Building an Angry Grandmother

Kevin Dill
Lockheed Martin Global Training and Logistics
35 Corporate Drive, Suite 250
Burlington, MA 01803
kevin.dill@lmco.com

Oliver Dreger
SAIC Corporation
10260 Campus Point Drive
San Diego, CA 92121
oliver.b.dreger@SAIC.com

***ABSTRACT**: Historically, the Infantry Immersion Trainer has relied heavily on human role-players to serve as the occupants of a simulated village located in Iraq or Afghanistan. While this works for many situations, it is expensive. Additionally, there are some vital roles and situations that aren't easily portrayed with live role-players (such as those of children or the elderly).*

The facility currently includes support for the use of mixed reality characters, which are displayed on life-sized screens in some areas of the village. These characters have been designed primarily for use in kinetic "shoot or no-shoot" scenarios requiring relatively short interactions with the trainees. Under the Future Immersive Training Environment Joint Capability Technology Demonstration, one goal was to develop technology which could be used to create characters who can force the trainees to make a variety of complex decisions in non-kinetic situations, as part of scenarios that last for an extended period of time (10 minutes or more). These characters need to be visually and behaviorally plausible, culturally appropriate, and responsive to the actual situation and the actions of the trainees. They should be able to act autonomously, but also be responsive to operator control.

Our test character portrays an elderly Afghan woman who is the grandmother of an insurgent. She is built around a modular artificial intelligence which works in conjunction with improved artwork, a more complex animation system, instrumentation of the room and the trainees, and direction from the Instructor/Operator Station. She can not only react appropriately to the actions of the Marines, but can also respond to questioning and engage in limited conversation (in her native language of Pashto), all while attempting to minimize the visible repetition of behavior. She has not only succeeded in forcing the trainees to make difficult decisions, but also leaves us well positioned to incorporate further improvements (such as more detailed instrumentation of the room, or a social/psychological model of the village state) in future characters.

1. Introduction

Infantry immersion training has come into prominence in recent years due to its perceived effectiveness in preparing Marines and soldiers for modern combat operations. This training is conducted at a variety of locations, including the Infantry Immersion Trainer (IIT) at Marine Corps Base Camp Pendleton, California.

Since its inception in 2007, the IIT has relied predominantly on human role-players to occupy its training area - a simulated Iraqi or Afghan village. This reliance on roleplayers has not been entirely satisfactory. While roleplayers have proven necessary across most training contexts, their use requires expenditures of thousands of dollars each day to cover their administrative, recruiting, salary, and supervisory support costs. Minimizing these costs through increased use of technology has become an important focus in recent years.

To be successful, such technology must accomplish key training objectives while saving costs, without significant loss in the trainees' feeling of immersion within the training environment. In addition, there is hope that the consistent reproducibility of technological training experiences will prove superior to the widely varying quality seen with human roleplayers. Finally, there are certain vital roles, such

as those of children or the elderly, that can't easily be portrayed with live role-players.

The IIT currently supports the use of mixed reality characters, which are displayed on life-sized screens in some areas of the village. The Artificial Intelligence (AI) for these characters is heavily scripted, making them most appropriate for use in kinetic "shoot or no-shoot" scenarios requiring relatively short interactions with the trainees.

The overall goal of the Future Immersive Training Environment (FITE) Joint Capability Technology Demonstration (JCTD) was to demonstrate that "by simulating the immersive conditions of the combat environment… we can improve tactical and ethical decision-making by small unit leaders, increase small unit operational effectiveness, improve individual resiliency, and reduce casualties." [1] Under this JCTD we developed a series of five training scenarios which were used for live training at the IIT. These scenarios focused on non-kinetic training, and employed a broad array of new technology which, among other things, reduced the reliance on live role-players.

This paper focuses on the fourth scenario in the series. In this scenario, the trainees are forced to change their tactical task on the fly, after receiving credible intelligence that an insurgent stages regular illicit meetings in his home at the edge of the village. Tasked or acting on their own initiative, the trainees (with Afghan soldiers attached) isolate and then enter the home with the intention of detaining any enemy forces present and conducting a search of the home. Upon entry, they discover that the only occupants present are virtual characters: an elderly woman and her juvenile grandchild. The grandmother is hostile, but not violent, complicating their mission within the home.

The high level goal for the Angry Grandmother character (Granny) was to introduce trainees to a noncooperative character within the immersive training environment who would challenge their interpersonal skills (through the medium of an attached interpreter), force them to consider relevant cultural training, test their knowledge of detention criteria and tactical questioning procedures, create non-kinetic complications during the conduct of a house search, and stress their judgment regarding when and how to contact higher headquarters in an ambiguous situation. An elderly woman and a child were ideally suited for this role, putting the trainees in an ambiguous situation that blends a known hostile location with belligerent, but obviously nonthreatening, occupants. Due to age limitations within the roleplayer staff, these ideal roles could not have been portrayed age appropriately by live role-players.

In order to meet the above goals, we needed to create a character who could maintain the Marines' immersion in the scenario for an extended period of time (10 minutes or more). This character needed to be visually and behaviorally plausible, culturally appropriate, and responsive to the actual situation and the actions of the trainees in real time. It needed to act in such a way as to create the desired training experience. Finally, it needed to be able to act autonomously, but also to remain responsive to operator control when necessary.

The character that we built makes use of a modular AI [2] which works in conjunction with improved artwork, a more complex animation system, instrumentation of the room and the trainees, and direction from the Instructor/Operator Station (IOS). Our character can not only react appropriately to the actions of the Marines, but can also respond to questioning (in her native language of Pashto), all while attempting to minimize the visible repetition of behavior. She has not only succeeded in forcing the trainees to make difficult decisions, but also leaves us well positioned to incorporate further improvements (such as more detailed instrumentation of the room, or a social/psychological model of the village state) in future characters.

The remainder of this paper will discuss Granny's hardware and art asset requirements, our approach to cultural authenticity, the instrumentation and operator controls that guide her actions, and the AI and animation systems that tie it all together into an immersive performance. Finally, we will discuss areas for improvement and other possible future work.

2. Audio/Visual Fidelity

Our first task was to make our characters look and sound real. This requires high quality art assets such as high polygon models with multiple textures, motion captured animation, and professionally recorded audio. In addition, we used a high end projector and painted the screen bright white, rather than projecting onto a tan background, in order to maximize visual quality.

We used DARPA's RealWorld Ground Client [3] (RealWorld) to drive the simulation. RealWorld provides top-quality rendering, animation, and sound engines.

Previous efforts simply projected the characters on the wall, with a piece of furniture (such as a table or bed) in front of them. The furniture allowed the projector to be set closer to the wall, since the image didn't have to

Figure 1: The Angry Grandmother in her nook.

go all the way to the floor, and also physically blocked the Marines from getting too close, making it less likely that a Marine would step between the projector and the screen, creating a shadow in the image.

We chose a different approach for this effort. Instead of a conventional projector and furniture, we used a short throw projector, which can be placed very close to the screen (in our case, 31 inches back from the screen, and 8½ feet above the floor). This allows us to project an 8' x 6' image on the wall, which goes all the way down to the floor, with little chance of obstruction.

Without the physical furniture present in front of the screen to give perspective, we found that simply projecting the characters on a white screen made them seem to float in space. In addition, the screen stood out, since it was much brighter than the surrounding wall. To improve this, we created a virtual backdrop, shown in Figure 1, which displays a nook, roughly 3' deep. This creates a small virtual space for the characters to stand in. We decorated this space with a stove, rug, and wall hanging. The models for these props, as well as the textures for the floor and walls, were taken from actual objects in the village, so their look and feel matched that of the rest of the IIT.

3. Cultural Authenticity

Although we had a predefined scenario, within the bounds of that scenario we wanted Granny to be as culturally authentic as possible. That is, not only did we want her to speak Pashto, the language most spoken in the represented region, but we wanted her choice of words and phrases, her gestures, and her general attitude and demeanor to be appropriate for a native Afghan.

Toward this end, both our voice actress and our motion capture actress were Afghans. Our motion capture was done with one of the role-players working at the IIT. For the voice acting, we didn't provide a strict script, but instead provided a general idea of the types of lines that we wanted, a list of suggested possibilities, and a

Figure 2: Instructor/Operator controls.

lengthy description of the situation, and then had our actress ad-lib lines that she felt were appropriate.

4. Instructor/Operator Control

One important goal for the Angry Grandmother was to create a character which could operate autonomously. The exercise controllers, referred to in this paper as instructor/operators or just operators, are typically quite busy. With 13 Marines moving through the village at a time, the instructor/operators not only manage the exercise but also role-play the platoon commander and adjacent squads communicating over radios, as well as preparing the After Action Report (AAR). With all of these responsibilities and distractions, if we relied entirely on human control it would be easy for Granny to fail to react at a critical moment.

On the other hand, humans are much better at picking up subtle cues and selecting an appropriate response. Particularly in a physical simulation, where the AI's knowledge of the world is largely limited to tracking the positions of the Marines, and where we would like to be able to respond to spoken utterances, it is highly beneficial to keep a human in the loop.

We addressed this dichotomy by creating a character that was capable of acting autonomously, and then providing the operator with an array of commands that they could use to control Granny's actions. These commands are executed immediately, interrupting the current line of dialog if necessary, giving precise control over the timing of Granny's responses.

Commands can be sent to Granny using the Instructor/Operator Station (IOS), which is a central control station that supports all of the actions necessary to successfully guide the execution of a training exercise. Using the IOS, a single instructor/operator can control the execution of the training.

Figure 2 shows the commands that were used during the JCTD, although our data-driven, modular design makes it relatively easy to add and remove commands.

The operation of the commands is fairly simple. The IOS uses High Level Architecture (HLA) network messages [4] to control the states of the various virtual and facility-based systems, and to monitor the movement and actions of the trainees. In the case of Granny's AI, the IOS uses HLA interactions to set the values of certain variables (referred to as the "brain variables") within the simulation. The AI monitors these values, and takes the appropriate action if it sees one change. The system diagram in Figure 3 shows the major interactions of FITE systems which influenced Granny, including the Agile FOM (Federation Object Model) Interface (AFI) Brain Variable Requests and Brain Variable Responses.

Commands fall into three categories: behavioral states, situational responses, and single lines of dialog. In addition, there are a number of other systems which can affect Granny's performance, such as the shot tracking and live character tracking systems.

4.1. Behavioral States

The Angry Grandmother has five behavioral states:

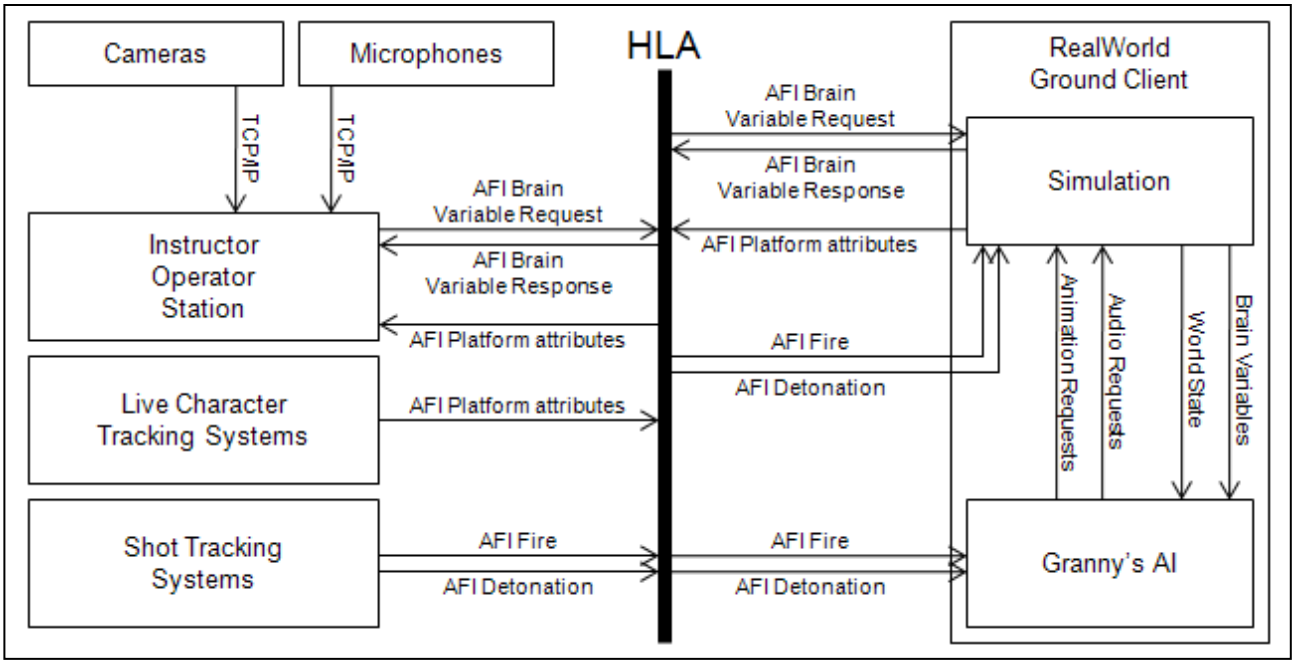

Figure 3: System diagram.

- **Stand and Listen**: Granny stands quietly, as if listening to what the Marines are saying. This state is used before the Marines enter and at the end of the scenario when they are questioning her.
- **Rant**: Granny berates the Marines in an attempt to get them to leave. This state is used while the Marines search the room.
- **Inconsolable**: Granny cries inconsolably, beseeching Allah to help her poor grandson. This state is used if the Marines find contraband.
- **Surrender**: Granny places her hands on her head and kneels on the floor. This state is used if the Marines try to arrest her.
- **Dead**: Although this state cannot be selected from the IOS, if the Marines should shoot Granny then she will die.

Once Granny is placed in a particular state, she stays there until told otherwise. Thus, if you place her in the Rant state, for example, and then hit any of the buttons for situational responses or single lines of dialog, she will execute the action associated with that button and then go back to ranting.

4.2. Situational Responses

There are a few specific situations which often occur while the Marines are going through the scenario. Each of these can occur multiple times, and we don't want Granny to do the same thing repeatedly, so we created a variety of responses. When the button is pressed, the AI selects an appropriate response (in some instances at random, in others using a predefined sequence).

These situations are as follows:

- **Marines Knock**: if the Marines knock on Granny's door, or if they are standing outside getting ready to go in for an extended period, this button will cause Granny to call out to them. Example utterances include: "go away, don't bother us," "my husband isn't home," and "may God destroy you, do not come in!"
- **Marines Talk to Granny**: Sometimes the Marines will try to talk to Granny while she is ranting at them. The instructor/operator can use the single lines of dialog to respond, but often it is simpler to just make her scream louder, going off on an extended rant. This button does that.
- **Marines Threaten Granny**: If the Marines should point their weapons at Granny or overtly threaten her, this button cues her to respond. She will also play one of these responses if they fire their weapons. Her responses vary depending on her current emotional state (i.e. how close are the Marines to finding contraband) and whether they have previously threatened her or not.

As described above, each of these commands will select a single utterance appropriate to the current

situation, play it, and then return to the performance associated with the current behavioral state.

Although these commands improve the quality of Granny's performance, making her more responsive to the situation, their use is not essential. Thus, they fit in with our vision of having Granny be autonomous, but providing the operators with controls that allow them to improve her performance.

4.3. Single Lines of Dialog

Initially, we envisioned the above as being the full extent of the Angry Grandmother's performance. We envisioned her as so by the potential discovery of contraband that she was essentially unresponsive to anything that was said. However we found that there was a powerful desire to be able to question her once contraband has been found. In addition, we found that the more interactive we could make her, the more relevant and to a certain extent the more real she was to the Marines.

In order to support more direct interaction, we provided Granny with responses that she could speak on command. These vary from generic answers (e.g. "yes," "no," "how would I know?") to very specific responses (e.g. "I don't know when he'll be back," or "that's just medicine for my headaches!"). In all we had 41 different lines, as shown in Figure 2, and a total of 263 variants of those lines so that we can avoid repetition.

Although there are a lot of lines to choose from, the instructor/operator's ability to select a line and respond in a timely fashion is enhanced by the fact that Granny speaks only Pashto, so the Marines are working with a translator. As a result, once a question is asked, Granny has to wait for the translator to repeat it in Pashto before she responds. This provides sufficient time to select an appropriate response and find the corresponding button.

These lines are primarily intended to be used toward the end of the scenario, after the Marines have found some or all of the contraband in the room. With that said, they are available for use at any time that the operator feels they are appropriate.

4.4. Shot Tracking Systems

The FITE JCTD included two shot tracking systems. The first detects shots that are fired using the Instrumented – Tactical Engagement Simulation System (I-TESS). When a shot is fired, an AFI Shot interaction is broadcast over the HLA.

The second shot tracking system was developed by the Naval Air Warfare Center Training Systems Division (NAWCTSD). It uses an infrared camera to detect the location of any hit on the screen from an I-TESS laser or from a SESAMS (Special Effect Small Arms Marking System) round. It broadcasts AFI Detonation events, which specify the character hit (if any). [5]

The AI listens for these interactions. The AFI Shot interactions cause Granny to play a "Marines Threaten Granny" reaction, while the AFI Detonation interactions are used to cause her to die (assuming they hit her, of course).

The simulation also listens for these events, and plays a visual effect at the location of the impact, which helps to enhance immersion.

4.5. Live Character Tracking Systems

The FITE JCTD also included two different tracking systems which can be used to determine the locations of all live participants in the exercise, whether they are Marines or role-players. The first, a Radio Frequency Identification (RFID) tracking system, is accurate to roughly 1 meter. The second is a 6 DOF (Degree of Freedom) system, which is accurate to within a few centimeters. [6] In both cases, the tracking data is transmitted to the IOS via HLA. The IOS displays it on a map of the village as shown in Figure 4.

On the map, green dots are civilians, blue dots are Marines, and red dots are insurgents. The red circle around each dot indicates the current health of that participant (a full circle for unhurt, a partial circle for wounded, no circle if they have been killed). You can see the Angry Grandmother Room on the lower left part of the map. Granny's screen is located on the bottom wall, covering the left half of the room.

Instructor/Operators can use this display to track the Marines' progress through the village. In addition, the IOS can trigger events in the scenario when certain thresholds are crossed, as follows. Using the Master Scenario Events List (MSEL, or "measle") system, we define a polygonal area on the map (a "trigger zone"). These trigger zones are displayed with blue outlines; there are several of them visible in Granny's Room in Figure 4. These zones can trigger a variety of events, such as sounds, scents, or animatronics actions, when live actors enter or leave them. One such event is to sent an AFI Brain Variable Request, just as we do when one of the command buttons is pressed.

With this data, we can configure Granny to automatically play "Marines Knock" reactions from time to time when the Marines are standing just outside

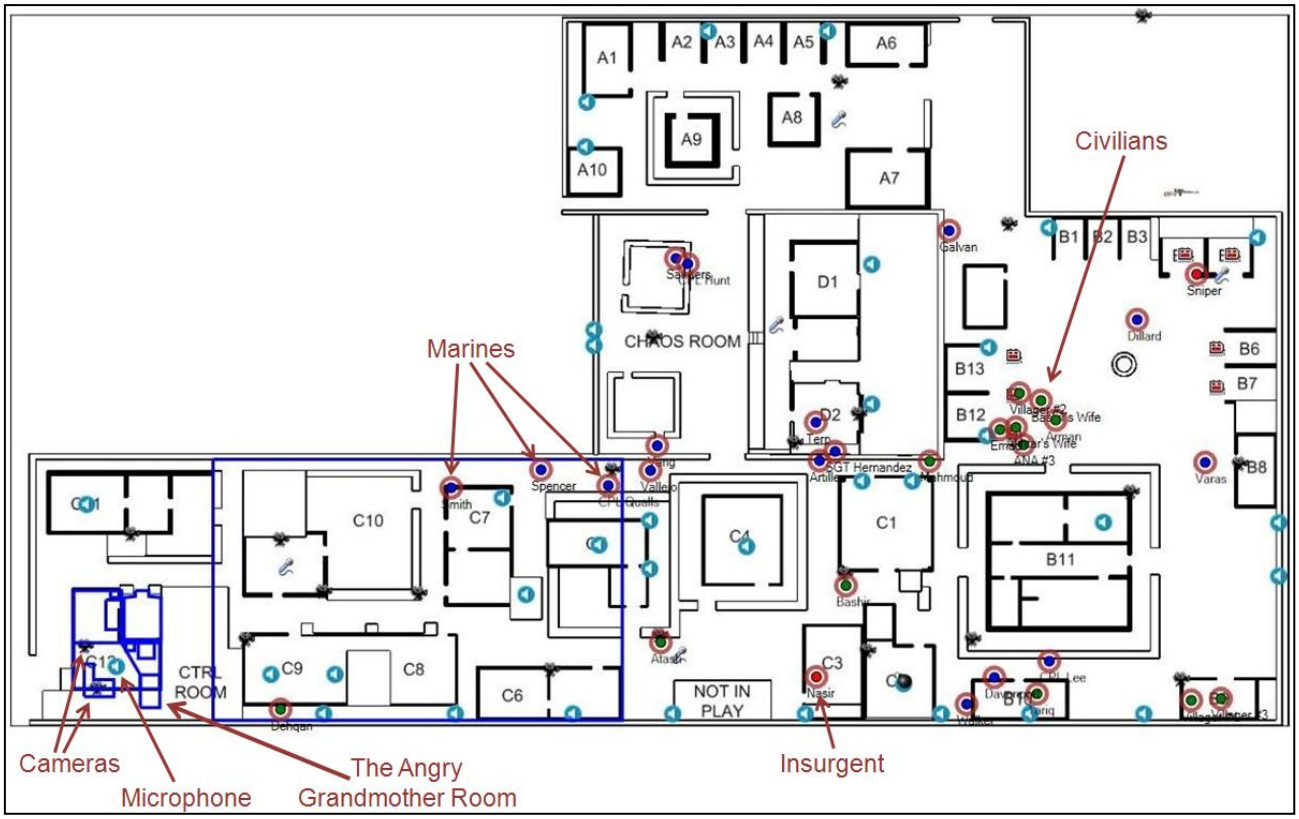

Figure 4: The IOS's map view.

her door, to automatically enter the Rant state when the Marines first enter her room, and to become more or less anxious as the Marines get closer or farther from the various caches of contraband hidden around the room. This gives her a core of autonomous actions which can be supplemented with operator commands.

4.6. Observing the Marines

In addition to knowing the positions of the Marines, the instructor/operator also needs to see and hear them. Toward that end, we placed two fixed cameras and a microphone in the room. Their approximate locations can be seen in Figure 4. There is also a pan/tilt/zoom camera which is primarily intended for the AAR system, but can be observed by the instructor/operator. Finally, the Marines' squad and team leaders wear head-mounted cameras and lapel microphones, which can also be accessed from the control center.

The camera and microphone systems share their output using streaming IP, generated using Axis encoders. [7] They are integrated into the IOS, and can also be viewed in a browser window.

5. Maintaining the Suspension of Disbelief

As discussed in the introduction, the ability to "simulate the immersive conditions of the combat environment" was a prerequisite for the success of the FITE JCTD. Therefore, we needed to make Granny immersive, to maintain the suspension of disbelief, to make her believable and relevant to the Marines who were being trained.

All of the features that we mention above – photorealistic graphics, professionally recorded sound, motion captured animation, culturally appropriate motions and dialog, operator guided actions, and so forth – contribute to our ability to accomplish this. By themselves, however, they are not sufficient.

It is our belief that one of the key factors which can make virtual characters feel artificial, even when they are photorealistic, is the repetition that we find in their motion. Real people never move in exactly the same way twice. There are always at least subtle variations. Virtual characters, on the other hand, are limited to a finite (and usually small) set of animations from which they can select. What's more, we tend to rely on short looping animations (also known as idle animations) to make up much of the motion of our characters. The end

result is repetition and uniformity of motion which the human eye quickly recognizes, and which breaks the suspension of disbelief.

With that said, abandoning keyframed animation in favor of a fully procedural approach is unlikely to be an improvement. Procedural animation techniques are generally not yet sufficiently mature to maintain suspension of disbelief. Instead, we prefer hybrid approaches that start with keyframed animation and then find ways to reduce or disguise repetition.

The approach that we took on this project was to break Granny's AI up into several distinct subsystems, and then use animation blending to combine those subsystems. As a result, while individual portions of Granny's body might be playing animations that have been used previously, those animations will be combined with others in ever-changing ways, resulting in an overall motion that is unique.

It's worth mentioning that RealWorld includes an early version of Natural Motion's Morpheme animation engine, which has extremely robust blending support. [8] This allowed us to create complex animation blends with relative ease.

The AI subsystems which help to create Granny's performance include: a *dialog system*, which decides what to say, a *gesture system* which controls her arms and upper body, a *pacing system* which controls her lower body, and an *emotion system* which controls facial animation and the speed and intensity of her hand-wringing motion. In addition, we play a *lipsynch animation* which makes her mouth move in time to her speech. All of this is combined with a line of dialog in an *utterance*.

5.1. Utterances and Dialog Selection

Most of the time, when it isn't reacting to external events such as operator commands or weapons fire, the high level AI for Granny is selecting a series of utterances to play. An utterance contains everything we need to play a line of dialog, including the filename of the sound file, along with metadata specifying when and how to play that dialog and how the various animation systems should control her body if we do.

The factors used to select an utterance include:

- The current behavioral state. Obviously, Granny uses different utterances when she is ranting or when she is inconsolable, and she doesn't use any utterances at all when she's dead. When she's standing and listening, or surrendering, she only plays utterances if prompted to do so from the IOS.
- Granny's anxiety level. This can be determined randomly, or it can be based on the positions of the Marines. In the latter case, her anxiety goes up as the Marines get closer to the locations of hidden contraband.
- Repetition avoidance. Once Granny plays a particular line of dialog, she avoids using it again for a short time.

5.2. Gestures and Gesture Sequences

The usual approach to dialog animation is to have a full-body animation that is played in conjunction with each utterance, ensuring that the character gestures appropriately, and that the timing of the gestures matches the timing of the speech. Unfortunately, due to the fact that we have only a finite number of utterances, doing this would result in visual repetition.

Instead, we had an Afghan role-player record a number of typical gestures for us, which were divided into a few distinct categories. We then had the AI select a sequence of gestures to play in conjunction with each utterance. At the end of the utterance, we simply blended out of whatever gesture was playing, and back into an appropriate upper-body pose (and then into the first gesture for the next utterance). Thus, even when a gesture was reused it often blended out early, changing the overall motion.

By eliminating the correlation between audio and video during the utterances, we made it less obvious that we're really just playing a sequence of prerecorded sound clips, and made it feel more like she was genuinely speaking, saying the same thing but in a slightly different way. The audio wasn't different, but the motion was, which helped to maintain the illusion of life.

One concern was that, with this change, her gestures would no longer match her speech. In particular, if her speech had strong moments of emphasis, and her gestures also had strong moments of emphasis, those moments might not match up. Two things prevented this. First, those moments are often at the start of an utterance, in which case they will match. Second, it turns out that when they don't match, we tend not to notice – but if they do match, even occasionally, the eye latches onto that. The end result is that viewers will perceive that the sound and animation are synchronized, even when this is not the case. We have actually had viewers comment that they like the fact that the gestures match the speech, even though we are doing nothing to make this happen.

With that said, it is important to select gestures that at least roughly match the content of the dialog. For instance, it is appropriate to reach dramatically for the sky in a praying motion when beseeching Allah for support, but not necessarily when muttering quietly to yourself. We accomplished this by dividing our gestures into categories. Each utterance could then specify the type(s) of gestures that are appropriate. Examples of these categories include:

- **Plead**: a pleading gesture, typically used when directly addressing the Marines, asking them to leave.
- **Get Out**: point at the door, or sharp horizontal waves, appropriate when demanding that the Marines leave.
- **Query**: questioning gestures, appropriate when asking the Marines questions (e.g. "Why are you doing this to us?").
- **Pray**: hands steepled at the chest, or raised toward the sky. Appropriate when beseeching Allah.

Additionally, gestures were divided into three categories based on Granny's level of anxiety. Thus the gestures used when the Marines first enter, and she is mostly angry, are different from the ones used when they have been searching a while and are getting close to finding something.

The end result is that while Granny has individual gestures that are recognizable, they don't always play the same way (because we blend into them, and blend out early), they don't always play in the same sequence, and they don't always play in conjunction with the same lines of dialog. This creates significantly more variation than we would have if we had only one gesture animation for each line of dialog.

5.3. Pacing

While the gesture system controls Granny's upper body and arms, we wanted a separate system that would allow her to move around her nook a little bit. This not only makes her seem less stiff and automated, but also further increases the variety of her performance by blending separate upper and lower body motions. We accomplished this through the use of pacing animations that included half steps (where only one foot moves) and full steps in various directions.

We do still have a few full-body animations that are intended for use in special circumstances. For example, there is a special animation that plays in conjunction with a long line of dialog when the Marines first enter the room, and another when they discover contraband and she enters the Inconsolable state. There are also similar moments that happen the first time that the Marines threaten her or try to talk to her. These moments are important, and only happen once. The custom animations, which have more freedom of motion than her normal performance, allow us to emphasize them.

These full-body animations caused a problem for the pacing system, however, because we didn't want to have to worry about Granny colliding with the other objects in her nook when one of them plays. Thus, for every pacing animation that moved Granny away from her starting position, we created a corresponding return animation that would bring her back. The responsibility of the pacing system was to pick from among the pacing animations, keep track of which one was most recently used, and ensure that we played the corresponding return animation after each one. Although this doesn't guarantee that Granny will be in a specific location at the start of a full-body animation, it does limit her position enough that she won't collide with her surroundings.

5.4. Expressions and Emotions

Although much of Granny's face is covered by her veil, we still wanted to be able to display her emotion through subtle cues such as the movement of her brow.

Our approach was, first, to create animations which would display an extreme expression (such as anger or shock). We then set up blend weights which allow us to blend in a little or a lot of each expression. For example, if we set the shocked expression slider to 1 then her face will show an extreme caricature of surprise. If we set the same slider to 0.3 or 0.4 then we will get a more subtle, mildly surprised expression. If we set all sliders to 0 then we will get a neutral expression. We can also control the speed and intensity of the hand-wringing motion that's built into her standing idle in the same way.

Next, we created five emotional states: neutral, worried, terrified, upset, and furious. Each emotional state defines random ranges for the blend weights of each of the expressions. For example, the furious emotion sets the anger expression's weight to a value between 0.75 and 1, it sets the shocked expression's weight to 0, and it sets hand wringing weight to a value between 0 and 0.25. Conversely, the terrified expression sets the furious expression's weight to 0, and the shocked expression and hand wringing to values between 0.75 and 1.

Each utterance could specify the emotion that should be used while that utterance is being played. This

would result in small movements of the face even when the same emotion was selected (since the random values would be slightly different), and larger movements as she transitioned from being furious at the Marines' intrusion to terrified that contraband will be found.

5.5. Lipsynch

The lipsynch system is responsible for moving the lips in correlation with the speech. There are two common approaches to lipsynch animation. The first is to use software to evaluate the sound file, and create a matching animation. The second is to record the animation at the same time as you record the audio, and create the animation that way. Of the two, the latter is generally felt to produce better results.

Unfortunately, the recording studio we used didn't have the capability to record the lip animation alongside the audio, which would have been our first choice. Fortunately, due to the fact that Granny is wearing a veil, you can't see the precise lip motions, just the general opening and closing of the jaw. As a result, we were able to create a very simple jaw-flapping animation which, when obscured by the veil, was just as good as a more accurate lipsynch animation. Not only did we have several people comment favorably on the fact that the lip motion matches the spoken words, but we also found that even people who were primed to look for the differences could not generally distinguish between the two.

Aside from the fact that it helped to reduce the cost of developing Granny, this also has the distinct advantage that it makes it much easier to add new audio files. For example, when we added the single lines of dialog late in the project, we didn't need to do any extra work to create lipsynch animations for them – we just used the "jaw flapping" animation that we already had.

6. Results

We ran a number of squads through the IIT using the FITE technology over the course of this effort. Our sample size is small, and the results of the independent evaluation are not yet complete, but we did see some nice indications of success.

Perhaps the most promising sign is that one squad elected not to search Granny's room at all. They simply said "Sorry to have bothered you, ma'am," and moved on. This would seem to indicate that Granny is real enough, and provocative enough, to cause these Marines to change their intended course of action.

In addition, another of the squad leaders that came through commented that "the angry woman acted exactly like women I experienced in country." This suggests that Granny was sufficiently culturally authentic to be believable.

Another success was the ease of modification. The modular approach taken, as well as the data-driven design of the IOS, allowed us to quickly change the AI and associated controls as the design of the scenario changed. For example, late in the project we were asked to produce an alternate scenario in which Granny had a gun. Building the AI and controls for this scenario only took us two days.

With that said, once significant challenge that will need to be overcome with this sort of character is the expense of generating the art assets – particularly the audio and animation assets. Although we were able to reconfigure the AI and scenario to support a gun very quickly, creating the associated animations took quite a bit longer – and ultimately, we were unable to produce assets that we were happy with in the time available, and this scenario was not shown at the JTCD. In the long run, the obvious solution here is to build up a library of audio and animation assets that can be reused across characters and across projects.

One area where our results were not as good as we had hoped was in our instrumentation. While we initially hoped to drive most of Granny's performance, as well as her mood, using the character tracking data, we found this to be problematic. The more accurate, 6 DOF system was only worn by a few participants (typically the squad and team leaders). In particular, the role-players acting as attached Afghan military members were tracked using RFID, and doctrine requires that these be the first people to enter the room. This made it impossible to accurately detect the moment when the room was breached. Even after the Marines had entered, there were simply too many people moving around in a small space to make it realistic to adjust her mood based on where they were positioned. As a result, we got better results using operator controls to cue her when the room was breached, and when contraband was discovered, and using random selection to vary her mood over time.

Had time permitted, this problem could have been addressed to some extent by placing sensors directly on the doors to the room, as well as the covers for the contraband caches, rather than relying on two-dimensional tracking data for this information. Another solution would be to use the bulkier 6 DOF system for all participants (including the role-players playing members of the Afghan military).

7. Future Work

In addition to those discussed above, there are a few areas for future work which we would like to pursue it he opportunity presents itself.

First, the 6 DOF tracking system is capable of tracking not only the position of the user, but also the orientation of their head (i.e., what they are looking at) and weapon (i.e., what they are aiming at). That information could be used to change Granny's performance depending on whether or not anybody is paying attention. For instance, when the Marines ignore her she might pace back and forth, muttering angrily to herself, and only occasionally say anything to them directly. If they turned to look at her, she could launch into a tirade similar to the one she uses now. It would also be culturally appropriate to have her look away any time a Marine looks at her directly. Finally, we could have her play an appropriate response if the Marines aimed their weapons at her or at her grandson, rather than relying on the operator to cue that.

A large part of the Granny's success was a result of a fact that (a) she spoke only Pashto, and (b) she was so upset as to be irrational and a little bit incoherent. This allowed us to select sequences of utterances that didn't necessarily make perfect sense. If we were to use this system for an English speaker, or a calmer character, it would be necessary to annotate our utterances and build sequences that flow more naturally.

Our primary solution to reducing the repetition inherent in the use of keyframed animation was to blend multiple animations, each controlling a different part of her body. While this is a good start, many of the repeated gestures are still recognizable. Further work could be done to find ways to control some parts of her body procedurally, or to procedurally modify the keyframed animations so that they play back differently each time they are used. For instance we could speed up or slow down certain parts of the animations, or we could use parameterized inverse kinematics or Perlin Noise [9] to adjust the bone positions within the animations.

Finally, this is a standalone simulation which neither accounts for nor influences other events that occur during the squad's training. It could be interesting to build a larger village state, possibly up to and including a social/cultural model of all characters – whether live, virtual, or constructive – and use that information both to shape Granny's performance and to shape the performances of other characters depending on the way in which the Marines handle her.

8. Conclusion

There is a rapidly increasing demand for believable, immersive, culturally authentic constructive characters for training, and for game AI technology to drive those characters. In this paper we have discussed the approach that we took to creating one such character. We feel that we succeeded at creating a character that was sufficiently believable and culturally authentic to provide effective training, and that the approaches here provide a good starting point for future efforts.

9. References

[1] P. Muller: "The Future Immersive Training Environment (FITE) JCTD: Improving Readiness Through Innovation," Intraservice/Industry Training, Simulation & Education Conference, 2010.

[2] K. Dill: "A Pattern-Based Approach to Modular AI for Games," Game Programming Gems 8, Cengage Learning, Boston, 2010.

[3] Total Immersion Software: RealWorld, http://www.totimm.com, 2010.

[4] IEEE 2000, September, Standard for Modeling and Simulation (M&S) High Level Architecture (HLA) - Framework and Rules. Document 1516-2000, 2000.

[5] Naval Air Warfare Center Training Systems Division: "MEFTRC Weapon Integration for IIT User Manual," 2010.

[6] S. Samarasekera et. al.: "Seamless Indoor/Outdoor 6 DOF Tracking of Trainees and Weapons," Intraservice/Industry Training, Simulation & Education Conference, 2010.

[7] Axis Communications: Axis video encoders, http://www.axis.com/products/video/video_server/index.htm.

[8] Natural Motion: Morpheme, http://www.naturalmotion.com, 2006.

[9] K. Perlin: "Ken Perlin's Homepage," http://mrl.nyu.edu/~perlin/.

10. Author Biographies

KEVIN DILL is a staff software engineer with Lockheed Martin Global Training & Logistics. Prior to coming to Lockheed Martin, Kevin was an AI programmer in the games industry for 9 years, and he has 7 published titles to his credit. Kevin was a section editor for "AI Game Programming Wisdom 4," the technical editor for "Behavioral Mathematics for Game AI" and "Introduction to Game AI," and has taught classes on game development at Harvard and Boston University.

OLIVER DREGER is a defense consultant with SAIC Corporation, and is also a USMC Reserve Officer. Prior to his current position, he spent ten years as a USMC Officer on active duty. Oliver has served as an intelligence officer for a Marine infantry battalion on combat deployments, and has deployed and served with aviation units and higher headquarters staff as well.