Raising the Aesthetic Quality of Character Interaction in Cinematic Videogames

Christopher Maraffi and Arnav Jhala
University of California, Santa Cruz
Santa Cruz, CA 95064
{topherm, jhala}@soe.ucsc.edu

ABSTRACT

In this article we review the state of the arts in virtual character control research, proposing a critical technical practice for making aesthetics a central concern. We argue that increased acting logics and affordances for playing a role in popular videogames are pushing character interaction towards the performing arts domain. To align procedural play with viewer preferences for skilled performance in media, we propose an expressive role play model based on classical acting. Our performative ontology draws from theory and practices in the arts, artificial intelligence, and psychology to provide a firm theoretical grounding for our approach. We use our model to analyze several examples of playable media, and describe initial experiments to infer the aesthetic quality of poses in motion capture data as a first step towards generating expressive character features.

"Art is not, as is said, an imitation of nature."
- Francois Delsarte (Stebbins, 1902).

1. Character Interaction as Artistic Performance

In popular media professional artists set aesthetic quality standards for performed and animated characters. As media consumers, we judge character quality by how much affect we experience as spectators. Skilled character artists routinely have a strong emotional impact on audiences, whether through live acting or animation. Who hasn't shed an involuntary tear watching movies like Cameron's Titanic or Disney's The Lion King? We showed a large university class video of actor Edward Olmos portrayal of Adama's reaction to the death of Starbuck in Battlestar Galactica. After Adama is told how Starbuck seemingly commits suicide, the scene cuts to him drinking alone in his cabin where in a fit of grief he destroys a miniature wooden sailing vessel constructed over the previous season. Olmos' acting is very physical, with little dialogue, and almost entirely improvised according to DVD commentary (Moore et al, 2007). Afterwards, we asked if anyone unfamiliar with the series had been affected by watching the scene. Quite a few hands were raised. This hints at what professional artists do to create an aesthetic experience for spectators: they capture our attention and emotionally move us with how they animate a character's body, even with minimal narrative context. It behooves us to therefore ask why interactive characters in videogames have yet to consistently achieve a comparable impact, and how can we close the expressive gap between traditional and new media? In this article we propose that the player's experience of videogames relies on the artistic quality of interaction, which require character control methods that generate the visual aesthetics of cinema.

Popular videogames that feature narrative and character portrayal has increasingly moved towards cinematics, from non-interactive cut scenes to the latest research in procedurally generated story, lighting, and cinematography. Understanding cinema aesthetics for character control is critical for improving the player experience because embodied characters are the focal interest of scripts, lights, and cameras, with playing a role becoming the dominant interface for game interaction. In the Mass Effect franchise (Bioware, 2008-2012), for instance, the procedures of fighting, questing, puzzle solving, and resource management, are playable through the role of Commander Shephard. Players improvise a paragon or rogue character style by how they control their Shephard avatar when interacting with system controlled non-player characters (NPCs), which in turn modifies game generated content and future play. These minimal acting affordances improve the player experience through performativity, by determining cinematic scenes to be played, and through evolving character relations. Wardrip-Fruin introduced the term operational logics to explain how an "experience can be playable" (Wardrip-Fruin, 2005) in new media, and further expanded the concept to include families of logics (Mateas and Wardrip-Fruin, 2009). Popular videogames like Elder Scrolls: Skyrim (Bethesda, 2011) and Portal 2 (Valve, 2011) may vary widely in genre, narrative, and mechanics, but display an emerging family of performative logics related to playing a role. Acting logics like skeleton control and dialogue choices support embodied character interaction through dramatic role play in the performing arts tradition.

Avatars and NPCs in cinematic games function as semi-autonomous embodied interfaces for role play, like electronic puppets or animatronics on a movie set. Control affordances should support character acting, and although engineers design such tools, skilled performing artists have always been the preferred operators. Artistic skill is essential because the main purpose of character behavior is to affect spectators, which require subtle gestural techniques for capturing and holding viewer attention (Barba and Savarese, 2005). Without first engaging viewers, story information cannot be affectively conveyed to the player. Professional performers learn how to present their bodies to spectators in ways that make them visually interesting to watch. Though some have unique skills that can be attributed to individual talent, there are more general conventions taught to professional performers that pertain to the aesthetic quality of gesture. More indirect character artists, such as puppeteers and animators, employ similar techniques for animating bodies because they are also playing for an audience (Thomas and Ollie, 1981). Media consumers intuitively judge artistic quality because we are constantly exposed to the work of professional performers. The fact that we are willing to spend time and money to watch the pros on stage and screen indicates we recognize there is a quality

UCSC-SOE-13-01

difference between them and ordinary people, and shows our general preference for what they do. But researchers have yet to computationally study the connection between performer skill and audience affect, much less apply their methods to autonomous character control.

The artistic quality of character behavior impacts the aesthetic experience of videogames because players interact as both performers of avatars and spectators of NPCs. So we are approaching this problem from a computational aesthetics point of view, an area of research that has drawn from the arts and psychology. A significant body of work in this area has been in aesthetics of photography and virtual camera control for digital media. Computational aesthetic studies start with an initial model based on domain knowledge from the arts. Then controlled studies are conducted using statistical methods, which for us takes the form of aesthetic experiments on artistic gesture to infer spectator perception of performance quality. Such knowledge can inform intelligent character control to generate more expressive interaction between avatars and NPCs in games. This article details our performative approach, starting with a critique of dominant artificial intelligence (AI) practices that attempt to simulate realistic human behavior through intelligent virtual agents (IVAs). We then define an alternative research space by mapping a performative ontology from AI to performing arts metaphors, with which we formulate a definition for expressive role play in videogames that is consistent with embodied performance in the arts. Our ontology focuses on the dual player experiences of flow and affect, relating performer skill to the spectator perception of aesthetic quality. We close by showing how we are applying our performative model in computational aesthetics experiments for virtual character control.

2. Aesthetic AI Critical Technical Practice

The dominant approach in AI research has been to simulate realistic or natural human behavior, not the artificial behavior of performing artists. Natural everyday behavior is utilitarian and efficient, while performative behavior cannot be entirely natural because the performer employs artifice to manipulate the attention of spectators. Professional physical performers learn their craft through years of training, and according to performance theory, embody gesture conventions that require subtle body control, and extra effort to perform (Barba and Savarese, 2005). Studies have shown that, contrary to expectations, skilled actors do not actually generate natural looking behavior when simulating emotions (Krahmer and Swerts, 2008). Perhaps non-performers don't recognize that the pros are doing anything unusual because their gestural codes have become naturalized to media literate consumers. But we intuitively recognize that they are more appealing and expressive to watch than everyday people, and we generally prefer their behavior over untrained amateurs in media. So why wouldn't we want interactive characters in media to display similar features and have the same affect on players?

Game AI for character interaction should be considered a different class of behavior than real human behavior, and IVAs that generate character behavior should be designed differently than utilitarian or helper agents. All cinematic videogame characters fall into this class, and require a performing arts model of behavior to align with aesthetic conventions and player preferences for skilled acting. But the arts in general have been relegated to the margins of mainstream AI research, with psychology and linguistics methods dominating the field. In IVA2012, there was only one paper related to the arts (Talbot and Youngblood, 2012). With the exception of some notable cases (Bates et al., 1993; Mateas, 2004; Seif El-nasr, 2005; Magerko et al., 2010), AI researchers have not considered physical acting models from the performing arts for embodied agent behavior. Of these, work done by the Oz and Digital Improv projects to study physical actors has since stalled, while Façade and Mirage implemented more literary drama theory than embodied acting technique. Very few have computationally studied what skilled performers do in electronic media or motion capture data (mocap). Animation based approaches to derive gesture style have treated artistic performance as pre-processed data to be manipulated and blended [Oshita, 2008], rather than artistic behavior rules to be learned. Most approaches to procedurally generated character animation have focused on psychological models of natural gesture [Kipp et al., 2007], or realistic physics-based tools (Horswill, 2008). Procedural animation tools for character behavior have been explored, but have since stalled before realizing expressive acting (Perlin and Seidman, 2008). Sustained work has yet to be done to simulate the symbolic methods of embodied performers in the arts domain. To shift the performing arts from the margins to the center of AI requires a critical technical practice (CTP) to privilege aesthetics over realism in the IVA design of character behavior.

Agre introduced the concept of a CTP to counteract dominant AI assumptions he termed mentalist, taking an interactionist approach that privileged intentionality and embodiment (Agre, 1997). Character AI in videogames should of course be interactionist, but also needs to be artistic because a primary goal of interaction is to control the aesthetic experience of players through role play. Agre also focused on roles, but in a more utilitarian sense. His approach leveraged contextual knowledge to define procedural roles as relationships in a stable system of embodied cultural practices. Roles have inherent intentionality, because when an agent casts an entity into a role their relationship is instantly defined. Computationally, role interaction is more portable and efficient than relating to more objective identities, because maintaining a role requires less information to process. For instance, a bus driver's role is constrained to taking money, driving, and maybe answering questions about their route. Designing affordances for roles minimizes system complexity because of the principle of machinery parsimony, which says that agents should only be endowed with the minimum knowledge required to fulfill their function. This approach contrasts with traditional AI that seeks explicit representation and greater generality, but is consistent with playing a character role in the arts. For instance, actors often draw character features from stereotypes, using common tropes and props, but then improvise variations through art practices (Magerko, 2010).

It is no more necessary for actors to create a real personality than for movie set builders to create real starships. The arts only require the surface illusion of a persona or starship within the context of a scene, a much more tractable problem to simulate than multidimensional reality. Since all acting in cinema is filtered through a single camera view, the search space for generating affective character behavior can be reduced to an aesthetic relationship between actor and spectator. For any performer to be affective, even virtual ones, they must play to their audience. The history of theater and cinema show that character believability is dependent on this relationship, and that it doesn't require realism to achieve it. This underscores two play strategies that are coming together in videogames; the aesthetic play of cinema or video is merging with the procedural play of games. Traditional game genres show the dominance of procedural play in the utilitarian nature of the player's role, wherein shooters you shoot, with little regard for the aesthetics of the shooting mechanic. A CTP for aesthetics should raise performing arts style of play to at least equal status with procedural game play.

To implement a CTP, Agre first suggested a strategy of reversal to reorient generative metaphors from the margins to the center of AI. In our case we begin by replacing technical terms from procedural rhetoric with more expressive terms from the arts. For instance, operational logics can be replaced by compositional craft and unit operations can become act compositions. A composition implies a composer, which indicates an embodied artist. The point is to privilege metaphors that imply the performing arts tradition of embodied aesthetics over the mentalist AI tradition of disembodied code. We shy away from literary dramatic theory that reduces acting to text, because it is not sufficient for explaining what actors do to play a role. Scripts are always underspecified in theater, requiring that actors generate significant content through physical interaction and improvisation (Sawyer, 2005, p.46). Text may supply performers with a loose framework from which to start a characterization, but actors flesh out a role in practice with aesthetics in mind. Cognitive and literary AI research is important for generating narrative constraints, but is not sufficient for determining how an actor delivers the story to spectators. The generative aspects of embodied performance has yet to be seriously studied in mocap data, to find out what professional performers do to craft an appealing character representation.

While the first step in our CTP is a strategy of reversal, the second step is to address the old center in terms of the new margins. Agre, who centered interactionism as a counterpoint to mentalist AI practices, questioned how an interationist would explain the insides of bodies (Agre, 1997, p. 306). For the traditional AI practitioner, insides are their cognitive model for agent behavior and emotions, but for an interactionist insides function as a black box. Spectators of characters can only know what they perceive, since the insides of actors cannot be directly experienced due to the encapsulation of embodiment. Acting has been called the art of lying, and it is likely that even the most naturalistic method actor is aware that they are performing a simulation. But for a classical actor, the inside of a role is how the character is intended to be perceived by the spectator, through the interface of embodied gesture. Insides as the intentionality of a role are derived from acting conventions, which shape how the "scripted body" is performed and interpreted (Shepherd, 2006, p. 27). Actors functioning as black boxes may facilitate spectator belief by manipulating attention away from the control source to the performed character. A CTP for aesthetics therefore has to study how artistic methods for articulating a character's body can manipulate a spectator's attention.

3. A Performative Ontology for Interactive Characters

Procedural rhetoric is central to traditional AI practices, with generative metaphors that define the research space. Procedural terminology shows the background of researchers in psychology and linguistics, and influences how problems are formulated and approached. Since we want to create a bridge between system interaction and artistic gesture in videogames, we first map common procedural metaphors to arts methods. We intend these conceptual transformations to reveal overlapping areas of interest that may have been obscured due to boundaries between the sciences and arts. To begin, one phenomenon found in science and the arts is the frequent use of dual metaphors. Bogost pointed out generative dualisms in the humanities and informatics, which he traced from his system-unit operations dialectic through Heideggar's enframing-poiesis to Aristotle's dual nature of form-matter (Bogost, 2006). Agre found dual metaphors at the center of AI, tracing the mentalist abstraction-implementation approach to Descartes and Plato, and then substituted his own intentionality-embodiment reversal as the center of his interactionist CTP (Agre, 1997).

Similar dual metaphors can be found throughout procedural rhetoric; Zhu related intentionality and material agency to Heideggar's aboutness-throwness dualism (Zhu, 2009); Chesher focused on the invocation-avocation capabilities of interfaces to generate roles that define user identities (Chesher, 2003); Wardrip-Fruin pointed out a data-process distinction between static and dynamic data in expressive authoring for digital media (Wardrip-Fruin, 2009); Hecker considered structure-style decomposition as instrumental in recent graphic successes of media rich games, and suggested a similar approach to generating better AI character behavior (Hecker, 2008); Mateas incorporated a neo-Aristotelian poetics approach to designing system interaction as dramatic beats which function as an action-reaction pair (Mateas, 2004). Dualisms in the performing arts take the form of imitation-improvisation, where apprentices train by mimicking an experienced performer, and then develop individual style variations through practice. Hall characterized all form in theatre as a mask that shapes performance, and noted the dual relationship between poetic words and dramatic acts (Hall, 2000). Shepherd traced similar script-performance dualisms in theater from Diderot's ideal-manifestation to Roach's textual-physiological, noting his own discourse-biology dual metaphor to point out the opposition between the "scripted body" in dramatic discourse and the "polymorphous thinking body" in performance (Shepherd, 2006, p. 28). We add our own performer-spectator dialectic for videogames that corresponds to a player experience of procedural flow and aesthetic affect.

Organizing dual metaphors from abstract procedural rhetoric to embodied artistic practices shows a pattern of thought that brings a performing arts model into focus. We have inserted additional dualisms from the sciences of biology and semiotics, as well as performance and videogame theory, to smooth out the transitions between metaphors:

> {(Inside,Outside), (Form,Matter), (Ideal,Manifestation), (Abstraction,Implementation), (Data,Process), (Structure,Style), (System,Unit), (Global,Local), (Planner,Reactive), (Signified,Signifier), (Iconic,Indexic), (Aboutness,Throwness), (Pretext,Context), (Word,Act), (Textual,Physiological), (Discourse,Biology), (Code,Body), (Intentionality,Embodiment), (Replication,Variation), (Propagation,Innovation), (Cooperation,Competition), (Convention,Invention), (Invocation,Avocation), (Artificial,Naturalized), (Ghosts,Liveness), (Poetics,Dramatics), (Scripted,Interacted), (Imitated,Improvised), (Aesthetic,Procedural), (Affect,Flow), (Spectator,Performer), (Character,Actor), (Mask,Face), (Puppet,Controller), (Avatar,Player), (NPC,IVA), (Role,Play), (Media,Playable), (Video,Game)}

Notice there is an evolution from abstract concepts of disembodied code or text to automated system processes, which then transforms to embodied methods in the arts, and finally to videogame play as character control. By tracing through the left side of each dualism, we can better answer Agre's question about the insides of bodies, and understand the aesthetic nature of cinematic play. The right side shows how system processes relate to external acting methods in reactive media. One dualism in particular stands out towards the arts end of the transformation, between the aesthetic-procedural and spectator-performer metaphors: affect-flow. We are interested in the dialectic relationship between performer flow and spectator affect because it describes the dual nature of the player experience, and because both concepts are already topics of scientific research in the arts and videogames. Flow has been related to skilled performance in theater and games, and computational studies have been done on how affect relates to aesthetic quality. However, there has been little research on an explicit connection between performer flow and spectator affect.

The term flow was coined by Csikszentmihalyi after descriptions of creative activity studied in the 1970s (Csikszentmihalyi, 1975). Since then, the flow experience has been reported by performers in the arts, sports, and board games (Engeser, 2012). Csikszentmihalyi's flow model has been associated with interactionism, where motivation is in context to interacting with other embodied performers and the environment (Csikszentmihalyi, 2002). Researchers have documented consistent features of the flow experience across many performative domains, including videogames. A player in flow is associated with a feeling of agency, loss of ego, and a time distortion. A flow experience in game play has been related to designing a good ratio of procedural challenge to performance skill, producing system interaction that affords a strong sense of agency. Chen advocated system dynamic difficulty adjustment as a game design technique to keep players in the "flow zone" (Chen, 2007). In addition, we suggest that playing a character role provides similar conditions for the flow experience. A well-cast role sets clear boundary conditions for players to interact, while controlling an embodied avatar provides aesthetic feedback for a greater sense of agency. New acting logics that leverage the natural user interface (NUI) control of characters, such as through the Kinect and Wii, are moving game play towards the embodied methods of live performers. Therefore the experience of flow when playing a role in videogames becomes directly relevant to character acting.

Features common to the flow experience have close similarities to classical acting methods associated with Diderot, who argued that actors should only outwardly play emotion by "rendering scrupulously the external signs of feeling" (Sawyer, 2005 p.51), as well as ritual performance practices going back to the ancient Greeks. For instance, the experience of personal ego loss when engaged in a flow experience is considered a trait of skilled character acting by performing arts theorists (Craig, 1963), and also correspond to descriptions of semi-autonomous states or "double" consciousness experienced by acting induced trance states (Johnstone, 1992, p.152). Trance acting has performers divide their attention between manipulating their body like a puppet and projecting how their character is perceived by spectators. Practitioners have reported the experience of watching themselves perform from the outside. Greek masked theater evolved from ritual practices, as did the classical acting techniques of the first professional actors in the Commedia dell'Arte. Copeau taught trance improv using Commedia style masks, followed by Decroux, influencing a generation of physical actors (Leabhart, 1995). Murray has suggested that a Commedia model would be conducive for virtual character interaction because it combines stereotypes with improvisation (Murray, 1998 p.235). Flanagan showed that automatism techniques were practiced by the Surrealists to stimulate creativity in language and drawing games, negating thought and reason to facilitate improvisation, which she also equated to a flow state (Flanagan, 2009). Ritual control methods have also been associated with puppetry, where the puppeteer is physically and emotionally distanced from the character, but deftly projects an illusion of life into an artificial body.

Performing arts theorists have suggested that theater in general reflects this spirit medium model, with conventions shaping actors like a Commedia mask to channel the ghostlike images of iconic stereotypes forward in time (Hall, 2000; Rayner, 2006). Theater rhetoric in general is full of ghost in the machine metaphors that should be studied to understand how they can inform virtual character control. Classical methods treat the performer's body as a semi-autonomous acting instrument to be manipulated, while the dual awareness of actors in flow may afford better body control. This is how professional artists approached the problem of character performance for thousands of years, and we consider it a promising model for NPCs and avatars because players also divide their attention into performing and viewing during game play. We also recommend classical acting over more realistic methods because performers must project meaningful gesture through the medium of computation. The broad physical acting of Jim Carey tends to appear more expressive in mocap animation than the more subtle technique of Tom Hanks. The most successful mocap actor to date, Andy Serkis, is skilled in stage acting methods. We suggest that procedural performance in videogames should follow a similar aesthetic path as film acting, which initially adopted the exaggerated technique of late 19th century live theater. Our CTP against realistic agent behavior extends the arguments of innovative theater practitioners who first conceptualized acting as an art form, from the corporeal mime of Decroux to the Uber Marionette of Craig (Leabhart and Chamberlain, 2008; Craig, 1963, Maraffi, 2011). Strongly influenced by Asian theater practices like Indian and Balinese dance, they advocated more symbolic acting methods.

There was a western symbolic acting method that was popular in America and Europe at the turn of the century that shaped the aesthetics of early cinema. The Delsarte system of expression was a semiotics of gesture that segmented the actor's body into playable units to perform emotions as a series of poses. Though Delsarte never published, his modernist acting method was spread to the American stage by his student Steele Mackaye, who established the first acting school in New York city. Hollywood actors adopted Delsarte's method through the DeniShawn dance school in Los Angeles, where silent movie stars like Lillian Gish were sent by the studios for movement training. It is purported that Craig based many of his Uber-Marionette concepts on the dance technique of his lover Isadora Duncan, who was a celebrated Delsarte instructor. Some computational research on using aspects of Delsarte for agent generated behavior has shown promising results (Marsella et al., 2006; Nixon et al, 2010). A professional actor demonstrated Delsarte's method to the UCSC Immerse project group to determine expressive viability (Hollander, 2013). The stylized movements required a high degree of control to clearly articulate each body part in order, but the results were believable and affective when skillfully done. That the participants could clearly read the actor's intention suggests that Delsarte aesthetics are part of our media literacy, and warrants further study.

Videogame AI research on affect has been approached in the context of player experience (Yannakakis, 2009), and computational inference work has been done on viewer preferences for still image quality (Joshi, et al., 2011). But to our knowledge no player experience research has focused on the aesthetic quality of character interaction, and no computational work has been done to study if there is a direct relation between performer flow and spectator affect. Though a connection between flow and affect has not yet been established, there has been a correlation between flow and aesthetics, and between aesthetics and affect in flow research (Engeser, 2012). Flow has also been directly related to performance skill, and actors have been known to perform better in front of spectators, with acting skill being related to audience affect (Sawyer, 2005). Studies have shown that spectators describe the aesthetic experience of being affected by art in similar terms to how skilled performers describe the flow experience (Csikszentmihalyi, 1990), especially a loss of ego and time distortion. But the emotional impact experienced by spectators is not experienced in the same sense by performers in flow, which makes intuitive sense as it could interfere with procedural control. Researchers have noted that subjects are not highly affected while acting in flow, but often have strong feelings in retrospect.

As the basis of our performative model, we propose a new role play definition specifically for character interaction in cinematic videogames that makes an explicit connection between flow and affect:

- **Expressive role play is procedural flow with a performing body to generate artistic interaction that engages spectator attention through aesthetic affect.**

Our definition raises aesthetic play to an equal footing with procedural play, and is consistent with playing a role in the performing arts. Player flow implies improvisational performance skill that affords a strong sense of agency. Interacting artistically with a performing body implies affordances for playing an avatar like a semi-autonomous instrument, with artistic conventions built in through character role constraints. Skilled artists always perform with the spectator view in mind, so virtual character control should be designed to work with virtual camera composition to visually move players to an affected state.

Our performative model can be used to analyze the acting logics of a system, to check for expressive role play affordances. It moves artistically generated behavior to the center of AI concerns, with significantly different aesthetic goals than traditional AI approaches. Another CTP that offers a literary alternative to authoring system behavior is Wardrip-Fruin's expressive processing approach (Wardrip-Fruin, 2009). Since expressive authoring in playable media combines narrative with performance, our model compliments his because both provide intuitive affordances that are natural for artists (Wardrip-Fruin, 2009, p.234). He points out that for experienced writers there is an obvious quality difference between how real people and characters talk. It is equally apparent to experienced performers that real people do not move the way characters animate in professionally produced media. Character behavior should not model reality, because most people are not skilled performing artists. Or as Wardrip-Fruin puts it, the "work of an actor is to take actions that communicate, not those that are correct" (Wardrip-Fruin, 2009, p. 221, 326).

In his book on expressive processing, Wardrip-Fruin analyzed several landmark systems from an authoring point of view, and we close this section by briefly examining a few of the same systems from our Performatology point of view. For instance, let's consider the expressive role play affordances in the Eliza Doctor chatbot program that gave rise to the famed Eliza Effect in the 1960s. Weizenbaum designed the system behavior to parody a psychoanalyst as portrayed in popular media, which cast the system and player into the stereotypical roles of doctor and patient. Interaction was in the form of a therapy session dialogue, in which the system used simple logics to transform player responses into leading questions. Interaction generated a short-term anthropomorphic affect for many players, coined the Eliza Effect, which was not intended by the author. Weizenbaum, a psychologist, described the effect as "powerful delusional thinking" prompting subsequent work to "rob Eliza of the aura of magic" (Wardrip-Fruin, 2009, p. 95). His reaction underscores the difference between a mentalist and artistic approach to AI.

Wardrip-Fruin pointed out that Eliza's main appeal was its entertainment value. From a performing arts perspective, we see Eliza as an early success of simple acting affordances, within the context of a therapy scene. Eliza invoked iconic roles that the player was familiar with from television, provided a performative instrument that simulated imitation and improvisation, and engaged the player's attention through an artificial style of dialogue. Once players recognized the artifice, by realizing how the system worked, like all theatrical tricks they had to accept the limitations of their role to continue, or break the flow of play by typing in an unconventional response. It is well known in the arts that performer and spectator collaboration is necessary to maintain character illusions, known as Coleridge's suspension of disbelief, which is also required for most media "magic". Compared to today's cinematic games, however, Eliza's primitive acting logics had too shallow affordances, quickly collapsing the doctor-patient roles into caricatures that could not retain viewer interest.

Wardrip-Fruin proposed a Sim City Effect alternative to the Eliza Effect, characterized by purposely exposing the system to players (Wardrip-Fruin, 2009, 301). From a performative perspective, the Sims franchise of games still cast players into a role, but they function more like a director who plays the entire system like a semi-autonomous instrument. Directors in theater and movies set scenic constraints and provide motivation for actors, done in the Sim's world through iconic displays that manipulate properties of objects and characters to be influenced. Because players are dual spectator-performers, this type of control is similar to Japanese Bunraku puppetry, where several puppeteers in black manipulate a character in full view. Exposing the control mechanism requires active spectator belief in the characterization by separating attention from the visible actors to the animated puppet, a kind of doubling experience. The Sim City Effect can be compared to a Brechtian role play model, where players and spectators are always aware that they are collaborating in an artistic performance. Brecht called his method an alienation effect, adapted from Asian theater practices (Brecht, 1964), which also originated from ritual performance and have properties that resemble flow.

Although the Eliza and Sim City effects had performative elements, their designers did not intentionally implement a performing arts model. However, Mateas and Stern did implement a Neo-Aristotelian drama model in Façade, which featured acting logics such as a drama manager and well-cast character roles

(Mateas, 2004). The drama manager encouraged artistic interaction through a dramatic arc, while also ensuring the player didn't perform outside the boundaries of their role as a guest in a cocktail party scene. If the player performed as a bad guest, the drama manager had one of the hosts eject them from the party. Player interaction was in the form of social games with two antagonistic hosts, Trip and Grace, who induced players to cooperate or compete with the other host. Mateas and Stern's expressive AI approach was critical because they did not try to model general intelligence, but instead character interaction, a more tractable problem in the arts domain. Acting logics in Façade included dramatic beats, which drove the interaction towards a point of catharsis, associated in classical theater with spectator affect. Façade provided many more degrees of freedom than Eliza for artistic performance, integrating traditional media such as scripted animation and voice acting into procedural play. However, Wardrip-Fruin pointed out that the main drawback of their approach was the cost of authoring the media for Trip and Grace (Wardrip-Fruin, 2009, p.346). This underscores the need for more robust acting logics that move procedurally generated character control towards the aesthetic quality of cinema acting.

4. Computational Inference of Performer Aesthetic Quality

Our experimental approach closely relates to work already done in computational inference studies and player preference modeling for intelligent virtual camera control. The goals of cinematography and performance are similar in that both aim to optimize the aesthetic experience of viewers. Actors predict how the lines of their body will read when framed by the proscenium in theater, or when captured by the camera on film, and compose their gesture for maximum affect. Animators create keyframe poses that incorporate similar aesthetic principles to make characters readable and appealing. Intelligent virtual character control in games will also need to calculate the most aesthetically interesting poses per frame in relation to the virtual camera composition.

At a base level, continuous gesture in media breaks down to a series of captured or rendered poses. This is true in film and animation, where the frame or keyframe is the atomic element of visual information. This is also true in choreographed live performance, where poses are planned as moments of relative stillness compared to transition gestures, and may have greater visual affect. A professional performer like Michael Jackson displayed body control that generated very clear poses, which are often imitated to convey his iconic stage persona. Dance and physical acting practices train artists to have uncanny control over how they present their body to an audience. They learn how to isolate the movement of individual body parts in order to compose a series of visually interesting shapes. Disney animators, who studied filmed actors in the early part of the 20th century, derived similar principles for animated characters related to pose composition (Thomas and Ollie, 1981). Lead artists would apply these principles to keyframe drawings to make gestures more readable and appealing. Therefore, before studying the aesthetic features of continuous gesture, we are computationally studying how performers control their body to strike an interesting pose.

For understanding pose composition, we need to develop models that compute the visual quality of a pose according to embodied acting conventions, and ideally also for individual player preferences. To this end we are implementing a gesture game to generate poses that are scored and annotated for key features from performing arts theory. This is similar to photo composition studies that rank image composition using visual features related to balance, thirds alignment, and symmetry (Swanson et al. 2012). We attain our initial feature set from character principles in acting and animation theory related to pose composition, which include gesture practices that emphasize dynamic balance, asymmetrical lines, and silhouette readability. Our features employ metrics that score poses in a gesture corpus, initially from pre-recorded mocap data, and later from streaming 3D data captured by a Kinect camera during game play. The goal is to understand learned conventions that make skilled performers interesting to watch, so that interactive characters can be designed to optimize these features in procedurally generated key poses.

Many of the composition rules used in photography are also utilized in the related domains of cinematography and graphic design. For instance, the rule of thirds can be generalized as a structural feature of any framed composition, including ones that feature figures (Khan and Vogel 2012). Other rules used to infer aesthetics in images, such as balance and symmetry, have comparable principles in embodied art practices like acting and dance. We hypothesize that composition rules in graphic and figurative arts have similarities because they have the same goal of catching and holding the viewer's attention through visual interest. Good composition through the arrangement of shapes and lines stimulates eye movement through the framed image space, which reflects visual processing theories from psychology (Reber et al. 2004, Peters 2007). Pose composition rules are based on general conventions in the performing arts that make a performer interesting to watch independent of narrative content, sometimes called pre-expressive technique (Barba and Savarese, 2005), which is believed to generate stage presence. Professional actor Jaron Hollander demonstrated stage presence as an open soda bottle precariously balanced on the edge of a table so that no one in the room could take their eyes off it (Hollander, 2013). This is a metaphor to show the kinetic potential of an actor's body when switched on to perform in front of an audience. In studying the aesthetic quality of poses in mocap data, we hope to better understand how a posed body can engage viewer attention in order to establish a character channel for conveying narrative content.

Since similar composition rules inform both cinematography and performance, we are drawing from recent aesthetics research in photography. Professional photographers learn composition rules that indicate pro quality to the general public, and it is assumed that amateurs can improve their photos by learning to apply such aesthetic conventions. Studies on visual composition quality for camera control have designed algorithms that approximate photography rules to optimize camera directives for aesthetic scoring (Swanson et al. 2012). In camera aesthetic meters have been invented to improve amateur photos through real-time quality feedback (Joshi et al. 2011). Similarly, we want to identify features that contribute towards the compositional quality of artistic gesture, and assess their relative importance in predicting perceived quality of character acting in games. Then we can design NPC behavior that maximizes the same features, and design avatar feedback that functions as an aesthetic meter for better player acting.

Our experiments are designed to test three intuitive assumptions:

- Professionals trained in performing arts conventions produce higher quality gesture than untrained amateurs, and mocap data contains aesthetic features that indicate actor skill level.

- Exposure to professionally produced media has created performance literacy in the general public, so that aesthetic preferences reflect artistic conventions, and spectators can distinguish between pro and amateur quality performers by the aesthetic features they display.
- In videogames, animated avatars and NPCs can generate higher quality aesthetics by training players and agents to incorporate performing arts conventions, so that interactive characters display general features of skilled acting.

Image quality assessment and semantics inference studies draw from aesthetics in the visual arts, especially photography and painting, as well as from psychology work on the states of mind involved in the aesthetic experience (Joshi et al. 2011). Surveys suggest that the main challenge is an "aesthetics gap" relating low-level computable visual features and high-level human semantics of perception. Feature representation has been influenced by the Gestalt psychology concept of goodness configuration, where perception is organized according to properties like symmetry and simplicity. Processing fluency theory suggests that the more fluently a perceiver can process an image, the more positive their aesthetic experience will be, so features that make images more clear or readable are important (Reber, 2004). The aesthetic experience of visual art is related to domain knowledge, in that skilled artists have a different perception of aesthetics than people not trained in the arts. Cross-cultural studies have shown that the general public has less appreciation for conceptual or abstract art, placing more emphasis on surface aesthetics than underlying ideas (Joshi et al. 2011). For these reasons aesthetic studies focus on formal composition rules, like the Rule of Thirds, which have been displayed in the popular arts of painting, photography, and cinema. Such rules have been implemented as quality inferring features that computational studies have used to distinguish professional from amateur photography (Swanson et al. 2012).

Computational inference studies have approached the problem of inferring image quality by taking both top-down and bottoms-up approaches. Top-down approaches use composition rules from photography literature to design algorithms that can predict features of high quality images, while bottoms-up approaches use statistical methods to learn viewer preferences for appealing images. Core problems for computational studies are predicting aesthetic and emotional responses for cliques in the general population, and understanding individual preferences that make some images more appealing than others (Joshi et al. 2011). The bulk of aesthetic and emotion inference approaches have used classification, with a two-class aesthetic prediction for "high" versus "low" quality being the most tractable approach, and a similar "interesting" versus "boring" scheme for emotional appeal.

The closest work related to our implementation was a mixed approach that used composition rules to generate and annotate images with target features, and then employed statistical methods to correlate viewer preferences (Swanson et al. 2012). They used gameplay scoring to generate landscape images that were annotated according to aesthetics derived from photography, and then implemented crowd sourcing on the resulting corpus to measure general and individual image preferences. Color coded badges in the game functioned as an aesthetic meter, encouraging players to take pictures that rated highly for rule of thirds, balance, and symmetry. Generating images in the target domain gave them additional control over feature dimensions through design abstraction, such as removing color or minimizing cultural references in the game's representation of landscapes. User studies were done through Mechanical Turk on preferences for image quality using pairwise comparisons and a Support Vector Machine (SVM) classifier.

We are taking a similar mixed approach to inferring the aesthetic quality of poses, with the top-down design of pose algorithms with target features taken from performing arts theory (Barba and Savarese, 2005), followed by bottoms-up preference studies for spectator affect. Since we are interested in scoring the aesthetic quality of posed figures, we chose expressive features that are displayed by a variety of professional performers, including actors, dancers, models, and animated characters. Our hypothesis is that players skilled in artificial gesture conventions targeted for media will consistently score higher for our metrics than unskilled players, and therefore will produce poses that rank higher for quality by consumers of media. Performing artists train extensively in counteracting gravity, pushing the limits of imbalance, so our first target feature is balance. Poses that are evenly weighted are considered static, while poses on the edge of falling are considered dynamic. Precarious balance shifts the hips and shoulders in opposite directions, towards an S shape, considered pleasing to view. Dynamic balance has similarities to photo composition rules that move objects off the center line to power points in thirds regions, to create shapes reminiscent of the golden triangle, which has been featured in visual aesthetics work (Khan, 2012; Swanson, 2012).

Asymmetry is our second target feature, where a performer has limb variations across the body center, so that one limb may be straight while the other is bent. Limb asymmetries create angles that move the eye of the viewer across the form of the figure in interesting ways. Symmetrical poses are considered static, while asymmetrical poses express dynamic movement in stillness (Barba and Savarese, 2005). It is interesting to note that aesthetic studies on facial beauty showed that functional asymmetries in the faces of models scored higher for attractiveness than digitally mirrored faces (Zaidel, 2005). Readability is our third target feature, a principle in animation and performance practices that relates to how well a pose communicates to viewers (Thomas 1981). Psychology studies have shown that how easily an image can be cognitively processed is related to aesthetic pleasure (Reber, 2004). Skilled animators often pay special attention to the silhouette of a posed character, and actors are continuously aware of their body positioning in relation to viewers. If a pose is unreadable, it cannot convey significant semantic information. Therefore, expressive artists subtly exaggerate poses by moving limbs away from the torso to minimize visual occlusions, and to maximize the body's volume for viewing.

We are implementing our performatology approach by first modeling pre-expressive aesthetic rules for posing figures that generate viewer interest. We have conducted a pilot study by training a binary classifier on our target features using a corpus of poses extracted from skilled performer mocap data, with encouraging results. As part of these initial experiments, we have designed a Kinect posing game that scores players according to our metrics, generating a corpus of annotated poses for use in preference studies on aesthetic quality. Future work will involve coordinating camera and character composition through combined pre-expressive metrics, and then by moving on to more expressive gesture rules as found in classic cinema acting. In this article, we have argued for a CTP to bring aesthetics to the center of AI character research for videogames. We have presented theory and practices from the performing arts, AI, and psychology that provide a firm theoretical grounding for such an approach. Finally, we have proposed a specific methodology for determining aesthetic quality of embodied gesture that has the potential to improve the player experience of character interaction in playable media.

5. REFERENCES

Agre, Philip. (1997). Computation and Human Experience. Cambridge University Press, New York, NY.

Barba, Eugenio and Nicola Savarese. (2005). A Dictionary of Theatre Anthropology: The Secret Art of the Performer. New York: Routledge.

Bates, Joseph et al. (1993). Dramatic Presence. Presence‘93.

Bethesda. (2011). Elder Scrolls V: Skyrim. Bethesda.

Bioware Corp. (2008-2012). Mass Effect, 2, 3. Electronic Arts.

Bogost, Ian. (2006). Unit Operations: An Approach to Videogame Criticism. MIT Press, Cambridge, MA.

Brecht, Bertold. (1964). Brecht on Theatre: The Development of an Aesthetic. John Willett, tr. Hill and Wang, New York.

Chen, Jenova. (2007). Flow In Games. MFA Thesis. www.jenovachen.com/flowingames/thesis.htm.

Chesher, Chris. (2003). Layers of Code, Layers of Subjectivity. Culture Machine, vol 5: The E-Issue. www.culturemachine.net/index.php/cm/article/view/255/238.

Craig, Edward Gordon. (1963) The Theatre Advancing. Benjamin Blom, Inc. New York, NY.

Csikszentmihalyi, Mihaly. (1975). Beyond Boredom and Anxiety: Experiencing Flow in Work and Play. Jossey-Bass, San Francisco, CA.

Csikszentmihalyi, Mihaly and Rick Robinson. (1990). The Art of Seeing: An Interpretation of the Aesthetic Encounter. J Paul Getty Trust, Malibu, CA.

Engeser, Stefan (editor). (2012). Advances in Flow Research. Springer Science and Business Media, LLC.

Flanagan, Mary. (2009). Critical Play: Radical Game Design. MIT Press, Cambridge, MA.

Hall, Peter. (2000). Exposed by the Mask: Form and Language in Drama. Theatre Communications Group, Ltd, New York.

Hecker, Chris. (2008) Structure Vs Style. GDC’08 Talk. http://chrishecker.com/Structure_vs_Style.

Hollander, Jaron. Emotion and Personality in Corporal Movement. UCSC Immerse Project Workshop, Feb.11th, 2013. www.kineticartscenter.com.

Horswill, Ian. (2008). Twig: A Simple, AI-friendly, Character World for Believable Agents. In Proceedings of AAAI’08.

Johnstone, Keith. (1992). Impro. Routledge, New York.

Joshi, Dhiraj et al. (2011). Aesthetics and Emotions in Images: A Computational Perspective. Tutorial in IEEE Signal Processing Magazine, Sept’11.

Khan, Shehroz and Daniel Vogel (2012). Evaluating Visual Aesthetics in Photographic Portraiture. In Computational Aesthetics in Graphics, Visualization and Imaging ‘12.

Kipp, M., M. Neff, et al. (2007). Towards Natural Gesture Synthesis: Evaluating gesture units in a data-driven approach to gesture synthesis. In Proceedings of IVA’07.

Krahmer, Emiel and Marc Swerts. (2008). On the Role of Acting Skills for the Collection of Simulated Emotional Speech. In Proceedings of INTERSPEECH’08.

Leabhart, Thomas. (1995). The Mask As Shamanic Tool in the Theatre Training of Jacques Copeau. Mime Journal, Issue: Incorporated Knowledge.

Leabhart, Thomas and Franc Chamberlain. (2008). The Decroux Sourcebook. Routledge, New York.

Magerko, B, C. Fiesler, et al. (2010). Bottoms Up: Improvisational Mico-Agents. In 3rd Workshop on Intelligent Narrative Technologies, FDG-INT3’10.

Maraffi, Christopher. (2011). Roots of Performatology: From Craig’s Uber-Marionette to Performative Embodied Agents, In Proceedings of DH'11.

Marsella, Stacy et al. (2006). An Exploration of Delsarte's Structural Acting System. Proceedings of IVA’06.

Mateas, Michael. (2004) A Preliminary Poetics for Interactive Drama and Games, In First Person. Eds: Wardrip-Fruin and Harrigan. Cambridge, Massachusetts: The MIT Press.

Mateas, Michael and Noah Wardrip-Fruin. (2009). Defining Operational Logics. Proceedings of DIGRA’09: Breaking New Ground.

Moore, Ronald et al. (2007). Battlestar Galactica: Maelstrom DVD Commentary. Season 3, Ep. 17. Universal Studios.

Murray, Janet. (1998). Hamlet and the Holodeck. MIT Press, Cambridge, MA.

Nixon, Michael, et al. (2010). DelsArtMap: Applying Delsarte's Aesthetic System to Virtual Agents. Proceedings of IVA’10.

Oshita, Masaki. (2008). Smart Motion Synthesis. In Computer Graphics Forum’08, vol27, Issue7, pp1909-1918.

Perlin, Ken and Gerry Seidman. (2008). Autonomous Digital Actors. In Proceedings of MIG’08.

Peters, Gabriele. (2007). Aesthetic Primitives of Images for Visualization. In Proceedings of IEEE’07.

Rayner, Alice. (2006). Ghosts: Death’s Double and the Phenomenon of Theatre. University of Minnesota Press, Minneapolis.

Reber, R, N. Schwartz, et al. (2004). Processing Fluency and Aesthetic Pleasure: Is Beauty in the Perceiver’s Processing Experience? In Persp. Social Psychol. Review, vol8, no4.

Sawyer, Keith.(2005). Acting. In Creativity Across Domains. Eds: Kaufman, James. Lawrence Erlbaum Assocs, London.

Seif El-Nasr, Magy. (2005). “Applying Principles from Performing Arts for an Interactive Aesthetic Experience”, Intelligent Agent Magazine, vol5, issue1.

Shepherd, Simon. (2006). Theatre, Body, and Pleasure. Routledge, New York.

Stebbins, Genevieve. (1902). Delsarte System of Expression. Werner Publishing & Supply Co, New York.

Swanson, R, A. Jhala, et al. (2012). Learning Visual Composition Preferences from an Annotated Corpus Generated through Gameplay. In Proceedings of IEEE'12.

Talbot, Christine and Michael Youngblood. (2012). Spatial Cues in Hamlet. Proceedings of IVA'12.

Thomas, Frank and Ollie Johnston. (1981). Disney Animation: The Illusion of Life. New York: Disney Editions.

Valve. (2011). Portal 2. Electronic Arts.

Wardrip-Fruin, Noah. (2005) Playable Media and Textual Instruments. Dichtung Digital 1. www.dichtung-digital.com

Wardrip-Fruin, Noah. (2009) Expressive Processing. MIT Press, Cambridge, MA.

Yannakakis, Georgios. (2009). Preference Learning for Affective Modeling. . In Proceedings of IEEE'09.

Zaidel, D. and J. Cohen (2005). The Face Beauty, and Symmetry: Perceiving Asymmetry in Beautiful Faces. In Intl. Journal of Neuroscience, vol115, no8.

Zhu, Jichen. (2009). Intentional Systems and the AI Hermeneutic Network. PhD Dissertation, Georgia Institute of Technology.