Interactive Training Applications using Responsive Virtual Human Technology

Robert C. Hubal & Geoffrey A. Frank
Research Triangle Institute
Research Triangle Park, North Carolina

ABSTRACT

Intelligent agents are being used in fields as diverse as computer generated forces, manufacturing, medicine, and theater. Where intelligent agents have not been employed, though, is in interaction skills training. But interaction skills--interviewing, negotiating, tactical communications, eliciting information--are critical to today's soldiers, police, and many professionals (e.g., doctors, lawyers) in our service-oriented economy. Sample scenarios include interviewing refugees, handling some forms of asymmetric threats (such as hostage negotiations), and encounters in high-stress military situations (such as negotiating a passage of arms through a checkpoint held by coalition forces). When it occurs, interaction skills training usually relies on peer-to-peer role playing or passive learning through videos. These forms of training lead to a critical training gap, because the students are limited in the practice time and the variety of scenarios that they encounter. But it is exactly this practice, studies show, that leads to significant on-the-job benefits.

We have developed responsive virtual human technology (RVHT) that allows natural, interactive dialog between the soldier and system. RVHT can improve training by reducing the necessary infrastructure (e.g., personnel), by providing soldiers with more practice time and consistent interaction experiences. RVHT is a relatively recent advance in training technology. Portraying emotions in a virtual human requires clearly defined emotional state, action that shows thought processes, and accentuation to reveal feelings, yet lifelike virtual humans can lead to improvements in problem-solving ability and can engage and motivate students. Most importantly, RVHT can open entirely new capabilities for computer-based training of interpersonal skills, and can provide the benefits of reduced training costs, increased student-teacher ratios, individualized tutoring, and greater student convenience that are associated with computer-based training.

Biographical Sketches

Robert C. Hubal holds an M.S. in Computer Science from North Carolina State University and a Ph.D. in Cognitive Psychology from Duke University. He conducts research at RTI on Technology Assisted Learning, focusing on development, presentation, and evaluation of materials and identifying approaches to improve learning and training effectiveness. Dr. Hubal headed up an experimental evaluation for the Army National Guard of the effectiveness of a virtual maintenance trainer, and is leading an evaluation of a constructive driving skills assessment application. He has investigated the interaction effects among cognitive task demands, expertise, and presentation of information. Most recently Dr. Hubal has developed behavioral software that enables virtual humans to act and behave realistically in controlled learning contexts.

Geoffrey A. Frank is a Principal Scientist at RTI. He has a Ph.D. from the University of North Carolina at Chapel Hill. He is Project Engineer for the JUST-TALK project supported by the Office of Science and Technology at the National Institute of Justice. He is co-author of a chapter on virtual reality training in the American Society of Training Developer's Handbook, and has written a military handbook for the Army Material Command on electronic systems design. Dr. Frank was the Project Engineer for advanced learning environments created for the Command and General Staff College, Ft. Leavenworth, the University of Mounted Warfare, Ft. Knox, and the Depth and Simultaneous Attack Battle Lab, Ft. Sill, OK.
Interactive Training Applications using Responsive Virtual Human Technology

Robert C. Hubal & Geoffrey A. Frank
Research Triangle Institute
Research Triangle Park, North Carolina
[rhubal, gaf]@rti.org

INTRODUCTION

As computing becomes more mobile and pervasive, the use of intelligent assistant agents is becoming practical. Intelligent agents are being used in fields as diverse as computer generated military forces [16], manufacturing [30], medicine [29], and theater [28,31]. Some agents are used for information management and data mining, some for autonomous control of simulated entities (e.g., aircraft in a wargame, or characters in an accident scene [27]), and some for electronic commerce. Some agents act as background demons, whereas others present to the user a human face or caricature. Some follow pre-programmed routines, others employ natural language processing (NLP), still others run via a behavior or simulation engine.

Where intelligent agents have not been employed, though, is in interaction skills training. But interaction skills--interviewing, negotiating, presenting, eliciting information--are critical in practically all fields, and advanced technologies for training these "soft skills" offer tremendous returns.

Responsive virtual human technology (RVHT) is a relatively recent advance in training technology. RVHT uses an intelligent agent framework to combine several information technologies, including virtual reality (VR), NLP, and an emotion engine. Some few researchers have begun integrating emotion models with agents [2,7,12,23], but none for interaction training. Portraying emotions in a virtual human requires clearly defined emotional state, action that shows thought processes, and accentuation to reveal feelings [1]. Computer-based interaction training requires "embodied interface agents", that is, virtual humans who use gaze, gesture, intonation, and body posture as well as verbal feedback during the interaction, since reading body language is part of the training [4,5]. Lifelike "pedagogical agents" can engage and motivate students [20,26]. Most importantly, RVHT opens entirely new capabilities for computer-based training of interpersonal skills, and can provide the benefits of reduced training costs, increased student-teacher ratios, individualized tutoring, and greater student convenience that are associated with computer-based training [8].

THE NEED FOR RVHT

Interaction skills are increasing in importance as the service economy grows, as people seek human contact in an increasingly technological world, and as inevitable conflicts arise. The greater mobility and increased interdependence, and expanded communications capabilities of modern society mean that people are frequently dealing with strangers and need interaction skills to achieve their goals. As we shift from agrarian and manufacturing economies to a service economy, interaction skills are required by more and more workers and are critical to achieving success in the workplace.

The most pressing need for interaction skills is in those situations where successful negotiations are essential to de-escalation of potentially explosive situations, setting the other person at ease, obtaining valid responses, and successful transmission of information. For example:

- Law officers must often respond to situations involving the mentally disturbed or requiring de-escalation of domestic disputes [6].
- Medical practitioners must learn to take patient histories (including asking personal questions about sensitive information) and interact with children [21,25].
- Military officers must establish situation awareness and effectively communicate operations orders [8,9].

Interaction skills trainers often lack the resources to provide extensive training in interpersonal skills, yet police officers, soldiers, medical practitioners, and others need this training to improve their
interactions with subjects, civilians, patients, and each other.

The most common method for learning interaction skills is through on-the-job experience. In most professions where interaction skills are critical, a form of apprenticeship is used to reduce the risks of ineffective interactions causing a tragedy. The quality of learning from apprenticeship varies with the quality of the mentor’s skills both as a practitioner and a teacher, and with the variety of learning situations that the apprentice encounters.

Formal interaction skills training, when it occurs, usually relies on peer-to-peer role playing or passive learning through videos. This leads to a critical training gap, because the students are limited in the practice time and the variety of scenarios that they encounter. But it is exactly this practice, as VR and computer based training studies show, that leads to significant on-the-job benefits.

TECHNOLOGY OVERVIEW

RVHT enables natural interaction with computer generated virtual humans, linking theory of human behavior with VR, knowledge representation, and NLP. In our architecture (see Figure 1), virtual humans act realistically without human intervention. Underlying virtual human action are a behavior engine, a virtual environment controller, and a natural language dialog processor [17,18]. We have developed RVHT products for training to conduct data collection interviews [3], for tank maintenance training [14], as virtual standardized patients [19], and for Emergency Medical Technician training [22].

RVHT applications provide an opportunity for practice with numerous case-based scenarios in a reproducible, objective learning environment prior to the challenge of actual engagement. The primary goal for RVHT is providing more cost-effective training. It is useful for initial training, sustainment training, and ongoing assessment of interaction skills. Training and sustainment benefits include enhanced adaptability, availability, and assessment, and reduced loss of effectiveness for students at distributed locations. In fact, RVHT architectures can be implemented with software that is designed to run on a relatively inexpensive laptop computer, so that it can be used on widely available personal computers, with distribution via compact disc or a network.

Although the scenarios are pre-defined, the interaction itself is unscripted. The scenario establishes initial conditions, but the student's responses to the virtual human, as well as inherent flexibility in how the virtual human is allowed to react, cause the conversational flow to vary from interaction to interaction. We believe this leads to a realistic learning application wherein the student must learn to handle each interaction individually.

RVHT Capabilities

RVHT integrates three basic capabilities:

**Behavior Modeling.** Models of human behavior are integrated into the architecture. These models specify how the emotional, physiological, and cognitive states of the virtual humans change based on user input and time course. They also specify how the virtual human should act given its new states. For instance, the virtual human should know to shake its head or hold up its hands when disagreeing with the user, but the emotional or physical state can temper or amplify the reaction. Similarly, an answer to a query on how the virtual human feels will depend on whether it represents a depressed person, a confused person, an injured person, or someone in a neutral state.

**Virtual Reality and Virtual Humans.** VR technology enables the construction of virtual humans that act realistically as if they are, for example, sad, confused, serene, or in pain. Using the behavior models, action takes the form of observable behavior, choice of utterances, conversational expectancies, and branching logic within the application. Activities occur and contextual cues reside in virtual worlds. Conversation does not take place in a vacuum; instead, the environment plays a large role in shaping conversational flow.

Figure 1. An RVHT Architecture
Natural Language Processing. NLP incorporates the ability to recognize natural, unscripted speech and to understand speech based on the content of the discourse. RVHT applications are designed to expect relevant, reasonable speech from the student, similar to that which occurs during regular conversation. As in regular conversation, expectations mature as the conversation progresses.

RVHT Architecture Considerations for Interactive Training

We use our RVHT software to create knowledgeable, emotional, expressive virtual humans, who populate virtual worlds both as actors within the interaction and as tutors. Our intelligent tutors can be customized by the user and can serve as demonstrator, trainer, coach, mentor, or observer [18].

An intelligent tutoring system includes a student model, instructor model, and expert model (see [24,32,33]). The simulation of virtual humans in our RVHT applications (see Figure 2) relies on the capabilities (and completeness) of these models.

Another key student model element is tracking whether or not the student has obtained the information necessary to make an appropriate diagnosis. For physician training, this means getting enough of the health history of the patient and information on the current symptoms. For law enforcement training, this means getting enough information on the subject to decide whether or not a crime has been committed, and whether or not the subject is dangerous to others or to himself. Student persistence for obtaining information from uncooperative subjects is rewarded, but repetition of questions which have already been answered is not, unless the student is trying to establish the consistency of subject responses.

The student model also tracks behavior skills that the student should be acquiring and practicing, such as politeness, responsiveness to the virtual human’s questions, and empathy for the virtual human’s feelings. Information on these skills is extracted through context-specific linguistic analysis of the student’s verbal responses to the situation. The student model also tracks negative student habits, such as the use of gratuitous profanity, impoliteness, and overuse of technical terminology and jargon.

Expert Model. The engine underlying virtual human behavior provides appropriate semantic and emotional reactions to the student’s inputs, following a flexible script. For training situations, the student needs to practice skills by applying them in a varying set of scenarios. However, the scenarios should provide consistent feedback and results to the student. Subject matter experts (SME’s) provide the basic inputs for the scripts, and also review the scripts to ensure appropriate and consistent results.

The semantic models allow the virtual humans to respond with answers, denials, objections, and challenges to the student’s requests, questions, and commands that are consistent with the script. Student input is analyzed to select the most appropriate group responses. The actual response depends on both the topic of the student input and the current emotional state of the virtual human. More generality in responses is possible with generated speech, but recorded speech is more realistic and provides much better feedback on the emotional state of the virtual human, so it is preferred for most training applications.

Figure 2. An Intelligent Tutoring System using RVHT Architecture

Student Model. A key element of the student model is ensuring that the student has succeeded in achieving the overall goals of the scenario. For a survey interviewer application, this involves getting inside the residence with a chance to administer the survey. For law enforcement training, this may mean persuading the subject to get into the patrol car without resorting to force.
The virtual human’s emotional state is updated at discrete intervals and also after each student behavior (such as a modification to the virtual world or verbal input). The update after student behavior is based on the semantic analysis of the student’s input and on the student model data. The virtual human’s emotional state is used to select a semantically appropriate, consistent response.

**Instructional Model.** The virtual tutor can operate in a variety of modes. In Demonstrator mode, the tutor acts as a surrogate SME to demonstrate good practices and techniques, showing the appropriate steps of a task and what operations need to be performed at each step. This mode is typically used to familiarize the student with the skill or skills being learned. In Coach mode, the tutor prompts the student through the sequence of steps. For example, the tutor may ask the student to select a topic for an interview, and then provide general recommendations for how to pose questions. This mode is typically used when the student is acquiring a skill or skills. In Mentor mode, the tutor offers suggestions, remediation, or critiques on the request of the student. The most common questions asked by the student are: “What do I do now?” and “How am I doing?” This mode is typically used when the student is practicing skills already acquired. In Observer mode, the tutor records and evaluates the student’s actions but does not interfere with the student’s efforts unless the student has acted outside the acceptable forms of behavior, particularly if the student’s actions in the real world could risk harm to the student or subject or harm to people or valuable objects in the virtual environment. The tutor either provides an after-action review (AAR) of the student’s performance after the student has completed the scenario, or provides the data collected back to the instructor so that the instructor can provide the review. This mode is typically used when the student is validating skills already acquired and practiced.

The instructional model provides feedback to the student in the form of dialogs with the student [11]. When the student is acquiring or practicing skills, the tutor provides recommendations for future student actions, and immediate and direct feedback on the student’s previous actions. The tutor uses the information collected on the student model, such as information on topics that have been discussed in an interview, to make suggestions to the student.

**Simulation:** A behavior engine simulates the changes in the behavior of the virtual humans and the virtual tutor. Of particular interest are the emotion engine that updates the emotional state of the virtual humans and the virtual tutor, and generates the appropriate facial expressions and gestures for the virtual humans; and the psychological and physiological models that govern level of pain, level of consciousness, and, when necessary for the application, observable characteristics such as breathing rate, blinking rate, and gestures such as scratching and coughing.

**Scripts:** The scripts capture the expert model, the student model, and the instructional model. With the help of SME’s, the scripts are created to define how the virtual human behaves at a specific point under specific conditions within the scenario. It easy to add scenarios or to adapt a script with variations in initial states, conversational flow, and virtual environment activities.

**User Interface:** We use a combination of VR and NLP for the user interface. For feedback from the virtual humans, either generated speech or recorded speech is used in combination with the virtual models. The user speaks into a microphone as the primary input mechanism, but we also use mouse clicks for navigation and synchronization.

**TRAINING CONSIDERATIONS FOR USE OF VIRTUAL HUMAN SYSTEMS**

**Integrating RVHT into a Training Regimen**

RVHT architectures are appropriate for learning-by-doing approaches to mastering interaction skills. However, RVHT is not a panacea, and should be used in combination with other training methods to achieve the most cost-effective training.

We analyze interaction skills training in terms of a four step process [10], considering the training methods appropriate for four steps in the learning process for each task to be learned. The four steps are:

**Familiarization:** Acquiring knowledge about the task by absorbing a presentation, watching a demonstration (e.g., by a tutor in Demonstrator mode), or by reading. This is a relatively passive process for the student.
Acquisition: Learning techniques and procedures by being tutored. The tutor (in Coach mode) guides the student through each step of the process, prompting the student to perform the action required for each step. If a student makes a mistake, the tutor provides immediate feedback.

Practice: Internalizing techniques and procedures by doing the skill with access to help from a tutor (in Mentor mode). The student performs the actions of the procedure without prompting from the tutor. At any point, the student may ask the tutor for help. If the student makes a mistake, the tutor provides feedback shortly after the incorrect action.

Validation: Testing the ability to perform the skill without help from a tutor. The student is on his/her own until either the task is successfully completed, or the tutor (in Observer mode) determines that the student cannot complete the task successfully. When the performance test has ended, either with success or failure, the tutor provides an AAR, interacting with the student to determine what went right, what went wrong, and how to improve his/her performance.

A typical approach to training interaction skills is to provide lectures, reading materials, or video tapes for familiarization, and then have the student acquire and practice the skills on the job. This can be an expensive and risky approach, but the costs and risks are hidden in operational costs and failures.

A more sophisticated but much more expensive approach is to again use lectures, reading materials, and video tapes for familiarization, but use an experienced worker as a mentor for the apprentice worker while the apprentice acquires and practices the skills. Validation is often a subjective evaluation by the mentor. This approach suffers from two drawbacks, the expense of the one-on-one training, and the variability of training experience due to its unstructured nature.

A third approach uses group sessions where carefully designed scripts are acted out. This is also an expensive approach, particularly for large classes, since these interactions are one-on-one. This cost severely limits the number of scenarios in which a student will participate.

RVHT can be used in combination with more traditional familiarization techniques to allow the student to acquire and practice basic skills in parallel, a variety of scenarios. By allowing many students to acquire and practice basic skills in parallel, RVHT can reduce the time needed for one-on-one interaction with human tutors and actors.

In many interaction skills training situations, students are being taught knowledge, skills, and attitudes. While we see RVHT as useful for acquiring and practicing skills, we believe that attitude changes must be taught through interactions with people with experience, not through the use of RVHT, so we expect that RVHT is best employed as a training aid, not as a stand-alone training course. We are experimenting with the use of RVHT for pre- and post-testing of students for law enforcement as a test of how broadly this technology can be used.

Training Benefits Achievable with Responsive Virtual Humans

Benefits of RVHT over traditional training methods that we have identified address both effectiveness and efficiency of training. For instance, RVHT represents more effective training through:

- Increased availability and diversity of role plays.
- Facilitated individual skill practice for greeting, identifying needs & concerns, handling diverse populations (e.g., in age, ethnicity, gender), dealing with difficult behaviors, and concluding interactions.
- Numerous virtual environments in which contextual cues can be added or changed.
- Repetitive, motivational practice at the student’s pace.

As with most computer based training, including VR-based training [8,15], RVHT represents more efficient training through:

- Reduced on-the-job learning of key interaction skills.
- Reduced training and travel time.
- Ease of modification, upgrading, extension, to new conditions, different types of virtual humans, psychological models, and physiological models.

Training Outcomes Achievable with Responsive Virtual Humans

These identified benefits naturally lead to some outcomes of RVHT training methods, to include:
Practice and feedback on specific behaviors, such as:
• concise sentences with clear enunciation,
• politeness and empathy,
• pausing during listening,
• restatement of needs & concerns, and
• confirmation of actions to be taken.
Learned flexible approaches.
Reduced job error rates.
Improved consistency by capturing best practices and expert knowledge.
Confidence gained before first real patient experience.

We believe that RVHT combined with traditional interaction skills training approaches results in some or all of these outcomes at considerably less cost of training time and dollars.

TRAINING APPLICATIONS OF VIRTUAL HUMAN SYSTEMS

We have created or are creating several effective, engaging training applications using RVHT.

Advanced Maintenance Assistant and Trainer
The Advanced Maintenance Assistant and Trainer (AMAT), developed by Research Triangle Institute (RTI) for the Combat Service Support Battle Lab [14], is a spoken-dialogue assistant and trainer for the maintenance of line replaceable units in the M1A1 Abrams tank. AMAT (see Figure 3) allows a mechanic to conduct a verbal dialog with a virtual tutor, who provides cues on how to find appropriate sections within Technical Manuals. Using AMAT, the mechanic can access important diagnostic information and procedures using voice input and output. During training, AMAT also allows the soldier to speak to the system to manipulate the view in the virtual tank.

Figure 3. AMAT Uses a Cartoon Virtual Human as a Synchronization Aid

Household Survey Interviewer Training
RTI created a survey interviewer training module, AVATALK-Survey (see Figure 4), that addresses survey nonresponse, a critical training need in survey research. Nonresponse to household surveys is increasing, despite extraordinary measures being taken to counter the trend [13].

Figure 4. AVATALK-Survey Allows the User to Interact with Virtual Respondents and Coaches

Research suggests that, to train how to solicit participation in surveys, effective training programs must address respondent concerns, train interviewers to develop strategies to adapt to cues provided by the respondent, and create a realistic learning environment [3]. Current survey practice does not offer a solution, instead leading interviewers to follow complex, standardized interviewing procedures. AVATALK-Survey, on the other hand, generates a variety of respondents showing a range of emotions, creates a virtual environment in which contextual cues can be
added or changed, and can be used for home study to supplement current training agendas.

**JUST-TALK Police Training**

RTI is developing an application for the National Institute of Justice, JUST-TALK, using RVHT to provide a computerized virtual person that interacts with the student in a similar way to the role-playing approach to training. JUST-TALK (see Figure 5) is designed to train civilian police officers to handle mentally disturbed individuals. It will be available on demand for practice, will allow self-paced study, and be usable on a home computer or on a computer at the police station.

An officer responding to a situation where the subject is mentally ill must quickly make a number of difficult decisions. In situations involving criminal actions, the officers are trained to use aggressive verbal techniques to quickly bring the situation under control. However, interaction with the mentally requires very different verbal interaction skills to de-escalate the situation. Students will learn in JUST-TALK to decide which verbal approach is most effective in a particular situation.

Figure 5. JUST-TALK Allows the User to Interact with Emotionally Unstable Virtual Humans to Learn De-escalation

The emphasis in JUST-TALK is on applying the technology to support law enforcement training, and determining the best way to deliver the technology to the many small police departments in the country. The technology is applicable to other related training problems where effective interviewing with the subjects is a critical skill for successful resolution of incidents.

**RESEARCH ISSUES**

There remain multiple research issues that RTI is pursuing which must be solved if responsive virtual humans are to reach the level of sophistication required for robust interaction skills training. These issues include:

- Under normal interaction conditions, when the virtual human is a calm adult, how is agent behavior modeled? What model parameters dominate when instead the virtual human represents an angry person, or confused, or a schizophrenic, or someone in pain? What if the virtual human is supposed to be someone who doesn’t speak the language? Or a child?
- Which sets of behaviors (e.g., facial expressions, gestures, body movement, intonation) will users interpret as serene, angry, confused, schizophrenic, pained, foreign, or childlike?
- What methodology can streamline the process of converting expert knowledge into agent behavior? Do the methods generalize across RVHT applications?
- Do responsive virtual humans make learning more accessible? How willing are students to accept virtual humans as interactive partners in learning?
- What skills can be acquired, practiced, and validated using RVHT? 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?

Addressing these research issues will open up a broad range of training and educational opportunities. We do not anticipate RVHT-based training to replace instructor-led training, but, based on our experience with VR-based maintenance training [15], we expect that combinations of RVHT-based training and instructor-led training will significantly reduce training costs and increase the number of people who can be effectively and consistently trained. As an additional return-on-investment, RVHT-based training can provide inexpensive, focused sustainment training. The key to opening up this broad range of applications is more robust and effective RVHT models and more efficient means of creating the models, as well as a better understanding of how to use RVHT in combination with other training methods to provide cost-effective training on critical interaction skills.
CONCLUSIONS

Responsive virtual human technology links theory of human behavior with virtual reality, knowledge representation, intelligent tutoring, and natural language processing. In RVHT applications, students may use visual cues as well as react to verbal responses to successfully complete each training scenario. We believe RVHT offers significant cost-effectiveness outcomes compared to traditional interaction skills training. Training and sustainment benefits include enhanced adaptability, availability, reusability, and assessment, and reduced loss of effectiveness for students at distributed locations.

ACKNOWLEDGEMENTS

Work on the JUST-TALK project is supported under Award number 2000-RD-CX-K002 from the Office of Justice Programs, National Institute of Justice, Department of Justice. Points of view in this document are those of the authors, and do not necessarily represent the official position of the U.S. Department of Justice.

REFERENCES


