

Research on the Use of Puppeteering to Improve Realism in Army Simulations and Training Games

Tami Griffith¹, Tabitha Dwyer², Corey Kinard²(✉),
Jeremy R. Flynn³, and Vic Kirazian²

¹ U.S. Army Research Lab, Human Research & Engineering Directorate,
Advanced Training & Simulation Division, Orlando, USA

tamara.s.griffith.civ@mail.mil

² Cole Engineering Services, Inc., Orlando, FL, USA

{tabitha.dwyer, corey.kinard,
vic.kirazian}@cesicorp.com

³ University of Central Florida, Institute for Simulation and Training,
Orlando, FL, USA

jflynn@ist.ucf.edu

Abstract. Virtual environments are increasingly used to provide training in military applications. These virtual spaces can represent large complex operational environments or simply a simple room. In either case the realism of the interpersonal interactions can be an important factor that influences the trainee's decisions. Interactions may include micro-expressions, small-motor movements, eye movement and voice intonation. This paper describes research conducted by the U.S. Army Research Lab, Human Research & Engineering Directorate, Advanced Training & Simulation Division along with Cole Engineering Services, Inc. and University of Central Florida, Institute for Simulation and Training on the use of tools that allow an individual to take ownership of an avatar in the Unreal 4 game engine. This is being described as puppeteering and is expected to improve realism and engagement. Potential uses for this technology are also described.

Keywords: Puppeteering · Virtual realism · Real-time interactions

1 Introduction

Neil Stevenson's *The Diamond Age: Or, a Young Lady's Illustrated Primer* [1] describes a future where a young girl receives an interactive book capable of providing her vast knowledge. The book provides instruction through allegories and simulations. Unfortunately, not everything can be taught in that way. Some learning must take place through interactions. In the book, interactive movies evolved because artificial intelligence was unable to meet the interactive needs of entertainment. "Ractors" are actors in those movies because they react directly to the participant, creating realistic interpersonal experiences from anywhere in the world. It was through this technology that the young girl continued to gain learning. A common theme in the book is the vast difference

between Artificial Intelligence (AI), which is renamed “pseudo-intelligence” in the book, and human intelligence. Interactive movies are expensive but highly valued because people are willing to pay for human intelligence and can quickly identify pseudo-intelligence. The Turing Test – which is used to distinguish human from machine – is alluded to throughout the story. This concept resonated with the teams at the U.S. Army Research Lab, Human Research & Engineering Directorate, Advanced Training & Simulation Division along with Cole Engineering Services, Inc. and University of Central Florida, Institute for Simulation and Training, and drove them to explore ways to improve the realism of human interactions within games and simulations.

Virtual environments are becoming more realistic, with textures and lighting that can make you feel like you’re actually in a desert or trudging through a cave. Virtual characters are also increasing in realism. Characters have scars, skin imperfections and asymmetry that makes them relatable. Cut-scenes and pre-canned animations look more life-like. The technology described in this paper explores ways to improve realism when interacting with game or simulated characters.

The team explored various off-the-shelf tools that would allow an individual to take ownership of a virtual character using the Unreal 4 (UE4) game engine. Facial micro-expressions, such as wincing, avoiding eye-contact, frowning or smirking have the potential to be important in deciding how you will respond to an individual. Additionally, gestures such as shrugging, pointing or shifting weight could potentially make a case for an individual’s frame of mind or intentions. Having the ability to replicate these movements in real-time can greatly increase the level of realism in a virtual environment. Characters must react realistically to actions or verbalizations of the player character or trainee. This goes far to suspend disbelief and immerse the trainee in the scenario in an evocative way.

This paper will explore the topic of puppeteering in simulated environments. For the purpose of this paper, this includes simulators, games and virtual environments. As such, the term simulation is intended to be inclusive and synonymous with game and virtual environment. The focus of this paper is on the use of puppeteering to support various types of training in virtual environments.

2 The Importance of Realism

2.1 The Role of Fidelity

Fidelity can be described as the level of realism displayed in a simulation [2]. Improved computer graphics cards and processing speed has made it possible to render virtual environments that appear very life-like. However realism still comes at a cost. Creating realistic environments takes a large amount of artist time and building realistic character behaviors takes a great amount of developer time. Increased realism also comes at a higher processing cost. Each effort within a virtual or simulated environment must include a trade-off analysis between the level of realism and performance [3].

Research [4] indicates that greater fidelity is associated with improved engagement and sense of presence. Higher fidelity can influence the extent to which users are able to suspend disbelief that the virtual environment is real and that what happens within it is

meaningful with respect to the learning goals of the developer. In fact, research [5] has shown that if the simulation cannot provide the appropriate level of fidelity in relation to real-world cues, the result can influence training by providing negative training transfer or negative training.

Despite arguments supporting the value of higher fidelity environment to support training, there are also training goals that can be met without a high-fidelity simulation. For example, Norman, Dore and Grierson [6] showed that a low-fidelity simulation of heart sounds functioned as well, and in some cases better, than higher cost and higher-fidelity simulators. The important take-away from that research is that “the relationship between simulation fidelity and learning is not unidimensional and linear” [6].

One compelling argument for improved fidelity comes from Vice et al. [7]. The research was focused on the US Marine Corps Combat Hunter training program which is focused on battlefield situational awareness and observation skills. They found that subject matter experts may be proficient at a task, but may not be able to articulate the cues necessary to support decision making. Using eye-tracking, electroencephalogram (EEG), cognitive workload and attention allocation, the researchers were able to explore event-related potentials (ERPs) based on slight variations that occurred, in some cases outside of the participant’s awareness. Performance was varied, but higher amplitude ERP waveforms were detected in the virtual condition which indicated higher levels of processing taking place.

Based on the previous discussion, simulation fidelity is important in association with cues that may stimulate action on the part of a simulation, such a defensive driving. However, fidelity should also be considered in association with its role in immersing a trainee in the training scenario, or establishing presence, and the associated level of engagement in the scenario.

2.2 The Role of Presence, Immersion and Engagement

Presence is a concept that can be defined as the “extent to which a person’s cognitive and perceptual system are tricked into believing they are somewhere other than their physical location” [8]. Brown & Cairns [9] make the argument that presence is the same as total immersion. They describe examples where gamers create distraction-free environments, low-light and high game-volume, that enable them to suspend disbelief in the game world. Researchers in learning and psychology may consider this state cognitive engagement [10]. Cognitive engagement involves “seeking, interpreting, analyzing and summarizing information, critiquing and reasoning through various opinions and arguments; and making decisions” [11]. Engagement has long been correlated with positive student outcomes [12, 13]. But it is important to ensure that students engage with the learning material rather than the mechanism for providing the learning material [14].

2.3 The Role of Experiential Learning

Simulations allow students to experience situations before being faced with them in the real-world. Learning that is grounded in experience is the concept behind experiential learning. Kolb [15] described how learning transformation takes the form of internal

reflection or through active manipulation of the external world. Knowledge gained through experiential learning, and understood at a deep, conceptual level, would be expected to be transferred and generalized more readily [16].

Puppeteering applied to simulated training events provides an opportunity for learners to experience a wide range of experiential training scenarios. Changing the scenario is simply a matter of changing the activities of the puppeteers.

One concern with using puppeteering in simulated training is that live actors increase the support burden of training events. This is an important and valid concern. However, there are live training events that occur in the Army where entire villages of people are paid to support week-long live training activities. Employing puppeteering would allow a much smaller group of actors to support the training. These puppeteers could be anywhere in the world while they support the training events. This greatly reduces the overall support costs of this type of event.

3 Technical Solutions

3.1 Art and Animation

Facial performance motion capture, or “mocap,” has been used extensively in both the visual effects and game industries for over a decade. While there are many different types of capture systems, ranging from marker-based to marker-less, most of them do not offer a real-time solution for driving an avatar, or what we’re calling “virtual puppeteering.” This isn’t a shortcoming of the technology; rather, it is a realization that most end-users of facial animation systems are not using the technology in real-time, but applying it to use-cases that allow post-processing. In a traditional production an actor’s or actresses’ facial performance is captured, then a team of animators will take that raw data and polish it into a believable facial animation to be used in the film or video game [17]. If the capture system offers any real-time capabilities, it generally is used for previewing and not for final animation. The need for real-time tracking narrows the field of software solutions to only a few.

Our chosen development environment is Epic Game’s Unreal Engine 4[®] (UE4). The team conducted market research and found two different facial tracking software systems of varied tracking quality and total cost to the end-user. The two systems used different methods for tracking a target face. One used a depth camera, such as a Kinect 1.0 or a Primesense Carmine, to scan a person’s face and track facial expressions. This method required each person to go through a set of 24 calibration expressions before facial tracking could begin. Calibration success was determined by how well the person could make the required expression shapes. The second software solution used a standard off-the-shelf web camera and required no calibration. The actor simply oriented themselves in view of the camera and the system began tracking. It is worth mentioning that in both cases, it was best to have a camera that performed well in low lighting conditions.

Character’s faces were animated via morph targets rather than via joint-based deformation. The team created 51 morph targets representing various facial positions during the production of phonemes and facial expressions in a content-creation package such as Maya, which were used to control a character in real-time. This could be accomplished through multiple methods such as modeling the facial expressions by

hand or by utilizing scan data and photogrammetry. The 51 morph targets developed, along with neutral expression and 4 eye movements (up, down, left and right), are shown in Fig. 1.

The expressions were imported as morph targets and mapped to the skeletal mesh in UE4 to prepare the assets for the software team. The range of each shape goes from zero to one, with zero being the neutral, default face, and one being the full extent of the expression. Each shape is named in the game engine so they can be referenced and driven by the game code.

Another focus area for the team was to combine real-time face tracking with real-time, full-body motion capture in order to achieve a holistic natural performance. This allows the actor to express emotion through body language, which is critical in a training exercise. We chose a new-to-market, 32-sensor motion capture system to capture full-body motion for a few reasons. The wearable sensors and accompanying software is very low-cost at only \$1,500, whereas some of the competitors range up to ten times that amount. This cost makes it possible to field multiple suits if a training exercise requires

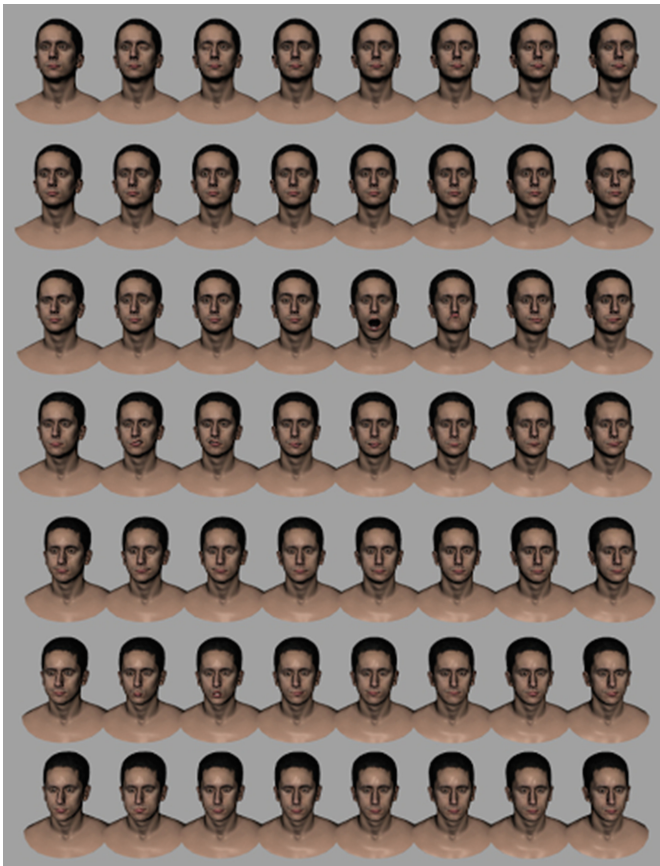


Fig. 1. Morph Targets generated

more than one role-player. The system is marker-less, which means cameras aren't needed to capture the motion. This is important because it allows for more flexibility and portability to the capture process, and you are not limited to a small area to do the performance. The software package already works with the UE4 skeleton for animation. This means when using UE4 to rig characters there is no additional art or animation necessary to apply the technology. Lastly the full body motion capture set captures and streams in real-time, enabling us to send animations to UE4 via community created plugin.



Fig. 2. Wearable markers for full body virtual puppeteering

3.2 Software Development

As the team conducted market research, a clear technical division emerged. It was necessary to combine two separate technical solutions, one for the face and another for the body. With current technology, any real-time facial animation solution needed the camera to be too close to the face to also detect the body, and anything that could handle detailed body movements would not be able to process the facial expressions with high enough fidelity.

First the team focused on facial tracking. The goal was to reach high fidelity facial expressions to improve simulated training for the US Army. After working with several commercial products, it became clear that two of the products allowed us to stream live facial micro-expressions to the gaming engine. As described in the previous section on art and animation, real-time puppeteering in a game engine is an emerging technology. Unfortunately neither real-time facial tracking solution had a direct pipeline for

controlling an avatar inside of UE4. The candidate software packages came with network interface specifications and had examples running in Maya, but a native plugin needed to be developed for UE4 to read the network data and drive the facial animations.

Thanks to thorough documentation, the team was able to create a plugin for UE4 within a couple of weeks. The plugin handles the receipt of network packets from the facial tracking software and translates it into data that are usable within the engine. A packet is sent from the tracking software to the engine once per frame. The packet includes an array of values between 0 and 1 for each face shape. It also includes eye rotation information as well as head rotation and translation. The eyes and head are skeletal movements, while the rest are applied to morph targets in the engine [18].

On the UE4 software side, a single idle animation is applied to the puppet avatar. Then a single animation is manipulated by changing each of the over-fifty values of the morph targets and skeletal joints. The frame rate is based on the camera in our case, which captures data at 30 frames per second. Since the morph target values come every frame, there is no interpolation. In each frame, the value (between 0 and 1) of each morph target is set to the value of each of the corresponding morph targets on the character. The correct bones are identified for head and eyes, and the incoming rotation and translation values are used for the animation. The currently playing animation is asked to update every frame, so we essentially create a real-time animation based on human input with nothing more in game than a single idle animation. Figure 3 shows a sample of the 51 data points being displayed on a character at one point in time. Each value is represented by the vertical lines. The teal line that is highlighted shows the brows upper center (BrowsU_C) value as 0.71 (Fig. 4).



Fig. 3. Data from face tracking software



Fig. 4. Resulting animation within UE4

After implementing the real-time facial puppeteering solution we confirmed our suspicions, that without a moving body the level of immersion and suspension of disbelief was compromised. As a short-term solution we applied pre-defined gesture animations the puppeteer could call upon with on-screen visual cues using a gamepad. This is a similar approach to many commercial video games that allow the player to “emote” using series of predefined animations. Although this may work in a scripted scenario, it was very limited in terms of the number of gestures, it took longer to train the puppeteer, and it didn’t look natural.

Using the newly released motion capture sensor set described in the art section, the team took a similar approach to accomplish full body implementation in UE4. The human-worn sensors send their location translation and rotation data across the network each frame. In this case, the data is collected at between 60 to 120 frames per second. The plugin handles the packet consumption and translation into software objects. Once again, a single animation is played on the avatar in-game and manipulated by setting the skeletal bone translations and rotations every frame. This gives the actor natural motions, such as shrugging, pointing, bending over laughing, kicking, or even dancing, and translates it one to one onto the in-game avatar. The display is at 30 frames per second to maintain realism and manage data throughput. The current limitation is a lack of higher level calibration and collision, such as crossing your arms or clapping. These issues may be mitigated in a software update to the motion capture software [19], and may potentially be overcome with an inverse kinematics implementation within UE4. Figure 5 shows the sensor system being applied to the UE4 skeleton system.

The challenge, then, is to bring these technologies together; facial expressions, body motion capture, and game controls to navigate the game environment. The solution must allow the actor to move and talk at the same time, so the camera must



Fig. 5. Real-time motion capture sensor system applied to UE4 skeleton

move with the actor. Various camera mounting strategies are being explored to allow this, such as a chest harness (as seen in Fig. 2) and helmet mount. Navigation is controlled using an analog navigation controller such as the Playstation® Move Navigation Controller (Fig. 6). It is hung at the side of the actor to be grabbed and dropped as needed so as not to interfere with hand and arm motions.



Fig. 6. Playstation® move navigation controller [20]

4 Applications

There are many potential applications for this technology. They fall into two basic, overlapping themes: direct, potentially emotional, interpersonal interactions; and breathing life into a virtual landscape. Interpersonal interactions have been applied to a wide range of leadership training. For example, a simulation can train leaders to recognize the signs of sexual harassment/assault, drug or alcohol abuse or Post-Traumatic Stress Disorder (PTSD), to name a few. In addition, avatars can be controlled through AI within an urban environment. Trainees can approach an AI character sweeping their front porch. A puppeteer takes ownership of that character, so that they are able to engage in a meaningful discussion, answer questions, act as a threat, or request assistance. This mirrors a true tactical environment where anything can happen and the trainee experiences the immediate and long-term consequences of their choices. As soon as the trainee disengages with that character, the AI would take over the behaviors and the puppeteer could step into the role of the next character the trainee encounters. One person could play many roles. This can provide the sense that the town is teeming with life with extremely low costs to support.

There are many more applications, such as PTSD therapy, exposure therapy, entertainment and the list goes on in areas this team has not even considered.

5 Discussion/Conclusions

The real-time connection between facial tracking software and body motion tracking linked to UE4 that is controlled through natural motions by an actor, is now a reality. So, how well does the system work in practice? It functions well, but there is room for improvement. While there is great power in the ability to drive character performance in real-time in a way that is completely unscripted, the downside is that there isn't an artist in the loop to clean up the peculiarities that may appear in the data. Real-time puppeteering can quickly enter the "uncanny valley." The uncanny valley is a theory used in reference to a sense of unease or revulsion caused when a computer-generated figure or humanoid robot bears a near-identical resemblance to a human being, then perceiving an indication that it isn't human [21, 22]. There is little actual data to support the theory, however the idea is widely supported. Virtual puppets that express complex emotions, such as talking and laughing at the same time, can bring about unusual results such as odd mouth shapes and showing far too much of their teeth. The uncanniness can be reduced by either visually stylizing the character or by training the actor on how to keep their performance within behaviors that appear more realistic.

There is a need for realistic human characters in virtual environments for various applications. The focus of this paper is on US Army training, however, the process can be applied to a wide range of uses. The process for integrating various technologies to provide real-time puppeteering have been described along with their strengths and weaknesses. The technology is in its infancy, but the expectation is that, with more investments in virtual and augmented reality, the need for puppeteering technology will increase.

References

1. Stephenson, N.: *The Diamond Age: Or, a Young Lady's Illustrated Primer*. Bantam Spectra Book, New York (2000)
2. Feinstein, A.H., Cannon, H.M.: Fidelity, verifiability, and validity of simulation: constructs for evaluation. *Dev. Bus. Simul. Experiential Learn.* **28**, 58–67 (2014)
3. Breslau, L., Estrin, D., Fall, K., Floyd, S., Heidermann, J., Helmy, A.: Advances in network simulation. *Computer* **33**(5), 59–67 (2000)
4. McMahan, R., Bowman, D., Zielinski, D., Brady, R.B.: Evaluating display fidelity and interaction fidelity in a virtual reality game. *IEEE Trans. Vis. Comput. Graph.* **18**, 626–633 (2012)
5. Summers, J.E.: Simulation-based military training: an engineering approach to better addressing competing environmental, fiscal and security concerns. *J. Wash. Acad. Sci.* **98**, 9–22 (2012)
6. Norman, G., Dore, K., Grierson, L.: *The minimal relationship between simulation fidelity and transfer of learning*. Medical Education, Blackwell Publishing, Ltd (2012)
7. Vice, J., Skinner, A., Berka, C., Reinerman-Jones, L., Barber, D., Pojman, N., Tan, V., Sebrecchts, M., Lathan, C.: Use of neurophysiological metrics within a real and virtual perceptual skills task to determine optimal simulation fidelity requirements. In: Shumaker, R. (ed.) *Virtual and Mixed Reality, HCII 2011, Part I. LNCS*, vol. 6773, pp. 387–399. Springer, Heidelberg (2011)

8. Patrick, E., Cosgrove, D., Slavkovic, A., Rode, J.A., Verratti, T., Chiselko, G.: Using a large projection screen as an alternative to head-mounted displays for virtual environments. In: *CHI Letters* (2000)
9. Brown, E., Cairns, P.: A Grounded investigation of game immersion. In: *CHI 2004 Late Breaking Results*, London, UK (2004)
10. Li, M., Jiang, Q., Tan, C.-H., Wei, K.-K.: Enhancing user-game engagement through software gaming elements. *J. Manage. Inf. Syst.* **30**, 115–150 (2014)
11. Zhu, E.: Interaction and cognitive engagement: An analysis of four asynchronous online discussions. In: *Instructional Science*, pp. 451–480 (2006)
12. Kuh, G.D., Kinzie, J., Schuh, J.H., Whitt, E.J.: Never let it rest: lessons about student success from high-performing colleges and universities. *Change Mag. High. Learn.* **37**(4), 44–51 (2005)
13. Pascarella, E.T., Terenzini, P.T.: *How College Affects Students: A Third Decade of Research (Volume 2)*. Jossey-Bass, San Francisco (2005)
14. Ke, F.: A case study of computer gaming for math: Engaged learning from gameplay? *Comput. Educ.* **51**, 1609–1620 (2008)
15. Kolb, D.A.: *Experiential Learning: Experience as the Source of Learning and Development (Second Edition)*, Upper Saddle River. Pearson Education Inc., NJ (2015)
16. Molanis, C., Burns, D., Assudani, R., Chinta, R.: Assessing experiential learning styles: A methodological reconstruction and validation of the Kolb Learning Style Inventory,”. *Learning and Individual Differences* **23**, 44–52 (2013)
17. Lowther, J., Shene, C.-K.: Rendering + modeling + animation + postprocessing = computer graphics. *J. Comput. Small Coll.* **16**(1), 20–28 (2000)
18. *Faceshift, Interfacing with Other Applications* (2014)
19. *Perception Neuron, Perception Neuron Axis Forum* (2016)
20. *Playstation, Playstation Move Navigation Controller* (2016)
21. Tinwell, A.: *The Uncanny Valley in Games and Animation*, pp. 165–170. CRC Press, Boca Raton (2014)
22. Mori, M.: The Uncanny Valley. *IEEE* **19**(2), 98–100 (2012)
23. Summers, J.E.: Simulation-based military training: an engineering approach to better addressing competing environmental, fiscal and security concerns. *J. Wash. Acad. Sci.* **98**, 9–22 (2012)
24. McMahan, R.P., Bowman, D.A., Zielinski, D.J., Brady, R.B.: Evaluating display fidelity and interaction fidelity in a virtual reality game. *Trans. Vis. Comput. Graph.* **18**, 626–633 (2013)