Creating Micro-expressions and Nuanced Nonverbal Communication in Synthetic Cultural Characters and Environments

Marjorie A. Zielke, Ph.D	**Frank Dufour, Ph.D**	**Gary Hardee, MA**
University of Texas at Dallas	**University of Texas at Dallas**	**University of Texas at Dallas**
Richardson, Texas	**Richardson, Texas**	**Richardson, Texas**
margez@utdallas.edu	frank.dufour@utdallas.edu	ghardee@utdallas.edu

ABSTRACT

Understanding how to observe and analyze nonverbal communication of virtual humans in synthetic environments can assist warfighters with determining source credibility, deception detection, behavior observation, process training and other training objectives. The First Person Cultural Trainer (FPCT) is a high-fidelity, game-based simulation that trains cross-cultural decision making and is also a platform for the creation of synthetic characters which display culturally accurate facial and micro-expressions. These facial and micro-expressions serve as a channel for factors such as invariants and affordances within the synthetic environment that also communicate critical nonverbal information. This paper will discuss technical challenges and solutions in creating facial and micro-expressions and environmental invariants and affordances such as those needed in a high-fidelity training project. The paper will also provide an overview of the state-of-the-art for creating nuanced nonverbal communications from the gaming industry and will conclude with the description of one solution for developing facial and micro-expressions and nuanced nonverbal communication.

ABOUT THE AUTHORS

Marjorie A. Zielke, Ph.D., is an assistant professor of Arts and Technology at the University of Texas at Dallas and the principal investigator on the First Person Cultural Trainer. Dr. Zielke has served as principal investigator and project manager on a series of culturally related military-funded simulations over the last several years. Her areas of research are cyberpsychology and hyper-realistic simulations, and she works in the cultural training, health and marketing sectors. Projects on which Dr. Zielke has been principal investigator, including the First Person Cultural Trainer, have won awards from the National Training and Simulation Society, the Department of Defense and the Society for Simulation in Healthcare. Dr. Zielke received her Ph.D. from the University of Texas at Dallas, and also has an MBA and a master's in international business.

Frank Dufour, Ph.D., teaches sound design for digital arts. He created one of the first all-digital recording studios in France specialized in digital sound restoration for audio publications and sound design for animation movies and video productions. His work on the First Person Cultural Trainer focuses on cognitive game development and sound design.

Gary Hardee, MA, is the research project manager for the Institute for Interactive Arts and Engineering and a doctoral student at the University of Texas at Dallas. He has many years of professional media experience and manages several UT Dallas projects involving cultural training simulations and virtual patients.

The following colleagues at the University of Texas at Dallas also contributed as authors: Dr. Robert Taylor, Dr. Bruce Jacobs, Desmond Blair, Adam Buxkamper, Jumanne Donahue, Brandon Keown and Dan Trinh.

Creating Micro-expressions and Nuanced Nonverbal Communication in Synthetic Cultural Characters and Environments

Marjorie A. Zielke, Ph.D

University of Texas at Dallas

Richardson, Texas

margez@utdallas.edu

Frank Dufour, Ph.D

University of Texas at Dallas

Richardson, Texas

frank.dufour@utdallas.edu

Gary Hardee, MA

University of Texas at Dallas

Richardson, Texas

ghardee@utdallas.edu

INTRODUCTION

Today's irregular warfare environment increasingly requires training for engaging and communicating with multi-cultural non-combatant populations. How well warfighters understand the verbal and non-verbal communications of non-combatants can influence their success in missions. Traditionally, non-combatant engagement training has been achieved through sessions with live actors. While such training can be valuable, it also can be limited and costly. Alternatively, training with virtual agents in game-based simulations can be an efficient, highly composable and effective solution.

To be effective, virtual agents, or non-player characters (NPCs), in a game-based simulation must demonstrate cognitive and emotional complexity through both verbal expressions as well as high-fidelity, nuanced visual representations that include facial expressions and much briefer micro-expressions. Further, the concept of emotional flow suggests that nonverbal communication is a continuum of behavioral actions displayed within the environmental milieu to which the NPCs respond with appropriate behaviors. This paper focuses on the development of facial and micro-expressions in synthetic environments and the creation of the overall affordances and invariants that set the context for interpreting behavior. Facial expressions, micro-expressions, affordances, invariants and emotional flow are explained further below and in Figures 1 and 2.

A facial expression is a configuration of muscle movements in the face and comes into existence in a number of ways. In the context of communication, facial expressions can be simulated, ritualized and completely voluntary (Russell and Fernández-Dols, 1997), genuinely connected to an emotional state and freely expressed, or used to mask an involuntary emotional state (Porter and Brinke, 2008). The final circumstance is the one in which micro-expressions emerge. A micro-expression occurs when a person cannot totally suppress or mask an involuntary emotional facial expression. The resulting leaked expression is brief, taking place in 1/5 to 1/25 of a second (Porter and Brinke, 2008). Figure 1 below illustrates the difference between facial and micro-expressions.

Figure 1. A facial expression (left) is a configuration of muscle movements in the face that expresses an emotional state. A micro-expression (right) is a quick burst – 1/5 to 1/25 of a second – of involuntary emotional expression.

The full understanding of facial and micro-expressions needs to be in the context of invariants and affordances. Invariants express the importance of the environment to understanding and synthetically portraying nonverbal communication. Both structural and transformational invariants are used in the model. Structural invariants are properties of objects that at some level of description can be considered the same. (Michaels, Carello, 1981) Structural invariants provide an overall sense of the combined elements of the environment to put the nonverbal communication of NPCs in context. Structural invariants might be understood as the overall and common feel of the environment as they are available to the entire population.
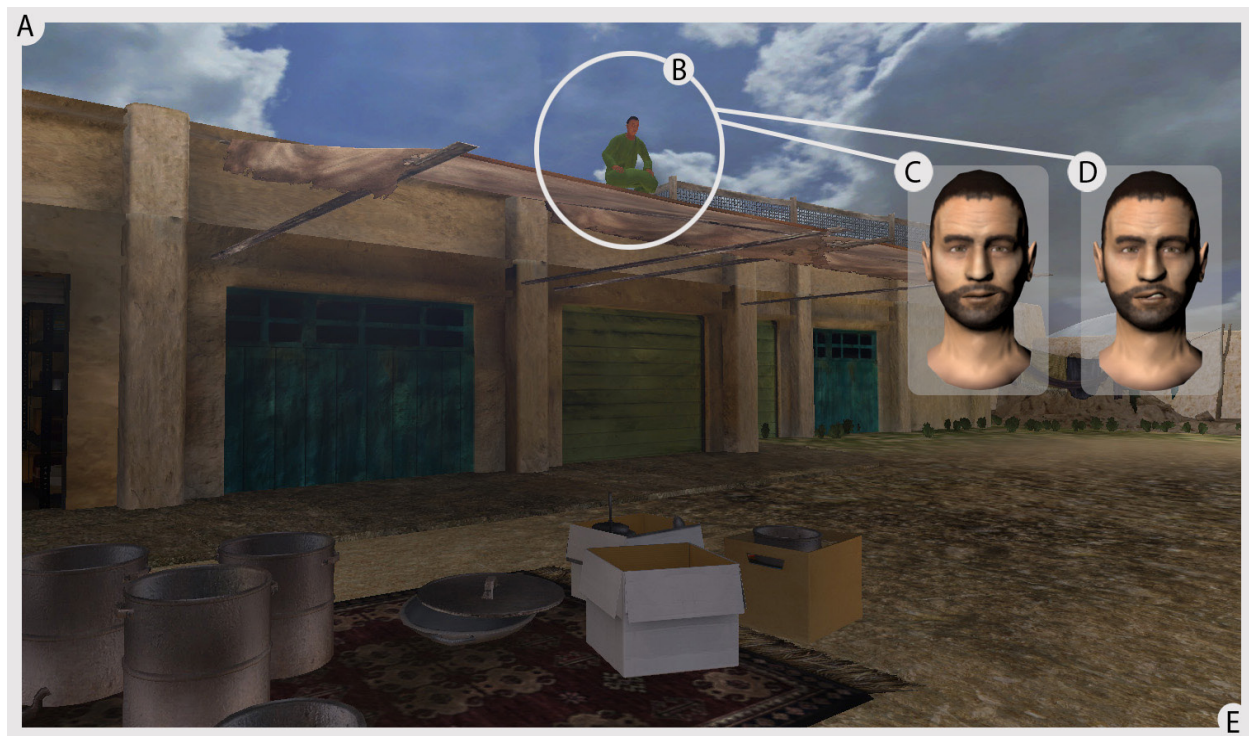
A transformational invariant is change occurring in or to an object or a collection of objects in the environment. (Michaels, Carello, 1981) A micro-expression, for example, is a transformational invariant that describes a brief change in a facial expression.

Affordances, in their simplest form, are the behavioral implications directly gathered by the player from

structural and transformational invariants attached to static objects and places in the environment. These elements can communicate information based on a player's perception and therefore might be different between users. For example, a bombed-out building represents more than charred rubble; it delivers information of a previous action and, depending on the perceiver's intention, location, culture and personal background, its structural invariants afford for various actions such as hiding, exploring or avoiding.

Finally, facial and micro-expressions cannot be presented as discrete representations, unconnected from the synthetic living-world environment in which NPCs exist. Instead, they should act in accordance with the affordances and invariants of the environment as defined above. This could be described as emotional flow. Emotional flow transitions a character's emotional equilibrium and may be a combination of facial and micro-expressions as well as other types of nonverbal communication. Figure 2 below summarizes these concepts.

Using the ideas presented above, this paper will discuss the technical challenges and solutions for creating visual representations of facial and micro-expressions. The paper begins with a discussion of how training in nonverbal communication can assist warfighters in distinguishing between emergent threats and non-threats. An overview of the gaming industry's state-of-the-art for developing facial and micro-expressions is presented. For example, new commercial software allows game developers to stream video capture of an actor's performance onto the face of a 3D model, producing high-fidelity, life-like animations of human emotion in videogame characters. The paper then concludes with a discussion of one approach to developing facial and micro-expressions in the First Person Cultural Trainer (FPCT), a 3D game-based simulation sponsored by TRADOC G2 Intelligence Support Activity that trains cross-cultural decision-making. The strategy for FPCT is based on five criteria identified as essential in developing cultural, psychological and behavioral complexity within virtual humans and synthetic societies. The criteria include: the need for composable training scenarios that are not dependent on actor

Figure 2. This image illustrates nonverbal information in a synthetic environment. The overall feel (A) is displayed through structural invariants – the combination of environmental objects. Change occurring in or to one object or a collection is a transformational invariant. Examples are movement by the man on the roof (B), changes in his facial expressions (C), or brief display of a micro-expression (D). The transition from state C to D is emotional flow. The entire environment indicated by (E) includes meanings attached by the player. To some, this scene might seem dangerous. To others, it may feel typical.

performance; the desire to develop modules for use across multiple technology platforms; the ability to vary the appropriate level of visual fidelity depending upon training objectives; the ability to provide in-game and after-action training feedback and assessments; and the goal to develop end-user tools for subject matter experts (SMEs). Whereas new technology can provide hyper-realistic facial and micro-expressions, the strategy developed for FPCT trades off some visual realism to support composability and other objectives.

SIGNIFICANCE AND CHARACTERISTICS OF NONVERBAL COMMUNICATION TRAINING

Many researchers posit that the majority of communication between people is nonverbal. Nonverbal communication encompasses a variety of different modalities. Gestures, or haptic touch, is one means. Posture is another. Finally, facial expression and movement, and eye contact are critical to understanding nonverbal communication. Although a great deal of nonverbal communication is based on idiosyncratic and sometimes arbitrary symbols which can vary from region to region or culture to culture, the vast majority of such communication is remarkably universal. Charles Darwin (1872) was perhaps the first to recognize this quality in his path-breaking work, *The Expression of the Emotions in Man and Animals.* (Darwin, 1872) In more recent times, the work of Paul Ekman is generally cited as contemporary evidence of the universality of facial expressions. Ekman links facial expressions – and specifically micro-expressions – to six universally human and well-established basic emotions—anger, fear, sadness, enjoyment, disgust and surprise (Sabini and Silver, 2005). There are other candidate basic emotions, like excitement, guilt and shame, but only six well-established ones (Sabini and Silver, 2005).

Nonverbal communication is particularly relevant in high-velocity, time-pressured situations imbued with threat, such as today's irregular warfare environment. Warfighters, much like police officers, must increasingly develop perceptual shorthand to reveal emergent threats from non-threats and to assist in determining direct, effective action. This perceptual shorthand is phenomenological and relies heavily, if not almost exclusively, on nonverbal appraisals. Phenomenology essentially is a study of how people typify behavior through pattern analysis. The cues investigated may be interactional, but they can be measured and quantified

and then assessed in dynamic environments. Although phenomenological interactionism is considered by some to be subjective, the ability to quantify things like body lean, arm angle, eye gaze duration, blinking, feet shuffling, arm swinging, degree of physical co-presence, and bodily stance make it abundantly empirical. The strategy in FPCT is to create a phenomenological training environment through the use of high-fidelity NPC facial and micro-expressions, as informed by the work of Ekman, and through environmental representations of invariants and affordances as defined above. The next sections provide a discussion of facial and micro-expressions and their connection to communicating emotion and intent in virtual characters and synthetic environments.

MICRO-EXPRESSIONS: A LINK TO EMOTION

As shown in Figure 3, micro-expressions are particularly informative in interpreting nonverbal communication, but also challenging to model in synthetic environments. One reason for this is that micro-expressions are strongly linked to emotion. As stated above, Ekman links micro-expressions to six universally human and well-established basic emotions—anger, fear, sadness, enjoyment, disgust and surprise (Sabini and Silver, 2005). Ekman argues that the six core emotions meet a number of criteria, including distinctive presentation, appearance in other primates and occurrence without conscious effort (Sabini and Silver, 2005). Ekman also asserts that each basic emotion is surrounded by a "family" of closely related ones (Ekman, 1999). For example, sadness might be related to states like feeling depressed or sullen. The six established emotions have unique micro-expressions associated with them. These expressions typically occur rapidly and are brief in duration. These expressions at times are blended to express multiple or conflicting emotions. Some work has been done exploring how simulated emotion might be translated into micro-expressions in a virtual human (Schaap, 2009), though this research does not account for the influence of culture and context.

Any approach to understanding or simulating micro-expressions must represent two types of emotion: felt emotion, which is experienced by the individual, and expressed emotion, which can be affected by personality, culture or context (Ekman and Keltner,

2003). This two-dimensional definition is especially useful when one notes that emotional expressions can be masked or faked to some degree, such as during an attempt to deceive another, or blended when a person is in an emotionally conflicted state (Ochs, Niewiadomski, Pelachaud and Sadek, 2005). When the nonverbal indicators conflict with verbal communication this might indicate deception. The literature on "deception clues" is rich and varied. The conventional wisdom that emerges from this research is that verbal cues may profess truth, but nonverbal cues can cause deception to "leak." Liars who attempt to compensate for the leakage may make their deception more glaring. Although there is debate in the literature as to which nonverbal movements indicate leakage of deception, researchers have concluded that distinctions can be made, they are empirical, and they can be quantified with appropriate data collection techniques (Vrij, 2008).

The strategy with micro-expressions is to communicate a greater degree of nuanced emotion and deception, and the development takes into consideration three major physiological factors:

  • Proxemics: how the NPCs manipulate, navigate and perceive the space around them. Space in nonverbal communication can be disaggregated into four distinct components: intimate space, personal space, social space and public space.

  • Kinesics: the use of posture, gesture, stance and movement in communication. Posture, for example, can signal a multitude of possible interactions and can evince levels of hostility or receptiveness to communication. Factors such as the angle of body lean, the degree of bodily openness to a speaker, and the configuration of fingers, hands and arms are all potentially relevant in this regard. Because these signals are nonverbal, they are communicated instantaneously—which is precisely why nonverbal communication is so efficient in conveying real-time information.

  • Oculesics: Eyes have long been considered windows to the soul, and much can be intuited about inner motives from the study of eyes. Variation in the extent, nature, frequency, intensity and duration of eye contact, for example, betrays a variety of emotions, such as indifference, fascination or hostility.

The facial and micro-expressions of the virtual agents reflect emotions that are elicited by three forms of stimuli: cultural rules, interactions with the player and other NPCs, and invariants and affordances in the synthetic environment. The next section discusses
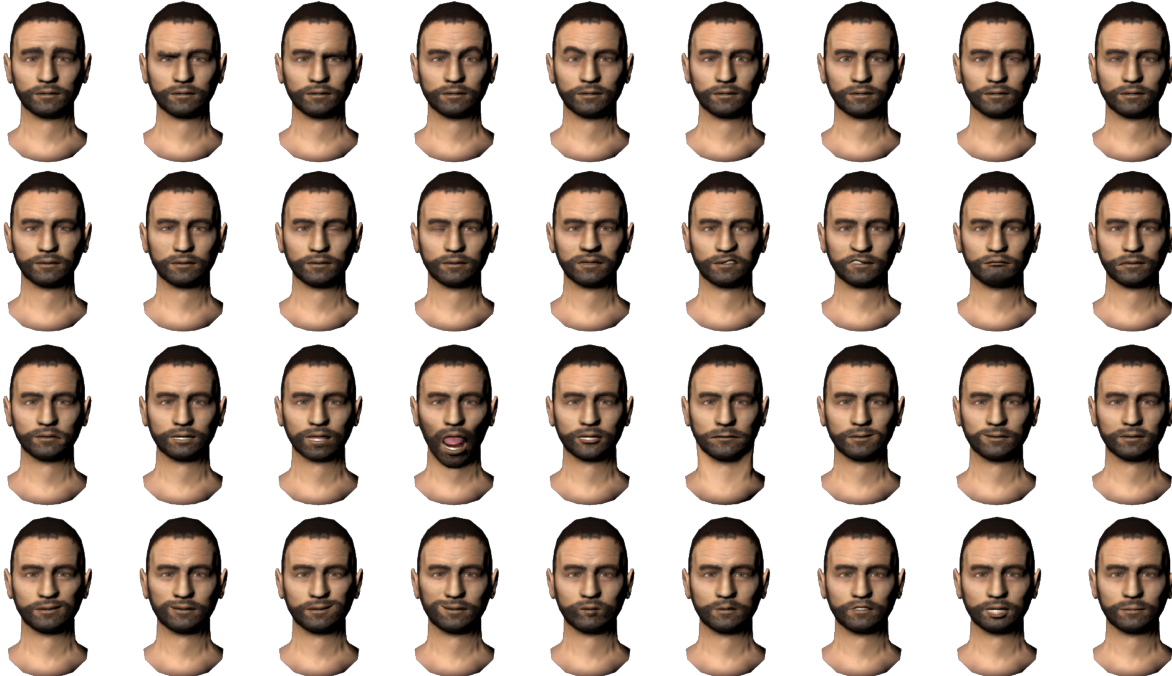
Figure 3. While at first glance the images above may all appear to be the same, if one looks closely, each face reveals a slightly different micro-expression related to emotion.

advantages and disadvantages of the state-of-the-art technology which focuses on enhancing the realism of both facial and micro-expressions.

NEW LEVELS OF REALISM

Overall, motion capture systems continue to improve the capability to capture subtle facial and body gestures that are critical to generating micro-expressions and the corresponding synthetic environments. However, in the past few years, several new facial animation approaches have been developed to generate accurate facial animation data onto virtual characters. In particular, two new systems focus on using video capture of actor performances to produce realistic facial animations -- MotionScan technology developed by Depth Analysis for Rockstar's new videogame L.A. Noire and Image Metrics' Faceware software and pipeline. This section discusses the basics of widely used traditional optical motion capture methodology and these two new state-of-the art approaches.

Optical Motion Capture Systems – The Basics

Optical motion capture is currently the most widely used facial motion capture solution for games and movies. The ability of the system to create a complex facial marker set on an actor allows for accurate facial capture of a given performance. The cameras in optical motion capture systems continue to improve in their capability to capture subtle facial and body gestures. A leader in the motion-capture industry, Vicon, utilizes cameras that differentiate between high and low - contrast on a gray scale. LED lights are strobed onto the capture area, where reflective markers placed on the motion capture actor reflect the light back. The camera captures the reflectivity of the markers in order to give positional data to the computer. This allows for the creation of a 3D scene containing the motion of the actor.

At this time, the newest motion capture camera that Vicon has available is the T160 series. The camera has a 16 megapixel resolution, which allows the smaller facial markers to be seen from a traditional wall-mounted camera setup. Newer Vicon cameras have made improvements over older models by increasing the pixel resolution allowing facial markers to be seen

from a distance. (Vicon T-Series, n.d. Para. 1-12.) One disadvantage of optical motion capture systems is the large amount of processing that must be done after capture in order to place the data onto a 3D character. This data cleaning invariably removes some of the original performance and replaces it with an animator's interpretation.

MotionScan

As mentioned above, MotionScan Technology was developed by Depth Analysis for the new Rockstar Games' videogame, L.A. Noire. MotionScan technology uses an array of cameras positioned around an actor's head to capture a performance from multiple angles simultaneously. With multiple angles and cameras, a full 3D mesh can be created from the live-action performance. This process allows MotionScan technology to turn a real-life performance into a virtual one with one-to-one accuracy. No markers, reflective paint or any other form of makeup, as found in traditional optical motion scan systems, need to be applied to the actor. The actor simply sits down in a chair and gives a performance, which is captured for use in the eventual game development. (Stuart, 2011) Figure 4 below is indicative of the type of production possible with MotionScan Technology.

The process is fairly simple. The data that is collected by the capture does not need to be cleaned by an animator and is therefore not altered as occurs in traditional motion capture. However, with current MotionScan technology, an actor's performance cannot be retargeted onto a different character. The performance can only be used on the mesh created by the capture session. This means that each capture can only be used for a single character. It also means that the character in game will look almost exactly like the actor giving the performance. (Stuart, 2011)

MotionScan textures are streamed onto the virtual character's face as video, which requires considerable hardware processing speed. In an interview, developers explained that, to optimize rendering of the video on all of the current generation of gaming consoles, no more than three virtual characters in L.A. Noire can be talking at the same time. To accommodate for this, other virtual characters face away from the camera and do not talk. (Stuart, 2011)

Figure 4. MotionScan, which was developed for use in the development of the L.A. Noire videogame, streams video of an actor's performance onto the face of the virtual character. The image above illustrates an in-game scene developed with this technology. This game was developed by Team Bondi for Rockstar Games. L.A. Noire screenshot courtesy of Rockstar Games. All rights reserved.

Image Metrics

Image Metrics technology has been used in numerous commercial videogames, such as Grand Theft Auto IV, Assassin's Creed II, NBA 2K11 and Halo Reach. As illustrated in Figure 5, Image Metrics uses a simple capture system using video from any camera. This simplifies the capture process and allows it to be done almost anywhere. The results of Image Metrics animations are accurate to the performance given on the video. (CGSociety Feature Article, n.d. Para. 1-13.) Image Metrics' process does not require the rigs to be similar. Any performance can be transferred/retargeted to any type of rig or character. The rigs can be full blend shapes, bones or a combination of both. Image Metrics' process is also faster than the traditional key-frame approach in terms of time spent both animating and smoothing out transitions between the extreme animation points in an actor's performance. (FaceWare, n.d. Para. 1-2.).

Unlike the MotionScan technology, the quality of Image Metrics' final animation depends largely on the skill of the animators. This is because the process for Image Metrics requires animators to create poses using facial rigs that match the most extreme gestures performed by the actor in the reference video. This process requires reference video to be sent to Image Metrics for in-house processing before a final animation is returned. Slight tweaks to the poses that the artist sets may be needed to influence the entire performance. But video or poses do not need to be sent back to Image Metrics for iterations. (CGSociety Feature Article, n.d. Para. 1-13.).

Facial capture technology that incorporates video, such as that of MotionScan or Image Metrics, produces a great degree of visual realism, and Image Metrics' technology does allow artists to change the performance to their needs. An approach that uses blend shapes, discussed in the next section, supports several development goals, including composability and flexibility, while still producing realistic facial and micro-expressions.

The FPCT Approach: Blend Shapes

Gameplay in FPCT focuses on the effects of player interactions on the emotions of NPCs. Conversation, gestures and events can influence NPC emotions positively or negatively, which in turn influence NPC mood and cooperation levels. The player's goal is to establish trust through culturally correct interactions, thereby enhancing the player's ability to discover mission-critical information or golden nuggets.

As mentioned earlier, FPCT development is based on five major criteria. First, developer tools have been designed to allow a high degree of composability to dynamically generate variable cultures and subcultures, each with its own variable sets of visual representations of emotion. This allows for the generation of culturally accurate micro-expressions in multiple NPCs as well as invariants and affordances from any type of environment. Second,
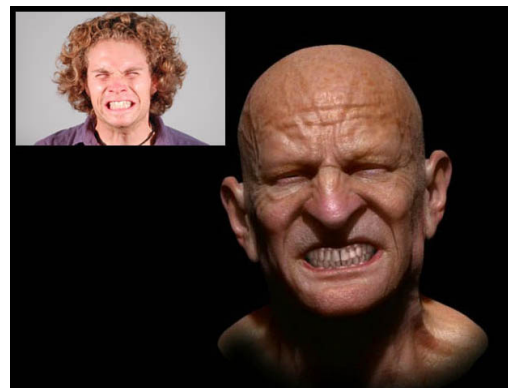
Figure 5. Image Metrics software creates facial expressions from video of an actor's performance. The software does not require facial markers such as those used in traditional motion-capture, but rather is based on video.

the development is designed to be platform independent, meaning that it is not tied to any one videogame system or engine. This requirement could restrict use of resource intensive technology which might not work optimally on mobile platforms, for example. Third, the development is being designed to accommodate varying levels of 3D fidelity, depending upon the training objective. For example, if the training objective is to analyze facial and body gestures of a culturally accurate NPC for signs of deception, then NPC micro-expressions must achieve a level of realism that transfers the knowledge. Fourth, in-game feedback, or during action review, is an important aspect of the training simulation. Therefore, the player must be able to receive feedback based on the information communicated through all the criteria set forth – facial and micro-expressions, invariants, affordances and emotional flow. Last, the FPCT development philosophy includes a goal for end-user tools for SMEs so they may design immersive virtual scenarios based on their real-world experiences.

Given these criteria, development work in FPCT is focused on using blend shapes to create facial and micro-expressions. Blend shapes are deformations on 3D models that are created from the topology of an already existing mesh with the intention of creating a new representation. This is achieved by manipulating the vertices on a duplicate mesh into the desired shape. The new representation can be blended between the original shape and the newly created model linearly, allowing morphing between the two models to any degree.

In general the advantages of blend shapes are their ability to be hand-crafted by artists. Their flexibility lies in the fact that they are created by vertex manipulation, meaning that the complexity of the blend shape is only limited by the number of vertices on the mesh and the ability of the artist creating the shape. They are also widely supported by most game engines. In FPCT, blend shapes are advantageous because they allow us to break expressions into multiple pieces. This gives us the flexibility to create expressions during run time instead of having to have a prerecorded performance for all interactions that take place in the game. This methodology also allows us to have high-fidelity expressions, although they may be slower to create than in the MotionScan or Image Metrics process. In general, blend shapes afford the flexibility needed to create a character that can react to the environment dynamically instead of through a scripted scenario.

The blend shapes approach also allows us to separate emotional representations into upper and lower facial expressions. As illustrated in Figure 6, this allows us to visually represent a conflict between felt, or elicited, emotions and expressed emotions, which are influenced by a virtual agent's goals, personality and culture. Such a conflict in emotions could represent, for example, an attempt to mask a felt versus an expressed emotion, as in deception. The design of this emotional model draws on Ekman's concepts that the upper face is harder to control. Thus, the upper face shows more felt emotion, while the lower face displays more expressed emotion.

EMOTIONAL FLOW

Blend shapes require no bones, instead using the changes in vertices on the actual mesh of the 3D character to create different expressions on the face. This allows for greater artist, and ultimately SME, control over what expressions look like. With this flexibility, FPCT's NPCs can display complex expressions that portray multiple emotions, as well as do full transitions from one emotion to another. This allows for the simulation of quick flashes of unintentional emotion as well as the ability for the NPCs to transition dynamically in-game from one emotional equilibrium to another in response to varying intensities of stimuli. This is called visual emotional flow, which, instead of being attached to a pre-recorded performance, is generated real-time during gameplay. The concept of emotional flow can


Figure 6. Through the use of blend shapes, NPCs can simulate blended micro-expressions, such as the combination of a felt emotion like sadness – conveyed in the upper part of the face – with an expressed emotion such as happiness – communicated in the lower portion of the face.

be explained best by using the acoustical concepts for resonance, damping, and dynamic profile.

The emotional flow of NPCs is an oscillating system. The temporal manifestation of the emotional flow is based on variations around a state of equilibrium, which, instead of being fixed and attached to a pre-recorded set of snapshots, is generated real-time during gameplay. For example, the visual display of happiness does not consist of pasting a grin over the character's mouth, but in organizing the flow of muscular events materialized by blend shapes according to the character's actions and responses to the stimuli from the environment. This oscillating system responds to stimuli in two ways: by resonating to them or by damping them. The system is similar to an acoustic resonator: when the stimuli are in sympathy with the state of the system, the system integrates and transfers more of the energy and produces a greater response. Conversely, the system damps the energy of the stimuli by not transferring it.

As shown in Figure 7, an emotional flow model that integrates the display of micro-expressions can be described by an acoustical model of a dynamic profile known as the ADSR envelope: attack, decay, sustain and release. ADSR is a model used in sound synthesis to replicate the dynamic profile of the sounds of natural musical instruments. This model represents the four phases of the response of the acoustic oscillator, i.e. the musical instrument to a mechanical stimulus. The "attack" represents the initial energetic input: depending on the type of instrument, it can be sharp, explosive, progressive, or continuous. The "decay" expresses a partial drop in intensity from the peak and represents the integration of the energy brought by the attack to the body of the instrument. The "sustain" phase corresponds to the main body of the sound after the transient "attack-decay" sequence and represents the resonant mode of the instrument. The release represents the dissipation of the energy and the end of the resonating mode. The release expresses the transition back to immobility, or in the emotional flow model, the transition to a new emotional equilibrium for the virtual character. This model provides a theoretical foundation for programming variable blend shapes into micro-expressions that can be altered by time and intensity and reflect the relationship between the stimuli and the current state of emotion or flow of emotion and its visual display.
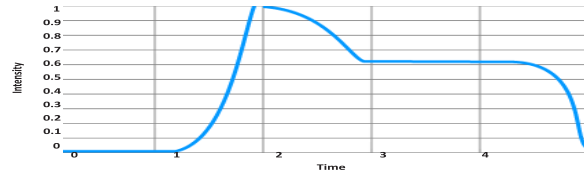
Figure 7. The ADSR (attack, decay, sustain and release) envelope models micro-expression display. Time t = 0 to t = 1 represents emotional equilibrium. Time t = 1 to t = 2 represents the attack, a rapid increase in emotional intensity. Time t = 2 to t = 3 represents the decay, as emotion wanes. Time t = 3 to t = 4 represents sustain, a flattened but still heightened emotional state. Time t = 4 to t = 5 represents release, the natural intensity decay.

A FINAL COMMENT ON THE IMPORTANCE OF INVARIANTS AND AFFORDANCES

Today's game engines are capable of handling hundreds of thousands of polygons, making them capable of producing virtual characters and environments that almost mirror real world counterparts. However, fidelity in virtual humans is not defined solely by the visual quality of a re-created human form. Fidelity is also an issue of understanding the foundational principles that have defined character animation for more than a century. The principles defined by Disney animation draw from traditional theatre and remain essential for creating virtual characters that grasp an audience's attention and evoke emotional responses. (Frank, 1981) The idea that emotion is expressed not only from characters themselves, but also from their environment, goes back to the very beginning of early animation.

Disney artists developed a system of guidelines for capturing motion, and although most of these guidelines apply to traditional 2D animation, there are two that are particularly valid when attempting to increase the fidelity of 3D animation as well. The first is anticipation. This principle involves creating counter motions that allow a viewer to follow or anticipate the action of a character on screen. Anticipation allows the viewer to remain engaged while alerting about actions that are about to take place. Anticipation is not always obvious and may not always reveal a character's intentions, but rather is a device used to inform the viewer. The "illusion of life" concept identified by Disney animators states that

"anticipatory moves may not show why [a character] is doing something, but there is no question about what he is doing." (Frank, 1981) Anticipatory motions serve to counter balance the primary action taken by an on-screen actor. For example, before a pitcher releases a ball he winds up alerting the audience to his intent to throw. If the same scenario were recreated in a game environment without the anticipatory motion, the viewer would lose the information required to remain engaged with the characters action.

Another Disney 2D principle that remains relevant in 3D animation is exaggeration. Disney animators found no matter how accurately they would reproduce an animation it seemed to lack the essence of the action. This essence is a key element required for engaging the viewer -- making them feel what was about to take place on screen. (Frank, 1981) This became more of a problem as animations became more subtle and nuanced, as is being done in FPCT. Nuanced animations are possible with current technologies although there is still a need to push or exaggerate character animations for them to read accurately. This exaggeration is not a distortion of shape or visual elements but an emphasis on a particular expression. Even when used in virtual simulations, virtual characters must be thought of as actors. Increasing the visual fidelity of a character helps the player identify specific features of that character. However, creating motion that is readable is what allows the player to connect with that character. These two principles are particularly significant to the expression of emotion through facial gestures and to events occurring in the virtual environment.

SUMMARY

Today's irregular warfare environment requires greater training for interaction with multi-cultural, non-combatant populations. Game-based simulations designed to meet these training needs must have tools for composing a wide range of expressions that offer the appropriate degree of animation and visual fidelity necessary to display nuanced nonverbal information. This nuanced nonverbal information should be displayed in virtual characters, through micro-expressions, and in the environment, through invariants and affordances. Micro-expressions can be combined and blended to create life-like transitions of emotional equilibrium. Such transitions can be considered emotional flow. One approach to generate such high-fidelity visual

representations is through the use of blend shapes. New commercial software allows game developers to stream video capture of an actor's performance onto the face of a 3D model, producing extremely realistic animations of human emotion. The use of blend shapes, however, provides greater flexibility and composability, criteria essential to a development strategy that emphasizes cultural, psychological and behavioral complexity in synthetic worlds. Other key criteria in such a development strategy include the capability to develop training scenarios for use across multiple technology platforms; the ability to apply the appropriate level of visual fidelity depending upon the training objectives; the ability to provide in-game and after-action training feedback and assessments; and ultimately, end-user functionality for SMEs.

ACKNOWLEDGEMENTS

REFERENCES

Aiken, Jim. (2004). *Power Tools for Synthesizer Programming*. Backbeat Books, San Francisco, CA. © 2004 ISBN: 0879307730

CGSociety: Society of Digital Artists Feature Article. "An Image is worth a thousand words, but the face says them all. CGScociety explores Image Metrics." (n.d.). Retrieved June 16, 2011, from http://features. cgsociety.org/story_custom.php?story_id=4480

Darwin, Charles. (1872). *The Expressions of the Emotions in Man and Animals*. London: J. Murray.

Davison, John. (2010). "The Most Impressive Thing I Saw at E3." Retrieved June 16, 2011, from http:// www.gamepro.com/article/news/215667/the-most-

impressive-thing-i-saw-at-e3/

Ekman, Paul. (1999). "Chapter 3: Basic Emotions." (1999) *Handbook of Cognition and Emotion*. Ed. T. Dalgleish and M. Power. Sussex, U.K.: John Wiley & Sons. Retrieved June 13, 2011, from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.123.1143&rep=rep1&type=pdf.

Ekman, Paul, and Dacher Keltner. (2003). "Introduction: Expression of Emotion." *Handbook of Affective Sciences*. Ed. R.J. Davidson, K.R. Scherer, and H.H. Goldsmith. New York: Oxford UP. 411-14. Retrieved June 13, 2011, from http://www.paulekman.com/wp-content/uploads/2009/02/Intoduction-Expression-Of-Emotion.pdf.

FaceWare. Facial Animation Software. (n.d.). Retrieved June 16, 2011, from http://www.image-metrics.com/pdf/FaceWare%20Overview.pdf

Kong, Jennie. (2010). Groundbreaking New MotionScan Technology set to Redefine 3D CGI Performances. Press announcement retrieved June 16, 2011, from http://depthanalysis.com/downloads/pr/depth_analysis_announcement_20100301.pdf
L.A. Noire Trailer – Behind the Scenes. Retrieved June 16, 2011, from http://www.youtube.com/watch?v=ZY7RYCsE9KQ

Michaels, Claire F., and Carello, Claudia. (1981). *Direct Perception*. Prentice-Hall, Inc. P. 42

Ochs, Magalie, Radoslaw Niewiadomski, Catherine Pelachaud, and David Sadek. (2005). "Intelligent Expressions of Emotions." Affective Computing and Intelligent Interaction. Vol. 3784. Springer, 2005. 707-14. Lecture Notes in Computer Science. Retrieved May 9, 2011, from http://www.springerlink.com/content/7007280gtq412j0h/.

Porter, Stephen, and Leanne Brinke. (2008). "Reading between the Lies: Identifying Concealed and Falsified Emotions in Universal Facial Expressions." *Psychological Science* 19.5 (2008): 508-14. EBSCO. Retrieved June 22, 2011, from .

Richardson, Stuart (2011). The art of lying: LA Noire. Retrieved June 16, 2011, from http://www.develop-online.net/features/1166/The-art-of-lying-LA-Noire

Russell, James A., and José Miguel Fernández-Dols. (1997). *The Psychology of Facial Expression.* Cambridge: Cambridge UP.

Sabini, John, and Maury Silver. (2005). "Ekman's Basic Emotions: Why Not Love and Jealousy?" Cognition and Emotion 19.5: 693-712. Retrieved March 24, 2011, from http://ivizlab.sfu.ca/arya/Papers/Others/Emotions/Ekman%20-%20Why%20Not%20Love%20and%20Jealousy.pdf.

Schaap, Robert. (2009). "Enhancing Video Game Characters with Emotion, Mood and Personality." Thesis. Delft University of Technology, 2009. Retrieved March 9, 2011, from http://graphics.tudelft.nl/Game_Technology/Schaap?action=AttachFile&do=get&target=Thesis_Schaap.pdf.

Schutz, Alfred. (1967). *The Phenomenology of the Social World.* Northwestern University Press. Evanston, Illinois. Originally published in German under the title Der sinnhafte Aufbaus der sozialen Welt. © 1932 by Julius Springer, Vienna; © 1960 by Springer-Verlag, Vienna. English translation © 1967 by Northwestern University Press.

Thomas, Frank. (1981). *Disney Animation: The Illusion Of Life.* New York: Abbeville.

Vicon T-Series: The world's next generation motion capture camera. (n.d.). PDF retrieved from website: http://www.vicon.com/products/documents/TSeriesSedition.pdf

Vrij, Aldert. (2008). *Detecting lies and deceit: Pitfalls and opportunities.* Chichester, United Kingdom: Wiley.