An Evaluation of Virtual Human Technology in Informational Kiosks

Curry Guinn
UNC-Wilmington
601 South College Rd
Wilmington, NC 28403-
(910)-962-7182
guinnc@uncw.edu

Rob Hubal
RTI International
3040 Cornwallis Road
RTP, NC, USA 27709
(919)-541-6045
rhubal@rti.org

ABSTRACT
In this paper, we look at the results of using spoken language interactive virtual characters in information kiosks. Users interact with synthetic spokespeople using spoken natural language dialogue. The virtual characters respond with spoken language, body and facial gesture, and graphical images on the screen. We present findings from studies of three different information kiosk applications. As we developed successive kiosks, we applied lessons learned from previous kiosks to improve system performance. For each setting, we briefly describe the application, the participants, and the results, with specific focus on how we increased user participation and improved informational throughput. We tie the results together in a lessons learned section.

Categories and Subject Descriptors
H.1.2 [**Information Systems**]: User/Machine Systems – *human factors, software psychology*

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *evaluation/methodology, graphic user interface, natural language, voice i/o.*

H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems– *evaluation/methodology, artificial, augmented and virtual realities.*

General Terms
Experimentation, Human Factors.

Keywords
Virtual humans, natural language, spoken dialogue system, gesture, evaluation, virtual reality.

ICMI'04, October 13–15, 2004, State College, Pennsylvania, USA.

1. INTRODUCTION
An "accessible" user interface is one that is easy to learn and easy to use, and can result in measurable goals such as decreased learning time and greater user satisfaction (i.e., acceptance) [28]. Characteristics of easy to learn and easy to use interfaces have been described as having navigational and visual consistency, clear communication between the user and application, appropriate representations, few and non-catastrophic errors, task support and feedback, and user control [15,20,21,28].

We have been particularly interested in how accessible responsive virtual human technology applications are. Usability testing, commonly conducted for commercial software to ensure that it meets the needs of the end user, is likewise vital to creating effective training and assessment software employing innovative technologies. This paper presents findings from a series of information kiosks investigating how users accept and evaluate virtual human applications.

1.1 Background on Responsive Virtual Human Technology
In the past decade, researchers have created on a series of PC-based applications in which the user interacts with responsive virtual characters. Virtual humans have been used in training, assessment and marketing ([2,4,7,16,17,19,22,24,25]). Applications have ranged from trauma patient assessment [[13] to learning tank maintenance diagnostic skills [[9] to gaining skills in avoiding non-response during field interviews [3]. In these applications, the PC simulates a person's behavior in response to user input. Users interact with the virtual characters via voice, mouse, menu, and/or keyboard.

1.2 Our Responsive Virtual Human Architecture
We have developed a PC-based architecture that enables users to engage in unscripted conversations with virtual humans and see and hear their realistic responses [10]. As seen in Figure 1, among the components that underlie the architecture are a Language Processor and a Behavior Engine. The Language Processor accepts spoken input and maps this input to an underlying semantic representation, and then functions in reverse, mapping semantic representations to gestural and speech output. Our applications variously use spoken natural language interaction [[9], text-based interaction, and menu-based interaction. The Behavior Engine maps Language Processor output and other environmental stimuli to virtual human behaviors. These behaviors include decision-making and problem

solving, performing actions in the virtual world, and spoken dialog. The Behavior Engine also controls the dynamic loading of contexts and knowledge for use by the Language Processor. The virtual characters are rendered via a Visualization Engine that performs gesture, movement, and speech actions, through morphing of vertices of a 3D model and playing of key-framed animation files (largely based on motion capture data). Physical interaction with the virtual character (e.g., using medical instruments) is realized via object-based and instrument-specific selection maps [29]. These interactions are controlled by both the Behavior Engine and Visualization Engine.

We keep track of domain knowledge via state variable settings and also by taking advantage of the planning structure inherent in our architecture [11]. Our virtual humans reason about social roles and conventions (what can be stated or asked at any point in the dialog) [23] and grammar definitions (how it gets stated or asked). The architecture was designed to allow application creators flexibility in assigning general and domain-specific knowledge. Hence, our virtual humans discuss relevant concerns or excuses based on specific setup variables indicating knowledge level and initial emotional state. Our personality models and emotion reasoning are based on well-accepted theories that guide realistic emotional behavior [1,4,23,24,26]. After user input, we update emotional state based on lexical, syntactic, and semantic analyses [11]. For the applications presented here, we only used emotional state for presenting facial expression.

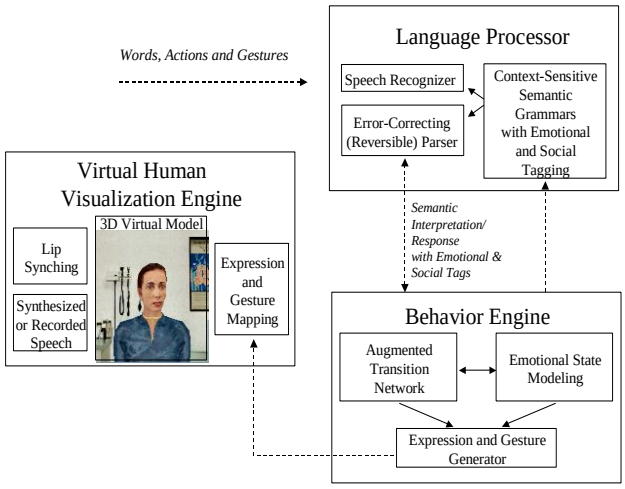

Figure 1 System Architecture

1.3 Overview of Paper

We present findings from studies of three different information kiosk applications. For each we briefly describe the application, the participants, and the results, concentrating on results that get at accessibility, engagement, and usability. We tie the results together in a lessons learned section.

2. Virtual Tradeshow Attendant Applications

2.1 Methods

We created a product that was positioned as a virtual tradeshow attendant. It was put into operation as a kiosk, drawing attention to the booth, augmenting the sales and marketing staff, and providing engaging dialog with visitors regarding the company and company products. We report on user data collected at three particular venues, the Exhibitor Show held in February 1999, the Space Congress held in April 1999, and the American Society for Training and Development (ASTD) International Conference & Exposition held in May 1999. Kiosks were configured as illustrated in Figure 2 with a computer screen embedded in a kiosk with a microphone. A placard was present to invite participants to talk with the avatar (i.e., virtual character). The virtual character would also occasionally "bark" to try to attract participants to a conversation: *Hey, now, you look like an intelligent sort. Why don't you stop and talk to me?* We had both male and female characters (see Figures 3 and 4). At the Exhibitor Show two kiosks were used, one with the male; the other with the female character. At other two shows, a single kiosk was used with primarily the female character. No attempt was made to determine which character participants preferred.

Figure 2 Tradeshow Kiosk

Figure 3 Virtual Tradeshow Attendant

2.2 Results

The virtual tradeshow attendant did attract visitors to the booths and answered visitors' questions, a definite advantage on the competitive tradeshow floor. For each show, we looked at the number of users, the time and length of conversation, and the content of the conversations.

2.2.1 Time and Length of Conversations

A summary of the time and length data may be found in Table 1. The first application of our virtual attendants was at the Exhibitor Show. The Exhibitor Show consists primarily of vendors who market to people who exhibit at tradeshows – it is a show for marketers. Users were not prompted in what to say by the virtual character. Over the 2 ½ days of the show, we had relatively few visitors who spoke with our virtual character. Forty-five users spoke with the system for an average of 21.4 seconds and spoke 2.8 utterances. As will be discussed more in the next section, 25% of the users utterances were understood as requests for help.

Table 1 Summary of User Interactions

Show	# Users	Avg. # of Turns	Avg. Time (sec.)	% Yes/No Answers
Exhibitor	45	2.8	21.4	N/A
ASTD	197	3.3	28.4	51
Space Congress	335	5.6	61.4	66

To improve the system, we realized we needed to provide more structure for the user. Therefore, for our next venue, ASTD[1], we added prompts of two kinds: screen prompts and avatar prompts. Screen prompts (illustrated in Figure 4) presented a changing list of topics that the participant may want to ask the system. After each user utterance, this list would change, offering new suggestions. A typical prompt might be

ASK ME ABOUT:

Spaceport Florida.

Employment opportunities at NASA.

My name.

Virtual reality and training.

[1] Although the ASTD occurred after Space Congress, the delivery cycle for these two shows was such that ASTD was developed a full month before Space Congress.

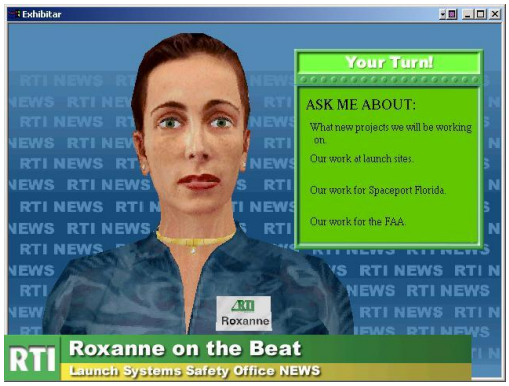

Figure 4 Attendant with Screen Prompts

In addition, the virtual character would follow each answer to a user question with a yes/no question prompting the user. A typical question-answer-prompt might look as follows:

USER:

Where is your company located?

AVATAR:

Our main campus is located in Research Triangle Park, North Carolina. We also have offices throughout the United States and several foreign countries. Would you like to hear about how we use virtual reality in training?

We also used the screen area where the prompts were displayed to present other information. For instance, when the virtual character was speaking about a particular subject, an image or map might be displayed in that region as illustrated in Figure 5.

Figure 5 Displaying Images When Speaking

At ASTD, over three days, 197 visitors spoke to the virtual characters for an average of 28.4 seconds and 3.3 conversational turns. In this venue, 51% of the user utterances were interpreted as Yes/No replies to the prompted questions.

One problem with the prompting system that we used at ASTD is that the topic prompts were randomly chosen from the remaining topics that had not been discussed. Often, this would mean the next prompt would be on a topic completely unrelated to the previous content resulting in a disjointed conversation. For Space Congress, we grouped prompts by topic and chose prompts

based on the previous context. In the following example, the follow-up question flows naturally in the conversation.

USER:

> Where is your company located?

AVATAR:

> Our main campus is located in Research Triangle Park, North Carolina. We also have offices throughout the United States and several foreign countries. Would you like to hear about our Launch Safety System Office in Florida?

At the Space Congress show, in front of a reasonably technical audience over four days, the application attracted 335 visitors, who conversed with the virtual characters an average of 61.4 seconds with 5.6 conversational turns.

2.2.2 Variance in User Questions

We analyzed not only the number of visitors and the structure of their conversations, but also the content of the conversations. For each venue, every single one of the possible conversation topics was covered at least once (Exhibitor, 26 topics; ASTD, 63 topics; Space Congress, 39 topics). The distribution and frequency of discussion changed somewhat across venues. Because of the lack of prompting within the Exhibitor application, users had a much higher incidence of asking for help or talking about the character itself ("you are pretty") rather than the information that was trying to be presented at the tradeshow. Table 2 presents the ten most common topics/questions posed by users at the Exhibitor Show.

Table 2 Top 10 User Questions in Exhibitor

Semantics	Example	% Among Users
Help_me	"How do I use this?"	28
company	"What does your company do?"	24
What_is_it	"What is this thing?"	15
Goodbye	"See you later"	13
What_can_I_say	"What can I talk about?"	8
How_much	"How much does this cost?"	8
My_face	"Can I get a different face?"	6
Avatar_name	"What is your name?"	6
Current_time	"What time is it?"	6
Customer(interested)	"This is cool."	4

For ASTD, we see a much different distribution of topics covered because of the use of Yes/No answers. Sometimes a user may ask, "What does your company do?" or they may answer, "Yes, I would" when asked "Would you like to hear about what we do?" Both of these user inputs would map to the semantics "what_do_you_do". Table 3 and Table 4 present the top user questions in ASTD and Space Congress respectively.

There are several broad semantic groups that we could classify the users' comments and requests. We have chosen to look at the distribution of five groups: 1) Greetings/goodbyes, 2) Help requests, 3) Comments about the virtual character, 4) Chatty comments (the weather, the time of day, the meaning of life), and 5) Informational (what do you do for NASA, what is virtual reality). As can be seen in Table 5, each successive application was able to provide more informational content while reducing the amount of user requests for help. A side effect of having a more avatar-directed conversation seems to be that the user also makes fewer comments about the avatar itself. Nonetheless, still roughly 10% of user utterances were about the character ("you are pretty/ugly/stupid", "will you marry me?", "how do you work?"). Also the percentage of cocktail party talk ("what time is it?", "how do you like Atlanta?") still ranged between 10%-20%.

Table 3 Top 10 User Questions in ASTD

Semantics	Example	% Among Users
hello	"Good afternoon."	24
company	"What does your company do?"	18
Current_time	"What time is it?"	14
My_face	"Can I get a different face?"	13
What_are_you	"What is this thing?"	9
Rti_sylvan	What's the relationship between RTI and Sylvan?"	9
Current_day	"What is today's date?"	8
clients	"Who are some of your clients?"	7
hardware	"Does this run on special hardware?"	7
Attendant(wrong)	"You misunderstood me."	7

Table 4 Top 10 User Questions in Space Congress

Semantics	Example	% Among Users
What_do_you_do	"What does your company do?"	50
greeting	"Hello"	33
Current_time	"What time is it?"	19
How_work	"How does this thing work?"	19

lsso	"Tell me about the Launch Systems Safety Office."	17
Nuclear_propulsion	"What are you doing in the field of nuclear propulsion?"	14
What_should_I_say	"What kinds of things can I say?"	13
Launch_sites	"What support do you provide for launch site?"	13
nasa	"Do you work for NASA"?	13
Launch_vehicles	"What launch vehicles have you worked with?"	12

Table 5 Distribution of Question/Comment Categories

	Greet	Help	Character	Chatty	Information
Exhibitor	11%	25%	28%	6%	30%
ASTD	11%	8%	13%	18%	50%
Space Congress	11%	6%	8%	10%	64%

3. Conclusions and Lessons Learned

Visitor data from virtual human marketing applications are not conclusive of usability or acceptability, but suggestive. Less technical users were sufficiently engaged to converse with the virtual characters for just under half a minute, and more technical users for just over a minute. Given prompting, the users covered the range of topics designed into the applications. It is important that these users had never before seen the applications, had no training or practice time, had to learn to use the applications at that moment, and yet stuck with the conversation for a significant period of time.

Some specific lessons learned include

- It is critical in applications to be able to detect and respond appropriately to "bad" or inappropriate input. In all our applications, users (often but not always intentionally) spoke utterances that were outside the range of what was expected in the context of the dialog. This occurred most frequently in the tradeshow exhibit application where users would try to test the limits of the system.
- Without explicit prompting by the virtual character, users often seemed lost as to what to say next. We found that explicit statements or questions by the virtual character helped to supply the user with the necessary context. This also helped to prune the language processing space.
- Our greatest difficulties in understanding the system occurred when the user replied with very complex compound sentences, multiple sentence, and even paragraph long utterances. This phenomenon led us to set user expectations through the prompting mechanism.
- Ultimately, because of the limitations in language understanding, the user would adapt to environment, adjusting the manner in which they spoke.

We are encouraged by results so far, but feel it is important to continue to investigate more robust and effective virtual human models and more efficient means of creating the models, to better understand user preferences and acceptance of virtual humans. We propose several areas of active research:

- Usability and acceptability studies across different populations. Are there differences in acceptance of virtual characters across boundaries of age, gender, education level, and cultural divides?
- Usability and acceptability studies with varied degrees of visual realism. How realistic do virtual characters have to be in order to receive high ratings of acceptability by users? What is the contrast in user impressions between video of actual humans versus more cartoon-like animated characters?
- Usability and acceptability studies with multimodal input. Currently our systems make no attempt to use the user's vocal affect, facial expressions, eye movement, body gesture, or other physiological input (such as heart rate) in interpreting the user's emotional state and intentions. We would like to introduce these elements into our systems to assess whether such input can create more realistic characters.

4. REFERENCES

[1] André, E., Klesen, M., Gebhard, P., Allen, S., & Rist, T. (2000). Exploiting Models of Personality and Emotions to Control the Behavior of Animated Interface Agents. Proceedings of the International Conference on Autonomous Agents (pp. 3-7). Barcelona, Spain.

[2] André, E., Rist, T., & Müller, J. (1999). Employing AI Methods to Control the Behavior of Animated Interface Agents. International Journal of Applied Artificial Intelligence, 13 (4-5), 415-448.

[3] Camburn, D.P., Gunther-Mohr, C., & Lessler, J.T. (1999). Developing New Models of Interviewer Training. Proceedings of the International Conference on Survey Nonresponse. Portland, OR.

[4] Dahlbäck, N., Jönsson, A., & Ahrenberg, L. (1993). Wizard of Oz Studies – Why and How. Knowledge-based Systems, 6(4), 258-266.

[5] Frank, G.A., Guinn, C.I., Hubal, R.C., Stanford, M.A., Pope, P., & Lamm-Weisel, D. (2002). JUST-TALK: An Application of Responsive Virtual Human Technology. Proceedings of the Interservice/Industry Training, Simulation and Education Conference. Orlando, FL.

[6] Frank, G.A., Helms, R., & Voor, D. (2000). Determining the Right Mix of Live, Virtual, and Constructive Training,

Proceedings of the Interservice/Industry Training Systems and Education Conference. Orlando, FL.

[7] Graesser, A., Wiemer-Hastings, K., Wiemer-Hastings, P., Kreuz, R., & the Tutoring Research Group (2000). AutoTutor: A simulation of a human tutor. Journal of Cognitive Systems Research, 1, 35-51.

[8] Groves, R., & Couper, M. (1998). Nonresponse in Household Interview Surveys. New York, NY: John Wiley & Sons, Inc.

[9] Guinn, C.I., & Montoya, R.J. (1998). Natural Language Processing in Virtual Reality. Modern Simulation & Training, 6, 44-45.

[10] Hubal, R.C., Deterding, R.R., Frank, G.A., Schwetzke, H.F., & Kizakevich, P.N. (2003). Lessons Learned in Modeling Pediatric Patients. In J.D. Westwood, H.M. Hoffman, G.T. Mogel, R. Phillips, R.A. Robb, & D. Stredney (Eds.) NextMed: Health Horizon (pp. 127-130). Amsterdam, Holland: IOS Press.

[11] Hubal, R.C., Frank, G.A., & Guinn, C.I. (2003). Lessons Learned in Modeling Schizophrenic and Depressed Responsive Virtual Humans for Training. Proceedings of the Intelligent User Interface Conference. Miami, FL.

[12] Hubal, R.C., & Helms, R.F. (1998). Advanced Learning Environments. Modern Simulation & Training, 5, 40-45.

[13] Kizakevich, P.N., McCartney, M.L., Nissman, D.B., Starko, K., & Smith, N.T. (1998). Virtual Medical Trainer: Patient Assessment and Trauma Care Simulator. In J.D. Westwood, H.M. Hoffman, D. Stredney, & S.J. Weghorst (Eds.), Art, Science, Technology: Healthcare (R)evolution (pp. 309-315). Amsterdam, Holland: IOS Press.

[14] Klein, G. (1998). Sources of Power. Cambridge, MA: MIT Press.

[15] Kolb, D.A. (1984). Experiential Learning. Englewood Cliffs, NJ: Prentice Hall.

[16] Lester, J., Converse, S., Kahler, S., Barlow, S., Stone, B., & Bhogal, R. (1997). The Persona Effect: Affective Impact of Animated Pedagogical Agents. Proceedings of the Human Factors in Computing Systems Conference, (pp. 359-366). New York, NY: ACM Press.

[17] Lindheim, R., & Swartout, W. (2001). Forging a New Simulation Technology at the ICT. Computer, 34 (1), 72-79.

[18] Link, M., Armsby, P.P., Hubal, R., & Guinn, C.I. (2002). A Test of Responsive Virtual Human Technology as an Interviewer Skills Training Tool. Proceedings of the American Statistical Association, Survey Methodology Section. Alexandria, VA: American Statistical Association.

[19] Lundeberg, M., & Beskow, J. (1999). Developing a 3D-Agent for the August Dialogue System. Proceedings of the Auditory-Visual Speech Processing Conference. Santa Cruz, CA.

[20] Nielsen, J. (1993). Usability Engineering. Boston: Academic Press.

[21] Norman, D.A. (1993). Things That Make Us Smart. Reading, MA: Addison-Wesley.

[22] Olsen, D.E. (2001). The Simulation of a Human for Interpersonal Skill Training. Proceedings of the Office of National Drug Control Policy International Technology Symposium. San Diego, CA.

[23] Ortony, A., Clore, G.L., & Collins, A. (1988). The Cognitive Structure of Emotions. Cambridge, England: Cambridge University Press.

[24] Rickel, J., & Johnson, W.L. (1999). Animated Agents for Procedural Training in Virtual Reality: Perception, Cognition, and Motor Control. Applied Artificial Intelligence, 13, 343-382.

[25] Rousseau, D., & Hayes-Roth, B. (1997). Improvisational Synthetic Actors with Flexible Personalities. KSL Report #97-10, Stanford University..

[26] Russell, J.A. (1997). How Shall an Emotion Be Called? In R. Plutchik & H.R. Conte (Eds.), Circumplex Models of Personality and Emotions (pp. 205-220). Washington, DC: American Psychological Association.

[27] Sugarman, J., McCrory, D.C., Powell, D., Krasny, A., Adams, B., Ball, E., & Cassell, C. (1999). Empirical Research on Informed Consent: An Annotated Bibliography. Hastings Center Report. January-February 1999; Supplement: S1-S42.

[28] Weiss, E. (1993). Making Computers People-Literate. San Francisco: Jossey-Bass.

[29] Zimmer, J., Kizakevich, P., Heneghan, J., Schwetzke, H., & Duncan, S. (2003). The Technology Behind Full Body 3D Patients. Poster presented at the Medicine Meets Virtual Reality Conference. Newport Beach, CA. Available at http://www.rvht.info/publications.cfm.