

The Future of Communication between Real and Virtual Bodies
Justine Cassell
Professor & Director, Human Computer Interaction Institute

There was a time when communication was about talking face to face, person to person, in close proximity, in real time. There was a time when collaboration was about working together in the same room, assisting each other in concrete tasks. Today, communication is about iChat or Skype or leaving each other messages on the Facebook wall. And collaboration is about using a robot to do surgery in South America while I move a robot arm in Sydney Australia; it is about each contributing elements of an article on Wikipedia. While the communication and collaboration technologies of today are exciting and raise our hopes about what will be possible one day, they also raise fears about the future. They raise fears about a day in which we might lose the skills of personal relationships and the ties to community that make us human. In my work I ask how we can preserve and develop those skills that are most representative of our human existence, those values we take to be most important, most dear to our hearts, most essential to our sense of what it means to be human. I pursue these questions by asking how we can use what we know about the human body, about human social interaction, and about the human mind to develop new kinds of communicating bodies – new technologies -- that don't sacrifice what we hold most dear about human existence. And I rely on the fields of anthropology, developmental psychology, literary theory, and linguistics to inform the development of those newest technologies so that we maintain a link with the past, a link with those human qualities and values that we hold dear.

My methodology in pursuing these questions through the multidisciplinary lens outlined above is to begin with the study of real humans and then to model virtual humans on those real humans. I begin by trying to understand those things that we take for granted about ourselves and our interactions with others. Simple things, like what we do with our eyebrows when we talk, and more complex things like how our eyes smile when we are truly happy, whereas only the corners of our mouths turn up when we want to look happy. Why do I sometimes look you in the eye and sometimes look away when we are talking? It turns out that eyebrow raises accompany important points in conversation, and that eye gaze is a way of managing turn taking. Studying these minute behaviors of human communication microscopically serves two purposes: first of all, it allows us to better understand humans themselves – how we function, how we differ from other animals, how we interact with one another. And, secondly, only by studying these minute but utterly human details of communication can we maintain them in our virtual human interlocutors of the future. In turn, only by maintaining these utterly human details of communication in virtual humans, can we ensure that virtual humans will draw out the very human patterns of communication in us, their human interlocutors.. So I use such studies of humans to build formal models of human conversation, of the details of communication and their function. Those formal models then serve as the basis for algorithms that are implemented in virtual humans, and by watching real humans interact with the virtual humans, by watching what works and – more interestingly – what fails, I get a sense of where the gaps lie in our knowledge of real humans. Then I can return to studying real humans, and the iterative process of study begins again.

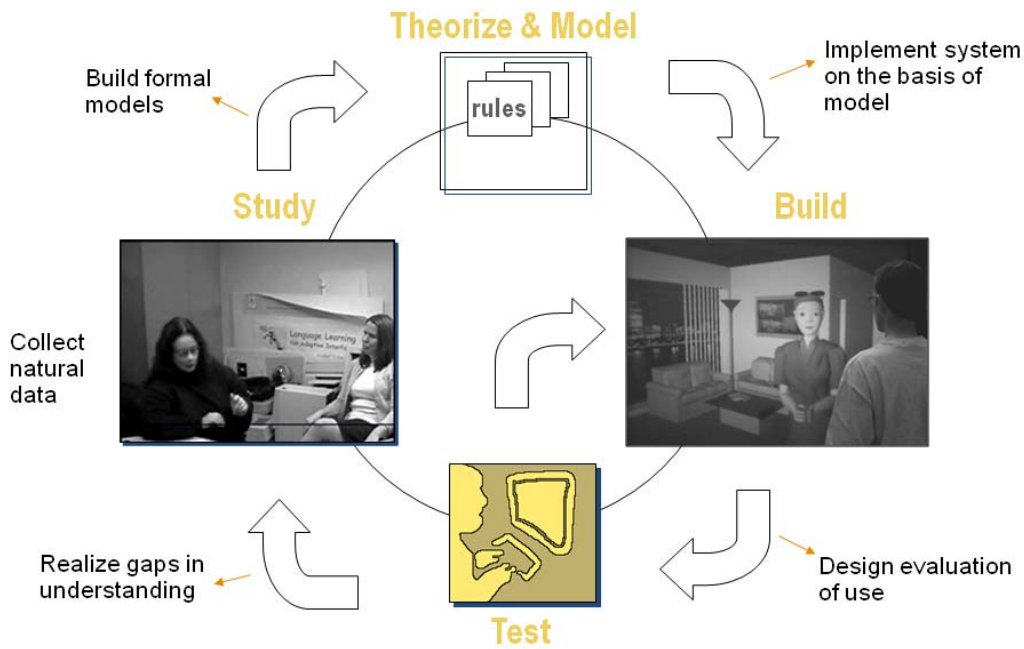


Figure 1: Methodology

In looking at conversational behavior over the years it has become clear that meaning is made up out of the sum of all the parts, where the parts are language and intonation and hand gestures and posture and facial movements (among other sources). And, perhaps surprisingly, it is the case that virtually all people use their faces and their hands and their bodies and their eyes (and so forth) to make meaning in conversation. And in all people in all countries, these embodied resources join in tight configurations to convey particular meanings and particular stances. The meanings that we make may differ from person to person and from culture to culture, but the ways that we make meanings are the same. So, for example, it appears to be the case that virtually all people in all cultures use gesture in order to clarify what they are saying in words. And virtually all people use eye gaze to manage the conversation. And virtually all people modulate the tone of their voices to add meaning to the content of the words they are saying. Thus last night in talking to Wolfgang Wahlster I said “Wolfgang, it took me only 20 minutes to get here tonight” and this morning I also said “Wolfgang, it took me only 20 minutes to get here this morning.” The words were identical, but yet Wolfgang knew that I got here first by subway and the next time by taxi. How did he know this? By virtue of my gestures, which for *last night* showed two fingers wiggling as if they were legs walking, and for *this morning* showed two hands on a wheel as if I was driving. Research shows that while people mostly do not remember the form of the gestures that they see, these gestures serve to add meaning that is complementary to speech.

These kinds of studies of human communicative behavior have been carried out for decades, but only 17 years ago did we first use such studies of human communicative behavior to build a virtual human that behaved in the same way as we do. And only 15 years ago did we first use these studies of human behavior to build a virtual human that could communicate with real humans in some of the same ways as humans communicate with one another.

Let me give a concrete example. In the figure below, a young woman is giving a description of a home.



Figure 2: Experimental Participant Describing a Home

After watching hundreds of such descriptions, we were able to draw generalizations about how people use speech and gesture together to contribute to their communicative goals – communicative goals which have to do with describing a house to the person sitting across from them. The chart below shows some of those generalizations and it shows how accurate the generalizations are, given what we have viewed in the data.

Table 1: Model of Contributions of Speech and Gesture to Communicative Goals

Major Communicative Goal	Gesture Type	Semantic Features in Speech	Semantic Features in Gestures	Accuracy in Data
Introduce a single Object	Complementary	Existence	Shape, Location	78.8% (26/33)
Introduce multiple Objects	Complementary	Existence, Number	Relative Position	100.0% (9/9)
Describe the configuration of multiple objects	Complementary	Existence	Relative Position, Shape	93.8% (15/16)
Describe location of an object	Redundant	Location	Location	80.0% (20/25)
Describe a general impression	Redundant (Metaphoric)	Impression	Impression	100.0% (6/6)
Describe the shape of an object	Redundant	Shape	Shape	100.0% (5/5)

Based on this model of communication, we were able to implement a virtual human – a realtor, in fact – who understood the questions of the people talking to her, and who responded to them, based on the model of human-human speech and gesture.



Figure 3: Virtual Human Describing House

It is important to note that the virtual human's speech and gesture were not scripted, they actually came from her *understanding* of what the human asked her, her *reasoning* about the best answer to make in response to him, her *knowledge* of human language and gesture and facial movements, and her *ability* to synthesize all of those into conversation.

Now, you will notice that REA (the Real Estate Agent) is not as beautiful as the actresses from the movie Final Fantasy – in fact, she does not even look as realistic. However, while Final Fantasy was filmed by animating the performance of real actors, REA is her own woman (one might say!). REA is based on an artificial intelligence engine and a natural language generation engine that understands the notion of thinking about space and communicating one's own representation of the world, as you can see in the figure.

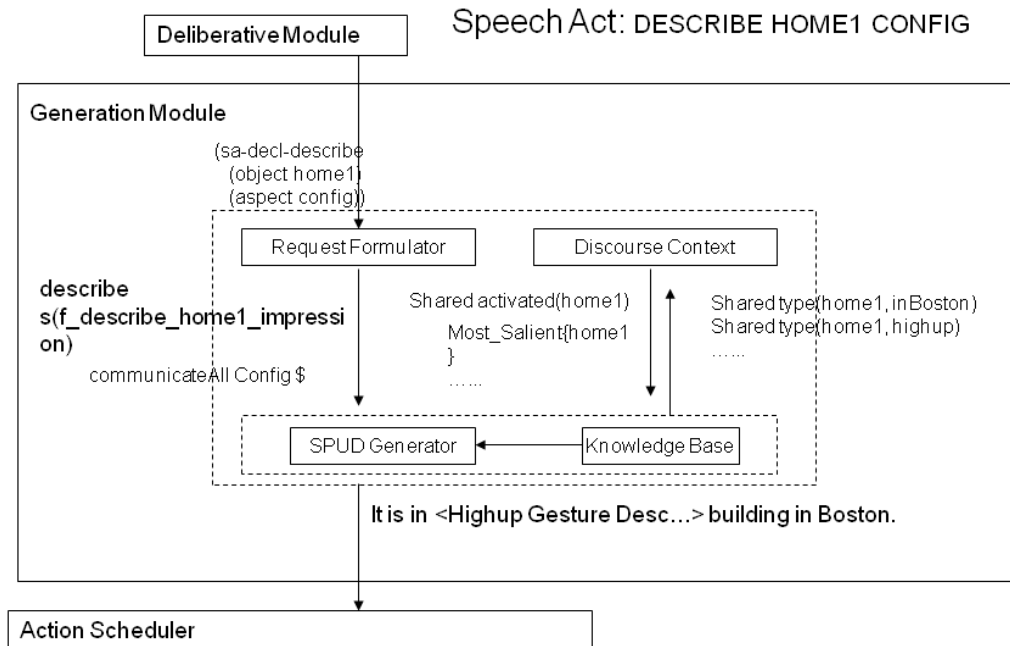


Figure 4: Generating Speech and Gesture in REA

REA constituted a break-through in how humans and computers communicate with one another. But the REA system was just a first step in understanding what makes us human through the use and study of virtual humans, and in maintaining those abilities and values in the digital world. In the years since REA was introduced in 1996, we have pursued the methodology described above to further analyze human behavior, and to implement it into virtual humans. We have better understood that no human conversation only concerns facts and goals; all of our conversations have a social component, so a next version of REA gave her the ability to engage in small talk, or social chit-chat. In order to understand those abilities we followed a genuine real estate agent around her world and analyzed the ways in which she used social chit-chat on the job. We found that the realtor – and other people whom we observed – used small talk in a very subtle way to establish a relationship, to avoid embarrassment, and to get past conversational impasses. On the basis of those data, we were able to build a model of the places in the conversation in which she was likely to use small talk and the places in which she was likely to stick to the task. And that model of small talk that we built from hundreds and hundreds of hours, not just looking at this real estate agent but also looking at sales people, allowed us to build a virtual human who used small talk in her interactions with real humans.

By this stage we were fairly confident in the accuracy of our models of human communication, based on the way in which we collected the data, the hundreds of hours of video we analyzed, and the accuracy of the models we implemented. But at this point we had to ask ourselves whether the virtual humans we were implementing were *effective*. That is, are virtual humans who act like real humans attractive conversational partners? Do they get the task done? Are their skills judged in the same way as real human skills judged? In order to evaluate our real estate agent we therefore asked people to work with REA to look at apartments for rent. Some of the people worked with a version of REA with social skills and some worked with a version of REA who got right down to business. In fact, even though none of the experimental participants knew about the differences between the two versions, and none of them knew the purpose of our study, we discovered that there was quite a difference between the two

REAs: people preferred the version of REA who was able to use small talk, they thought she was smarter, and that she understood their needs better – just like in the real world! Even more strikingly, however, when we looked at the personality of the people in our study, it turned out that it was extroverts who most preferred the version of REA who used small talk, while introverts didn't care which version they used. And this too is very similar to what happens in the real world, where extroverts engage in small talk and appreciate small talk, while introverts might prefer to be left alone.

For example, an extrovert who used the version of REA that engages in small talk, and who did not know about the purpose of the study, said to us

“I thought she was pretty good. You know, I can small talk with somebody for a long time. **It's how I get comfortable with someone, and how I get to trust them, and understand how trustworthy they are,** so I use that as a tool for myself.”

And an introvert said:

“REA exemplifies some things that some people, for example my wife, would have sat down and chatted with her a lot more than I would have. **Her conversational style seemed to me to be more applicable to women, frankly, than to me. I come in and I shop and I get the hell out.** She seemed to want to start a basis for understanding each other, and I would glean that in terms of our business interaction as compared to chit chat. I will form a sense of her character as we go over our business as compared to our personal life. Whereas **my wife would want to know about her life and her dog, whereas I really couldn't give a damn.**”

We were pleased to have been able to evoke such strong feelings in the people who interacted with our virtual humans, and pleased that our study of human-human behavior resulted in virtual humans who were so realistic in their behaviors (even if not in their looks!).

As time has gone on, our study of human behavior, and our implementation of models of human behavior into virtual humans, has advanced beyond the building blocks of turn-taking and acknowledgement and introducing new topics, past gesture and eye gaze and posture, to more social phenomena such as social chit-chat and, most recently, culture and identity. As time has gone on, we have been able to understand more and more about those phenomena that make us most human, and that we most value, and the behaviors that signal those phenomena. And as we understand more about these phenomena in humans, and as we collect more data about the behaviors that make them up, we are better able to build those behaviors into virtual humans.

Most recently, then, we have begun to think about the thorny but preeminently important questions of identity and culture. How do we show others who we are? How do we demonstrate our alliance and affiliation to particular groups? As we have begun to study this, and to read the literature from social psychology and anthropology, it has become clear that “national origin” is only one aspect of who we are – in fact, each of us belongs to a number of different “cultures.” I am American, and I am also female, a professor, from New York City, with years of living in France. Each of these aspects of my identity is more important to me, and to the people around me, at some moments than others. And I highlight the importance of each aspect of identity as I move from context to context throughout the week. Not only do I dress differently, but my accent changes, my gestures adapt, I move with more excitement or with more reserve. I might speak differently with my elderly parents than with the college students I teach. From this perspective, cultural identity can be seen as the demonstration in a

particular context of a set of behaviors and practices that index to other members of the group, and to members of other groups, one's cultural community membership.

With this in mind, we set about exploring issues of identity and culture that were more subtle than national origin, and that come up frequently in people's perceptions of who they are. Most languages are spoken in different ways in different parts of the country. In Germany, for example, German is spoken quite differently in the North than in the South, and the way one speaks German plays quite an important role in how others see you, and how you identify yourself. The US is no exception, and so we studied some American sub-cultures and dialect use. Of course, while language is spoken differently in different parts of the country, it is often the case that one dialect is felt to be the most appropriate for use in school, and in the US this dialect is called "mainstream American English" or MAE. The question we asked is how children use the different dialects they hear around them? Do they learn to switch between the dialect spoken at home and the dialect spoken at school? We read in the educational literature that children who do learn to switch into MAE at school are more likely to do well on their schoolwork – not because MAE is a better version of English, but because it is more accepted. Following our usual methodology, then, we asked if we could observe how dialect is used in culture, if we could build a model of its use, and of switching between dialects, and then if we could build that model of dialect and cultural identity into a virtual human.

That virtual human – what we call a virtual peer – is shown in the figure below.



Figure 5: Virtual Peer using Multiple Dialects

Our research with this virtual peer showed that children recognize the cultural identity of the virtual peer as being the same as their own cultural identity. And when we built the ability to switch dialects into the virtual peer, we found that children are willing to switch dialects to match the dialect of the virtual peer. We believe that by maintaining the important aspects of cultural identity that we all prize, we may have built an educational tool that could help children learn the mainstream dialect that they need for school.

The implications of what I have shown so far are that in order to maintain those aspects of human identity that we prize, virtual humans must have minds, they must have social skills, and they must fit into communities of identity. In the long run, I hope I am demonstrating that we needn't fear that we will lose the face-to-face nature of interaction and be consigned to only using text to communicate. Virtual humans are an important aspect of human-computer interaction today, and their importance continues to grow. And we needn't fear that we must give up our humanity to live in the future. Studies of real humans can and do play an essential role in the development of the technologies for the future. And, studies of those technologies can shed light on our own human behavior, as well as helping us develop technologies that maintain those behaviors and values that we prize. In fact, communication between real and virtual beings can be important for teaching and learning.

I started with history and I would like to end with history. In 1772 the Droz brothers, Swiss clockmakers, built a series of automatons that were able to carry out real human tasks, such as writing and drawing, in the way that humans do. Others followed the same path, and automatons became quite popular. Although these automatons were based on gears rather than software, they were not so different from the virtual humans of today - and their presence as entertainment in drawing rooms of the time began to worry people. As the German novelist E.T.A. Hoffmann wrote, "The story of the **automaton** had struck deep root into their souls and, in fact, a pernicious **mistrust of human figures** in general had begun to creep in."

I would claim that every time a new technology comes on the scene, whether it's an automaton or a virtual human, our tendency is to fear it. But if we make sure that those automatons or virtual humans are based on us and not solely on the capacity of gears or computers, without reference to our minds, hearts and communities, then we will carry our humanness far into the next century and beyond.