# Usability and Acceptability Studies of Conversational Virtual Human Technology

**Curry Guinn[1], Robert Hubal[1], Geoffrey Frank[1], Henry Schwetzke[1], James Zimmer[1], Sarah Backus[1], Robin Deterding[2], Michael Link[1], Polly Armsby[1], Rachel Caspar[1], Laura Flicker[1], Wendy Visscher[1], Amanda Meehan[1], Harvey Zelon[1]**

**[1] RTI International, Research Triangle Park, NC**
**[2] University of Colorado Health Sciences Center, Denver, CO**

## Abstract

Acceptance, accessibility, and usability data from a series of studies of a series of applications suggest that most users readily accept responsive virtual characters as valid conversational partners. By responsive virtual characters we mean full-body animated, conversant, realistic characters with whom the user interacts via natural language and who exhibit emotional, social, gestural, and cognitive intelligence. We have developed applications for medical clinicians interviewing pediatric patients, field interviewers learning about in-formed consent procedures, and telephone interviewers seeking to obtain cooperation from respondents on federally-funded surveys. Usage data from informational kiosks using the same underlying technology (e.g., at conference exhibits) provide additional corroboration. Our evidence suggests the technology is both sufficient to actively engage users and appropriate for consideration of use in training, assessment, and marketing environments.

## 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 de-scribed 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].

As part of our Technology Assisted Learning (TAL) initiative, we have been particularly interested in how accessible responsive virtual human technology (RVHT) 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 studies investigating how users accept and evaluate RVHT applications.

### 1.1 Background on RVHT and TAL

Since approximately 1996, we have worked on a series of PC-based applications in which the user interacts with responsive virtual characters. 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, which we collectively categorize as involving RVHT, 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. We are certainly not alone in developing training, assessment, marketing, and other RVHT applications (see, e.g., [2,4,7,16,17,19,22,24,25]), but the breadth across do-mains and combination of technologies is unusual.

The RVHT applications are representative of those developed in our TAL division. We define TAL as "proactively applying the benefits of technology to help people train more safely, learn better, retain skills longer, and achieve proficiency less expensively". We develop TAL applications for jobs requiring complicated knowledge and skills, complex or expensive equipment or work material, a high cost of on-the-job training or failure on the job, jobs where safety or spatial awareness is essential, and for large student throughput requirements [6,12].

Practicing skills in a safe and supportive environment allows the student to learn flexible approaches. Flexibility is critical for interaction skills [8] and for performing well under time constraint, information-poor, and other difficult conditions [4,14]. The consistency that is gained by repeating this practice in virtual environments leads directly to good decisions on the job [24]. By practicing skills in safe, computer-generated settings, students have the opportunity through repetition to develop practical experience and skills which would otherwise be difficult to acquire. Practice also leads to increased confidence prior to the first real on-the-job experience.

### 1.2 RVHT Architecture

We have developed a PC-based architecture, Avatalk, that enables users to engage in unscripted conversations with virtual humans and see and hear their realistic responses [10]. 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 under-lying 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 inter-action, 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 al-low 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 mod-els 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].

### 1.3 Overview of Paper

We present findings from studies of four different applications. The applications are, in order of presentation, a virtual pediatric standardized patient, a trainer for practicing informed consent procedures, a telephone surveys interview trainer, and two implementations of a tradeshow booth marketing product. 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 Pediatric Patient

Training and assessment in pediatrics is complicated by the poor reliability of children to behave in a consistent manner. Consequently, curriculum is difficult to develop, performance assessment is restricted, and practice opportunities are limited. Our goals using RVHT have been to develop specific interactive training sessions using virtual pediatric characters and to explore related educational issues [10].

Figure 1. Virtual Pediatric Patient

One educational issue in pediatric medicine is instruction. Medical students rotating through pediatrics have limited exposure to children and are given limited one-on-one faculty observation time, hence the curricular material is mostly passive, while on-the-job learning involves variable experiences with behaviors or problems and dispersed learners. Another educational need in pediatric medicine is associated with assessment, since there is no reliable or valid authentic assessment in young children (it is currently text-based or multimedia videos) as is possible with standardized patients for

adults, and since interaction skills with children may not be valued by the student.

Our use of virtual pediatric patients follows models of experiential learning, where abstract conceptualization leads to active engagement and experimentation, which leads to concrete experience, which leads to reflective observation, which leads back to the beginning of the cycle [15,21]. By adding virtual characters, we are adding experiential learning to the traditional classroom, discussions, and rounds.

Our work supports training and assessment not only of verbal interaction skills, but also of medical diagnostic skills, dealing with the spectrum of behavioral responses, and other types of high-level problem solving.

### 2.1 Methods

Three specific interactive pediatric scenarios have been developed to date in our virtual pediatric standardized patient (VPSP) application. In one scenario, the clinician is tasked with obtaining an ear exam in a very young girl. The girl may be helpful if she is healthy but whiny if she has an ear infection. In another scenario, the clinician is asked to examine the lungs of a pre-teen boy. In the last scenario, the clinician must obtain a high-risk behavior history from a teenage girl.

Educational issues that we are addressing include defining – and identifying – pediatric interactive strategies, program validity, scoring performance, and providing feedback. Our goal is to provide information for a "gold-standard" setting language acquisition to improve the robust nature of the interaction, and to address face, content, and construct validity. We hypothesize that expert and novice users will provide valuable development information about language and strategies in these scenarios, and the differences will exist based upon expertise with children and technology experience.

### 2.2 Results

Interactive pediatric scenarios were created and shown to content and educational experts. The first rounds of feedback from experts, on the girl and boy scenarios, came at exhibits sessions at the Association of American Medical Colleges annual meeting in November 2002 and the Medicine Meets Virtual Reality conference in January 2003. From comments at these sessions we revised the scenarios and added the adolescent scenario. The latest round of feedback, and the results described here, came from the Council on Medical Student Education in Pediatrics (COMSEP) annual meeting in April 2003.

Fourteen attendees at the COMSEP meeting were recruited to use the VPSP application. The attendees were first given a questionnaire asking about their experience with completing ear exams, lung exams, and adolescent social history, and also about their computing experience. They were then given brief instructions on how to use the application, told to choose whichever of the scenarios they wanted, and handed headphones to avoid distraction. Finally, they were given a questionnaire asking their perceptions of the realism of the application in comparison to clinical experience.

In a way, this was the toughest group of all we've tested. These participants were true experts, unaware of the technology (until a debriefing at the end of each session), and presented with an application prototype. Given this rationalization, the data are acceptable. On average, these participants rated the realism of the response time and the realism of the objections, concerns, and questions posed by the virtual characters as "somewhat" realistic. They rated the realism of the overall conversation as a little better than "not very" realistic. However, somewhat surprisingly, when asked to compare the simulated clinical experience with real clinical experience, the participants rated the comparison as somewhat challenging, that is, the comparison is reasonable. Four of the participants even found the simulated experience "moderately" or "extremely" challenging. Analysis of the participants' log files shows they spent an average of almost 4 1/2 minutes in the scenarios, taking eight conversational turns, and collectively covering 32 topics (of a possible 130 topics across all scenarios, and with no prompting). The participants were observed to take the cases seriously, asking strategic questions to get the virtual character to cooperate, and becoming frustrated when their questions were misinterpreted. (We are pleased by frustration, as it implies engagement, though anxious, too, to make the application work better.) We take these data as encouraging, but fully understand the need to revise in depth the language and behavior of the virtual characters to satisfy acceptance criteria.

## 3 Practice on Informed Consent

### 3.1 Methods

Under grant funds to enhance our IRB program, we created a virtual reality simulation for teaching proper informed consent procedures. In the application, a potential survey respondent poses questions regarding the survey, the sponsor, confidentiality, privacy, compensation, and contact information [27].

In November 2003, we presented the trainer to a group of five experienced telephone or field interviewers who were being trained for a study intended to better understand the health effects of exposure to smoke, dust, and debris from the collapse of the World Trade Center. We observed the participants and also had them, after completing the interactions, fill out a short questionnaire on their familiarity with computers and their impressions of the application.

### 3.2 Results

The application forced the respondents to touch on all aspects of informed consent before finishing. The only way an interaction could be cut short was if the participant replied incorrectly to a question (e.g., giving the wrong sponsor name, or indicating that participation was mandatory rather than voluntary). Participants interacted no fewer than three times with the virtual character and up to six times.

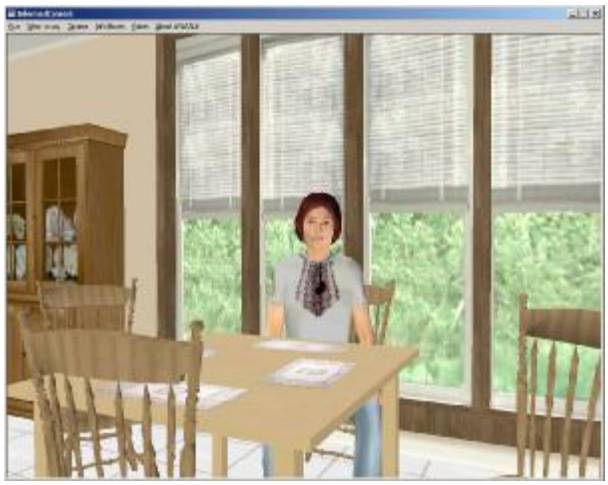

Figure 2. Informed Consent Training

The results were generally positive, particularly in the subjects' assessment of usability and enjoyment. The realism of the character was consistently rated by the participants as moderately realistic (average of 5.2 on a 7 point scale), a decent rating given the virtual character's relatively few body movements and facial gestures. Ease of use (5.8), enjoyment (5.6), and effectiveness (5.4) were all rated moderately to very easy, enjoyable, or effective. An observer also rated the level of engagement of the interaction. Engagement, verbalization, and information seeking were all moderately or highly demonstrated. Participants were judged either relaxed or amused by the interaction, they responded in a moderate amount of time, and they appeared to comprehend what was being asked. As would be expected, they were also judged to find the interaction not at all provocative and needed very little negotiation.

## 4 Telephone Survey Interviewer

One of the most difficult skills for a telephone interviewer to learn is gaining cooperation from sample members and avoiding refusals. In telephone interviewing in particular, the first half-minute on the telephone with a sample member is crucial. Sample members almost automatically turn to common phrases to avoid taking part in surveys: "How long will this take?" "How was I selected?" "I don't do surveys." "I don't have time." "I'm just not interested." "What is the survey about?" Non-response research suggests that the best approach to obtaining participa-tion is for the interviewer to immediately reply with an appropriate, informative, tailored response [2,9].

We tested an RVHT application designed to simulate the first 30-60 seconds of a telephone interview [21]. Interviewers begin with an introduction and then need to respond to a series of objections or questions raised by the virtual respondent. Ultimately, the virtual character ends the conversation by either granting the interview or hanging-up the telephone. The emotional state of the virtual respondent varies from scenario to scenario. A total of six basic objections were recorded in four different tones of voice for both a male and female virtual respondent.

Figure 3. Telephone Survey User Interface

### 4.1 Methods

The assessment provided here of the interviewer training module is based on researcher/instructor observations, and user debriefings in the form of a questionnaire. Empirical data were collected on users' observed ability to interact with the application as well as their perception of the interaction. The training application was tested with a group of 48 novice telephone interviewers during Spring 2002.

To evaluate the accessibility of the application we focused on the users' understanding of the application's basic features, their ability to complete each task, and capabilities shown by different users (e.g., based on ethnicity, job level, and education level. To evaluate acceptance of the application by the trainees, we debriefed participants using a structured questionnaire and moderator-facilitated focus groups to gauge reactions and engagement in the application. We were interested in the virtual humans' realism, speed and accuracy of the speech recognition, and detection of changes in the emotive states of the virtual human.

Finally, each training session was observed by either the researchers or training instructors, who made notes

of their observations. These observations are included as part of the analysis.

### 4.2 Results

**Ease of Use:** Users of the RVHT application found it very accessible, with 84 percent indicating the software was either extremely easy or very easy to use (52% extremely, 31% very, 13% somewhat, 4% not too, 0% not at all). Only eight (17%) of the 48 trainees indicated that they required additional assistance to use the training software (after the initial training received by all trainees).

**Realism of the Training Environment:** The promise of RVHT-based training tools is that they can simulate a real environment, thereby allowing trainees repetitive practice in conditions that are as close as possible to what they will encounter on the job. For this particular application, the virtual respondent needed to mirror the behaviors and emotions of real respondents encountered when doing live interviewing. This means delivering an array of objections to the trainees in different tones of speech and emotional levels in a fast-paced manner. Interviewers were asked a series of questions to try to assess how well they accepted the virtual environment as a substitute for real work conditions.

The answer is somewhat mixed. In general, trainees did not find the virtual environment to be realistic and they cited two primary reasons: the slowness of the response of the virtual respondent and the limited number of different objections/questions offered by the virtual respondent. They did, however, find the responses that were offered to be realistic and stated that they could detect and respond to changes in tone and emotional cues offered by the virtual respondents. A majority of the trainees also indicated that they felt the sessions helped them to improve their skills needed at the outset of an interview either "somewhat" or "a lot".

When asked how realistic they found the overall conversation with the virtual respondent, 17 percent of participants said they thought it was "extremely" or "very" realistic, and 44 percent said it was "somewhat" realistic. Slowness of the virtual respondents in replying (due to the lag caused by the speech recognizer as it interpreted the interviewer's responses and determined the next script to launch) was the primary problem cited by interviewers. Over three-quarters (77%) of the users felt the response time was too slow (4% felt it was too fast and 19% indicated the speed was just right).

The trainees were, however, more positive when evaluating the realism of the objections and questions offered by the virtual respondent. A plurality (48%) indicated that the content of what was said was either "extremely" or "very" realistic, with 40 percent saying it was "somewhat" realistic. They also felt it was relatively easy to determine the emotional state of the virtual respondent based on the tone of voice they heard (23% "extremely" easy, 44% "very" easy, 29% "somewhat" easy, and 4% "not too" easy). Likewise, the content of the speech used by the virtual character was also a good cue to trainees as to the virtual human's emotional state (8% "extremely" easy to tell, 54% "very" easy, 27% "somewhat" easy, 10% "not too" easy).

Being able to recognize changes in the emotional state of the virtual respondent changed how the interviewer approached the situation. Nearly 60 percent indicated that they behaved differently in the practice scenario based on the tone of the virtual respondent's voice. Thus, the content of the objections raised by the virtual respondent and the emotional behavior of the virtual human were generally accepted by the trainees and caused them to react differently within the various training scenarios. It appears, however, that while the interviewers do recognize and react to emotional cues, they do not necessarily process these as being very distinct. They focus more on the actual content of the argument (regardless of the tone of voice or gender) when considering how diverse the scenarios offered are.

**Enjoyment and Reuse:** An effective training tool is also one that trainees should enjoy using, would use again, and recommend to others. Approximately two-thirds (65%) of the users said that they found using the RVHT software to be fun and enjoyable. Nearly three-quarters (73%) said they would like to use the software again. In addition, 83 percent said they would recommend the program as a training tool for other interviewers. In open-ended responses, a number of interviewers indicated that it would be a very good practice vehicle for new or less experienced interviewers.

## 5 ExhibitAR Applications

### 5.1 Methods

Using earlier versions of the same underlying technology we created a product called ExhibitAR 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.

### 5.2 Results

The ExhibitAR product did attract visitors to the booths and answered visitors' questions, a definite advantage on the competitive tradeshow floor. 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 five conversational turns. At ASTD, with less technical attendees, over three days, 197 visitors spoke with the virtual characters for an average of 28.4 seconds and 2.6 conversational turns.

Figure 4. Virtual Tradeshow Attendant

We analyzed not only the number of visitors and their conversations, but also the content of the conversations. For ASTD, every single one of the 63 topics of conversation was covered at least once. The average per topic was almost nine occurrences (i.e., nine different visitors asked about the topic). For Space Congress, again every topic was covered, the average number of times per topic for the 39 topics was 35 occurrences. The most common topics for both applications were a greeting, asking about the company and its associates, asking what to do or say, asking how the technology worked, and asking the current date or time, but topics specific to each application were also discussed.

The Exhibitor data are less telling, but this was the show at which ExhibitAR was introduced, and this was the only venue where the visitor was not at all prompted. The application attracted 45 visitors over 2 1/2 days, each visitor averaging 2.5 turns and 21.4 seconds. Though each of the 25 topics was covered at least once, the only topic that was covered considerably more often than any other was a request for assistance. (This led us to devise a prompting feature.)

Visitor data from RVHT marketing applications are not conclusive of usability or acceptability, but suggestive. Even at the time these data were collected (five years ago), 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, yet stuck with the conversation for a significant period of time.

It is only anecdotal data, but RVHT continues to attract visitors to exhibit booths. The various applications described in earlier sections, and others, have been shown since 1999 at least a dozen times to audiences varying from educators to medical practitioners to public health workers to military service personnel. Visitors are increasingly less surprised (skeptical?) to encounter virtual characters, and more impressed with the state of the art. They appear willing to accept virtual characters as sensible for training, assessment, and marketing uses.

## 6 Conclusions and Lessons Learned

In this paper we describe usability and acceptance data derived from a number of studies using a number of different RVHT applications. No data suggest that our applications are completely accessible yet to these users, but the data in aggregate suggest we are moving in the right direction.

The different studies involved various user groups, from experts (medical clinicians) to novices (field and telephone survey interviewers) to “common folk” (exhibit visitors) in greatly different domains. A common finding was for our participants to suggest additional potential audiences, also ranging from novice to expert. Further, the majority of participants said they enjoyed using the applications – and/or were observed to be engaged with the virtual characters – despite technical obstacles, prototype-stage content, and conspicuous presence of the investigators.

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. But we even found that in the training applications that users would often express frustration by cursing or otherwise verbally mistreating the virtual character.

- 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. . In the ExhibitAR domain, a subset of possible relevant questions was always present on the screen.

- Because of shortcomings in speech recognition technology, we found that typed input was often needed to overcome the limitations of large grammars. This was particularly true in the more open-ended pediatric trainer. We also found typed input to be invaluable in development stage even in applications that were ultimately going to be speech-driven. The typed inputs in development helped us to derive grammars that

we could later use to improve the speech recognizer.

- 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 in the training environment prior to their using the system.
- Anecdotally we found that pre-recorded speech was much more acceptable than any currently available speech synthesizer. This effect seemed to be less noticeable the longer the user spoke with the system. We would like to conduct a study comparing the use of the two technologies.
- 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 RVHT models and more efficient means of creating the models, to better understand user preferences and acceptance of RVHT, and to determine how best to use RVHT in combination with other approaches to produce cost-effective training, assessment, and other applications. 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 input modes. What are the tradeoffs between using a typed natural language interface versus a spoken interface? We found that a typed interface improved the computer's ability to comprehend the user which leads to more cohesive dialog. On the other hand, a typed interface reduces the naturalness of the dialog, the believability of the character, and the usability of the system.
- 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 galvanic skin response) 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.

## Acknowledgements

The studies described here were performed under awards # 290-00-0021 from the Agency for Healthcare Research and Quality, # 1-S07-RR18257-01 from the National Institutes of Health, and # R9898-002 from the Research Triangle Institute. Preparation of this paper was supported under award # EIA-0121211 from the National Science Foundation. Points of view in this document are those of the authors, and do not necessarily represent the official position of any of the above-listed agencies.

## References

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.

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.

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.

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

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.

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.

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.

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

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

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.

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.

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

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.

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

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

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.

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

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.

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.

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

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

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.

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

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.

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

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.

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.

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

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.