UNIVERSITY OF CALIFORNIA

SANTA CRUZ

# Inference of Performer Artistic Skill from Artistic Pose Features in Motion Capture Data

A thesis submitted in partial satisfaction

of the requirements for the degree of

MASTER OF SCIENCE

in

COMPUTER SCIENCE

by

**Christopher Maraffi**

June 2014

The Thesis of Christopher Maraffi is approved:

_____

Assistant Professor Arnav Jhala, chair

_____

Associate Professor Noah Wardrip-Fruin

_____

Professor Michael Mateas

_____

Tyrus Miller

Vice Provost and Dean of Graduate Studies

# Table of Contents

# List of Figures

# Abstract

Inference of Performer Artistic Skill from Artistic Pose Features in Motion Capture Data

Christopher Maraffi

This thesis proposes a computational aesthetics methodology for measuring the design quality of poses in animation data, and then for predicting the composition skill of the source artist. We draw from animation and performing arts principles to select pose features and design metrics for supervised learning on a corpus of extracted 3D poses. Though our approach is designed to be general enough to apply to aesthetic features in any performative figure data, we choose three features to investigate and conduct specific experiments using motion captured data from live performers in the domain of dance. An initial pilot study is conducted on pose data from a dance instructor to assess our metrics, and then a formal experiment is conducted on performance captured data from participants playing a popular Kinect dance videogame. Principal component analysis is utilized to identify low-level skeletal features, and then supervised learning experiments are conducted to infer performer skill from figure composition quality based on metric scores. Results show statistical correlations between intuitive skill rankings, game score distributions, and metric ratings. This thesis provides a methodological foundation for future work in scientifically studying the arts to formalize principles of figure representation.

# Acknowledgements

The author would like to thank my advisor, Arnav Jhala, for mentoring me through the transition of thinking like a technical artist to thinking like a computer scientist. I would also like to thank my project partner Sascha Ishikawa, for his help with conducting the experiments, and the rest of UCSC's Computational Cinematics Studio team for their support in this research. Additional thanks to members of UCSC's Expressive Intelligence Studio, including faculty committee members Michael Mateas and Noah Wardrip-Fruin, for their interest in my work. Also, the author would like to thank faculty in UCSC's Games and Playable Media Program who gave me advice on my research, including Marilyn Walker and Alex Pang.

# Chapter 1

# Introduction

This thesis presents a computational methodology and initial experiments on analyzing artistic gesture data to learn figure design rules. The motivation is that understanding how skilled artists create compelling figure representations could inform the design of procedural character acting in videogames that incorporate cinema aesthetics. Popular online lists of the most cinematic videogames include titles from a variety of game genres, from RPGs (Walking Dead, 2011) to action-adventures (Mass Effect 3, 2012) to puzzle-platformers (Portal 2, 2011) to first-person shooters (Bioshock Infintite, 2013), with a noted trend towards increased cinema aesthetics in subsequent releases of popular franchises like Halo (Microsoft Studios, 2001) and Final Fantasy (Square Enix, 1987). Procedural content generation (PCG), through real-time physics simulations and movie quality visual effects, are blurring what used to be clear aesthetic distinctions between interactive gameplay and non-interactive cinematics. Furthermore, academic research in the areas of computational cinematography (Swanson et al, 2012) and interactive narrative (Mateas, 2004) have been laying the foundations for virtual camera control and story generation that formalizes arts procedures. Since cinema aesthetics in movies are dependent on the combined skills of many different arts professionals, from script writers to cinematographers, simulating a comparable interactive experience in videogames will likely require PCG for each artistic component. One visual arts

component not well researched or computationally modeled, however, is figure design procedures used by artists to create appealing gesture in media.

A promising methodology for the scientific study of the arts can be found in the field of computational aesthetics, where machine learning has been used to classify structural design features in artist generated data, such as Rule of Thirds in a corpus of photographs (Khan and Vogel, 2012) or brush stroke patterns in paintings (Bressan et al, 2008). For instance, data mining and viewer preference studies are combined to identify best practices in photography that may contribute to viewer perception of visual production quality. Much of the photography work is on learning general composition features that contribute to image appeal or interestingness (Joshi et al, 2011), rather than focusing on creativity or unique artist style variations. For figure design in videogames, the former could provide a general model of industry craft conventions employed by skilled professionals in representing characters, allowing for the prediction of popular appeal for PCG of avatars and non-player characters (NPCs). Modeling the production and reception of professional quality aesthetics in popular art may be more tractable than modeling creativity that produces artistic novelty or genius, which is beyond the scope of this thesis. Since there is prodigious amounts of artistic figure data being produced by professionals in the entertainment industry, this thesis proposes a methodology for analyzing the aesthetic quality of a subset of this data, specifically motion captured data in the domain of dance.

Though it is intuitive that skilled graphic illustrators, photographers, and cinematographers learn to apply design rules in their work, it is less obvious that

gesture artists are trained in comparable procedures. But intercultural performance studies have theorized that actors and dancers learn postures for promoting visual interest, sometimes characterized as pre-expressive because they precede more expressive gestures that convey plot information (Barba, 2003). Pre-expressive figure practices contribute to the spectacle of cinema by designing visually interesting poses within the proscenium or frame. Because visuals are immediate and story constructs take time, elements of spectacle both precede and support narrative by attracting, manipulating and holding viewer attention. Barba documented the use of these design principles across a variety of performative practices, including live acting, dance, puppetry and animation (Barba and Savarese 2005). Disney animators developed comparable principles from rotospcope studies of filmed dancers and actors early in the 20th century (Thomas and Johnston, 1985), which have since become industry standard practices for professional animators. In practice, Disney's principles have been applied to a wide variety of biped characters, independent of narrative context, from very abstract dance scenes to emotive acting. The aesthetic goal of professional performing artists and animators is to compose visually affective figures for spectator viewing, with the perceived result in digital media from both sources being identical in rendered pixels. Figure composition procedures have been associated with audience perception of heightened actor presence in cinema (Barba, 2003) and an illusion of life in animation (Thomas and Johnston, 1985), but to our knowledge there has been no computational data analysis of the design conventions related to this aesthetic phenomenon.

Therefore this thesis takes a computational aesthetics approach to studying figure composition in digital media, using statistical and machine learning to better understand what skilled artists do to make characters appealing to spectators. Specifically, in this thesis we apply our methodology to study figure composition in motion captured animation data of live performers playing a dance game, to quantify features related to pose quality and qualify the perceived skill of the performer. Our experiments are intended to address the question: can figure composition metrics that measure figure quality according to arts theory, combined with preference studies on spectator perception of pose quality, statistically classify the artistic skill level of the performer? The intuitive hypothesis we are testing is that there exists measurable features in animation data that can be rated for general aesthetic quality, and that there also exists a causal connection between perceived figure aesthetics and the skill ranking of the artist that generated the data.

Chapter 2 discusses related work in the field of computational aesthetics that studies the perception of quality in photography. Methodologies in this emerging field of research are reviewed, including computational cinematography work that most influenced the experimental design in this thesis. Chapter 3 reviews figure design principles from the domains of animation and live performance that influenced our feature selection and metric design. Chapter 4 discusses feature selection from figure design principles, metric design for measuring pose quality, as well as experimental design and study results. Chapter 5 relates motivations that influenced the

performative design methodology adopted for this thesis, and proposes future work that is intended to broaden and deepen this area of computational research.

# Chapter 2

# Computational Study of the Arts

Computational Science has been distinguished as "applied" Computer Science (NITRD, 2005), where computational methodologies are used to address problems in other areas of science or the humanities. Data scientists conduct machine learning on a target domain, such as Biology or Psychology, to infer structure and build predictive models for multi-dimensional problems that are hard to understand and visualize. Computational Aesthetics is a subfield that studies user experiences of media artifacts, and since aesthetics are most often crafted by artists, it is where data analysis is starting to be done on design practices. The majority of aesthetic analysis has been conducted in the artistic domains of painting (Berezhnoy, et al., 2005), photography(Yeh et al, 2010), and cinema(Adams et al, 2002), including spectator preference studies on perceived quality. Research on 2D static figure composition in photography has been minimal, with some work on figure composition in portraits (Khan and Vogel, 2012), but no aesthetics analysis has previously been done on 3D animated figure data.

While there has been other work in computer science on what actors do in the performing arts (Bates 1993, Seif El-Nasr 2004-2005, Marsella 2006, Tanenbaum and Tanenbaum 2008, Perlin and Seidman 2008, Neff and Fiume 2008, Magerko 2010), none of these studies took a computational aesthetics approach to modeling figure composition. Some data analysis work has been done to derive style and skill features of performing artists in motion capture data (Brand and Hertzmann 2000), but they did not focus on quantifying composition features associated with visual appeal. Because the hypothesis of this thesis seeks to test figure design principles not dependent on narrative or character context, other studies related to choreography and cognition are outside the scope of this work (Calvert et al. 2005, Tanenbaum 2011, Kipp et al. 2008, Neff et al. 2008, Thiebaux 2008). Instead, this thesis adapts work used to infer 2D image quality in photography and cinematography to analyzing 3D figure quality in animation data.

## 2.1 Related Work in Computational Aesthetics

The emerging field of computational aesthetics was surveyed in an image processing tutorial on the inference of aesthetics and emotions in images (Joshi et al, 2011). Image quality assessment and inference studies have focused on visual design practices in the arts of photography and painting, as well as drawing from psychology work on the states of mind involved in the aesthetic experience. Cross-cultural psychology studies have shown that the general public has less appreciation for conceptual art, placing more emphasis on structural design features prevalent in popular art than the underlying concepts behind abstract or "high" art. One such

structural design feature that was the focus of many photography studies in the field was the popular Rule of Thirds, where an image is composed so salient content is placed close to intersecting "thirds" lines (see Figure 1). Such structural design features may indicate professional image quality to viewers, according to the survey, "Photographers generally follow certain principles that can distinguish professional shots from amateur ones. A few such principles are the rule of thirds, use of complementary colors, and close-up shots with high dynamic ranges" (Joshi et al, 2011). In this thesis, we study structural design principles in the domains of animation and live performance that may indicate professional figure quality.


Figure 1:Structural design principles like Rule of Thirds, where the eye is directed to content placed on intersecting "thirds" lines, has been studied in computational aesthetics (Wikimedia Commons image by Chaky, 2009).

Computational frameworks for this area of study have been influenced by psychology methods and visual art theory. Art theorists hypothesize that there are multiple levels of aesthetic perception, starting with surface resemblance to objects, followed by a deeper level of cognitive interpretation that depends on the cultural environment and habitual semantic conventions. Pictorialist photographers in the 19th century borrowed composition practices from the domain of painting to develop photography as an art form. A common practice of landscape photographers is to use film that exaggerates color intensity, rather than film that produces realistic or natural colors. Studies show that color intensity, contrast, and blurriness are features that imply artistic quality to viewers, and suggests that artistic practices push features beyond natural parameters.

The dominant machine learning strategy for aesthetic inference has been supervised support vector machine (SVM) classification. A simple regression framework for quantitative scoring and prediction is challenging due to noise in user ratings. Instead, the bulk of aesthetic and emotion inference approaches have used binary classification, such as “high” versus “low” for quality, or “interesting” versus “boring” for appeal. In one such study, crowd sourced ratings data from the DPChallenge web site were sorted into “preferred” and “non-preferred” classes, and then ListNet was used for feature extraction and training their binary classifier (Yeh et al. 2010). Their personalized ranking system used a “late fusion” method to produce an aesthetic score as a linear combination of optimal weightings for their top features, such as simplicity and contrast, with a reported 93% accuracy. Users were

presented with an initial ranking of photos in two types of interfaces: one where they manipulated feature sliders to re-rank the photos, and another where they clicked on preferred example photos for automatic re-ranking. Included in the interfaces were more subjective personal features like portrait, color, and aspect ratio. Unsupervised learning has also been done using K-means clustering for visual vocabulary generation, graph-based region segmentation, and image clustering to form topical groups. Bayesian network classifiers have also been used for generative learning, and to classify facial expressions for emotion inference. Real-world photography applications of visual learning algorithms include camera systems that filter data into an aesthetics meter for composition quality.

Other related work that specifically focuses on inferring quality in amateur portrait photos was done by (Khan and Vogel 2012). Their top-down approach reduced the classification space to only 7 composition and color features that specifically target portraiture. They make the point that most statistical methods take a bottom-up approach by calculating a large amount of global features across a general space of images to correlate a subset of relevant features to visual aesthetics. In a template-based method they combine composition principles like the rule of thirds and golden mean into a two dimensional lookup table, and claim better classification results with 1/9th the size of the traditional feature set. The human photo data set with ratings from Mechanical Turk was their training data. Using the median of all ratings for ground truth, they labeled portraits for a binary classification of “high” and “low” quality using five Weka algorithms in a 10-fold cross validation process. The top

performing features were all composition related, with rule of thirds being consistent with other more general feature sets.

## 2.2 Computational Cinematography Studies

Other research in the related field of computational cinematography takes a similar approach to inference of image quality, but targets videogames for data acquisition and metrics applications. The UCSC Panorama research project designed a videogame with game play mechanics that encouraged players to take virtual landscape photos, with composition features that could be measured for aesthetic quality (Swanson et al, 2012). They implemented metrics for three high-level composition features (balance, symmetry, and thirds), and designed an intuitive scoring meter to provide real-time aesthetic feedback to the player (see Figure 2). Since the 2D landscape photos were generated from a 3D object-oriented virtual environment, additional low-level image features could be easily added to their multivariate classification algorithm. Generating virtual photos reduced the feature dimensions through design abstraction, such as removing color or minimizing cultural references, which would have been difficult using a corpus of real photos from internet databases.


Figure 2:Panorama research project used videogame play to generate a corpus of virtual landscape photos for computational cinematography analysis (Swanson et al, 2012).

The game play of Panorama encouraged players to capture photos according to photography conventions, and the resulting corpus of images was conveniently annotated for their target features. Crowd-sourced preference studies for viewer perceived quality were done on the images using Amazon's Mechanical Turk. Viewers did pair-wise comparisons of four-alternative forced choice (4AFC) ratings to learn individual as well as general preferences. Their multi-class SVM algorithm got the best accuracy by reducing the choices to a binary "good" vs "bad" classification task. Their methodology of following composition ratings with viewer preference classification provides two layers of analysis for interpretation and correlation of data from both sides of the arts process. First, artistic performance is measured by "inside" or industry standard rules of production, and then second, the

aesthetics of the artifact are measured according to "outside" or spectator perception of quality. In this thesis, the first half of this methodology is applied to figure arts composition in frames of media.

# Chapter 3

# Figure Design Studies in Traditional Media

In cinema, performers and cinematographers collaborate towards the common goal of designing an interesting visual. Live performers are taught to be aware of where the camera or spectator is positioned in relation to their own bodies, and routinely adjust their gestures to maximize figure appeal and readability. Camera operators are in turn always following the figurative action, framing the performers for better composition. The professional interaction between cinematographers and performers is guided by artistic conventions that have been studied by both theatre and animation practitioners. This thesis draws on principles derived from live figure studies done by Disney animators in the early part of the twentieth century (Thomas and Johnston, 1985), as well as intercultural performance studies done in the latter part of the twentieth century on the aesthetic phenomenon of stage presence (Barba and Savarese 2005).

## 3.1 Disney's Animation Principles

The modern style of Disney characters developed out of in-house figure studies and experimentation in rotoscopy (painting on film) in the 1930s when the studio was scaling up to produce feature length animations like Snow White (1937). Rotoscoping individual film frames had limited success due to the realistic constraints of everyday human gesture. According to Disney, "The fact is that humans can't move as freely and gracefully and comically as we can make animated figures move" (Thomas and Johnston, 1985). Instead, they adapted live action gesture conventions that allowed animators to push beyond realistic parameters, but still be perceived as believable in an entertainment context. As Disney animator Dan McManus put it, "Our business is to present something in an unreal way to make it seem more real... create it as the eye thinks it should look" (Thomas and Johnston, 1985). 12 principles of animation were developed and taught to junior animators, and a new hierarchical method of "pose to pose" animation was developed that offered more control of timing and narrative. The pose to pose method had a lead animator draw all the main "key" poses while junior animators would complete the in-between transition poses. The original 12 principles have since become industry standards, being taught in most animation schools, and adapted to computer animation by John Lasseter of Pixar (Lasseter, 1987). Here are the principles organized and paraphrased by the author:

1. Straight Ahead Action/Pose to Pose - two methods for animating figures.
2. Follow Through/Overlapping Action - moving body's reaction to momentum.

3. Slow In/Slow Out - a moving body's reaction to gravity.
4. Arcs - how joints are constrained to move on a skeleton.
5. Solid Drawing (or Weight) - the solidity of a body's mass to forces.
6. Squash and stretch - the flexibility of a body's mass to forces.
7. Timing - number of poses to create attitude with tempo and rhythm.
8. Secondary Action - secondary layer of design to add visual complexity.
9. Anticipation - preparatory motion to direct viewer attention to main action.
10. Staging - composing figures so nonverbal attitudes are clearly readable.
11. Exaggeration - extending poses beyond realistic parameters for clarity.
12. Appeal - creating visually interesting gesture to hold spectator attention.

The list of principles stand out as being generalized to the point of being structural rather than expressive, in that they are not story or character specific, but have supported expressive animated characters in numerous Disney features. Animators have applied these principles to a wide variety of figures in an equally broad range of scenic actions, from intense drama in Snow White to comedic dancing hippos in Fantasia. Principle #1 is less principle than methodology, defining the two animation methods used by Disney animators over the last century. In 3D computer animation software, a type of Straight Ahead Action is utilized in motion capture systems, while software interpolated Pose to Pose keyframe controls are standard in most professional rigs. Principles #2-5 are more concerned with simulating believable physical properties of moving bodies, where skeletal joints limit movement to arcs and the weight of flesh is deformed by classical physics. Many of these physical effects are automatically built into the structure of digital rigs that incorporate soft

and rigid body dynamics. Principles #7-12 reflect Disney early film studies of live dancers and actors, adapting figure composition conventions of the time to animation practices. These six design principles are concerned with creating clearly readable attitudes and visually interesting poses, in order to capture and manipulate spectator attention for aesthetic affect. They are the only principles that have not been automated in computer animation, and while still structural or pre-expressive, support character personality and narrative. Principal #6 or Squash and Stretch, considered by Johnston and Thomas to be the most important principal, can be interpreted in both ways. A physical interpretation of this principal is simulating the volumetric scaling of deformable characters, however, a more performative interpretation would be any opposing representation of the body's flexibility. This thesis proposes that structural principles like these are what makes animation productions look professional to spectators, and are important pre-expressive design procedures in the figure representation process. To further understand the more performative principles (6-12), the next section will discuss performing arts theory.

## 3.2 Barba's Performing Arts Principles

While the structural design of figure poses is regarded as an essential component of animation production, similar composition procedures are less obvious in live performance. Posture in live performance is either choreographed in pre-production rehearsals or improvised in real-time from the performer's years of training. By contrast, animation is a slower externalized process, where an animator must explicitly compose each movement, either by drawing individual poses or by

manipulating a rig into each pose for rendering. Even though both processes generate composed animated figures in media, the design process in live performance is often obscured within the performer's craft. Called the "secret art of the performer" by theater arts practitioner Eugenio Barba, who has conducted intercultural performance studies to reveal the underlying principles of performer technique (Barba and Savarese 2005). Since Disney's animation principles were developed at least in part from studying live performance footage of dancers and actors, it is likely that stage performers utilize some overlapping procedures with animators. Barba's work studies how physical performers attract and hold spectator attention through stage presence, which he considers a result of the pre-expressive training process. Pre-expressive techniques are found in both Eastern and Western performance practices, and are considered an essential component of crafting the body as a tool for performance. Conditioned responses, sometimes called "muscle memory" in practice, allow the performer to clearly display attitudes and intentions through learned postures, which Barba characterizes as presenting a "dilated, decided, and fictive" body:

- Dilated - A body extended beyond everyday parameters, with a larger than life personality that projects a clear attitude in gesture.
- Decided - A body that directs viewer attention towards the flow of action through muscle tensions and counter-body oppositions.
- Fictive - A body that systematically employs artistic gesture conventions to abstract reality for heightened visual affect.

These three aspects of a performing body are further described by the following 9 principles (paraphrased by the author):

1. Opposition - directing viewer attention with counter body movements.
2. Montage - composing gestural timing for heightened visual interest.
3. Rhythm - organization of movement and pauses to form a rhythmic flow.
4. Energy - visible muscle tensions create movement potential and expectation.
5. Kinaesthesis - awareness of visible tensions that indicate impending action.
6. Equivalence - presenting an alternate symbolic representation of reality.
7. Omission - abstracting something real to symbolically represent it with mime.
8. Balance - actively countering gravity to create a dynamic posture.
9. Nostalgia - representing archetypal theatrical conventions to affect viewers.

Each of Barba's principles can be related to his three body types, and to Disney's principles. For instance, Opposition and Kinaesthesis contribute to representing a Decided Body, and relate to Disney's Anticipation. Preparatory and oppositional movements indicate the performers decided trajectory, and allow spectators to anticipate and follow the flow of action. For this thesis we will reduce these many overlapping principles to three measurable features of a posed figure: balance, asymmetry, and readability. Balance relates to Barba's extra-daily Balance and Disney's Staging principle, and is indicated by the torso angle. Asymmetry relates to Barba's Opposition and Disney's Anticipation, and is measured by the difference between limb angles. Readability is related to Barba's Energy and Disney's Exaggeration principle, and is measured by limb visibility in relation to the torso. In the next chapter, the design of metrics from these pose-level features will be discussed in detail.

# Chapter 4

# Aesthetic Analysis of Performance Captured Animation Data

There are a few intuitive assumptions about performers trained in the arts that the computational experiments in this thesis are designed to test. We hypothesize that individuals skilled in the performing arts employ figure composition procedures that relate to Disney and Barba's principles, and that metrics designed to quantify pose features in animation data can be used to qualify the artistic skill level of the performer. Developing a methodology for measuring the aesthetic features of figure composition in media could provide structural knowledge of artistic procedures that have yet to be formalized. Metrics that accurately predict pose quality and performer skill can be used to inform algorithms that are designed to simulate figure composition in virtual applications like interactive characters in videogames.

## 4.1 Feature Selection

In this thesis we focus on analysis of pose features that display off-center torso angle, counter-body limb rotations, and outward extension of gestures. These features relate to Disney's Staging, Exaggeration, and Appeal, as well as to Barba's Balance, Opposition, and Energy. We label our features Balance, Asymmetry, and Readability

(BAR), and each is hypothesized to contribute to a visually interesting pose in the performing arts and animation (see Figure 3). We also hypothesize that more than a random display of these features indicates some performing arts training. The BAR features may be a small subset of what skilled performers do to create appealing poses, but they still encompass a range of full body attributes and will be expanded in future work. Though selection was informed by domain knowledge, the features tested in our experiments are considered less important than the experimental design for applying computational aesthetics methodology to figure data.

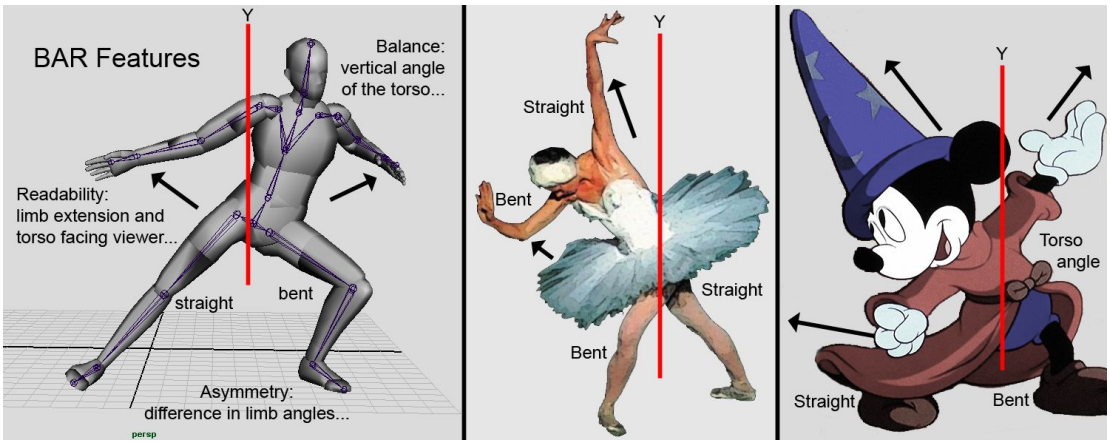



Figure 3:BAR features of balance, asymmetry, and readability in a pose from our corpus, and in ballet dance (Barba and Savarese, 2005), and in a Mickey Mouse drawing (Fantasia, 1940).

1. Balance - One of the "extra-daily" skills that Barba studies in performing arts practitioners is the ability to engage gravity through "precarious balance" (). Practitioners in many dance and acrobatic forms train to move their shoulders and hips in opposing direction on the horizontal axis, countering imbalance, and creating a dynamic S shape in the torso posture. The closer the

practitioner gets to imbalance, the more they must engage muscles of the torso, leading to a visible tension or "liveness" in the figure. When engaging gravity, they clearly display the solidity of the body by distributing weight onto one foot or the other. As a figure composition procedure, moving the upper and lower parts of the body off the central vertical axis has some similarity to how Rule of Thirds has been shown to be an important visual design feature (Khan 2012, Swanson 2012). Both the upper and lower torso are angled towards a thirds line or power point in many precariously balanced figures. This may stimulate spectator eye movement across the body more than an equally balanced vertical posture, contributing to figure appeal through a more dynamic aesthetic.

2. Asymmetry - Another performing arts skill is the ability to separately control the opposite sides of the body, sometimes called counter-body movement, and related to Barba's Oppositions principle. Mirroring the two sides of the body is not only easier to perform, it also displays a less complex pose to the spectator because limb information is duplicated across the vertical axis. Varying the limb rotations on either side of the body creates variations, and can extend the visual lines of a precariously balanced torso. More complex poses may stimulate spectator eye movement, which is aesthetically appealing according to visual processing theory (Reber et al. 2004). Additional psychology studies on symmetry and attractiveness scored functional asymmetries in the face as more appealing than digitally mirrored facial features (Zaidel 2005). Limb

rotations that structurally support the dominant torso angle reflect design principles related to Disney's secondary actions and staging, potentially displaying a figure with a clear attitude.

3. Readability - A third skill studied in performing artists is their constant awareness of the spectator or camera position, so that the figure is reoriented and composed for best viewing. Limbs are projected outward, away from the torso, minimizing occlusions and adding energy to the pose through muscle tensions. A readable pose relates to Disney's Exaggeration and Staging principles, subtly going beyond natural parameters to display a clear idea or attitude. If a figure is not composed to be readable, it is unlikely the spectator will get enough semantic information to be affected by it. Therefore, animators often check the readability of their key poses by squinting their eyes or stepping back to see if the figure is instantly recognizable from a distance. Live performer's are also known to subtly project their gestures outward towards spectators at the back of a theater, and use theatrical conventions for presenting clear attitudes. Such practices contribute to a dilated, decided, and fictive body on stage, which Barba attributes to enhanced presence.

## 4.2 Metric Design

In our experiments we implemented three BAR metrics to measure how well a single frame of 3D skeletal data displays our three target features. Focusing on individual poses in animation data, rather than trajectories of joints across a series of frames,

was both a practical and theoretical choice. This allowed us to reduce computational dimensionality by removing temporal features, and reflects animation practices that consider individual key frames as the foundation of figure composition. We took a supervised learning approach where we sampled a corpus of key frames from a performance captured sequence to statistically register how well a performer rated according to our metrics. Each pose is measured relative to the default skeletal values taken when the figure was in the classic T-pose, with torso vertical in Y, arms outstretched in X, and with the legs shoulder width apart so the weight is evenly distributed to both feet.

The Balance metric is designed to measure the amount a pose appears to be engaging the skill of balance, by calculating the amount of "lean" on the three main components of the torso, specifically the hips, shoulders, and backbone. The lean measurement calculates the difference between the total angles and the default, so that a torso that is angled further from the upright T pose is leaning more, and therefore requires more balancing ability by the performer to counter gravity. The angle of the hips are measured by drawing a line between the left and right shoulder joints, as is the angle of the shoulders. The angle of the backbone is calculated by drawing a line between the center of the hips and shoulders (see Figure 4 A).

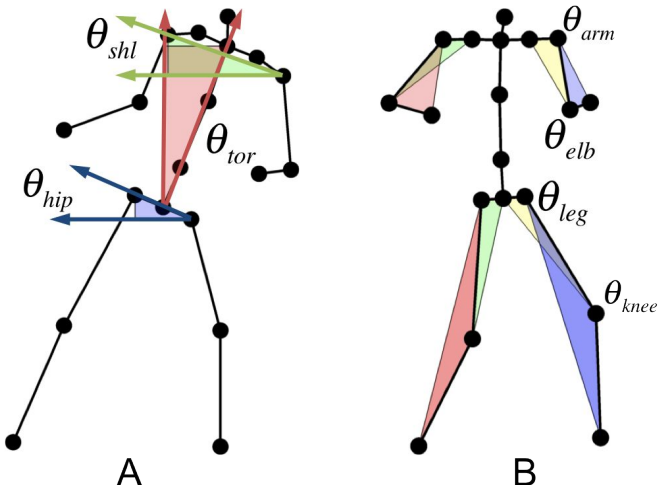



Figure 4: The Balance (A) and Asymmetry (B) metrics calculate angles of the torso and limbs (Maraffi et al, 2013).

The Asymmetry metric is intended to measure the ambidextrous counter-body skill a performer has in varying the limb rotations across the vertical Y axis, by summing the difference between right and left limb angles (see Figure 4 B). A dynamically posed figure will often have one leg bent while the other is straight, especially if the torso is angled so that the weight of the figure moves onto one foot. Likewise, the lines of the arms are often varied to support the dominant torso inflection. Specifically, our metric design as reported in AIIDE, "Let $\theta_{h,a}$ denote the rotation of the joint, where the subscripts h and a specify the left/right side of the body and the axis along which the angle is measured. Then the rotational asymmetry of the j-th pair of joints is given by (see Figure 5), where Nj is the total number of joint pairs. Finally, the rotational asymmetry and the joint asymmetry are summed and normalized over the total range of degrees they can vary." (Maraffi et al. 2013).

$$a_r^{(j)} = \sqrt{\left(\theta_{L,x}^{(j)} - \theta_{R,x}^{(j)}\right)^2 + \left(\theta_{L,y}^{(j)} - \theta_{R,y}^{(j)}\right)^2 + \left(\theta_{L,z}^{(j)} - \theta_{R,z}^{(j)}\right)^2} \qquad A_r = \frac{1}{N_j}\sum_{j=1}^{N_j} a_j$$

Figure 5: Asymmetry is calculated as the sum of the differences between the left and right joint angles (Maraffi et al, 2013).

The Readability metric is designed to measure the performer's ability to compose the torso and limbs for viewing by the audience or camera, so that the parts of the figure are clearly visible and not occluded. Because readability is dependent on the viewer position relative to the figure, we use the camera data in our performance captures to calculate the "foreshortening that occurs in the projection of the skeletal features onto the camera plane. Since the motion capture system normalizes the skeletal features of each performer to standard lengths, an estimate of readability can be obtained by calculating the ratio of foreshortened lengths to the actual lengths of the skeletal features" (Maraffi et al. 2013). For instance, Figure 6 shows a partially occluded pose with an asymmetry score of 89.2, while a pose with the torso angled towards the camera so the arms are more visible achieved a score of 92.5 by our metric.

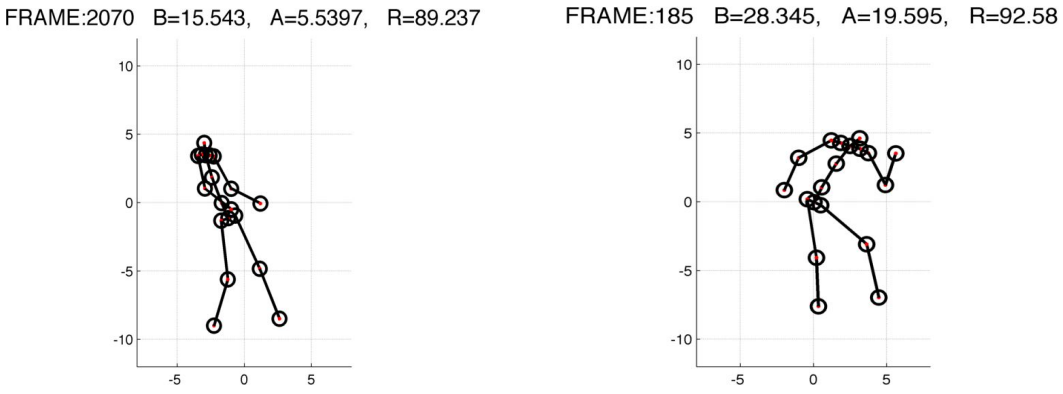



Figure 6: Application of our metrics shows Asymmetry being scored higher in the figure angled towards the camera so it is less occluded (Maraffi et al, 2013).

## 4.3 Pilot Study Metrics Analysis

We tested the efficacy of our three metrics in a pilot study by applying them to a corpus of poses from a skilled performing artist. We motion captured the subject performing a series of stretching exercises designed to test the motion capture system, and then some dance and martial arts movements. Figure 7 shows a plot of the BAR features as our metrics score them for a part of the sequence, and some corresponding poses of the skeletal joints.

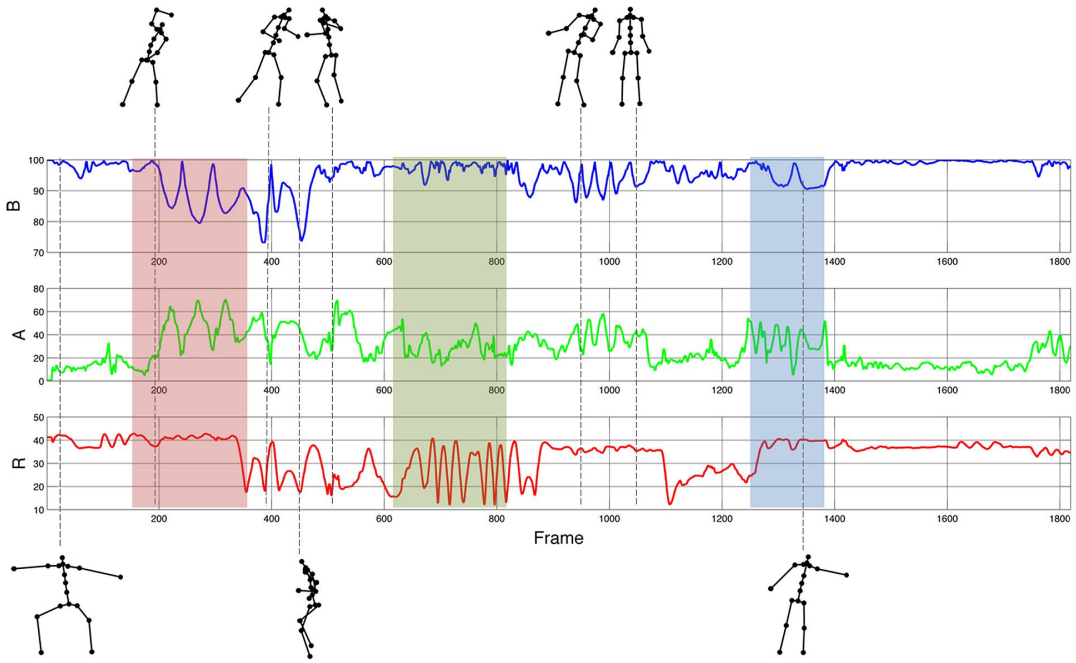
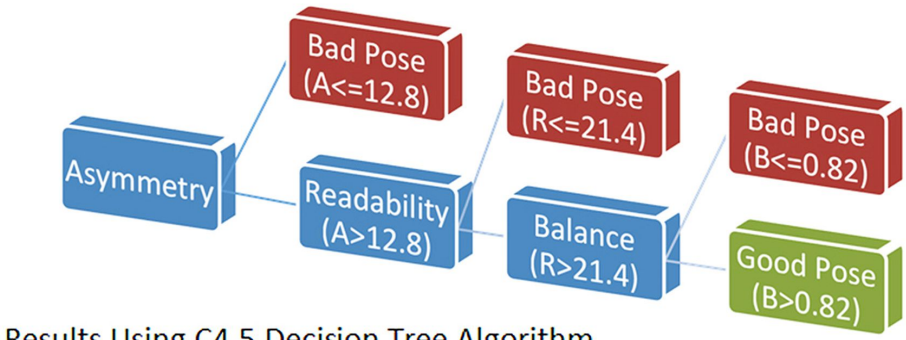



Figure 7: Plotting our BAR metrics to score poses of a skilled dancer in an initial pilot study (Maraffi et al, 2013).

The sequence starts with a T pose as the baseline, and then shows a lateral stretch, followed by a 360 degree spin, and some dance movements. Note how more occluded side poses during the spin from frames 400-500 cause the Readability score to drop.

The dance pose around frame 950 measures relatively high in all three features, while the at rest pose around frame 1050 measures low in balance and asymmetry, but high in readability. A second phase of the pilot study involved supervised machine learning on the perceived quality of poses from the corpus. We had the skilled participant and some less skilled subjects rate rendered skeletal poses as either "good" or "bad" on a training subset of the corpus. We then trained a C4.5 decision tree learning algorithm in Weka on the binary pose ratings, as well as on our BAR ratings. We then ran the trained algorithm on a test set from our corpus, and "using a 10-fold cross-validation procedure, the resulting decision tree was able to rate “unseen” poses with an accuracy of 79.7 percent" (Maraffi et al. 2013). While just initial results on a small scale, Figure 8 shows a confusion matrix that supports our hypothesis that poses ranked higher for our BAR metrics are qualified as "good" poses by spectators.




Results Using C4.5 Decision Tree Algorithm

| | Good Pose | Bad Pose |
|---|---|---|
| Good Pose | 45 | 17 |
| Bad Pose | 9 | 57 |

**79.7% Correct**

Figure 8: Confusion matrix showing the results of our pilot study support the hypothesis that poses perceived as "good" are ranked higher by our BAR metrics *(Maraffi et al, 2013)*.

## 4.4 Dance Game Experiment to Quantify Performer Skill

We conducted a more formal study of our target features using performance captured data of participants playing a Kinect dance game (see Figure 9). We chose a dance game because they are designed to be easily learned by the general public, and because dance incorporates artistic movement while minimizing narrative components. 20 participants volunteered for our experiment, ages 18-48, with a variety of experience in dancing and videogames. Each learned the same choreography for the popular song *Beat It* in the Michael Jackson Experience (MJE) game. Each participant was given a brief warm-up tutorial where they learned the choreography, and then they danced to the song twice in our Organic Motion markerless motion capture studio. The round with their highest game score was saved for further processing and analysis, along with the score itself. In addition, they were given a self-assessment survey with questions designed to assess their performing arts training. From these questionnaires we initially ranked the dance data as 10 skilled and 10 unskilled performances.


Figure 9: We collected dance data for a formal study that had participants play a Kinect dance game in our markerless motion capture studio *(Maraffi et al, 2013)*.

Dance data was captured directly into Autodesk Motion Builder, and then processed in Autodesk Maya to a normalized skeleton that had no distinguishing features. We then extracted a corpus of 56 target poses from each sequence, as well as 10 neighboring frames to account for slight deviations from the choreography, to analyze according to the BAR metrics. Some initial unsupervised learning in the form of Principal Component Analysis (PCA) was done in Weka to better visualize the structure of the data, and to reduce the dimensionality. As per our report, "The top-three principal components, or PCs, accounted for 89.1 percent of the overall variance in the data (51.5, 26.9, and 10.7 percent, respectively)" (Maraffi et al. 2013). Figure 10 shows a plot of the data according to the top two PCs, with colors indicating performer skill level according to our initial survey assessment. The data from the two groups displays enough clustering to suggest "there may exist some linear

combinations of the 114 pose parameters that are capable of distinguishing between skilled and unskilled poses." (Maraffi et al. 2013).

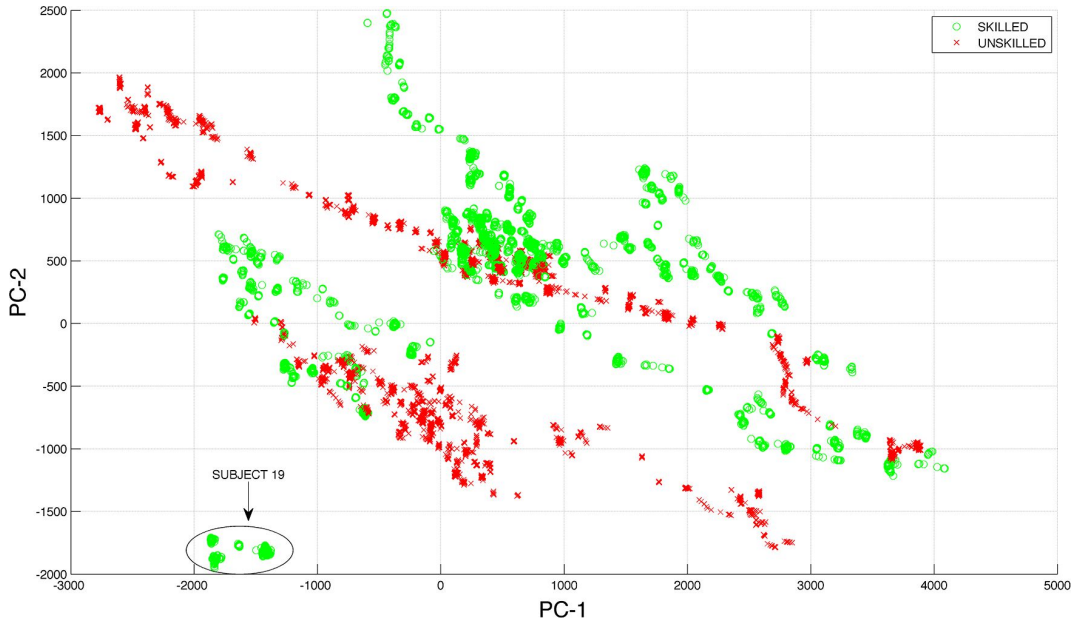



Figure 10: PCA plot of the top two principal components shows clustering that indicates players labeled skilled (green) and unskilled (red) are performing somewhat differently *(Maraffi et al, 2013)*.

While the PCA of the data indicates that the participants we labeled as skilled and unskilled seem to be doing something different, it should be noted our binary classification covers a range of skills, in that most of the participants were not professional level performing artists. Only one participant had an extensive enough dance background to be considered professional skill level, or Subject 19 on the plot, and who's data clearly clusters separately from the other subjects. Future studies should include more data from known professionals as well as amateurs. To further

visualize the structure of the data, "we counted the number of times each parameter was used as a term in the linear transformation of the top 10 PCs" (Maraffi et al. 2013). We interpret the bar plot of the top 10 PCs as being strongly influenced by the MJE choreography, where the right leg is featured prominently (see Figure 11). However, some of the top PCs could be related to BAR features, such as spine angles indicating backbone lean may be a contributing factor.

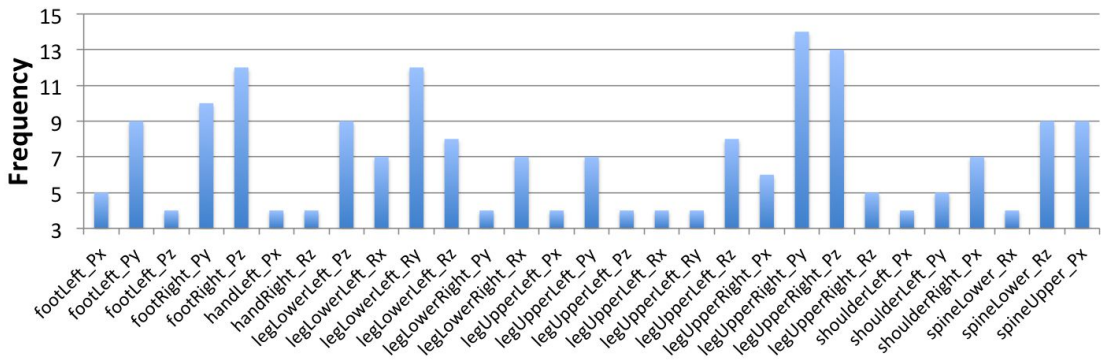



Figure 11: Bar plot of the top ten PCs shows that the legs and spine have the highest frequency in the dance data *(Maraffi et al, 2013)*.

We again used a C4.5 decision tree for supervised learning in Weka, using half the data from an equal number of skilled and unskilled performances to train the classifier, and then to qualify the remaining test set. " The decision tree was able to predict the skill level of the remaining (unseen) performers' poses with up to 81.3 percent accuracy" (Maraffi et al. 2013). Figure 12 shows that that the classifier was better at classifying skilled over unskilled performances.

| | Skilled | Unskilled |
|---|---|---|
| **Skilled** | 2872 | 208 |
| **Unskilled** | 946 | 2134 |

Figure 12: Supervised learning was able to predict "skilled" and "unskilled" performances in the dance data with 81.3 percent accuracy using a C4.5 decision tree classifier *(Maraffi et al, 2013).*

To further investigate the intuitive relationship between performance quality and perceived skill, in Weka we checked to see if game scoring supports our classification of "skilled" and "unskilled players. "The scores in the skilled group ranged from 110K to 144K, and in the unskilled group ranged from 84K to 128K. To determine if these scores came from different distributions, we ran a two-tailed test" (Maraffi et al. 2013). Figure 13 shows that game scoring from the two groups fall into two distinct distributions (with a p-value = 0.0016).

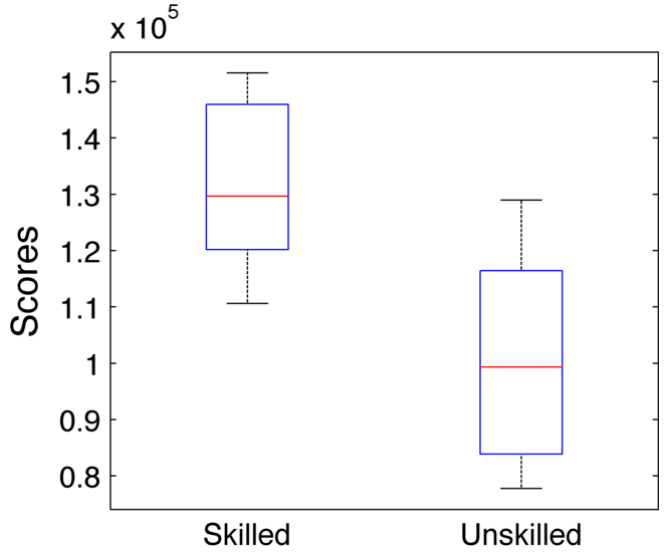



Figure 13: Game scores show two distinct distributions for dancers classified as "skilled" and "unskilled" by our study. *(Maraffi et al, 2013)*.

In these initial experiments we began using statistical methods to study aesthetic features in animation data, to both quantify pose features and qualify performer skill. The methodology we are developing uses a combination of unsupervised and supervised learning on performer and spectator data to better understand how skilled figure compositions can affect viewers. Because figure arts practices have not been studied computationally to formalize design principles, much more work needs to be done in this area. The last chapter of this thesis will discuss motivations and future work.

# Chapter 5

# Research Motivations and Future Work

In this final chapter, some of the motivations for conducting computational aesthetics research on figure data is discussed, and future work in this area is outlined. We consider the artistic, cultural, and technological reasons for developing a methodology for studying figurative design practices using statistical and machine learning. We also outline the theoretical foundations for such a methodology in the larger context of computational analysis of the arts, and what foundational work in this area may facilitate in interactive media. Lastly, we propose future work to deepen and broaden this area of research, such as a systematic study of Disney's animation principles, and interaction of figure and camera metrics for enhancing player experience in videogames.

## 5.1 Computational Performatology Approach

One of the foundational aspects of using computational aesthetics to study figure arts data is to scientifically test arts theory. One fundamental hypothesis implied in both Barba and Disney's principles is that a body represented by artistic gesture techniques is aesthetically different than how bodies are represented in real life. Specifically, artistic training is believed to develop non-natural skills that enable figures to be posed beyond everyday parameters, making them more believable and appealing than realistic representations. If these claims can be scientifically verified through experiments like the ones presented in this thesis, then it could have profound implications for designing interactive characters, because realistic gesture would no longer be the goal of AI design. Behavior designers for interactive characters could stop trying to mimic reality, and instead work at simulating arts conventions practiced by animators, puppeteers, and actors. Accurate metrics for predicting the perception of pose quality by viewers would mathematically formalize what skilled artist already do routinely, which can then inform the design of algorithms for generating character gesture that will align with media consumer preferences. Such algorithms would continue the technological trend in the arts towards externalizing character control through technology (see Figure 14). It would also start to develop a formal knowledge base, or performative ontology (aka performatology), for pre-expressive figure design principles in new media.

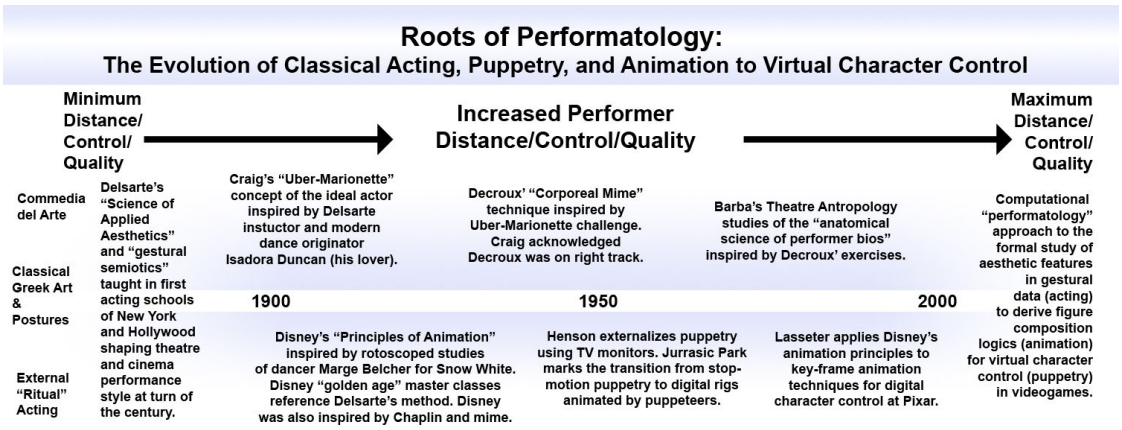



Figure 14: This thesis is motivated by a technological trends in the arts towards virtual control of animated characters.

We don't anticipate that a performative ontology for pre-expressive figure design will have the same structure as an ontology for expressive narrative, but instead would feature a general set of mathematical heuristics for generating appealing figure compositions in relation to the camera. Disney's principles are a knowledge base used throughout the animation industry to generate appealing animated figures. A formal ontology with similar functions would do the same, only as pre-expressive gesture logics calculated in real-time. Another motivation behind this thesis is the broader hypothesis that figurative arts have been converging through technological innovations since the beginning of cinema (see Figure 15), and that learning figure logics from artist generated data is the next intuitive paradigm for representing characters in media.

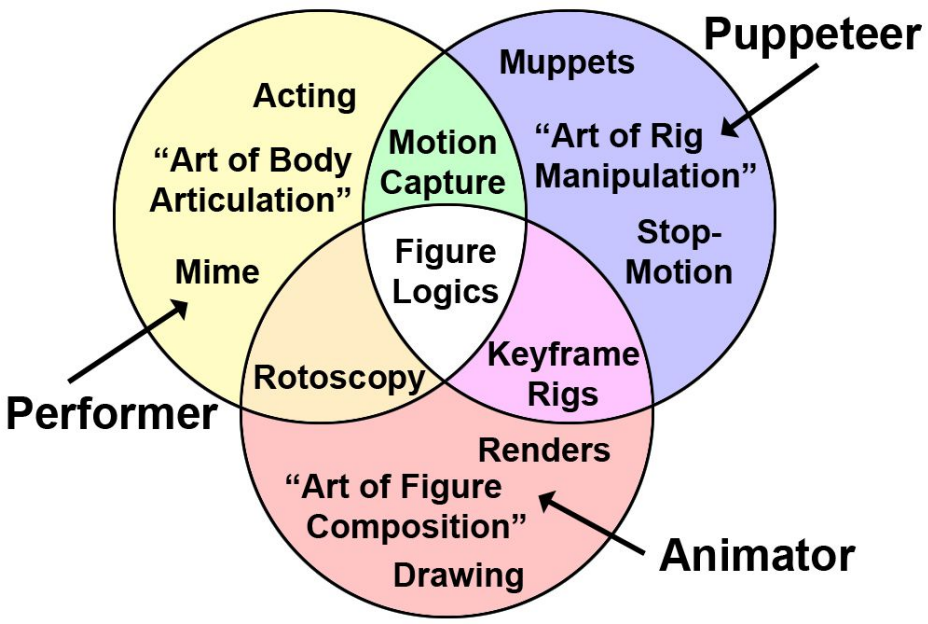



Figure 15: Motivations include a convergence of animation, puppetry, and live performance techniques in figure logics learned from data in motion capture and keyframe animation.

## 5.2 Methodological Design

In addition to related work in computational aesthetics and figure arts principles, the performatology approach in this thesis is motivated by videogame design methodologies. One of the few interaction design methods that incorporates aesthetics is the popular MDA (Mechanics, Dynamics, Aesthetics) framework (Hunicke et al, 2004), which the author expanded to better define the components for teaching purposes (see Figure 16).

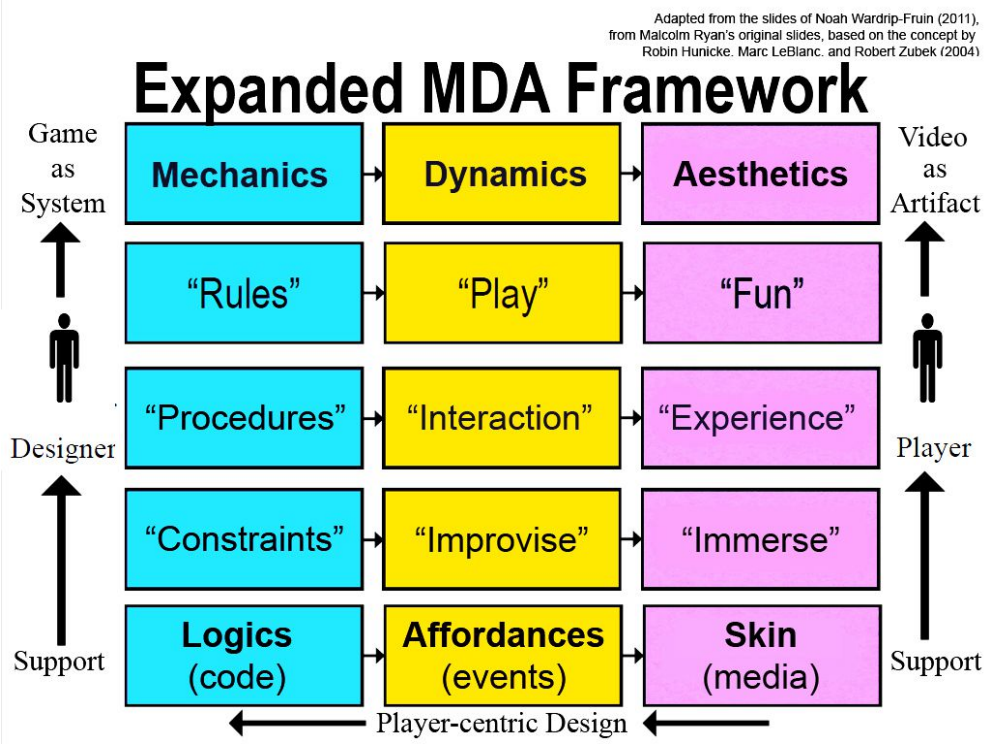



Figure 16: The methodology in this thesis was influenced by an expanded version of the popular MDA videogame design framework.

The author's MDA revisions led to an ASO Performatology framework adapted for figure interaction design in videogames (see Figure 17). The new framework places Aesthetics foremost in designing figurative player interaction, and replaces Mechanics with Semeiotics, and Mechanics with Ontology. The methodology followed in this thesis can be seen in the third row, where viewer preference studies qualify aesthetics affect ( as binary classifications of skilled/unskilled), figure data analysis quantifies gestural skill (as BAR ratings), and the formalization of arts principles to support a performative ontology (as formalized design principles). The

ASO framework includes other important concepts in arts and videogame research, such as flow and affect, which will be the subject of future computational studies.

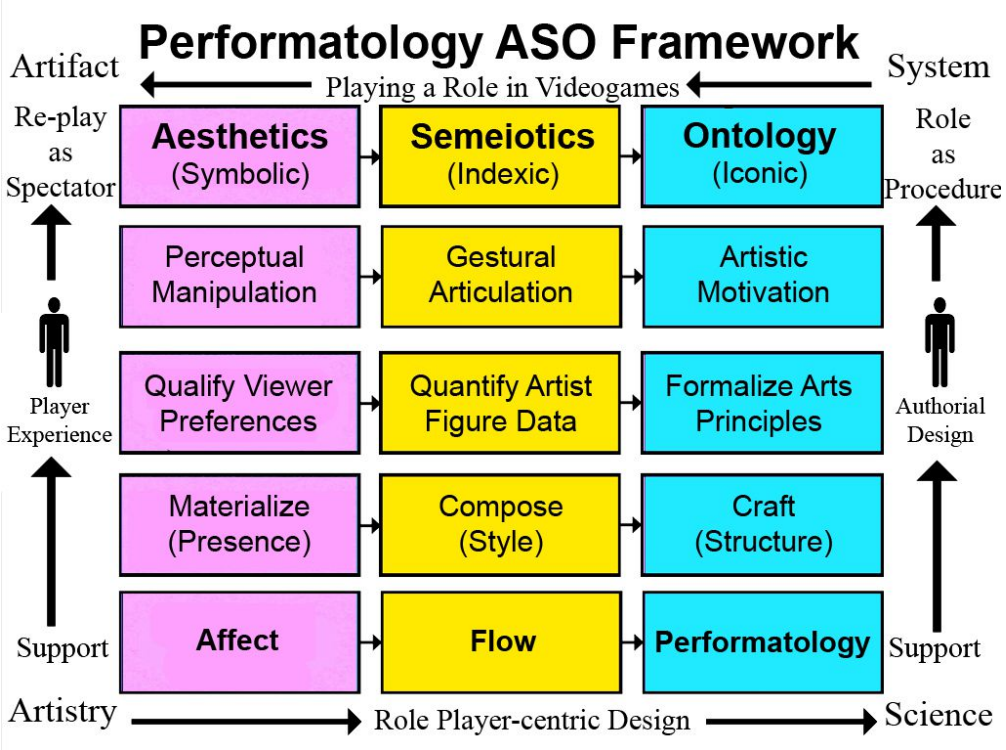



Figure 17: The methodology in this thesis is based on a new ASO framework that takes a performatology approach to interactive figure design.

## 5.3 Future Work

While this thesis begins laying the foundations for the computational study of figure arts data to formalize pre-expressive design rules, there is still much future work to be done in this area. We derived the BAR features from Disney's and Barba's principles, but a more systematic study would clarify what skill features are relevant to the perception of figure quality in each domain, and exactly which features overlap. The most logical place to start would be to analyze the use of Disney's principles in classic

Disney animation, such as scenes from Snow White. This would remove unknown factors related to skill and quality, since history has proved the animation to be affective, and instead allow the study to focus on how the principles are applied by professional animators. The 2D animation would have to be "lifted" to a 3D skeleton, either by hand or through image processing (Jain et al, 2009), and then PCA could be run to analyze the structure of the data, as well as supervised learning using metrics for principles like squash and stretch. The goal would be to quantify features related to each principle, and how the principles relate to each other in the context of the scene. It would be interesting to see if animator style is recognizable and predictable through metric scores that quantify how the principles are combined in a figure composition. Similar work could be done with Barba's principles by lifting the figures of professional live performers from 2D video to 3D animation for aesthetics analysis. A thorough investigation of Disney and Barba's principles would require expanding the methodology presented in this thesis for poses to incorporate the dimension of time, and the structure of a sequence of poses for motion analysis to be done.

Another area for collaborative future work is to computationally analyze how cinematographers and live performers work together to generate appealing visual experiences in movies. Combined camera and figure metrics that can accurately score the quality of framed animated figures would have useful applications in cinematic 3D videogames. Since the camera is a general representation of the spectator in cinema, with the player being both spectator and actor (through avatars) in

videogames, combined such studies could have a direct impact on player experience. The relationship of these two aspects of player experience has not been thoroughly studied, especially the potential connection between the performer's experience of flow and the spectator's experience of affect. Though both flow and affect have been the subject of psychology studies related to media and videogames (Csikszentmihalyi, 1970; Csikszentmihalyi et al, 1990; Chen, 2007; Engeser, 2012), the complex relationship between the two that likely exists in playable media has not been investigated to our knowledge. We hypothesize that a combined cinematography and performatology approach to computationally studying both these aspects of player experience could reveal features which could impact future videogame design.

In summary, this thesis adapts computational aesthetic methodologies from 2D image composition studies to the analysis of pose design in motion captured 3D figurative data. Analysis consists of unsupervised PCA applied to skeletal data to reveal low-level structural features, and supervised machine learning using metrics designed to rate high-level pose features according to professional arts practices. Initial results from our dance game experiments to quantify features of balance, asymmetry, and readability in selected poses warrant further study. Broadening the scope of this work, through more data collection and exploring additional features, has the potential to correlate artistic gesture metrics with viewer preference studies on pose quality. Since similar data analysis techniques seem to be applicable across related design domains, 2D image composition and 3D figure posing, future work in a broader scope may lead to a more general methodology for studying design features in entertainment artifacts.

# References


Adams, Brett et al. (2002). Towards Automatic Extraction of Expressive Elements from Motion Pictures: Tempo. IEEE Transactions on Multimedia, vol.4, no.4, pp.472-481.

Barba, E. (2003). The Ripe Action. In Mime Journal '02-'03.

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.

Berezhnoy, I. E, et al. (2005) Computerized visual analysis of paintings. In Proceedings of 16th Int. Conf. Association History and Computing, pp. 28–32.

Brand, Matthew and Aaron Hertzmann. (2000). Style Machines, In Proceedings of SIGGRAPH 2000.

Bressan, M et al. (2008). An analysis of the relationship between painters based on their work. In Proceedings of IEEE ICIP '08, pp. 113–116.

Calvert, T., et al. (2005). Applications of Computers to Dance. Computer Graphics and Applications, IEEE , vol.25, no.2, March-April'05.

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

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.

Hunicke, R., et al. (2004). MDA: A Formal Approach to Game Design and Game Research, AAAI Workshop on Challenges in Game Design.

Jain, Eakta et al. (2009). Leveraging the talent of hand animators to create three-dimensional animation. In Proceedings of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation '09.

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

Lasseter, John (1987). Principles of Traditional Animation Applied to Computer Animation. ACM SIGGRAPH '87.

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.

Maraffi, C, and Jhala, A. (2013). Raising the Aesthetic Quality of Character Interaction in Cinematic Videogames, Technical Report, UCSC-SOE-13-01, UCSC.

Marsella, S., 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.

Neff, M, et al. (2008). Gesture modeling and animation based on a probabilistic re-creation of speaker style, In ACM Transactions on Graphics, 27.

Neff, M. and E. Fiume. (2008). From Performance Theory to Character Animation Tools. In journal, Human Motion Computational Imaging and Vision Volume 36, 2008, pp 597-629.

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.

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.

Seif El-Nasr, Magy. (2004). A user-centric adaptive story architecture: borrowing from acting theories, In ACE '04 Proceedings of the 2004 ACM SIGCHI.

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

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

Tanenbaum, J and Tanenbaum. (2008). "Improvisation and Performance as Models for Interacting with Stories", In Proceedings of ICIDS '08.

Tanenbaum, J. (2011). Being in the story: readerly pleasure, acting theory, and performing a role. In Interactive Storytelling '11.

Thiebaux, M., et al., (2008) SmartBody: Behavior Realization for Embodied Conversational Agents. Proceedings of AAMAS'08.

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

Yeh, Che-Hau et al. (2010). Personalized Photograph Ranking and Selection System. MM'10.

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