# Virtual Humans with Secrets: Learning to Detect Verbal Cues to Deception

H. Chad Lane[1], Mike Schneider[1], Stephen W. Michael[2], Justin S. Albrechtsen[2], and Christian A. Meissner[2]

[1]Institute for Creative Technologies
University of Southern California

[2]Department of Psychology
University of Texas at El Paso

**Abstract.** Virtual humans are animated, lifelike characters capable of free-speech and nonverbal interaction with human users. In this paper, we describe the development of two virtual human characters for teaching the skill of deception detection. An accompanying tutoring system provides solicited hints on what to ask during an interview and unsolicited feedback that identifies properties of truthful and deceptive statements uttered by the characters. We present the results of an experiment comparing use of virtual humans with tutoring against a no-interaction (baseline) condition and a didactic condition. The didactic group viewed a slide show consisting of recorded videos along with descriptions of properties of deception and truth-telling. Results revealed that both groups significantly outperformed the no-interaction control group in a binary decision task to identify truth or deception in video statements. No significant differences were found between the training conditions.

**Keywords:** virtual humans, deception detection, intelligent tutoring systems.

## 1 Introduction

Animated pedagogical agents are often designed as tutors [1] or peers [2] in virtual learning environments. In these roles, the agent typically works alongside the learner to solve problems, hold conversations, and provide guidance. Recently, intelligent agents have expanded their roles to become the *object* of practice. That is, it is the interaction with the agent that is intended to be educational. For example, virtual humans [3] have been used to provide practice for intercultural communication [4-5], clinical interviewing [6-7], police officer training [8], and healthy play for children with autism [9], to name only a few examples.

In each of these cases, the virtual human acts as a role player in some social interaction with the learner. The primary goal is to simulate specific communicative patterns (verbal and nonverbal) in realistic ways to give the learner a chance to assess what they see and hear, then respond, all within a social context. Live role playing exercises have a long history in education [10] and there is strong evidence to believe that learners naturally interact with virtual humans as if they are real [11-12]. In this paper, we present the results of a small research project to investigate the use of virtual humans for teaching deception detection. We describe two virtual humans that exhibit common traits of truth-telling and deception along with an accompanying tutoring system for teaching diagnostic cues to deception. We also present results from an initial evaluation of the system's effectiveness.

## 2 Deception detection

How does one detect the difference between the truth and a lie? The ability to detect deception is a critical skill for a number of professions, including school administrators, reporters, therapists, and law enforcement officers. The results of an investigative interview or interrogation of non-cooperative suspects can have profound consequences on society. Unfortunately, research shows that lie detection is extremely difficult, and individuals tend to perform only slightly above chance levels [13]. Even law enforcement officials, who routinely encounter deception in their daily work and receive training in this task, perform similarly at chance levels [14-15].

It is possible, however, for training to improve one's ability to detect deception by helping learners identify the right *cues* on which to focus. Cues to deception are often divided into one of three categories – verbal, nonverbal, and vocal. *Verbal cues* are cues that come from the content of the speaker's statement (e.g., admitted lack of memory, textual embedding, self-references). *Nonverbal cues* can be observed solely by the behavior of the individual (e.g., eye contact, posture, hand/arm movements). Finally, *vocal cues* are behaviors that are related to speech production (e.g., speech hesitations, pitch of voice, response latency). DePaulo et al. [16] conducted a meta-analytic review of over 150 cues hypothesized to be related to deceptive or truthful statements from 116 research studies. Their results identified 23 cues with large effect sizes for significant differences between liars and truth tellers - of these 23 cues, 21 involved verbal or vocal cues to deception (e.g., level of detail, spontaneous corrections, admitted lack of detail, negative statements). On the other hand, many of the nonverbal cues, including those commonly believed by lay persons and law enforcement to be diagnostic (e.g., eye contact, posture, and blinking), were unrelated to deception.

In a meta-analytic review of 11 training studies (20 comparisons) in the deception detection literature, Frank and Feeley [17] determined that overall training showed a minimal effect (4% increase) in improving performance; however, the authors argued that reasons for the weak and inconsistent results may lie within the research designs and stimulus materials (e.g., relevance of the deception detection task, adequacy of the training materials, appropriateness of pre- and post-test). In a more recent meta-analytic review of the training literature, Hauch, Sporer, Michael, and Meissner [18] examined the effect of training on verbal, vocal, and non-verbal cues to deception 22 published and 8 unpublished studies involving N = 3,638 participants. Overall, there was no significant effect of training found for vocal cues ($d$ = .11), while small effects were found for both training on nonverbal cues ($d$ = .18) and the combined training of nonverbal and vocal ($d$ = .21). In contrast, a robust, medium-sized effect was found for training involving verbal cues ($d$ = .62).

**Figure 1.** Virtual humans Victor and Amber can be truthful or deceptive in their responses. Art assets are variants of those used in the virtual patients project [6].

## 3 Deceptive virtual humans

Most training programs for deception detection involve lecture-based seminars, recorded videos, role-playing with peers, and group discussion [19]. A frequently missing element in these approaches is the opportunity for realistic practice for investigative interviewing skills, which provides the context for eliciting and assessing interviewee responses that reveal deception when it is present. To explore the feasibility of using virtual humans for this purpose, two characters were developed to exhibit common traits related to truthfulness and deceit. Character data was augmented with information about deception cues and tutorial feedback. This allowed the creation of a simple tutoring system that supports the learner in asking the right questions and identifying common properties of truth-telling and deceit. In this section, we describe the development of the characters and implementation of the tutoring system.

### 3.1 Character design

The first step in designing characters was to decide on incidents that would provide the backdrop for the interviews. For this, we chose two common (but serious) law enforcement situations: a bombing and a shooting. A character for each scenario was created (see figure 1): Victor, who was in the area during the bombing, and Amber, who witnessed the shooting. Both characters have the ability to be truthful or deceitful in their responses, thus providing four distinct practice opportunities. Both characters use basic nonverbal behaviors in their utterances and idol behaviors, but these are not intelligently modeled as distracters or indicators of truth or deceit in this version of the system.

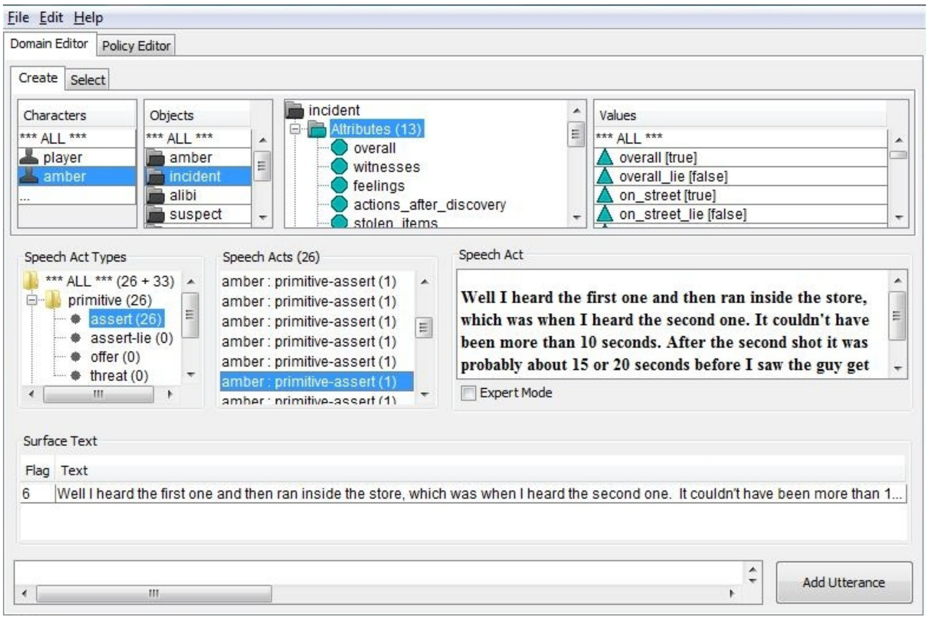


**Figure 2.** Screenshot of the domain editor for Amber.

To build characters we relied primarily on the Tactical Questioning domain editor [20]. For other aspects (i.e., animation, speech generation, nonverbal behaviors) we leveraged the Virtual Human Toolkit.[1] The domain editor requires an author to create a set of domain objects that the character knows about with attributes for each object. Figure 2 shows the authoring tool for Amber: the author created the domain object *incident* with attributes such as *witnesses* and *feelings*. The domain editor automatically generates basic speech acts for these attributes, which correspond to what the character can say about them (and what he or she might want to keep secret, as shown in the upper right corner of the screenshot by the true and false values).

In Victor's domain model includes objects and attributes related to the women's clinic and the bombing; thus, *clinic* as an object, along with an attribute *location*. The system automatically generates dialogue acts that enable him to answer questions such as "Where did the bombing occur?", however, to support robust natural language understanding, authors should provide a variety of surface forms for the questions users may ask. So, other examples for the location attribute would be "Where did the bomb go off?" and "Where did the bombing happen?" At run-time, the system uses a statistical matching algorithm to match the user's questions to those in the character's knowledge base. The matching is done in such a way that exact matches are not required to the set of pre-authored questions [21]. Below we show how the similarity measures from the classifier are used during learner interactions with the characters.

[1] http://vhtoolkit.ict.usc.edu/

**Table 1.** Example statements from Victor and their underlying deception cues.

| Truthful | Verbal cues |
|---|---|
| **T1.** We were walking down Whitman road and I remember hearing a few cars screeching to stop in the middle of the street. There were a fare number of people around. A few started to run towards the building… | cognitive states, level of detail |
| **T2.** Umm, I can't remember all the details of what I have heard… I do remember them talking about "causing damage" a few times, but I could have easily heard wrong. | forthcoming, ordinary imperfections |
| **Deceitful** | |
| **D1.** I think there were a few cars that just stopped in the middle of the road. | uncertainty |
| **D2.** Well I was arrested once for just some minor offense. But, you know, that was just kids getting into mischief. I don't think that that's on my record anymore. | distancing |

Victor and Amber each have two versions of approximately 70 responses available to answer learner questions. They can assume either a truth-telling or deceptive mode for a given session, thus providing 4 different experiences available for use during training. Utterances were carefully crafted to consist of diagnostic verbal cues based on the content. Table 1 shows several examples from Victor's utterance library along with some of the most prominent cues:

- In T1, Victor brings up his state of mind during the event (cognitive states) along with a high level of detail. Both are suggestive of truth telling.
- In T2, he is honest about his lack of memory (forthcoming) and suggests that he may have remembered wrong (ordinary imperfections). Both are common indictors of truth-telling.
- D1 is the deceptive version of T1. "I think" indicates uncertainty, which is often present in deceptive statements.
- Victor distances himself in D2 by suggesting the incident is no longer on his record. Even though not directly related to the event, it still suggests deception.

A primary goal of deception training is to help the learner identify such cues in utterances, and understand what they imply for the deception judgment. Although nonverbal behaviors also play a critical role in deception, both as indicators and distracters (e.g., nervousness is not a reliable sign of deception [16]), our prototype currently uses them to a limited degree.

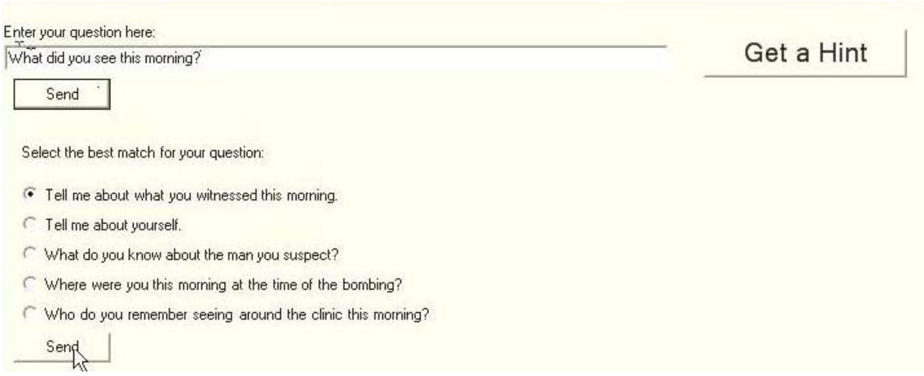


**Figure 3.** Learners type in their question for the character then select the best match.

## 3.2 Interaction

For learners to pose questions of the characters, we chose to use a typed interface along with suggested question matches. Although automated speech recognition and understanding would increase fidelity, our focus was primarily on helping the learner (1) ask the best questions (the skill of investigative interviewing) and (2) identify diagnostic cues (the skill of deception detection). Thus, we decided a typed interface with animated and oral responses from the characters would be sufficient. In addition, to reduce the frustration of asking a question that was not recognized, we provide the learner with the top five matches from the statistical matcher for the student to select (see figure 2). A dialogue history was also maintained in a window below the question asking area.

## 3.3 Guidance and feedback

A simple tutoring system provides help to the learner during their interview with the character. The system responds to help requests from the learner by clicking on the hint button shown in figure 3. It also provides unsolicited feedback on the first occurrence of every deception cue when they are present in character responses (about ¼ of all utterances do not contain cues). As mentioned, the utterances were authored and tagged with their relevant cues (see table 1). In addition, the utterances are also tagged with with phases indicating when particular questions should be asked. When a hint is requested, the current phase of the interview is consulted in order to suggest an appropriate action. The phases are: (1) greetings, (2) background on character, (3) information about the incident, (4) identifying responsible party, and (5) closing. Hints are associated with each phase. Feedback about cues present in character utterances is based on the tags described earlier that describe properties of the utterance and how they relate to the character's veracity. Table 2 shows several examples of tutor messages. All tutor messages are delivered via a pop-up window and they require that the student click to OK to close them.

**Table 2.** Examples of tutor utterances.

| **Tutor hints** | **Relevant Phase** |
|---|---|
| You may want to ask the individual about certain beliefs he has that might give you an idea of how he feels about the incident. | background |
| If he was involved, he probably would have been busy for the last day or so, not just during the time of the explosion. You should ask about this. | alibi |
| **Tutor feedback** | **Relevant Cue** |
| People telling the truth will often use a greater level of detail in their descriptions. | level of detail |
| Admitted lack of memory for details actually occurs more commonly in truthful statements | ordinary imperfections |

# 4 Method

The objective of the current study was to determine the utility of two deception detection training programs that differed with respect to the training approach (rather than the content of the training materials). The performance of these two training approaches was also compared with a no-training control condition. Participants completed both a pre- and post-test assessment of their deception detection performance on Days 1 and 5. Day 3 involved interaction with the training program for those in the training conditions, or completion of an innocuous (irrelevant) task for those in the control condition. Given research on the validity of various cues to deception [16], training focused on the most valid verbal and vocal cues to deception.

## 4.1 Participants

One hundred and five undergraduate psychology students from the University of Texas at El Paso participated in exchange for credit in their introductory psychology courses.

## 4.2 Design

The experiment employed a 3 x 2 mixed factor design in which participants were randomly assigned to one of two training conditions or a no-training control condition (between-subject factor, N = 35 per condition), while all participants completed both a pre- and post-test assessing their deception detection accuracy (repeated measure).

**Videotaped stimuli.** Videotaped alibi statements were collected by the authors for the purposes of developing a pre- and post-test measure of deception detection accuracy. Individuals were randomly assigned to conditions in which they were instructed to provide either a truthful or deceptive account of their whereabouts three nights prior to the interview. Participants were interviewed on video regarding their statement. Interviews ranged from approximately one and a half minutes to seven minutes ($M$ = 2.47 minutes). Forty-two videos (21 truthful and 21 deceitful) were collected in total. Videos were then pilot tested by a separate group of participants (N = 18), and 20 videos were selected for use in the pre- and post-test (10 deceitful, 10 truthful). The pre- and post-test stimuli were shown to have equivalent discriminability, and presentation of the stimuli were randomized and counterbalanced across participants.

**Training programs.** To contrast with the Virtual Human based system described above (VHuman), a non-interactive, didactic presentation was created (Didactic). This program involved training on the same cues to deception, but presented the information through a less immersive format. In the presentation, each cue to deception is explained via both text and auditory information. Following each explanation, a video example of the cue is shown. Statements were taken from the scenarios developed for the VHuman condition to provide comparable exposure. The presentation of cues was automated so that time on task would be comparable across participants. Participants in the VHuman condition were also presented with an abbreviated didactic presentation prior to interviewing the virtual characters that provided participants with a very basic introduction to the various cues to deception (excluding any video depictions of the cues). Learning and exposure to the cues were then subsequently reinforced by the VHuman system as each participant interacted with both characters who were randomly assigned to lie or tell the truth regarding their given scenario. Time on task for both the Didactic and VHuman training conditions were comparable (i.e., approx. 40 min).

## 4.2 Procedure

The experiment was conducted across three sessions involving a pre-test, a training session, and a post-test, with sessions separated by 48 hours. The pre- and post-test sessions involved presentation of 10 videotaped statements (described previously) for which participants had to determine veracity (i.e., truth vs. lie) and provide a confidence estimate (i.e., 0 to 100% scale). Following each deception detection task, participants also completed a questionnaire assessing their knowledge of the relative value of various cues to deception – this measure served as an assessment of content learning for the two training conditions. Participants were randomly assigned to one of the two training conditions, or completed an innocuous filler task (i.e., participants completed an unrelated face recognition study in which they studied a series of faces and were later tested in their recognition performance) in the control condition.

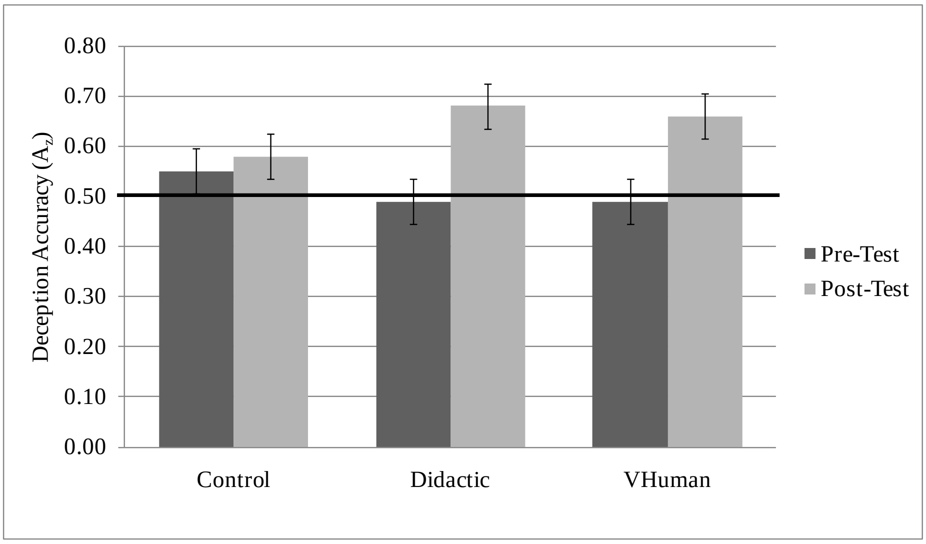


**Figure 4.** Deception detection performance ($A_z$) as a function of training condition across pre-test and post-test. Bolded line represents "chance" performance.

# 5 Results

A measure of discrimination accuracy was computed via signal detection theory [22]. Specifically, $A_z$ was computed via the following formula:

$$A_z = (d'/\sqrt{2})$$

$A_z$ provides an estimate of discrimination accuracy that ranges from 0 to 1, with .50 being equivalent to chance performance in the present study. $d'$ is computed as the difference between the Z-score for the "hit" (i.e., proportion of deceptive statements correctly identified) and "false alarm" (i.e., proportion of truthful statements incorrectly identified as deceptive) estimates. Results of the experiment are displayed in Figure 3. A 3 x 2 mixed Analysis of Variance (ANOVA) was used to assess the influence of training across the pre- and post-tests. A significant interaction was observed, $F(2,102) = 3.31$, $p < .05$, $\eta_p^2 = 0.06$. Follow-up pairwise comparisons indicated that both the Didactic, $t(34) = 3.88$, $p < .001$, $d = 0.88$, and the VHuman, $t(34) = 3.37$, $p < .01$, $d = 0.65$, training conditions significantly improved detection performance. In contrast, the control condition showed no significant learning effect across the pre- and post-test, $t(34) = 0.39$, *ns*., $d = 0.08$.

Participants estimates of confidence were similarly analyzed using a 3 x 2 mixed ANOVA, resulting in a significant interaction, $F(2,102) = 3.37$, $p < .05$, $\eta_p^2 = 0.06$. Pairwise comparisons demonstrated that both the Didactic, $t(34) = 2.10$, $p < .05$, $d =$

0.27, and the VHuman, $t(34) = 4.98$, $p < .001$, $d = 0.66$, conditions significantly increased their confidence as a function of training, while the control condition showed no significant effect, $t(34) = 1.19$, *ns*., $d = 0.13$.

Finally, we also assessed participants' learning of the validity of various cues to deception as a function of training condition. A 3 x 2 mixed ANOVA demonstrated a significant interaction, $F(2,102) = 18.28$, $p < .001$, $\eta_p^2 = 0.26$. Once again, pairwise comparisons showed that both the Didactic, $t(34) = 8.08$, $p < .001$, $d = 1.95$, and the VHuman, $t(34) = 7.34$, $p < .001$, $d = 1.83$, conditions significantly improved their knowledge of diagnostic cues to deception, while the control condition showed no significant learning effect, $t(34) = 1.78$, *ns*., $d = 0.29$.

## 6 Conclusions and discussion

Based on these results, it appears that both forms of training had an equivalently positive effect on ability to detect deception when compared to a baseline control (with no training). This suggests that teaching learners about cues is an effective method for enhancing their ability to detect deception. It is also suggests there is no apparent value of interactive practice with feedback for the skill of detecting deception in recorded statements.

It is common practice in deception studies to use recorded statements as pre- and post-tests. They are passive experiences that are easy to repeat and compare since participants simply watch the video and make a decision. In this study, the pre- and post-tests are similar in structure to the didactic condition (which consisted of a 40 minute presentation on cues to deception with recorded examples), just without the instructional content. This may explain why the didactic condition was sufficient to achieve a significant learning gain. In many ways, the didactic condition has higher fidelity than the VHuman condition since recorded statements show real people, using intonation, facial expressions, and so on. Even though virtual humans also simulate these aspects of human communication, in our prototype, they were identical between truth-tellers and deceptive versions of Victor and Amber. Given this, our results can even be considered positive since virtual human-based training was able to produce equivalent learning to a very strong didactic condition closely aligned to the test.

A weakness of using recorded videos is that they only tap recognition skill; they do not evaluate a learner's ability to conduct an investigative interview. This represents a difference between the training conditions in our study. Specifically, learners who interacted with virtual humans were required to generate questions for the characters (assuming they didn't game the system). They also received hints on what to ask about, if requested. If a positive difference is to be found between conditions, it may be revealed in having learners conduct an independent interview.

The virtual characters and tutoring system were developed in about 4 months, and so there are many opportunities for improving and extending it. Perhaps most importantly, as virtual human technology matures, it will become easier to simulate human behavior with higher fidelity which will enable our system to address a greater range of novice misconceptions. For example, nervousness is commonly interpreted as a sign of deception when it is, in fact, not a reliable indicator [16]. A nervous, but truthful, virtual human would provide an interactive example to demonstrate that there are often many causes of nervousness, such as being asked questions. Beyond this, Victor and Amber fall into the category of question / answer virtual humans, meaning they do not possess a realistic model of the interaction with the user beyond the simple phase markers used by the tutor (see section 3.2). Although sufficient for some learning goals, to tap more deeply into investigative interview process, learning would likely benefit from models of emotions, proxemics, and consequences of what is being said (some of these aspects appear in other virtual human research [3, 12]).

In summary, virtual human technology is in the early stages of being applied to the problem of teaching social interaction skills. Our prototype system and evaluation suggests a virtual human-based system can increase learners' deception detection skills. The question of whether it is necessary, and if there are other benefits beyond detection, depends on new measures of investigative interviewing skill and richer models of virtual human behavior.

## Acknowledgements

The project or effort described here has been sponsored by the U.S. Army Research, Development, and Engineering Command (RDECOM). 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. We extend special thanks to Patrick Kenny for sharing art assets from the virtual patient project at ICT and technical support from Sudeep Gandhe and Jina Lee.

## References

1. Johnson, W.L., Rickel, J., Lester, J.C.: Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments. Int. J. Art. Int. in Ed. 11, 47--48 (2000)
2. Kim, Y., Baylor, A.: A Social-Cognitive Framework for Pedagogical Agents as Learning Companions. Ed. Tech. Res. and Dev. 54, 569--596 (2006)
3. Swartout, W., Gratch, J., Hill, R.W., Hovy, E., Marsella, S., Rickel, J., Traum, D.: Toward virtual humans. AI Magazine 27, 96--108 (2006)
4. Kim, J.M., Hill, R.W., Durlach, P.J., Lane, H.C., Forbell, E., Core, M., Marsella, S., Pynadath, D.V., Hart, J.: BiLAT: A Game-based environment for practicing negotiation in a cultural context. Int. J. Art. Int. in Ed. (in press)
5. Johnson, W.L., Valente, A.: Tactical language and culture training systems: Using artificial intelligence to teach foreign languages and cultures. Proc. 20th Nat. Conf. on Inn. App. Art. Int., vol. 3, pp. 1632--1639, AAAI Press, Chicago (2008)
6. Kenny, P., Parsons, T., Gratch, J., Rizzo, A.: Virtual humans for assisted health care. Proc. 1st Int. Conf. on Pervasive Technology Related to Assistive Environments 1--4, ACM Press, New York (2008)

7. Johnsen, K., Raij, A., Stevens, A., Lind, D.S., Lok, B.: The validity of a virtual human experience for interpersonal skills education. Proceedings SIGCHI Conf. on Human Factors in Comp. Sys. 1049--1058, ACM Press, New York (2007)
8. Hubal, R.C., Frank, G.A., Guinn, C.I.: Lessons learned in modeling schizophrenic and depressed responsive virtual humans for training. Proc. 8th Int. Conf. on Int. User Interfaces. 58--92, ACM Press, New York (2003)
9. Tartaro, A., Cassell, J.: Playing with virtual peers: bootstrapping contingent discourse in children with autism. Proc. 8th Int. Conf. Learning Sciences, vol. 2, 382--389 (2008)
10. Kane, P.E.: Role playing for educational use. Comm. Ed. 11, 320--323 (1964)
11. Reeves, B., Nass, C.: The media equation: How people treat computers, television, and new media like real people and places. Cambridge University Press (1996)
12. Gratch, J., Wang, N., Gerten, J., Fast, E., Duffy, R.: Creating rapport with virtual agents. Proc. 7th Int. Conf. on Intelligent Virtual Agents, LNCS, vol. 4722, pp. 125--138. Springer, Heidelberg (2007)
13. Bond, C.F., DePaulo, B.M.: Accuracy of deception judgments. Personality & Social Psychology Review 10, 214--234 (2006)
14. Meissner, C.A., Kassin, S.M.: He's guilty!: Investigator bias in judments of truth and deception. Law & Human Behavior 26**,** 469--480 (2002)
15. Kassin, S.M., Meissner, C.A., Norwick, R.J.: "I'd know a false confession if I saw one": A comparative study of college students and police investigators. Law & Human Behavior 29, 211--227 (2005)
16. DePaulo, B.M., Lindsay, J.J., Malone, B.E., Muhlenbruck, L., Charlton, K., Cooper, H.: Cues to deception. Psychological Bulletin 129, 74--118 (2003)
17. Frank, M.G., Feeley, T.H.: To catch a liar: Challenges for research in lie detection training. J. Applied Comm. Res. 31, 58--75 (2003)
18. Hauch, V., Sporer, S.L., Michael, S.W., Meissner, C.A.: Does training improve detection of deception? A meta-analysis. Unpublished manuscript (2009)
19. Colwell, L.H., Miller, H.A., Lyons, P.M., Jr., Miller, R.S.: The Training of Law Enforcement Officers in Detecting Deception: A Survey of Current Practices and Suggestions for Improving Accuracy. Police Quarterly 9, 275--290 (2006)
20. Gandhe, S., DuVault, D., Roque, A., Martinovski, B., Artstein, R., Leuski, A., Gerten, J., Traum, D.: From domain specification to virtual humans: An integrated approach to authoring tactical questioning characters. Interspeech (2008)
21. Leuski, A., Kennedy, B., Patel, R., Traum, D.: Asking questions to limited domain virtual characters: How good does speech recognition have to be? Proc 25th Army Science Conf. (2006)
22. Wickens, T.: Elementary Signal Detection Theory. Oxford University Press, New York (2001)