

# Developing INOTS to Support Interpersonal Skills Practice

Julia Campbell, Mark Core, Ron Artstein, Lindsay Armstrong, Arno Hartholt,  
Cyrus Wilson, Kallirroi Georgila, Fabrizio Morbini, Edward Haynes, Dave Gomboc,  
Mike Birch, Jonathan Bobrow, H. Chad Lane, Jillian Gerten, Anton Leuski,  
David Traum, Matthew Trimmer, Rich DiNinni, Matthew Bosack, Timothy Jones  
University of Southern California Institute for Creative Technologies  
12015 Waterfront Dr.  
Playa Vista, CA 90094  
310-574-5713  
{campbell, core, artstein, hartholt, cwilson, kgeorgila, morbini, haynes, gomboc,  
mbirch, bobrow, lane, gerten, leuski, traum, trimmer, dininni, bosack}@ict.usc.edu,  
armstrong@ucla.edu, tjones.media@gmail.com

Richard E. Clark &  
Kenneth A. Yates  
Center for Cognitive Technology  
University of Southern California  
Rossier School of Education  
250 N. Harbor Dr., Suite 309  
Redondo Beach, CA 90277  
(310) 379-0844  
clark@usc.edu  
kenneth.yates@usc.edu

**Abstract**—The Immersive Naval Officer Training System (INOTS) is a blended learning environment that merges traditional classroom instruction with a mixed reality training setting. INOTS supports the instruction, practice and assessment of interpersonal communication skills. The goal of INOTS is to provide a consistent training experience to supplement interpersonal skills instruction for Naval officer candidates without sacrificing trainee throughput and instructor control. We developed an instructional design from cognitive task analysis interviews with experts to serve as a framework for system development. We also leveraged commercial student response technology and research technologies including natural language recognition, virtual humans, realistic graphics, intelligent tutoring and automated instructor support tools. In this paper, we describe our methodologies for developing a blended learning environment, and our challenges adding mixed reality and virtual human technologies to a traditional classroom to support interpersonal skills training.<sup>1 2</sup>

## TABLE OF CONTENTS

1. INTRODUCTION.....	1
2. THE INTERPERSONAL SKILLS DOMAIN .....	2
3. INTEGRATED SYSTEM DESIGN .....	3
4. CONCLUSIONS .....	10
REFERENCES .....	11
ACKNOWLEDGEMENTS .....	12
BIOGRAPHIES .....	12

## 1. INTRODUCTION

### *The Problem*

Interpersonal communication skills are critical to effective leadership in the United States Navy. Leaders employ effective communication strategies to reach mutual understanding, clearly relay performance expectations, and to help resolve personal and performance issues. Honing interpersonal communication expertise requires practice, and although these skills have been identified as vital for

successful leaders, they are often learned on the job. There are several challenges to providing novices with interpersonal communication skills training before reaching their first job assignment.

Communication skills training is most effective when it is goal-based, and when the training explicitly defines the strategies for reaching the goal and the communication skills for supporting the strategies [1]. Major challenges for communication skills training, however, are a failure to clearly define the communication skills and a failure to provide a framework to organize communication strategies and skills being trained [1]. Additionally, role-play practice with feedback has been shown to have the greatest effect on performance scores [2, 3]. Yet role-play sessions, especially with novices, will be inconsistent, and may not present opportunities for practicing all of the target skills. Finally, feedback following role-play sessions will also be inconsistent without an organizational framework that identifies the appropriate skills and strategies required to reach a successful outcome.

The specific problem identified for new Naval officers was that they were not prepared to face difficult personal and professional issues from their subordinates. What is needed for Naval officer communication skills training is a framework for learning the strategies and skills to help a subordinate solve a personal or performance problem. To practice applying the skills and strategies, they need a controlled role-play environment designed specifically to allow the trainees to demonstrate they have acquired the strategies and skills. In order to support feedback following the practice session, trainees must have a clear understanding of the criteria with which to assess their peers. Naval instructors must also be provided with tools to objectively assess interpersonal communication skills during a role-play session for an entire classroom.

Our solution to the challenges inherent to traditional role-play practice is to replace the live novice role player with a virtual human subordinate. The virtual human's behavior and dialog are tailored to provide the cues that indicate to

<sup>1</sup> 978-1-4244-7351-9/11/\$26.00 ©2011 IEEE.

<sup>2</sup> IEEEAC paper #1321, Version 2, Updated October 26, 2010

trainees when to apply specific strategies and skills. The simulation role-play is integrated with class instruction in a blended environment, allowing the entire class to participate in a single role-play, stimulating discussion and peer-to-peer evaluation, and allowing the instructor to assess class performance in real time.

The University of Southern California (USC) Institute for Creative Technologies (ICT) developed the Immersive Naval Officer Training System (INOTS) in order to address the training challenges previously described. INOTS offers an environment for officer candidates to practice resolving realistic personal or performance issues using interpersonal skills embedded in authentic dialog. The practice takes place in a mixed reality setting where a single trainee playing the role of an officer participates in a one-on-one interaction with a virtual human subordinate. The virtual human responds with appropriate dialog and gestures. Using a virtual human role player, we ensure a consistent interaction experience because the virtual human behavior and dialog is designed to trigger the human role player's application of the appropriate interpersonal communication strategies and skills. The strategies and skills are represented by the human role player dialog choices and delivered through a conversation between the human and virtual human. The secondary interaction area is the classroom, where a class of trainees observes the primary interaction and takes part in the practice via the use of Audience Response System (ARS) input, or "clickers." The clickers allow trainees to follow the experience and make their own choices. Linking the classroom experience with the individual practice session encourages class participation, tracks individual and group performance in real time, and guides the instructor through this interpersonal communication skills training session.

The following sections outline the approach to developing the INOTS blended learning environment. First, we define the interpersonal skills domain and the instructional design elements that drive training and system development. In the second section, we describe the integrated system design including how we developed authentic branching dialog, managed the spoken interactions between the virtual human and human, modeled the virtual human, tracked student performance during the experience, and created automated instructor support tools to organize the experience.

## **2. THE INTERPERSONAL SKILLS DOMAIN**

We employed several methods to gain an understanding of the training challenges associated with the interpersonal skills domain. First, we reviewed the existing curriculum for a Navy officer leadership course and surveyed literature relevant to leadership and interpersonal communication skills. Then, we conducted cognitive task analysis (CTA) to generate the instructional design. The instructional design includes the framework for clearly defining the communication skills and organizing the communication strategies and skills being trained. This framework also

supports system development. Cognitive task analysis is an overarching term to describe any number of information gathering and interview strategies to capture expert decision-making and cognitive processes [4, 5]. CTA methods also elicit and represent the underlying expert knowledge and skills that inform training programs and system design [6].

### *Leadership and Interpersonal Skills*

Preliminary research and anecdotal evidence suggested that newly commissioned officers were not prepared to face difficult personal and professional issues from their subordinates. While interpersonal leadership skills training is a domain covered by the Navy's leadership course curriculum, sufficient practice of these skills is determined by the exercises used by each instructor and the skill and motivation of each trainee who acts as a role player. Practice exercises do not focus on the process of applying interpersonal leadership skills in communication to successfully achieve a desired result (e.g., motivating a subordinate to improve performance or guiding a subordinate through a personal issue). A survey of relevant research provided strong support for the role of interpersonal communication skills in team leadership. Leaders with strong interpersonal skills maximize performance and subordinate motivation [7, 8]. The interpersonal skills necessary for effective leaders include communication that is behavior-oriented, not person-oriented, corresponding verbal and nonverbal messages (maintaining eye contact, leaning in toward the speaker, maintaining a neutral posture and neutral facial expression), positive feedback towards the individual, and active listening (listening closely to the speaker without interruption, summarizing what the speaker has said, and asking the speaker to confirm understanding) [9, 10]. Once we identified these skills, we needed a process for determining how Navy expert leaders use these skills to help solve performance or personal issues with subordinates.

### *CTA to Support Instructional Design*

We employed comprehensive CTA interviews to develop a step-by-step approach for resolving personal and performance issues using interpersonal skill strategies. The CTAs were conducted using the Concepts, Processes and Principles (CPP) [4] CTA method and Critical Decision Method (CDM) [11]. We asked expert Navy officers to describe the cues for recognizing when a subordinate had a performance or personal problem, what strategies they used to address the problem, what courses of action they pursued to solve the problem, what alternatives they could have chosen to solve the problem, and what variables in the situation would have caused them to make a different decision or take alternative actions.

Once the CTAs were conducted, we cross-referenced that information with the Navy's officer leadership curriculum and relevant literature in order to identify the INOTS training goal, supporting learning objectives,

communication strategies and skills, and cues for initiating and applying the strategies and skills.

We developed an acronym for the strategies: ICARE. The strategies are: Initiate conversation, Check for causes, Ask questions, Respond with course of action, Evaluate by following up. The ICARE acronym is designed to help candidates remember the strategies and the order in which the strategies should be performed (e.g. one should not respond, or provide a course of action, before determining the causes of the problem). Under each strategy are communication skills, or specific observable actions, to achieve each step. It is not as crucial for these actions to be performed in order. For example, under the strategy “Initiate” are skills associated with communicating performance expectations. The rule of thumb during this step is to focus on performance feedback that bridges gaps between behavior and the desired goal. There is flexibility, however, regarding when to describe the target behavior versus when to explain how the subordinate’s behavior impacts the individual, the team and the mission. The context of the conversation may determine which action is more appropriate.

The foundational skills referenced in a majority of the interviews were active listening and non-verbal communication. Active listening and appropriate non verbal communication are essential to understanding and applying the ICARE strategies and skills which are the foundation of the INOTS system. This information is integrated into the instructional design blueprint which provides instructor support materials in addition to driving system design.

### **3. INTEGRATED SYSTEM DESIGN**

The INOTS system is developed to support instructional design objectives while integrating emerging ICT and commercial educational technologies, in order to provide an engaging and educational user experience for instructors and trainees. The INOTS software design functionality includes direct interaction in ordinary speech with a virtual human, who responds with speech and body gestures. The mixed reality environment represents a workspace setting and life-sized virtual human that are both rear-projected onto a screen. In the first section, we discuss the conversation manager of this virtual human, how it decides what to say next in the conversation. We then describe the visualization of character, how it was brought to life on screen. The third section deals with the understanding of the spoken language of the learner in the mixed reality environment. Last, we

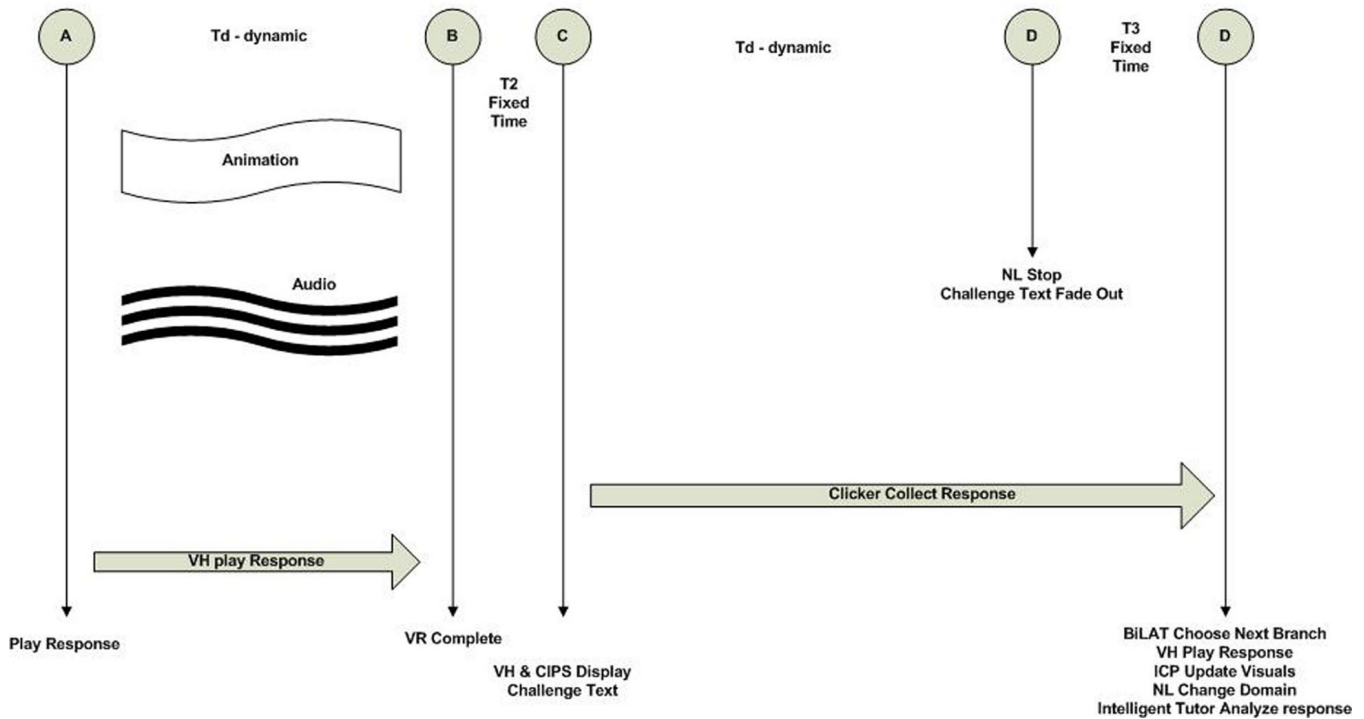
describe the intelligent tutoring technology, which tracks the trainee in the mixed reality environment as well as trainees in the classroom who are participating along in the scenario using clicker technology. This automated instructor support facilitates monitoring of the exercise and conducting an After Action Review through a component called the Instructor Control Panel which, among other capabilities, allows visualization of performance through graphs as well as video replay. The goal of these sections is to provide a high-level overview of the authoring processes and corresponding system functionalities. Each of these major functionalities corresponds to one or more software components that run independently and communicate via messages but we omit the specific implementation details here.

#### *Conversational Resources and Management*

There are a number of standard approaches to implementing a virtual human that interacts through speech with humans. One requirement here was that the virtual human’s responses reflected the conversational context. Unlike some approaches where only what was last said is considered, this virtual human had to be a realistic simulation of a subordinate with a personal/performance problem. Thus, emotions should build over time and the problem should be dealt with a piece at a time rather than being solved with a single response. Another requirement was authoring by novices meaning that non-computer scientists could understand the logic used by the virtual human to decide upon its response and use this understanding to create the necessary conversational resources for the character.

To meet these requirements we chose to use a branching narrative as the central representation of our conversation manager. The logic of the branching narrative is simple to understand; each utterance by the virtual human constitutes a decision point with a fixed number of possible responses by the human. Each response is linked to another decision point in the narrative representing the virtual human’s reaction and the next set of possible utterances for the human. In the conclusions section, we address the drawbacks of the approach: learners are not formulating responses to the virtual human spontaneously and instead are reading from a menu of permitted responses. Here, we note the features of the approach. It allows novice authors to write responses as English text and know exactly what could have been said before the response. This allows the psychologists that conducted the CTA to be directly involved with the authoring, and allows the full conversational context to be used in crafting the responses.

## Typical INOTS Interaction Timeline



**Figure 1—INOTS Interaction Timeline.**

The authoring process started with the creation of a linear script guided by the instructional design that represented the best possible course of action for resolving the virtual human's performance/personal issue. This linear script was loaded into an open source tool [12], which allowed the authoring of non-optimal responses and the creation of alternative story lines and branches for these choices. To keep the process manageable we would recombine branches of this tree or create a bad ending for a branch with several non-optimal choices. As described in the Automated Instructor Support section, choices in the narrative were linked to the skills identified in the CTA as part of the authoring process. Once the branching narrative was complete it was translated into a format understandable by our conversation manager. When the virtual human software is running, this software module keeps track of the made by the learner in the mixed reality environment.

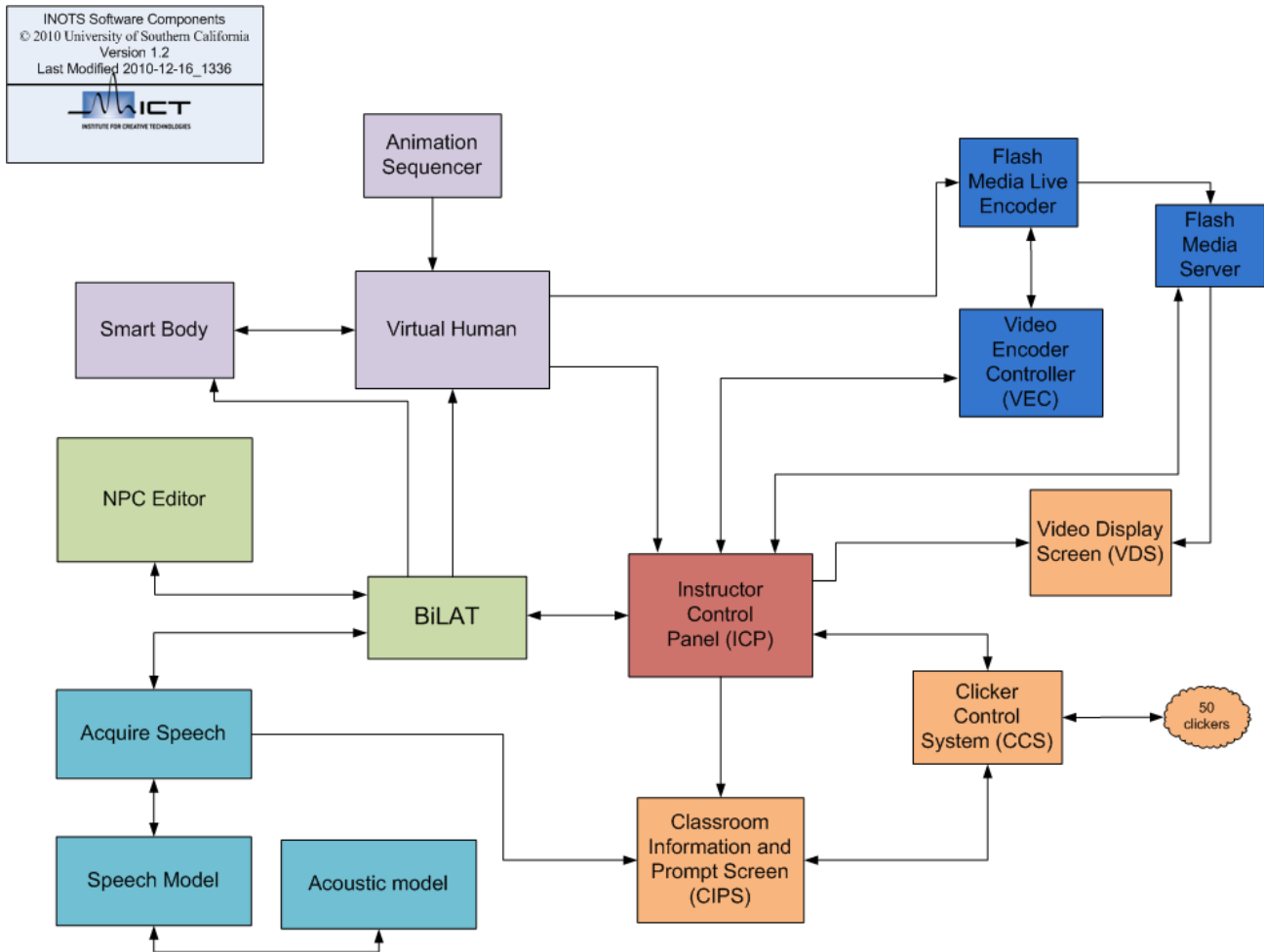
### *System Architecture Overview*

We refer to the conversation manager described in the previous section as BiLAT since it is a version of the bilateral negotiation training system described in [13]. BiLAT is the center of the INOTS architecture and handles all communications with other components in the system and maintains the conversational state. At each node in the branching narrative, BiLAT broadcasts the state or domain to all other modules in the system. It also triggers the virtual human to speak the next statement (also known as a challenge) in the narrative. After the virtual human speaks, the system displays several possible responses for the

student to choose. The underlying messaging architecture utilizes an open source messaging broker, [14] which multicasts all system messages. This allows each component to be aware of all relevant messages.

BiLAT receives the speech recognizer output from the Acquire Speech module described in the Audio Acquisition section. BiLAT then sends this raw data to the classifier described in the Interpretation section. Once the classifier identifies the most likely answer, it broadcasts this classified response back to BiLAT. BiLAT then advances to the next appropriate node in the dialogue tree.

During this interaction, the instructor control panel (ICP) sends a request to the Classroom Information and Prompts Screen (CIPS) classroom display system to present the response choices in the classroom. The ICP communicates with the wireless clicker system and collects classroom responses tracking each trainee's response. The ICP analyzes each response for accuracy and updates its real-time metric displays as described in the section Automated Instructor Support. Figure 1 summarizes the steps described above. Additionally, the intelligent tutor component generates a suggested customized discussion plan for the AAR. During the entire experience, the ICP collects video feed of the trainee and virtual human for live display during the interaction and for replay during AAR. Figure 2 is a component diagram showing the components introduced here and discussed in detail in the following sections.



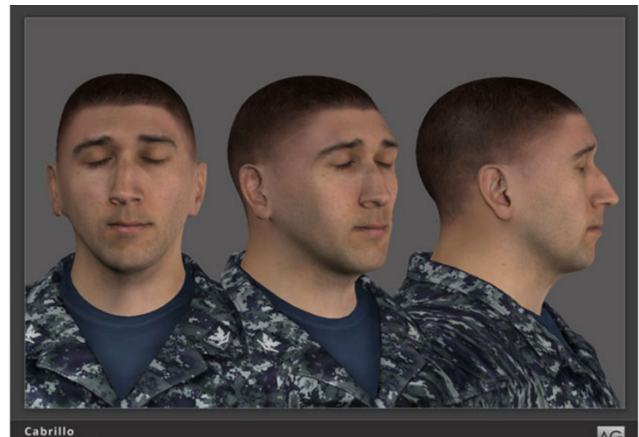
**Figure 2—INOTS Technical Diagram.**

### *Character Visualization and Animation*

For the virtual human character to be an effective aid in exploring interpersonal issues, it is important that the character expresses emotion not just through dialogue, but also through facial expressions and body language. In addition, the character needs to be situated in a realistic context. These aspects are explored in the following sections.

*Face*—Visualization and animation of a realistic, expressive face remains one of the most significant challenges in character animation. To confront this challenge, we constructed our character’s face model from extensive measurements of a real face. Our overall approach was based on [15], with significant modifications to meet the constraints of a real-time, interactive simulation

The face measurement process involved photographing the subject under a set of illumination patterns projected by a sphere of 160 LED lights, as described in [16]. This set of photographs, taken through a pair of digital SLR cameras, was then aligned to compensate for movement of the live subject, using techniques from [17]. The images were then used to compute several channels of information about how the face reflects light, including the overall shape, the color texture, and fine variations in surface orientation arising from details such as wrinkles and skin pores. Given the level



**Figure 3—INOTS Virtual Human facial model.**

of detail encoded in these data, it is then possible to use the data to generate realistic images of the measured face, under arbitrary lighting and viewpoint. In our simulation, this operation is performed in real-time on modern graphics hardware, using a custom shader, as described in [18].

In order to achieve a face model that could be animated, we repeated the measurement process for a variety of facial poses, so that we could recover information on how the face deforms into each pose, similar to [15]. We computed rough correspondences between neutral and non-neutral scans

using optical flow [19] on the scan data. We used the rough correspondence information to produce an initial estimate of how the neutral mesh should deform into each target pose, which we then refined through manual artistic manipulation. The preliminary results can be seen in Figure 3.

*Body*—The body model consists of an artist-created outer mesh that is driven by an articulated skeleton. This skeleton is manipulated by SmartBody, an in-house developed procedural animation system [20]. SmartBody controls the character, based on behavioral messages defined in the Behavior Markup Language (BML) as part of the SAIBA framework [21]. The requested BML behaviors are converted to a schedule of real-time animation controllers and timed system messages. For every animation frame, the animation controllers are run, resulting in a new posture for the character skeleton, sent to the renderer via a combined TCP/UDP protocol. For INOTS, these BML behaviors were hand-authored for each utterance, in order to have complete control over the virtual human character’s body language (in order to capture precise nonverbal communication, as warranted by the instructional design). The animations themselves are individual, key-framed animations, created by in-house artists, which are connected together in real-time by SmartBody. Rather than make full animations for each of Cabrillo’s responses, this allows us to create a suite of ‘building-block animations’ that can be reused in many different situations in novel combinations to create unique full body behaviors.

*Environment*—The INOTS instructional design required that the interaction with the virtual human take place in a superior’s office – ostensibly that of an officer being “played” by one of the trainees. This office is modeled on reference pictures and provides an appropriate, realistic, yet non-specific, context for the classroom training sessions.

Leveraging previous ICT work creating immersive environments, we developed a mixed reality environment for the INOTS project. The INOTS mixed reality space includes physical props such as an officer’s desk, chair, computer screen and telephone. The digital environment includes the virtual human character sitting in a chair opposite the physical desk, and a virtual background of the office including a virtual file cabinet.

### *Natural Language Interaction*

Interaction with the INOTS character is accomplished using conventional speech. The character communicates using pre-recorded audio clips which are animated off-line (see the Character Visualization and Animation section), and the dialogue itself follows a scripted branching dialogue, managed by the conversation manager (see the Conversational Resources and Management section). The Natural Language components described in the following sections are the ones used to interpret speech input from the user; they include audio acquisition, speech recognition and interpretation.

### *Audio Acquisition*

The user talks to the virtual human through a microphone that feeds directly into the computer, using a mouse as a press-to-talk device. A custom-built speech acquisition client interprets press and release of the mouse at the beginning and end of user speech, and sends the audio stream to the speech recognizer when the button is pressed. The client retains a 500 millisecond buffer before and after button press and release to allow for a small margin of error in the user’s operation of the mouse. The client also handles the communication between the speech recognizer and the rest of the system, receiving speech recognizer output and sending it to the conversation manager over the messaging system.

### *Speech Recognition*

We use PocketSphinx [22] as our speech recognition engine. PocketSphinx receives user audio from the acquisition client and returns the most likely string of words. The search space is constrained by acoustic and language models, which represent expected probability distributions of sounds and words, respectively. At present we use the default acoustic models that come with PocketSphinx; as the system approaches deployment, we will adapt these models with recordings from the field. Language models are trained from the data in the user prompts stored in the conversation manager (Conversational Resources and Management section), with additional paraphrases entered manually. As we collect more user utterances, these are added to the training data in order to make speech recognition more accurate.

Since the expected utterances are different for each state of the conversation manager, the language models can be state-specific, so as to recognize only utterances appropriate to the conversation state. However, our experiments so far show that a single language model works better than individual, state-specific models. There is likely too little training data for each state, so the state-specific language models are too small, causing the recognizer to run out of hypotheses very quickly. We will revisit the idea of state-specific recognition when we have collected more training data.

### *Interpretation*

Each user utterance is interpreted as one of the three options presented to the user at the current conversation state. This functionality is carried by NPCEditor, a statistical natural language classification tool [23]. The classifier learns a mapping between sample input-output pairs; for each new input the classifier uses the learned mapping to compute a language model of the hypothesized output, and selects the closest available output to the hypothesized language model. Each conversation state constitutes a separate instance of classifier training data, with the output being the three possible user options; the input for training are the same

manual paraphrases used for training the language model for speech recognition.

For a given speech input, NPCEditor generally returns the best fitting meaning output. It also has the capability to indicate that the best output is not a good fit, signaling that the utterance is un-interpretable and leaving it to the conversation manager to decide how to handle it (a typical action in this case would be for the character to repeat his previous statement).

Making this option available generally has the effect of reducing misunderstandings—that is, utterances that are interpreted incorrectly—at the cost of increasing non-understandings—that is, utterances that receive no interpretation, even if a forced interpretation would have turned out correct. In our testing so far, we found that classification is generally very reliable, so we opted to force an interpretation on every utterance, accepting an increase in misinterpretations for the benefit of eliminating non-interpretations. This decision may change if classifier performance on field data turns out to be substantially different than in the lab.

### Automated Instructor Support

We are using intelligent tutoring system technology to assist instructors in the context of the unique challenges of conversation simulation and the large volumes of data generated by having learners participate via ARS, or “clickers.” In the first section, we discuss assessing actions in the simulated conversation through expert modeling. Unlike a simulation of a physical process (e.g., flying an airplane) it may not be obvious to instructors how choices in the simulation link to the instructional material. The second section describes the INOTS instructor control panel, which facilitates an instructor reviewing records of learner actions and their assessments, and using this data to conduct an After Action Review.

### Expert Modeling for Scripted Characters

In typical simulation-based training, the results of the learner actions on the simulation are used to judge the correctness of their actions. When conversations, rather than physics are being simulated, it becomes more difficult to judge correctness. The conversation partner may simulate being upset and emotional and thus correct answers will not always result in positive reactions. It is also the case that the conversation partner may not provide clues why actions are correct or incorrect since doing so could break character. Thus, it is crucial to be able to link learner actions in the simulated conversation to the instructional design.

Specifically in the case of INOTS, the goal is to link concrete communication skills (actions) to the strategies or steps that an expert leader would use to solve certain types of interpersonal problems. Each strategy is composed of actions and decision steps which may be further decomposed into additional actions and decision steps

reflecting how a complex task breaks down into simpler ones, or a complex rule has exceptions. Table 1 shows some of the actions and decision steps of the procedure “Initiate Conversation (when dealing with a performance problem)”. The rule of thumb during this step is to focus on performance feedback that bridges gaps between behavior and the desired goal. There is flexibility, however, regarding when to apply each action.

**Table 1—Actions & Decision Steps.**

STRATEGY	Initiate Conversation: State performance issue.
SHORT NAME	State performance issue.
ACTIONS AND	State the performance problem focusing on behavior affecting performance.
DECISION STEPS	<p>Ask if the person is aware of the problem.</p> <p>If the person is not aware of the problem, re-state performance problem and use active listening.</p> <p>If the person is aware of the problem, describe target behavior to eliminate the performance problem.</p>

The scripted dialogue of the virtual human and the possible responses for those playing the officer were designed to allow practice of these procedures. In addition to opportunities to correctly follow the procedures, the scenario developers had to create plausible alternatives such that learners who had not mastered the procedures would be tempted to select these choices. The goal is to link these correct responses and the alternatives to actions and decision steps of the procedures to be taught. Positive links indicate the response embodies an action or decision step performed correctly and in the appropriate context. Negative links either classify the type of error made, or indicate an action or decision step that should have been performed but was not.

Table 2—shows a decision point from a draft scenario. The decision occurs after the dialogue with the officer candidate has already begun and the first row contains the virtual human’s reaction to a previous utterance by the officer. Rows two through four are possible responses the learner chooses from. Row two is a correct response and has positive links to two actions in the Active Listening procedure: summarizing what the other person is saying and asking for confirmation. Row three reflects an incorrect answer, because in no way does it match the procedure’s description of how to deal with such an issue. It is labeled with a negative link to the correct action, summarize. Row

four is a mixed response and illustrates that the procedures do not dictate a unique solution to every situation.

**Table 2. Example of a decision point.**

Virtual Human	Potential Learner Response
Yes, Sir, but I can't focus. That's why I need to get off the ship.	
	Are you saying that being underway is what's hurting your performance?
	You can do this, GM2. It just takes some discipline.
	If you focus on work, we might run the chit. Okay?

Here, the officer is reminding the virtual human of the performance problem; he is distracted and performing poorly. This is a reasonable response and is positively linked to the action, confirm performance expectations. However, the officer is also saying he/she may put in a request for shore leave for the character. This is not the best response because the officer has not diagnosed the underlying cause of the problem, and will either lose a good sailor unnecessarily or fail to solve the problem if the shore leave request is denied. As noted in table 1, one action in the strategy, state performance issue, is to identify the target behavior you want your subordinate to achieve. Here, the officer is identifying the wrong target behavior (shore leave) and thus we make a negative link to this action. The combination of both positive and negative links mean this response is mixed and neither totally correct nor totally incorrect.

Although the primary purpose of the linking process is supporting the intelligent tutoring system (here embodied as the instructor control panel), it is also a useful cross-check for the authoring process. Authoring such scenarios is a delicate balance between holding learner interest, creating realistic dialogue and plausible conversational options, and enabling practice of the procedures to be taught. The links can be used to check whether learners are being given ample opportunities to choose the different actions and decision steps of the procedures. It also allows the instructional designers to control factors such as: should multiple correct answers be allowed, should there be a continuum of answers (e.g., good, so-so, bad), and should there be decisions with no completely correct response.

Although we also used this linking process in the BiLAT training system [13], there are a number of interesting differences. One difference is that the domain of BiLAT, negotiation in an Iraqi cultural setting, contains non-procedural aspects such as small-talk and avoiding cultural taboos. To address these non-procedural aspects, BiLAT

implements a conversation simulation by providing a relatively large set of conversational options (30-60 on average) from which a learner can choose their next conversational move during the dialogue parts of the scenario. By necessity most of these options are fairly generic (e.g., "flatter host") and do not get into the details of what would actually be said in the conversation. INOTS takes a slightly different approach where each step of the conversation has a limited set of choices, but the choices are highly contextual and contain the exact words that the learner is choosing to say. Because these choices and the virtual human character's response are scripted in advance, the positive and negative links to the procedures are all that is needed to assess the learner's choices. A choice with all positive links will be deemed correct; all negative links mean incorrect, and a combination of positive and negative links means the choice is mixed.

#### *Instructor Control Panel*

After a learner interacts with the life-sized virtual human in the mixed reality environment, and the class votes on their preferred choices at each decision step, there is a large quantity of data for an instructor to consider and potentially use during an After Action Review (AAR). The instructor control panel (ICP), is the instructor's interface for accessing this data, and the key design consideration was to avoid overloading the instructor while still allowing them to explore the data. In this section, we describe the main features of the ICP. The application is broken into sub-windows corresponding to different views of the data.

Instructors can focus on views of interest to them, and additional data can be accessed by interacting with the sub-windows. Once we are able to observe the system in use we will be able to evaluate how much the individual sub-windows are used as well as whether we achieved the goal of a simple interface that allows browsing the full set of data.

Figure 4 shows a mock-up of the instructor control panel and its five sub-windows. In the upper left quadrant is the video window. This split screen allows the instructor to view the interaction between the virtual human (left) and learner (right) as well as show selected aspects of the interaction during the AAR. Underneath the video window is the narrative window, which contains a textual transcript of the interaction as well as a graph allowing the instructor to explore what other options the learner interacting with the virtual human character could have chosen.

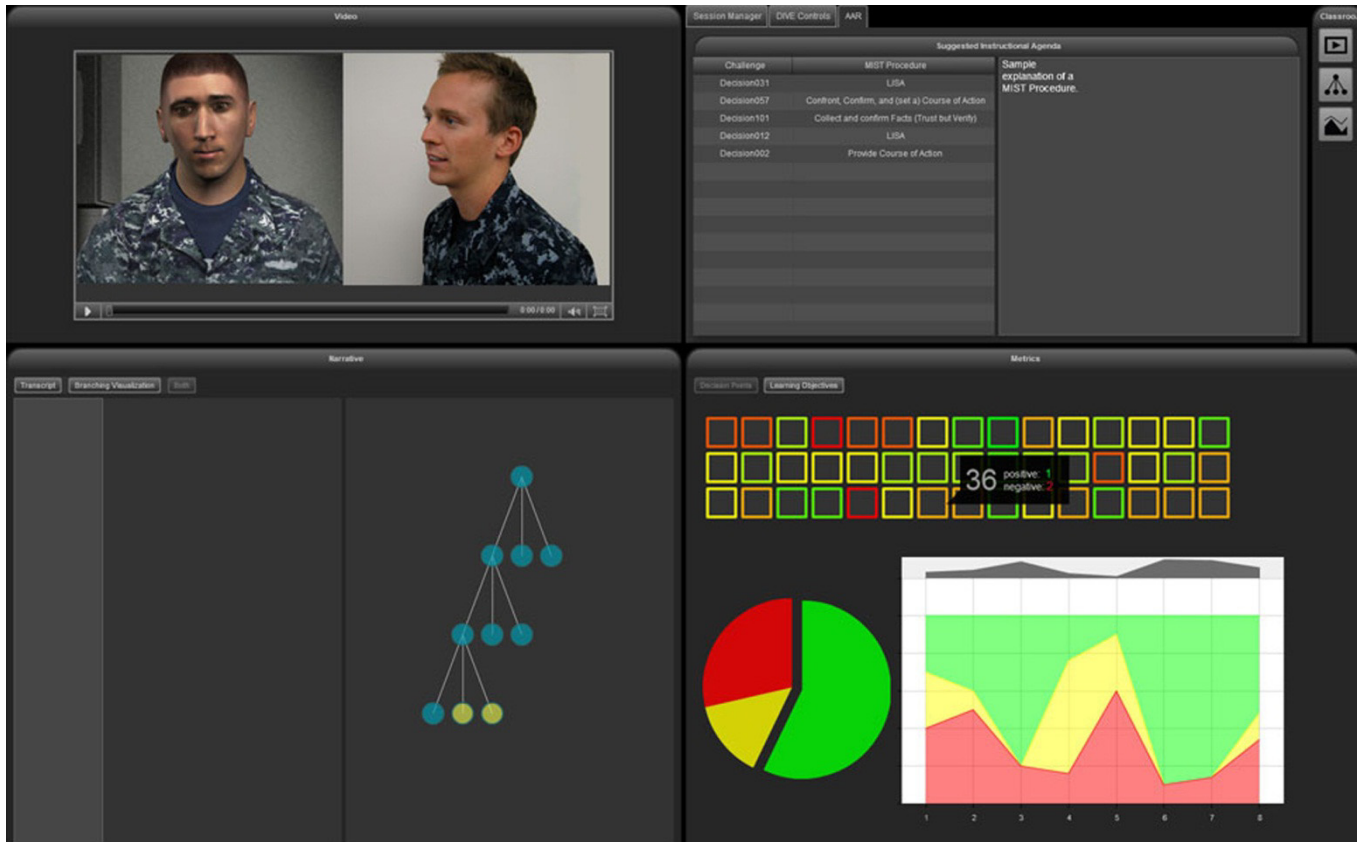
On the lower right quadrant is a seating map window, and a graph window. The colors on these displays follow the convention of green=correct, red=incorrect, and yellow=mixed. For example, the pie chart shows the proportion of learners in the classroom selecting each of the three possible choices with each choice colored according to its correctness.



The line chart shows the aggregate performance of the class over time (i.e., how many correct, incorrect and mixed options they have selected). These charts can be switched out at the push of a button, and we are also planning to have a chart of aggregate performance organized by procedure. The seating map displays the aggregate performance for individual learners with seats getting more red as errors are made and moving through the spectrum towards green as choices with positive links are made.

on these monitors. During the AAR, the instructor can decide what data to present on the monitors.

One design principle underlying the ICP was linking the sub-windows to enable intuitive browsing of the data. For example, instructors are not required to use the video player controls to reach a particular decision point. Instead they could find the decision point in the transcript and click there to move the video to the proper location. The instructor could also click on a decision point in the graph window in



**Figure 4—Instructor control panel with quadrants visualizing runtime data.**

The upper right quadrant contains a sub-window that houses the actual controls of the control panel. During the AAR, as shown in Figure 4, it is also used to display suggested topics for the AAR in the form of a list of the noteworthy choices made. For each of these choices, links to the relevant procedures are listed, and documentation of the relevant procedures are provided. Empirical evaluation will be needed to determine how best to narrow down the list of decisions made in the scenario to the most important. We currently plan to consider correctness of decisions made by learners (prefer to address errors) as well as how much learners agree on what decision to make (prefer to address disagreement). Although this is the instructor's private display, the instructor can mirror the video, narrative, and graph sub-windows on monitors in the classroom. While one learner is interacting with the virtual human character in the mixed reality environment, the rest of the class will be seeing the possible choices at each step as well as the video

an area of interest, see the different choices in the narrative window and then watch the video of the relevant dialogue.

These displays within the instructor control panel serve two roles. They allow the instructor to see the intentions of the scenario developers in creating the different responses for each decision point, and they allow instructors to sort through data from an entire class of learners to plan and conduct an AAR. Instructors can see whether responses are correct, incorrect or mixed using the coloring of the graphs. Decisions in the suggested AAR topics will have their links to procedures clearly shown allowing for an additional level of detail. Instructors will be able to see in real time the decisions of the learners: whether the class reaches a majority agreement, whether a majority of the class agrees with the learner in the mixed reality environment, and whether the class or learner in the mixed reality environment was correct. Before starting the AAR, the instructor will have a list of suggested topics as well as

graphs of learner behavior over time to assist in organizing the AAR. The instructor can use these graphs during the AAR as well as using the video playback and narrative windows to remind learners what happened.

#### 4. CONCLUSIONS

At the time of this paper's submission, we have yet to conduct comprehensive user reaction and learning assessment studies. We anticipate conducting these tests May – June 2011. Once instructors and trainees have had an opportunity to use the INOTS to support their training activities, we will discover which components of the system are more useful than others and which should be modified or removed altogether.

The system integration described here creates a platform for this type of empirical research. The visuals described above and the audio provided by a voice actor allow the investigation into the importance of realistic depiction of an emotional setting that a leader might face in the future. The system is modular so that different components can be swapped out as needed with new components supporting the messaging protocol. The integration of intelligent tutoring system technology allows us to explore the area of instructor support or in the future consider a PC-based version that learners could use outside of the classroom. Many software challenges were met to integrate the various components, and an additional challenge/opportunity was getting different groups with their own vocabulary and culture to truly work together. INOTS can be considered a "serious game" as it uses many technologies from the video game industry; but with this project, instructional designers were not merely consultants on the project but actively changed the way the character in the "game" behaved. One challenge is that we lack proper authoring tools to support the complexities of having multiple authors, and moving forward it is important to streamline this process and eliminate the need to reconcile multiple versions of the same data.

A persistent concern is evaluating the methodology of having a fixed set of responses for each utterance spoken by the virtual human. Having the users read predetermined prompts helps achieve good performance prior to collecting data on what users say when not prompted. However, with data collection, the methodology could be made more free form by replacing the utterance prompts with more general instruction prompts. For example, instead of prompting the user with an utterance like, "I understand you are concerned about being away from your family," the user will be prompted with an instruction such as "Acknowledge Cabrillo's concerns," and allow the trainee to formulate the phrasing of the utterance. Class members participating via the Audience Response System will select one of the generalized responses using their clickers. In this data collection stage, the instructor will act as a wizard, bypassing the natural language understanding and selecting the appropriate user choice manually. After we have

collected sufficient instances of users' formulations of utterances for such prompts, we can retrain the natural language understanding module to identify this wider set of user responses.

A second type of data collection would eliminate the prompts altogether, leaving the users to not only formulate the utterance the way they want to, but to also choose whatever content they want. This would allow the instructional design team to see what types of actions learners perform in the scenario. However, one drawback of using a branching narrative for conversation management is that increasing the number of alternatives for each decision quickly becomes too complex for human authors or results in unrealistic stories (e.g., stories that end suddenly or many choices leading to the same place). Future work on this limitation will involve using more complex models of emotion and memory rather than relying upon location in a branching narrative to implicitly encode these factors for the virtual human.

The Navy's Immersive Naval Officer Training System (INOTS) is the first blended learning environment to incorporate a life-sized virtual human to support the instruction, practice and assessment of interpersonal communication skills for Navy officer candidates. Ultimately, the instructors and trainees will determine how and to what extent they perceive the INOTS as an effective training system. With user feedback, we will continue to refine our methodologies in order to develop similar blended learning environments for additional training domains.

## REFERENCES

- [1] Richard F. Brown and Carma L. Bylund, "Communication Skills Training: Describing a New Conceptual Model," *Academic Medicine*, 83, 37-44, January, 2008.
- [2] Gordon E. Mills and R. Wane Pace, "What Effects Do Practice and Video Feedback Have on the Development of Interpersonal Communication Skills?" *Journal of Business Communication*, 26, 159-176, March 1989.
- [3] M. Berkhof, H.J. van Rijssen, A.J. Schellart, J.R. Anema and A.J. van der Beek, "Effective Strategies for Teaching Communication Skills to Physicians: An Overview of Systematic Reviews," *Patient Education and Counseling*, (in press) July 2010.
- [4] Richard, E. Clark, David F. Feldon, Jeroen Van Merriënboer, Kenneth A. Yates, and Sean Early, "Cognitive Task Analysis," in *Handbook of Research on Educational Communications and Technology*, 3rd ed, J. M. Spector, M. D. Merrill, J. JG. Van Merriënboer and M. P. Driscoll, Eds. 2007. Available: [http://www.cogtech.usc.edu/recent\\_publications.php](http://www.cogtech.usc.edu/recent_publications.php).
- [5] Jan Maarten Schraagen, Susan F. Chipman, and Valerie L. Shalin, *Cognitive Task Analysis*, Mahwah, NJ: Lawrence Erlbaum Associates, 2000.
- [6] Susan F. Chipman, Jan Maarten Schraagen, and Valerie L. Shalin, "Introduction to Cognitive Task Analysis," in *Cognitive Task Analysis*, J.M. Shraagen, S.F. Chipman, & V.L. Shalin, Eds., Mahwah, NJ: Lawrence Erlbaum Associates, 2000.
- [7] David A. Hofmann, and Lisa M. Jones, "Leadership, Collective Personality and Performance," *Journal of Applied Psychology*, 90, 50-522, 2005.
- [8] Andrew J. Vinchur, Jeffrey S. Schippmann, Fred S. Switzer, III, and Phillip L. Roth, "A Meta-analytic Review of Predictors of Job Performance for Salespeople," *Journal of Applied Psychology*, 83, 586-597, August 1998.
- [9] Robert Hogan, and Robert B. Kaiser "What We Know About Leadership," *Review of General Psychology*, 9, 169-180, 2005.
- [10] Michael J. Stevens and Michael A. Campion, "The knowledge, Skill and Ability Requirements for Teamwork: Implications for Human Resource Management," *Journal of Management*, 20, 503-530, 1994.
- [11] Gary A. Klein, Roberta Calderwood, and Donald MacGregor, "Critical Decision Method for Eliciting Expert Knowledge," *IEEE Transactions on Systems, Man, and Cybernetics*, 19(3), 254-276, May/June 1989.
- [12] ChatMapper. <http://www.chatmapper.com>
- [13] Julia M. Kim, Randall W. Hill, Jr., Paula J. Durlach, H. Chad Lane, Eric Forbell, Mark Core, Stacey Marsella, David Pynadath, and John Hart, "BiLAT: A Game-Based Environment for Practicing Negotiation in a Cultural Context," *International Journal of Artificial Intelligence in Education*, 19(3), 289-308, 2009.
- [14] ActiveMQ. <http://activemq.apache.org/>
- [15] Oleg Alexander, Mike Rogers, William Lambeth, Jen-Yuan Chiang, Wan-Chun Ma, Chuan-Chang Wang, and Paul Debevec, "The Digital Emily Project: Achieving a Photorealistic Digital Actor," *IEEE Computer Graphics & Applications* 30, 20-31, July 2010.
- [16] Wan-Chun Ma, Tim Hawkins, Pieter Peers, Charles-Felix Chabert, Malte Weiss, and Paul Debevec, "Rapid Acquisition of Specular and Diffuse Normal Maps from Polarized Spherical Gradient Illumination," *Eurographics Symposium on Rendering Proceedings*, 2007.
- [17] Cyrus A. Wilson, Abhijeet Ghosh, Pieter Peers, Jen-Yuan Chiang, Jay Busch, and Paul Debevec, "Temporal Upsampling of Performance Geometry Using Photometric Alignment," *ACM Transactions on Graphics*, 29, March 2010.
- [18] Bill Swartout, David Traum, Ron Artstein, Dan Noren, Paul Debevec, Kerry Bronnenkant, Josh Williams, Anton Leuski, Shrikanth Narayanan, Diane Piepol, H. Chad Lane, Jacki Morie, Priti Aggarwal, Matt Liewer, Jen-Yuan Chiang, Jillian Gerten, Selena Chu, and Kyle White, "Ada and Grace: Toward Realistic and Engaging Virtual Museum Guides," *10th Int. Conf. on Intelligent Virtual Agents Proceedings*, September 20-22, 2010.
- [19] Thomas Brox, Andrés Bruhn, Nils Papenberg, and Joachim Weickert, "High Accuracy Optical Flow Estimation Based on a Theory for Warping," *Proceedings of the European Conf. on Computer Vision*, 2004.
- [20] Frederic I. Parke and Keith Waters, *Computer Facial Animation*, Second Edition, Wellesley, A K Peters, 2008.
- [21] Marcus Thiebaut, Andrew N. Marshall, Stacy Marsella, and Marcelo Kallmann, "SmartBody: Behavior Realization for Embodied Conversational Agents," *7th Int. Conf. on Autonomous Agents and Multi-agent Systems Proceedings*, May 12-16, 2008.

[21] Stefan Kopp et al, Towards a Common Framework for Multimodal Generation: The Behavior Markup Language, in 6th International conference on Intelligent Virtual Agents, Marina del Rey, CA, 2006.

[22] Pocket Sphinx. <http://cmusphinx.sourceforge.net>

[23] Anton Leuski and David Traum, "NPCEditor: A tool for building question-answering characters," in Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC), Valletta, Malta, May 2010, 2463-2470. Available: [http://www.lrec-conf.org/proceedings/lrec2010/pdf/660\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2010/pdf/660_Paper.pdf)

## ACKNOWLEDGEMENTS

The project described here has been sponsored by the Office of Naval Research (ONR). Statements and opinions expressed do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

The authors would like to thank Kim LeMasters for his guidance and support; Dan Wright for his subject matter expertise; Nicholas Palmer-Kelly for participating as the face model for the virtual character; Abhijeet Ghosh, Pieter Peers and Jay Busch for assistance with the face scanning process; Matt Liewer, Benny Garcia and Waylon Dobson for processing the resulting data and creating the additional art assets; Grace Benn and Michael Padilla for their artistic efforts; Julia Kim and Todd Richmond for lending the use of their voices; and Glenn Storm whose patience and hard work made this effort possible.

## BIOGRAPHIES



**Dr. Julia Campbell** is a Research Associate at the ICT. She received an M.A. in Communication and Ed.D. (in Education) from USC in 2006 and 2010, respectively. Her dissertation work focused on developing cognitive-task-analysis-supported instruction and self-efficacy surveys for surgical residents and medical students.



ICT in 2004.

**Dr. Mark Core** is Research Scientist at the ICT specializing in artificial intelligence in education. He received his Ph.D. from the University of Rochester in 2000 under the direction of Dr. Lenhart Schubert, and was a Research Fellow at the University of Edinburgh working with Dr. Johanna Moore until joining the



**Dr. Ron Artstein** is a Research Scientist at the ICT, specializing in linguistic data analysis.



**Lindsay Armstrong** is a Project Specialist at ICT. She assists in mixed reality training, scenario development, and instructional design. She supports authoring and documentation efforts for various ICT projects. She earned a Master of Fine Arts in Creative Writing at New Mexico State University in 2007.



where he earned both a bachelor's degree and master's degree.

**Arno Hartholt** is Project Leader of the Integrated Virtual Humans group at ICT and acting Project Director of the central ICT Art Group. He manages technology, art, processes and procedures related to the creation of virtual humans. Hartholt studied computer science at the University of Twente in the Netherlands



organize to produce coordinated movement at the much larger scale of a whole cell.

**Cyrus Wilson** is a Senior Research Associate at ICT. He completed his Ph.D. in Biochemistry at Stanford University in June 2006, where he combined molecular manipulations, video microscopy, and computer-vision techniques to elucidate how simple processes at the molecular scale could



**Dr. Kallirroi Georgila** is a Research Scientist at ICT specializing in spoken dialogue processing. Previously she held research appointments at the Educational Testing Service and the School of Informatics of the University of Edinburgh.





**Dr. Fabrizio Morbini** is a Research Programmer in the natural language group at ICT.

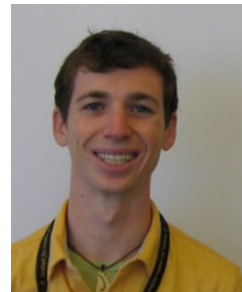
University of Alberta, and has also recently begun Ph.D. studies at the University of California, Riverside.



**Mike Birch** is an ICT Research Programmer working in the area of artificial intelligence in education. He has conducted research in public school classrooms, holds a California teaching credential and received an M.S. in Computer Science from the University of Southern California



**Edward Haynes** is a Senior Program Analyst at the ICT and serves as the project lead for Scalable Visualization. The main focus of his current work at the ICT is the integration of the ICT's AI research and technology with mixed reality virtual environments using game engines.



**Jonathan Bobrow** is Programmer Analyst at the ICT. He holds a BA in Design|Media Arts from UCLA, with a minor in mathematics (2008). His contributions to INOTS include significant effort working on the Instructor Control Panel, Graphical User Interface, and Human-Computer Interactions with a focus on user experience.



**Dr. Richard E. Clark** is Director of USC's Rossier School of Education Center for Cognitive Technology (CCT). His interest is in the design and application of research on complex learning, performance motivation and the use of technology in instruction. His work, "Turning Research Into Results: A Guide to Selecting the Right Performance Solutions," won the 2003 International Society for Performance Improvement Award For Excellence



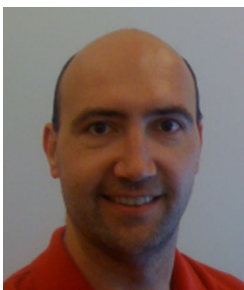
**Dr. H. Chad Lane** is an ICT Research Scientist who conducts research in the areas of intelligent tutoring systems, cognitive science, serious games, and the learning sciences. His research at ICT has focused on the role of feedback in immersive learning environments, both during and after practice. He received his M.S. in Computer Sciences from the University of Wisconsin-Madison and his Ph.D. in Computer Science from the University of Pittsburgh in 2004 under the direction of Kurt VanLehn.



**Dr. Kenneth Yates** is Senior Research Associate at CCT. He develops training and evaluation programs and conducts research in and applications of cognitive task analysis methods to improve human performance, instructional design, and educational technology.



**Jillian Gerten** is a Project Specialist at the ICT. She served five years on active duty in the U.S. Army as a linguist in Military Intelligence. She is a graduate of the Korean language program at the Defense Language Institute, CA, and the Crypto-linguistic course at Goodfellow Airforce Base, TX. During her service, she was promoted to the rank of Sergeant. She received an honorable discharge in 2003.



**Dave Gomboc** is a Research Programmer at ICT who has specialized in designing and implementing intelligent tutoring, explainable AI, and lifelong learning systems, addressing domains such as leadership, negotiation, and cultural awareness. He holds a M.S. in Computing Science from the



*analysis, natural language understanding and generation.*

**Anton Leuski** is a Research Assistant Professor at the Computer Science Department at USC and a Research Scientist at the ICT. His research interests include interactive information access, human-computer interaction, and machine learning. Dr. Leuski's recent work has

and was honorably medically discharged in 2004. He received a Bachelor's of Arts degree in Sociology from the University of California, Los Angeles.



**Timothy Jones** served in the role of Project Manager to support the INOTS effort. He has been instrumental in providing domain knowledge, shaping the user experience, and supporting the implementation of the instructional design. He is currently pursuing his PhD in Film & Television Studies,

School of Arts & Humanities, University of East Anglia, Norwich, UK.



**Dr. David Traum** is a research scientist at ICT, leading the natural language dialogue group, and research faculty in Computer Science at USC. His research is on natural language dialogue between all combinations of human and artificial agents.



concentrations in Cinema-Television and Technology Commercialization.

**Matthew Trimmer** is a Project Director at the ICT. He has directed the creation of several award-winning mixed reality and game-based training projects, and is currently the project leader for INOTS. He received his B.S. and MBA degrees from USC's Marshall School of business 2002 and 2007, respectively, with



(TRADOC) sponsored projects that support the Army's training corps.

**Richard DiNinni** has been a Project Director at the ICT since March 2001. His responsibilities include oversight of technical workshops and military conference participation as part of the institute's Army outreach and transition efforts. Richard also leads a series of US Army Training and Doctrine Command



**Matt Bosack** is currently serving as Project Manager on several ICT projects, including ELECT UrbanSim and the Military Terrain for Games Pipeline, and assisting in the development of the Virtual Officer Leadership Trainer (VOLT). He was formerly Active Duty in the United States Air Force