THE SIMULATION OF A HUMAN SUBJECT FOR INTERPERSONAL SKILL TRAINING

Dale E. Olsen and
William A. Sellers
The Johns Hopkins University Applied Physics Laboratory
11100 Johns Hopkins Road
Laurel, MD 20723-6099

Robert Garland Phillips, Jr.
Federal Bureau of Investigation
FBI Academy
Quantico, VA 22135

Abstract

We have developed a unique, PC-based training tool that emulates human behavior using a computer-simulated person in a realistic scenario. The tool was developed for the FBI to help agents develop interview skills by providing meaningful experience in detecting deception during interviews. The self-paced, multimedia courseware enhances learning while delivering an effective, relatively low-cost, interactive experience. The interview-training module gives the trainee experience in asking proper questions and distinguishing between deceptive and truthful responses. It also provides a critique and numerical score for the interview. As their skills develop, law enforcement students can see their critiques improve and their scores rise. Since the implementation of the simulation within the FBI, new agent trainees have increased their interviewing practice.

This type of training tool supplements and reinforces traditional classroom instruction by giving the trainee an opportunity to practice. The technology could also be used to help develop a large range of interpersonal abilities including investigative skills and peer-pressure resistance skills taught as part of a drug prevention program.

Introduction

Government and industry spend heavily to design and develop sophisticated simulations so trainees can obtain ample practice and acquire experience without risking lives or expensive equipment. Pilots practice on in-flight simulators before flying aircraft; military personnel use war-gaming simulators to practice executing missions; medical personnel use computer simulations to practice triage as part of their training. For a wide range of skills, practice is required for the training to be effective. Training simulation technology has progressed to the point where it can successfully be used to help develop a variety of interpersonal skills.

Use of Simulations to Teach Interpersonal Skills

Simulators are needed to train people to deal with social and behavioral issues and situations. To maximize their effectiveness, they must provide an engaging environment where the student can experience various realistic situations and provide different responses. These simulators must rely on the recent developments in modeling and simulation, sociology, psychology, and other fields.

The system described in this paper, designed for use on a multimedia PC, was developed to teach FBI agents to detect deception during a series of practice interviews with Mike Simmen, a simulated person, by watching and listening for what are sometimes subtle clues revealed through his speech and body language. The interviews are engaging partly because Mike is unpredictable. Records show that students voluntarily practice for hours, and it is believed that they develop valuable experience and skills beyond those gained from lectures. This training experience will prove useful in many criminal investigations.

Detection-of-Deception Training

For years, law enforcement agents have used verbal and nonverbal clues to detect deception. Much of the original research showing the validity of the techniques was done by Reid and Associates (Reference 1) and perhaps is most readily available as part of course work provided by their company. More recently, Paul Ekman (References 2

and 3), Stan Walters (Reference 4), and many others have made significant contributions to that body of knowledge. The techniques taught using Mike Simmen were carefully designed to match those taught at the FBI academy.

Interviewing Mike Simmen

In the training scenario, the FBI student/agent interviews Mike Simmen about a crime to determine Mike's involvement. He may or may not have committed the crime. The student conducts the interview by selecting from an extensive, scripted list of questions and observing the subject's verbal and nonverbal responses. A stochastic model of the subject selects responses to the student's questions based on logical and emotional factors associated with human behavior. Whereas the subject's behavior and responses are determined by a computer model of Mike's brain, the visual and audible responses are presented by video sequences using an actor. This provides a realistic, two-way conversational interview.

While carefully listening to and observing the subject's responses, the student must also plan a line of questioning and judge the content as truthful, deceptive, or not informative. Because most questions have several possible responses and the simulated subject may or may not be guilty during a simulation, the interview proceeds differently each time it is conducted. Like a real interview, the simulated interview is expected to take over an hour.

Modeling[1] of Mike's Behavior

The model for Mike's behavior was developed with specific attributes that help the user develop solid interview skills. Mike "remembers" the nature of the interviewer's questions and statements and responds

[1] The model of the brain is not based on scientific research. The model was constructed specifically to meet the needs of this project and has proven successful in that endeavor. The terminology used is describing the model was chosen to convey ideas and may not be consistent with that used within certain disciplines.

based on typical behavior patterns related to his guilt or innocence and the content of the interview. His "brain" has both logical and emotional components. The logical component tracks the responses and keeps them reasonable and consistent. It selects one of a series of likely responses considering the question and circumstances, which are affected by Mike's status (e.g., guilt or innocence) and his emotional state.

Mike's emotional component is critical in the selection of the response to a question. His emotional state is determined primarily by the trainee's questions. A mathematical model determines the flow of Mike's emotions as questions are asked and responses are given. The model parameters can be easily adjusted to affect his personality. For example, the emotion model can be tuned so that Mike may be forgiving of a poorly worded question, or may be easily upset and slow to forgive.

Logical Component

The logical component of Mike's brain is saved in a database. It contains all available questions and all possible responses to those questions. Different questions can share responses, and each question can have many responses, so that data fields written into the database are used to link questions and responses.

There are over 500 possible questions that can be asked in almost any order. To reduce the search for the desired questions, those that have been asked are eliminated from the list. Also, similar questions are deleted. As replies are provided, new information is revealed and new questions become relevant and available. Fields in the database (logical component) are used to identify which questions and replies are opened and closed as a result of each question and Mike's replies.

One particular difficulty is to make sure Mike replies consistently. For example, suppose Mike is asked if he likes his supervisor and he replies, "She's okay." If the next question is "Do you ever socialize with her?" the reply "No, I can't stand her." must be dropped from the

set of possible responses. In other situations, a different or inconsistent reply is required. For example, if Mike is asked, "What do you like to do in your spare time?" he may reply that he plays golf where the possible replies are that he plays golf, reads, or skis. If he is then asked, "What else do you like to do?" the golf reply is a consistent reply, but must be dropped. To accommodate all of the possibilities, a complex set of data fields is made available to the database developer who is known as the scriptwriter.

Emotional Component

Mike will be in any one of five emotional states. The states are five levels of rapport, including worst possible, poor, average, good, and excellent. As rapport deteriorates, Mike provides short uninformative replies, while with better rapport, Mike provides more complete and revealing replies. For another application, the states could be anger, denial, depression, bargaining, and acceptance and the subject would respond accordingly. Mike moves through his emotional states as a result of the interactions with the student. Each question is coded by the scriptwriter in accordance with its effect on Mike's emotional state. The code may be rapport-dependent, so that a question may be a good choice when rapport is good or excellent and a bad choice when rapport is poor or the worst.

The Stimulating Question

The stimulus of a question or statement affects the flow of emotions. The rapport code in the database is used to compute the stimulus value and ultimately to determine how Mike's emotions change. When rapport is good, a single awkward question most often will not destroy it, but a series of offensive questions will. The rapport code is entered in a spreadsheet by the scriptwriter as a number between 0 and 9. This value is first converted into a rapport value between –5 and +5. Negative values represent poor question selection. After each question, the stimulus value, S_r, is computed using the average rapport value and the current rapport value, S_q,as follows*:*

$$S_r = 0.8\,[\text{memory} * S_{r-1} + (1 - \text{memory}) * S_q]\ (1) + 0.2\ (\text{average rapport value}).$$

In this equation, S_{r-1} is the stimulus before the last question. The memory quantity in the equation is a parameter that can be tuned to change the behavior of the subject and is nominally set to 0.45. This parameter controls the influence of the last question on the stimulus. The stimulus is influenced most by the value of the last question, but the most recent history and the entire history also influence it.

When the interview is initiated, quantitative emotional values are allocated to each of the five rapport states, and these values are constrained to add up to 1. The questions act as stimuli to affect the flow of these emotional values from state to state. The model of the emotional flow is complicated, yet can easily be modified to accommodate changing requirements. For Mike, the model modifies itself as the interview progresses. For example, each time you irritate him (rapport state deteriorates), he is progressively more easily irritated.

Mike's state is the rapport state with the largest emotion value. The emotion model must perform two fundamental functions: it must determine the direction of emotion flow as well as the amount of the flow. The stimulus S_r, is the input to this model.

As questions are asked, the emotions will flow towards a target rapport from all other rapport states until either the limit for that state is reached or the sign of S_r is changed. If good questions are being asked, S_r will be positive, and when the flow reaches the limit, the next highest rapport state will normally become the target state. Flows of emotion toward an improved rapport are positive and those toward a weaker rapport are negative. The sign of S_r determines the direction of the flow. If the sign changes as questions are asked, the direction of the flow immediately changes and a new target rapport is selected. The target rapport selected is usually the next higher or lower state depending on the sign of S_r, but that is not a requirement. The selection is determined using transition probabilities and occasionally

another state is selected. Thus, poor questions usually cause rapport with Mike to deteriorate smoothly, but occasionally it turns sour suddenly.

The memory constant described above is only one of several parameters that allow Mike's personality to be tuned. The limit before a transition to another rapport state is set separately for each state. These state-related parameters are used to make it difficult to move away from the worst rapport level and easy to slip from others.

Another parameter that makes it difficult to leave the worst level is the "stickiness" parameter. This parameter determines the portion of a rapport state's emotion that is allowed to flow from that source to a target state. A sticky state yields less emotion. The forward and backward speed parameters also help determine the amount of emotion flowing from a given state. These speed factors provide ways for the user to regulate the speed of the flow of emotions that affects all states equally. Since rapport is slow to build and can be destroyed rapidly with poor questions, the forward speed (improved rapport) is much slower than the backward speed (deteriorating rapport). More detail is provided in Reference 5.

Logical and Emotional Component Interface

When the interview begins, Mike is randomly assigned a number of initial conditions and these initial values are used to select responses from the database. He may be guilty or innocent. If he is guilty, he could be motivated by need or hate. His initial rapport with the student is also randomly assigned with most of the emotion being assigned to the poor and average states. The selection of his guilt status affects his behavior throughout the interview. The student must identify and improve the rapport while determining his guilt.

Mike's replies are selected using his rapport state, his guilty status, and a pseudo-random number generator. For each guilt status and each rapport state, interface fields in the logical database provide likelihoods of each response. These likelihoods are numbers between

0 and 9 and are used to compute probabilities associated with the available responses. When a question is asked, the likelihoods for all available replies for Mike's current state and status are summed. Each likelihood is divided by the sum to produce a series of probabilities for the available replies. These probabilities and a pseudo-random number generator are used to select Mike's response.

Interview Evaluation

Since the interview can be conducted in many different ways, and Mike's behavior is different for each interview, percentage scores or traditional evaluations are not used. To make comparisons, the students can compare their score to their own previous evaluations or to the high score of another student using that computer. The students accumulate points by making good decisions. The points are awarded based on a number of criteria, reflecting the objectives of the training.

An important objective of this training module is for the student to develop rapport with the subject by creating an environment where he or she is comfortable and therefore likely to provide informative and reliable responses. Statements or questions that make Mike comfortable or less threatened contribute to rapport building and generate positive rapport ratings. Those that make him feel defensive or offended contribute to a negative rapport. However, some of the questions that make Mike feel uncomfortable may help the investigator determine if the subject is attempting deception and therefore provide important diagnostic information. Therefore, questions are also evaluated according to their diagnostic value. The diagnostic ratings of the questions are stored in the database and used in evaluating the interview.

Other factors that are used to evaluate the interview include correct determination of the subject's status (truthful or deceptive) and the detection of clues. Some of Mike's replies reveal attempts to deceive and should be identified as an indicator of deception. Other replies are rare for a deceptive person and indicate that Mike is truthful. In a typical interview, both types of clues will be observed, but most of the

clues will reveal the truth. Points are awarded for the proper identification of all clues.

Software Options

The program has four levels of difficulty: beginner, intermediate, advanced, or professional (fewer clues are provided at the more challenging levels). When a question is selected, it is read before Mike responds, reinforcing the question and allowing the student to observe any delays in Mike's answers. Another option allows the user to have the questions read either using a female or a male voice. The system stores the sequence of questions and responses from the interview so that the entire interview can be replayed and reexamined. During replay, the system identifies and records any unusual behavior in the subject.

Results of the FBI's Use of the Simulation

A major factor prompting the FBI to support the development of the Michael Simmen simulation was the need to provide additional interviewing practice for new agent trainees during their training program. The challenge was to provide this opportunity within the 16-week training program and without reducing other curriculum areas. What was needed was a meaningful exercise that would be compelling enough that the agents would voluntarily practice on their own time. The paragraphs below summarize the experience to date of the use of the simulation within the FBI.

Starting with fiscal year 1999, the Michael Simmen simulation was made an integral component of the training provided to all FBI new agents. Trainees are introduced to the simulation during their first week of training. Their use of the simulation to hone their interviewing skills is voluntary beyond the completion of two

interviews of Michael Simmen, which are assigned as ungraded homework assignments.

Although the two assignments could be completed in less than half an hour, thorough and thoughtful execution of each interview takes about an hour, for a total of two hours of work. Systematic tracking of new agent use of the Michael Simmen simulation shows that trainees spend, on the average, between five and six hours interviewing Michael Simmen. Thus, the trainees, on average, voluntarily spend three to four hours longer than required engaged in the simulated interviews. Since the FBI trained over 600 new agents during fiscal year 1999, agents received more than 3,000 additional hours of meaningful interviewing practice. Over 1,800 of these practice hours were completed voluntarily.

Another measure of utilization of the Michael Simmen simulation is the number of simulated interviews conducted by new agents. The trainees are required to conduct two interviews with Michael Simmen. During fiscal year 1999, new agents conducted an average of seven interviews. Thus, the challenge of providing a meaningful practice opportunity compelling enough that new agents would practice on their own time has been met.

Changes in agent trainee performance on the Michael Simmen simulation from the first homework assignment to the second homework assignment were tracked. Among the measures of trainee effectiveness in interviewing Michael Simmen are the change in trainees' scores on the overall exercise, the percentage of clues indicating deception/truthfulness correctly identified, and the extent to which Michael Simmen's truthfulness or deception was correctly assessed. For trainees during fiscal year 1999, the results were as follows:

Measure	Percent Improvement First to Second Interview
Interview score	16
Clues correctly identified	30
Deceptive/truthful interviews correctly assessed	16

Note that there is no intention to attribute any improvement in interviewing effectiveness simply to the use of the simulation. The simulation was intended from the outset to be an integral part of a well-established regimen including classroom instruction, readings, role plays, etc., designed to provide the most effective possible set of experiences in support of the development of effective interviewing skills. The purpose of the simulation was to provide an inviting opportunity for additional practice outside of the formal classroom experience.

As of the writing of this paper, the Michael Simmen simulation package has been provided to each of the FBI's field offices where it is being made available for FBI employees who have already graduated from the academy. Some of the more seasoned agents included in this group graduated from the academy prior to the time when nonverbal indicators of deception became part of the curriculum. For these agents, the availability of the Michael Simmen simulation provides an opportunity to augment their existing set of skills with an additional tool to assist them in conducting effective interviews. Dissemination of the simulation to other federal law enforcement agencies as well as state and local law enforcement agencies is under way.

Preparing for Future Developments

Many possible futuristic upgrades to the system have been discussed. Our goal for the next generation of systems is to develop successful systems by not pushing the technology to the point where the number of users will be limited or the system will be too complex. This means

that we will continue to target PCs, perhaps with DVD readers, but we are not planning on targeting higher-end computers typically used for simulations.

Several new training systems are in the early planning stages. To prepare for their development, we determined that the most difficult part of the project will be in developing the database for the logical part of the brain. This part of our effort was the most time-consuming and will continue to be the most difficult part of the development of future systems.

Authoring Software

To apply what we have learned about developing the logic database, an authoring software system was written to assist future developers in several ways. It will prompt the scriptwriter for questions, replies, and all the special codes. It automatically checks for missing data and detects some types of inconsistencies. Perhaps its most useful feature is its ability to help trace through the free-form logic paths of questions and responses.

Speech Recognition

As PCs become faster and speech recognition improves, it appears that the ability to use speech recognition systems in the next generation of PCs could be practical. To verify this possibility, we are now using the Dragon Naturally Speaking Software Developers Kit ActiveX components and a 450-MHz Pentium II PC to ask questions in a special version of the training system. The student verbally asks Mike Simmen a question from the given list further reinforcing the learning of appropriate questions.

The speech recognition component converts the read question into a text string. It was decided that it would be easier to implement the software if it were treated as an information retrieval problem and not a natural language processing problem. Standard information retrieval techniques were applied to the text string to determine the best match

from the list of available questions. The text string is parsed into individual words and a weighting factor is determined for each word in the question. The weighting factor W_f is based on the following formula:

$$W_f = \text{Log}(N_{\text{total}} / N_{\text{word}}) \quad (2)$$

where N_{total} is the total number of unique words in the list of all questions and N_{word} is the number of times an individual word appears. If a match is not found in the weighting-factors list, a weighting factor of zero is assigned. A score is calculated for each question in the list of available questions. The question with the highest score is selected as the question to ask Mike.

As a result of this implementation, questions can be asked using somewhat different words than those suggested by the program. Mike usually provides the correct response. As technology develops, the next step we may take beyond speech recognition is to incorporate language understanding. With this capability, the student will be able to ask questions of the simulated subject using completely different words and the system will be able to relate the actual question to one anticipated in the database of responses and then will provide an appropriate response.

Conclusion

This type of training tool has a broad range of potential applications. The sophistication of the simulation and programming make the system a viable candidate for many kinds of training that aim to increase our understanding and awareness of human interaction. The basic approach developed for the FBI could be used for training by any investigative agencies that need to use techniques that detect deception. A few obvious examples are employment interviews, security investigations, legal testimony, and insurance claims

investigations. The computer-based, interactive approach used here also has applications in situations other than detecting deception. It could be used effectively to help teach people how to resist peer pressure and other influences.

Author Biographies

Dale E. Olsen received a A.B. in mathematics from California State University, Chico, in 1965 and an M.S. and Ph.D. in statistics from Oregon State University in 1969 and 1973, respectively. He joined The Johns Hopkins University Applied Physics Laboratory in 1973 and was responsible for developing techniques to evaluate the reliability and accuracy of submarine-launched ballistic missiles. Most recently, Dr. Olsen has jointly developed an automated scoring system for polygraph data. This new system is accurate and objective and is being widely used by law enforcement agencies.

William A. Sellers is a Senior Professional Staff Physicist in APL's Strategic Systems Department. He obtained a B.S. degree in physics from the University of Maryland in 1978 and an M.S. degree in computer science from the Johns Hopkins University in 1985. From 1978 to 1980, he was a laser tracking system analyst with Bendix Field Engineering Corporation. Since joining APL in 1981, he has worked at designing and developing data analysis and simulation tools for GPS missile tracking systems, NEAR spacecraft ground systems, digital short-range communications systems, an informatics system for the pain management doctors at the JHU School of Medicine, and an FBI interview training simulation system.

R. Garland Phillips received a B.S. in business/transportation economics from the American University in 1968 and an M.A. in education from the George Mason College of the University of Virginia in 1972. From 1971 to 1982, while with the U.S. Office of Personnel Management, he developed models for assessing training needs and for determining the economic return-on-investment of training for public sector employees. Since joining the FBI in 1982, he

has conducted program evaluation studies and has worked to utilize new technologies to enhance organizational productivity. Mr. Phillips currently manages the FBI Academy's Office of Information and Learning Resources.

References

1. Reid, John E., *The Reid Technique of Interviewing and Interrogation*, John E. Reid and Associates, Chicago, 1991.
2. Walters, Stan B., *Principles of Kinesic Interview and Interrogation*, CRC Press, Boca Raton, FL, 1995.
3. Ekman, Paul, *Telling Lies: Clues to Deceit in the Marketplace, Politics and Marriage*, W. N. Norton and Co., Inc., New York, 1985.
4. Ekman, Paul, "Who Can Catch a Liar?" *American Psychologist*, 46, 913–920, 1991.
5. Olsen, Dale E., "Interview and Interrogation Training Using A Computer-Simulated Subject," 19th Interservice/Industrial Training, Simulation and Education Conference, Orlando, FL, December 1-4, 1997.