the rules involved in transferring control [9]. However, although these features have gained widespread recognition, human-computer conversation has only recently become more than a metaphor. That is, only recently have human-computer interface designers taken the metaphor seriously enough to attempt to design a computer that could hold up its end of the conversation with a human user.

Here, I describe some of the features of human-human conversation being implemented in this new genre of embodied conversational agent, exploring a notable embodied conversational agent—named Rea—based on these features. Because conversation is such a primary skill for humans and learned so early in life (practiced, in fact, between infants and their mothers taking turns cooing and burbling

to one another), and because the human body is so nicely equipped to support conversation, embodied conversational agents may turn out to be a powerful way for humans to interact with computers. However, in order for embodied conversational agents to live up to this promise, their implementation must be based on the study of human-human conversation, and their architectures reflect some of the intrinsic properties found there.

Embodied conversational interfaces are not just computer interfaces represented by way of human or animal bodies. They are not just interfaces in which human or animal bodies appear lifelike or believable in their actions and their reactions to human users. Embodied conversational agent interfaces are specifically conversational in their behaviors and specifically humanlike in the way they use their bodies in conversation. That is, they may be defined as having the same properties as humans in face-to-face conversation, including four very human abilities:

- Recognizing and responding to verbal and nonverbal input;
- Generating verbal and nonverbal output;
- Dealing with conversational functions, such as turn-taking, feedback, and repair mechanisms; and
- Giving signals that indicate the state of the conversation, as well as contributing new propositions to the discourse.

Embodied conversational agents represent a new slant on the argument about whether it is wise to anthropomorphize the interface. Clifford Nass and Byron Reeves of Stanford University and others have 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 and view some computers as more authoritative than others, even that some are experts and others generalists. In many other ways, users react to the computer as if it were another human [10]. Nevertheless, the fact that we react to computers in this way begs the question of whether interface designers should accede to such illogical tendencies by building computers that look like humans. Critics (such as Ben Shneiderman of the University of Maryland [12]) have asked what function building humanlike computers really serves, pointing out that anthropomorphized interfaces have never succeeded in the past and may even lead to slower user response times and confusion.

We might say that only conversational embodiment—giving the interface the appearance and the function of the human body in conversation—allows us to evaluate the function of embodiment in the interface. Simply building anthropomorphized interfaces that talk (but don't use their talk in humanlike ways) sheds no light on the debate about embodiment. However, well-designed embodied conversational interface agents address particular needs not met in conventional keyboard, mouse, and screen interfaces. Such needs include making dialogue systems robust despite imperfect speech recognition, increasing bandwidth at low cost, and supporting efficient collaboration between humans and machines and between humans mediated by machines. These missing ingredients in conventional human-computer interfaces are exactly what bodies bring to conversation.

Embodied conversational agents also bring a new dimension to discussions about the relationship between emulation and simulation, as well as to the role of foundational principles in interface design, that are true to real-world phenomena in the interfaces consumers actually buy. The first wave of interface agents with bodies and autonomous embodied characteristics in the early 1990s—often called autonomous synthetic characters—did not focus on conversation but on more general interactional social skills. Researchers developing these characters discovered, sometimes to their surprise, that the best way to produce believability and lifelikeness may not be through the modeling of life. They found themselves turning to insights from Disney animators and other artists about caricaturization and exaggeration as a way of getting users to suspend disbelief and attribute reality to interactive characters. For example, the OZ project at Carnegie-Mellon University in Pittsburgh enlisted artists and actors in the early process of developing their interactive characters to help convey features of personality in a compelling way [2].

This drama-based design approach has carried less 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 they have to turn to theories of human-human interaction and investigate the nitty-gritty details of conversation to ensure their interfaces share the conversational skills of human users. Thus, for example, James Lester's COSMO character [8] at North Carolina State University refers to the objects in its environment by using pronouns, descriptions, and pointing gestures, according to a complex algorithm based on the linguistic theory of referential ambiguity.

Elisabeth André and Thomas Rist of DFKI in Germany associate particular gestures with aspects of

Figure 1. A user interacting with Rea.

Rea's domain of expertise is real estate; she has access to a database of available condominiums and houses for sale in Boston. She can display pictures of these properties and 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 while Rea is gazing about idly. As Mike moves within range of the two cameras mounted above the screen, Rea turns to face him and says:
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 the 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, so Rea does not continue, 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 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 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 bedrooom?
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.

planning [1]. These researchers generate pointing gestures as a subaction of the rhetorical action of labeling, as in, "This is a hotel," which is in turn a subaction of the action of elaboration. Similarly, Jeff Rickel and W. Lewis Johnson of the University of Southern California have their pedagogical agent move to objects in the virtual world and then generate a pointing gesture at the beginning of an explanation about the object [11].

We shouldn't be debating whether anthropomorphization is good or bad. Instead, we should be emphasizing the implementation of precisely described, motivated characteristics of human conversation into the interface. This perspective would encourage researchers developing embodied conversational agents to address the adequacy of their theories of human behavior when implementing effective interfaces.

What conversational skills can embodied conversational agents display? Let's start with the scenario in Figure 1 between a human user and an embodied conversational real-estate agent—Rea, short for real

estate agent—then turn to the behaviors that characterize it.

Conversational Models

Why is the scenario in Figure 1 so exciting yet so difficult to achieve? Because Rea is engaging in subtle humanlike conversational patterns. And because a set of five properties of human conversation had to be modeled for the system to be able to demonstrate these patterns.

Function rather than behavior. Even though conversation seems orderly, governed by rules, no two conversations are exactly alike, and the set of behaviors exhibited by the people doing the conversing differs from person to person and from conversation to conversation. Therefore, to 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 (see Table 1). These elements are described in terms of their role or function in the exchange. Typical discourse functions include conversation initiation, turn-taking, feedback, contrast and emphasis, and breaking away.

This distinction is especially important, because particular behaviors, such as raising eyebrows, can be employed in a variety of circumstances to produce a variety of 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, the availability of modalities, such as the face and hands, type of conversation, cultural patterns, and personal style. Thus, in the dialogue in Figure 1, Rea nods to indicate that she is listening (as a way of providing feedback). She might have said, "Uh huh" or "I see." Note that in a different context, these behaviors might carry different meanings; for example, a head nod might indicate emphasis or a salutation, rather than feedback.

Synchronization. Behaviors that represent the same function or achieve the same communicative goals occur in temporal synchrony with one another. This property leads humans to assume that synchronized phenomena carry meaning. That is, the meaning of a nod is determined by where it co-occurs in an utterance, with even 200msec making a difference; consider the difference in meaning between: "You did a [great job]" (brackets indicate the temporal extent of the nod) and "You did a [. . .] great job"). In the dialogue in Figure 1, Rea says, "It has a nice garden," at exactly the same time she sketches the outlines of the garden; 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, as defined by social scientists. Propositional information, which corresponds to the content of the conversation, includes meaningful speech, as well as hand gestures and intonation used to complement or elaborate on speech content; gestures indicating size, as in "It was this big," or rising intonation, indicating a question, as in "You went to the store?" Interactional information consists of cues regulating the conversational process and includes a range of nonverbal behaviors: quick head nods to indicate that one is understanding; bringing one's hands to one's lap and turning to the listener to indicate that one is giving up the turn. It also includes regulatory speech, such as "Huh?" and "Go on."

Table 1. Examples of conversational functions and how they are represented in bodily behaviors (taken from [6]).

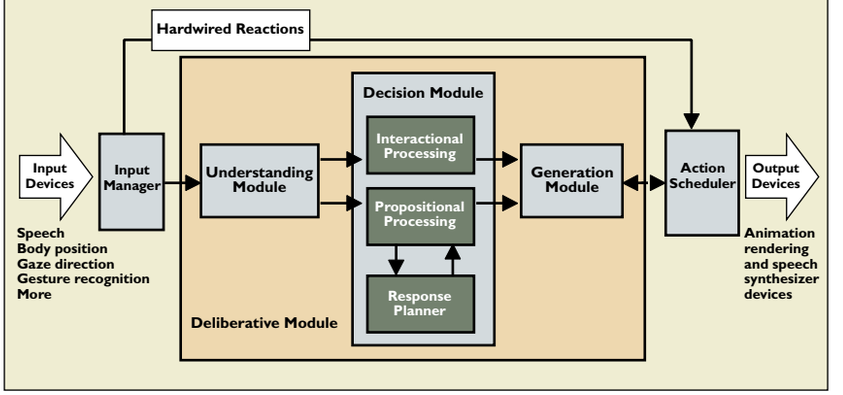
Functions	Behaviors
Initiation and termination	
React to new person	Short glance at other
Break away from conversation	Glance around
Farewell	Look at other, head nod, wave
Turn-taking	
Give turn	Look, raise eyebrows (followed by silence)
Want turn	Raise hands into gesture space
Take turn	Glance away,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

The interactional discourse functions are responsible for creating and maintaining an open channel of communication between participants; propositional functions shape the actual content. Both functions may be fulfilled through either verbal or nonverbal means. In the dialogue in Figure 1, Rea's nonverbal behaviors sometimes contribute propositions to the discourse, such as a gesture indicating that the house in question is five minutes on foot from the T stop, and sometimes regulate the interaction, such as the head-nod indicating Rea understands what Mike just said.

Multithreadedness. Interactional behaviors tend to be shorter in duration than their propositional

counterparts. Conversation among humans is striking for the variety of time scales involved; for example, a 500msec 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 continues to deliver his contribution to the conversation, which may go on for as long as several minutes. This multithreadedness means that only some conversational behaviors—specifically the longer ones, such as deciding what to say—are deliberate (or planned); others—specifically the shorter ones, such as producing a feedback nod—are simply reactive (carried out unconsciously). Thus, in the dialogue in Figure 1, only 200msec into Mike's speech, Rea nods that she is following. However, her later verbal response to the same message from Mike takes more than one second to plan and deliver.

Figure 2. Rea architecture.
(codeveloped with the Fuji-Xerox Palo Alto Laboratory)



Entrainment. "Entrainment" is one of the most striking aspects of human-human conversation. Through gaze, raised eyebrows, and head nods, speakers and listeners collaborate in the construction of synchronized turns and smooth conversation. Over the course of a conversation, participants increasingly synchronize their behaviors with one another. Entrainment ensures that conversation proceeds efficiently; it is also one of the functions Susan Brennan of the State University of New York at Stony Brook and Eric Hulteen of Interval Research suggest are needed for more robust speech interfaces [3]. Rea cannot yet entrain her nonverbal behaviors to those of the listener. But human users quickly entrain to her, beginning to nod and turn their heads in synchrony with her within one or two conversational turns.

Rea's Verbal and Nonverbal Behaviors

Beyond these essential properties of embodied human-human conversation and the benefits of incorporating them into human-computer interfaces, we still need to figure out how to implement them. Successful embodied human-computer conversation depends on our ability to incorporate these insights into every stage of the architecture of an embodied conversational agent. Again consider Rea, whose verbal and nonverbal behaviors are designed in terms of these properties.

Humanlike body. Rea has a humanlike body and uses it in humanlike ways during a conversation. That is, she uses eye gaze, body posture, hand gestures, and facial displays to contribute to, organize, and regulate the conversation. She also understands some aspects of how her human interlocutor uses these modalities.

Feedback and turn requests. Thanks to multithreadedness, Rea can watch for feedback and turn requests, while the human user can send these signals at any time through various modalities. The Rea architecture has to be flexible enough to track these threads of communication in ways appropriate to each thread. Because different threads have different-response-time requirements, the architecture has to allow different processes to concentrate on activities at different time scales.

User discourse model. Dealing with propositional information requires building a model of the user's needs and knowledge. The architecture includes both a static knowledge base that deals with the agent's domain (in Rea's case, real estate) and a dynamic discourse knowledge base that deals with what has already been said. To generate propositional information, the system plans how to present multisentence output and manages the order of the presentation of interdependent facts. To understand interactional information, the system builds a model of the current state of the conversation with respect to the conversational process, including knowing who the current speaker is and who the listener is and determining whether the listener has understood the speaker's contribution.

Conversational functions. The core modules of the Rea system operate exclusively on functions (rather than, say, on sentences), while other modules at the edges of the system translate input into functions and functions into outputs. This arrangement also produces a symmetric architecture, because the same functions and modalities are present in both input and output. Similar models have been proposed for other conversational systems [3]. Rea extends this

work by developing a conversational model that relies on the function of nonverbal behaviors, as well as on speech, making explicit the interactional and propositional contribution of these conversational behaviors.

Figure 2 shows Rea's architectural modules, demonstrating the following key points:

Input. Input is accepted from as many modalities as there are input devices. However, the various modalities are integrated into a single semantic representation passed from module to module.

Semantic representation. This semantic representation frame has slots for interactional and propositional information, so the regulatory and content-oriented contribution of every conversational act is maintained throughout the system.

Categorization of behaviors. Categorizing behaviors in terms of their conversational functions is mirrored by the architecture, which centralizes decisions in terms of functions—the modules for understanding, response planning, and generation—and moves to the periphery decisions in terms of behaviors—the input manager and action scheduler.

The input manager collects input from all modalities and decides whether the data requires instant reaction or deliberate discourse processing. A hardwired reaction component handles spontaneous reaction to such stimuli as the user's appearance. These stimuli can then be used to 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 requiring a discourse model for proper interpretation. This input includes many of the user's interactional behaviors, as well as all his/her propositional behaviors. The action scheduler is responsible for scheduling motor events to be sent to the animated figure representing the agent Rea. One crucial scheduler function is the prevention of collisions among competing motor requests. The modules communicate with each other using the Knowledge Query and Manipulation Language, a speech-action-based interagent communication protocol that makes the system modular and extensible.

Rea Hardware and Software

The system includes a large projection screen on which Rea is displayed; the user stands in front of it. Two cameras mounted on top of the screen track the user's head and hand positions in space. A user wears a microphone for capturing speech input. A single SGI Octane computer runs the graphics software (written in SGI OpenGL) and the conversation engine (written in C++ and C Language Integrated Productions System, CLIPS), a rule-based expert system programming language. Several other computers manage speech recognition (until recently IBM ViaVoice, now moving to SUMMIT, a probabilistic framework for feature-based speech recognition from the MIT Spoken Language Systems Group) and generation (until recently Microsoft Whisper, now moving to British Telecom's Festival text-to-speech system) and image processing (the Stereo Interactive Vision Environment, which uses stereo cameras to perceive human gestures).

The Rea implementation attends to the conversational model's propositional and interactional components. In the propositional component, Rea's speech and gesture output is generated in real time. The descriptions of the houses she shows, along with the gestures she uses to describe them, are generated using the Sentence Planning Using Description natural-language-generation engine, modified to be able to generate natural gesture [5]. In this key aspect of Rea's implementation, speech and gesture are treated on a par, so a gesture is as likely to be chosen to convey Rea's meaning as a word is. This approach follows psychology and linguistics research suggesting a similar process in humans [4]. For example, in the dialogue in Figure 1, Rea indicates the extent of the garden with her hands, while conveying its attractiveness in speech. Rea's other responses, such as greetings and off-hand comments, are generated from an Eliza-like engine, which has the ability to understand particular key-

Table 2. Rea's output functions.

State	Output Functions	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
	Sign off	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

words and generate responses based on them.

Rea's interactional processing component, as in Figure 2, provides at least three functions:

Acknowledging the user's presence. The user's presence is acknowledged through posture and turning to face the user;

Feedback. Rea gives feedback in several modalities; she may nod her head or emit a paraverbal (such as "Mmhmm") or a short statement (such as "Okay") in response to short pauses in the user's speech; she raises her eyebrows to indicate her partial understanding of a phrase or sentence.

Turn-taking. Rea tracks who has the speaking turn, speaking only when she holds the turn. Rea always allows verbal interruption, yielding the turn as soon as the user begins to speak. She interprets user gestures as an expression of a desire to speak, halting her remarks at the nearest sentence boundary. At the end of her speaking turn, she turns to face the user.

These conversational functions are realized physically as conversational behaviors. For example, in turn-taking, the specifics are: Rea generates speech, gesture, and facial expressions based on the current conversational state and the conversational function she is trying to convey. When the user first approaches Rea (the "user present" state), Rea signals her openness to engage in conversation by looking at the user, smiling, 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 is ready to give the turn back to the user, she looks at the user, drops her hands out of the gesture space, and raises her eyebrows in expectation. This formalization of conversational turn-taking comes directly from the social science literature on human-human conversation [4]. Table 2 summarizes Rea's interactional output behaviors.

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

A Deep Understanding of Conversational Function

Embodied conversational agents are a logical and needed extension to the conversational metaphor of human-computer interaction, as well as to the anthropomorphization of the interface. Following Raymond S. Nickerson, a researcher at Bolt, Beranek and Newman [9], 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." I argue instead that, since conversation, anthropomorphization, and social interfaces are so popular in the interface community, attention should be paid to their implementation. That is, embodiment needs to be based on a deep understanding of conversational function, rather than an additive—and ad hoc—model of the relationship between nonverbal modalities and verbal conversational behaviors.

The qualitative difference between these two views is that the human body enables certain communication protocols in face-to-face conversation. Gaze, gesture, intonation, and body posture all play essential roles in the execution of many conversational behaviors, including initiation and termination, turn-taking and interruption handling, and feedback and error correction; such behaviors enable the exchange of multiple levels of information in real time. Humans 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, and gaze behavior can significantly alter the message a speaker is sending.

Of particular interest to interface designers is that these communication protocols are available for free. Users need no training; all native speakers of a given language have these skills, using them daily. Thus, an embodied interface agent that is able to exploit them has the potential to provide communication of greater bandwidth than would be possible otherwise. The flip side is that these protocols have to be executed correctly for the embodiment to bring benefit to the interface.

To date, few researchers have empirically investigated embodied interfaces, and their results have been equivocal. As Shneiderman points out, there is ample historical evidence—in the form of a junk pile of abandoned anthropomorphic systems—against using anthropomorphized designs in interface design [12].

The recent evaluations of animated interface agents by Doris M. Dehn, a psychology professor at the University of the Saarland in Germany, and Susanne van Mulken, a researcher at DFKI in Germany, concluded that the benefits of these systems are arguable in terms of user performance, engagement with the system, and even attributions of intelligence [7]. However, they also pointed out that virtually none of the evaluated systems managed to

exploit the affordances, or special characteristics, of the virtual human bodies they inhabit. This design paradigm of embodying the interface, according to Dehn and van Mulken, "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 pretty graphics alone is unlikely to work.

Note too that only recently have embodied conversational agents been implemented with anywhere near the range of conversational properties I've outlined here. For this reason, we are only beginning to be able to carry out rigorous evaluations of the benefits of conversational embodiment. In my research lab, we have been encouraged by the results of early comparisons of embodied conversational agents to an embodied interface without conversational behaviors and to a menu-driven avatar system.

Comparing one of Rea's ancestors [4] to an identical body uttering identical words, but without nonverbal interactional behaviors, we found that users judged the version with interactional behaviors more collaborative and cooperative and better at exhibiting natural language (though both versions had identical natural language abilities). On the other hand, the user's performance on the task they were asked to complete with the computer was not significantly different between user groups. An evaluation of one of Rea's cousins—a 3D graphical world where anthropomorphic avatars autonomously generate conversational behaviors—did show positive benefits on task performance. Users in this study preferred the autonomous version to a menu-driven version with all the same behaviors [6].

Computers Without Keyboards

Another factor motivating interface designers to develop embodied conversational agents is the increasing computational capacity in many objects and environments beyond the desktop computer, such as smart rooms and intelligent toys, in environments as diverse as military battlefields and children's museums, and for users as diverse as we can imagine. It is in part our desire to keep up with these new computing applications that we pursue the vision of computers without keyboards accepting natural untrained input. In such face-to-face situations between humans and computers, we'll need increasing robustness in the face of noise, universality and intuitiveness, and bandwidth greater than speech alone.

These benefits may result from embodied conversational interface agents, as the Rea system demonstrates. Capable of making content-oriented, or propositional, contributions to a conversation with human users, Rea is also sensitive to the regulatory, or interactional, function of verbal and nonverbal human-human conversational behaviors. Rea is also able to produce regulatory behaviors that improve interaction by helping the user be aware of the state of the conversation. Rea is an embodied conversational agent that is increasingly able to hold up her end of the conversation. C

References

1. André, E., Rist, T., and Mueller, J. Employing AI methods to control the behavior of animated interface agents. *Appl. Artif. Intel. 13,* 4–5 (June–Aug. 1999), 415–448.
2. Bates, J. The role of emotion in believable agents. *Commun. ACM 37,* 7 (Jul. 1994), 122–125.
3. Brennan, S. and Hulteen, E. Interaction and feedback in a spoken language system: A theoretical framework. *Knowl. Based Syst. 8,* 2–3 (Apr.–June 1995), 143–151.
4. Cassell, J. Nudge, nudge, wink, wink: Elements of face-to-face conversation for embodied conversational agents. In *Embodied Conversational Agents,* J. Cassell, J. Sullivan, S. Prevost, et al., Eds. MIT Press, Cambridge, Mass., 2000, 1–28.
5. Cassell, J. and Stone, M. Living hand to mouth: Theories of speech and gesture in interactive systems. In *Proceedings of the AAAI Fall Symposium: Psychological Models of Communication in Collaborative Systems* (Cape Cod, Mass., Nov. 5–7). AAAI Press, Menlo Park, Calif., 1999, 34–43.
6. Cassell, J. and Vilhjálmsson, H. Fully emobodied conversational avatars: Making communicative behaviors autonomous. *Auton. Agents Multiagent Syst. 2,* 1 (Mar. 1999), 45–64.
7. Dehn, D. and van Mulken, S. The impact of animated interface research: A review of empirical research. J. *Hum.-Comput. Stud. 51* (2000).
8. Lester, J., Towns, S, Calloway, C., and FitzGerald, P. Deictic and emotive communication in animated pedagogical agents. In *Embodied Conversational Agents,* J. Cassell, J. Sullivan, S. Prevost, et al., Eds. MIT Press, Cambridge, Mass., 2000, 123–154.
9. Nickerson, R. Some characteristics of conversations. In *Man-Computer Interaction: Human Factors Aspects of Computers & People,* B. Shackel, Ed. Sijthoff & Noordhoff, The Netherlands, 1981, 53–64.
10. Reeves, B. and Nass, C. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places.* Cambridge University Press, Cambridge, U.K., 1996.
11. Rickel, J. and Johnson, W. Task-oriented collaboration with embodied agents in virtual worlds. In *Embodied Conversational Agents,* J. Cassell, J. Sullivan, S. Prevost, et al., Eds. MIT Press, Cambridge, Mass., 2000, 95–122.
12. Shneiderman, B. *Designing the User Interface: Strategies for Effective Human-Computer Interaction,* Third Ed. Addison-Wesley, Reading, Mass., 1998.

Justine Cassell (justine@media.mit.edu) is an associate professor in the Media Laboratory at the Massachusetts Institute of Technology in Cambridge, Mass.

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.