Ada and Grace: Direct Interaction with Museum Visitors

David Traum[1], Priti Aggarwal[1], Ron Artstein[1], Susan Foutz[3], Jillian Gerten[1], Athanasios Katsamanis[2], Anton Leuski[1], Dan Noren, and William Swartout[1]

[1] USC Institute for Creative Technologies, Los Angeles
[2] USC Signal Analysis and Interpretation Laboratory, Los Angeles
[3] Institute for Learning Innovation, Edgewater, Maryland

Abstract. We report on our efforts to prepare Ada and Grace, virtual guides in the Museum of Science, Boston, to interact directly with museum visitors, including children. We outline the challenges in extending the exhibit to support this usage, mostly relating to the processing of speech from a broad population, especially child speech. We also present the summative evaluation, showing success in all the intended impacts of the exhibit: that children ages 7–14 will increase their awareness of, engagement in, interest in, positive attitude about, and knowledge of computer science and technology.

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

1 Introduction

Ada and Grace [14], a pair of twin virtual humans that were designed to engage visitors and increase their knowledge and appreciation of science and technology, have been in operation at the Museum of Science in Boston since December 2009. The Twins are physically located in a kiosk inside Cahners ComputerPlace, a room dedicated to exhibits about computers and related technologies (such as robots). They have been designed to answer a variety of questions about science and technology, museum exhibits, and themselves. The characters use speech recognition to recognize the words in a question and then use a statistical classifier to select the most likely answer from a pre-existing set of approximately 150 answers. The answer is then presented using coordinated speech, gestures, eye gaze and body movement.

In the first version, Ada and Grace did not directly interact with visitors. Instead, a museum staff member would wear a head-mounted microphone and pose questions to the Twins. Staff members could either ask their own questions, or they could field questions from the audience and relay them to the Twins. In this paper, we report on our efforts to add two additional interaction use cases with the exhibit: a *direct interaction* condition, in which museum visitors could talk to Ada and Grace without staff member intervention, and a *blended* condition in which both staff and visitors can talk to the Twins. Allowing visitors to interact directly with the characters raised a number of issues, which we address in Section 2. In Section 4, we present a summary of some of the results of an independent summative evaluation that was performed by the Institute for Learning Innovations (ILI).

Y. Nakano et al. (Eds.): IVA 2012, LNAI 7502, pp. 245–251, 2012.

2 Improvements to the Twins to Enable Direct Interaction

Allowing visitors to interact directly with the visitors raised a host of issues related to the wide usability and performance of the system, the most important of which are described below.

Hardware. Initially, museum staff used a standard wired USB microphone to talk to the Twins. This was replaced in June 2010 by a wireless Sennheiser microphone, to allow more mobility and cut down on failure points in the constant use of the museum environment. For visitor use, we needed a fixed microphone that could focus on the voice of the speaker. We experimented with three microphones, a Sennheiser ME66, Shure SM58, and Shure 522. The narrow beam pattern of the Sennheiser ME66 gives superior pickup of quiet signals in a noisy environment, but we found that it was very difficult to get visitors to remain at the optimal distance and orientation. The Shure SM58 alleviates this problem, but visitors still had difficulty in associating the microphone with the separate, wireless press-to-talk switch. The current setup (since February 2011) uses the Shure 522, a desktop microphone with integrated press-to-talk button which is typically used for paging and dispatching applications.

Data Collection. In order to handle questions that visitors frequently ask, as well as be able to recognize them well, data was collected in the museum, which was then transcribed and coded for the speaker type (child, adult male, adult female, or no speech) [1]. A sampling of 17,244 utterances from April and May of 2011 revealed the following composition (identified by listening to the voice): 47% children, 13% adult male, 8% adult female, and 31% no speech. The questions were annotated with the correct answer if possible, or were marked either as questions for which answers could be constructed or questions that would be treated as "off-topic" [9]. Additional details about the corpus collected are given in [1].

Speech models. Speech recognition was performed by the SONIC toolkit [11] until December 2010, and thereafter by OtoSense, a recognition engine that is currently being developed by USC. The transcribed audio recordings from the museum visitors have been used for the adaptation of three separate acoustic models, namely for children's, adult male and adult female speech [12]. Adaptation was performed using Maximum Likelihood Linear Regression while the original children's models were trained on the Colorado University children's speech database [5] and the two adult speech models were trained on the Wall Street Journal corpus [10]. In the deployed system, the three recognizers are used in parallel, each providing an n-best list with confidence scores.

Audio acquisition. Incorporating both the multiple microphones and multiple acoustic models into the Twins system posed a significant engineering challenge. The original speech acquisition subsystem (Acquirespeech) was designed to support a single microphone connected to a single ASR engine. We redesigned Acquirespeech to work with multiple simultaneous inputs and multiple concurrent ASR engines. The current system starts five instances of the OtoSense ASR engine, connects the first two (generic male and female) to the staff microphone and the other three (child, adult male, adult female) to the visitor microphone. Both microphones operate in a push-to-talk mode with separate physical control buttons. Acquirespeech handles both the control button clicks and audio inputs from multiple microphones independently. In theory, both the visitor and the staff member can be talking to Twins at the same time and it's the task of the response selection process to select between two inputs.

When Acquirespeech receives audio input from a microphone, it forwards it to all linked ASR engines concurrently. For example, audio from the visitor microphone goes to all three visitor speech engines simultaneously. Acquirespeech waits for the text response from each engine, selects the one with the highest confidence score, and passes it along to the response selection component (NPCEditor [8]).

Content enhancements. The exhibit was also extended and improved in several ways, based on experiences during the initial period. Captions were added for the Twins output to aid the hearing impaired. We also provided optional indicators on the processing stages (listening, thinking, responding), so that visitors could regulate the timing of asking questions. Many animations were also improved.

An optional "idle" dialogue behavior was added, such that if no one talks to the Twins for a threshold time (usually set to 10 minutes), then the Twins would start talking to each other, to cue visitors that they could ask questions.

A number of changes were made to the language understanding and response selection component. One set of changes was to remove directions to exhibits that had been removed from the space, such as the AI Dome and the Computer Build Bench. However the information about the scientific knowledge related to the exhibits (such as cell phones) was retained. Additional answers about new exhibits, such as Coach Mike's arrival in Robot Park [7] were added. The data collection was also used to identify frequent questions that were not understood or did not have a good answer. We added several answers to handle these classes, such as people speaking to the Twins in other languages, insults and hazing, and asking about dinosaurs.

We also provided a card of about 10 example questions, to help visitors get a sense of the types of questions they could ask. In a sample of over 20,000 utterances asked by visitors, 30% were identical to one of the posted suggestions [1].

3 Performance Evaluation

Automatic speech recognition using the three acoustic models has been systematically evaluated only for a small but representative portion of the Twins corpus comprising 1003 utterances recorded on a single day. The average word error rate was found to be 57%, when automatically selecting the model that has the highest confidence score (the initial configuration in the museum, used at the time of the summative evaluation); this resulted in a response selection accuracy of 42%. The best performing individual model is the child model, with a 53% overall word error rate and response accuracy of 45%; however, using an oracle to choose the best performing model for each utterance lowers the word error rate to 43% and raises the response accuracy to 53%, suggesting that better speaker identification should lead to improved performance overall. The relatively high word error rate is linked to the challenges of the museum setting: (1) Speech is spontaneous, i.e., with frequent hesitations and mispronunciations. (2) Speech is coming mainly from children (76% of the sample). (3) There are no vocabulary constraints on what visitors can say.

4 Summative Evaluation

The redesigned exhibit, together with an accompanying exhibit highlighting the science behind the Twins, was subject to a summative evaluation from an external, independent evaluator, the Institute for Learning Innovation (ILI). The study was designed to address two primary questions: 1) What is the nature of visitors' interactions with the Twins and Science Behind exhibits? and 2) In what ways do interactions with the exhibits impact visitors' knowledge and awareness of, engagement and interest in, and attitudes and perceptions towards computer science and technology? The intended impacts are shown in Table 1. Overall, 15 indicators were identified across the four impact areas. The summative evaluation was designed to determine whether the exhibits achieved these indicators, and therefore, the visitor impacts. The evaluation demonstrated that 14 of these indicators were achieved, as shown in Table 1.

Table 1. Intended impacts of the Twins and Science Behind exhibits

Impacts	Indicators	
Children (ages 7–14) and adults will	Tested	Achieved
– increase their **engagement and interest** in computer science and technology.	5	5
– have a **positive attitude** about computer science and technology.	2	2
– increase their **awareness** about computer science and technology.	5	4
– increase their **knowledge** about computer science and technology.	3	3

Two conditions were tested: direct visitor interaction, and blended staff and visitor interaction. Three methods were used in the study: 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 using the exhibit, the number and types of questions that the visitor addressed to the Twins, categorization of the Twins' responses, and visits to the Science Behind exhibit. Interviews were conducted after visitors engaged with either exhibit with the goal of collecting a paired observation and interview with the same participant. Children under 16 years of age were interviewed only after the data collector obtained permission from an adult family member in the visiting group. 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. Only adult participants were asked to complete retrospective-pre/post-experience ratings in order to measure change in attitude and awareness as a result of the experience.

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%). A total of 61 follow-up online questionnaires were collected (for a response rate of 42%).

In this paper, we present a selection of the results from the summative evaluation study showing the combined results for both condition and illustrating each of the 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 [4] for the complete results of the summative evaluation.

Engagement and Interest. Time spent in the exhibit ranged from 19 seconds to just nearly 18 minutes, with a median time of 3 minutes and 7 seconds ($N = 221$). 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” on the 4-point scale. Participants were asked this same question six weeks later in the follow-up online questionnaire; researchers compared ratings from the interview and follow-up questionnaire. Ratings remained the same six weeks following the original visit (Wilcoxon Signed Rank Tests).

Attitudes. The same quantitative rating scale question was also used to determine if participants had positive attitudes towards speaking with the Twins, by asking them to rate the statement “Being able to speak with the Twins” . The overall rating for all participants was a median of 3, or “pretty good” on the 4-point scale. As with the engagement questions above, ratings remained the same six weeks following the original visit (Wilcoxon Signed Rank Tests).

An additional quantitative approach to assess visitors’ attitudes was used only with adults, to determine if interacting with the exhibit impacted self-reported agreement with the 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 a significantly higher rating for all of these measures of attitudes towards computers/virtual humans directly after their interaction with the exhibit.

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 post than retrospective-pre. In answers to open-ended questions, over 90% 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% of participants noted aspects of the connection between the Twins and the main subjects of Cahners (computers, communications, robots) or described the Twins as guides for the space.

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). 97% of all participants mentioned at least one aspect, while 73% mentioned two or more aspects. The most commonly mentioned aspect was natural language, mentioned by 86% of participants. 64% of onsite participants named at least one technology needed to build a virtual human; this rose to 90% in the follow-up six weeks later. Finally, 84% of participants gained at least one additional understanding about STEM (Science, Technology, Engineering, and Mathematics) domains related to the Twins, while 59% of participants indicated they learned something new about computers or technology from interacting with the exhibit.

5 Related Work and Discussion

There are other efforts, both past and present, to put virtual characters in museums as guides. An early system is Max [6], who was placed at the Heinz Nixdorf Museums-Forum in Paderborn, Germany in 2004. Like the Twins, Max is projected life-size on a screen, and communicates using speech and body animations. Max can engage in both reactive and deliberate conversational behavior, and visitors communicate with him by typing on a keyboard. An analysis of interactions between visitors and Max showed that visitors treat Max conversationally as a person, evidenced by conventional strategies of beginning and ending conversations and general cooperativeness [6].

Another virtual museum guide is Tinker [3], who has been situated since April 2008 in the Museum of Science in Boston, just around the corner from the Twins. Tinker builds relations over time, recognizing a visitor who returns for a second conversation; relations bring about gains in visitors' attitudes toward, engagement with, and learning from Tinker [3]. Users communicate with Tinker through menus on a touch screen.

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. Systems similar to the Twins in this regard are Sergeant Blackwell [13], exhibited at the Cooper-Hewitt Museum in New York in 2006–2007 as part of the National Design Triennial exhibition, and Furhat [2], who was shown for four days in 2011 at the Robotville exhibit in the Science Museum in London. 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.

This paper described extensions to the Virtual Human Museum Guides, bringing the system from one that can be demonstrated by an expert to one that can interact directly with visitors (or blending the two, having both interact). This involved hardware, software and content changes. The independent summative evaluation showed that despite some remaining challenges in language understanding performance, the exhibit successfully impacts visitors, as intended, realizing some of the capability for virtual humans to aid in informal science education.

Acknowledgments. This material is based upon work supported by the National Science Foundation under Grant 0813541.

References

1. Aggarwal, P., Artstein, R., Gerten, J., Katsamanis, A., Narayanan, S., Nazarian, A., Traum, D.: The Twins corpus of museum visitor questions. In: LREC 2012, Istanbul, Turkey (May 2012)

2. Al Moubayed, S., Beskow, J., Granström, B., Gustafson, J., Mirnig, N., Skantze, G., Tscheligi, M.: Furhat goes to Robotville: A large-scale multiparty human-robot interaction data collection in a public space. In: Edlund, J., Heylen, D., Paggio, P. (eds.) LREC Workshop on Multimodal Corpora, Istanbul, Turkey, pp. 22–25 (May 2012)
3. Bickmore, T., Pfeifer, L., Schulman, D.: Relational Agents Improve Engagement and Learning in Science Museum Visitors. In: Vilhjálmsson, H.H., Kopp, S., Marsella, S., Thórisson, K.R. (eds.) IVA 2011. LNCS, vol. 6895, pp. 55–67. Springer, Heidelberg (2011)
4. Foutz, S., Ancelet, J., Hershorin, K., Danter, L.: Responsive virtual human museum guides: Summative evaluation. Tech. rep., Institute for Learning Innovation, Edgewater, Maryland (2012)
5. Hagen, A., Pellom, B., Cole, R.: Children's speech recognition with application to interactive books and tutors. In: Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU 2003), pp. 186–191 (2003)
6. Kopp, S., Gesellensetter, L., Krämer, N.C., 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)
7. Lane, H.C., Noren, D., Auerbach, D., Birch, M., Swartout, W.: Intelligent Tutoring Goes to the Museum in the Big City: A Pedagogical Agent for Informal Science Education. In: Biswas, G., Bull, S., Kay, J., Mitrovic, A. (eds.) AIED 2011. LNCS, vol. 6738, pp. 155–162. Springer, Heidelberg (2011)
8. Leuski, A., Traum, D.: NPCEditor: Creating virtual human dialogue using information retrieval techniques. AI Magazine 32(2), 42–56 (2011)
9. Patel, R., Leuski, A., Traum, D.: Dealing with Out of Domain Questions in Virtual Characters. In: Gratch, J., Young, M., Aylett, R.S., Ballin, D., Olivier, P. (eds.) IVA 2006. LNCS (LNAI), vol. 4133, pp. 121–131. Springer, Heidelberg (2006)
10. Paul, D.B., Baker, J.M.: The design for the Wall Street Journal-based CSR corpus. In: Proceedings of the DARPA Speech and Natural Language Workshop, pp. 357–362. Harriman, New York (1992), http://acl.ldc.upenn.edu/H/H92/H92-1073.pdf
11. Pellom, B., Hacıoğlu, K.: SONIC: The University of Colorado continuous speech recognizer. Tech. Rep. TR-CSLR-2001-01, University of Colorado, Boulder (2001/2005), http://www.bltek.com/images/research/virtual-teachers/sonic/pellom-tr-cslr-2001-01.pdf
12. Potamianos, A., Narayanan, S.: Robust recognition of children's speech. IEEE Transactions on Speech and Audio Processing 11(6), 603–616 (2003)
13. Robinson, S., Traum, D., Ittycheriah, M., Henderer, J.: What would you ask a conversational agent? Observations of human-agent dialogues in a museum setting. In: Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC), Marrakech, Morocco (2008)
14. 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., White, K.: Ada and Grace: Toward Realistic and Engaging Virtual Museum Guides. In: Allbeck, J., Badler, N., Bickmore, T., Pelachaud, C., Safonova, A. (eds.) IVA 2010. LNCS, vol. 6356, pp. 286–300. Springer, Heidelberg (2010)