Virtual Humans for Learning

William Swartout, Ron Artstein, Eric Forbell,
Susan Foutz, H. Chad Lane, Belinda Lange,
Jacquelyn Morie, Dan Noren, Skip Rizzo, David Traum

■ *Virtual humans are computer-generated characters designed to look and behave like real people. Studies have shown that virtual humans can mimic many of the social effects that one finds in human-human interactions such as creating rapport, and people respond to virtual humans in ways that are similar to how they respond to real people. We believe that virtual humans represent a new metaphor for interacting with computers, one in which working with a computer becomes much like interacting with a person and this can bring social elements to the interaction that are not easily supported with conventional interfaces. We present two systems that embody these ideas. The first, the twins are virtual docents in the Museum of Science, Boston, designed to engage visitors and raise their awareness and knowledge of science. The second, SimCoach, uses an empathetic virtual human to provide veterans and their families with information about PTSD and depression.*

For more than a decade, we have been building virtual humans (VHs) at the University of Southern California (USC) Institute for Creative Technologies. Ultimately we want to be able to create virtual humans that look, communicate, and behave like real people as much as possible. Specifically, these characters would be autonomous, thinking on their own, modeling and displaying emotions, and interacting in a fluid, natural way using verbal and nonverbal communication. While that ultimate vision is still in the future, it is already possible to build characters that realize parts of this vision — characters that can be practically incorporated into a variety of useful systems.

Our initial goal in creating virtual humans was to construct characters that could act as replacements for human role players in training and learning exercises, but their potential is far more profound: virtual humans are able to connect with real people in powerful, meaningful, and complex ways. Because they mimic the behavior of real people, virtual humans can add a rich social dimension to computer interactions, providing not only a wealth of information, which computers already do well, but a means to present that information in more personal ways. Virtual humans can, for example, add social elements that great teachers often employ, such as establishing rapport, building relationships, expressing enthusiasm about a subject, or providing encouragement to a struggling learner. Such inherently human interactions can serve to increase user engagement and one's sense of connection to the virtual character. In fact, studies have shown repeatedly that people respond to virtual humans in much the same way as they do to real people (Reeves and Nass 1996; Krämer, Tietz, and Bente 2003; Gratch 2007b). In sum, we believe the emergence of practical virtual humans opens up an entirely new metaphor for how we interact with computers, one that has important consequences for a wide range of domains, but perhaps most importantly, for how we learn.

 ISSN 0738-4602

In this article we will describe two virtual human systems we have constructed that embody these ideas. The first, the Twins, named Ada and Grace, are virtual characters that act as digital docents and STEM (science, technology, engineering, and mathematics) role models in the Cahners Computer Place at the Museum of Science, Boston. They tell visitors about things they can see and do there, can answer general questions about information technology, and can even explain how they work. However, more than just providing information, the twins are embodied social beings. They banter back and forth showing signs of sibling rivalry. They answer questions about their backgrounds, likes and dislikes, and even whether they have boyfriends. They were expressly designed in this way to appeal to children and young teens.

Our second system, SimCoach, uses a virtual character to help returning veterans confront problems such as depression and PTSD by engaging them in a dialogue to assist them in finding appropriate mental health resources. While the information provided by SimCoach is similar to what could be obtained from a website such as WebMD, the use of conversational interaction with a highly approachable virtual character allows us to create rapport, establish trust, and encourage people to find the help they need. The use of virtual humans, we believe, lowers the barriers to care by providing an engaging guide to potentially daunting mountains of medical information and by removing the stigma that is sometimes associated with seeking help.

Thorough evaluations of these systems have shown that people who interact with them (1) respond to virtual humans in ways that are similar to how they respond to real people, (2) are enthusiastic about interacting with the characters, and (3) gain knowledge and information based on what the characters seek to convey.

Although the two systems described cover very different learning domains, they are united by their use of virtual humans. Because they bring social elements to computer information systems, virtual humans tap into a way we have learned well for millennia — through talking to people — and thus open up new possibilities for how we can use computers as part of learning and education.

In the remainder of this article we first describe how the twins were designed to engage visitors, explain their technology, and report on the results of an extensive evaluations of the accuracy and performance of their natural language-processing technology as well as their impact on museum visitors. We then describe how SimCoach was designed, its underlying technology, and summarize its evaluation. We close by discussing the implications of the virtual human as a new metaphor for learning and interacting with computers.

Virtual Human Museum Guides (The Twins)

The virtual human museum guides were part of an NSF-funded effort in collaboration with the Boston Museum of Science that had two purposes: first, to see if virtual humans could serve as a form of intelligent interface between museum visitors and the exhibits in the museum's Cahner's Computer Place, and second, to create a living exhibit that educated the museum public about how virtual human characters were built and how they actually functioned.

Related Work

There have been some previous installations of virtual humans in museums. In January 2004, the Max agent was installed in the Heinz Nixdorf Museums Forum (HNF), a public computer museum in Paderborn (Germany) (Kopp et al. 2005). Max is humanlike in size on a static screen, standing face-to-face to visitors of the museum. Acting as a museum guide, Max's primary task is to engage visitors in conversations to provide information about the museum, the exhibition, or other topics of interest. However, Max only allows keyboard input. In Cahners Computer Place, the Tinker exhibit (Bickmore et al. 2008; Bickmore, Pfeifer, and Schulman 2011), an embodied conversational agent (ECA) in the guise of a robot, builds relationships and tracks engagement with the user, and includes novel user ID through analysis of hand images. Visitors must select options from a menu to interact with Tinker.

The main difference between the above two systems and the twins is the input modality — the twins understand human speech, allowing unmediated, naturalistic interaction with visitors. There are several systems similar to the twins in this regard. The "pixie" system (Bell and Gustafson 2003) that was part of a 2003 exhibit in the Swedish Telecom museum called "Tänk Om" ("What If"), where visitors experienced a full-size apartment of the year 2010. The visitors could help Pixie perform certain tasks in the apartment or ask the agent general questions about herself or the exhibition. Sergeant Blackwell (Robinson 2008), exhibited at the Cooper-Hewitt Museum in New York in 2006–2007 as part of the National Design Triennial exhibition, and Furhat (Al Moubayed et al. 2012), who was shown for four days in 2011 at the Robotville exhibit in the Science Museum in London, also supported spoken interaction with visitors. These systems do not share the twins' educational goals, but they do understand speech input and employ a variety of techniques to overcome noisy and difficult-to-recognize speech.

Designing for Engagement

We wanted to make a character that visitors to Cahner's Computer Place would find engaging. But engagement is not determined by a single factor. Instead, it is a function of many aspects, such as appearance, behavior, and content all working together. Any single weak aspect may lead to disengagement. Thus we had to consider carefully how our characters would look, how they would interact, and what content they would deliver.

For the character's appearance, we wanted to design a look that would appeal to the broad demographics of the Museum of Science's visitors. Virtual humans in such settings are far from commonplace, so it was difficult to find previous research that would support the choice of any particular character type over another. Would our visitors respond better to a young person their own age, an elder wise-looking expert, someone slightly older and "hip," or someone funny? Would they prefer to interact with a digital character that was very realistic in appearance, or one that was more graphic and "cartoony"? Should the virtual human look like a superhero or a normal person? Because there was already one virtual character in residence at Cahner's, a personable digital robot called Tinker (Bickmore et al. 2008; Bickmore, Pfeifer, and Schulman 2011), we felt that the guide character should look more realistic to provide variety. We also felt that it should resemble the (human) museum interpreters, as it would be performing a function similar to theirs. Initial discussions led us to decide on a female character, based on anecdotal evidence in informal education venues that indicated young (twentyish) female role models would better attract and engage the target audience of 7- to 14-year-old visitors, especially young female visitors. It was also noted that young adult females were often perceived as more sociable and therefore more approachable that other potential role models.

Selecting the Character Model

To ensure that we were creating a figure that really appealed to the visitor base, we decided to present a survey of selections to people to see how they responded (Swartout et al. 2010). We selected a number of images of young females from a modeling agency catalog that reflected a cultural ambiguity that might appeal across a wide range of ethnicities. The museum personnel performed an internal formative research survey to see which of these most appealed to museum visitors. Museum visitors of different ages, genders, and ethnicities (selected at random) were presented with these images and were asked to choose the one with which they would be most comfortable interacting. We also collected visitor demographics and their reasons for selecting the particular human model. Based on survey results we selected a young model, about 19 years old. She was brought to the ICT to be scanned to obtain the three-dimensional (3-D) geometry and detailed facial textures that would be used to create a digital character that closely resembled her.

Twin Characters

Parallel to this, discussions continued about how to make the character most engaging to the audience. While the project initially specified a single VH museum guide, we came to understand that museum visitors, especially young visitors, might not approach such a character because they might not know what they were seeing, nor understand that they could actually interact with the VH guide. Kim LeMasters, ICT's creative director, suggested it might make sense to use two virtual characters because they could interact with each other as well as with the audience. We discussed the analogy of a "party group" where a newly arrived guest would more likely approach two people interacting than a lone individual. Presenting responses to user queries as dialogue rather than monologue has several advantages, such as allowing opportunities for humor — teasing each other rather than the visitor, and follow-up responses to known lead-ins, in cases when the user response might not be fully predictable. Moreover, as described by Piwek (2008), a number of empirical studies have found presentation in dialogue more effective than monologue for educational and persuasive purposes. Having an interactive conversation with two guides would be intriguing, and to our knowledge had never been done in a museum, but it would also be twice the work, and therefore cost prohibitive. We came up with the idea to portray them as identical twins so that the art assets could be reused, giving us approachable interacting characters, but without much of the extra cost of building a second character. It also allowed us to have the twins conversing with each other while waiting for visitors to approach, which would keep them lively and draw visitors in to see what they were saying.

Now that we had determined the look, and had scanned the model and translated her into 3-D digital twins, it was time to think about the voices we would implement for each of the individual characters. There are two approaches to create voices for intelligent virtual humans: speech synthesis or recording a human voice actor. To make the conversation between the twins as engaging as possible, and to support believability, we decided to use the recorded human voice solution. This required hiring a professional voice actor, as our model was not a voice talent. It also required us to ensure that the voice actor's speaking voice was not only appealing but actually sounded like it could belong to the visual characters.

We made recordings of several voice talents reciting a subset of phrases from the full range of responses the twins would speak. ICT and museum project teams were polled to select the voice talent recordings that objectively met certain voice and speech criteria, including 7–14 year old friendly; matches twins persona; pitch, cadence, tempo; pronunciation; conversational; fun / interesting; comfortable / natural; and Bostonlike.

We also considered using different voice talents for each twin, and doing signal processing to adjust the speaking rate and pitch, to make the two twins sound different. After reviewing simulated dialogues of each combination, the unprocessed recordings from one of the voice talents was the top choice for both twins (and we recorded a slightly altered performance for each twin voice).

Naming the Twins

Because the Cahner's Computer Place section of the Boston Museum of Science (where the twins have been installed) is dedicated to computers, robots, and electronic communication, the museum staff often referenced Ada Lovelace and Grace Hopper, two female computer pioneers from the 19th and 20th centuries (respectively) in interactions and interpretations with visitors. The names Ada and Grace therefore came up naturally in the selection process. After some discussion that they might be a bit "old fashioned" for the target audience, it was agreed there was much to be gained by using them. We felt that while they were not "typical" or modern names, they did bring a historical component to the project, as well as providing a springboard for responses and phrases that allowed us to include

Figure 1. Kiosk for the Twins.

conversations that could promote STEM role models of famous women in computer science.

Personalities / Backstory Content for VH Twins

To enhance the ability of the twins to attract and continually engage the target audience, distinct personalities along with backstories for each were created. The twin named Ada is artsy and tends to be more logical and somewhat serious; Grace is more geeky and loves to joke. The phrase/response content, as well as body and facial expressions, were developed and refined to support these personalities. Family background, school subjects, individual likes and dislikes, physical and personality traits, dress, and favorite things were decided upon and implemented within the conversational repertoire for each of the 19-year-old twin sisters.

Responses, Content, and Phrases for VH Twins

The content for the virtual twins' knowledge base was defined by the Cahners Computer Place personnel, then edited and expanded by ICT, with creative input from entertainment industry writers. The raw content and the final phrasing were optimized for the target audience of young visitors: to attract and then engage for an extended period, as well as to encourage exploration of the other exhibits and activities in the space, and even outside of the museum. In addition, multiple responses to some questions were developed and used to ensure some variety, even in the face of repeated or similar questions. This attention to detail provided the virtual characters with believability and facilitated lifelike interactions with the museum visitors. All responses were typical of young but knowledgeable female twin sisters and incorporated many personal responses as well as informational ones. The full range of responses included several distinct types of questions, including questions related to the museum itself ("What can I see here?"), questions about the space ("Can I program a robot here?"), questions about STEM ("How does a cell phone work?"), questions about the characters' technology ("How do you understand what people are saying?"), questions about the characters' background and preferences ("What's your favorite color?" or "Do you go to school?"). There were also "off topic" responses to be used if the virtual human did not fully understand a question.

	Speaker	Text
1	Visitor:	hello (ASR: HELLO, WER: 0.00)
2	Ada:	Hey there!
3	Grace:	How can we help you?
4	Visitor:	what are your names (ASR: WHAT ARE YOUR NAMES, WER: 0.00)
5	Ada:	My name's Ada.
6	Grace:	And I'm Grace. We're your Virtual Museum Guides. With your help, we can suggest exhibits that will get you thinking! Or answer questions about things you may have seen here.
7	Ada:	What do you want to learn about?
8	Visitor:	what's a virtual human (ASR: WHAT'S A VIRTUAL HUMAN, WER: 0.00)
9	Grace:	A virtual human looks and acts like a human, but it isn't really alive.
10	Ada:	That's just a clever way of saying we're a lot like you, just on this side of the screen.
11	Grace:	When you talk to us in Natural Language, we try to make sense of your words, then figure out what we should say and do in response.
12	Visitor:	how old are you (ASR: HOW OLD ARE YOU, WER: 0.00)
13	Ada:	We were first compiled just months ago.
14	Grace:	My program finished twenty minutes prior.
15	Ada:	You were a test run. But... interaction with our knowledge base does replicate conversation with nineteen year olds.
16	Grace:	Correction. With one nineteen year old, and one nineteen and twenty minutes year old.
17	Visitor:	why is there no such thing as white light (ASR: WHY IS THERE NO SUCH THING AS WHITE LIGHT, WER: 0.00)
18	Ada:	It looks white or a certain color because you can't see the individual colors that make it up. Your eyes and your brain merge the colors together, as if the color were being mixed.

Table 1. Excerpts from an Interaction between a 9–10 Year Old Male and the Twins.

These included requests to repeat or rephrase a question, as well as other responses, often using humor to redirect the conversation to another topic.

Physical Display and VH Twins' Behaviors

For maximum impact and for more natural interactions, we decided to project the twins at life size on a semitransparent 5' 10" screen material that was built into a rugged kiosk (see figure 1). A full array of body and facial naturalistic behaviors such as "fidgets" were incorporated to make the twins appear more humanlike during interaction, as well as during idle times between visitor interactions. These fidgets were continually enhanced over the project based on museum staff and visitor anecdotal input.

Interaction with the Twins

An example of the twins interacting with a visitor where speech recognition was working perfectly is shown in table 1. In the text column for the visitor utterances we first list a (human) transcription of what the visitor said and then in parentheses the speech recognizer's interpretation of what was said, and the word error rate compared to what the visitor actually said. In table 2 we show a more typical interaction where speech recognition had errors. This shows that the twins could tolerate moderate inaccuracy in recognition.

How the Twins Work

The twins have some aspects that were built in advance, and some that operate in real time as the user interacts with them. The aspects built in advance include the character bodies, animations, textual content, and spoken output. The speech recognition, natural language understanding, and dialogue management decisions of what to say are computed in real time, as is the scheduling and rendering of spoken and gestural outputs. Speech recognition, natural language understanding, and dialogue policies also make use of knowledge sources constructed in advance, using supervised machine learning.

The twins are based on and extend the approach used in the SGT Star (Artstein et al. 2009). The architecture is composed of several components that can run on one or multiple machines. The current version of the twins includes the following eight software components: The VHMSG message passing system; an audio acquisition client; one or more speech-recognition engines (ASR); a natural language understanding (NLU) component; a dialogue manager (DM); a behavior sequencer (BML Sequencer); the Smartbody behavior realization system; and the Gamebryo game engine.

People interact with the twins by pressing a push

	Speaker	Text
1	Visitor:	are you a computer (ASR: THE, WER: 1.00)
2	Grace:	Could you say that another way for us?
3	Visitor:	are you a computer (ASR: ARE YOU A COMPUTER, WER: 0.00)
4	Grace:	Yes, you have caught us. We are so busted.
5	Ada:	Our plans for world domination are ruined!
6	Grace:	Actually, we're better described as virtual humans.
7	Visitor:	who are you named after (ASR: ARE YOU NAMED AFTER, WER: 0.20)
8	Ada:	Our namesakes are Ada Lovelace and Grace Hopper.
9	Grace:	Yeah, they may be old fashioned names. But both are from the most modern computing women of their time.
10	Ada:	In the 1800s, Ada Lovelace predicted that computers would some day help us create music and create art.
11	Grace:	And in this past century, Grace Hopper created a type of computer program called a compiler, which helps people tell computers what to do.
12	Ada:	It could be said that Ada inspired Grace.
13	Grace:	Or that Grace did all the work in which case our names our perfect.
14	Visitor:	what is artificial intelligence (ASR: TELL ITALIAN, WER: 1.00)
15	Grace:	What did you say?
16	Visitor:	what is artificial intelligence (ASR: WHAT IS ARTIFICIAL INTELLIGENCE, WER: 0.00)
17	Grace:	One example of AI, or Artificial Intelligence, is 20Q, an online computer activity here at ComputerPlace that asks you questions to guess what you're thinking.
18	Ada:	I wish we'd been programmed to do that. Nah... on second thought, I prefer just answering your questions.
19	Grace:	That takes AI too. Check out 20Q and tell a volunteer what you think we have in common and what makes us different.

Table 2. Interaction with a 9–10 Year Old Female and the Twins.

to talk button and speaking into a microphone. The audio acquisition client monitors button pushes and releases, and sends audio from the accompanying microphone to one or more speech-recognition engines that are linked to the microphone. The acquisition client allows arbitrary connections between audio input devices and recognition engines. In the current configuration, we have two microphones attached: a wireless Sennheiser microphone for museum staff to demonstrate the twins to visitors, and a more rugged table-mounted Shure 522 microphone with integrated push-to-talk button for visitors to use. Initial testing with Museum visitors indicated a broad range of voice types and that we could achieve better performance by running different speech models rather than trying to build a "one size fits all" model to cover all speakers. We are currently running multiple instantiations of the Otosense speech-recognition engine, developed by the USC SAIL lab.[1] The models were adapted with transcribed audio recordings from the museum using Maximum Likelihood Linear Regression, from original children's models trained on the Colorado University children's speech database (Hagen, Pellom. and Cole 2003) and adult speech models trained on the Wall Street Journal corpus (Paul and Baker 1992). The acquirespeech client can also communicate with other speech-recognition engines, including SONIC, PocketSphinx, and several commercial speech-recognition engines.

The wireless staff microphone can be connected to a recognition model trained on that staff member or (by default) to generic adult male and female models. The visitor microphone is connected to three models — one for adult males, one for adult females, and one for children. The acquisition client retrieves output hypotheses and confidence scores from each connected recognizer. The acquisition client sends begin and end speaking signals (from button pushes) and the results from the highest confidence recognizer to the language understanding component.

The language understanding (NLU) and dialogue manager (DM) are both built using the NPCEditor software (Leuski and Traum 2010). The NLU uses cross-language information retrieval techniques to score and rank the set of possible outputs given the words recognized by ASR module. The dialogue manager uses recent dialogue history and the scores to decide which response to produce or whether to say something not directly related to the input.

The NLU works by analyzing the text of the sample utterances and the text of the responses and creating a statistical model of the "translation relationship" that defines how the content of an input utterance determines the likely appropriateness of a response. Specifically, it learns how to compute a conditional likelihood of observing a particular word in a character's response given an operator's utterance (Leuski and Traum 2008). When the NLU receives a new (possibly unseen) utterance, it uses this translation information to build a model of what it believes to be the best response for the utterance. The NLU then compares this representation to every stored response to compute a score for each response. In contrast, a traditional text classification approach would compare a new question to sample questions and then directly return the corresponding responses, ignoring the actual text of the response. We have observed that this "translation-based" classification approach significantly increases the effectiveness of the NLU for imperfect speech recognition (Leuski et al. 2006). We can see an example of NLU robustness in table 2, line 7, with a 20 percent word error rate, but correct response returned.

The DM uses the rankings from the NLU, as well as context information to decide what to say. In many cases, this will just mean returning the top-ranked NLU response. However, the twins have a large but finite set of responses (currently 158), so the characters might repeat themselves if the top answer were always selected. Thus the DM keeps track of local history and will use a lower-ranked answer if the top answer has been said recently, and the other answer has a high enough score. Also, there are cases when no answer has a high enough score to be deemed acceptable. This can happen either if the user asks a question for which there is no good answer, or speech is not understood well enough by the ASR module, or when the NLU cannot find the connection between the formulation of the question and the answer, given the training material. When there is no highly ranked answer, then the DM selects what we call an "off-topic" response. These can be either requests for the user to repeat the question (for example, line 2 or 15 in table 2), to prompt for information the twins do know how to provide, to finally giving up and moving on. Responses can consist of a sequence of dialogue from both twins, rather than just a single utterance. The DM keeps a schedule of pending utterances, and sends them one at a time to the animation components, waiting for a callback signal from Smartbody before sending the next one. If the characters are interrupted by user speech (either visitor or staff member) before the schedule has completed, the DM can cancel the remaining sequence and attend to the new utterance. Finally, the DM keeps track of how long the system has been idle, and after a fixed threshold time will select an utterance designed to attract visitors to engage in dialogue.

The SmartBody (SBM) behavior realization system (Thiebaux et al. 2008) is used to play audio files and animate the bodies for Ada and Grace. The input is in the form of behavior markup language (BML) (Kopp et al. 2006). SmartBody develops a timing schedule for audio, viseme mouth movements, and other gestures, and passes this information to the renderer through the "Bonebus" communication protocol. For the twins, the Gamebryo game engine is used as a renderer.

SmartBody currently only allows BML messages for a single virtual character. This makes it more challenging to animate the dialogue of the twins, who are highly reactive to each other's dialogue behavior. For this purpose, we developed the BML Sequencer (Aggarwal and Traum 2011), which allows artists to create synchronized behavior sequences for multiple characters on a timeline, and can convert the sequences to BML (including recursive BML calls for other characters at appropriate synchronization points). The sequencer thus takes the place of a real-time behavior generator, such as NVBG (Lee and Marsella 2006) and allows the NPCEditor dialogue manager to use the same FML/BML messages without caring about whether the animations are created by artists and retrieved at run time by the BML sequencer, or created in real time by the NVBG.

Results and Evaluation

We conducted several evaluations of the twins, including how they were received by visitors, system performance, and the like. An initial evaluation of system performance was based on data collected shortly after the first deployment, when interaction with the twins was mostly mediated by handlers at the Museum of Science (Swartout et al. 2010). This was followed by an evaluation based on data collected after the twins began interacting directly with museum visitors (reported in brief in Traumet al. [2012]), and with more detail in Aggarwalet al. [2012)]). An independent, summative evaluation also was conducted, focusing on the twins' impact on museum visitors through user observation and follow-up interviews (Foutz et al. 2012). Highlights from these evaluations are presented next.

Speech Recognition

The measure we use for assessing the quality of speech recognition is word error rate (WER), defined as the total edit distance in words (additions, deletions, and substitutions) between the actual (transcribed) speech and the speech-recognizer output, divided by the length of the actual speech; a lower score implies better performance. For the initial system, used by staff members who were familiar with the system, we had a WER of 0.26 for 6000 utterances collected between February 10 and March 18, 2010.

Recognition of museum visitors is much more challenging, however, because (1) there is a wide range of visitors, most of whom are children, (2) vis-

Model	In-domain (N = 843) Correct %	Off-topic %	Incorrect %	Out-of-domain (N = 160) Off-topic %	Incorrect %
Child	45	42	13	71	29
Female	39	47	15	72	28
Male	31	50	19	76	24
Autoselect	42	44	14	74	26
Oracle	53	35	12	75	25
Transcription	72	22	6	77	23

Table 3. NLU Performance Using the Different Speech-Recognizer Models

itors can ask anything, (3) visitors are deciding in the moment what to say, and (4) their speech has frequent disfluencies and hesitations. For the visitor condition we systematically evaluated speech recognition for a small but representative portion of the twins corpus comprising 1003 utterances recorded on a single day (in June 2011). The average word error rate was found to be 57 percent when automatically selecting the model that had the highest confidence score. The best performing individual model is the child model, with a 53 percent overall word error rate; however, using an oracle that chooses the best performing model for each utterance lowers the word error rate to 43 percent.

Response Selection Accuracy

For the initial system used by staff members, almost 70 percent of utterances were known to the system, and over 80 percent of those were responded to correctly, with only 2.5 percent answered incorrectly. To evaluate the effect of the different speech-recognizer models on the twins' responses to museum visitors, we ran the NLU on each of the speech-recognizer outputs using the same test set of 1003 utterances as in the previous section. The NLU results were compared to a gold-standard manual annotation, where each user utterance is marked as either in-domain or out-of-domain, and those utterances that are in-domain are mapped to the desired responses (Aggarwal et al. 2012).

Of the 1003 utterances in the test set, 843 are identified in the gold standard as in-domain, while the remaining 160 are out-of-domain utterances that should not receive an on-topic response. For the in-domain utterances the classifier should return a correct response; failing that, an off-topic response (indicating that the classifier did not understand the utterance) is preferable to an incorrect response. For out-of-domain utterances there is no correct response available, and therefore the appropriate classifier behavior is to give an off-topic response. NLU performance using the different speech-recognizer models is given in table 3, with percentages calculated separately for in-domain and out-of-domain utterances.

Performance on out-of-domain utterances is similar for all models, correctly identifying these utterances as out-of-domain 71–77 percent of the time. For the in-domain utterances, NLU performance matches that of the speech recognizer: the children's model yields the most correct responses, slightly outperforming automatic selection due to the large number of children's utterances in the test set. A substantial improvement is gained by using an oracle to select the appropriate speech-recognition model. Using manual transcriptions improves on the oracle performance by 19 percentage points, suggesting that improvements in speech recognition will lead to improved overall performance.

Summative Evaluation

The full twins exhibit was subject to a summative evaluation from an external, independent evaluator, the Institute for Learning Innovation (ILI). The study was designed to assess the nature of visitors' interactions with the twins, and the ways these interactions affect how children (ages 7–14) and adults relate to computer science and technology. Overall, 15 indicators were identified across the four impact areas shown in table 4, and the evaluation demonstrated that 14 of these indicators were met (Foutz et al. 2012).

Two experimental conditions were used. In the first, visitors interacted directly with the Twins; in the second, a visitor and staff member together interacted with the characters. These conditions were tested using three methods: observation of visitors while they interacted at the exhibits, in-depth interviews with visitors after their interaction, and follow-up online questionnaires 6 weeks after the initial interaction. Observational data included group size and composition, stay time, types of social interaction (between the target visitor and other visitors and between the target visitor and museum staff/volunteers), usability issues encountered while interacting with the exhibit, the number and types of questions that the visitor addressed to the twins, categorization of the twins' responses, and visits to an accompany-

Figure 2. Visitors Interacting with the Twins.

Impacts	Indicators	
	Measured	Achieved
Increased engagement and interest	5	5
Positive attitude	2	2
Increased awareness	5	4
Increased knowledge	3	3

Table 4. Summative Evaluation.

ing exhibit showing the science behind the twins. Interviews were conducted after visitors engaged with either the twins or the science behind exhibit, with the goal of collecting a paired observation and interview with the same participant. Interviews included open-ended questions and rating scale questions for use with all visitors designed to elicit visitor interest, attitudes, awareness, and knowledge of themes related to the visitor impacts. Children under 16 years of age were interviewed only after the data collector obtained permission from an adult family member in the visiting group.

Observational and interview data were collected at the museum between July 21 and September 11, 2011; online questionnaires were collected between August 20 and October 26, 2011. A total of 225 observations were collected, 180 of which were paired with interviews (for a refusal rate of 20 percent). A total of 61 follow-up online questionnaires were collected (for a response rate of 42 percent). The dialogues in tables 1 and 2 were part of this evaluation.

In this article, we present a selection of the results from the summative evaluation study showing the combined results for both conditions and illustrating each of the four impact areas. In most cases the trends are the same for the direct and blended condition; however, in some cases there are significant differences between the conditions. See Foutz et al. (2012) for the complete results of the summative evaluation.

Engagement and Interest

Time spent in the exhibit ranged from 19 seconds to nearly 18 minutes, with a median time of 3 minutes and 7 seconds ($N = 221$) (see figure 2). Quantitative rating scale questions were used to determine whether participants had a positive experience at the exhibit. Participants were asked to rate the statements "Interacting with the exhibit" and "Learning more about computers by interacting with the twins" on a four point scale, where 1 was "boring" and 4 was "exciting." The overall rating for both statements was a median of 3, or "pretty good." Participants were asked this same question six weeks later in the follow-up online questionnaire. Ratings remained the same six weeks following the original visit (Wilcoxon Signed Rank Tests).

Attitudes

The same quantitative rating scale was also used to determine if participants had positive attitudes towards speaking with the twins. When rating the statement "Being able to speak with the twins," the overall rating for all participants was a median of 3, or "pretty good." As with the engagement questions above, ratings remained the same six weeks following the original visit (Wilcoxon Signed Rank Tests).

Retrospective Pre-/Postinteraction Ratings

These were used only with adult visitors to assess change in attitudes as a result of interacting with the twins. Adults rated their agreement with four statements: (1) "I enjoy being able to speak to a computer as a way to interact with it," (2) "Having a computer with a personality is a good thing," (3) "In the future, there will be new and exciting innovations with smarter computers," and (4) "In the future, interacting with computers will be easier." Adults reported significantly higher ratings for these measures of attitudes towards computers/virtual humans directly after their interaction with the exhibit (Wilcoxon Signed Rank Tests).

Awareness

Five indicators were used to indicate awareness. For the statements "I understand what a virtual human is" and "Women have made important contributions in the field of computer science," adults showed significantly higher agreement ratings postinteraction than retrospective-preinteraction. In answers to open-ended questions, more than 90 percent of visitors were able to describe the twins as a computer that acts like a human and recognize interaction characteristics of the twins. However, only 39 percent of participants noted aspects of the connection between the twins and the main subjects of the exhibit space (computers, communications, robots) or described the twins as guides to the space. The fact that more people did not notice this connection is perhaps ironic, since we had originally thought that a major role the twins would play would be to inform visitors about other exhibits in the space. This could be due to the fact that visitors only received guidance to other exhibits if they asked about them and that as the twins' range of responses grew they became significant sources of domain knowledge themselves and engaging characters in their own right, so visitors were less likely to get responses that referred to other exhibits. Future versions of the twins could address this issue if desired by making them more proactive about suggesting other exhibits to see.

Knowledge

To determine whether study participants recognized aspects of computer science needed to create a virtual human, open-ended responses were coded for the presence of five aspects (communications technology, artificial intelligence, natural language, animation/graphics, and nonverbal behavior). Ninety-seven percent of all participants mentioned at least one aspect, while 73 percent mentioned two or more aspects. The most commonly mentioned aspect was the use of natural language, mentioned by 86 percent of participants. Sixty-four percent of on-site participants named at least one technology needed to build a virtual human; this rose to 90 percent in the follow-up six weeks later. Finally, 84 percent of participants gained at least one additional understanding about STEM (science, technology, engineering, and mathematics) domains related to the twins, while 59 percent of participants indicated they learned something new about computers or technology from interacting with the exhibit.

These results were encouraging to us as they indicated that such virtual human museum guides were successful as engaging characters and could also be used to increase interactions with, and understanding of, the museum content.

SimCoach

The next virtual human we present was designed for a very different purpose than our fun-loving, personable, and informative museum guides. SimCoach was designed to assist the large number of postdeployment military personnel returning from conflicts in the Middle East with a wide range of physical and mental health issues. As discussed below, there are a number of other differences between these virtual human systems, including user input modality (spoken versus typed), display platform (museum exhibit versus website), degree of initiative (user versus mixed system and user). Nevertheless, several architectural components are used in both systems, and both share an objective of teaching subject matter through a personal connection with a virtual human.

Problem Statement and Design

The primary goal of the SimCoach project is to break down barriers to care (for example, stigma, unawareness, complexity) by providing military service members, veterans, and their significant others with confidential help in exploring and accessing health-care content and, if needed, for encouraging the initiation of care with a live provider.

Recent advances in computer and information technology combined with the urgency of the Operation Enduring Freedom and Operation Iraqi Freedom conflicts have driven development of innovative military-focused clinical assessment and treatment approaches. This has resulted in a diversity of novel applications ranging from computerized prosthetic limbs[2] to virtual reality exposure therapy for posttraumatic stress disorder (PTSD) (Rizzo et al. 2011) to mobile smartphone apps (Luxton et al. 2011) designed to assist patients in self-management of clinical symptoms between treatment sessions. However, the Department of Defense (DOD) and U.S. Department of Veterans Affairs (VA) dissemination and delivery system needs improvement to promote awareness and access to these evolving and already

Figure 3. SimCoach Characters: Retired Sergeant Major, Civilian, Aviator, Battle Buddy.

established health-care options. The SimCoach project aims to address this challenge by supporting users in their efforts to anonymously seek health-care information and advice by way of online interaction with an intelligent, interactive, embodied virtual human health-care guide.

Rather than being a traditional web portal, the SimCoach project allows users to initiate and engage in a dialogue about their health-care concerns with an interactive VH (sometimes also called a SimCoach). A SimCoach uses speech, gesture, and emotion to introduce the capabilities of the system, solicit basic anonymous background information about the user's history and clinical/psychosocial concerns, provide advice and support, present the user with relevant online content, and potentially facilitate the process of seeking appropriate care with a live clinical provider. An implicit motive of the SimCoach project is that of supporting users who are in need to decide to take the first step toward initiating psychological or medical care with a live provider.

It is not the goal of the SimCoach project to break down all of the barriers to care or to provide diagnostic or therapeutic services that are best delivered by a live clinical provider. Rather, SimCoach was designed to foster comfort and confidence by promoting users' private and anonymous efforts to understand their situations better, to explore available options, and initiate treatment when appropriate. Coordinating this experience is a VH SimCoach, selected by the user from a variety of archetypical character options (see figure 3), that can answer direct questions and guide the user through a sequence of user-specific questions, exercises, and assessments. This interaction between the SimCoach and the user provides the system with the information needed to guide users to the appropriate next step of engagement with the system or with encouragement to initiate contact with a live provider. Again, the SimCoach project is not conceived as a replacement for human clinical providers and experts, but rather aims to start the process of engaging users by providing support and encouragement, increasing awareness of their situation and treatment options, and in assisting individuals who may otherwise be initially uncomfortable talking to a live care provider.

The options for a SimCoach's appearance, behavior, and dialogue have been designed to maximize user comfort and satisfaction, but also to facilitate fluid and truthful disclosure of clinically relevant information. Focus groups, "Wizard of Oz" studies, and iterative formative tests of the system were employed with a diverse cross section of our targeted user group to create options for a SimCoach interaction that would be both engaging and useful for this population's needs. The SimCoach system underwent formative user testing at regular intervals throughout the iterative design and development process. During the period of the project, formative testing was performed with more than 280 veterans and clinicians who treat veterans in order to collect feedback on the technical aspects, content, and interaction usability of the system to inform the design. The feedback was used to make changes to the interaction style, layout, and underlying structure of the system. Results from these user tests indicated some key areas that were determined to be important including user choice of character archetypes across gender and age ranges, informal dialogue interaction, and interestingly, a preference for characters that were not in uniform. Also, interspersed within the program are options that allow the user to respond to simple screening instruments, such as the PCL-M (PTSD symptom checklist) that are delivered in a conversational format with results fed back to the user in a supportive fashion. These screening results serve to inform the SimCoach's creation of a model of the user to enhance the reliability and accuracy of its output to the user, to support user self-awareness through feedback, and to better guide the delivery of relevant information based on this self-report data.

One way in which SimCoach characters attempt to maintain user engagement is by delivering health-care content that is relevant to persons with a mili-

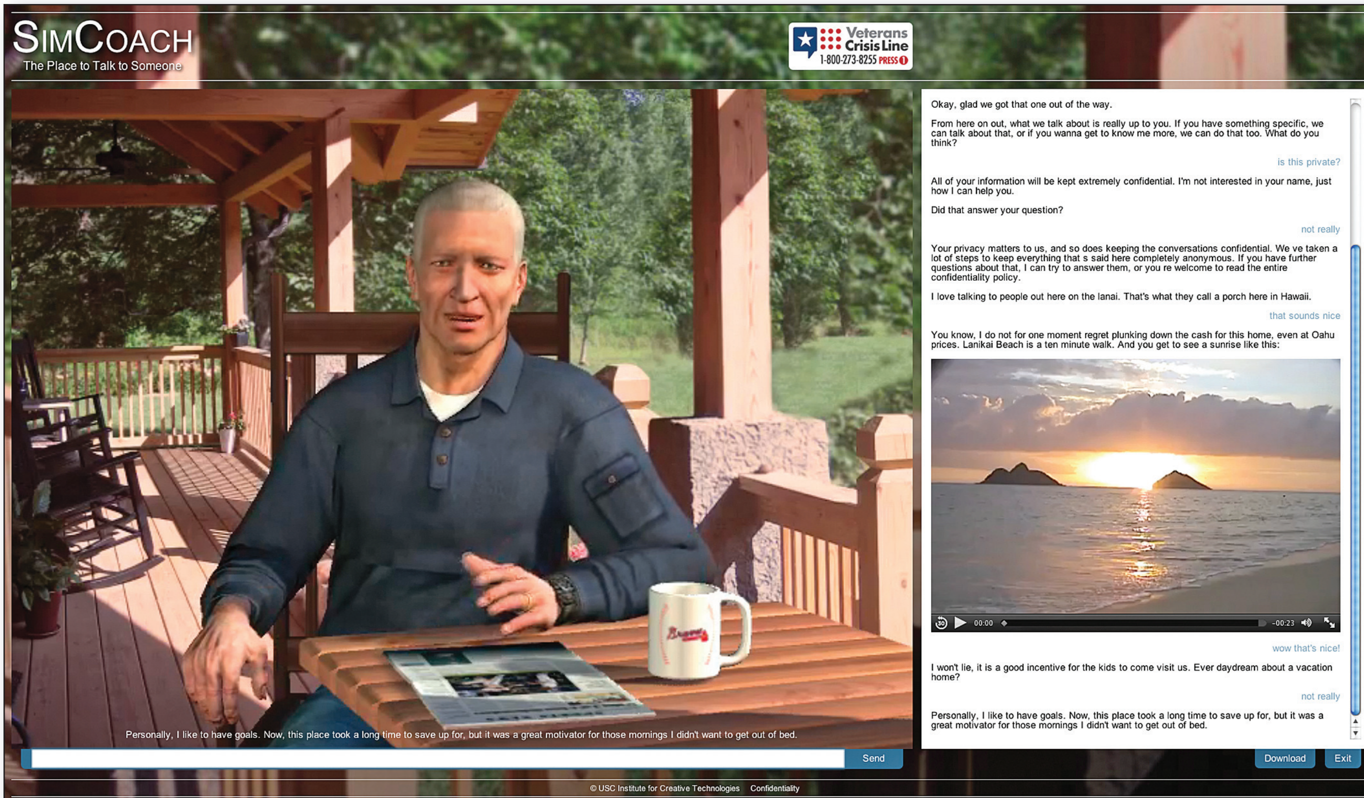

Figure 4. Interaction with SimCoach Character Bill Ford.

tary background and their families. This was addressed by leveraging content assets that were originally created for established DOD and VA websites specifically designed to address the needs of this user group (for example, Afterdeployment, Military OneSource, National Center for PTSD). Our early research with the user group indicated a hesitancy to directly access these sites when users sought behavioral health information; a common complaint was a fear that their use of those sites may be monitored and might jeopardize advancement in their military careers or later applications for disability benefits. In spite of significant efforts by the DOD and VA to dispel the idea that user tracking was employed on these sites, the prevailing suspicion led many of the users in our samples to conduct such health-care queries using Google, Yahoo, and Medscape. To address this user concern, supplemental content presented by the SimCoach (for example, video, self-assessment questionnaires, resource links) is typically "pulled" into the site, rather than directing users away to those sites.

While the Twins mostly react to visitor questions, SimCoaches are designed to go beyond purely reactive behavior, and lead the dialogue at some points, for example, delivering the PCL-M questionnaire. Further, to enhance engagement and build rapport, SimCoaches can introduce themselves and engage in small talk with users. However, SimCoaches must also be responsive to user input. This is necessary not only to put users at ease when discussing a sensitive topic, but also to accommodate the needs of a diverse user base, where some users will have specific questions and concerns, but others will prefer to let the system present suggestions and available options. This requires mixed initiative, where both the system and the user can be in control and drive the interaction at different points within a session.

The SimCoach character and interaction design process led to a virtual human–based application governed by a set of key requirements: (1) universally accessible by users, (2) relatively confidential and anonymous, (3) easy to use, (4) able to directly leverage web-based health-care resources, (5) able to be developed and improved by a variety of content authors, and (6) able to have flexible interaction strategies with mixed initiative to support a variety of users. The first four requirements strongly encouraged a web-based delivery platform, while the latter two suggested the need for a flexible dialogue management approach. Developing a web-delivered virtual human capable of mixed-initiative interactions developed by nontechnical authors became the primary technical hurdle for the SimCoach effort.

	Speaker	Utterance
1	User	<connects>
2	System	Hi there. My name's William Ford. You can call me Bill.
3	User	hello bill
4	System	First thing: the suits want me to explain that I'm a virtual human, which means that I'm based on the real experiences and personalities of actual warfighters and their families. Make sense?
5	User	Cool
6	System	Secondly, since I am virtual, if I'm a little slow on the uptake, be patient. Right now, I can pretty much only talk about PTSD and Depression.
7	User	K
8	System	Phew, glad to have that part done.
9	System	From here on out, what we talk about is really up to you. If you have something specific you wanna talk about, we can talk about that, or if you wanna get to know me more, we can do that too. What do you think?
10	User	what are the symptoms of ptsd
11	System	Everybody experiences PTSD differently, but there are a few major symptoms I'd like to mention. Avoiding anything that reminds you of the trauma, or re-living the trauma in some way, like nightmares or flashbacks, are significant signs of PTSD. Or, you could end up being on constant alert, have serious difficulties sleeping, or getting seriously startled when you really shouldn't be.
12	System	So, is PTSD something you're worried about. I only ask, because you've been asking about it.
13	User	Oh
14	System	I'm having a bit of comprehension trouble
15	System	So, is PTSD something you're worried about. I only ask, because you've been asking about it.
16	User	no its not
17	System	Do you want more info about it?
18	User	no im good
19	System	Okay. Do you want a link to a website? Or do you want to see a video?
20	User	Video
21	System	This is a video about PTSD, from afterdeployment. http://www.youtube.com/v/Emq7ouPMvuA
22	User	Thanks
23	System	You are welcome

Table 5. Interaction Between Bill Ford and an Evaluation Subject.

How SimCoach Works

Toward achieving the goals outlined above, the SimCoach virtual human architecture builds upon prior ICT virtual human efforts, including TACQ (Traum et al. 2007), ELECT-BILAT (Hill et al. 2006) and the Twins (Swartout et al. 2010). The biggest innovations from these prior systems are a web-delivered character, with more aspects compiled in advance, and a new dialogue manager to support mixed initiatives. Figure 4 shows an image of the web client, with the SimCoach "Bill Ford" to the left, a box below for the user to type in, and the dialogue history to the right. Users also have the option to print out a PDF summary of the SimCoach session. This is important for later personal review and for access to links that the SimCoach provided in the session and perhaps to provide a starting point when dealing with human clinical care providers and experts. Parts of a dialogue between Bill and a test subject are shown in table 5.

The primary system components include a web user interface (UI) for delivering animation video and accepting user text input; a text-based language classifier for natural language understanding (NLU); a dialogue manager (DM) for managing conversation information state and deciding future system actions; and an action generation process that translates the system's dialogue act response to a realized system action. This action can be a character animation performance or the insertion of an interactive web video into the client's transcript window

Offline-rendered animation video was elected in order to keep the web client standards based, in contrast to alternative approaches (Gutiérrez, Vexo, and Thalmann 2002; Rossen, Lind, and Lok 2009) and because native browser support for real-time 3-D rendering is not universally supported across browsers and therefore a risk for our target population. Web-delivered video is highly scalable given available content delivery networks and modern video codecs, allowing for high quality with minimal bandwidth requirements.

Our approach to address the notoriously challenging task of mixed-initiative dialogue is to model dialogue management using a forward-looking reward seeking agent (Morbini et al. 2012), similar to that described by Liu and Schubert (2010), but with support for complex dialogue interaction while keeping the dialogue policy authoring process accessible to those without expertise in dialogue systems. Our dialogue manager combines several methods of dialogue reasoning to promote the twin goals of flexible, mixed-initiative interaction and tractable authoring by domain experts and creative authors. Authoring involves design of local subdialogue networks with preconditions, effects, and rewards for specific topics. The dialogue manager can locally optimize policy decisions, by calculating the highest overall expected reward for the best sequence of subdialogues from a given point. Within a subdialogue, authors can craft the specific structure of interaction.

The dialogue manager is composed of four main modules. The first module is the information state (Traum and Larsson 2003), a propositional knowledge base that keeps track of the current state of the conversation. A second module is a set of inference rules that allows the system to add new knowledge to its information state, based on logical reasoning. Forward inference facilitates policy authoring by providing a mechanism to specify information state updates that are independent of the specific dialogue context. The third module is an event-handling system that allows the information state to be updated based on user input, system action, or other classes of author-defined events (such as system timeouts). Finally, the fourth module is a set of operators. Operators represent local dialogue structure, and can also be thought of as reusable subdialogues. Each state within the subdialogue can include a reward for reaching that state and ultimately determine what to do when there is more than one applicable operator. Operators have preconditions and effects. Effects specify changes to the information state. The preconditions define when an operator can be activated.

To support content development for nontechnical subject matter experts and other support staff, an authoring system was developed to support creation and validation of character data sets for this architecture. While the authoring tool was not fully instantiated prior to the development of the original SimCoach virtual character, that process provided lessons learned that drove design and development. The authoring tool, called Roundtable, is itself a web application and provides capabilities that empower many types of authors and team makeups. The system provides the ability to select from a set of configured 3-D character models, model the dialogue policy through behavior templates and more direct subdialogue editing, train the natural language understanding component, refine realized system action language and render animation performances, and test text-based and fully animated interactions within the same browser environment (figure 5). The complete character data set can be exported and deployed to a live, highly available server environment.

Finally, an animation workflow system was developed to support a variety of content authors not fluent in 3-D modeling and character animation. This subsystem utilizes a distributed computational grid and work queue to execute the following pipeline steps in batches at authoring time: (1) generation of speech audio from text (optional), (2) analysis of speech audio and generation of visemes schedule, (3) lexical and semantic analysis of text, (4) generation of behavior schedule, (5) realization of visemes and behavior schedule on a skeleton as animation keyframes, (6) import of keyframe data onto skinned rig, (7) rendering of 3-D scene with animation keyframes as stills, and (8) encoding of stills as video with merged audio. This pipeline uses existing ICT virtual human components such as a rule-based nonverbal behavior generation (Lee and Marsella 2006) (steps 3 and 4) and the SmartBody animation system (Lee and Marsella 2006, Thiebaux et al. 2008) for fused behavior realization with additional procedurally driven gaze, saccade, head nods, shakes, and blinks, for example (step 5). Additionally, we rely on commercial off-the-shelf software for speech audio analysis (step 2; FaceFX — OC3 Entertainment), 3-D modeling and rendering (steps 6 and 7; Autodesk Maya) and H.264 video encoding (step 8; FFmpeg). To allow greater levels of scale, the bottleneck to this process — preparing and rendering the 3-D scene with animation — has been parallelized such that for an animation job resulting in N frames, N / k subjobs are forked and merged at step 8, where k is a configurable parameter set appropriately for the size of the cluster available (currently set to $k = 32$).

Evaluation

Between July and December 2011, a total of 111 participants took part in user testing sessions of an alpha version of the SimCoach character. Participants were invited to take part in the testing if they were over the age of 18 and were an active military service member, veteran, or family member of a service member or veteran. The participants were asked to complete a demographic survey before interacting with one of the SimCoach characters, Bill Ford (Retired Sergeant Major). Following the interaction, participants completed a postinteraction survey and structured interview exploring their thoughts on the interaction, the character, the content and probing for issues and potential changes to the system. Eighty-eight males and 23 females, with a mean age of 41 years (range 18–76 years old) took part in the user testing sessions. The majority of the participants were discharged service members (60 percent, $n = 62$). The remainder of the participants were active duty ($n = 21$), reserve ($n = 15$), or retired ($n = 4$) service members, and 8 percent were family members of a service member or veteran ($n = 9$). Sixty-one percent of participants were members of the U.S. Army ($n = 68$), 30 percent were U.S. Navy ($n = 23$) or U.S. Marine Corps ($n = 10$), and the final 9 percent were members of the U.S. Air Force ($n = 9$) and National Coast Guard ($n = 1$). Overall the response to the interaction with the SimCoach character was positive. More than 60 percent

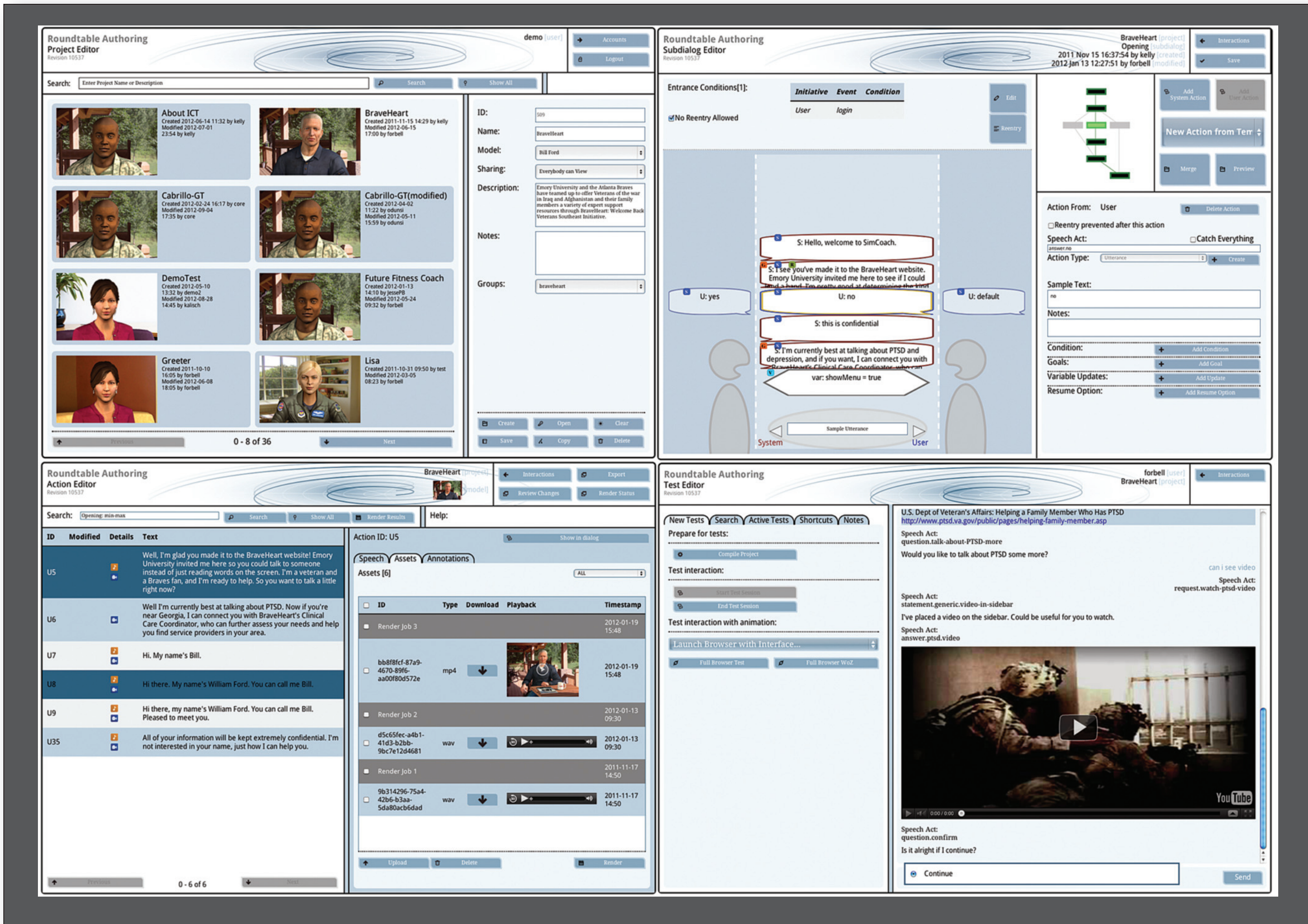

Figure 5. Roundtable Web-Based Authoring System.

of the participants responded that the SimCoach character, Bill Ford, was well informed, caring, and trustworthy. Sixty-seven percent of participants stated they felt comfortable providing the SimCoach with personal information about themselves. Only 5 percent of respondents felt that Bill gave the perception of being dishonest or unreliable. Interestingly, participants often referred to the character by his first name when describing their experience with the system, rather than talking about the program, the system, or the character. Users reported that the character responded appropriately (61 percent) and provided the appropriate information (55 percent). The majority of users felt SimCoach could be successful because of anonymity, confidentiality, and objectivity.

The current version of SimCoach is presently undergoing beta testing with a limited group of test-site users. Results from this user-centered testing will serve to advance the development of a SimCoach system that is expected to undergo a wider release in 2013.

Discussion

As the system evolves, it is our view that engagement would be enhanced if the user was able to interact with the SimCoach repeatedly over time. Ideally, users could progress at their own pace over days or even weeks as they perhaps develop a "relationship" with a SimCoach character as a trusted source of health-care information and feedback. However, this option for evolving the SimCoach comfort zone with users over time would require significant content production for revisits such that the SimCoach would be capable of describing the information acquired from previous visits and to build on that information as in a human relationship. Moreover, the persistence of a SimCoach memory for previous sessions would also require the user to sign into the system with a user name and password. Such functionality might be a double-edged sword as anonymity is a hallmark feature to draw in users who may be hesitant to know that their interactions are being stored, even if it resulted in a more relevant, less redundant, and perhaps more meaningful interaction with a SimCoach over time. Likely, this would have to be a clearly stated "opt-in" function. Given some of these hurdles, it became evident that content authoring should be sufficiently flexible and accessible to clinical professionals to enhance the likelihood that the program will evolve based on other care perspectives and emerging needs in the future.

Although this project represents an early effort in this area, it is our view that the clinical aims selected can still be usefully addressed within the limits of current technology. However, we expect that SimCoach will continue to evolve over time based on data collected from ongoing user interactions with the system and advances in technology. Along the way, this work will afford many research opportunities for investigating the functional and ethical issues involved in the process of creating and interacting with VHs in a clinical or health-care support context. As advances in computing power, graphics and animation, artificial intelligence, speech recognition, and natural language processing continue to develop at current rates, we expect that the creation of highly interactive, intelligent VHs for such clinical purposes is not only possible, but probable.

Conclusion

In this article we have described two virtual human systems that seek to engage people in meaningful ways in order to support pursuit of their own goals. These systems leverage techniques from artificial intelligence, modern 3-D gaming environments, and computer animation to simulate realistic social interactions. We argued that the social affordances of virtual humans have the potential to dramatically improve the capabilities of computers to produce successful outcomes. For example, to interact with a virtual human, users can rely upon more natural methods of communication, whether it be speech or typed input. Further, embodied virtual humans greatly expand the channel of communication back to the user through the use of nonverbal behaviors, facial expressions, speech intonation, and more. Indeed, multiple studies (Gratch et al. 2007a; Huang, Morency, and Gratch 2011) have found that people are accepting of such signals from virtual humans and that they readily adopt their usual social assumptions and behaviors in such interactions. We believe this finding has profound implications for how computers might positively affect the lives of people now and in the future.

The twins and SimCoach characters share the goals of disseminating information and conveying knowledge based on user needs, interests, and questions. They seek to emote and to establish a more powerful connection with users than would be possible with traditional UI elements, like buttons or printed text. While we are strongly opposed to simply replacing human teachers and health providers with virtual counterparts, we see no reason why virtual humans can't become part of the chorus of influences people encounter every day. For example, a broad goal of informal science education is to inspire interest in science — to increase the number of children who want to pursue science-related careers as well as to raise awareness of the scientific challenges that face the modern world. Just as fictional characters have influenced people since the introduction of novels, plays, and most recently television, we believe virtual humans could be used to enhance the pursuit of important societal goals. The studies of Ada and Grace revealed that many visitors told other people about their experience. Visitors ended up speaking with others about virtual humans outside of the museum walls and said they remembered their experience six weeks later. Surveys suggested that visitors liked the twins' senses of humor and became less concerned about the downside of technology after meeting them. We find these self-reported outcomes intriguing and certainly worthy of further attention.

We are continuing to investigate questions related to the impacts of virtual humans on learning, information seeking, and behavior change. Just as the enthusiasm and passion of top educators can be infectious, we believe virtual humans can achieve similar effects. Similarly, we think virtual humans can achieve some of the effects of the best counselors, who are able to create comfortable social atmospheres and convey genuine empathy. To achieve these goals, it is clear that continued research is needed in a variety of AI areas, such as natural language processing, emotional and cognitive modeling, nonverbal behavior generation, and pedagogical reasoning. Further, since they only receive the utterances and questions of their users, the virtual humans presented in this article have a necessarily limited view of the users interacting with them. Along with many in the intelligent virtual agents research community, we are also aggressively pursuing research on advanced user sensing and assessment to enable more meaningful interactions and greater understanding of user needs. All of this together suggests a new breed of virtual humans is coming — virtual humans that can not only entertain and educate you, but also understand you in a way that computers never have before. We hope that through judicious and psychologically informed use of these advanced technologies that this line of research will open up entirely new ways to use AI to benefit society.

Acknowledgements

The work presented here was sponsored by the National Science Foundation under Grant 0813541, by the Defense Center of Excellence for Psychological Health and Traumatic Brain Injury, and by the US Army. We also thank the inspiring staff and volunteers at the Museum of Science, Boston, as well as the extremely creative virtual human and animation teams at ICT. Any opinions, content, or information presented does not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

Notes

1. See sail.usc.edu.

2. See ScienceProg, 2007, Embedded Electronics in Prosthetic Limbs. (www.scienceprog.com/embedded-electronics-in-prosthetic-limbs.)

References

Aggarwal, P.; Artstein, R.; Gerten, J.; Katsamanis, A.; Narayanan, S.; Nazarian, A.; and Traum, D. 2012. The Twins Corpus of Museum Visitor Questions. In *Proceedings of the 8th Language Resources and Evaluation Conference*. Paris: European Language Resources Association.

Agarwal, P., and Traum, D. 2011. The BMLSequencer: A Tool for Authoring Multi-Character Animations. In *Proceedings of the 11th International Conference on Intelligent Virtual Agents,* 428–430. Berlin: Springer.

Al Moubayed, S.; Beskow, J.; Granström, B.; Gustafson, J.; Mirning, N.; Skantze, G.; and Tscheligi, M. 2012. Furhat Goes to Robotville: A Large-Scale Multiparty Human-Robot Interaction Data Collection in a Public Space. Paper presented at the International Workshop on Multimodal Corpora, Tools, and Resources. Istanbul, Turkey.

Artstein, R.; Gandhe, S.; Gerten, J.; Leuski, A.; and Traum, D. 2009. Semi-Formal Evaluation of Conversational Characters. *Languages: from Formal to Natural. Essays Dedicated to Nissim Francez on the Occasion of His 65th Birthday,* ed. O. Grumberg, M. Kaminski, S. Katz, and S. Wintner, Lecture Notes in Computer Science 5533. Berlin: Springer.

Bell, L., and Gustafson, J. 2003. Child and Adult Speaker Adaptation During Error Resolution in a Publicly Available Spoken Dialogue System. Paper presented at the 8th European Conference on Speech Communication and Technology, Geneva, Switzerland, 1–4 September.

Bickmore, T.; Pfeifer, L.; and Schulman, D. 2011. Relational Agents Improve Engagement and Learning in Science Museum Visitors. In *Proceedings of the 11th International Conference on Intelligent Virtual Agents*. Berlin: Springer.

Bickmore, T.; Pfeifer, L.; Schulman, D.; Perera, S.; Senanayake, C.; and Nazmi, I. 2008. Public Displays of Affect: Deploying Relational Agents in Public Spaces. In *Proceedings of the ACM CHI 2008 Conference on Human Factors in Computing Systems,* 3297–3302. New York: Association for Computing Machinery.

Foutz, S.; Ancelet, J.; Hershorin, K.; and Danter, L. 2012. Responsive Virtual Human Museum Guides: Summative Evaluation. Edgewater, MD: Institute for Learning Innovation.

Gratch, J.; Wang, N.; Gerten, J.; and Fast, E. 2007a. Creating Rapport with Virtual Agents. In P*roceedings of the 7th International Conference on Intelligent Virtual Agents,* Lecture Notes in Computer Science. Berlin: Springer.

Gratch, J.; Wang, N.; Okhmatovskaia, A.; Lamothe, F.; Morales, M.; Van Der Werf, R.; and Morency, L.-P. 2007b. Can Virtual Humans Be More Engaging Than Real Ones? In *Proceedings of the 12th International Conference on Human-Computer Interaction,* Lecture Notes in Computer Science. Berlin: Springer.

Gutiérrez, M.; Vexo, F.; and Thalmann, D. 2002. A MPEG-4 Virtual Human Animation Engine for Interactive Web Based Applications. In *Proceedings of the 11th IEEE International Workshop on Robot and Human Interactive Communication,* 554–559. Piscataway, NJ: Institute of Electrical and Electronics Engineers.

Hagen, A.; Pellom, B.; and Cole, R. 2003. Children's Speech Recognition with Application to Interactive Books and Tutors. In *Proceedings of the IEEE Workshop On Automatic Speech Recognition and Understanding*. Piscataway, NJ: Institute of Electrical and Electronics Engineers.

Hill, R. W.; Lane, H. C.; Core, M.; Forbell, E.; Kim, J.; Belanich, J.; Dixon, M.; and Hart, J. 2006. Pedagogically Structured Game-Based Training: Development of the Elect Bilat Simulation. Paper presented at the 25th Army Science Conference, Orlando, FL, 27–30 November.

Huang, L.; Morency, L.-P.; and Gratch, J. 2011. Virtual Rapport 2.0. In *Proceedings of the 11th International Conference on Intelligent Virtual Agents,* Lecture Notes in Computer Science. Berlin: Springer.

Kopp, S.; Gesellensetter, L.; Krämer, N.; and Wachsmuth, I. 2005. A Conversational Agent as Museum Guide — Design and Evaluation of a Real-World Application. In *Proceedings of the 5th International Conference on Intelligent Virtual Agents,* Lecture Notes in Computer Science, 329–343. Berlin: Springer.

Kopp, S.; Krenn, B.; Marsella, S.; Marshall, A.; Pelachaud, C.; Pirker, H.; Thorisson, K.; and Vilhjálmsson, H. 2006. Towards a Common Framework for Multimodal Generation in ECAS: The Behavior Markup Language. In *Proceedings of the 6th International Conference on Intelligent Virtual Agents,* Lecture Notes in Computer Science, 329–343. Berlin: Springer.

Krämer, N. C.; Tietz, B.; and Bente, G. 2003. Effects of Embodied Interface Agents and Their Gestural Activity. In *Proceedings of the 4th International Conference on Intelligent Virtual Agents,* Lecture Notes in Computer Science, 292–300. Berlin: Springer.

Lee, J., and Marsella, S. 2006. Nonverbal Behavior Generator for Embodied Conversational Agents. In *Proceedings of the 6th International Conference on Intelligent Virtual Agents,* Lecture Notes in Computer Science, 292–300. Berlin: Springer.

Leuski, A.; Patel, R.; Traum, D.; and Kennedy, B. 2006. Building Effective Question Answering Characters. Paper presented at the 7th SIGdial Workshop on Discourse and Dialogue, Sydney, Australia, 15–16 July.

Leuski, A., and Traum, D. 2008. A Statistical Approach for Text Processing in Virtual Humans. Paper presented at the 26th Army Science Conference, Orlando, FL, 1–4 December.

Leuski, A., and Traum, D. 2010. NPCEditor: A Tool for Building Question-Answering Characters. In *Proceedings of the 6th Language Resources and Evaluation Conference*. Paris: European Language Resources Association.

Liu, D., and Schubert, L. 2010. Combining Self-Motivation with Logical Planning and Inference in a Reward-Seeking Agent. In *Proceedings of the Second International Conference on Agents and Artificial Intelligence,* 257–263. Setubal, Portugal: Institute for Systems and Technologies of Information, Control and Communication.

Luxton, D. D.; McCann, R. A.; Bush, N. E.; Mishkind, M. C.; and Reger, G. M. 2011. mHealth for Mental Health: Integrating Smartphone Technology in Behavioral Healthcare. *Professional Psychology: Research and Practice* 42(6): 505–512

Morbini, F.; Devault, D.; Sagae, K.; Nazarian, A.; and Traum, D. 2012. Flores: A Forward Looking, Reward Seeking, Dialogue Manager. Paper presented at the International Workshop on Spoken Dialogue Systems, Paris, France, November 28–30.

Paul, D. B., and Baker, J. M. 1992. The Design for the Wall Street Journal-Based CSR Corpus. In *Proceedings of the DARPA Workshop on Speech and Natural Language*. Stroudsburg, PA: Association for Computational Linguistics.

Piwek, P. 2008. Presenting Arguments as Fictive Dialogue. Paper presented at the 8th Workshop on Computational Models of Natural Argument, Patras, Greece, 21 July.

Reeves, B., and Nass, C. 1996. *The Media Equation*. Cambridge, UK: Cambridge University Press.

Rizzo, A.; Parsons, T. D.; Lange, B.; Kenny, P.; Buckwalter, J. G.; Rothbaum, B.; Difede, J. A.; Frazier, J.; Newman, B.; and Williams, J. 2011. Virtual Reality Goes to War: A Brief Review of the Future of Military Behavioral Healthcare. *Journal of Clinical Psychology in Medical Settings* 18(2): 176–187.

Robinson, S.; Traum, D.; Ittycheriah, M.; and Henderer, J. 2008. What Would You Ask a Conversational Agent? Observations of Human-Agent Dialogues in a Museum Setting. In *Proceedings of the 6th International Language Resources and Evaluation*. Paris: European Language Resources Association.

Rossen, B.; Lind, S.; and Lok, B. 2009.

Human-Centered Distributed Conversational Modeling: Efficient Modeling of Robust Virtual Human Conversations. In *Proceedings of the 9th International Conference on Intelligent Virtual Agents,* 474–481. Berlin: Springer.

Swartout, W.; Traum, D.; Artstein, R.; Noren, D.; Debevec, P.; Bronnenkant, K.; Williams, J.; Leuski, A.; Narayanan, S.; Piepol, D.; Lane, C.; Morie, J.; Aggarwal, P.; Liewer, M.; Chiang, J.-Y.; Gerten, J.; Chu, S.; and White, K. 2010. Ada and Grace: Toward Realistic and Engaging Virtual Museum Guides. In *Proceedings of the 10th International Conference on Intelligent Virtual Agents*, Lecture Notes in Computer Science, 292–300. Berlin: Springer.

Thiebaux, M.; Marshall, A.; Marsella, S.; and Kallmann, M. 2008. SmartBody: Behavior Realization for Embodied Conversational Agents. In *Proceedings of the 7th International Conference on Autonomous Agents and MultiAgent Systems*. New York: Association for Computing Machinery.

Traum, D.; Aggarwal, P.; Artstein, R.; Foutz, S.; Gerten, J.; Katsamanis, A.; Leuski, A.; Noren, D.; and Swartout, W. 2012. Ada and Grace: Direct Interaction with Museum Visitors. In *Proceedings of the 12th International Conference on Intelligent Virtual Agents,* Lecture Notes in Computer Science, 292–300. Berlin: Springer.

Traum, D.; Roque, A.; Leuski, A.; Georgiou, P.; Gerten, J.; Martinovski, B.; Narayanan, S.; Robinson, S.; and Vaswani, A. 2007. Hassan: A Virtual Human for Tactical Questioning. Paper presented at the 8th SIGdial Workshop on Discourse and Dialogue, Antwerp, Belgium, 1–2 September.

Traum, D.; Larsson, R. and S. 2003. *The Information State Approach to Dialogue Management. Current and New Directions in Discourse and Dialogue*, 325–353. Dortrecht, The Netherlands: Kluwer.

William Swartout is director of technology for USC's Institute for Creative Technologies (ICT) and a research professor of computer science at USC. Swartout has been involved in the research and development of artificial intelligence systems for more than 35 years. His particular research interests include virtual humans, explanation and text generation, knowledge acquisition, knowledge representation, intelligent computer-based education, and the development of new AI architectures. In July 2009, Swartout received the Robert S. Engelmore Award from AAAI for seminal contributions to knowledge-based systems and explanation, research on virtual human technologies and their applications, and service to the artificial intelligence community, He is a Fellow of the AAAI, has served on the Executive Council of the AAAI, and is past chair of the Special Interest Group on Artificial Intelligence (SIGART) of the Association for Computing Machinery (ACM).

Ron Artstein is a research scientist at the Institute for Creative Technologies, University of Southern California. He received his Ph.D. in linguistics from Rutgers University in 2002. His research focuses on the collection, annotation, and management of linguistic data, analysis of corpora, and the evaluation of implemented dialogue systems.

Eric Forbell currently directs the research and development of the SimCoach virtual human platform at USC's Institute for Creative Technologies, bringing virtual characters to the web to support a variety of applications most notably in health care. His current focus is on building creational tools to make virtual humans commodities in the applications marketplace. Forbell received two B.A. degrees from Bowdoin College in computer science and neuroscience.

In her ten years as an evaluator, **Susan Foutz** has conducted front-end, formative, summative, and remedial evaluations of informal learning expereinces including museum and library programs, exhibits, multimedia presentations, and websites. Her research areas include public engagement with science, positive youth development, and the use of technology in informal learning settings. She holds a BA in anthropology and sociology from Ohio Wesleyan University and an MA in museum studies from the Univrsity of Nebraska-Lincoln. She is an independent evaluation consultant based in Annapolis, MD.

H. Chad Lane is a research scientist at the University of Southern California's Institute for Creative Technologies. His research is highly interdisciplinary and involves the application of entertainment and intelligent technologies to a variety of challenges, including education and health behavior. A significant portion of this work investigates the many roles virtual humans can play in virtual learning environments, such as coach and colearner, as well as in the context of mobile educational games. Lane received his Ph.D. in computer science from the University of Pittsburgh in 2004. Currently, he serves on the Artificial Intelligence in Education (AIED) Society executive committee and as program cochair for the AIED 2013 Conference.

Belinda Lange is a research scientist at the Institute for Creative Technologies and a research assistant professor in the School of Gerontology at the University of Southern California. She received her Ph.D. and degree in physiotherapy from the University of South Australia and her science degree from Flinders University. Lange's research interests include the use of interactive video game and virtual reality technologies for motor rehabilitation, exergaming, cognitive assessment, postoperative exercise, and virtual human character interactions. She is on the board of directors of the International Society for Virtual Rehabilitation and is an associate editor for the *Journal of Computer Animation and Virtual Worlds* (CAVW).

Jacquelyn Ford Morie was instrumental in the founding of the USC Institute for Creative Technologies, where she was senior research scientist from 2000–2013. She has two decades of expertise in immersive environments and their intersection with social media. She is currently a visiting scholar at Stanford University and founder and chief scientist at All These Worlds, LLC, building custom virtual worlds for health, training, art, and research.

Dan Noren was program manager of Cahners Computer Place at the Museum of Science, Boston, during the design, development, and installation of the twins at the museum. Now retired from the Museum of Science, he is an independent consultant for virtual human knowledge base content, exhibit design, and operational support.

Albert "Skip" Rizzo is a clinical psychologist and associate director at the University of Southern California Institute for Creative Technologies. He is also a research professor with the University of Southern California's Department of Psychiatry and at the University of Southern California Davis School of Gerontology. Rizzo conducts research on the design, development, and evaluation of virtual reality systems targeting the areas of clinical assessment, treatment, and rehabilitation across the domains of psychological, cognitive, and motor functioning in both healthy and clinical populations.

David Traum, Ph.D., is a principal scientist at the Institute for Creative Technologies (ICT), and a research assistant professor in the Computer Science Department, both at the University of Southern California. He completed his Ph.D. in computer science at University of Rochester in 1994. His research focuses on collaboration and dialogue communication between human and artificial agents.