SURVEY AND STATISTICAL COMPUTING IV. THE IMPACT OF TECHNOLOGY ON THE SURVEY PROCESS
Edited by R. Banks *et al*

Accessibility and Acceptance of a Virtual Respondent-Based Interviewer Training Application[1]

Polly P. Armsby, Michael W. Link, Robert Hubal, Curry I. Guinn, Laura Flicker, and Rachel A. Caspar

Abstract

Training applications which use responsive virtual-human technology (RVHT) – training tools based on sophisticated voice recognition and behavior modeling technologies – have great potential for improving training of interaction skills essential to effective interviewing, such as refusal avoidance, probing, and addressing questions related to informed consent. However, our understanding of how to model the behavior of responsive virtual humans and how people interact with them is limited. The overall effectiveness of this technology as a training tool depends upon its ability to provide appropriate learning experiences, its ability to engage the student, and its acceptability to disparate users. This research assesses the accessibility and acceptance of a training application based on RVHT as a tool for teaching refusal avoidance skills to telephone interviewers. The assessment focuses on users' ability to understand the basic features of the application, whether diverse users are able to use the application equally, how users react to problems, whether the virtual humans are realistic enough for the users, and ultimately, whether users accept the virtual environment as a valid proxy for the real work environment.

Keywords

Interviewer training, computer-based training, responsive virtual-human technology-based training

1. Introduction

Survey research is in an era of great challenge. Response rates across all modes of data collection have been in decline, threatening the validity and utility of the information collected in surveys. As it becomes more difficult to convince sample members to participate in surveys, it is essential that the interviewers who are on the front lines of collecting these data are given the tools they need to be successful in their jobs. While interviewer training encompasses a variety of topics (CATI/CAPI skills, gaining cooperation, respondents' rights, questionnaire administration, etc.), training tools built using responsive virtual human technology (RVHT) hold the promise of offering interviewers a simulated, realistic environment for developing and practicing basic interviewing skills – such as gaining respondent cooperation, probing, administering informed consent – and honing those skills

[1] This work was supported by a research grant from the National Science Foundation (Grant No. EIA-0121211).

over time. RVHT reduces the amount of learning that must occur on the job, by allowing repetitive practice in a virtual environment.

RVHT is admittedly in its developmental infancy and requires additional improvements before it can be deployed as a fully mature technology in a production environment. The research presented here is one small part of a larger research program shepherding the growth and development of these technologies. RVHT involves the use of natural language processing and an emotive behavioral engine to produce natural, interactive dialogues with intelligent, emotive virtual-reality (VR) agents. RVHT has great potential for use in training interaction skills, such as those required for effective survey interviewing. However, our understanding of how people interact with responsive virtual humans (a.k.a. intelligent agents) is quite limited.

Better understanding requires employing RVHT in training applications and conducting systematic use, usability, perception, and training-effectiveness assessments. Important questions yet to be answered include: determining whether intelligent agents make learning more accessible; determining whether students are willing to accept intelligent agents as interactive partners in learning; determining what skills can be acquired, practiced, and validated using RVHT; determining what is involved in providing a convincing simulation of human interaction, realistic enough for the student to suspend disbelief and acquire skills that will transfer to a "live" environment.

Users' interactions with RVHT applications are little studied and poorly understood. The research presented here (and the larger research program from which it is drawn) provides an initial assessment of some of the issues associated with user interface design, user acceptance of computer-based training, and perceptions of the realism and effectiveness of the training tool. The assessment of these issues was conducted in two phases. The first phase was conducted within a controlled environment using a dozen subject matter experts as test subjects and collected repeated assessments of the perceived performance and realism of the virtual training environment across several different hardware platforms. The second phase of the assessment was conducted within a live production environment, using approximately 50 telephone interviewers of varying backgrounds to test the application. A longer, structured questionnaire was then used to capture their evaluations and perceptions of the training tool. Both assessments involved the use of an RVHT-based training tool for refusal avoidance at the outset of a telephone interview. The findings seem to indicate that RVHT has great promise as a tool for training survey interviewers; however, the underlying technologies need further development before such applications are robust enough to be fully production-ready.

2. Mechanics of the RVHT Survey Interviewer Training Application

Successful interviewers must employ a variety of skill sets including standardized interviewing practices, proficiency using a computer and relevant software, a thorough knowledge of the current survey instrument, and interpersonal, interaction and active listening skills. Research has shown that flexibility is critical for developing effective interaction skills (Groves & Couper, 1998) and for performing well under time constrained, information-poor, and other difficult conditions (Klein, 1998). In order to acquire flexible and effective approaches to gaining respondent cooperation, new and experienced interviewers require a learning environment that realistically simulates the environment they face in an interviewing situation. The consistency that is gained by repetitive practice in virtual and constructive learning environments leads directly to effective decisions in the production environment (Ross, Pierce, Haltermann, & Ross, 1998). Practice also leads to increased confidence before the first real on-the-job experience, minimizing the amount of on-the-job learning

that is necessary. In the survey world, on-the-job-learning can translate into numerous unsuccessful interview attempts at the start of a study by a new interviewer, leading to lower response rates, lower quality data, delayed schedules, and increased costs.

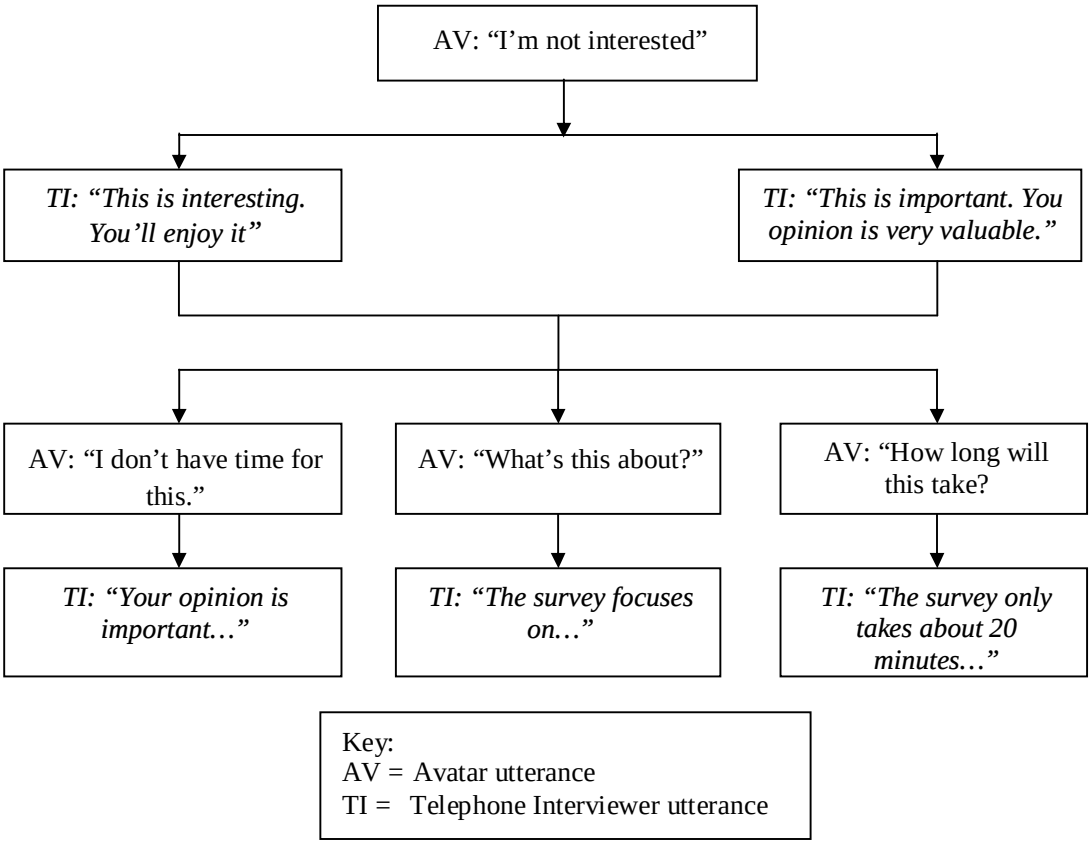
This is exactly the type of scenario in which RVHT can be most effective. The outset of any interview is generally very fluid, despite the fact that interviewers are nearly always provided with an introductory script or set of bullet points for making the introduction. Sample members often interrupt interviewers with a barrage of questions or remarks, such as "I'm not interested," and "I don't have time." Non-response research suggests that the best approach to obtaining participation is for the interviewer to immediately reply with an appropriate, informative, tailored response (Camburn, Gunther-Mohr, & Lessler, 1999; Groves & Couper, 1998; Groves, 2002). Generally, such skills are taught through a combination of lecture, paired-mock practice with other interviewers, and by using multimedia to listen to real or mock audiotapes of exchanges between interviewers and sample members. RVHT allows us to take skill building to the next level, by providing a realistic, simulated environment in which an interviewer can practice and hone his or her skills.

The application tested here involves the use of an RVHT-based application to simulate the environment a telephone interviewer faces during the first thirty to sixty seconds of a telephone survey interaction. The training tool allows interviewers to practice their skills in gaining cooperation in a self-paced, realistic environment. The software is designed such that interviewers begin with an introduction and are then required to respond to a series of objections and questions raised by the "virtual respondent." The interviewer's responses are captured electronically and processed by a natural language speech processor. Based on the content of the interviewer's speech, the software launches another objection/question or ends the conversation by either granting the interview or hanging-up the telephone (see Figure 1).

The application uses spoken natural language interaction (Guinn & Montoya, 1998), not text-based interaction (except for data collection during development from subject-matter experts). The speech recognizer uses a basic dictionary of common words as well as a specific dictionary for each turn of a conversation. The specific dictionary consists of up to 200 words based on behavioral observations of real world events. These specific dictionaries are dynamic, therefore, changing with each turn of the conversation.

The application tested here is not designed to present training content to the user. The basic skills needed for gaining cooperation are presented and initially acquired during an instructor-led classroom session. However, one of the primary potential benefits of the application is the enhanced ability to practice these critical skills using repetitive, structured, and standardized scenarios and to conduct practice sessions outside of the traditional classroom setting. While RVHT applications can be used to direct content to the trainee, this research focuses primarily on the practice component of the training curriculum. Finally, the current system is not yet equipped to provide intelligent tutoring. That is, coaching and feedback (aside from an indication of success and failure) are provided by direct observation from supervisors and trainers during or after use of the application, rather than from the application itself.

Figure 1. Example of Dialogue Flow



3. Assessment 1: Controlled Test Environment

Overview

The primary purpose of our first assessment was to capture users' evaluations of the overall performance and perceptions of realism of the application across several different computer platforms. Assessing gains in the persuasion skills of the test subjects was not within the scope of this (or the second) assessment; rather the focus is on perceived performance and perceptions of realism of the application itself. The experiment consisted of testing four hardware platforms with twelve volunteer subjects who were blinded to the machines being tested. Each test consisted of three separate conversations with the Avatar on the selected hardware platform. After completing those three conversations, the subject was asked to rate the experience with respect to realism of the simulation using a three-question assessment form.

Analysis Measures

Analysis variables were derived from two sources: (1) coded responses from the transcripts of interactions between the Avatar and the subject and (2) evaluations made by the subjects themselves.

Transcript-derived Measures

The taped conversations were first transcribed (with the transcriptions being verified by the test administrator). Then each conversation was coded to indicate unique conversation exchanges and the semantic meaning or focus of each exchange. In all, there were a total of 910 unique exchanges that were coded from the 264 conversations (which represent 88 different trials across the four machines). From the coded transcripts, three measures were developed to measure the behavior of the RVHT application:

- **Conversation Exchange:** measures the number of Avatar-subject conversational interactions. An "exchange" is defined as the pairing of an Avatar "objection" and a subject "response."
- **Conversation Semantic:** measures the content or meaning of the exchange between the Avatar and the subject. Initially all exchanges were coded into one of 35 possible "semantic" categories. These 35 categories were then collapsed into six general conversation semantics: Introduction, Survey Content, Time Concerns, Selection Criteria, Survey Attributes, and Setting
- **Conversation Complexity:** measures the number of unique semantics observed during the course of a conversation. A conversation with a larger number of unique semantics is considered to be a more "complex" conversation than one with fewer unique semantics.

Subject-derived Measures

Three additional measures were developed from observations made by the subjects themselves. For each machine trial, subjects completed three separate conversations with the Avatar. After each set of conversations the subjects were asked to rate the realism of the trial in terms of responsiveness, overall conversation, and the objections raised. Each of these dimensions was rated on a seven-point scale, where 1 = not at all realistic and 7 = extremely realistic.

- **Realism of Response Times:** Did the application respond quickly enough to mirror the way in which sample members actually respond over the telephone?
- **Realism of the Overall Conversations:** Did the dialogue that took place during the three conversations generally reflect the types of dialogues (in terms of flow and content, pace and tone) that take place with sample members at the outset of a telephone interview?
- **Realism of the Objections Raised:** Were the objections raised by the Avatar realistic and reflective of those encountered in exchanges with reluctant sample members during actual interviews?

Findings

The analysis was conducted in two parts: (1) subject evaluations of the application's performance across four platforms and (2) evaluation of the relationship between subjects' evaluations of realism and the behavior of the application in terms of exchanges, semantics, and complexity of the conversations.

Subject Rating of Application Realism by Platform

First, we examined how the subjects themselves rated their practice experiences across the four platforms. The unit of analysis in this section is the "trial" level (i.e., a trial equals three conversations conducted on a single machine). In all there were 88 trials conducted across the four machines. In terms of evaluating the realism of the response times across the platforms, the differences (while not statistically significant at the traditionally excepted level of $p < .05$) are suggestive of a significant difference (given the relatively small sample size of 88 and a $p < .096$ value). In terms of response time, the Dell laptop rated the highest (5.06 average rating), followed by the IBM ThinkPad (4.87), the

Gateway PC (4.38), and the IBM PC (3.96). The laptops, therefore, ranked higher than the PCs in terms of subjects' ratings of their response time.

Subjects demonstrated little difference in their evaluations of the other two dimensions of realism – assessment of the overall conversation and the objections raised. There were no significant differences noted across each measure in terms of platform used or testing trial. Outside of possible differences in perceptions of response time, therefore, subjects found little difference in their ratings of the realism of the practice conversation generally and of the specific content of those conversations. Likewise, there appeared to be little "educating" of the subjects between trials 1 and 2.

Subject Rating of Realism Based on Application Behavior

Our second area of interest focuses on how subjects' ratings of realism may have been affected by the behavior of the application itself. We might expect that if the application is perceived to behave in a more "realistic" way that we should see differences in subjects' ratings of response time, conversation flow, and content. We found, however, surprisingly few differences in ratings on these dimensions across the different measures of Avatar behavior (conversation exchanges, semantics, and complexity). Again, the results presented here are at the "trial" level.

There were no significant differences seen across these three dimensions based on the average number of exchanges per conversation within a trial. Trials with an average of 1.0 to 2.9 exchanges per conversation were not rated significantly higher or lower in terms of response time than those with 4.0 to 5.0 exchanges. The same is true when we look at ratings of the overall conversation and the objections raised during the exchanges.

Likewise, there was little variation across the three realism dimensions when we consider the six general conversation semantics. The only statistically significant difference was noted in terms of evaluation of the realism of the objections raised. When "setting a callback" was a topic of a trial, that trial tended to be rated higher in terms of the realism of the objections made, than did trials where setting a callback was not a focus.

Finally, and somewhat surprisingly, the complexity of the conversations across a trial was not related significantly to ratings of response time, nor of the realism of the overall conversation and objections raised. One might have expected that interactions that are more complex would have led to higher ratings on either or both the overall conversation or objections raised dimensions. This, however, was not the case.

Summary of Controlled Environment Assessment

In terms of user perceptions, there were few notable differences discerned. Subjects did not vary significantly in their evaluations of the realism of the response time, overall conversation, or objections raised across different platforms and trials for the experiment, nor across differences in the types of exchanges they encountered (shorter/longer, more/less semantically complex). In part, this may be due to the low number of observations resulting from this analysis being conducted at the trial-level (the trial level was used for analysis since that is the level at which the perception evaluations of realism were made). Further analyses will be conducted using more sophisticated statistical modeling (nested data analyses) at the conversation and exchange levels to determine if significant differences in perceptions are revealed at those levels.

4. Assessment 2: Operational Test within Live Production Environment

Overview

A primary goal of the overall research program of which this study is a part is to determine if RVHT can be an effective technology for interaction training across a broad spectrum of ethnic and socio-economic backgrounds, jobs, and job levels. The effectiveness of this technology depends upon its ability to provide appropriate learning experiences, its ability to engage the trainee, and its acceptability to disparate users.

In the second phase of our assessment, we examine this aspect by collecting data from a group of approximately 50 telephone interviewers of varying ages, races, experience and education levels, who used the refusal avoidance training module within a production environment. To evaluate the accessibility of the application we focused on the following:

- Do users understand the basic features of the application?
- Are users able to complete each task and exit the application?
- Are different users (e.g., based on ethnicity, job level, and education level) equally able to use the application?

Analysis of these questions will provide clues as to how smoothly the application runs, or when and why difficulties arise in its use.

The question of whether and why participants "accept" or "reject" the virtual training environment is also central to this research. 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. In particular we are interested in the following:

- Are the virtual humans realistic enough for the users? Why or why not?
- How fast and accurate is the speech recognition?
- Could trainees detect changes in the emotive states of the virtual human using only audio cues?
- Would they use the application again and/or recommend its use by others?

As part of this second phase of the evaluation process, data were collected using a questionnaire filled out by the interviewers and notes made by instructors and researchers who observed the training sessions. The questionnaire asked questions related to users' perceptions of the realism of the interactions with the "virtual human," ease of use of the software, the perceived effectiveness of the training sessions, and some basic background characteristics of the users. In all, a diverse group of 48 interviewers filled-out the questionnaires (96% of the software users).

Findings

The questions posed to the interviewers were designed to assess their perceptions and experiences in using the RVHT training tool in four basic areas: ease of use of the software, realism of the training environment, impact on skill development, and desire to recommend or use the software again. Although this is the first detailed look at how users interact emotive intelligent agents for soft-skills development, we can formulate some hypotheses regarding how different types of users might respond based on how users generally differ in their use and acceptance of other computer-based tools. For example, we might expect to find that trainees who are younger, have more education, and are more comfortable using computers in general to have fewer difficulties in using the system. Likewise, we might expect that more experienced interviewers might not find the training tool as useful as inexperienced interviewers because the more experienced interviewers will have already developed and honed their refusal avoidance skills (a supposition that mirrors the finding of Groves, 2002). To

examine possible differences in accessibility and acceptance of the program, we cross-tabulated all of the closed-ended questions in the questionnaire with demographic characteristics including sex, education level, age, race, and work experience. Significant differences are noted below.[1]

Ease of Use of the Application

Users of the RVHT software seemed to find it very accessible to use, with 84% 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). Nearly everyone found the written instructions (96%) and the verbal instructions (98%) that accompanied the training to be clear and accurate. 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. In other words, do they "buy-into" the virtual environment?

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, In general, how realistic did you find the overall conversation with the 'virtual respondent,' 17% said they thought it was extremely or very realistic, 44% said it was somewhat realistic, 17% not too realistic and 23% not at all 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. Perhaps not surprisingly, trainees who describe themselves as "fast-touch typists"were more likely than those who indicated they were "slow-touch typists" to say the response time was too slow (82% fast-touch vs. 67% slow-touch; $p < .08$ chi-sq.).

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% saying it was somewhat realistic, 8% not too realistic, and 4% not at all realistic. They also felt it was relatively easy to determine the emotional state of the

[1] Because of the small number of observations (N=48) we also created dichotomous variables for both the dependent variables (collapsing scales where possible) and independent variables (collapsing or combining variables with 3 or more values). These variables were also examined to determine if significant differences among subgroups could be identified.

virtual respondent based on the tone of voice they heard (23% extremely easy, 44% very easy, 29% somewhat easy, and 4% not too easy; no one indicated that they could not determine the avatar's emotional state from the tone of the "virtual human's" voice). Likewise, the content of the speech used by the avatar 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, 0% not at all easy.

Nearly 60% indicated that they behaved differently in the practice scenario based on the tone of the virtual respondent's voice. Interestingly, a higher percentage of women than men reported reacting differently to the changing tone of the avatar's voice (women 67% v. men 33%, $p < .04$ chi-sq.). Similarly, 54% said they treated the situation differently based on the actual words used by the avatar in expressing a concern or voicing an objection. There were, however, no differences between men and women on this question. It seems, therefore, that the both 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.

Impact on Skill Development

The purpose for allowing trainees to operate within a virtual environment is to allow them to develop and hone essential skills before entering the "real" environment, thereby reducing the amount of "on the job" skill development required. New interviewers can do considerable damage at the outset of a telephone study, generating a large number of refusals as they gain comfort and confidence on the telephone. If practice within a virtual environment at the beginning of a project can reduce the numbers of initial refusals even modestly, then the training program will have value. While longer-term assessments of the effectiveness of the RVHT software will need to include examination of more objective measures of improved performance, this preliminary assessment focused on the user's assessment of the impact of the training on their own skill development.

Trainees were asked to evaluate if they thought the RVHT software increased their abilities in six different areas. Nearly three-quarters of the trainees felt that the practice sessions increased a lot or somewhat their ability to respond to questions and concerns by sample members. Approximately 56% felt it helped them a lot or somewhat in better gaining respondent cooperation at the outset of an interview. Likewise, over half felt it helped in their ability to adapt to differences in respondents' tone or voice or perceived moods and to adapt to differences in the speed and pace of different sample members' speech. About half of the trainees also thought that the sessions helped them a lot or somewhat in avoiding refusals at the outset of an interview.

Would They Use the RVHT Training Tool Again?

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. Interestingly men were significantly more likely than women to say that they found the sessions to be enjoyable (92% men vs. 56% women, $p < .05$ chi-sq.). Nearly three-quarters (73%) said they would like to use the software again. In addition, 83% 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.

Summary of Live Environment Assessment

This initial assessment of an RVHT-based training tool for telephone interviewers provides some valuable insights into how trainees access and accept virtual environments as practice labs and "virtual

humans" as training partners. There were aspects of the training program that interviewers clearly liked, such as the ability to do repeated practice of frequently asked questions, being able to distinguish different emotional states from the tone of voice and speech content of the virtual respondent, and the opportunity to learn to think on their feet in a simulated environment before being placed into a live interviewing situation.

There were also aspects that the interviewers did not like, such as the slowness of the response of the virtual respondent and the perceived lack of variety in the scenarios that were presented. This provides constructive feedback for the engineering and improvement of the software. While adding additional scenarios is a relatively easy process, involving research into the "normal" flow of such scenarios , the responsiveness issue is a more fundamental matter, reflecting the current state-of-the-art in speech recognition. For virtual training partners to be more readily accepted, the underlying speech recognition technology needs to be improved, providing, faster processing of the input from interviewers and launching of responses by the virtual respondent.

5. Conclusions

A considerable amount of basic research is still required to make RVHT applications robust, viable training tools within production environments. RVHT can hold one of the keys, however, for improved training of interviewers – both telephone and field-based staff. The research provided here offers additional information allowing developers and application designers a greater understanding of how RVHT applications respond under repeated test conditions and will hopefully help speed the development of these much needed training tools.

We do not anticipate RVHT-based training will replace instructor-led training, but we expect that combinations of RVHT-based training and instructor-led training will continue to offer advantages for presenting training exercises that are more uniform and realistic than those that can be reproduced in the classroom alone. Additionally, RVHT-based training can provide easily implemented, focused sustainment (i.e., refresher) training.

Future research will continue to examine if the presence of the RVHT applications allows trainers to increase the amount of time spent on skill acquisition, by reinvesting time spent on individual practice into classroom sessions focused on existing or additional content. Now that initial assessments are complete, efforts can be made to improve the realism of practice sessions by compiling a more robust corpus of Avatar objections, thus allowing the user to engage in a richer conversation with the application. In an effort to address technical issues that may detract from the realism or the reliability of the training application, further testing and analyses will be conducted in an effort to determine the source of variability in application behavior across hardware platforms. Furthermore, we hope to make improvements to the engineering of the application itself as the underlying speech recognition technology improves. Last, in future experiments involving more rigorous testing with a larger group of users and including controlled experiments comparing users' and non-users' performance, we hope to discern whether an RVHT application contributes to learning beyond what is offered through classroom, multimedia, and computer-assisted instruction.

We 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 training methods to provide cost-effective training on critical interaction skills.

References

Bates, J. (1994). The Role of Emotion in Believable Agents. Communications of the ACM, Special Issue on Agents, July, 1994.

Becheiraz, P., & Thalmann, D. (1998). A Behavioral Animation System for Autonomous Actors Personified by Emotions, Proceedings of the First Workshop on Embodied Conversational Characters (WECC '98), Lake Tahoe, California.

Camburn, D.P., Gunther-Mohr, C., & Lessler, J.T., (1999). Developing New Models of Interviewer Training. International Conference on Survey Nonresponse, Portland, OR, October 28-31, 1999.

Cassell, J., & Vilhjálmsson, H.H. (1999). Fully Embodied Conversational Avatars: Making Communicative Behaviors Autonomous. Autonomous Agents and Multi-Agent Systems: 2, 45-64.

Cole, R., et. al. (1999). New tools for interactive speech and language training: Using animated conversational agents in the classrooms of profoundly deaf children. In Proceedings of ESCA-MATISSE ESCA/SOCRATES Workshop on Method and Tool Innovations for Speech Science Education, London, UK.

Elliott, C. (1993). Using the Affective Reasoner to Support Social Simulations. In Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence, pp. 194-200, Chambery, France, August 1993. Morgan Kaufmann.

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.

Gratch, J. (2000). Modeling the Interplay Between Emotion and Decision-Making. Proceedings of the Ninth Conference on Computer Generated Forces & Behavioral Representation, May 16-18, 2000, Orlando, FL.

Groves, R. (2002). Principles and Practices in Non-response Reduction. Presentation at the 2002 Respondent Cooperation Workshop sponsored by the Council for Marketing and Opinion Research, New York, NY.

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

Guinn, C.I., & Montoya, R.J. (1998). Natural Language Processing in Virtual Reality, Modern Simulation and Training, pp. 44-55.

Helms, R.F., Hubal, R.C., & Triplett, S.E. (1997). Evaluation of the Conduct of Individual Maintenance Training in Live, Virtual, and Constructive (LVC) Training Environments and their Effectiveness in a Single Program of Instruction. Final Report, September 30, 1997. Submitted to Battelle RTP Office, Subcontract # TCN 97031, Delivery Order #0027, Dated April 16, 1997.

Hubal, R.C., Kizakevich, P.N., Guinn, C.I., Merino, K.D., & West, S.L. (2000). The Virtual Standardized Patient-Simulated Patient-Practitioner Dialogue for Patient Interview Training. In J.D. Westwood, H.M. Hoffman, G.T. Mogel, R.A. Robb, & D. Stredney (Eds.), Envisioning Healing: Interactive Technology and the Patient-Practitioner Dialogue, 133-138. IOS Press: Amsterdam.

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

Lester, J., Converse, S., Kahler, S., Barlow, S., Stone, B., & Bhogal, R. (1997). The Persona Effect: Affective Impact of Animated Pedagogical Agents. In S. Pemberton (Ed.), Human Factors in Computing Systems: CHI'97 Conference Proceedings. 359-366. New York: ACM Press.

Loyall, A.B., & Bates, J. (1997). Personality Rich Believable Agents That Use Language. Proceedings of the First International Conference on Autonomous Agents, February 1997, Marina del Rey, CA.

Lundeberg, M., & Beskow, J. (1999). Developing a 3D-Agent for the August Dialogue System. Proceedings from AVSP '99, Santa Cruz, CA.

Maggart, L.E., & Hubal, R.C. (1998). A Situation Awareness Model. In S.E. Graham & M.D. Matthews (Eds.), Infantry Situation Awareness: Papers from the 1998 Infantry Situation Awareness Workshop. U.S. Army Research Institute: Alexandria, VA.

Massaro, D.W., Cohen, M.M., Beskow, J., Daniel, S., & Cole, R.A. (1998). Developing and Evaluating Conversational Agents. In Proceedings of Workshop on Embodied Conversation Characters (WECC), Lake Tahoe.

Psotka, J. (1995). Immersive Training Systems: Virtual Reality and Education and Training. Instructional Science 23, 405-31.

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

Shlechter, T.M., Bessemer, D.W., & Kolosh, K.P. (1992). Computer-Based Simulation Systems and Role-playing: An Effective Combination for Fostering Conditional Knowledge. Journal of Computer Based Instruction 19, 110-114.

Shneiderman, B. (1992). Designing the User Interface: Strategies for Effective Human-Computer Interaction. Reading, MA: Addison-Wesley.

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

About the Authors

Polly P. Armsby is a Survey Director at RTI International. She coordinates tasks such as design, implementation, and data collection and plans data collection schedules, supervises data collection activities, manages data quality control, trains data collection staff, and conducts methodological research. She can be contacted at RTI International, 3040 Cornwallis Rd., P.O. Box 12194, Research Triangle Park, NC 27709-2194, USA; tel. (919) 248-4514, fax (919) 316-3863; e-mail: armsby@rti.org.

Michael W. Link, Ph.D. is a Senior Survey Director and Methodologist at RTI International. He directs and manages survey-related projects and conducts methodological research in the areas of survey operations, interviewer training, and uses of new technology in data collection efforts.

Robert Hubal, Ph.D. is a Senior Research Engineer at RTI International. He is responsible for exploring technology assisted learning methodologies and identifying approaches to improving learning and training effectiveness with an emphasis on cost-effective, technology-enhanced innovations in the design, development, delivery, evaluation, and management of training and education.

Curry I. Guinn, Ph.D. is a Research Computer Engineer at RTI International. He leads the development of interactive virtual humans with emotions that affect their body and facial gestures, decision-making, and language generation.

Laura Flicker is a Survey Methodologist at RTI International. She provides expertise in survey design and implementation and has helped develop many original survey instruments on a wide range of topics including healthcare, drug abuse, welfare, education, housing, and transportation.

Rachel A. Caspar is a Senior Survey Methodologist with RTI International. She specializes in designing questionnaires and developing data collection procedures for surveys of sensitive topics, has conducted methodological studies to improve data quality, and has conducted studies of both unit and item survey nonresponse.