Ada and Grace: Toward Realistic and Engaging Virtual Museum Guides

William Swartout[1], David Traum[1], Ron Artstein[1], Dan Noren[2], Paul Debevec[1], Kerry Bronnenkant[2], Josh Williams[1], Anton Leuski[1], Shrikanth Narayanan[3], Diane Piepol[1], Chad Lane[1], Jacquelyn Morie[1], Priti Aggarwal[1], Matt Liewer[1], Jen-Yuan Chiang[1], Jillian Gerten[1], Selina Chu[3], and Kyle White[3]

[1] USC Institute for Creative Technologies
[2] Museum of Science, Boston
[3] USC Speech Analysis and Interpretation Laboratory

Abstract. To increase the interest and engagement of middle school students in science and technology, the InterFaces project has created virtual museum guides that are in use at the Museum of Science, Boston. The characters use natural language interaction and have near photoreal appearance to increase and presents reports from museum staff on visitor reaction.

Keywords: virtual human applications, photoreal characters, natural language interaction, virtual museum guides, STEM, informal science education.

1 Introduction

A well-informed guide or interpreter can have a tremendous influence on the quality of a museum visitor's experience. The best guides not only provide information but also engage the visitor in an interactive exchange that can lead to deeper understanding and promote excitement about museum content. Unfortunately, human museum guides are often in short supply. Many studies have shown that people react to virtual humans in much the same way that they react to real people [1-3]. Could virtual humans be used to create museum guides that can engage visitors with museum content? The InterFaces project, a collaboration between the USC Institute for Creative Technologies (ICT) and the Museum of Science, Boston (MoS), has been exploring exactly that question.

Fig. 1. Guides at the Museum of Science, Boston

Set in Cahners ComputerPlace (CCP) at the Museum of Science where most of the museum's information technology exhibits

J. Allbeck et al. (Eds.): IVA 2010, LNAI 6356, pp. 286–300, 2010.

are located, the virtual human guides are housed in exhibit called InterFaces (Figure 1) designed to promote interest in Science, Technology, Engineering and Mathematics (STEM). The primary audience that we sought to reach was children between ages 7 to 14, and we were particularly interested in engaging females and other groups under-represented in STEM. We chose middle school aged children as our primary audience for two main reasons.

First, recent studies such as [4] suggest that children's level of interest in science during middle school or even earlier can have a strong effect on ultimate career choice. Our hypothesis is that interacting with Virtual Humans will help pique and engage children's interest in what computer science and related STEM can offer because not only can they provide knowledge and advice, as computers typically do, but the fact that they are embodied as Virtual Humans adds a social element that can create greater rapport and involvement.

Second, it is sometimes difficult to get younger museum visitors to engage with the exhibits, particularly when in school groups. Museum personnel report that school group behavior tends to be "run in, run around, run out." Students spend relatively little time actually engaged with the exhibits and teachers struggle to guide them to interact with the many choices available. In our project, however, visitors encounter Ada and Grace, twin virtual museum guides who are a focal point of the space, when they first enter the ComputerPlace. Ada and Grace are life-sized, photo-realistic characters that interact in natural language, complete with gestures and other forms of non-verbal communication. In our current implementation, a museum staff member interacts with the virtual museum guides in natural language and the visitors who can pose their own questions to the guides through the staff member. The virtual museum guides answer general questions about ComputerPlace and the information sciences, and based on visitor's expressed interests, they can suggest exhibits to check out. Our hope was that the cutting edge technology of the virtual humans as well as the social rapport they could establish would engage the students and "stop them in their tracks."

1.1 Creating Engagement

Creating an engaging experience for museum visitors is a central goal of the Virtual Museum Guides project. Several facets of the virtual humans' design are intended to increase engagement:

- **Broadly appealing appearance.** We conducted a study (described in Section 4.1) to select a (human) model for our guides that was not clearly identified with any one ethnic group and had broad appeal to museum visitors.
- **Two characters.** When virtual humans have been used as information agents or guides in the past, in most cases, there is a single virtual human interacting with one or more real people. To enhance engagement, we decided to use two characters so that they could dialogue with each other as well as the visitors. In section 4.3, we describe our rationale for this approach in more detail.
- **Natural language interaction.** We chose to have Ada and Grace interact using natural language input and output rather than a menu-based or type-in interface because it makes the interface more transparent and the characters more realistic.

- **Near photoreal appearance.** Studies have shown that a more realistic depiction of either a simulated scene [5] or virtual human [6] can create greater participant involvement in a virtual experience. As we will describe below, Ada and Grace, our virtual museum guides, were created using Light Stage technology. That technology has been used to create photorealistic non-interactive characters for movies such as *Avatar, Hancock,* and *The Curious Case of Benjamin Button*. Ada and Grace are the first examples of deployed, interactive virtual humans that make extensive use of Light Stage technology (See Section 4.2).

1.2 Related Work

There have been some previous installations of Virtual Humans in museums. The "pixie" system [7] 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. In January 2004, the Max agent was installed in the Heinz Nixdorf Museums Forum (HNF), a public computer museum in Paderborn (Germany) [8]. Max is human-like 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 ComputerPlace, the Tinker exhibit [9], an embodied conversational agent (ECA) in the guise of a robot, builds relationships and tracks engagement with the user, and includes novel user ID via analysis of hand-images. Visitors must select options from a menu to interact with Tinker.

In the next section, we present a sample interaction with Ada and Grace, followed by a more detailed discussion of how they work and were developed. We then present evaluations of the Twins' natural language processing, and reports from the museum staff about visitor reaction to the deployed virtual museum guides.

2 Sample Scenario

In our current implementation, a museum staff member uses speech to interact with the guides. Visitors may pose questions to the guides through the staff member. We choose to take this approach rather than having visitors speak directly to the guides to ensure speech recognition accuracy, which is less accurate when there is wide variability in dialects among speakers. This first group of interchanges illustrates a drill-down exchange in which the twin characters, Ada and Grace, are leading a visitor (via the Handler) towards information the visitor is interested in:

HANDLER: What are your names?
ADA: My name's Ada.
GRACE: And I'm Grace. We're your Virtual Museum Guides. With your help, we can point you towards exhibits that will get you thinking! Or answer questions about things you've seen here.
ADA: What do you want to learn about?

HANDLER: I'm interested in computers.
GRACE: Here at Cahners ComputerPlace, you can *build* your own model computer, solve a *puzzle* like a computer, examine computer *brains*, and learn how to *think* like a computer. Which one of these do you want to put through your processor?

HANDLER: I want to learn how computers think.
GRACE: At Logic Land you can learn how computers think by trying to decode a bitmap, and interacting with Boolean logic gates.

HANDLER: What are Boolean logic gates?
GRACE: Boolean logic gates are the basic building blocks of computing! Remember these three words; AND, OR, NOT. Each of these is a type of logic gate. Like real gates you walk through, some need a key, or even two identical keys to let you through. Head over to Logic Land, or ask one of our volunteers if you want to learn more.

Other common exchanges include information about the Museum of Science and Cahners ComputerPlace:

HANDLER: Why is this place named Cahners ComputerPlace?
ADA: Cahners ComputerPlace is named after Norman Cahners, a publisher based in Boston, and longtime supporter of the museum.
GRACE: Welcoming 300,000 visitors annually, Cahners ComputerPlace offers a one-stop resource for software that inspires people to create, explore and learn.
ADA: Did you read that in the brochure?!

The twins are also capable of responding to questions about their own exhibit and supporting technology:

HANDLER: What is your technology?
ADA: We're virtual humans. We use speech recognition technology to identify your words…
GRACE: [Finishing her twin's sentence] …and use statistics to figure out the meaning of the words based on context. Once we know what you're talking about, we'll reply appropriately.

Grace and Ada are also intended to highlight the major role that women have played historically in the development of computer science:

HANDLER: Who are you named after?
ADA: Our namesakes are Ada Lovelace and Grace Hopper.
GRACE: Yeah, they may be old-fashioned names, but both from the most modern computing women of their time.
ADA: In the 1800s, Ada Lovelace predicted that computers would some day help us write music and create art.
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.
ADA: It could be said that Ada inspired Grace.
GRACE: Or that Grace did all the work, in which case our names are perfect.

Finally, Ada and Grace can answer questions that middle schoolers might be interested in:

HANDLER: Do you have boyfriends?
ADA: Why? [Tongue in cheek] Is there another virtual human you think we should meet?
GRACE: I hope they're into neural networks and amino acid sequences.
ADA: Grace?! What's wrong with just charming and generous?

3 How the Guides Work

The Twins are based on and extend the approach used in the SGT Star character [10] and the Gunslinger project [11]. To interact with the characters, an operator presses a push-to-talk button and speaks into a microphone. An audio acquisition client then sends audio to the automatic speech recognizer (ASR), which creates hypotheses of the words that were said, and then sends this text to the Language Understanding (LU) module. The ASR module uses the SONIC toolkit [12], with custom language and acoustic models (see Section 4.4). The LU module uses a statistical text classification algorithm to map the utterance text onto character responses. It selects a set of responses it believes to be appropriate to the utterance from a domain-specific library of scripted responses and passes them to the dialogue management (DM) module. The DM module uses that response set and the recent dialogue history to select one response for the characters to perform. The response is sometimes a sequence of utterances, as seen in section 2. In this case, the DM keeps a schedule of pending utterances, and sends them one at a time to the animation components, waiting for a callback signal before sending the next one. If the characters are interrupted by more speech from the operator before the schedule has completed, the DM can cancel the remaining sequence.

The LU/DM module pair uses the NPCEditor software [13]. The NPCEditor classification algorithm analyzes the text of the sample utterances and the text of the responses and creates 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 [14]. When NPCEditor 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 classifier then compares this representation to every stored response and returns the best match to the DM part of NPCEditor. 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 classifier for imperfect speech recognition [14]. NPCEditor has been fielded in a number of applications and has been shown to be successful in noisy classification tasks [13].

The Twins have a large but finite set of responses (currently about 400), so the characters might repeat themselves. One of the tasks of the DM is to match the

classifier selection to the recent dialogue history and choose responses that have not been heard. The DM also handles cases when the classifier returns no responses. This happens when the operator asks a question for which the characters have no answer or speech is not understood by the ASR module. In that case, the classifier decides that none of the known answers is appropriate. The Twins database contains a number of responses that we call "off-topic." These responses range from prompts for repetition "Could you ask that again?" to utterances indicating that the characters do not know how to answer the questions "I really wish we had an answer for that."

The animation process is revised from that used by SGT Star and employs the Smartbody (SBM) behavior realization system [15] and a new sequencer module, as well as the Gamebryo animation engine. The sequencer module retrieves Behavior Markup Language (BML) [16] animation schedules for each of the utterances coming from the DM. Since BML as interpreted by SBM only animates a single agent, the sequencer schedule includes a number of synchronization points that are broadcast back to the sequencer. When the sequencer receives these callbacks it sends additional BML schedules to animate the other agent, so that Ada and Grace can each react appropriately while the other is speaking. SBM uses several behavior controllers and blending to realize the specific combination of motion, and sends the resulting commands to the Gamebryo engine to generate the motion.

4 Building the Guides

In this section we outline the main steps we took to create the Twins. We start with the appearance, then the content and output expression, and finally the resources for speech understanding.

4.1 Formative Study on Visitor Preferences

To support the goals of engagement and ability to serve as role models for young girls potentially interested in STEM, we decided to base the characters' appearance on a young adult female of indeterminate racial background. We conducted a formative study with 75 museum visitors from the target audience of 7-14 year olds (with parental consent), which was used to inform the choice of character appearance as well as impressions the visitors associate with the person shown. Six photos (selected from a larger set provided by a modeling agency) were presented and visitors were asked to select the one whom they would most want to speak to in Cahners ComputerPlace and provide reasons for their choices. In addition, visitors were probed for ideas about what the virtual human might do with their free time, what characteristics of a virtual human guide would be most important to them, and what interested them about computers and robots. Results from the exit survey indicate that one photo (actress/model Bianca R.) was the overwhelming choice of the museum visitors, and thus she was selected to be recorded in the Light Stage. Visitors rated the following traits as being most important for their virtual human: Friendly, Smart and Patient. Visitors reported that their virtual human occupied their time: having a job/occupation, having pets, going to the mall/shops, hanging out with friends, and playing sports. This information was used in the content development (described in Section 4.3) to help craft the backstory for the characters.

4.2 Light Stage: Capturing the Model

Fig. 2. Light Stage 5

We recorded Bianca Rodriguez using ICT's Light Stage 5 high-resolution facial scanning system (Figure 2) that enables the creation of characters that appear and animate realistically, and look substantially better than standard video game characters that visitors might be familiar with. Light Stage 5 is a two-meter diameter sphere with 156 evenly-spaced white LED light sources pointing toward the center from all angles. A stereo pair of high-resolution digital still cameras photographs the actor's face under a variety of different lighting conditions as in [17]. Polarized lighting conditions allow us to independently measure the skin color, surface shine, and surface orientation at each point on the face with 0.1mm resolution. Using spherical harmonic lighting conditions – essentially bright-to-dark gradients of light across the sphere's X, Y, and Z axes – allows us to measure the surface orientation at each pixel, telling us the shape of skin pores, creases, bulges, and wrinkles. From that data we created a highly detailed 3D model of the subject's face.

The first use of the Light Stage system for creating a virtual character was for SGT Star. Based on the deployed version of the character, an advanced prototype was created by leveraging the hybrid normal rendering skin technique from [17] where the diffuse and specular reflectance components of the skin are rendered with different surface orientation maps as measured from the photometric data.. This significantly increased the realism of the character with little impact on rendering speed(Figure 3). The face looked quite realistic in a neutral pose but less convincing as it animated, since it distorted unnaturally when morphed to form expressions.

Fig. 3. Hybrid Normal

To improve upon SGT Star's facial quality for the Twins, we acquired scans of the actor in a variety of facial expressions to be used as blend shapes in the animated character rig. Thus, when Ada or Grace exhibit an expression, the shape of her face is based on the actual shape of the actor's face in that expression. The complexity of facial expression is difficult to model manually, making this blend shape data very valuable for creating believable digital characters.

We had previously built a character from light stage scans in a variety of expressions, most notably in the "Digital Emily" project [18] in collaboration with facial animation company Image Metrics. However, Digital Emily was rendered offline using computationally intensive light transport simulations. For the museum guides, we further developed the real-time skin shader using the hybrid normal maps technique of [17] to render a faithful rendition of skin reflectance in the Gamebryo game engine. The deployed Twins at the museum use this rendering technology. A detail of one of the Twins' faces produced by a more advanced version of

the shader is shown in Figure 4[1]. The neutral pose of the guides was based directly on fitting an animatable facial mesh to the original high-resolution scan data of the model's neutral pose. The blend shapes for the expressions were created semi-automatically by using approximately seven expression scans as reference in a 3D modeling program. The resulting models were exported in formats compatible with off the shelf tools such as Maya.

While the Twins' faces were being created using Light Stage technology, the rest of the characters (bodies, clothing, hair, eyes etc) were created by digital artists. These elements were then brought together with the Light Stage models to create the characters shown in Figure 1.

Fig. 4. Twins Detail Enhanced Shader

4.3 Developing the Content

As described above, the main goals for interaction were to involve 7-14 year old kids in natural, engaging conversation related to STEM and the Museum's related exhibits in Cahners ComputerPlace. Rather than a single guide, we decided to use twins to enhance visitor engagement in several ways. First, some character responses are quite information-rich and inherently lengthy. Having one character deliver long responses can seem long-winded and tax the attention span of younger visitors. By having two characters share such responses (as exemplified in Section 2), we can better maintain the pace of the conversation and visitor interest. Second, two characters can be an obvious source of differing opinions and behaviors, which allows a dialectal approach to providing information [19] as well as allowing the characters to act as foils for each other's humor. There is also some evidence that presenting different types of information as coming from different agents may enhance learning over having all information come from a single agent [20]. The decision to use twins as opposed to two distinct characters was mainly to reduce production costs and allow maximal reuse of resources, but it also provides a good backstory for their interaction.

There were six content areas developed for the twins:

1. Cahners ComputerPlace (CCP) exhibits, activities and exhibit space namesake
2. General computer, robot, and cell phone communications
3. Overview of the Museum of Science
4. Backstory about the characters (favorite color, pets, etc)
5. Technology of the Virtual Human Guides
6. Off topic responses (triggered by un-interpretable inputs)

For areas 1-3, the CCP staff collected typical visitor questions and interpretation responses given by the staff and volunteers. From a base of over 300,000 visitors per

[1] This advanced version has not yet been released in the museum. We expect to release it in June 2010.

year, we were able to compile comprehensive, detailed questions and answers from the viewpoint of many different visitor demographics - with a range of Computer STEM skills and knowledge. For topics 4-6 we were able to rely on ICTs previous experience with Virtual Humans, such as SGT Star. An iterative process involving both groups led to the final content.

A single voice actor was cast to create the slightly different voices of Ada and Grace. These recorded lines were then used as the basis for animation, using the authoring component of the sequencer to allow artists to select appropriate animations for the characters. The animations were designed to match the personalities of the characters as well as engage the visitor.

4.4 Speech Recognition

The SONIC toolkit [12] and the SRI Language Modeling Toolkit (SRILM) [21] were used to create the acoustic speech and language models, as well as to provide an API for on-line speech recognition. Language models were constructed by combining a large vocabulary (5-15k words) with the full set of inputs used for classifier training.

Several acoustic models were built, customized to individual Museum staff member's voices, using gender-dependent three state triphone context HMM acoustic models trained from the Wall Street Journal corpus as well as around 250 utterances for each speaker. Twelve Mel Frequency Cepstral Coefficients (MFCCs) and normalized frame energy, along with the velocity and acceleration of these features, are used for audio frame representation. Systems were built via three iterations of Maximum a Posteriori adaptation on the baseline gender dependent model (e.g., [22]). The performance (word error rate) of the adapted systems improved by 33% (from 15% to 10%) compared to the baseline. A number of engineering optimizations helped improve the overall performance robustness. For example, allowing for a larger beam path in the speech decoding process, yielded a performance gain. Similarly, optimizing the adaptation function parameters for frame count threshold and silence count threshold increased performance as well, especially in the presence of acoustic variability such as background noise.

5 Additional Project Elements

As part of the project, we are developing two additional exhibit elements aimed primarily at our secondary audience, older teen-agers and adults:

The Science Behind Virtual Humans. Because many sophisticated computer science research areas are required to create virtual humans, in addition to serving as a guide, a virtual human can *itself* serve as an exhibit of technology. In its current form, the "Science Behind" exhibit (Figure 5) consists of flat panel displays on the side of the virtual guides kiosk (Figure 1) that dynamically show the virtual human's speech recognition and statistical NLU text classifier in operation as the characters interact. It also includes several posters that describe different stages of the installation design and construction. Visitors can watch as the system recognizes the words in the handler's speech and see how the classifier ranks and then selects a response. Another window shows a transcript of recent interactions. These supporting exhibits engage

visitors by allowing to see first-hand the cutting edge of technology and grasp the promise and limitations of current virtual humans.

Living Laboratory. As part of the exhibit, we include a "Living Laboratory", which engages the museum visitors in the scientific method (as applied to virtual humans) in three different ways. First, visitors can be experimental subjects, just by interacting with the virtual humans. In a standard university research laboratory one of the most difficult aspects of advancing the state of user interaction with virtual humans is finding enough appropriate subjects to evaluate the system. Thus, only a small set of the experimental conditions that are worth testing can actually be accomplished. The thousands of visitors to the museum provide a much larger pool from which to test a number of issues, such as performance of the speech understanding, coverage of the domain, appropriateness of the dialogue strategies, and effectiveness at teaching and motivating interest in STEM. Secondly, visitors can help evaluate the data and analyze the results. Finally, through interaction with the museum staff and on the Exhibit website, visitors can suggest new experiments.

Fig. 5. Museum visitors exploring the Science Behind exhibit

6 Evaluation

The Museum Guide Twins were first displayed to the public on December 8, 2009. We have since conducted evaluations of the Twins' natural language performance and we have reports from museum staff about how visitors are reacting to the Guides, which we discuss below. In the future, we will conduct summative evaluations to assess the impact that the Guides have on a museum visitor's experience, and their engagement and interest in STEM topics.

Evaluation of Natural Language Performance

We evaluated the Twins' performance based on data collected at two venues: ongoing live sessions at the Museum of Science between February 10 and March 18, 2010, and a demo at the AAAS annual meeting in San Diego on February 19-21. The data

consist of system logs and audio recordings of utterances spoken to the characters; the vast majority of the utterances are by trained museum staff, though occasionally a visitor spoke directly to the Twins. All recordings were transcribed manually.

Utterances spoken to the characters can be divided into those that appear in the classifier training data (*known* utterances) and those that are not in the training data (*unknown* utterances). Since speech input to the characters is provided primarily by museum staff familiar with the Twins, we found a large proportion of known utterances (about 70%); unknown utterances usually come about when the interpreter diverges from the standard questions, for example, by posing a question asked by a visitor. For known utterances we can automatically determine whether the response was correct (by seeing if it is linked to the utterance in the classifier), incorrect (an on-topic response that is not linked to the utterance), or off-topic. For unknown utterances there are no defined correct responses, but we can automatically determine whether the response was on-topic or off-topic. Table 1 shows the breakdown of responses.

Table 1. Responses from the Museum of Science, February 10 to March 18, 2010

Question	Response	N	%	WER
Known	Correct	3516	56.8	0.1726
Known	Incorrect	106	1.7	0.8261
Known	Off-topic	629	10.2	0.5829
Unknown	On-topic	1444	23.3	0.2222
Unknown	Off-topic	498	8.0	0.4543
Total		6193	100.0	0.2597

The results show that performance on the known utterances is good, with over 80% of known utterances receiving a correct response; those known utterances that received off-topic and incorrect responses typically had higher word error rates (WER), so the failure of the classifier is likely due to poor speech recognition. Unknown utterances also result in mostly on-topic responses. To better understand the performance on the unknown user utterances we used a sample of the data (all the data collected at the museum between February 10 and February 19) to perform two manual annotation tasks: separating the unknown utterances into in-domain and out-of-domain utterances, and rating the coherence of system responses.

Unknown user utterances can be divided into two types: in-domain utterances which have a good on-topic responses and out-of-domain utterances which do not have an on-topic response in the characters' repertoire. In-domain utterances are typically minor variations on known utterances, and for such input the classifier is expected to provide the correct response; out-of-domain utterances are often not related to any known utterance, and the dialogue manager should handle these by issuing an off-topic response. Since the definition of in-domain and out-of-domain utterances depends on the desired system output, determining which class an utterance belongs to is a somewhat subjective task which has to be performed manually. To ensure the annotations were meaningful we had the sample data marked by two annotators, and calculated inter-rater reliability using Krippendorff's alpha [23]; reliability was reasonably high at α=0.75 (observed agreement=0.89, N=264; alpha ranges from −1 to 1, where 1 signifies perfect agreement and 0 obtains when agreement is at chance level).

To assess the quality of the responses we conducted a separate rating study, similar to [10], where annotators rated utterance-response pairs on a scale of 1 to 5. All the utterance-response pairs collected at the museum between February 10 and February 19 were rated. A reliability study on a separate sample, the utterance-response pairs from the AAAS demo, showed that reliability was fairly high for the on-topic responses, with α=0.827 for unknown on-topic responses and α=0.596 for known incorrect responses, but negative for the off-topic responses, indicating that the ratings of the latter cannot be trusted. Table 2 shows the ratings of the 390 on-topic responses (out of 582 total utterances analyzed). The table shows that on-topic responses to unknown utterances are generally very good, especially for those user utterances that are in-domain.

Table 2. Coherence ratings for On Topic Responses to Unknown Utterances

Question	N	Mean	Median
In-domain	342	4.78	5
Out-of-domain	48	3.40	4

Interaction reports from museum staff

While we have yet to conduct formal studies, anecdotal reports from the museum staff are encouraging. Museum staff reports that the exhibit really does 'stop the kids in their tracks' when the Twins are talking. In idle mode, when the Twins are not interacting, most visitors pause then walk on, whether or not a handler is present. When the handler is interacting with the Twins and a visitor walks by, a significant percentage stop, with a majority of them staying and interacting.

For most exhibits in Cahners ComputerPlace, adults accompanying children in families tend not get involved themselves. In contrast, with the Twins, the entire family tends to get involved. Females seem to be attracted to the exhibit more than males, and they tend to stay longer. There is some reticence for visitors to ask questions, although females tend to ask questions more spontaneously than males and these questions tend to be more personal questions about the Twins.

Visitors do not immediately make the connection between what is going on at "The Science Behind" and what the Twins are doing, they think "The Science Behind" stands alone. This suggests a need clarify the connection in the exhibit design.

Fig. 6. Visitors engaging with Ada and Grace

The natural language technology tends to engage the visitors attention, and there is some real amazement at the exhibit. Staff has literally observed jaw-dropping reactions from visitors to the Twins.

An email sent by Dan Noren, director of Cahners Compter Place and a co-author of this paper, shortly after the Twins debut, sums up the initial observations:

> "Well, the young visitors are enchanted! Lots of *"Awesome", "Wow", "Really cool", "Neat"...,* lots of smiles / wide eyes, lots of questions - and both girls and boys are interested in looking at the Science Behind and the actual computers / networks driving the whole thing. I believe InterFaces is everything we have been working so hard to do - give the WOW factor to Computer STEM."

7 Future Work and Conclusions

In the near future, we intend to enhance ASR to support direct interaction between the Twins and museum visitors. We are investigating several approaches to rapidly selecting or adapting speech models to visitors. Other future enhancements include more expressive facial expressions and eye gaze, and idle behaviors in which the characters will interact with each other when no one is talking with them to help draw in visitors.

A major goal for this project was to create virtual guides that would truly engage visitors. We sought to do this through several means. We surveyed visitors to help design a character that would be broadly appealing. We used two characters instead of one so that the characters could interact with each other as well as the visitors and increase engagement. We used natural language input and output for a more natural interface, and we used Light Stage technology to capture highly realistic models of the characters' faces, and developed the technology to render those models in realtime within the Gamebryo game engine.

Our evaluation of the Twins' performance shows the feasibility of using natural language interaction, and we believe the pictures of visitors and the reports from the museum staff give strong evidence of success in creating engagement.

Acknowledgements. This material is based upon work supported by the National Science Foundation under Grant 0813541. We thank the staff and volunteers of Cahners ComputerPlace for their support. We also thank Kim LeMasters, Creative Director of the ICT, for the suggestion to use twins, Ed Fast for software support, and Stacy Marsella and Andrew Marshall for help with the Smartbody system. Finally, we would like to thank Arlene de Strulle for her continued support and enthusiasm.

References

1. Reeves, B., Nass, C.: The Media Equation. Cambridge University Press, Cambridge (1996)
2. Krämer, N.C., Tietz, B., Bente, G.: Effects of embodied interface agents and their gestural activity. In: Rist, T., Aylett, R.S., Ballin, D., Rickel, J. (eds.) IVA 2003. LNCS (LNAI), vol. 2792, pp. 292–300. Springer, Heidelberg (2003)

3. Gratch, J., Wang, N., Okhmatovskaia, A., Lamothe, F., Morales, M., van der Werf, R., Morency, L.-P.: Can virtual humans be more engaging than real ones? In: 12th International Conference on Human-Computer Interaction, Beijing, China (2007)
4. Tai, R., Liu, C., Maltese, A., Fan, X.: Planning early for careers in science. Science(Washington) 312, 1143–1144 (2006)
5. Slater, M., Khanna, P., Mortensen, J., Yu, I.: Visual realism enhances realistic response in an immersive virtual environment. IEEE Computer Graphics and Applications 29, 76–84 (2009)
6. MacDorman, K., Coram, J., Ho, C.-C., Patel, H.: Gender Differences in the Impact of Presentational Factors in Human Character Animation on Decisions in Ethical Dilemmas. Presence: Teleoperators and Virtual Environments 19 (2010)
7. Bell, L., Gustafson, J.: Child and adult speaker adaptation during error resolution in a publicly available spoken dialogue system. In: EUROSPEECH-2003, pp. 613–616 (2003)
8. Kopp, S., Gesellensetter, L., Krämer, N., Wachsmuth, I.: A conversational agent as museum guide - design and evaluation of a real-world application. In: Panayiotopoulos, T., Gratch, J., Aylett, R.S., Ballin, D., Olivier, P., Rist, T. (eds.) IVA 2005. LNCS (LNAI), vol. 3661, pp. 329–343. Springer, Heidelberg (2005)
9. Bickmore, T., Pfeifer, L., Schulman, D., Perera, S., Senanayake, C., Nazmi, I.: Public displays of affect: deploying relational agents in public spaces. In: CHI 2008, pp. 3297–3302. ACM, New York (2008)
10. Artstein, R., Gandhe, S., Gerten, J., Leuski, A., Traum, D.: Semi-formal evaluation of conversational characters. In: Grumberg, O., Kaminski, M., Katz, S., Wintner, S. (eds.) Languages: From Formal to Natural. LNCS, vol. 5533, pp. 22–35. Springer, Heidelberg (2009)
11. Hartholt, A., Gratch, J., Weiss, L., The Gunslinger Team: At the virtual frontier: Introducing gunslinger, a multi-character, mixed-reality, story-driven experience. In: Ruttkay, Z., Kipp, M., Nijholt, A., Vilhjálmsson, H.H. (eds.) IVA 2009. LNCS, vol. 5773, pp. 500–501. Springer, Heidelberg (2009)
12. Pellom, B., Hacioglu, K.: Sonic: The university of colorado continuous speech recognizer. University of Colorado, Technical Report# TR-CSLR-2001-01, Boulder, Colorado (2001)
13. Leuski, A., Traum, D.: NPCEditor: A tool for building question-answering characters. In: Language Resources and Evaluation Conference (2010)
14. Leuski, A., Traum, D.: Practical language processing for virtual humans. In: Conference on Innovative Applications of Artifical Intelligence (2010)
15. Thiebaux, M., Marshall, A., Marsella, S., Kallmann, M.: SmartBody: Behavior Realization for Embodied Conversational Agents. In: International Conference on Autonomous Agents and Multi-Agent Systems, Portugal (2008)
16. Vilhjalmsson, H., Cantelmo, N., Cassell, J., Chafai, N.E., Kipp, M., Kopp, S., Mancini, M., Marsella, S., Marshall, A.N., Pelachaud, C., Ruttkay, Z., Thorisson, K.R., van Welbergen, H., van der Werf, R.: The behavior markup language: Recent developments and challenges. In: Pelachaud, C., Martin, J.-C., André, E., Chollet, G., Karpouzis, K., Pelé, D. (eds.) IVA 2007. LNCS (LNAI), vol. 4722, pp. 99–111. Springer, Heidelberg (2007)
17. Ma, W., Hawkins, T., Peers, P., Chabert, C., Weiss, M., Debevec, P.: Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. In: Rendering Techniques 2007 (2007)
18. Alexander, O., Rogers, M., Lambeth, W., Chiang, M., Debevec, P.: Creating a Photoreal Digital Actor: The Digital Emily Project. In: Sixth European Conference on Visual Media Production (CVMP) (2009)

19. Piwek, P.: Presenting arguments as fictive dialogue. In: 8th Workshop on Computational Models of Natural Argument (in conjunction with ECAI 2008), Patras, Greece (2008)
20. Baylor, A., Ebbers, S.: Evidence that multiple agents facilitate greater learning. In: Hoppe, U., Verdejo, M., Kay, J. (eds.) Artificial Intelligence in Education: Shaping the Future of Learning Through Intelligent Technologies, pp. 377–397. IOS Press, Amsterdam (2003)
21. Stolcke, A.: SRILM-an Extensible Language Modeling Toolkit. In: Seventh International Conference on Spoken Language Processing, pp. 901–904 (2002)
22. Wang, D., Narayanan, S.: A confidence-score based unsupervised MAP adaptation for speech recognition. In: 36th Asilomar Conference on Signals, Systems and Computers, Asilomar, CA (2002)
23. Krippendorff, K.: Content analysis: An introduction to its methodology. Sage Publications, Inc., Thousand Oaks (1980)