INTERVIEW AND INTERROGATION TRAINING USING A COMPUTER-SIMULATED SUBJECT

Dale E. Olsen
The Johns Hopkins University Applied Physics Laboratory
Johns Hopkins Road
Laurel, Maryland 20723-6099
e-mail: dale.olsen@jhuapl.edu

ABSTRACT

New and experienced law enforcement agents, among others, need extensive training in techniques for interviewing and interrogation. However, it is often difficult for the student to practice those techniques before putting them to use in investigative work. Practice interviews using actors are not always realistic and are too expensive. As a result, interactive, multimedia software that involves a simulated subject has been created to help trainees develop their interview and interrogation techniques using personal computers. Although the trainee must select questions from a predetermined list, the available questions are many. Users are required to observe both verbal and nonverbal behavior and to make well-reasoned decisions. Chances to make errors are presented at every decision point. The simulated subject responds differently each time the system is used and will sometimes be deceptive or truthful. Even the truthful subject will react to questions and show signs of deception. The goal of the trainee is to navigate the subject through different behavioral states and then determine if the subject is truthful. (The FBI is funding this program.)

AUTHOR BIOGRAPHY

Dale E. Olsen received an 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 made responsible for developing techniques to evaluate the reliability and accuracy of submarine-launched ballistic missiles. Dr. Olsen has led several methodology and software development efforts. One effort to model the reliability of reentry bodies as a function of operational parameters successfully identified important mission parameters and accurately predicted system performance under then-untested conditions. Another effort to build accuracy models using the output of Kalman filters enabled the detection of small but important system modeling errors. Most recently, Dr. Olsen has jointly developed an automated scoring system for polygraph data. This new system is accurate and objective and is being used by law enforcement agencies. His discriminate/detection analysis capability is currently being applied to other problems as well. He is a member of the American Statistical Association, Biometrics Society, American Polygraph Association, and Maryland Polygraph Association.

OVERVIEW

Proper use of interview and interrogation techniques can help correctly resolve many criminal cases and significantly reduce investigative and judicial costs. However, it is difficult for the student to practice applying such techniques before attempting to use them in investigative work. In an interview, the investigator must listen to response content carefully, identify changes in the voice and speaking style, watch nonverbal behavior, and develop the next question, while simultaneously formulating a question strategy based on continually changing subject inputs. Our intent is to use a computer simulation to provide the student a means to practice the skills needed to effectively apply important interview and interrogation techniques.

Training normally requires instructor time and in many cases travel expenses, making the entire process for many federal law enforcement agents and thousands of state and local agents very expensive. Even when funding is available for paid actor-subjects, the experience is necessarily limited. Most often the practice comes only with cases involving real criminal offenses. Although a completely realistic computer-simulated interview is not possible, we can simulate human behavior in an investigative interview in such a way as to provide significant training value. To this end, we are developing interactive, multimedia software for interview and interrogation training purposes.

Our long-term plan is to develop the software in stages, ending in a rich variety of training scenarios, using state-of-the-art technology to provide basic lessons through experience in applying interview and interrogation techniques that are as realistic as possible. Our initial effort, which we describe here, focuses on developing software for training in the fundamental steps of interviewing. We use a series of recorded responses to about 400 available questions. These responses have been carefully planned and directed.

The interviewer attempts to build rapport with the subject while using diagnostic questions to determine if he or she is deceptive. The student should learn to (1) build rapport while maintaining proper professionalism, (2) listen carefully to verbal clues, and (3) detect important changes in both verbal and nonverbal behavior.

The most significant part of the training system focuses on experience with a single case, where the student works through the steps of an interview. At each step, the student is given an opportunity to make mistakes. Each time the system is used to practice, the simulated subject provides different responses, sometimes indicating truthful behavior and at other times subtly indicating deception motivated either by revenge or financial need.

Although our goal is to simulate a real interview, the possible questions and responses must be limited to those planned as the script was developed. What appears to be a serious limitation, however, is not as restrictive as it may seem. First, there is a limited set of standard questions that provide important diagnostic information. These questions are included in the script. Second, the script offers a reasonably rich variety of questions, giving the student the opportunity to practice question formulation. Third, even though the questions are limited, there will still be hundreds available, making it possible to provide paths representing many realistic interviews. Finally, responses will depend on how well the student has laid the foundation for the questions, making rapport development an essential part of the successful simulated interview.

There is, of course, a lack of spontaneity and realism in the simulated interview. The student must somehow input or select the desired question while the simulated subject waits. However, the delay does give the trainee time to think and to develop better habits.

As the interview proceeds, the student learns to determine when it is appropriate to ask certain questions while watching and listening for indicative responses. If and when the student feels that the interviewee is the likely perpetrator of the crime or will provide no new information, the student can terminate the interview or suggest that the interrogation begin. At that time, the student is asked to fill out an on-screen questionnaire that forces the user to make a decision on the question of the subject's honesty.

SYSTEM OPERATION

When the executable program is started, the user can select one of two options: the on-line manual, which can be called up and read, or the interview. The on-line reference manual includes a word-search capability and a few short portions of video-recorded interviews to illustrate certain points. The interview is designed to parallel the manual as much as possible. For example, the simulated subject acts out many different behaviors described in the manual. If the interview is selected, the student is presented two more options: (1) “check background” to read a short case history and (2) “interview subject” to start the interview.

The interview option allows the user to ask questions by selecting them in any of a number of ways. The student may enter a key word, which brings up a list of questions. Or the student may type a question, which prompts a list of other questions that most nearly match it. In addition, the user may page through a list of hundreds of available questions. Sometimes the subject may ask the student a question, in which case the student may choose to ignore the question or respond using a special window that can only be opened in special situations. The student may also elect to make supportive statements, helping to build rapport. These statements are made available throughout the question list at appropriate times. A critical part of the system is to provide easily available bad questions. Poor student performance ratings and information restrictions will result from the use of bad questions.

The student will be investigating the theft of $43,000 from an Automatic Teller Machine (ATM) at a bank. The subject of the interview will be a male loan officer who had the opportunity to take the money.

The subject is initially assigned any one of three guilt states: innocent, guilty and motivated by revenge, or guilty and motivated by financial pressure (states 1, 2, and 3, respectively). The investigator will have several possibilities to explore, including revenge for being passed over for promotion, financial pressure caused by a series of events, problems with alcohol, problems with drugs, or problems with a girlfriend. The innocent subject also has the motives of revenge and financial problems, but demonstrates a different pattern of responses to critical questions.

The system remembers the sequence of questions and responses from the interview so that when the interview is ended, the user can replay it. During replay, the system stops at the end of each response. Deceptive behaviors, if any, are identified in text form. For example, a change in voice or an evasive answer may be noted. By design, some of the behaviors are subtle and a few are misleading.

SYSTEM DEVELOPMENT

The Storyboard

A text-only storyboard was written and used for testing. It demonstrates screen formats planned and the integration of the modeling. Although questions were available and the system performed in a way similar to the final product, no video or audio was used. Rather than providing audio and video displays when questions were asked, text appeared that provided the verbal response and a list of behaviors to be acted out.

The Simulated Subject

Again, our goal is to make the simulated subject behave realistically and unpredictability, while allowing the student to ask a wide variety of questions. A distinction is made between anger directed at the interviewer and anger directed at some outside institution or person who has “done” our subject wrong. Anger at the interviewer will always be considered a negative, whereas anger at the outside person can provide the opportunity for bonding by making the subject feel that the investigator truly understands. ***The anger at the outside object is the anger in the subject mood. The anger at the investigator is represented by poor rapport.***

A subject may give good information inadvertently because of anger, but most often because the rapport is so good at that point. The script includes a few “slips” that will rarely occur, which requires that the user listen carefully to pick up cues. If the student hears the slip, he or she may choose to end the interview, since the required information will have been obtained.

The Script

The interview script consists of all available questions and responses. Both verbal and nonverbal behaviors are described for each of over 2000 responses. When an interview is initiated, only a limited number of questions made available to the student make sense, some because specific information has not yet been revealed and others because they may imply information that would not realistically be available. These questions are not made accessible to the user until the appropriate information is revealed. As certain information comes to light, some questions may no longer make sense or the available responses may not make any sense.

Codes

Perhaps the most interesting feature of the system is the modeling codes contained in the script. For each question, there is a list of key words, a question code, a mood value, a rapport value, and an information value (Table 1).

Table 1 Sample of a Scripted Question.

Question/Statement/ Response	Question Code	Mood Value	Information Value	Rapport Value	Key words
What do you think should happen to whoever took the missing money?	37448	3 4 4	9	4	Diagnostic, Punishment, Need help

The **question code** is used by the software to identify the question and the corresponding responses (Table 1). The responses given to each question depend on the mood (anger, denial, or compliance) of the subject and the rapport between the subject and the student. As a result, each question is rated separately, depending on the way it affects the mood and the rapport. The mood **value** is used to update the mood of the subject. The student's goal is to make the subject's mood compliant. In this mood, the subject gives the most informative responses. This score is affected by the entire question history and is used to filter out question responses. The mood score drives the subject mood and affects the mood value for a particular question (see the section on Modeling). For each question, three ordered mood values are given, one for each possible mood (Table 1). Some questions may be good for a compliant subject but inappropriate for a subject showing signs of anger. As questions are asked, the mood values (Table 1) are used to continually update the **mood score**.

Certain questions or comments improve rapport, even while being of little value in providing information or improving the subject's mood. A cumulative **rapport score** is developed using individual question **rapport values** (0 to 9). When this score is low, the subject will provide short answers with much less information. When it is high, the subject is more likely to provide more valuable information.

Some questions may hurt rapport or irritate the subject, but will provide useful information. Therefore, each question is also rated according to its **information value**.

The question, rapport, and information value scores are all used to help evaluate the student's performance. If the student consistently asks good questions, that information is combined with other information to produce a high overall evaluation of the student's technique.

As stated previously, for each question or comment, there is a series of possible responses (Table 2). Each response requires 12 pieces of information or codes. For many of the codes, "1" is used to indicate that a subject in the corresponding state could make this reply, and "0" indicates that this reply would never be made by a person in that state. The ordered set of 0's and 1's indicates the state the subject

must be in to provide the reply. For the guilt code, the order is (1) innocent, (2) guilty for revenge, (3) guilty for financial reasons. If only an innocent person would give a particular reply, then the code would be 100, with the single "1" indicating that this reply can be used for the innocent subject and the two 0's indicating that this reply would not be given by either type of guilty subject.

Table 2 Sample of a Scripted Response.

						Likelihood Codes					
						Guilty			Innocent		
Reply Code	Response	Guilt	Mood	Rapport	Cluster	Anger	Denial	Compliance	Anger	Denial	Compliance
37448-10 35378-20	The person should go to jail for a long time.	111	101	0111	H4, E6, A9, L1	2	2	1	9	8	8

The reply codes identify the questions for which a given response can be used. The two codes in Table 2 indicate that this reply can be used in response to two questions (37448-10 and 35378-20).

The guilt status (states 1, 2, or 3) for which the reply can be used is also identified. The code (111) in Table 2 indicates that the response can be given by any subject.

The mood codes identify the mood of the subject, given a particular response. The code shown in Table 2 (101) indicates that this response can be given by an angry or compliant subject, but not by the subject in denial.

Examples of behavioral and cluster codes are given in Table 3 and indicate how the response is delivered. They include codes for voice quality, voice pitch, and clarity of speech (behavioral) as well as head, eye, hand, arm, and leg movements (cluster). Simple movements are described with a few codes. More complex behavior is likely to represent a behavioral pattern and may be repeated. Special cluster codes are used to describe these behaviors.

Table 3 Sample of Clusters Codes.

Behavioral Codes	Cluster Codes	Description of Behavior
H4. E6, A9, L1	C1	Chin projected, knitted "V" over nose, crossed legs (barrier)
H7, E9, A2, L9	C2	Jaw flexed, eyes "hard," finger pointed, feet hidden (under chair)
E9, A5, L1	C3	Eyes "hard," "L" thumb under chin, finger to eye, crossed legs (barrier)
H9, E4, A11, L3	C4	Head tilted, zero blink rate, arms crossed, fist(s) clenched, "Fonz" attitude (slouched, arrogant)

Some cluster codes may indicate stress or habit, which will often emerge when the subject is in a particular mood. For example, a subject in denial may nonverbally express stress by tucking his chin, maintaining extended eye contact, and covering his mouth. His speech may slow or his voice soften. If these behaviors only occur when certain topics are discussed, then the investigator should be aware that such topics are sensitive and may require more discussion.

During the rapport-building portion of the interview, the investigator is able to observe the baseline behavior. This behavior includes two stress-like clusters, which vary randomly among

three different pairs. Without proper rapport building, the investigator cannot identify the clusters of behavior that have no significance.

The rapport levels are used to filter out some of the available responses. The levels include:

1. Very bad: Subject will provide little information.
2. Bad: Subject can be changed with work. Most responses are brief.
3. Neutral: Subject's demeanor does not change; responds with a mix of both vague and informative answers.
4. Good: Subject has some trust and respect for the investigator and is generally forthcoming.
5. Excellent: Subject sometimes volunteers information. Relationship is warm and friendly.

The rapport code shown in Table 2 (0111) indicates that the responses will not be given with a level 1 rapport. The likely answer to the question in Table 1 would be the statement "Don't know."

The last six codes in Table 2 (Anger-Guilty through Compliance-Innocent) are likelihood codes, which are used to determine how likely the response is, based on the mood and guilt status of the subject. Each of these six columns includes a number between 0 and 9. If the subject is guilty and in denial, and the response is common for a subject in that state, that response will be given a 7, 8, or 9. If the response would be rare, that response would be given a 1, 2, or 3. For the example shown in Table 2, the response is more likely to be given by an innocent subject.

The software will examine all the possible responses to the question given the current state of the subject. Some responses may be eliminated because they are inappropriate for the guilt status, mood, or rapport level. Of the remaining responses, the likelihood codes are summed and then each is divided by that sum to obtain a probability for each response. A random number will be used to select the response.

MODELING

The complete modeling of a subject's behavior is complex. A flowchart of our simplified information processing is shown in Figure 1. The system must keep track of the guilt status, subject mood, and rapport level as well as any limitation on the questions available. The guilt status is selected using a random number and remains constant throughout the execution of the program. The initial mood is also selected randomly, with probabilities dependent on

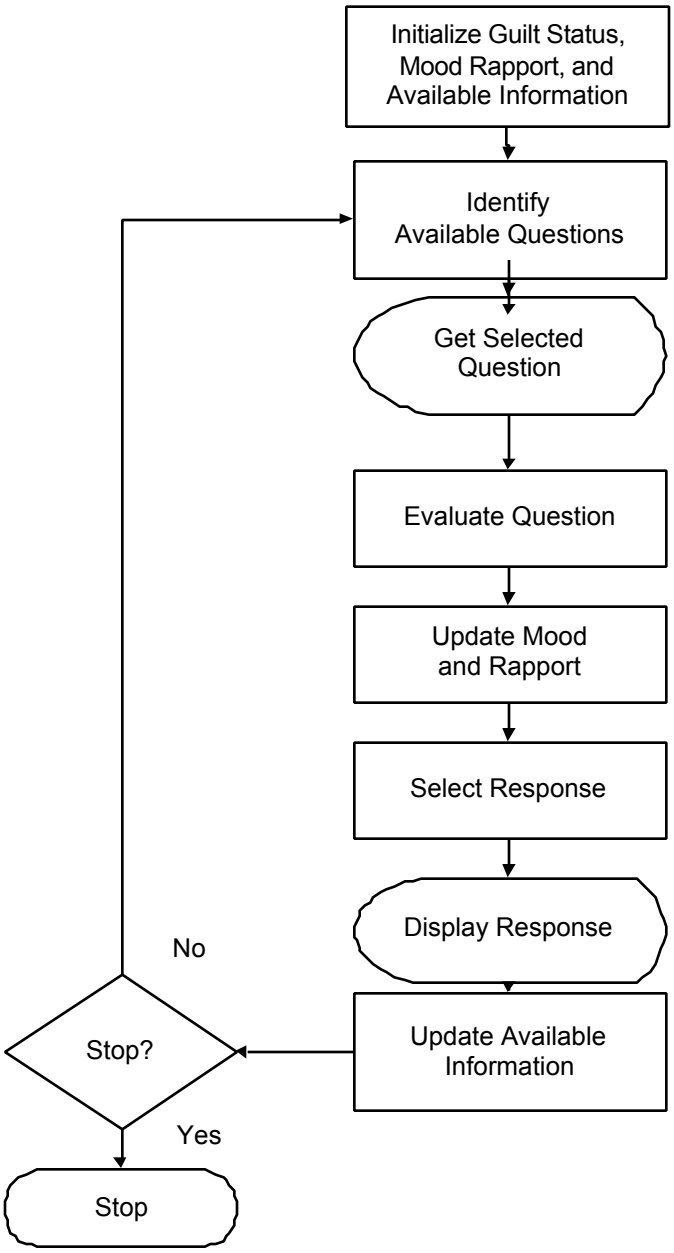

Figure 1 Logic Flowchart.

the subject's guilt status (e.g., an innocent person is more likely to be compliant). The rapport level will most often initially be 3, but occasionally will be 2 or 4. The issue of how the

mood and rapport states change is at the heart of the modeling. To explain this, we closely examine the more complex mood modeling in the following paragraphs.

The computer-simulated subject is presented with a series of stimuli in the form of questions that will affect mood. A model for using the scripted question values to produce the mood designators has been developed and implemented as part of this program. The logic for this model is contained in Figure 1 as part of "Update Mood and Rapport."

The mood value of the questions depends on the mood of the subject. For example, when in an environment where the subject is ready to confess (compliant), a certain question may be appropriate. However, that same question may diminish rapport and make the subject angry if it is asked too early in the interview. The script provides a quantitative value for each question and each mood. The model is needed to address mood changes as a result of the stimuli of questions. The inputs for this model are the current mood vector for the subject, the quantified mood value of the question, and the mood score representing the question history. These quantities will be explained. The outputs of the model are the updated values of these same quantities.

Recall that the subject will be in one of three states or moods: anger, denial, or compliance. Although one mood may dominate, elements of other moods are often present. Therefore, the moods are modeled using three numbers that add to 1 and measure the degree of the mood at any one moment during the interview. The model operates on these three numbers that in combination are called the mood vector. The largest element in the vector is the mood of the subject. For example, the vector

$$(P_{anger}, P_{denial}, P_{compliance})$$

could be

$$(0.30, 0.60, 0.10),$$

implying that the subject was primarily in denial with a value of 0.60, but with anger also present at a value of 0.30. Since the elements of the mood vector add to 1, they will be referred to as the mood probabilities. In fact, these numbers may or may not be used as probabilities.

The Mood Scores

The stimulus is received as a mood value or a number between 0 and 9. This value is first converted into a score between –5 and +5. Negative scores represent poor question selection. After each question, the mood value is also used to compute: total score, which is the sum of all the scores up to that time; average mood score, which represents the total score divided by the number of questions asked; and mood score S_M, which is used in determining the mood change and is computed as follows:

$$S_M = 0.8[\text{memory}*S_{M-1} + (1 - \text{memory})*S_q] + 0.2(\text{average question value}) .$$

In this equation, S_q is the current question value and S_{M-1} is the weighted score before the update. The memory quantity is a parameter that can be tuned to change the behavior of the subject and is nominally set to 0.45. It controls the influence of the last question on the mood score. The mood score is influenced most by the value of the last question, but the most recent history and the entire history also influence it.

Probability Flow in the Mood Vector

The mood model must perform two fundamental functions: it must determine the direction of probability flow as well as the amount of probability flow.

Direction of Probability Flow . The mood probabilities will always flow to a target mood from all other moods. The direction of the flow changes when the target mood changes. Changes from anger to denial, from denial to compliance, or from anger to compliance are considered positive. One goal of the investigator is to get the subject as far into compliance as possible. The probabilities can also flow in the opposite direction, producing a negative flow and ultimately an angry subject or a subject in strong denial. The investigator is not likely to get valuable information or a confession from a subject in either of these moods. Good questions produce a positive mood score and will most often produce a positive flow of the probabilities. If the probability is flowing toward denial, as long as the mood score remains positive, the probability flow will continue into that mood until that state reaches a limit set for

that subject for denial. Once the limit for that mood is reached, the forward transition probabilities are used to select the next target mood. These transition probabilities provide a method for selecting the next target mood and, with a positive mood score, will most often select a target mood to produce a positive flow.

Poor questions will produce negative mood scores that will almost always result in probabilities flowing to a less desirable mood. As long as the mood score is negative, the probability will continue to flow into a mood until a limit is reached, and then the next mood is selected using the backward transition probabilities. These transition probabilities will almost always produce a backward flow of the mood probabilities. If the score is too negative, the system will stop and indicate that the student has exceeded any reasonable anger level and will not get any more information from the subject.

As previously discussed, probability is always flowing toward a single target mood until the mood reaches the designated limit. When the limit is reached, the target mood will change. The target mood will also change if the sign of the mood score changes. Good questions will produce a positive mood score, causing the probability to flow into a mood. However, a short series of poor questions will cause the mood score to change to a negative value, and the subject will immediately start to move to a less desirable mood. A change in sign of the mood score does not require a mood limit to be reached before a new target mood is selected.

Negative mood scores work in a similar way. As long as the scores remain negative, the subject will work through states until the limit is reached. Then a new state is selected using the backward transition probabilities. If the mood score turns positive, a new mood is immediately selected on the basis of the positive transition probabilities.

The Limit. The limit-before-a-change or limit value can be adjusted or tuned to change the personality of the subject. For some, anger may be a state that is rarely seen and that would not last long if it were seen. For this type of personality, the limit would be set low (0.60), so the signs of anger would last only for a short period.

If poor questions are asked, a subject in denial who has previously been angered by other bad questions is more apt to get angry again. Each time the target mood changes because of a change in the sign of the mood score, it is assumed that the subject will more easily move out of that mood again. When a mood change occurs owing to a change in the sign of the mood score, the limit associated with that mood is diminished by 10% until it drops below 0.50. The subject's moods will become more volatile as the subject is stimulated by wildly varying question quality.

Transition Probabilities. Transition probabilities are used to determine the next mood. With a positive mood score, forward transition probabilities are used; with a negative mood score, backward transition probabilities are used. If a subject is in denial, has reached the limit, and the mood score is positive, then a new mood is selected. The forward transition probabilities for that state might be (anger, denial, compliance) = (0.25, 0.00, 0. 0.75). To select the next mood, a random number between 0 and 1 is generated. If it is less than 0.25, the transition is to anger. If the number is between 0.25 and 1.00, the next state is compliance.

A special feature of the model is that the transition probabilities are dynamically changing. A subject who has moved into denial several times because of poorly selected questions will more often return to denial. The same is true for anger. Whenever the subject moves to either of these moods as a result of a change to a negative mood score, the corresponding transition probability is modified to increase the probability that the subject repeats the same mood change.

Amount of Probability Flow. The four factors affecting the amount of the probability flowing from any mood to the target mood are question value, speed, stickiness, and amount of the probability available to move. The first three quantities are used to compute a factor F_i for each mood *i*. The factor times the mood probability is used to determine how much probability flows from that mood to the target mood. For example, if the target mood is denial (mood 2) and the mood anger (mood 1) has 0.5 probability, then $(0.5)F_1$ would be subtracted from anger and added to denial. It is now necessary to determine how to compute F_i to

provide needed flexibility to model different personalities.

Since the question scores are between –5 and +5, they are divided by 10 and the absolute value is taken to obtain a question factor Q, which is between 0.0 and 0.5. There are two speed factors, one for forward flow R_1 and one for backward flow R_2. These factors are between 0 and 2, so that after speed is used in the computation of F_i, the QR_j is between 0 and 1. Finally, the stickiness factor S_i is used to individualize the various moods by specifying how difficult it is to take probability from a particular mood. Its value is between 0 and 1. The update equation, $F_i = (1 - S_i)QR_j$, is used to compute the final factor.

The mood model was implemented and tested independently of the main software. Many of the tuning parameters (e.g., memory, transition probabilities, limits, stickiness) were adjusted to make the behavior realistic.

CONCLUSIONS

Much of the knowledge of the behavior of a subject being interviewed or interrogated is available only through experience. A few books have been written on the subject,[1–4] including one by a subject-matter expert supporting this program, S. B. Walters. We know of no rigorous mathematical behavioral models which would be suitable for implementation in this format. The primarily knowledge base is contained in the minds of the people with experience in the field. Our efforts are not to validate any technique or even to represent our models as valid, but only to present available beliefs and information in a fashion that can more easily be learned.

Training systems reviewed by the author use relatively simple branching logic to provide interactivity. The student selects one of a few options, and based on that selection, the software makes a deterministic decision. To properly teach interview techniques, the student selects from one of a large number of options at each point in the interview and, based on that selection, the software makes a stochastic decision.

The system must model some aspects of human behavior. To pick those aspects and to write a script, the expertise of those writing and practicing in the field have been utilized. The resulting multimedia system makes the training of new FBI agents and others in using interview techniques more effective by making them sensitive to behavior known to be typical of deceptive or truthful subjects.

REFERENCES

1) Walters, Stan B., *Principles of Kinesic Interview and Interrogation*, CRC Press, Boca Raton, FL, 1995.

2) Ekman, Paul, *Telling Lies: Clues to Deceit in the Marketplace, Politics and Marriage,* W. N. Norton and Co., Inc., New York, 1985.

3) Ekman, Paul, “Who Can Catch a Liar?” *American Psychologist,* **46**, 913–920, 1991.

4) Reid, John E., *The Reid Technique of Interviewing and Interrogation,* John E. Reid and Associates, Chicago, 1991.