A Study of How the Technological Advancements in Capturing Believable Facial Emotion in Computer Generated (CG) Characters in Film has Facilitated Crossing the Uncanny Valley

A Research Report submitted in partial fulfilment of the
requirement for the Degree of Masters of Arts in
Digital Animation at the
University of the Witwatersrand (School of Digital Arts)
Johannesburg, South Africa

By

Clare Louis

Supervisor: Bronwyn Horne

17 February 2014

Declaration

I declare that this research report and abstract are my own unaided work. Proper accreditation has been given to all outside sources.

This report has not been submitted to another university or institution of higher learning. No part of this report has been published in any journals or social media.

Clare Louis

on this 17th day of February 2014.

Acknowledgements

A very big thank you to my supervisor, Bronwyn Horne, without whose guidance and support this journey would have had a different ending.

Another very big thank you to my family for each of your contributions of support, especially my brother Sylvester Louis and nephew, Bradley Louis.

Shayleen, Tessa, Francois, Navan, Fawzia and Treneru Darijo, your encouragement throughout this year has been unfailing and appreciated more than you know. I thank you for keeping the year alive with possibility.

An extension of gratitude goes out to:

- Peter Busch, Vice-President of Faceware Technologies, division of Image Metrics, for taking the time for a very informative interview and the questions that followed over the months.
- Andre van der Merwe and Martin of Flying Circus Studios here in Johannesburg for the demonstration and patience with the questions that followed over the ensuing months.
- Werner Botha and Drew for the additional demonstrations of their motion capture set-up at AnimMate Animation Studios.
- Paul Debevec and his team at ICT at the University of California for the demonstrations on the different versions of the LightStage and detailed explanations of world-first facial scanning techniques used to generate stunning texture and light maps.
- Kan Anant, Systems Engineer at PhaseSpace Inc, San Francisco for his time and demonstrations on their facial capture technology.
- Julius Tuomisto, co-founder of NI Mate, based in Helsinki, Thibaut Weise, founder of Faceshift AG, based at the University of Zurich and Mikhail Nikonov, CEO of iPisoft.(Russia) Thank you for the interviews, explanations and introducing what the Microsoft Kinect is capable of in terms of low-end (but still good) facial motion capture. I'm very interested to see what the next few years will bring.

Clare Louis

Supervisor: Bronwyn Horne

WSOA 7044 Research Report - MA Digital Animation

15 February 2014

A Study of How the Technological Advancements in Capturing Believable Facial Emotion in Computer Generated (CG) Characters in Film has Facilitated Crossing the Uncanny Valley

Abstract

In recent years, the quest for capturing authentic emotion convincingly in computer generated (CG) characters to assist exceedingly complex narrative expressions in modern cinema has intensified. Conveying human emotion in a digital human-like character is widely accepted to be the most challenging and elusive task for even the most skilled animators. Contemporary filmmakers have increasingly looked to complex digital tools that essentially manipulate the visual design of cinema through innovative techniques to reach levels of undetectable integration of CG characters.

In trying to assess how modern cinema is pursuing the realistic integration of CG human-like characters in digital film with frenetic interest despite the risk of box office failure associated with the *uncanny valley,* this report focuses on the progress of the advances in the technique of facial motion capture. The *uncanny valley* hypothesis, based on a theory by Sigmund Freud, was coined in 1970 by Japanese robotics professor, Masahiro Mori. Mori suggested that people are increasingly comfortable with robots the more human-like they appear, but only up to a point. At that turning point, when the robot becomes too human-like,

it arouses feelings of repulsion. When movement is added to this equation, viewers' sense of the uncanny is heightened when the movement is deemed to be unreal.

Motion capture is the technique of mimicking and capturing realistic movement by utilising technology that enables the process of translating a live actor's performance into a digital performance. By capturing and transferring the data collected from sensors placed on a body suit or tracked from a high definition video, computer artists are able to drive the movement of a corresponding CG character in a 3-Dimensional (3D) programme. The attention of this study is narrowed to the progress of the techniques developed during a prolific decade for facial motion capture in particular. Regardless of the conflicting discourse surrounding the use of motion capture technology, these phenomenal improvements have allowed filmmakers to overcome that aspect of the *uncanny valley* associated with detecting realistic movement and facial expression. The progress of facial motion capture is investigated through the lens of selected films released during the period of 2001 to 2012. The two case studies, *The Curious Case of Benjamin Button* (2008) and *Avatar* (2009) were chosen for their individual achievement and innovative techniques that introduced new methods of facial capture.

Digital images are said to undermine the reality status of cinematic images by challenging the foundation of long held theories of cinematic realist theory. These theories rooted in the indexical basis of photography, have proved to be the origin of contemporary viewers' notion of cinematic realism. However, the relationship between advanced digital effects and modern cinematic realism has created a perceptual complexity that warrants closer scrutiny. In addressing the paradoxical effect that photo-real cinematic realism is having on the basic comprehension of realism in film, the history of the seminal claims made by recognized realist film theorists is briefly examined.

Contents

Introduction **9**

Chapter 1: Realism in Film **15**

1.1 Introduction to Realism 15
1.2 Formalist (or Formative) Theory versus Realist (or Photographic) Theory 16
1.2.1 Formalist (or Formative) Theory 16
1.2.2 Realist (or Photographic) Theory 17
1.3 Types of Cinematic Realism Affecting Modern Spectatorship 18
1.3.1 Cognitive Illusion 19
1.3.2 Perceptual Realism 19
1.4 Digital Visual Effects challenges Cognitive Illusions and Perceptual Realism 20
1.5 Conclusion: Realism in the Age of Digital Cinema 21

Chapter 2: How the *Uncanny Valley* Affects Viewer's Perception of Realistic Integration of CG Characters in Contemporary Digital Film **23**

2.1 Introduction 23
2.2 Characteristics of the Uncanny ('*Unheimlichen*') 25
2.3 Mori's Theory of the *Uncanny valley* translated to CG Human Characters 26
2.3.1 Mori's Graph 27
2.3.2 Detecting the *Uncanny valley* Response in Digital Animation 29
2.4 Avoiding the *Uncanny valley*. 31
2.5 The Digital Emily Project 32
2.6 Conclusion 36

Chapter 3: The Development of Facial Motion Capture Techniques **37**

3.1 Introduction to Motion Capture 37
3.2 Performance Animation and Facial Motion Capture Processes 38
3.2.1 Performance Animation 38
3.2.2 Facial Motion Capture Processes and Technology 38
3.2.2.1 Marker-Based Systems used for Facial Motion Capture 39
3.2.2.2 Marker-less Systems for Facial Motion Capture 40

3.2.2.3	Performance Driven Facial Animation (PDFA)	41
3.3	Technical Developments in Facial Motion Capture	44
3.3.1	Introduction of Facial Action Coding System (FACS)	44
3.3.2	Facial Action Coding System (FACS) -	45
3.4	Movies that fell into the Uncanny Valley yet contributed to advancing Facial Motion Capture Techniques	48
3.4.1	Final Fantasy (2001)	48
3.4.2	The Polar Express (2004)	49
3.4.3	Beowulf (2007)	50
3.5	Personal Graph depicting Technological Advances in Facial Motion Capture based on graph of Mori's Uncanny Valley Theory	51
3.6	Conclusion	52

Chapter 4: Case Studies **55**

	Introduction	55
4.1	**Principal Case Study – Case Study 1: *The Curious Case of Benjamin Button* (2008)**	56
4.1.1	The Facial Motion Capture Process: Identifying the Challenges	57
4.1.2	The Facial motion Capture Process: Overcoming the Technical Challenges to Convincing Facial Animation	59
4.1.2.1	Mova Contour	61
4.1.2.2	Image Metrics (Faceware Technologies)	62
4.1.3	Crossing the Uncanny Valley	64
4.1.4	Conclusion	65
4.2	**Case Study 2: *Avatar* (2009)**	66
4.2.1	Facial Performance Replacement on Avatar	68
4.2.1.1	Head-mounted Camera (Headcam)	68
4.2.2	Enhanced Muscle System and the dynamic FACS Solver	71
4.2.3	The Simulcam	73
4.2.4	Conclusion	76

Figures

Fig. 1.	Linear graph chronologically representing the eleven movies referenced in the report	13
Fig. 2.	Mori’s Uncanny Valley graph	28
Fig. 3.	A performance in a typical optical, marker-based motion capture stage	39
Fig. 4.	Stages of the photogrammetry process that Mova Contour perfected for *Benjamin Button*.	42
Fig. 5.	Dr Paul Eckman demonstrating 5 of the 72 facial expressions human faces are capable of.	45

Fig. 6. Calibrating phase - setting poses for the FACS Solver for *Monster House*. 46
Fig. 7. FACS muscle diagram, illustrating the breakdown of the underlying facial muscles 47
Fig. 8. Personal graph based on Mori's graph of the *Uncanny Valley* 53
Fig. 9. Digital version of Brad Pitt as Benjamin Button. 57
Fig. 10. Clear aging of Brad Pitt's character Benjamin Button with retargeted mesh. 61
Fig. 11. Brad Pitt filmed with four high resolution cameras during the performance capture process 63
Fig. 12. Sam Worthington as Jake Sully on the motion capture stage 70
Fig. 13. James Cameron filming with the Simulcam on location. 74
Fig. 14. View from the Simulcam of Jake Sully's Na'vi avatar (Sam Worthington) and Neytiri (Zoe Saldhana).
Fig. 15. Dr. Mark Sagar 89
Fig. 16. Peter Busch at Image Metrics Office in Santa Monica, California 95

Conclusion **78**

Works Cited

Appendices

i. **Appendix 1:** Dr Mark Sagar of Weta Digital, New Zealand -Developer of the FACS Solver and Muscle System [transcript of podcast - FXGuide.com] 89

ii. **Appendix 2:** Personal Interview with Peter Busch **-**Vice President of Business Development for Faceware® Technologies at Image Metrics Head Office (Santa Monica, Los Angeles) 95

iii. **Appendix 3:** Joe Letteri - Senior Visual Effects Supervisor at Weta Digital, New Zealand - FACS Solver and Muscle System [transcript podcast- Mike Seymour interviews Joe Letteri. FXGuide.com-Podcast] 100

iv. **Appendix 4:** Demonstration (photographs) and Interview with Kan Anant, Product Manager/ Systems Engineer at PhaseSpace Inc (San Francisco, California) at SIGGRAPH2012 in Los Angeles 109

Introduction

The stealthy manner in which the computer generated (CG) character Gollum spider crawls his way down the side of the mountain towards the sleeping hobbits in Peter Jackson's *Lord of The Rings: The Two Towers* (2002) had never been seen as convincingly on screen before, at the time of its release. A profound sense of curiosity at how the animators had achieved the clear behavioural complexity of emotion demonstrated in the schizophrenic soliloquy between the CG characters of the monstrous, sadistic Gollum as he taunts and mocks his alter-personality, the softer, frightened Sméagol serves as the impetus for this research. That unsurpassed level of believability of the movement and facial expressions that Gollum exhibits on screen is echoed in a research report by Kelly Christophers in her observation of him as "a CG character possessed of a personality and emotional appearance like no other CG character before" (48).

Audience reaction was equally similar to the visible expressiveness demonstrated by the digital lead character in David Fincher's *The Curious Case of Benjamin Button* (2008) and the Na'vi species in James Cameron's *Avatar* (2009). The success of these movies is contrasted with the general negative discourse in the digital effects industry towards other earlier animated CG characters also attempting realistic facial movement using similar technology, such as *The Polar Express* (2004) and *Beowulf* (2007). The clear dichotomy presented by the disparity in viewer response to these thematic films became the motivation behind this report. In exploring a plausible explanation to the question of why some movies succeeded while others were not successfully received, the research is divided into three contextually related parts. The first part will explore where notions of realism in the cinematic context originate and attempts to explain why this avenue of digital effects is being pursued with such vigour. The second part of the report will concentrate on a psychoanalytical analysis of the dangers of these pursuits when they fall short in the form of the *uncanny valley*, and the third will celebrate the technological achievements in facial motion capture that have made it possible to overcome the challenges posed by this pursuit.

In her book *Digital Storytelling,* Shilo McClean's remark that "spectacle for spectacle's sake or as an expansion of the stylistic devices available" (10) encapsulates the theoretical discourse on the prominent, flashy reputation exhibited by digital visual effects blockbusters of the 1980s and 1990s. Two decades later finds digital effects striving to be sensually immersive, and no longer the counter-narrative spectacle they were previously branded as. Chris Nolan, director of the latest Batman trilogy, famously tried to use as few CG elements as possible in the series. However, as a modern filmmaker, he explains that when he was obliged to turn to CG manipulation, he was always mindful that "… anything you notice as technology reminds you that you're in a movie theatre. Even if you're trying to portray something fantastical and otherworldly, it's always about trying to achieve invisible manipulation" (wired). In an effort to blend the fantastical with a photographic medium, audience's sense of the 'unreal' are heightened. The importance of understanding the impact of introducing something new into a fundamentally traditional medium like film, and how viewers perceive the new designs these digital tools have created, is one of the concepts at the core of this research.

As one of the predominant theorists in this report, Stephen Prince's[1] extensive research has integrated a wide range of empirical and theoretical insights on the consequence of visual effects in contemporary cinema. His contemporary books, in particular, *Digital Visual Effects in Cinema: The Seduction of Reality* (2012), *Movies and Meaning* (2013) and article *True Lies: Perceptual Realism, Digital Images and Film Theory* serve as a guide through this investigative, interpretive analysis while navigating the paradoxical effect that photo real cinematic realism has on the basis of our comprehension of realism in film. This report focuses on the emulation of realistic human movement, however an understanding of how realism operates on a deeper level of emotion and empathy need to be investigated to gain a broader understanding of why viewers react with revulsion and confusion when presented with an animated CG character that does not meet that expectation. Part of the history of our notions of realism is articulated in essays by major realist theorist Andre Bazin in his book, *What is Cinema?* (1969). Realist theory is based in the indexical[2] nature of film.

[1]Stephen Prince, a Professor of Cinema has published numerous books focused on digital visual effects and film studies.For over two decades, his research has been cited in diverse scholarly works
[2]Indexicality is defined as the relationship between the photographic image and its referent. Further explanation of this relationship is discussed in Chapter 1 of this report.

In order to grasp the tension within cinema "between its recording functions and the power it gives filmmakers to stylise and transform reality" (411), a documentation of how the advances in technology have freed the director from the indexicality of the photographic likeness is required. In assessing how digital cinema is now posing "a challenge for film theory when it attempts to locate a basis for realism and realistic film styles" (411), formalist and realist film theory is supeficially introduced and discussed in the context of this report in Chapter 1.

A more pragmatic measurement of realism derived from interest in David Bordwell and Janet Staiger's relevant approach to the *business* of realism in the digital age is skilfully summarised in Lev Manovich's papers, *Assembling Reality* and *New Language Media*. Research into Bordwell and Staiger's analysis of "the relationship between the character of realism in computer animation and the particular industrial organization of the computer graphics industry" (5) reveals an emphasis on cinema being an industry like any other -with stated goals and quantifiable elements that will be discussed in the report. Realism being one of these goals, makes their views on how the unavoidable business of technology affects facial motion capture in the industry.

The quest for imperceptible integration of CG characters into live-action film has inherited an unintentional psychological phenomenon; one that could negatively or positively impact on the outcome of these thematic movies if the viewers' notions of reality are not met. The technique of motion capture has allowed actors to interact with the film realities constructed on screen as never before but with results that are unexpected when they fall foul of the dilemma noted in the *uncanny valley* theory described in Chapter 2. Originally presented in the field of Robotics, Mori suggested in his article published in 1970, *The Uncanny Valley,* that people are increasingly comfortable with robots the more human-like they look and act, but only up to a point. At that turning point, when the robot becomes too human-like, it arouses feelings of creepiness and repulsion arising from the realisation that you are no longer looking at a human-like machine, but instead, at a person that isn't quite 'right' (98). An abundance of discourse on the impact of Mori's theory since it has been applied to the field of computer graphics can easily be sourced. A sense of unease due to a lack of believability portrayed in the robots translates on a psychological level to how viewers respond to CG digital humans - and is the reason cited for *The Polar Express'* poor reception in the animation industry. This conceptual dilemma relates to all areas of the report as it affects viewers' perception of realism. Evolutionary evidence and empirical psychoanalytical evidence regarding this theory will be included in the examination. Frank Pollick, Professor in Psychology at the University of Glasgow's research will serve as a measure for detecting and quantifying the 'uncanny effect' and the impact that movement has on perceptions. An investigation of characteristics of the uncanny will aid in explaining why films such as *Final Fantasy* (2001), *The Polar Express* and *Beowulf* fell into the *uncanny valley* and how *The Curious Case of Benjamin Button* successfully avoided it.

It must be acknowledged that advancement in the areas of digital compositing, 3D lighting -how the light convincingly bounces off the skins these CG creatures - rendering and texturing based on real world evidence have all contributed to enhance an illusion of the detail which viewers find so fascinating. However, the focus of this report is on movement, specifically facial, an area identified as highly problematic in Mori's theory if not executed correctly. Motion capture is a dynamic, time-saving process for capturing realistic human movement by utilizing technology that enables the process of translating a live actor's performance into a digital performance. Chapter 3 discusses the process of capturing and outputting the collected data from markers placed on a body suit or directly on the performer's skin. With that information, 3D artists are then able to drive the movement of a corresponding CG character in a 3D programme. Alberto Menache's book, *Understanding Motion Capture For Computer Animation (2nd Edition)* is recommended for further reading as it superbly details comprehensively the early history of all motion capture techniques. The scope of this report allows for a superficial treatment of the history of motion capture and touches on rotoscoping - an old animation technique of tracing over film footage, frame-by-frame. Rotoscoping is considered to be one of the origins of motion capture. The focus of this section will be on contemporary technological improvements in facial motion capture highlighting the use of the Facial Action Coding System (FACS) - developed by Dr Paul Eckman - and what has become commonly referred to as performance capture. The reason why marker-based motion capture systems do not generate enough facial data to satisfy the deep impulse in viewers accustomed to seeing faces everyday will also be investigated.

Through texts uncovered in a variety of media sources, eleven movies have been identified as best describing the change in the cinematic landscape over the last decade. Each movie in Figure 1 has been selected for their innovative approach to facial motion capture. Their individual pioneering contributions serve as a ground-breaker for eventual Academy Award winners. When James Cameron, director and producer of the highest grossing movie of all time, *Avatar*, remarks: "We stand on the shoulders of those that came before us. We ride on top of what *Polar Express* did and *Polar [Express]*was an incredible first step in [the motion capture] arena" (YouTube, Cameron), it is an acknowledgement that his success would not have been possible had it not been for the perceived failures of motion captured films from previous years. Following an engagement of each movies' contribution to the topic, they will be plotted on a graph as a visual representation of their role in overcoming the *uncanny valley*. These movies will be referenced throughout this report for their roles in paving the way for the next technological chapter.

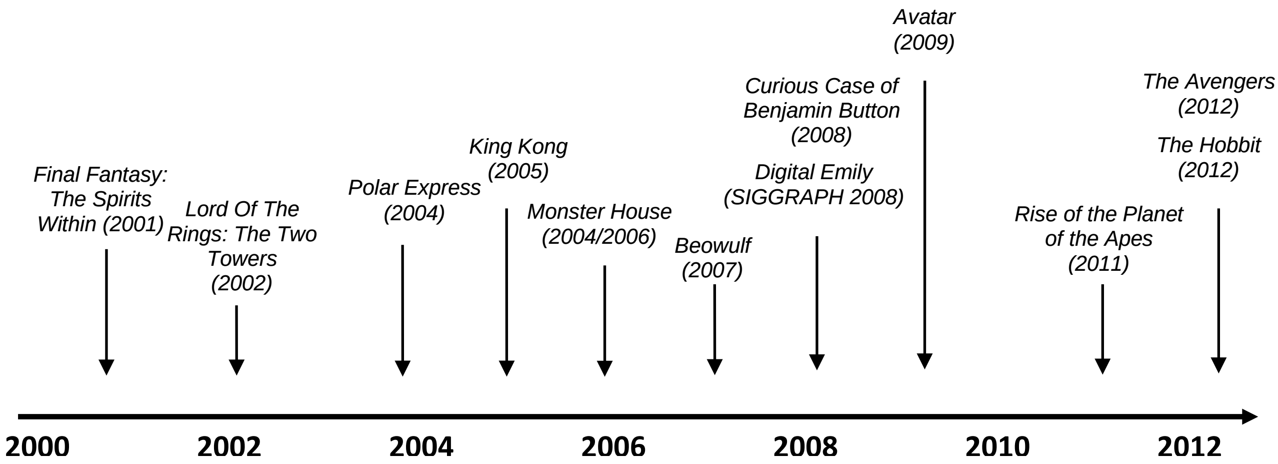


Fig 1. This linear graph chronologically represents the eleven movies and the Digital Emily project that will be referenced throughout the report for their pioneering contributions to facial motion capture of CG characters in film.

The final part of this report presents two case studies chronicling how the technological advancements in facial motion capture used on the lead character in *The Curious Case of Benjamin Button* were then employed and improved upon by James Cameron to spectacular effect in *Avatar*. *The Curious Case of Benjamin Button* will be highlighted as the principal case study for being the first Academy Award winning digital human to successfully cross the *uncanny valley*. Both movies have individually been a leap forward in innovative approaches to facial motion capture and animation. This report recognizes that the unique ontological[3] status of *Avatar* has been elevated by way of the 'simulcam'. This specially designed camera has superseded the function of a regular camera to one housing the live actor's performance and all motion capture data streams to be viewed simultaneously in real time *inside* the simulcam's unified space *before* transfer to a high end

[3] Ontology is the branch of metaphysics that studies the nature of existence or the state of 'being'

graphics programme. As the most innovative example of non-indexical cinematic technology at that time, *Avatar* will be evaluated for the contribution the team of digital artists have made to advancing facial motion capture techniques.

Several leaders in the field of motion capture in the United States and other countries were visited and interviewed in 2012. Included in the appendices is an account of an interview with Peter Busch of Faceware Technologies® - the facial motion capture product for Image Metrics. In the interview, Busch discusses their contribution to the *Digital Emily (*2008) project and *The Curious Case of Benjamin Button*. He also offered great insight into the progress of facial motion capture technology and how it has enhanced the possibilities of digital integrations to the level of believability that had never before been presented.

Chapter 1: Realism in Film

1.1 Introduction to Realism

In an informative TEDTalk presentation in February 2009, Ed Ulbrich, Chief Creative Officer of the LA-based visual effects powerhouse, Digital Domain stated that "to create a believable human form, and particularly the human head, has been considered the holy grail of our industry". He then elaborated on the Oscar-winning technology and the Herculean collaborative effort involved in creating the first convincing digital human in film, the lead character in David Fincher's *The Curious Case of Benjamin Button*. To investigate how the film industry achieved this triumphant feat of realistic integration of a CG character, a formal analysis of the litigious topic of realism follows.

The debate between reality and perception is historically exhaustive and the theme has come to encompass many variations in the context of film with the development of technology. In this research report, the intention is not to carry out extensive exegeses of realism in film, but to narrow the scope to situate realism in film in relation to developing cinematic technology. If a nuanced understanding of the manner in which perceptions are being challenged by this new development in the medium is to be reached, it is important to briefly engage with traditional film theory**.** Film theory often presents as a vast and tricky maze; - however its relevance is propagated by being one avenue of science that debates and discusses, until generally providing practical answers to the questions of those engaged in the business of creating films.

There is vast discourse on the common consensus that traditional film theory developed into two opposing ideologies based on the premise that all cinema has its roots in the work of either George Méliès (formative theory) or Auguste and Louis Lumiére (realist theory). The Lumiére brothers forged the actuality based, slice-of-life style films in the late

18th century when they took their cameras out of the studio and captured the reality around them on film for the first time. George Méliès is considered to be the progenitor of cinematic fantasy when he pioneered fundamental special effects by shooting films of fantasy - trips to the moon or adventures on the ocean floor. Contemporary discourse could define 'production' as the process employed by the Lumiére brothers. This would involve the actual shooting of the footage - and 'post-production' would be the term applicable to Méliès' process. These processes include editing, compositing, colour correcting and now, digital effects. By establishing the origins of the traditionally separate and clear margins of the rudimentary filmmaking process - that is now the film industry standard - this report will demonstrate how, with the introduction of digital effects, the boundary between these fundamental processes of filmmaking has been eroded. Stephen Prince's statement that "we will not be able to understand visual effects unless we overcome the dichotomy in our thinking represented by Méliès and Lumiére" (Prince, "Digital" 3) is correct if one is to arrive at a basis for interpreting the representational reality projected to us on screen via the "transformational function of cinema" (Prince, "Movies" 309). To aid in this endeavour to overcome that dichotomy, a select few cogitators of the enormous array of theorists stemming from either Méliès or the Lumiéres are drawn on.

1.2 Formalist (or Formative) Theory versus Realist (or Photographic) Theory

1.2.1 Formalist (or Formative) Theory

The essence of formalist theory is that film is art and not a tedious reproduction of reality. In his book, *The Major Film Theories*, Dudley Andrew elaborates on cinema's capacity to re-organise physical reality with his explanation that formative film theory "draws its strength from the burgeoning academic interest in film..." and is based on "the technical variables of film...camera, lighting, editing. Each chapter enumerates the various possibilities for artistic control, ... underscoring the 'cinematic' aspects of the medium" (76). Louis Giannetti, author of *Understanding Movies*, clarifies these cinematic aspects with his elucidation that the "art of cinema is possible because a movie is unlike everyday life ... more a translation of observed characteristics onto the forms of the medium" (4).

Formalists believe that the real world needs to be shaped and heightened for the audience. Their choppy montage[4] or very artistic style turns filmmaking into a formula of 'correct choices' that essentially "classifies formative theory as providing us with nothing but a catalogue of film effects" (Andrew 77). Their style of filmmaking relies heavily on establishing and designing their aesthetic after the film footage has been shot. Andrew warns that there is always something hollow in this kind of approach deeming "formative theory to be dangerous [as] it is centred entirely on the *techniques* of the film medium" [emphasis added] (77).

To understand the sense of the 'uncanny' that is experienced when perceptions are challenged, a more psychoanalytical approach will be required. The answer to that challenge to traditional notions of realism is rooted in the rubric of realist film theory -which is "more closely linked to a sense of the social function of the art [of filmmaking]" (Andrew 104) rather than the "distinctive technological capacity the cinema has for this kind of creative play" (Sweeney 177).

1.2.2 Realist (or Photographic) Theory

The fundamental idea of realist theory proceeds from the idea that the photograph is a recorded trace of an image captured onto celluloid by the camera. Realists' metaphysical term, *indexicality*, is used to embody that link between the photograph and the image that the camera 'preserved'. It has been a long held belief that "photographic indexicality is a commonly accepted basis for realism in cinema" (Prince, "Digital" 3) and cinema is best understood as being an indexical medium for the purposes of this report. Unlike paintings or line drawings, the photograph or film is a direct connection to its referent; a physical index or evidence of the scene and objects that were in front of the camera. An array of scholarly works consider indisputably the writings of French critic, Andre Bazin to be the most

[4] Montage is best described as the ordering of the image in time by the practice favoured by formative directors of juxtaposition editing ie building a scene out of many brief shots. In this context, Bazin opposed the *manner* of montage employed by the formative directors. (Pramaggiore and Wallis 221) He felt that viewers should arrive naturally at their own conclusions of a scene instead of being force-fed the director's view by the almost patronising manipulations of the editing of a scene.

important in realist film theory. He was the first critic to effectively challenge formative theory, by intelligently voicing the psychological influence of "the naked power of the mechanically recorded image rather than on the learned power of artistic control over such images" (Andrew 134). There is no doubt that according to Bazin, editing film amounted to nothing more than trickery. "The quarrel over realism ... stems from a misunderstanding [or] confusion between the aesthetic and the psychological; between true realism -the need to give significant expression to the world both concretely and in its essence- and the pseudo-realism, of a deception aimed at fooling the eye (and the mind)" (12) but this is the issue will be addressed in this report.

In summarising part of one of Bazin's eponymous essays, the '*Ontology of the Photographic Image'*, he rationalizes that, as man plays no part in the purely mechanical process of photography, the psychological satisfaction and acceptance of realism in this case is not in the result (photograph), but in the method used (camera) (9). This harks back to a tradition of the artist having a subjective input on his painting and a part of his reality will always be part of the form of that image. Traditional film re-produces rather than just represents the real world because of the mechanical, untouched-by-human-hand photographic method employed by the film medium. Another leading realist theorist, Roland Barthes, elaborates on this theory in his analysis of photography by noting that "unlike every other type of image, photographs can never be divorced from their referent. Photographs and referents are 'glued together' " (Barthes 5). This integral part of the philosophy of realist thinking has a bearing on this report for its value in describing the established belief of viewers' traditional relationship with realism in film. These images correspond very closely with their experience in the world, as opposed to formalist theory which is about escaping reality.

1.3 Types of Cinematic Realism Affecting Modern Spectatorship

Research has revealed a plethora of definitions and descriptions of realism in contemporary film. Transparency, Illusionism, Epistemic Realism and Photorealism- which is essentially the goal of digital visual effects- are but a few from the litany of scholarly definitions, related claims and exhaustive derivations of the branches of realism. In his chapter on *Realism* in the *Routledge Companion to Philosophy and Film*, Andrew Kania looks to Greg Currie, leading philosophical theorist. Currie has divided what Kania refers to as 'motion picture realism' (237) into two categories: cognitive illusion and perceptual realism. After extensive engagement with the literature, these definitions provide a clearer means to measure and evaluate responses to the nature of digital film more precisely. These definitions inform a pertinent investigative analysis of Mori's theory of the *uncanny valley* that is conducted in the following chapter.

1.3.1 Cognitive Illusion

"Film engenders a false belief in us, such that we are literally seeing the fictional events of a film unfold before us" (Currie 333). The viewer wills their conscious mind to accept what they imagine rather than believe what is fictionally represented on the screen. An interesting note states that one form of cognitive illusion holds that the viewer is fully aware that he is watching a film but while the film is playing, it is perceived as 'documentary footage' - as would be the case in watching a fantastical creature like the dragon, Smaug in *The Hobbit* (2013). Kania underlines that the awareness we carry does not allow us to forget that we are watching a film, as opposed to perceiving reality (Realism 239).

1.3.2 Perceptual Realism

Perceptual Realism is the difference between how a film *appears* and how it actually is, independent of individual beliefs about it. The idea that the closer the experience of film watching approximates our usual experience of witnessing the real world, "the more effectively film engenders in the viewers the illusion that they are actually watching the real world" (Currie 326). In the context of this report and in contemporary experience, Prince's definition of perceptual realism as "the replication via digital means of contextual cues designing a three dimensional world" is more tangible ("Digital" 32). Prince's perceptual realism offers a clearer explanation of how digital artists absorb information from the live-action plate to anchor a CG character into a scene - and the manner in which these images are evaluated and interpreted by viewers.

1.4 How Digital Visual Effects challenges Cognitive Illusions and Perceptual Realism

"Reality is seen not as something to be captured in the purest way possible by the camera lens but rather as something to be constructed" (Giralt 1). From the outset it is clear that Giralt's position on the viewer's relationship with reality in contemporary cinema is that "we perceive the filmic image not as a pure representation of reality" [but rather] "as a medium that communicates certain represented realities to the viewer regardless of its photographic or digitally-rendered status" (Lohmeyer 3). Lohmeyer's interpretation of Giralt's view that digital cinema does not supply an understanding of the real world, but instead extracts emotions from the audience in the form of attraction or repulsion explains how visual effects are responsible for introducing and fostering a new and radical film consciousness.

When Prince discusses the dinosaurs in Spielberg's *Jurassic Park* (1993) (Prince "Digital" 32), similar concepts of cognitive illusion can be applied to the Na'vi, the principal humanoid species in Cameron's *Avatar.* Viewers did not actually believe, nor could they be fooled into thinking that the blue-skinned inhabitants of the mythical planet of Pandora were truly alive. However, because the digital tools enabled reaching new heights of sensory detail in that exotic bioluminescent world, "viewers could be sensually persuaded to believe in the fiction and participate in the pleasures it offered" (33), invoking a sense of cognitive illusion.

To overcome these assumptions in the digital era, McClean offers by way of an explanation, that in the desire to create a CG human indistinguishable from live-action humans, "we are creating the cues that are needed to convince audiences to accept what they see" (60). This adds to the paradoxical element that the use of digital tools has ushered into contemporary film narrative, we are "creating credible photographic images that cannot *possibly* be photographed" (Prince, "True" 28) are created. French postmodern philosopher, Jean Baudrillard refers to this intangibility as *simulacra* in his book, *Simulacra and Simulation* (1981). Baudrillard defines simulacra as copies that portray things that no longer have an original -or had no precedent to an object in reality from the outset. Although this was in reference to holographs, this apposite term aptly describes this new radical consciousness that digital effects has introduced. As an outcome of digital imaging, the photo-realistic CG character of Caesar in Rupert Wyatt's *Rise of the Planet of the Apes* (2011) displays startlingly human characteristics. Based on Giralt's argument, irrespective of whether the film is indexical or not, it "does not detract from viewer acceptance of [Caesar's raw emotional] realism" (1) displayed when he has to leave the father figure played by actor James Franco. McClean further adds that "many of the images created by digital visual effects never appeared before a camera, even if they do incorporate photographically obtained elements" (63).

1.5 Conclusion: Realism in the Age of Digital Cinema

With an understanding of realism in a traditional indexical sense, the question at the heart of this chapter can be addressed: "When faced with digital images that have changed the entire landscape, will we need to discard entirely notions of realism in the cinema?" (Prince, "True" 31). In the course of researching this question, traditional theory was insufficient as it covered part of the answer. Manovich demonstrates that the differences between cinematic and synthetic realism begins on the level of ontology (12). It begins with reality, photography's antecedent and it becomes an additive process, rendering traditional theory as the beginning of the answer. However, this new cinematic realism renders traditional theory as partial and uneven. Several industry critics have expressed their displeasure to the effect that digital imaging is ringing the death knell of 'real' films because of the frequency with which profilmic reality[5] is becoming an increasingly stable part of films. A more satisfying, less abstract approach was found in Manovich's summary of David Bordwell and Janet Staiger's claim that they have "successfully reduced realism to a rational and a functional notion" (7). In an effort to define contemporary cinematic reality as more than a floating term inside virtual environments and in trying to assess the impact of digital visual effects, Bordwell and Staiger distinguishes cinema as an industry and denotes specific, modern, tangible parameters for measuring realism that is lacking in classical film theory.

[5] Much of the profilmic reality of The Lord of the Rings films for example, consists of miniatures or scenes acted against green screens. Hence, much of what is seen in the final film "never existed as such" (Jean Pierre Geuens, 20).

It would appear that these realist notions of indexicality and the negative idea of adding 'unnecessary components' are against the integration of CG objects and characters that produce a heightened illusion of reality itself. With digital manipulation, one would be forgiven for thinking that these theories are no longer tenable, however we are assured by McClean that, although Bazin's discussions and essays are "in relation to photography, with the advent of digitally created images, these ambitions remain of influence" (63). In support of McClean, Prince points out that digital imaging nestles side by side with cinema's traditional aesthetics, but is not confined or constrained by them. It has merely taken audiences to new visual domains previously foreclosed by the optics of real, non-virtual cameras” (Prince, “Digital” 98).

Research bears testimony to the fact that indexicality is still compatible with visual effects as realism is the desired goal behind the vast technological research and time being invested in facial motion capture.The importance that this holds for this report is in Lohmeyer's explanation that, "the realism inherent to the motion-captured body is grounded in the indexical motions of the live actor, [and] we are able to experience vicariously the emotions of the virtual body in a manner similar to that of the photographically recorded actor" (6).

Chapter 2: How the *Uncanny Valley* Affects the Realistic Integration of CG Characters in Contemporary Digital Film

2.1 Introduction

Realism is a concept that is "relative to a chosen point of reference" (Christophers 26) and as Prince points out, photographic indexicality is a commonly accepted basis for realism in cinema (Prince, "Digital" 3). In relation to this chapter, the point of reference for realism is primarily in the *movement* of CG characters and secondly, convincing digital image integration. Digital effects of buildings and inanimate environments have become adept at underpinning narratives by being imperceptible but facial expressions has proved to be the most difficult hurdle to get over. What ingredients must a digital character possess to warrant a positive reception and overcome "our extreme delicacy of perception" as Sigmund Freud terms it? (2) Part of the problem for why faces have proved to be so problematic, Prince contends, is the function of their complexity i.e. the multifarious yet subtle way "that the fifty-three facial muscles move to generate a multitude of expressions" ("Digital" 122). A particular issue that has arisen out of this pursuit that this chapter will investigate is whether increases in mimicking hyper realistic movement leads to increases in believability.

The concept of the *uncanny valley* theory, coined by a robotics professor at the Tokyo Institute of Technology, Masahiro Mori, has its roots in robotics but it is generating very serious considerations in the field of entertainment because the psychological theory is easily translatable to contemporary photorealistic human animation. His theory was based on the sense of unease that accompanies the sight of something that is almost, but not quite, human (Lay). Mori asks "why are we equipped with this eerie sensation? Is it essential for human beings [to experience this sensation]?" (99). At the time of Mori's research in the 1970's, little quantifiable, scientific evidence could be produced as to what cues trigger the *uncanny valley* reaction. There has since been a wealth of psychological, and recently, empirical evidence supporting the validity of Mori's theory in both robotics and within the field of computer graphics. Answers to the question of the uncanny in film garners potentially important economic outcomes of figures that reach high into the hundreds of millions of dollars. Failure to cross the uncanny valley have proved to be costly ventures for movies that did not pass the test of perceptual realism. It is alleged that the studios that produced both *The Polar Express* and *Final Fantasy: The Spirit Within* lost more than one hundred million dollars each despite the movies' detailed articulation.

Photorealism poses special challenges that the stylisation of caricature avoids. Disney and Pixar animators realised that caricature provides a way for digital characters to express emotion persuasively and still keep audiences engaged in the narrative. However, as this report is concerned with how movies pursuing realistic CG characters are confronting and overcoming the challenges posed by emulating realistic facial expressions and facial movement, it will not be concentrating on the stylised 3-Dimensional (3D) animated approach of movies in the style of *Shrek* (2001) and *Ratatouille* (2007). Christophers observes that "in recent films, higher levels of realism have been practiced, which is some cases, like special effects for live-action, were successful, while in other cases they failed in terms of believability" (23). Fully animated films, like *The Polar Express* and *Beowulf* attempted realistic facial movement and facial expressions in their characters to "strictly fit the conventions of live-action film" (Christophers 23) but they were not the pioneers of modelling digital characters based on the dynamics of a live performance. That distinction belongs to *Final Fantasy: The Spirits Within.*

In support of the importance of producing perceptually realistic characters in both aesthetic and their movement when integrating CG characters with live action, James Cameron says about the otherworldly Na'vi: "We don't have to necessarily believe that it is one hundred percent photoreal ... but we have to believe in them as emotional creatures" (Pramaggiore and Wallis 177). Prince has touched on the correct assessment with his model of perceptual realism in that it "produces a better integration of the tensions between realist and formalist theory" (Prince, "True" 28). This chapter explores why these movies have fallen victim to the psychologically eerie secret lying in the *uncanny valley*, briefly covering the origin and evolutionary implications of the 'uncanny' in the context of the viewers' relationships with CG characters that mimic humans.

2.2 Characteristics of the Uncanny ("*Unheimlichen*")

Sigmund Freud defined the 'uncanny' as "everything that arouses dread and creeping horror" (2) He also states that the word, translated from the German "*umheimlich*", is not always clearly definable and it tends to coincide with whatever excites and incites dread (214). The 'uncanny' is something hidden that ought to have remained so but which is brought to light. "...and the difficulty is that people vary so very greatly in their sensitivity to this quality of feeling" (Freud 2). Director, Stanley Kubrick's definition is uncomplicated: "Uncanny feelings arise when something familiar (a father, a house, a family member) starts to act a bit strange" (KubrickCorner).

The basis of cinematic perceptions are built on "correspondences between selected features of the cinematic display and a viewer's real world visual and social experience" (Prince, "Psychoanalytic" 80). To determine what psychological principles underlie the existence of that uncanny feeling will require an investigation into the field of psychoanalysis. However, critics of psychology are very sceptical of this as psychoanalytical theories have generally not been proved scientifically. Prince advises against basing theories of spectatorship on psychoanalysis because such theories would remain unsupported. He continues that "psychoanalysis is a discipline without reliable data" but he also concedes that "failure by cognitivists to take the psychoanalysis seriously can only result in limited and imperfect accounts" (Prince, "Psychoanalytic" 72).[6]

The sense of unease associated with the 'uncanny' is essentially an emotional issue affecting perceptions of realistic integration of CG characters into film, but this branch of psychoanalysis is relevant in the digital age. It advances an understanding of the quandary fuelling the emotional and intellectual battle that viewers wage while watching the constitutive elements expected of the photographic facets in film. For this reason, the characteristics of the uncanny that Freud contends within his explanation cannot be

[6] On page 73, Prince realises that when he makes this seemingly disparaging statement, "it may seem harsh or astonishing." He clarifies succinctly in the following paragraphs why he finds psychoanalysis in the vein of Freud, Lacan and others problematic. "Psychoanalysis as a discipline lacks established standards for interpretation that can ensure inter-analyst reliability." As one example of the overly subjective approach, he states that the published accounts of clinical cases present a grossly inadequate, summarized description of the therapeutic encounter between the patient and the therapist.

discounted, despite arguments against its plausibility. Bazin also acknowledged the ambiguous quality that reality possesses in his essay, *Ontology of the Photographic Image*. He felt that each person's perspective on the world, to a significant degree was their own and differed from the perspective of others (Prince, "Movies" 413). Frustrating as it may be due to the subjectivity of Bazin's expression here, the psychoanalysis of his opinion is of importance as it facilitates a systematic approach to defining the source of viewers' anxieties when presented with ambiguous digital images in film.

One theory holds that a sensation of the 'uncanny' exists as a gauge of humanity's highly evolved process of reading facial expressions. One of the more interesting findings at the root of the mysterious, eerie feeling experienced by humans has been from tests on primates at Princeton University. The monkeys exhibited the exact same response as humans - of revulsion and dislike - when presented with images of almost genuine looking digital monkeys. Researchers believe that they can prove that when a human face, the most recognised conduit of emotions is observed by another human, millions of years of evolution are being crossed. This provides one answer to deeper questions about the evolutionary basis of communication and perception (Steckenfinger and Ghazanfar 4).

2.3 Mori's Theory of the *Uncanny Valley* applied to CG Human Characters

Masahiro Mori put forth an essay on how he envisioned people’s reactions to robots that looked and acted almost like humans. He theorised that "a person’s response to a humanlike robot would abruptly shift from empathy to revulsion as it approached, but failed to attain, a lifelike appearance" (Mori 1). Mori suggested that people are increasingly comfortable with robots the more humanlike they look and act, but only up to a point. At that turning point, the robot becomes creepy, as if you are no longer looking at a human-like machine, but instead, at a person that isn't quite ‘right.’ This feeling leaks into the native consciousness of a viewer and alters our perceptions. Frank Pollick, professor in psychology at the University of Glasgow, relays that Mori called this precipitous drop *bukimi no tani* and the translation of *bukimi no tani* into *uncanny valley* has become popularized (In Search of the Uncanny Valley 7).

Mori's hypothesis received almost no attention in 1970 when the article first appeared in a Japanese journal. It practically lay dormant until recently when "the concept of the *uncanny valley* has rapidly attracted interest in robotics and ... researchers have explored its implications for human-robot interaction and computer-graphics animation" (Mori 98). Interest in the *uncanny valley* has now intensified, as technology has evolved to the point where researchers have built robots that look human. In the entertainment industry, the *uncanny valley* has entered the grammatical consciousness of digital artists. Their search for methodologies to integrate CG characters into film is increasing now that readily available digital tools allow animators the capacity to create photo-realistic digital humans on screen that look and act imperceptible from a human in that same environment.

2.3.1 Mori's Graph

As established in chapter one, perceptual realism and cognitive illusion are two types of realism that raise consciousness when assessing the barrier to unconditional belief in the digital characters being portrayed onscreen. Once we've established what the characteristics are for that 'creepy' feeling of unease, prudence dictates that we investigate a way to understand and systematically process those feelings in order for the data to be a useful analytical tool for filmmakers. Mori's graph is an easily understood visual representation of the dips in the relationship between affinity and realism when the eerie phenomenon is triggered first in the subconscious, then the conscious mind.

Mori put forth the proposition that as the robot became more human-like there would first be an increase in its acceptability and then as it approached a nearly human state there would be a dramatic decrease in acceptance.

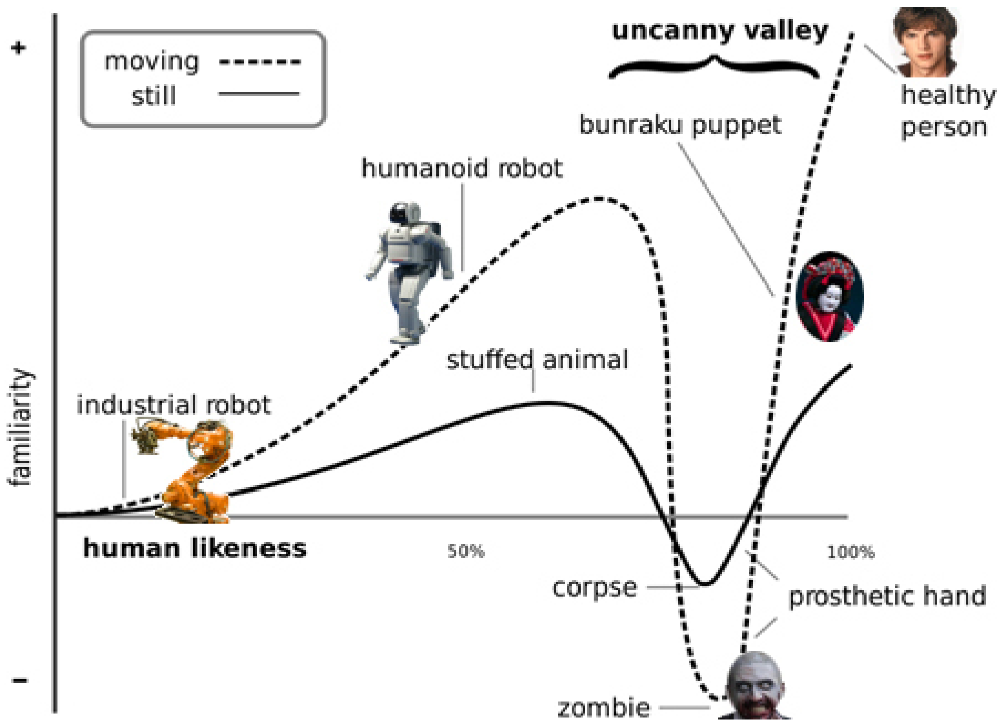


Fig. 2. Mori's Uncanny Valley graph represents a visual depiction of Mori's theory about what a robot needs to look like in order to trigger a positive versus negative response in viewers. The y-axis shows familiarity. A negative familiarity indicates dislike or repulsion, while a positive familiarity is associated with acceptance, comfort and liking. The x-axis is a continuum, from 0 to 100 percent, of how human and life-like the robot appears. The dramatic dip in familiarity near the end of the human likeness continuum indicates the uncanny valley. Important note particularly in the context of this report: The presence of movement steepens the slopes of the uncanny valley.
Source: Guizzo, Erico. *Who's afraid of the Uncanny Valley*? 2 Apr 2010, Web. 27 November 2013. wwww.spectrum.ieee.org.

Christophers' summary of the graph explains:

> human-likeness is mapped out against our familiarity (or appeal) to that entity. Mori illustrates how we share little familiarity with robots that have little or no human-likeness, and therefore we treat them as something mechanical and lifeless. As the human-likeness of these artificial beings increase, so does our familiarity with them and the viewer will find them appealing. However, as the artificial being reaches higher levels of human-likeness, to the degree where the artificial being resembles a realistic human, but does not move like one, there is a

> significant drop in its appeal, making it almost repulsive to the spectator. It is this dip in the graph that is termed the *uncanny valley*. It is only once the artificial being can mimic humans in all areas of representation *accurately* that the spectator will again begin to familiarise themselves with the artificial being and accept it as appealing. (35)

What makes the *uncanny valley* a formidable hypothesis, for believable emotion in facial capture in particular, is that the psychological root of the issues is biologically coded at a deep perceptual level. Rega and Sumida elucidate that species' survival has been dependant on the successful decoding of facial signals. "Unlike other body areas, the face will be subject to intense unconscious scrutiny deeply embedded in our biology" (1). This is not to say that the psychological barrier of the *uncanny valley* cannot be crossed over, more to highlight the magnitude of the task facing digital artists who are pursing realistic facial expressions in digital humans.

Some care is needed in the evaluation of the *uncanny valley* as despite being unsubstantiated by empirical evidence until recently, the theory remained very influential. There is widespread acknowledgement of the *uncanny valley* as a valid phenomenon, but the there appears to be a lack of a solid, single explanation for it. Discourse reveals that from its inception, the *uncanny valley* was never a perfectly defined concept and "audience perception is also influenced by social and cultural factors that govern how a spectator perceives a film" (Cassimus 2). Researchers still debate how it triggers such eerie sensations in the human mind. In an effort to determine what prompts the effect, detection of the response to the *uncanny valley* would be useful as a gauge.

2.3.2 Detecting the *Uncanny Valley* response in Digital Animation

A large part of the public discourse on where Robert Zemeckis went wrong with the likes of *Beowulf* , *Polar Express* and *A Christmas Carol* (2009) lies in his pursuit of realistic facial movement through new technology to the detriment of the perception of his characters, thereby invoking a sense of the 'uncanny'. In pursuit of an endeavour in detecting the *uncanny valley* response triggered in viewers while watching CG human characters, research using the base of Pollick's "four descriptions of relevant psychological processes that could predict the uncanny valley" (5) reveals that the 'trigger' is not a single, solid scientific explanation but a sequence of mental realisations. Cognitive and perceptual realism provide psychoanalytic insight that reveals evidence for the existence of the *uncanny valley*.

Pollick notes that a common explanation is related to the perceptual issue that increased realism is inextricably linked to increased information. However, once errors in the approximations to realism are perceived, these errors simply become *more* evident as more information is provided. He also admits that this explanation does not answer the question of *why* the errors would become more evident. The following trigger comes "once a lack of genuineness is discovered, the animation does not fit solidly into either the living or non-living category" (5). Prince explains the reason disconcertion arises is because this type of CG characters "simultaneously asks viewers to suspend disbelief and to embrace realms of fantasy while the image design erases distinctions between live action and the painted look of animation" (Prince, "Digital" 125). The characters in *Beowulf* , as an example, inhabit this stylistic zone that creates what Prince terms a "modal ambiguity". He attributes this confusion to the filmmaker not addressing or resolving the characters' cinematic design from the beginning (Prince, "Digital" 125). This incapability to distinguish and categorize where these characters like *Beowulf* and Aki Ross, the lead character in *Final Fantasy*, belong in this strange blend of animation and live action leads to confusion and a state of dissonance. This cognitive issue of classification cannot be avoided as it is part of human nature, "however, since category boundaries are not necessarily static, the possibility then arises that increasing exposure will lead to a third category being developed which resolves the dilemma" (Pollick 6) and leaves the viewer with an uncomfortable feeling.

That category, together with a possible final explanation inspired by the observation of Mori and categorised in his graph, is that motion exacerbates an uncanny situation already existing in form. Mori comments that movement is fundamental to human beings. Recent empirical evidence indicates being presented with a still image triggered the uncanny effect, but the *presence of movement* in a CG character - or realistic robot - changes the shape of the *uncanny valley* graph by amplifying the peaks and valleys (Mori 99). Pollick observes that "human actions consist of a wealth of different sensory cues. If these cues are not mutually consistent, then reconciling the differences among cues will lead to a state of unease and uncertainty about what is being observed" (6). A simplistic action for human beings, like a smile, becomes a dynamic sequence of facial deformations to be mimicked to make a CG character smile in a humanlike fashion. It was discovered that the speed of the deformations is crucial. When Mori adjusted the speed in an attempt to make the robot smile more slowly, his observation was that "instead of looking happy, its expression turns creepy. This shows how, because of a variation in *movement*, something that has come to appear close to human could easily tumble down into the uncanny valley" [emphasis added] (Mori 99).

2.4 Avoiding the *Uncanny Valley*

By gaining an understanding of the plethora of considerations that must be taken into account to ensure the successful integration of digital humans characters into live action film, our response to the image is authenticated not only by what appears in front of the camera, but also how that content is presented. One way to avoid the *uncanny valley* is not to fall into it. Mori's graph indicates clearly that his insight is consistent with the realisation of classical animators in the Pixar and Dreamworks Animation tradition, that "caricature provides an effective mode for expressing emotion and for eliciting audience participation in the narrative ... Caricature conveys emotion in concentrated forms" (Prince, "True" 122). Digital characters designed on a type of stylisation that avoids photorealism are very effective at reaching audiences and have never fallen into the uncanny valley. Prince explains that "viewers intuitively understand what the exaggeration conveys and no sense of the uncanny is evoked" (Prince, "True" 122), however, as Hodgkinson points out in his article, *The Seduction of Realism*, “it is thought that stylisation may reduce audience identification, in that not everyone will agree with the style.”

Another issue that instigates debate on whether animations are affected by the *uncanny valley* theory - or whether they avoid it all together- is the topic of anthropomorphism[7]. For the purposes of this report, anthropomorphism is understood to be the specific design of non-human faces to reflect the complex and nuanced characteristics and emotions of human beings. In reference to the fantasy characters in Tim Burton's *Alice in Wonderland* (2010) or *Avatar*, Christophers argues that "the reason behind films such as

---

[7] There are various competing definitions of anthropomorphism in the literature ranging from the attribution of *any* mental state to nonhumans (Kennedy, 1992) to the attribution of exclusively human characteristics (Noske, 1989; Shapiro, 1997). However, as Lehman (1997) points out, these distinctions matter primarily because of the common assumption that anthropomorphism, however it is defined, is necessarily erroneous or mistaken (Serpell 95).

these being successful, is because fully animated *humans* were avoided" (32). In the same paragraph, David Bennett, head of the facial animation department at Weta Digital Studios, agrees that "by changing the character to that of another species, means that we do not judge them in the same way that animated humans are judged" (Christophers 32). Despite its positive contribution to the discourse on the *uncanny valley* and photographic realism, there is an aspect of Christophers' reasoning that is problematic. It could be argued that the issue of whether Gollum in *Lord of the Rings* (2002), the Na'vi in *Avatar* and Caesar in *Rise of the Planet of the Apes* (2011) crossed the *uncanny valley* or not does not extend to these movies as their CG characters avoid the subject by having 'not human' characters. However, as McClean states, these type of humanoid CG characters did not attempt to "pass for human," although there was an intention to "pass for real" in the sense that these performances sought to be convincing representations of alien, fantasy monster, or epic fantasy creatures (60). This does present a discrete tension, but it could be argued that as these digital creations are interacting with live actors in an active and spatially convincing way, they are subject to the same governance of the *uncanny valley* as it applies to humanoid CG characters that are attempting to mimic realistic human movement. They also retain characteristics of the classical notion of indexicality by tricking the audience with real cues from the photographic environment that apply to that character.

2.5 The 'Digital Emily' Project

In 2008, Image Metrics[8] joined forces with the University of Southern California’s Institute for Creative Technologies (ICT) on a project that was conceived and specifically "aimed to cross the *uncanny valley* that divides a synthetic-looking face from a real, animated, expressive person" (Alexander, Chiang and Ma 1). Busch relates that, unable to discuss the work they were doing at the time with *Benjamin Button* due to standard non-disclosure agreements, the motivation behind the Emily Project was to "recreate a live human, at high-definition (HD) quality to basically showcase that our technology is capable of producing animation with that fidelity" (Interview). Animating this digital face would

[8] 'Image Metrics' is the UK-based motion capture company (with a branch in Santa Monica, Los Angeles) that provided the innovative technology that helped bring the digital version of *Emily* to life. They also collaborated on the CG version of Brad Pitt's character in *The Curious Case of Benjamin Button*.

provide the scope and environment to test the innovative, technological approaches in 3D facial capture and animation to produce a faithful one-to-one reproduction of actress Emily O'Brien’s performance. Oleg Alexander, lead technical artist at Image Metrics, reveals that the key technologies included a fast high-resolution digital face-scanning process using USC ICT’s Light Stage capture system and Image Metrics’ video based facial-animation system (Alexander et al. 1).

Mindful of the danger of the *uncanny valley*, the presentation of the 'Digital Emily' project at SIGGRAPH[9] was a resounding success. Emily was almost universally accepted as real and facial motion capture was heralded for the first time. The collaboration had yielded the first photorealistic digital face to speak and perform persuasively in a close-up. After viewing the demonstration video of Emily that astounded the animation community, Peter Plantec, writer for *VFX World* and self-professed toughest critic of facial capture announced:

> This is the first virtual human animated sequence that completely bypasses all my subconscious warnings. I get the feeling of Emily as a person. All the subtlety is there... I officially pronounce that Image Metrics has finally built a bridge across the *uncanny valley* and brought us to the other side.

When asked about the challenges they faced, Busch recounts that "Benjamin Button’s character is very old and has lots of wrinkles, so it's really forgiving. You have a bit more freedom. With Emily, she's flawless. Beautiful, has a great complexion [and] an even skin tone. To recreate her [the lack of texture] was a bit more of a challenge" (Interview). The very detailed scans that were provided by their collaborative partners, ICT, are one of several new fields in facial scanning technology that have contributed to Emily crossing the *uncanny valley*.

Another challenge that could have derailed the project, was inaccurate rigging[10]. Hodgkinson reaffirms Mori's theory that "if an animated character appears realistic, it also has to move realistically in order to avoid the *uncanny valley*" (Christophers 31). Ever mindful of this risk that amplifies the ambiguity in viewers' perceptions if they were even just a fraction off the mark, one of the breakthroughs on this project was the team at Image Metrics' ability to integrate the set of high-resolution face scans provided by ICT and create a

[9] For the computer graphics industry, the major professional organization is SIGGRAPH (Special Interest Group on Computer Graphics of the Association for Computing Machinery). Its enormous annual conventions combine a trade show, a festival of computer animation and a scientific conference where the best new research work is presented.
[10] Rigging is the 3D process of setting up the skeleton inside the digital character that will be used to drive the animation.

fully rigged, animatable face model for Emily. Busch elaborates further in the interview that the rigging process "is the ... key to [crossing] the *uncanny valley*...not the *entire* key but one of the biggest elements."

'Emily' is almost universally accepted as a video of a real person. From this result, it’s clear that a tractable process can achieve the goal of an emotionally compelling, photorealistic digital actor that also moves indistinguishably from a real human. Alexander finishes: "What remains to be done is to streamline the techniques and to further explore the uses of such technologies in applications such as film production, interactive entertainment, simulation and education" (Alexander, Chiang and Ma 29).

Conclusion

In order to overcome the visual element of the uncanny valley, the CG characters must retain characteristics of the classical notion of indexicality by tricking the audience with real cues from the photographic environment that the character sits in. After examining the *uncanny valley*, Hodgkinson concludes in agreement that “the pursuit of realism will not in itself improve the viewer’s relationship to a character nor improve the story. Realism can improve visual immersion but must be driven by and compatible with the visual style adopted for the characters.” With the advances in digital scanning it has become possible to integrate a human visually, as it has been proved with still shots from the movie of *Beowulf*, however this chapter proves that movement is the acerbating hurdle to acceptance of that CG character as real.

Anthropomorphism is one avenue around the uncanny valley, however in this chapter we have discussed a systematic approach to identifying triggers that bring on the uncanny valley effect in viewers. Critically, Digital Emily is a triumph in both technological and eventually the psychological victory over the valley. By recognising Pollick’s sequence of four descriptions that predict these psychological triggers, we can develop a system to overcome each step systematically, thereby repeating the success of Emily and having more photorealistic digital actors that sit and move comfortably in the film surrounding.

Chapter 3: The Development of Facial Motion Capture Techniques

Motion Capture is any "technology that enables the process of translating a live performance into a digital performance" (Menache 2). As a technique, it has its origin in the pursuit of realism by recording complex human actions and faithfully translating the data to the CG model that then mimics the movement and comes alive. It has been established that the origin of our notions of cinematic realism and the dangers presented by the *uncanny valley*, represent the biggest barriers to believable CG human-like characters. Ayse Saygin, a cognitive scientist at the University of California-San Diego, confirms Mori's theory on how movement in a digital character can increase a sense of unease if not animated correctly. She says: "if you look human-like but your motion is jerky or you can't make proper eye contact, those are the things that make them uncanny" (innovationnewsdaily).

The pioneering attempts at photorealism of *Final Fantasy*, *The Polar Express* and *Beowulf* were poorly received in the animation industry. They were perceived as technical failures because they had fallen into the *uncanny valley*, however, this chapter will demonstrate that each of these perceived failures were in fact giant steps forward for their individual contribution to the field of motion capture. Ordinarily, the assumption would be that a fully CG animated movie like *The Polar Express* and a movie with one central digital human inserted into live-action footage, as is the case with *The Curious Case of Benjamin Button*, cannot be compared as they differ substantially in stylistic execution. This chapter is not concerned with the aesthetic choices of these movies, as, in agreement, they differ greatly. The research focuses on the progression of the ground-breaking motion capture techniques of seven movies: *Final Fantasy*, *The Polar Express*, *Beowulf*, *Lord of the Rings: The Two Towers*, *King Kong* (2007), *The Curious Case of Benjamin Button* and *Avatar*. The lens through which a comparative analysis is conducted is fixed on their contribution to the technological development in the field of facial motion capture. In the case of the CG creatures, Gollum and King Kong, it could be argued that they have no place in this discussion on facial motion capture as their facial expressions were the result of the manual method of key-frame animation. Eventually, that did transpire, however both movies have undoubtedly earned their place in the report. In trying to achieve the photorealism, the facial techniques tested - but not implemented in final production due to immaturity of the technology (podcast, The Hobbit) - were great contributions to the overall arc of focus of this section: facial motion capture technology development. The advances that helped to cross the *uncanny valley* will be demonstrated throughout this chapter.

Motion Capture was originally developed as an application in medical science and quickly found its way to the fields of sports, healthcare and the military. As the field is so broad and diverse, the history and treatment of motion capture systems that are not part of the facial motion capture discussion will be superficial and brief. Alberto Menache is a leading figure in motion capture implementation, renowned for his work on pioneering motion capture blockbusters. He has produced some key texts in the research and pedagogy of motion capture. His book, *Understanding Motion Capture for Animation - 2nd Edition* constitutes a summation of his expertise by drawing together contemporary production examples that are referenced in the report. In addition to Menache, another source for contemporary technical discussion is Matt Liverman, author of the *Animator's Motion Capture Guide.* His book also exemplifies different approaches in motion capture technology while offering a critical commentary on styles and creative techniques within the field of motion capture.

The empirical information in this chapter has been compiled from a set of demonstrations and discussions carried out over visits to two motion capture studios in Johannesburg : Flying Circus Animation Studio[11] and AnimMate Animation Studio[12]. A compendium of accepted opinions garnered from visual effects experts at SIGGRAPH 2012 in Los Angeles or from papers presented at other SIGGRAPH conferences in this rapidly evolving field, has added positively to the discourse presented in this chapter.

[11] Flying Circus Animation Studio is a fully operational motion capture facility and animation studio in Kyalami, Johannesburg that opened in 2012. Their RAPTOR 4 system by Motion Analysis® is the same system that WETA Digital Studios based in New Zealand used for the capture of *Avatar* and *The Adventures of Tintin (2012)*.

[12] AnimMate Animation Studio houses the proprietary motion capture system from Viacon® - one of the leading motion capture manufacturers in the world. AnimMate is also the official Viacon agent for Africa.

3.1. Introduction to Motion Capture

Medicine had always been the largest user of the technology, but with recent investment and research, entertainment users are growing faster (Menache 38). As recently as a decade ago, the use of motion capture was limited to well-funded research universities and dedicated technicians in studios. However, Liverman informs that motion capture has since surmounted that small, highly specialised niche in computer animation. With recent advances in technology, motion capture (or 'mocap') systems and the software used to animate the data have become more user-friendly and commercially available yet remain sophisticated (xvii).

In the entertainment field, the technique became valuable in presenting many scenarios for emulating realistic movements that animators using traditional' key-frame methods' - defined in this chapter - would have a more difficult time accommodating. The earliest known mechanism for tracing human movement was 'The Rotoscope', a device patented in 1917 by cartoonist Max Fleischer, "projected live-action frames of film onto a light table in the animator's workspace," providing the animator with a guide for tracing the frame images onto paper (Gleicher 54). The attempt to retain a link to reality through tracing from life and realistic movement has a paradoxical resonance that accords with an understanding of Bazin's theories. Due to its indexical nature, research shows "this device is believed to be one of the origins of motion capture technology" (Menache 4).

In motion capture there has been a recent distinction between capturing facial motion and capturing body motion. James Cameron refers to the distinction as "uncoupling" which was part of the process followed on *Avatar* - having a separate system for capturing the face, called 'performance capture' and another for the body, called 'motion capture' (Duncan 119). The focus remains on facial capture systems and techniques which present a different set of challenges than body motion capture. For a comprehensive history of motion capture and a detailed account of the process of motion capture entailing explanations of the optical, mechanical, and magnetic systems, Menache's book, *Understanding Motion Capture for Animation - 2nd Edition* is an effective source of reference. For a deeper mathematical approach to body motion capture processes, Chapter 7 of Richard Radke's book, *Computer Vision for Visual Effects* is a rich source of information.

3.2. Performance Animation and Facial Motion Capture Processes

3.2.1 Performance Animation

The research and advances in the area of facial motion capture software has pushed the capabilities of these techniques past the boundaries of what discourse reveals was considered impossible a decade ago. In the time period from *The Polar Express* to *Avatar* the technology has evolved at such a rapid rate in the entertainment industry that the techniques had to split into Performance Animation and Motion Capture. Menache clarifies:

> Performance animation is not the same as motion capture, although many people use the two terms interchangeably. Motion capture pertains to the technology used to collect the motion, whereas performance animation refers to the actual performance that is [crafted out of the data] to bring a character to life, regardless of the technology used. (2)

3.2.2 Facial Motion Capture Processes and Technology

If the holy grail of the entertainment industry is to render realistic digital humans, the mission statement of every manufacturer of motion capture technology is to obtain the ultimate result: "real-time tracking of an unlimited number of key points with no space limitations at the highest frequency possible with the smallest margin of error" (Menache 38). Research has revealed that the majority of motion capture techniques for facial capture can be classified as either marker-based or marker-less capture.

Fig 3: A performance in a typical optical, marker-based motion capture stage. In this picture, the cameras on the wall behind the performer have infrared lasers that catch the reflection from the marker or sensor. The term optical, marker based derives from the camera 'seeing' the marker on the suit. Source: Menache, Alberto. *Understanding Motion Capture for Computer Animation (Second Edition).* Burlington, 2011. Print. 30 March 2013.

3.2.2.1 Marker-Based Systems used for Facial Motion Capture

Optical[13] marker-based systems use special visual markers on the performer and a number of special cameras positioned around the 'volume' (capture area) to determine the 3D location of the markers. Traditionally, the markers are passive objects, such as retro-reflective spheres, and the cameras are high-speed, monochrome devices tuned to sense a specific colour of light (Gleicher 53). Optical systems, so named because the camera 'sees' the marker on the suit and triangulates its position to the 3D programme that collates all the information and translates the data to a corresponding 3D point in computer space. State of the art optical systems often use many cameras - sometimes as many as 24 - in an effort to minimize the risk of markers not "being seen" - or occluded - by enough cameras. At Flying Circus Animation Studios, they utilise a total of 40 high resolution cameras in the 'volume' (the capture stage).

In a marker-based system for human face capture, several smaller markers were placed on key facial features such as on the eyebrows, around the nose and mouth, however,

[13] Because optical capture systems must address lost markers due to occlusion and correspondence as post-processes, magnetic systems have traditionally been the preference for performance animation. Improved software for optical processing is changing this. Similarly, both technologies are evolving rapidly, changing many of the historical trade-offs in their relative price-performance.

as Prince explains, optical motion capture systems cannot possibly grab all the data involving the subtle and intricate surface deformations over the distance between the markers and the face ("Digital" 122). The marker-based system creates more accurate and efficient data, but very often the recorded motions result in mechanical movement - because the subtle and intricate surface deformations that Prince mentions - the information *between* the markers cannot be captured by marker-based techniques. If the goal is photorealistic integration, that information is vital. In some attempts, the number of markers were increased to overcome this issue but to little effect. The reason for this, Cameron explains, is that "the face has almost as many muscles as the whole human body that drives every expression in a neuromuscular symphony" (Wolf, forbes). Joe Letteri, senior visual effects supervisor at Weta Digital, reveals that actor, Andy Serkis had up to 150 sensors patterned across his face for his performance in *King Kong* (2004) (podcast, Avatar-Weta), - and it was nowhere near enough to get the level of detail that was required to drive the CG model. After copious failed efforts documented in research papers and in institutions across the globe, motion capture industry professionals began the hunt for a better solution to obtaining higher resolution data for realistic movement in the face.

3.2.2.2 Marker-less Systems for Facial Motion Capture

Liverman describes marker-less systems as the natural estimation of human movement without any special clothing or markers (xxi). It would appear to be the answer to an alternative capture system however, "marker-less motion capture systems are rarely used for production quality animation for body capture since they are generally less accurate than marker-based techniques" (Menache 52). In his chapter on "Types of Motion Capture" (16-45), Menache mentions that in theory, this should lead to much better motions, but is rather a more difficult and time-consuming system to track thousands of pixels with arbitrary surface texture than with a few hundred retro-reflective balls. In spite of this, marker-less systems have had the highest impact on facial motion capture. Gleitscher elaborates on why vision-based technology is more suited for facial capture:

> A vision-based technology, instead of marker-based, is one that can analyse “standard” video streams, performing some form of image analysis to determine

> what the performer is doing. Efforts to perform facial motion capture using vision tracking technology are much more established because of the limited range of motion and slower movement rate of facial features. Video has been a much more workable methodology for facial capture. (54)

These technologies usually provide engineering solutions to standard vision issues - such as tracking and identification by video footage for example or using special cameras and lighting that make some type of marker drawn on skin obvious. (Gleicher 54) The newer optical systems are "based on what is now termed 'optical flow analysis' - such as the system that was used for the facial performance capture for *Avatar"* (Menache 38). In *Avatar*, they developed a head-mounted camera [headcam] that could constantly track the actors face to obtain the video footage (podcast, Avatar-Weta). These tracking systems are always a post-process as time must be allocated to analyse the data to be transferred to the CG model. Dr Michael Black, considered the world's leading researcher in the art of optical flow, defines optical flow as the apparent motion of image pixels from one frame to the next. Optical flow analysis incorporates tracking *every* pixel as it compares frames in a moving image, then tries to map those pixels from one frame to the next (fxguide, Optical). Throughout the research on facial motion capture, it was evident that optical flow analysis systems is the subject of intense computer vision research projects, both in several high-profile new research and development projects in the motion capture industry and in the interests of the academic community for other applications in fields other than entertainment.

3.2.2.3 Performance Driven Facial Animation (PDFA)

Many systems have not been mentioned in this report as they have little bearing on the topic, however one of the several exciting techniques that have emerged because of invested research into new facial capture systems is digital photogrammetry[14]. Menache defines digital photogrammetry as:

[14] Photogrammetry is a technique that first appeared in 1849 when French inventor, Aime Laussedat used photographs to make a map (motioncapturesociety). For documented photographs from a personal demonstration of the image-based facial capture system of leading motion capture manufacturer, PhaseSpace Inc- based in San Francisco, California - see Appendix 4

> a method of *building geometry* based on photographic images. Capturing facial motion using [digital] photogrammetry is a relatively new application, as it requires significant computing power. It is now called 'videogrammetry' because it utilises a set of video images instead of photographs. (34)

A model is constructed dynamically for *each frame* of the performance. These models are then used as the basis for the animated face. Data can be extracted to drive another character CG character - some of the geometry is utilised as very high resolution models for blendshapes[15]. This technology exists in a number of forms, but for facial motion capture we are concerned with two companies. US-based Mova Contour and UK-based Image Metrics, both considered leading proprietors of new techniques described in this chapter due to their work on *The Curious Case of Benjamin Button* specialize in vision-based facial performance capture are. Motionscan[16] is another company that also utilised a version of videogrammetry face scanning technology for the production of the game, *LA Noire.*

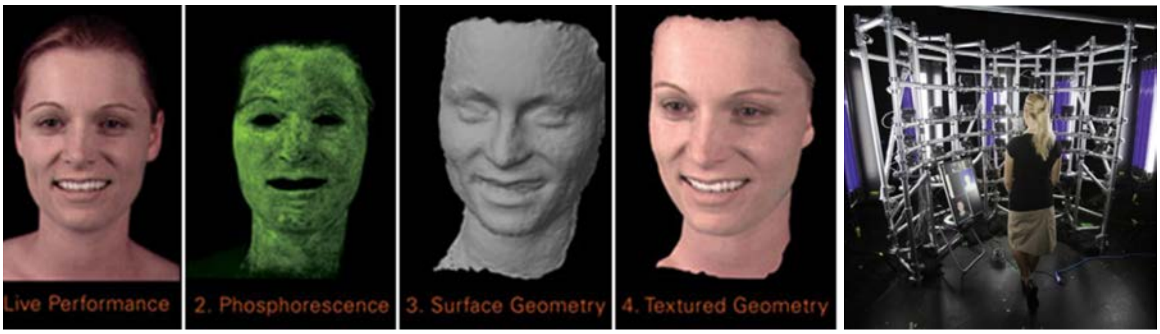


Fig. 4. The stages of the photogrammetry process that Mova Contour perfected for *The Curious Case of Benjamin Button*. **Far right**: The Mova 'capture stage' with black light that returns a high-resolution scan of the face for the data stream.
Source: Mova.com. Web. 27 March 2013.

Mova Contour requires subjects to paint their faces with phosphorescent 'make-up' that shows up under black light (Figure 4). This paint acts as millions of sensors from which to read the data of all the detailed and subtle dynamics of what the skin does when it glides across the facial bones and folds into itself. However, as mentioned earlier, the computational

---

[15] Blendshapes are a 'morph target' animation technique that allows you to change the shape of one object into the shape of another targeted object.

[16] Motionscan's videogrammetry records the actor's performance on video and creates the 3D meshes frame-by frame include eye information, however they are un-editable. Should the need arise to change an aspect of the performance at a later stage, it is not possible.

power to read this information is immense. In researching digital photogrammetry via demonstrations on the software, it was discovered that due to concentrated computer research and resources donated to this area of interest, data construction from close-range photogrammetric mapping systems - that was previously very time consuming - now has considerably reduced in processing time, making this area an attractive option.

Image Metrics' approach differs from Mova Contour. Busch explained in a demonstration during the interview that Faceware Technologies, the facial software spin off of Image Metrics, use a high-resolution video analysis system to get their data from analysing video footage of the performer's face. They run their proprietary processing software to obtain the 'tracks' (data stream information) to be augmented by the motion editor[17] (animator) once all the tracks are collated in a 3D programme. This provides highly accurate commencement points of the movements of the actor as it is translated to the CG character for the motion editor to begin augmenting. One of the focuses that Busch mentions while working on The Digital Emily Project was the importance of picking up the line of sight of the actor (Interview). Image Metrics' approach has the capability to capture more of the subtle eye movement that gets transferred to the CG model. The video footage of the performance will be used again by the motion editor or animator, but for reference purposes. Mova Contour's photogrammetry approach was limited in this aspect as - at the time of Emily - intricate eye movement could not be captured as easily (Figure 4).

[17] The position of an animator - usually specialising in facial animation - working with motion capture data is referred to as a *motion editor*, however this depends on the individual production facility or studio set-up. For this report, the term *motion editor* will be used.

3.3. Technical Developments in Facial Motion Capture

3.3.1 Introduction of Facial Action Coding System (FACS) [18]:

Micro-expressions are the key to crossing the *uncanny valley*. In 1978, Dr Paul Eckman, a renowned physiologist in the field of non-verbal communication, and W.V Friesen developed the Facial Action Coding System (FACS) to describe and classify nuanced differentiation between human facial expressions. FACS divides the face into Action Units (AU) to quantify the 72 different facial expressions Dr Eckman and Friesen had identified as the amount of expressions a face can display. These included inner brow raiser or a nose wrinkle to name but two 'poses' (Figure 5).

Although FACS was originally designed to analyse natural facial expressions, "it turned out

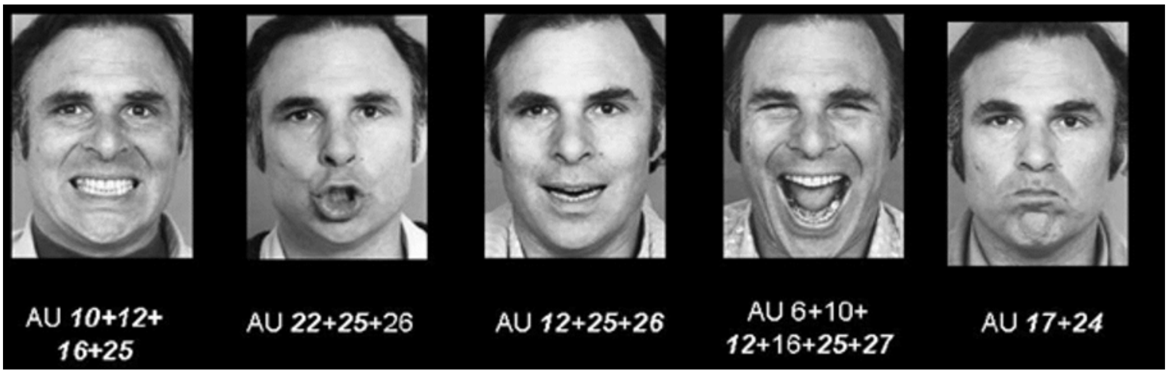


Fig. 5: Dr Paul Eckman demonstrating 5 of the 72 facial expressions that Dr Eckman and W Friesen's research concluded the human face was capable of making. The combination of letters and numbers listed below the line of photographs demonstrate how they divided each feature of the face into Action Units (AU). With these AUs, they were then able to assign each expression the face was capable of producing with quantifiable combinations of features in a logical, algorithmic approach. Source: Eckman, Paul. *Eckman.com.* Web. 28 November 2014.

[18] Informative featurette on the animation process, how information from the data is encapsulated in this 6 minute featurette on the "Making of Monster House" on the web. (http://youtube.com/watch?v=MY936inB_mM&feature=youtube_gdata_player). See Appendix 1 for an explanation by Mark Sagar, developer of the FACS solver on the initial history of his progression on implementing FACS on Monster House.

to be usable as a standard for production purposes too, especially for generating facial expressions displayed in virtual 3D characters" (Bee, Andre and Vogt 137).

FACS has emerged in contemporary pipelines as an ideal, efficient 'alphabet' of facial actions for the creation of realistic digital expressions. Being able to reduce these elusive expressions into a quantifiable entity has been a breakthrough. Remington Scott, FACS Supervisor on *Final Fantasy*, *Beowulf* and the *Lord of the Rings* trilogy, explains that FACS is the most versatile method for measuring and describing how the contraction of each facial muscle changes the appearance of the face (remingtonscott). While conducting this research, several studios visited were adopting and implementing this system to great effect.

3.3.2 Facial Action Coding System (FACS) - Facial Expression Solver

Mark Sagar, the developer of the facial expression solving system, used the FACS system of the facial expressions pioneered by Dr Paul Eckman, to engineer a facial expression solver and initially mapped the performance to Gollum for a test in *The Two Towers* (2002). He clarifies that this system was born out of the desire to work seamlessly with animation as at the time, standard motion capture techniques were unsuitable for the level of photorealism they were striving for (podcast, Sagar; Appendix 1). An improved FACS solver was implemented at Weta Digital for *King Kong* in 2005 by Sagar, and was eventually able to capture extremely subtle to highly dramatic expressions and faithfully retarget and translate Andy Serkis' performance to CG King Kong's face (Sagar 1). However, the solver needed time to mature. As a result, the facial expressions on King Kong were also key-frame animation.

In the same podcast, Sagar breaks down how the solver calculates the nerve signals that are sent to the muscles. It is exact information content facilitated by the action unit combinations that are sent to the individual muscles to read an expression. He admits that the first attempt to use the FACS solver for performance capture, then map those expressions onto a 3D character's face, was on the stylised animated feature, *Monster House* in 2004 while employed at Sony Imageworks. The movie was not released until July 2006, after *King*

Kong (2005). Parag Havaldar[19], computer supervisor at Sony Imageworks, explains that the system analyses the motion capture data for up to 96 facial expressions in the calibrating phase before they begin capturing, rather than analysing photographs of skin motion *after* the performance of the subject- as is the case with optical flow analysers. Figure 6 clearly depicts the actor from *Monster House* (2006) preparing in the calibrating phase, mimicking Dr Eckman's poses and the corresponding CG character. See Appendix 1 of this report for an additional explanation by Mark Sagar on the implementation of the FACS Solver.

(Below) Lip Stretcher Pose

(Below) Mouth Stretch Pose

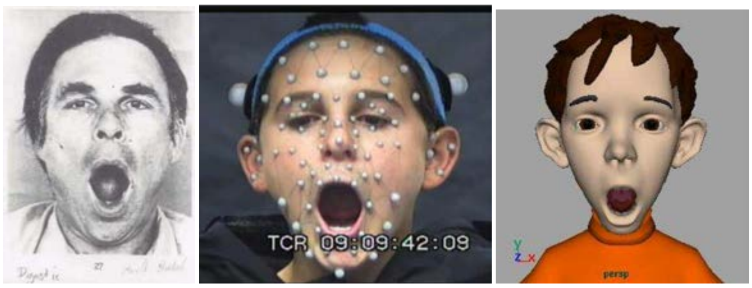


Fig. 6. Calibrating phase, setting poses for the FACS Solver for *Monster House*. The actor is copying the 72 different posed developed by Dr Paul Eckman for the FACS-based digital facial system. The reference AU image is shown on the left, the actor mimicking Dr Eckamn's pose and performance of the pose is shown in the middle and the same pose interpreted on the CG character is shown on the right with the CG character from *Monster House*. Source: Havaldar, Parag. *Performance Driven Facial Animation.* SIGGRAPH, 2006.

This represents the fundamental information of the performance which is then driven by an Imageworks proprietary animation key-frame tool for the motion editors to control the performance called Character Facial System. Once all the required facial poses had been recorded from the actor, only *then* was the performance captured and translation of this expression data onto an arbitrary CG character took place. From observation of a standard

[19] Parag Havaldar explains clearly what the FACS system is doing and in great detail beyond the scope of this report. I direct any interested party to this document. SIGGRAPH 2006 - Parag Havaldar - Performance Driven Facial Animation Sony Pictures Imageworks

set-up, the data would translate *directly* to the CG character with no interpretation of expressions in between.

Havaldar explains in his paper presented at SIGGRAPH that these digital tools used the basic, underlying FACS theory of muscle groups to group markers together and organize them into groups of muscles. In Figure 7, a clear designation of the muscle groups and breakdown of the corresponding underlying facial muscles responsible for the myriad of expressions viewers see on a daily basis.

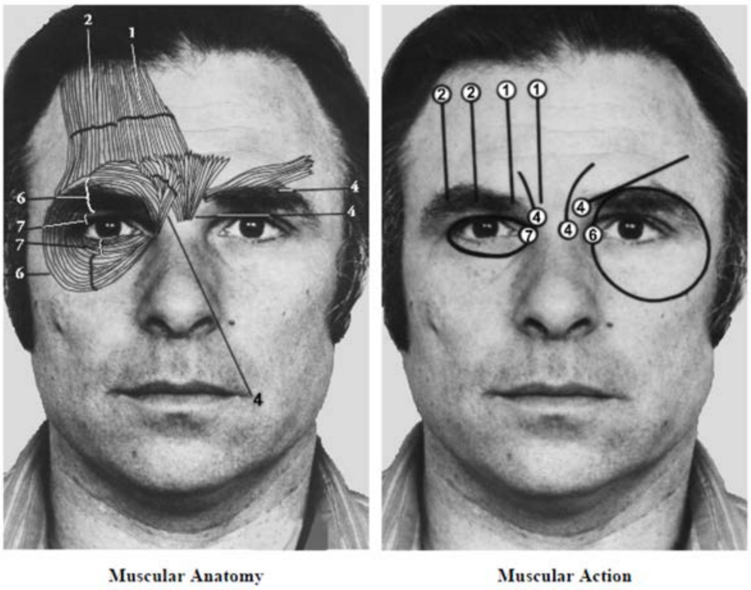


Fig. 7. FACS muscle diagram, illustrating the breakdown of the underlying facial muscles responsible for the FACS Action Units that is a driver in the FACS Solver.
Source: Watt, Kathleen. 04 February 2012. Web. 18 October 2013.
www.*wattwork.blogspot.com.*

This system was tested on Gollum, and then refined for *King Kong* (2006). Instead of directly mapping marker to marker, which creates the look of a human in a gorilla mask, motions are mapped onto the perceptual equivalent for a gorilla, resulting in more ape-like behaviour. In Chapter 4, this system is elaborated on to greater effect with regards to the enhancements that Mark Sagar and the team implemented at Weta Digital for *Avatar* based directly on the FACS muscle solver. One of the great breakthroughs that will be discussed is how the team combined optical flow analysis video footage with their proprietary version of a FACS solver "to implement a new learning-based facial capture system" (podcast, Avatar-Weta).

In support, Speech Graphics co-founder, Michael Berger[20] observes that "people are very sensitive to this when you get it wrong because we are all focused on faces" (20). In testament to the heights achieved in the animation industry with these recent advances, he admits in his professional capacity, that capturing the fluidity and accuracy in natural speech is one of the most difficult things to do convincingly in facial animation. Veridically sound elucidation makes the CG characters that have crossed the *uncanny valley* due to the contributions of the innovative digital artists more impressive.

3.4. Movies that fell into the *uncanny valley* yet contributed to advancing facial motion capture technology

3.4.1 *Final Fantasy* (2001)

Final Fantasy was the first fully CG animated feature film that attempted at photorealism. The combination of motion capture technology with key-frame animation was also an original approach. Scott shares that "initially *Final Fantasy* was to be a heavily key-frame animated film. Teams that were involved with *Titanic* (1997) and *Godzilla* (1998) cited those films as earnest attempts to use motion capture, however, [in those movies, their] data was unusable and the characters had to be key-framed instead" (remingtonscott).

In response, Scott teamed with researchers, development artists and technicians to prove that motion capture could be used for the performance of human characters. The team developed an algorithm to test this first. The result convincingly proved there was a significant difference in fidelity of character animation for human movement using motion capture. Scott admits, however, that a key-frame animation approach falls short in that respect and the facial data in the movie was insufficient for the required fidelity of quality for their endeavour to replicate photorealism (remingtonscott). In his article, *Crossing the Great Uncanny Valley*, Plantec critically points out that the "rich world of subliminal expression was completely missing. All nuances were missing. Because they looked so real, we expected

[20] Michael Berger is part of the award-winning team from the University of Edinburgh's School of Informatics and the Centre for Speech Technology Research,

that full spectrum of face and body language...” (awn.com, Crossing). That level of complexity, "... of fifty three face muscles moving in complex subtle ways to generate a multitude of expressions" could not be captured by hand, Prince adds by way of an explanation ("Digital"125). Scott admits that the project required extraordinary computational powers that were far beyond the resources of that period.

3.4.2 *The Polar Express* (2004)

"It wasn’t until 2004, when Robert Zemeckis released *The Polar Express,* that motion capture started to find its place in the film industry..." (Menache 49). Admittedly, it has been established in the court of public opinion that *Final Fantasy, The Polar Express* and *Beowulf* fell into the *uncanny valley* as articulated by Plantec:

> Zemeckis chose here, as he partially did in *Beowulf,* to have his animated characters look much like the real actors playing them... Tom Hanks, for example, as the conductor and six other characters. This exacerbated the situation. Not only do we expect the characters to convince us they're human, but also that they are Tom Hanks. Zemeckis wisely chose the non-photoreal shading used throughout *Polar Express,* which pulled everything back a bit from reality. But the look and voice characteristics still triggered disbelief. It was a wildly uncanny experience.

However, as Jerome Chen, visual effects supervisor on the *Polar Express* and *Beowulf* points out, “[*The*] *Polar Express* was, in terms of motion capture, probably the most ground-breaking. Not necessarily on the imaging side, but because of the technology that had to be put together. Nobody had actually captured body and face together at that point” (49). As Plantec alluded to earlier, for the visual execution of the characters, Zemeckis had chosen to remain faithful to the illustrations in Chris van Allsburg’s book (movies.about) but this report is concerned with how *The Polar Express* sought to re-organise the relationship between the actual body and the digital image "as one wherein the body is neither erased nor held prisoner by the animated, but rather controls it from a separate space entirely" (Aldred). *The Polar Express* was also the first project where Performance-Driven Facial Animation (PDFA)[21] was used exclusively for main characters - with very little key-frame animation - and was the reason cited for its lack of success. The motion capture supervisors learned that more information was needed between the markers on the face if the filmmakers were to achieve photo-real rendering. They also had to solve the issue of dealing with the interference deriving from several characters interacting in the one capture space as explained by Mark Sagar, motion capture supervisor on the project in Appendix 1.

3.4.3 *Beowulf* (2007)

The pioneering attempt of the *The Polar Express* generated excitement in the motion capture industry. *Beowulf* 's motion capture pipeline was almost identical to the one developed for *The Polar Express*, however in an attempt to achieve more realistic facial expressions, extra attention was lavished on facial animation. *Beowulf* has the distinction of being in the Guinness Book of Records for having the most amount of motion capture cameras in a capture volume - 234 cameras (MotionCaptureSociety). "The volume was bigger so all the capture was done in a single stage and the crew were more experienced after having worked on *The Polar Express* and *Monster House* (2006)" (Menache 72). The most strikingly observable detail when viewing the film is that the motion editors paid attention to the eyes in *Beowulf* (2007), as opposed to *Polar Express*. *Beowulf* authored a patent for Sony Imageworks: an algorithm that recorded micro-saccadic eyeball movement to a tiny computer with several electrodes placed on the actor's face (see picture 'C' in Fig. 8). The information supplied by the electrodes around the performer's eye drove the movement of the CG character's eyeball (remingtonscott). Menache explains that saccadic motions are fast simultaneous motions of both eyes moving in the same direction (51).

David Bennett, motion capture supervisor at Weta Digital in New Zealand, also worked on motion capture for *The Polar Express*. He tells Christophers that "there were

[21] Performance Driven Facial Animation is a type of animation when one CG object drives another 3D object. Blend-shapes would be a good example.

issues that needed to be worked out in the [motion capture] technology. Most movies these days are getting close to cracking the issues associated with the *uncanny valley*" (40). In Appendix 1, Mark Sagar provides an enlightening explanation of how the interference that Bennet is referring to affected the pipeline they had set up. Their research revealed that when motion capture moved away from a marker-based system to a marker-less system and the processing power of the technology increased, this was one of the areas that helped the industry traverse the *uncanny valley*.

3.5. Graph depicting Technological Advances in Facial Motion Capture Techniques based on graph of Mori's *Uncanny Valley* Theory

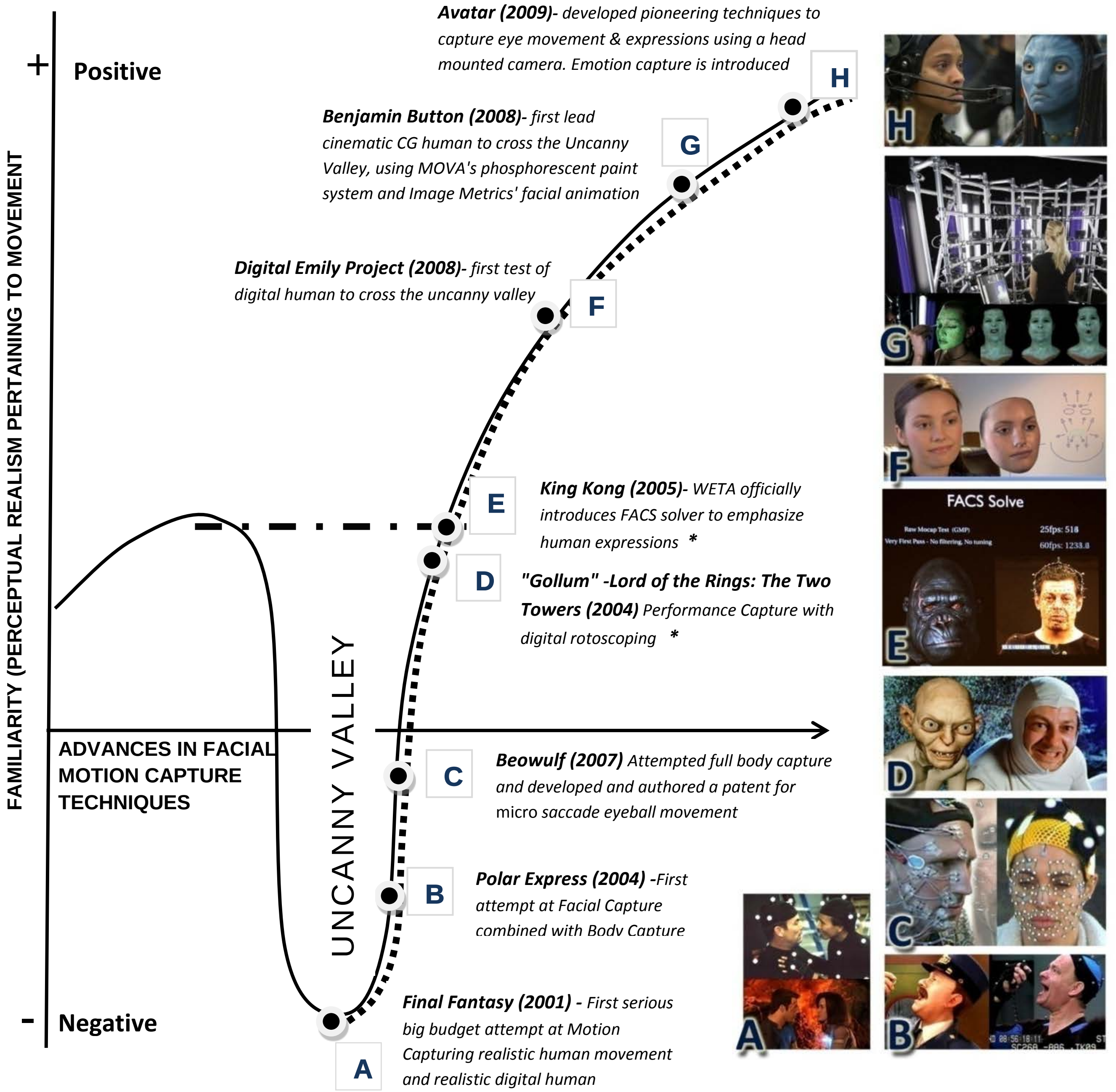
In Bazin's essay, the *Evolution of Cinema as a Language*, each person's perception of realism is based on their personal philosophy or understanding according to their experiences. The success of the viewer's investment in the virtual body highly depends on how perceptually realistic the body of each of these CG characters in the graph in Figure 9 appears and moves, or rather, how it might successfully link to the viewer's personal understanding of how a real body moves. The construction of the graph in figure 9 was achieved by using a combination of assessment of the technological advances made in motion capture, viewer and industry experts' opinion as well as personal perception after studying each movie depicted.

This graph (Figure 8) uses Mori's *uncanny valley* theory as a basis. This is a visual representation of the progress of facial motion capture technology and innovations represents the fact that the technology did not happen in isolation - and what eventually led to Digital Emily, CG Benjamin Button and the Na'vi of Avatar rising out of the uncanny valley.

Using the x axis measures to the Familiarity (Level of Perceptual Realism pertaining to movement) - as established in Chapter 1 - and the y-axis plots the Advances in Facial Motion Capture Techniques of the seven movies that have impacted facial motion capture and been expanded upon throughout this report. *Final Fantasy* is at the bottom of the *uncanny valley* indicating the negative perception and lack of realistic facial expression in the CG characters. *The Polar Express* and *Beowulf* fall below the line, into negative levels of Perceptual Realism as those CG characters are associated with dislike or repulsion. As progress and research documented in this chapter produced better technology, the curve continues on an upward trajectory, indicating increased positive levels of Perceptual Realism and Familiarity leading to viewers' acceptance of the CG characters with each technological improvement.

Prince cites that Gollum in *Lord of the Rings: The Two Towers* (2004), the titular case of *The Curious Case of Benjamin Button* and the Na'vi in *Avatar* as "three examples of emotionally expressive digital performance elicited in motion capture and that have achieved credible perceptual realism" ("Digital" 127). In his opinion, they have crossed the *uncanny valley*. Conversely, the graph constructed in Figure 8 demonstrates a disagreement with Prince about the creature Gollum. Prince cites the scene in *The Two Towers* (2004) where Frodo tells Gollum to release Sam or have his throat cut, "the saccadic eye movements and the way the eye jumps are augmented by eye movements... the psychological response that provides an index of emotion and thought" ("Digital "128). Gollum has been placed just below the precipice of the *uncanny valley* in the graph as he is not *entirely* convincing in *The Two Towers* (2004) as a CG creature that sits comfortably in the surrounding footage. CG King Kong is no less complex than Gollum, but his micro expressions are of a higher degree of believability, therefore the perception by viewers of CG King Kong as real creature is higher. However, King Kong is still in the *uncanny valley* as the appeal of King Kong in relation to the environment is still slightly less believable in several shots on his home island. It does not imbue the degree of conviction of Digital Emily, CG Benjamin Button and the Na'vi in viewers' minds

The dotted line (.........) illustrates the first documented attempt at using facial motion capture techniques for CG characters by director Robert Zemeckis - beginning with *Final Fantasy* - and subsequent efforts that improved in familiarity and acceptance but did not quite make it out of the *uncanny valley* until the *Digital Emily* project in 2008, where the viewers' uncomfortable experience becomes a pleasurable engagement with the CG characters. A description of each technological improvement associated with that movie is included in a text-box next to their corresponding place on the graph along with a visual aid of each technique on the far right in the form of a picture of each motion capture technique.



Key

indicates the beginning and end of the Uncanny Valley

illustrates the progress of Facial Motion Capture Technology

*** It must be noted that the facial expressions of the CG creature Gollum in *The Two Towers* were eventually key-framed, as was *King Kong*, but the intention to emulate CG characters grounded in realism have earned them their place on the trajectory of motion capture advancements in the graph. Facial Motion Capture technology was tested on these CG characters, but not used in final production. As demonstrated in the report to this point, due to their contribution to Facial Motion Capture *technology* employed at the time, they are represented on the graph.**

Fig. 8.: Personal graph based on Mori's graph of the *uncanny valley*. This graph depicts the evolution of facial motion capture technology with the progress out of the *uncanny valley* of CG characters that are perceived to move in a realistic manner. It illustrates *how the seven movies that have had the most impact with their innovative facial motion capture solutions have aided the progress of Facial Motion Capture technology in the 2001 - 2009 period.* As each new development was implemented from *Final Fantasy* (2001) to *Avatar* (2009), we observe the direct positive impact with the increase in appeal is observed until, first the Digital Emily Project, then Benjamin Button and the Na'vi from Avatar overcome the *Uncanny Valley*.

Note: *A Christmas Carol* (2009) was the first movie released in cinemas that featured a head-mounted camera, however it was *Avatar* that introduced the technology. *Avatar* was still in production at the time (Appendix 1). With regard to *Monster House* (2006), although the FACS system initially called Character Facial Coding was used, this is not included in the graph as the stylized animation avoids the *uncanny valley*. There is no intention for the characters and all their movements to "pass for real" as there is with *Final Fantasy, Polar Express* and *Beowulf*.

3.7. Conclusion

A motion capture performance is the culmination of hard work of many motion editors and technicians, celebrating successful collaborations in direct exchanges between visual effects specialists and university research departments on finding a means to capture the 'truth' of an actor’s performance. When one traces the individual successes of finally crossing the *uncanny valley*, it is clear that a multifaceted network of influences from basic academic research to industry professionals, all dedicated to the research and development that ultimately imbued the CG characters featured in this section with the same mannerisms as the performer in the most literal embodiment. A good performance in this medium would not be witnessed if all aspects of the visual effects pipeline were not in place to acknowledge the union of acting and technology.

The direction that facial motion capture will take is still unclear as the research has shown several avenues of approaches are being constantly evaluated for a better solution. There have been some breath-taking developments - in the past five years especially - that make it of significance in this report to address how motion capture is not just fitting into the classic structure of production pipelines, it is de-fining and re-fining those pipelines. The rhetoric of progress on facial motion capture presents the transformation of technological progress as a revolution that has become part of pre-production and production, rather than only in the area it was previously relegated to: post-production. The technique of motion capture is, in part, responsible for bridging the dichotomy presented earlier by George Méliès and the Lumiére brothers.

Chapter 4: Case Studies : *The Curious Case of Benjamin Button* (2008) and *Avatar* (2009)

Introduction

The Curious Case of Benjamin Button and *Avatar* are presented as the two case studies that embody the findings of the research. The principle case study, *The Curious Case of Benjamin Button* celebrated the advances in facial capture technology by being critically hailed as the first digital CG character in a feature film to cross the *uncanny valley*. In the quest for believable integration of CG characters into live-action film, *The Curious Case of Benjamin Button* has overcome the tension presented by realist theory and the fundamental indexicality of Bazin's realism.

The second case study, *Avatar* chronicles an ambitious undertaking by director, James Cameron, portraying an entire alien race and mythical planet. Technological advancements in facial motion capture used on the lead character in *Curious Case of Benjamin Button* were employed and improved upon in two ways to spectacular effect in the sweeping epic. The innovations in *Avatar* have merged the dichotomy presented by traditional film theories and ascended the ontological nature of the production with their technological breakthroughs.

4.1. Principal Case Study – Case Study 1: *The Curious Case of Benjamin Button*

The Curious Case of Benjamin Button, directed by David Fincher, offers a compassionate allegory about the reversed passage of time. The capricious nature of the story was considered too difficult to execute for over forty years "and bringing it to life on the screen was perceived as too ambitious, too fantastical to accomplish" (wildaboutmovies).

This section will be an analytical exploration of the herculean feat by that technical computer artists travelled over four years at visual effects powerhouse Digital Domain in innovative collaborations with Mova Contour, Image Metrics' Facial Animation software and the Institute of Creative Technologies (ICT) to convincingly age the lead character. The narrative succession portrays the protagonist as a man who is born old, then ages backwards. 'Digital make-up' is utilised in a previously unmatched height of realism to render the reverse aging process that the film's protagonist, played by Brad Pitt, goes through during his unusual lifespan.

Through research it is learned that Brad Pitt only agreed to fill the lead role if he could play the character throughout the natural life of Benjamin Button[22]. The dissonance Pitt felt viewers would experience having another actor play Benjamin at various ages, as was the case with Kate Winslet and Gloria Stuart playing young Rose and aged Rose in James Cameron's *Titanic* (1997) served to intensify the challenge to digital artists on the project. It was imperative to the actor that Benjamin's transformational arc from age eighty all the way back to twenty be anchored by a single actor's performance, before another actor stepped into the role for adolescence.

To grasp the understated but meticulous attention to detail that the digital artists knew would be required on this broad a canvas, Eric Barba - visual effects supervisor on *The Curious Case of Benjamin Button* - reveals the decision to step away from the traditional approach of prosthetics and make-up was taken very early in pre-production. Barba clarifies why this approach would not work; old-age makeup is added in layers whereas the reductive aging process is the exact opposite. "Everything is receding. There is no way to do that one hundred percent convincingly by adding prosthetics" (Duncan, The Unusual Birth of Benjamin Button 72).

Ed Ulbrich's feelings of extreme anxiety upon the realisation that "there was a giant chasm of the state of the art of the technology in 2004 and where it needed to be" (Ulbrich), engendered the decision to focus on motion capture. Through discourse and research, it is well known that the first 52 minutes -with 325 digital head-replacement shots - Benjamin Button is computer generated from the neck up. Precise motion tracking was required to place the head onto the three body doubles that were used to convincingly portray Benjamin

[22] David Fincher reveals in an interview documented in "The Curious Case of Benjamin Button - Behind the Scenes" that for Pitt, the only way to play the character was all the way through, at every age, which posed one of the film's most daunting challenges. Fincher explains that "Brad was only interested in playing the part if he could play the character through the totality of his life" (wildaboutmovies).

as he would be at that age. In Figure 10, the body doubles had to be the correct body mass for CG Benjamin's head to provide the solidity and visual personality from which the digital artists and Pitt could take what Prince terms "perceptually convincing visual cues" (True Lies 39) as they moved through the narrative space. In testament to the correct decision made by Digital Domain, Mike Seymour, visual effects artist and co-founder of fxphd[23], observed that when "he found himself wondering at which point in the film did the head shift from a digital effect to live-action footage..." (podcast, Benjamin). In this section, while analysing how *The Curious Case of Benjamin Button* avoided falling into the *uncanny valley*, David Fincher's pursuit of a very novel approach to virtual performing will also be discussed. He said "acting and performing are two different things. Acting is what you do; performance is the thing that you make from acting" (Prince, "Digital" 102).

Fig. 9. Digital version of Brad Pitt as Benjamin Button. Source: Ulbrich, Ed. “TedTalk - Ed Ulbrich: The Curious Case of Benjamin Button”. *Youtube*. Youtube, February 2009.Web. October 2013.

4.1.1 The Facial Motion Capture Process: Identifying the Challenges

The digital effects team working on the project knew the goal was not to create the lead character’s performance with animation, but rather to faithfully reproduce the complex emotions that resonated at the core of Pitt's performance onto the CG head that would be

[23] 'Fxguide' and 'fxphd' arerespected, comprehensive high-end online visual effects production and post-production training sites. Founded in 2005, one of the co-founders is Australian visual effects artist, Mike Seymour.

created. Once the decision was taken to turn to performance driven facial animation, research began on the challenging performance requirements. Barba knew there was no mono-causal solution in order to realize their visual strategy. A range of various technological approaches were tested with unsatisfactory results and the respective scopes available at the time were deemed too limited. The eventual results of their systematic research concluded that three areas needed to be isolated in order to cross the *uncanny valley*. They were: moving away from the traditional marker-based facial animation and hand animated key-frame techniques that had proved inadequate; a new form of tracking to situate Benjamin comfortably into the digital cinematography had to be introduced to the pipeline; and a comprehensive face scanning technique for rendering lighting needed to be developed. Steve Preeg, character supervisor in charge of facial movement on *The Curious Case of Benjamin Button*, recalls that to obtain the compelling level of believable detail that highlights Benjamin's peculiar predicament, the animators knew they "needed the detail between the markers for the micro-expressions" (Preeg, GRID09[24]).

A popcorn commercial in 2007 presented an opportunity to test new rendering and tracking technologies as research and development for *The Curious Case of Benjamin Button*, however the resulting aesthetics were controversial and disappointing. Fincher tested CG techniques to bring deceased Orville Redenbacher, maker of the iconic popcorn brand in the United States, back for a commercial - to disastrous effect[25]. Due to facial movement that did not resonate truthfully, the elderly CG Orville Redenbacher character tested in the commercial failed the basic perceptual challenge and the advertisement fell into the *uncanny valley*. Preeg explains that by a process of elimination, it was learned that a stronger focus on speech and detailed motions of the various *individual* components that comprise the face had to be attended to for CG Benjamin to be convincing (Desowitz).

At the time these tests were being administered, the traditional optical motion capture systems utilised to capture facial motions failed to robustly track eye movements. The software could not overcome the crucial component of accurate gaze estimation "in realising 'intent'" (Ulbrich). It could execute different actions but not decipher, for instance, between an ironic smile and a sarcastic smile. Peter Busch agrees that the most problematic elements in the facial motion capture tests being conducted that did not cohere with convincing facial

[24] GRID is a two day summit of international speakers hosted annually by Swedish company, Bonnier AB in Stockholm to inspire creativity, network and highlight innovative approaches in technology. A video of character supervisor, Steve Preeg giving a presentation on facial motion capture techniques developed for Benjamin Button can be found on www.vimeo.com

[25] The video of the completed commercial can be viewed at http://www.youtube.com/watch?v=Fcn4p213Zg8.

performance were the eyes and lip-synch. "You can't put markers on the inside of the lips and you can't put markers on the eyes" (Interview). At a presentation, this time at the fmx conference[26] in 2009, Preeg elaborated on the scheduling dilemma presented by the absence of information on the eyes and the inside of the mouth - that is, the teeth and the tongue - that extended to several months of extra work. In observation of the Redenbacher commercial, it is evident that this was an area for concern for the animators. Barbara Flueckiger, published faculty member of the University of Switzerland - Film Studies, explains in her 2011 article, *Zur digitalen Animation von Körpern in Benjamin Button und Avatar* translated[27] (10) that the models had to be constructed and added in manually by multiple modellers and animators. Through these challenges, a new process which Preeg and Barba dubbed "*emotion* capture" was born (Seymour and Dawes 9).

4.1.2 The Facial Motion Capture Process: Overcoming the Technical Challenges to convincing Facial Animation

Ulbrich mentions in the TedTalk presentation that the methodology of the development of Preeg and Barba's emotion capture pipeline incorporated several different technologies evolving out of academic institutions and fields of medicine and sports. This study will investigate and concentrate on the three individual types of 'capture' that were delineated: Model Capture, Camera Capture - which happened on set - and Performance Capture, which was image analysis from a new type of optical flow system.

As an initial starting point in the journey, the CG modellers scanned very detailed maquettes - or head sculptures - commissioned by Digital Domain to depict Pitt at age sixty, seventy and eighty. A common practice is to create and scan a maquette to serve as the basis for complex organic forms when modelling in a 3D software package. It allows for an easier commencement point than starting from scratch (Prince, "Digital" 138). By remapping Pitt's features, The CG modellers would know precisely what Benjamin would look like at those ages as ambiguity in interpretation would be a defeat before they even began. It was noted by Preeg that on the Redenbacher commercial, this was one area for improvement as the

[26] Character Supervisor, Steve Preeg gave a presentation on facial motion capture techniques developed for Benjamin Button at the fmx conference in Stuttgart, Germany in May, 2009 (http://09.fmx.de/ start.php?lang=E&navi=1&page=pages)
[27] Translated from German by Benjamin Letzler. The English title is "Computer Generated Characters in Avatar and Benjamin Button"

maquette on that advertisement lacked detail. Prince maintains that having Pitt's continuity of personality also adds to making the character conception work visually (139).

Preeg relates that Fincher mandated that it must be a direct translation of the specific performance emoted by Pitt and not the traditional interpretation of that performance by an animator to the digital model. Research shows that Fincher and the team's fear was that they would run the risk of 'volume loss' or inaccuracies (Preeg GRID09) ; it would no longer have a veridical connection to the indexicality of Pitt's performance, leading to a breakdown of sympathy by the audience for Benjamin Button. In answering his own simple question to his team, "how do we make him move?" Preeg was instrumental in designing a proprietary volumetric[28] deformation rig to overcome the risk of an animator interpreting in their own style what Pitt emoted in a particular shot. "Key-frame animation drawn from the reference library of Pitt's FACS expression poses shaped and modulated the character's responses" (Prince, "Digital" 138). A proprietary FACS solver formed a relational database of Pitt's micro-expressions with Mova Contour's volumetric capture data and Image Metrics' facial animating software.

4.1.2.2 Mova Contour

For a key component in the performance capture pipeline, Digital Domain collaborated with facial capture company Mova on their volumetric capture system named *Contour*. Surface meshes of Brad Pitt’s face in a multitude of FACS facial expressions were captured using Mova’s *Contour* facial-capture system. Although volumetric capture systems were neither new nor unique, Preeg points out that one of the distinguishing features of the *Contour* system was that the software "captured the pore-level detail of each movement with absolute fidelity. There was no rig[29], just [surface] mesh detail" (4). This afforded an artistic freedom for the animators to create an aged digital Benjamin with very detailed blend-shapes reinterpreted from the Mova meshes.

[28] Volumetric in this case is the construction of a 3D mesh head in that will be the CG character, Benjamin Button.
[29] Rigging is the process of inserting a skeleton of digital bones that essentially drive the movement of the mesh in much the same way a skeleton holds our bodies.

The second distinguishing aspect of the software that allows Benjamin's head to occupy the unified space with the actor enacting Benjamin's body movements is described by Eric Barba in an interview with Bill Dawson. He says:

> we didn't use Mova to capture [Pitt's] performance, but to capture facial expressions volumetrically. Twenty-eight cameras are mounted in an array around the actor. The cameras are aimed at the actor's face which is covered with phosphorescent make-up. This allows for frame by frame tracking of patterns and each point can be tracked in 3D space. [However] this is the first system to truly capture someone's face moving in real-time and provide a moving mesh that can be subdivided and rebuilt then retargeted to another mesh to drive a CG performance [see Figure 9] ... We used the [video] image analysis technology [from Image Metrics] to get animation curves and the timings that drove our proprietary deformation rig. (4).

Fig. 10. Clear aging of Brad Pitt's character Benjamin Button with retargeted mesh. Head 1 is Brad Pitt pulling one of the Eckman FACS poses in the phosphorescent makeup. Head 2 is the digital version of Brad Pitt's doing the same pose in real-time and Head 3 is Benjamin, aged and convincingly being driven by Brad's performance. The data from Brad at age 44 is transposed (or re-targeted) to the digital mesh of Brad at age 87.

Source: Ulbrich, Ed. "TedTalk - Ed Ulbrich: The Curious Case of Benjamin Button". *Youtube*. Youtube, February 2009.Web. October 2013.

Barba's understated explanation that this was the "first time a software provided a moving mesh that can be subdivided and rebuilt then retargeted to another mesh" is explained by Preeg as one of the breakthrough technologies that brought *The Curious Case of Benjamin*

Button across the *uncanny valley*. He clarifies that what differentiated Mova's technology for this project was the systematic manner in which it constructed the mesh from frame to frame. The mesh was not a haphazardly, random calculation of constructed geometry as the meshes produced by other photogrammetry software were at that time. The randomness deems the meshes unusable. Instead the geometry that was produced "allowed a temporal coherence to the mesh" (Preeg GRID09). That meant all vertices[30] were *consistently* in the same order from mesh to mesh, frame to frame. This is vital for the high detail re-targeting from Brad Pitt to aged, CG Benjamin Button as consistency was the key.

Each geometric mesh head produced by Mova had over 10 000 vertex points, alluding to the very detailed quality denseness of the facial geometry in comparison to the 150 markers that could fit on the face previously - as was the case with the marker-based technique utilized for *The Polar Express* and *Beowulf.*

4.1.2.3 Image Metrics (Faceware Technologies)

Months after principal photography (production), it was necessary to do an image analysis of Brad Pitt's facial performance to calculate facial actions, "to assess when the left brow moves up and his nostril flares on a certain word" (Preeg, GRID09). After the successful *Digital Emily* project, Image Metrics' proprietary video analysis and animation system was called upon as the quality of the software had several advantages over traditional performance-driven animation techniques at the time. This added significantly to the realism of CG Benjamin. The software is designed to allow animators to associate the character poses with a small subset of performance frames. "The analysis of the performance required video from standard video cameras in a performance designed to capture all the characteristics of the actor" (Alexander, Rogers and Lambeth 10).

Peter Busch explains to Barbara Robertson, editor of CGWorld.com, that once Faceware technology identified key features in the face, like the eyes and mouth, "it knew where the rest of the face was and tracked the motion. What used to take an hour or two now takes two minutes" (cgworld, facing). Faceware software automates the video tracking

[30] Vertices are the corner points on the mesh that describe and inform where that point on the mesh is lying in 3D space.

process. In the demonstration at the Image Metric offices, it was learned that once the data is tracked and analysed, the re-targeter software moved the facial tracking data onto the facial rig that had been constructed in order to direct the animation. Flueckiger explains that four high-resolution digital Viper cameras recorded Brad Pitt’s facial performance from multiple perspectives (10). See Figure 11.

With the provided animation curves for the Benjamin facial rigs, animation artists were given an advantageous start for a number of shots as it "leverages an appropriate division of labour between the animator and the automated algorithms" (Alexander et al. 10). This process saves valuable production time if the data is of high quality. The animators were then free to augment and elicit the nuances of Pitt's performance using standard key-frame techniques and blend-shape information from the meshes provided by Mova. Busch asserted that "the animator brings that human element that breathes life into the model" (Interview). Their creativity is enhanced by the software. There is an easier facilitation in ensuring "each of the key animation poses reads in an emotionally faithful way to the appearance of the actor in the corresponding frame of video" (Alexander et al. 10). Preeg freely admits that the analysis data was very helpful, however the primary results were far from perfect. He acknowledges that the bridge that connected the analysis data they were presented with and what emerged from the performance "was the most difficult part. To augment emotions is not a tangible, easily executed task" (Preeg, GRID09). It took many man hours to achieve the level of the performance that they did.

Fig. 11. Brad Pitt being filmed with four high resolution cameras as part of the performance capture process. Image analysis technology from Image Metrics was used to get animation curves and timings to transfer the data to the facial rig set up for the digital face model. Source: fxguide.com, December 2009. Web.

4.1.3. Crossing the *Uncanny Valley*

The research reveals the trepidation on the part of the creative team that the CG version of Benjamin Button would not convey that ultimate level depicting sadness or wonder and this drove their versatile approaches to startling new heights. In Lee Stranahan's video clip, Preeg admits to a heightened awareness that almost imperceptible changes in the human face elicit any number of emotions. "Every other project I had worked on, other than *Final Fantasy,* was a [fantasy] creature ...which allows you to get away with alot" (Stranahan). In the case of *The Curious Case of Benjamin Button* (2008), the phenomenal efforts of the team negated Fincher's fear that audiences would not respond to Benjamin; that viewers would arrive instead at an interpretation that would render their aesthetic sensibility ambiguous, sending the movie tumbling into the *uncanny valley*.

One of the crowning achievements in the pursuit to cross the uncanny valley has been the humanity depicted in Benjamin Button's eyes. To capture the subtleties of the inside of the eyes, - as they could not spray phosphorescent paint in Pitt's eyes -a sophisticated eye system was developed and improved upon from what was learned on *Beowulf* (2007). Ed Ulbrich shares that the intention was to have everyone focused on Benjamin's eyes - if viewers could feel the warmth and humanity, they knew they would succeed in anchoring aged Brad Pitt in that scene and supersede the boundary blocking dynamic extensions of digital performance. A special animator "was tasked solely with visualizing Benjamin's eyes for almost two full years" (Ulbrich). The empirical journey on the development of the saccadic eye system was very similar to the system developed for Gollum in *The Two Towers* (2002) however with CG Benjamin, they experimented further. Prince explains that one aspect was multi-pass rendering that was used to isolate the essential characteristics of ageing eyes ("Digital" 139). Detailed observations and methods included "...the amount of water in the eyes, the red conjunctiva of the eye ..." were rendered separately for ultimate compositing control, shot by shot" (Duncan, "Benjamin" 116).

Itti, Dhavale and Pighin proposed in their scientific research report that gaze estimation is an essential component of nonverbal communication and social interaction (1). To gain an understanding of the milestone achieved by all that meticulous attention to detail, Itti et al. clarify that "the crucial role of gaze estimation in determining intention and in interacting with social partners makes *any* inaccuracy in its rendering obvious to even the most casual observer." The animators working on *The Curious Case of Benjamin Button* knew the mammoth undertaking to pull off the illusion demanded animating Benjamin's eyes perfectly. They commented that minute shifts in the position of an eyelid, as little as one millimetre, could affect dramatic changes of expression, since every viewer is practiced at decoding facial expressions (Flueckiger 11).

An atypical opinion offered by Mark Christiansen[31], the author of a range of Adobe® practical books on visual effects, is that the filmmakers and visual effects crew were aided in avoiding the *uncanny valley*, albeit in a clever way. Christiansen proposed that by selecting an uncanny premise in the form of the unusual narrative framework that allows for an intense emotional investment, Fincher got *around* the *uncanny valley* (podcast, Benjamin). There has been no systematic examination of this view in the research thus far, yet Christiansen does present an alternative logic that merits interest. The temporal storyline is so novel, that the situation forces viewers to pay very careful attention to the surreal premise unfolding in that back-to-front narrative. Christiansen does clarify later in the same podcast that the visual effects of the CG character is extraordinarily well done and not a 'cheat', however he still maintains that viewers are already thrown off balance by their immersion in the story itself to pay deep attention to the effect of the digital head .

4.1.4 Conclusion

Kevin Sweeney, author of the chapter, '*Medium*' in *The Routledge Companion to Philosophy and Film,* expresses his interpretation of realist theoretical belief that "Bazin conceives the cinematic medium as being fixed and immutable. There seems little room for aesthetic creativity in inventing, extending or transforming our conception of film as a medium" (Medium 177). This would make it appear that Bazin's realist philosophies established in this report are incongruent with the integration of realistic CG characters in film. Prince refutes Bazin's idea by citing the example of Brad Pitt in *The Curious Case of Benjamin Button*. He argues - that CG characters inserted into live action footage retain their indexical properties beyond the inclusion of perceptually realistic cues - as follows in this compelling explanation:

[31] Mark Christiansen's latest book Adobe After Effects CS6 - Visual Effects and Compositing Studio Techniques reveals the secrets and how-to's of the latest blockbusters

> Benjamin Button is an animated computer model, not a photographic image, but this model was derived from a scan of Pitt's face ... So, although Brad Pitt does not physically appear as an actor in these sequences or as a photographically derived image, the computer based image we see is indexical. It persuades us that it is an age altered version of Pitt because it carries his trace. And if one accepts the translation of physical space into binary data, then one must accept that the Pitt head replacement is physically connected to the actor as an index (Prince, "Digital" 154).

Manovich emphasises Bordwell and Staiger's definition of realism and the pressure on the film industry in his reflection that the "specificity of industrial organization of the computer animation field is that it is driven by software innovation" (7) and, as in every industry, the producers of computer animation stay competitive by differentiating their products. From the beginning, it was evident that the amount of determination shown by the crew of digital animators on this project was going to yield successful results. Christiansen's portentous comment in 2008, that "the work that was done on Benjamin Button will potentially have an impact on the coming decade" (podcast#64), - has already been proven with the concept of some of those techniques used to enhance fellow award-winning blockbuster *Avatar*. In agreement with Manovich, it is clear that in order to attract clients, a company has to be able to offer some novel effects and techniques. The production team on *The Curious Case of Benjamin Button* have excelled in this regard. Digital solutions brought Pitt's performance to an unprecedented level in visual effects and "for this reason, *The Curious Case of Benjamin Button* is a film of historical importance" (Prince, "Digital" 137).

4.2 Case Study 2: *Avatar* (2009)

After the deserved splash that *The Curious Case of Benjamin Button* made in terms of overcoming the *uncanny valley*, that it would be uneventful for a longer period of time in terms of radical innovations in facial motion capture technology would be a reasonable assumption. However, one year after the release of *The Curious Case of Benjamin Button*, director James Cameron unleashed *Avatar*.

The film's narrative surrounds, Jake Sully, the protagonist who is a paraplegic soldier who travels to the fantastical world of Pandora. The extra-terrestrial moon is a lush jungle, filled with fluorescent life forms and home of the sentient blue-skinned humanoid race, the Na'vi. The ten-foot tall, blue-skinned Na'vi are fighting rapacious human corporations attempting to mine for precious minerals and remove the indigenous people from their native world, that is toxic to humans. As part of a scientific programme that Jake participates in on Pandora, human scientists have created genetic human-Na'vi hybrids known as 'avatars' allowing Jake and other scientists to infiltrate the moon's biosphere and initiate interaction with the local species.

Cameron's mandate was clearly delineated in an exclusive epilogue when he shared that “the goal [of the design] became to mix the familiar and the alien in a unique way. To serve the metaphor and create a sense of familiarity for the audience, but to always be alien in the specifics” (scriptPhd.com). This sentiment would appear to have carried over into the approach to the over-all production. The complexity and sheer quantity of never- before-seen elements including floating islands and mythical animals in the computer-generated universe, celebrates this movie as a revolutionary milestone for visual effects. Scott Bukatman confirms that cinema is already “an artificial and technological paradigm that will realise utopian fantasy” (McClean 207). With *Avatar*, Cameron waited fifteen years to usher in his technological paradigm that extended the flexibility of film, bridging the dichotomy separating traditional film theory as discussed in Chapter 1. His acquiescence stemmed from the implicit recognition that "the technology just wasn't there yet" (popularmechanics).

To realise his narrative vision, similar technological challenges that beset *The Curious Case of Benjamin Button* in the form of the counter-intuitive approaches of the contemporary technology needed solutions to be brought into existence in all areas of the production. As reward for Cameron's profound insistence on perfection, *Avatar* was considered more than a movie - it was an event. Cameron waited for motion-capture to develop to the level of motion capture for *Beowulf* (2007) before embarking on his project. Jim Tanenbaum, contributing writer for the Pandorapedia books and production sound mixer on *Avatar*, writes that four new features were added to the system [used on *Beowulf*]: real-time animation display, video facial capture, the virtual camera and the 'Simulcam' (soundandpicture). The analysis that follows will be a critical investigation focusing on a learning-based muscle system that minimised key-frame animation to augmenting expressions and two of the innovative cameras that revolutionised the approach to facial motion capture for the digital characters on *Avatar*.

4.2.1 Facial Performance Replacement on *Avatar*

Jon Landau, the producer on Avatar says that “over the years, some image-based facial capture was tested on [the never made movie] *Brother Termite* ... that proved to us that there was some validity there" (Menache 72). Landau's resolve that the collective efforts of James Cameron and himself could be the necessary impetus to drive the cog of technology was strengthened by the astonishing results of Gollum in *Lord of the Rings: The Two Towers* and the pioneering motion capture efforts of director, Robert Zemeckis. From *Beowulf* it was learned that the cameras reading the reflective markers were placed too far away for veridically effective facial analysis (popularmechanics). Cameron reveals that trying to use the same system with that set-up for the face and body yielded disappointing results and, "as the face has almost as many muscles as the human body" (Riley), that section needed to be monitored five times as closely, but separately. A capture system was created on *Avatar* that 'uncoupled' the facial performance from the physical performance. Cameron asserts in an interview with Forbes writer, Josh Wolfe, that image-based facial performance capture had never been used in a feature film before. On *Avatar*, the process for gross body motion with the facial performance "had to be separated to get two data streams for each character. One data stream [the head-mounted camera] for their image-based facial recording; the other one is a cloud of moving points that represent their body motion. The two are put together later in the pipeline, when that scene is ready for a virtual camera session" (forbes).

4.2.1.1 Head mounted Camera (Headcam)

For *Avatar*, a head-rig reworked from Cameron's idea in 1995 was created using a conformal carbon fibre helmet with a little standard-depth camera (headcam) that is trained to record only facial performance and provide all the expression inputs for the surface mesh of the 3D character (see Figure 12). Cameron needed the performers to be very agile in the large motion capture space known as 'the volume' and this rig solved the issue of mobility and unobtrusiveness. Through critical assessment and examination of alternatives, Joe Letteri, senior visual effects supervisor on *Avatar*, revealed to Mike Seymour that the team decided to remain with the single camera approach (podcast, Avatar-Weta)[32]. Having a camera attached to the performer also provided constant reference footage for the motion editors, documenting all the subtleties of the actor's performance. The term, motion editor, derived from those animators who use chiefly motion capture data with minor touches of traditional key-framed animation techniques to produce the high level of animation demanded by Cameron. That level of detail "had been lost for many of the CG characters that have come before. In those films [that fell victim to the *uncanny valley*], the process of motion capture served only as a starting point for motion editors, who would finish the job with digital brush strokes" (Cameron, popularmechanics).

[32] In this very informative FXGuide podcast Mike Seymour speaks with Joe Letteri detailing their facial motion capture technology, the decision to use a single head-mounted camera, the evolution of their pipeline, a description of the FACS muscle solver and other aspects of their work on *Avatar*. A transcript of part of the interview described in the topics listed is included in the Appendix 3. The podcast can be found at www.fxguide.com/fxpodcasts/Avatar_Weta_Digital/

Fig. 12. Sam Worthington as Jake Sully on the motion capture stage. Inset: Example of the projected image captured by the head-mounted single camera (headcam). Picture Right: The head-mounted camera visible to the actor with the array of fixed lights to offset the lights on the capture stage that interfere with the video capture data. Source: Discovery. “Avatar: Motion Capture Mirrors Emotions” *Youtube*. Youtube, 24 December 2009. Web.

Flueckiger states that "the images produced by the helmet camera – unlike the Image Metrics system used in Benjamin Button – did not provide reliable movement data" (18) and cites two reasons that prevented the team from using the footage as reliable movement data is the one-frontal perspective, distortion of the fish-eye lens, and that a very high-resolution image quality is required for pixel flow analysis to function" (18) Jody Duncan's article on *Avatar* in Cinefex digest, cited in Flueckiger, elucidates that for pixel flow analysis to function optimally, high-resolution quality footage is required - as the footage of Brad Pitt supplied to Image Metrics was. Should the resolution not be high enough, filtering problems arise similar to those created with markers. The single helmet camera displays a frontal 2-Dimensional perspective, distorted by a fish-eye lens (inset of Figure 12), as opposed to the multiple recording angles as demonstrated in the high-definition multiple views of Brad Pitt, shown in Figure 11. Mark Sagar renders Fleuckiger's statement inaccurate in the podcast on fxguide.com stating that the digital team at Weta eventually solved the issue and got FACS to work with the low resolution, therefore allowing for the low resolution image to be used. The digital team at Weta developed a real-time system for footage received from the head camera (Appendix 1). He also elaborates on the other digital solutions that were found to counter issues such as the instability of the camera when the performers used broad action. In the search for a viable strategy to overcome the low resolution limitations of the lightweight cameras at the time, Letteri recounts that the team had considered putting more cameras onto the helmet. But this caused the rig to became heavy and unbalanced. "It was cumbersome to synchronise each camera with the 3D analysing software - not to mention uncomfortable for the actors" (podcast, Avatar-Weta). Before the team knew they could overcome the limitations of low-resolution, should Cameron choose a take from the head camera, the mobility and flexibility the cameras afforded for the performance was worth it. Letteri counters that they "lived with the limitations and figured out a way to solve the 3D [issues] that arose from having the one camera" (Avatar, "podcast"). Throughout production, the testing process went through a rigorous process of trial and error until an eventual solution was found (podcast Sagar). The FACS solver was enhanced to accommodate the distorted image and that footage became useful.

A system was tested for reading facial expressions from the headcam that "used an edge detection algorithm" writes Tanenbaum explaining further how the system that Mark Sagar's team effected calculates the motion, "that senses pupil size and position, mouth shape, tongue and teeth position and any skin wrinkles that form" (soundandpicture). A similar process that was documented for *The Curious Case of Benjamin Button* ensued whereby the actors repeated their facial performance for image analysis, this time recorded by multiple high-resolution cameras (Prince, "Digital" 135). The recordings were called Facial Performance Replacement (FPR). Where the headcam was also used effectively, is the footage was used to help finely tune the facial features that proved problematic in facial animation. Having a camera so close to the face one hundred percent of the time granted access to the inside of the mouth and accurately read eye movement. Their eventual triumphant solution was to combine the image analysis data with a "dynamic learning-based muscle system cued from FACS that digital artists at Weta Digital were familiar with from *King Kong* (2005)" (podcast, Avatar-Weta).

4.2.2. Enhanced Muscle System and the dynamic FACS Solver

The previous belief was that if better surface scans of the actor could be obtained, there would be enough data to have a believable CG character - as with the method of the Mova Contour system utilised to enhance facial movement in the CG Benjamin Button. Cameron dismissed this blend-shape approach, saying "that wasn't the answer; we had to get *below* the surface, at the muscular level. We have as many muscles in our face as we do from the neck down. The face is an incredibly complex, fluid system" (Wolf, forbes). In appreciation of the genesis of facial capture from the digital rotoscoping technique employed for Gollum to *King Kong* (2005) culminating in the Na'vi in *Avatar*, Letteri acknowledges the connection and progression of technology between those projects (Avatar podcast). However, this biological analysis his team had expanded, orchestrated an organic approach tempered with technical discipline that supersedes traditional retargeting methods. Dating from the system developed for Gollum and King Kong, Letteri's team advanced to a robust fully-developed muscle system based on the FACS system to solve for the muscles (Avatar, podcast).

Cameron clarifies that "the information from the cameras produced a digital framework, or rig, of an actor's face. The rig was then given a set of rules that applied the muscle movements of each actor's face to that of the CG avatar" (Cameron, popularmechanics). The information received from analysed footage, combined with coding expressions in the face information is derived from the analysis data that is supplied. With knowledge of the FACS system, a deep and critical dissection of a particular facial expression is executed. This system mimicked "an expansible animation system" (podcast, Avatar-Weta).

Letteri expands on the inner working of the muscle system solver, explaining that 'weights'[33] were assigned to certain muscles and certain muscle groups in the face - and a description was assigned to how each group was activated. The strength value of each group of muscles was assessed - as was the combination of that muscle group - to create the facial expression that was presented to the motion editor. The system was designed to be an

[33] Each muscle is mathematically assessed and given a corresponding computer code for identification in the system

accumulative, learning-based muscle system. Addressing the power of the toolset that motion editors were now armed with, Letteri shares that "if we saw a new [muscle] shape that the system hadn't encountered before, we put it into the learning data base and added that to our solver. The next time [the system] encountered that shape, it would be able to remember it" (podcast, Avatar-Weta). Should the need arise to edit anything, he adds, the 3D animator can vacillate from the solver straight into key-frame animation . When asked what is was they were trying to achieve specifically with their solver, he continues:

> when we solve, all the points on the [actor's] face is assessed. From those points, the system is trying to derive which muscles are being engaged - and to what degree. That [information] is utilised to drive the muscles on the CG character. We are trying to understand what the emotion is by reading what muscles are engaged. That really *is* the key to the performance. (podcast, Avatar-Weta).

Other re-targeting systems work on the surface detail of the mesh. What differentiates Weta's muscle system from other re-targeting systems is that "the system is not just taking the geometric data from one face and targeting it to another face. It is actually solving for the underlying muscle activations and then activating those same muscles to the same degree on the new character" (podcast, Avatar-Weta).

4.2.3 The Simulcam

The realization of the nuanced interaction with the feline Na'vi or achieving a sense of sheer exhilaration as Jake swept off the cliff on his banshee was "facilitated by Cameron's ability to direct digital actors alongside real actors in real time" (Prince, "Digital" 134). Cameron explains that the sophisticated simulcam[34] perfected for A*vatar* combines virtual production toolset and superimposes it onto the physical production inside the camera, effectively combining disparate elements to inhabit a unified space. "The set gets turned into a capture volume and turned the physical camera into a capture-virtual camera, so we were able to integrate CG characters and environments into live action" (youtube, simulcam).

[34] For a clear demonstration by the simulcam supervisors on *Avatar* of the process that the simulcam officiates, see: http://www.youtube.com/watch?v=lyHa_0yJBlw

Fig. 13. James Cameron filming with the Simulcam on location.
Source: Lohmeyer, Edwin. *From Cellulloid Realities to Binary Dreamscapes.* Masters Thesis - Art & Visual Studies. University of Kentucky, 2012

Traditionally, in the conventions stemming from Méliès and Lumiére, the live action filmmaking process is strictly divided into production - shooting the live action footage - and post production, when, at a later period, the data is collated, translated into a 3D software package[35]. These are eventually composed in another digital tool in order to see the final performance - as was the process used with *Benjamin Button*.

As opposed to a director of a fully animated feature film who is accustomed to the organic, traditional post-production process for a live-action filmmaker like Cameron, the ontological element that the simulcam leverages, "a process that is more aligned with their experience of making movies" explains Casey Schatz, simulcam supervisor (youtube, simulcam). One of the numerous benefits of this process is that actors are able to react to elements that are physically present (Figure 14).

Jim Knight, motion capture producer on the project, also informs that the interaction between the CG characters viewed in real-time helped to eliminate one big hurdle with

[35] Autodesk, maker of high-end 3D software packages for animation, developed a motion capture data collection programme called Motionbuilder®. This programme translates the data collected from the markers attached to the actor during their performance and allows the motion editor (animator) to transfer that information to the 3D character.

motion captured performances, which is 'point of contact'[36] - especially with Avatar's multiple performers (cgchannel, Knight).

Fig.14. View from the Simulcam of Jake Sully's Na'vi avatar (Sam Worthington) and Neytiri (Zoe Saldhana). Source: Lohmeyer, Edwin. *From Cellulloid Realities to Binary Dreamscapes.* Masters Thesis - Art & Visual Studies. University of Kentucky, 2012

Through these indexical actors being captured in front of the camera, and the traces of the CG characters blending instantly with the live footage as Cameron looked through the eyepiece of the camera, the director was able to position the shot. His digital construct of cinematic realism in the camera displays the nuances of the performance and, he interacts with that reality, being able to immediately change what was not working. This afforded a final performance output that is far more expressive, natural and superior than other thematic movies. Alluding to Prince's terms, the facial movements of Jake Sully's Na'vi avatar and Neytiri are seen as creatures that are understood through perceived correlations with the viewer's reality because the movements were appropriated from an organic collection of performances from the actors, Sam Worthington and Zoe Saldana. The head mounted camera footage was reference for the motion editors who would be augmenting or adding the final touch to the digital Na'vi. Through the simulcam, Cameron has a vast array of camera shots

[36] As the data does not relate the size of the CG character, only skeletal movement, point of contact refers to when one CG character touches another CG character- or the CG character is of another body mass e.g. someone who is more portly- you must be careful that the characters are "touching" and not floating in space.

to view immediately and make his choice based on what he has already see, as opposed to having to wait to see the collated data days later.

4.2.4. Conclusion

The technologies that have aligned to culminate in the collaborative success of *Avatar* have been nothing short of revolutionary. There were many aspects of the industry that experienced shifts in the traditional procedure for achieving the result that was desired. When Steven Shaviro posits in his article, *Emotion Capture*, that the truthful character of the photographic image has vanished because "the very shift from analog to digital destroys the indexical nature of the photograph [and] there is no longer any ontological distinction between a 'true' image and a 'false' one" (65), discourse leaves a strong reminder that "a goal throughout [*Avatar*]was to take motion capture from being a post-production tool to one that was integrated with the filming and directing of live-actors" (Prince, "Digital" 134). By combining the two processes, the technological advances of the FACS Solver and the Simulcam, in particular, have overcome the ontological nature of the film medium and converged with traditional film theory established in Chapter 1.

It may be true that digital photographs can no longer function as indexes of traditional cinematic reality in the digital age or that "photographic images themselves are no longer objective in Bazin’s sense" (Shaviro 65) but that does not mean that digital images are not mimetic. The goal in these two case studies has been to achieve photorealism, and through an immense galvanizing effort on the part of digital artists and technicians, that goal has been attained.

As Roger Wyatt pointed out in regards to the ambiguous nature of these thematic digital films, "digital cinema images are taken from life and reworked into another image of reality that is often surrealistically interdisciplinary in nature. These images occupy a conceptual zone somewhere between videography and animation ... Reality becomes extended by abstraction into visions of reality" (Lohmeyer 13).

By repacking Christansen's loose premise of *The Curious Case of Benjamin Button's* unusual back-to-font narrative that allowed Fincher to circumvent the *uncanny valley*, a similarly aligned view could be applied in the case of *Avatar*. The *uncanny valley* theory may not strictly apply in this case, as the CG characters are not digital humans. However if we apply McClean's argument that the characters in the film do not 'pass for human', they attempt to 'pass for real', (MacClean 65) the digital Na'vi are governed by the same rules of perceptual and cognitive realism that governs the creators of *The Curious Case of Benjamin Button*. "The concept of ‘perceptual realism’ proposed by Prince, in which the viewer understands CG through the way certain perceptual cues programmed into the film’s represented reality match up to their real-world, visual experience" (Lohmeyer 9). Due to unprecedented advances demonstrated in these case studies, Cameron has achieved an allusion to perceptual reality which unequivocally places the Na’vi and their home planet of Pandora in narrative space that conforms to the viewer's understanding of cinematic realism.

Conclusion

It has been established that the shift from analog to digital technology has fuelled the paradigmatic shift in viewers' expectations of a cinematic experience. On the shifting perceptions of what viewers consider persuasive digital imagery, Bazin states in his essay, *The Myth of Total Cinema*, that "cinema is an idealistic phenomenon [that] exists fully armed in [viewers'] minds, as if in some platonic heaven…" (Bazin, “Evolution” 21). This interpretation, in effect, reasons that introducing any additional technologies to achieve these increasingly ambitious narratives would only continue to build upon the already conceived mental image of an ultimate cinematic illusion that each person already holds in their imagination. This could be viewed as one of the sources of why the quest for realistically integrated CG characters has intensified in recent years. As realised by James Cameron in Chapter 4, research shows that for many years the limitations of technology deemed it impossible to meet the demands of photorealism, as did the dual concern of the the lack of specialised technological hardware and software, paired with the lack of specialised technological expertise in the industry.

Digital film technologies have improved in all areas of production, so has the ease with which they can provide hyper facsimiles of reality that are perceived as photographic. The increase in the number of motion-captured characters that have proliferated modern cinematic narratives has been in direct correlation with the resurgent interest in Masahiro Mori's theory of the *uncanny valley.* It was evident in many areas of the research that, irrespective of the lack of scientific proof at the time it of its inception, the *uncanny valley* phenomenon had made an impact in computer graphics that were attempting photorealism and the theory was later proved to be valid. One of the findings of the report's research goal - to show how CG characters overcame the *uncanny valley* - was that the core meaning of the *uncanny valley* can be extended to the simulacrum of human-like CG creatures attempting photorealism, such as the Na'vi. That sense of *unheimlich* associated with the *uncanny valley* extends to the anthropomorphic King Kong and as well as the Na'vi as they may not be attempting to "pass for human" but they are trying to "pass for real" (MacClean 60). As such, the notion of perceptual realism holds these CG characters to the same intense amount of scrutiny to sit comfortably in a world on screen as would be applied to CG Benjamin Button extending the indexicality of the medium of film. As a photorealistic CG human, the exact technological cues to carry the same level of information as a photograph, would apply to Gollum, King Kong and the Na'vi.

Neuroscience espoused a solution to an on-going challenge faced by animators trying to animate realistic facial expressions. Dr. Paul Eckman and WV Friesen systematically broke down the human face into zones and muscle groups with their Facial Action Coding System (FACS). It was established with *Final Fantasy* that traditional key-frame animation could not capture the dense, elusive quality of facial micro-expressions and FACS was systematically applied as a reference for animators for many years. However, the degrees of success of these techniques varied, based on the skill or left to the whimsy of the animator. The breakthrough for facial animation in 2006 was yielded by Dr Mark Sagar's complex system of algorithms to develop a FACS 'solver'. This has been instrumental in the ambitions and ultimately the success of crossing the *uncanny valley*. The implementation of a systematic combination of Action Units (AU) dividing facial features into quantifiable zones and muscle groups for each micro-expression reduced, but did not fully remove, the intangibility of facial nuances to something that a logical 3D programme could understand. The motion editors had a tool that could better imitate those previously difficult to mimic micro-expressions.

The mastery of this tool and improved technology ushers in the broader critical debate of who owns the “performance” – the actor whose facial expressions are being recorded or the motion editor whose human creativity aids in interpreting that performance by choosing where to emphasise or de-emphasise motions of interest on the CG character? The answer lies in sharing the performance. Motion editors handle the technical issues of interpreting the data, but ultimately they place their personality into crafting the data. The challenge of the motion editor is to master the creative end of the motion capture scale as it is the technology coupled with the artistic interpretations of the data and the actor’s performance that overcomes the *uncanny valley*.

The success of the CG characters in both case studies, CG Benjamin Button and the Na'vi, points to facial capture technological solutions being an on-going collaboration between the field of computer research and the film production industry. In the first case study, *The Curious Case of Benjamin Button*, once it was realised that motion capture markers were not good enough for capturing realistic movement of the face, it was found that the alternative marker-less systems had the greatest results in tracking facial movement and expression. Digital photogrammetry and optical flow analysis being two techniques that helped CG Benjamin Button cross the *uncanny valley*. Despite cynical discourse regarding the unsuitability of this technique due to previous failed attempts, Digital Domain technicians saw the prudency in nurturing their collaboration with Mova Contour. Mova's development of a volume capture system was based on photogrammetry. Their solution was being able to track video footage and construct Brad Pitt's head in CG, frame-by-frame. Many other software programmes could perform the same function but where Mova succeeded was in being able to construct the geometry with the vertices consistently, frame-to-frame. The consistent geometric temporal-spacial transference was the key to keeping all the blend-shapes and animating conventions in their place. That was one of the innovations that helped CG Benjamin Button sit convincingly in the environment.

Optical Flow Analysis, like the Image Metrics system, is a sophisticated technique of tracing video footage. Every pixel from every frame is analysed and tracked and converted to data to be transferred to a CG character. This system did not work well with body movement, but with facial capture it was more accurate, due to the limited range of mobility of the facial features. It was also learned that this is an area of fervent interest and intense research as there is scope for enhanced facial tracking, especially trying to find a better solution for the complexity of eye features. This posed one of the biggest challenges to crossing the *uncanny valley*.

In the second case study it was revealed that the team at Weta Digital went a step further for *Avatar* and developed a proprietary FACS-based Muscle Solver that changed the geometry on a muscle level - *under* the surface mesh as those are the muscles in the face that are responsible for expressions (See Figure 7). This is a learning-based system that accrues its own data base of facial and micro expressions as the motion editors craft the CG performance. What differentiated this system, as opposed to other proprietary FACS solvers, is that it combined the data from image analysis and the muscle system to produce the levels of subtle detail that the motion editors needed for natural facial expressions. The stream of footage retrieved from the specially designed head-mounted camera provided invaluable reference in this endeavour - not solely for the actors who had to return to recreate their performance for high resolution cameras - but to see close up, inside the mouth and the eyes at the time that an expression was made in the process of action. The dynamism of the movement in the performance was what the motion editors tried to capture faithfully. The unobtrusiveness of the headcam allowed the performers freedom of movement. Facial replacement at a later stage took care of the expressions to be transferred to the CG characters.

The focus of the *Avatar* case study was on pioneering facial capture solutions. However a direct, paradoxical consequence of these solutions has been the evolution of the role of the traditional 3D animator to that of a motion editor. The limitations of traditional animation techniques were not conducive to producing photorealistic facial expressions and complex face movement found in the characters in a fantastical world such as Pandora. The advantage of motion capture is that although the techniques are very effective, the processing of the data still requires human intervention and interpretation. The function of 3D animators animating photo-real facial expressions includes absorbing and combining data streams from traditional marker-based motion capture data with data obtained from new technologies like digital photogrammetry and optical flow analysis or absorbing any number of other experimental techniques for capturing facial motion. These are then combined to essentially produce an animation that will eventually be augmented with key-frame animation. Managing and interpreting those extra sources of information requires versatility and specialised skill. In the pursuit of crossing the *uncanny valley*, the motion editor needs all the equipment that will assist with micro-expressions and saccadic eye movements for the enormous challenge posed by photorealistic integration of the CG characters into film. As Cameron points out, "the face has as many muscles as the rest of the body" (Wolf, forbes) and unlike many muscles in the body, these muscles are exercised with alarming frequency. It would therefore make sense to assign a motion editor to be responsible for just the face.

Looking ahead, the future of facial motion capture is an area that is being improved upon daily, rendering it an exciting field of interest. An area that could benefit from immediate research is a solution for realistic eye systems. There were elaborate CG eye solutions developed in both case studies, but presently, it is noted, that the final depiction of realistic eyes is still a painstaking and very precise art interpreted by the 3D animator. This area is receiving attention - and it has done so for quite some time - however a workable solution such as FACS for facial expressions has yet to be introduced.

Another future development that no doubt will be very carefully observed is underwater facial capture techniques that are presently being researched and tested for the

Avatar sequels. The movies are scheduled to be released in the coming decade and already the technology is being touted as the potential next step in faithful facial capture. There is already speculation and discourse on the new cameras that are being developed for all the challenges associated with underwater submersion.

In addition to a successful collaboration between a research institution and production company, Digital Domain with the Institute of Creative Technology (ICT) at UCLA, for a realistic skin system for CG Benjamin Button, an interesting observation throughout the research visits abroad revealed many collaborations between computer science research institutions and visual effects production companies on all aspects of the animation and visual effects industry. Peter Busch stated in the interview that their success with *Digital Emily* was in a large part due to their investment of time and energy in collaborations with institutes. Vision dictates that growth in the industry can be enhanced by a strategy that the South African animation industry can also adopt: investing in long-term collaborative projects with scientific resources of local computer and science research institutions to develop unique innovative solutions.

Works Cited

Aldred, Jessica. *From Synthespian to Avatar:Reframing the digital human in Final Fantasy and Polar Express*. Web. 15 November 2013.

Alexander, Oleg, Mike Rogers, William Lambeth, Matt Chiang and Paul Debevec. *The Digital Emily Project. Scientific Research Paper*. Los Angeles: SIGGRAPH 2009 (Course Notes), 2009. Print

Alexander, Oleg; Chiang, Jen-Yuan; Ma, Wan-Chun; Debevec, Paul. *The Digital Emily Project: Achieving a Photorealistic Digital Actor*, Computing Now (2010 July/August): 20-31. Print.

Andrew, Dudley. *The Major Film Theorists.* Oxford University Press, 1976. Print.

Bazin, Andre. *What is Cinema?* trans, and ed. Hugh Gray. Berkeley: University of California Press, 1967. Print.

Bee, Nikolaus, Elisabeth Andre, Thurid Vogt and Patrick Gebhard. *The Use of Affective and Attentive Cues in Empathic Computer Based Companions*. In Wilks, Yorick. Close Engagements with Artificial Companions. Amsterdam, Netherlands: John Benjamins Publishing Company (2010): 131-142. Print.

Bernard, Kaitlin. *Between Reality and Realism: CGI and Narrative in Hollywood Children"s Films*. MA Thesis. Ottowa, Canada: University of Ottawa, 2011. Print.

Bitzer, Josh. *A critical look at Motion Capture*. 12 May 2012. Web. 17 January 2013.

Bregler, Chris. *Motion Capture Technology for Entertainment.* IEEE SIGNAL PROCESSING MAGAZINE (November 2007): 158-160. Print.

Brown, Scott. *Wired Magazine - Dark Knight Director Shuns Digital Effects For the Real Thing.* 23 June 2008. Web. 20 November 2013.

Busch, Peter. *Motion Capture and the Uncanny Valley.* Personal Interview. 20 August 2012.

Cassimus, Paulina. *Digital Imaging and Film Theory.* 5 August 2011. Web. 21July 2013 www.pcassimus.tumbler.com.

Christophers, Kelly. *Realism in CGI Character Facial Performance: A Comparative Study of the Evolution of Key-Framed Animation and Motion Capture at WETA Digital Studios.* MA Research Report. University of the Witwatersrand, Johannesburg, 2011. Print.

Currie, Gregory. “Film, Reality and Illusion”. Bordwell, David and Noel Carroll. *Post Theory: Reconstructing Film Studies*. Madison, Wisconsin: University of Wisconsin Press, (1996) : 325 - 344. Print.

Desowitz, Bill. *Bringing 'Benjamin Button' to Life.* 1 January 2009. Web. 7 January 2013. www.awn.com.

Duncan, Jody. *The Seduction of Reality*. Cinefex (Vol. 120). 2010 : 68 - 146. Print.

Duncan, Jody. *The Unusual Birth of Benjamin Button.* Cinefex (Vol. 116). 2009: 70 - 99. Print.

Flueckiger, Barbara. *Computer Generated Characters in Benjamin Button and Avatar* (Translated from German by Benjamin Letzle). Harro Segeberg (ed.): Digitalität und Kino (2011) : 1 - 28. Print.

Fox Movie Channel. "Avatar Behind The Scenes." *Youtube*. Youtube, November 2009. Web 10 June 2013.

Freeman, Will. *Motion capture: Moving with the times*. 16 January 2013. Web. 20 January 2013. www.develop-online.net.

Freud, Sigmund. *The Uncanny*. London: Penguin Books, 2003: 213 - 234. Print.

Gianetti, Louis. *Understanding Movies* (10th Edition). Prentice Hall, 2004. Print.

Giralt, Gabriel. *Realism and Realistic Representation in the Digital Age*. Journal of Film and Video, Vol 62, No 3, 2010: 3 - 16. Print.

Gleicher, Michael. *Animation From Observation: Motion Capture and Motion Editing.* Computer Graphics 33 (4) - Applications of Computer Vision to Computer, 1999: 51 - 54. Print.

Gomery, Douglas. “Towards a New Media Economics”. Bordwell, David and Noel Carroll. *Post Theory: Reconstructing Film Theory*. Madison, Wisconsin: University of Wisconsin Press, 1996: 407 - 418. Print.

Hodgkinson, Gray *The Seduction of Realism*: New Zealand: Massey University, n.d. Print.

Hogan, Joseph. *Illuminating Realism: A Bazinian Analysis of Spike Lee's "Inside Man".* Cinesthesia: Vol. 1: Iss. 1, Article 2 December 1 2012: 1-6. Print.

Hsu, Jeremy. Innovation News Daily. *Why the creepy uncanny valley keeps us on the edge*. 4 March 2012. Web. 28 June 2013.

ImaginariumStudios. *Imaginarium Studios.* n.d. Web. 15 December 2012

Interview *Avatar Mocap Producer James Knight*. 24 January 2012. Web. 27 March 2013. www.cgchannel.com.

Itti, L., N. Dhavale and F. Pighin. *Realistic Avatar Eye and Head Animation Using a Neurobiological Model of Visual Attention. Research Report* . Los Angeles, California: Department of Computer Science, University of Southern California, n.d. Print.

Kania, Andrew. “Realism”. Livingston, Paisley and Plantinga, Carl. *The Routledge Companion to Philosophy and Film*. Oxon, UK: Routledge, 2009: 237 - 248. Print.

Lay, Stephanie. *Why We're freaked by zombies - Research illuminates the 'uncanny valley'.* Web. 19 January 2014. www3.open.as.uk.

Liverman, Matt. *The Animator's Motion Capture Guide.* Newtown Centre, MA: Charles River Media, 2004. Print.

Lohmeyer, Edwin. *From Cellulloid Realities to Binary Dreamscapes*. Masters Thesis - Art & Visual Studies. University of Kentucky, 2012. Print.

Manovich, Lev. *Assembling Reality: Myths Of Computer Graphics*. Web. 7 March 2013. www.manovich.net.

McClean, Shilo T. *Digital Storytelling.* Cambridge, Massachusetts: MIT Press, 2007. Print.

Meade, Benjamin. *Emotional Response to Computer Generated Special: Realism Revisited.* Journal of Moving Image Studies. 2002. Print.

Menache, Alberto. *Understanding Motion Capture for Computer Animation (Second Edition).* Burlington, MA: Morgan Kaufmann- Imprint of Elsevier Publishers, 2011. Print.

Mori, Masahiro. *The Uncanny Valley* (Translated by Karl F. MacDorman and Norri Kageki). IEEE Robotics & Automation Magazine June 2012: 98-100. Print.

MotionCaptureSociety. *Motion Capture Society*. n.d. Web. 13 February 2013 www.motioncapturesociety.com.

Murray, Rebecca. *Tom Hanks and Director Robert Zemekis Discuss “The Polar Express”*. Web. n.d. www.movies.about.com

Unknown. *The Curious Case Of Benjamin: Behind the Scenes*-2009. Web. 07 January2014 www.wildaboutmovies.com

Orange, B. Alan. *Joe Letteri Talks The Evolution of Gollum in The Hobbit: There and Back Again.* Web. 17 October 2013. www.movieweb.com.

Plantec, Peter. *Crossing the Great Uncanny Valley*. 19 December 2007. Web. 24 January 2014. www.awn.com.

—. *The Digital Eye: Image Metrics Attempts to Leap the Uncanny Valley*. 6 August 2008. Web. 13 February 2013. www.awn.com.

Podcast, fxguide.com. *Avatar-Weta Podcast - Joe Letteri.* Sydney. 09 January 2010. Web. 10 November 2013.

—. *Dr. Mark Sagar –* New Zealand. 18 February 2011. Web. 10 November 2013.

—. *The Hobbit with Dave Clayton* - New Zealand. 17 December 2012. Web. 10 November 2013.

Podcast, Vfxshow. *The Curious Case of Benjamin Button podcast #64*. 14 January 2009. Web. 10 October 2013. www.fxguide.com.

Pollick, Frank. *In Search of the Uncanny Valley*. Research Report. Glasgow, Scotland: University of Glasgow, 2009. Print.

Popular Mechanics : *Exclusive Interview With James Cameron.* 11 December 2009. Video. 23 May 2013.

Pramaggiore, Maria and Tom Wallis. *Film: A Critical Introduction*. London: Laurence King Publishing, 2011. Print.

Preeg, Steve. *The Curious Face Of Benjamin Button - GRID09 Conference*. Stockholm, Sweden: www.vimeo.com/m/26718845, 21 September 2009. Video. 07Jan 2013.

Prince, Stephen. *Digital Visual Effects in Cinema: The Seduction of Reality*. New Jersey, USA: Rutgers Press, 2012. Print.

—. *Movies and Meaning: An introduction to Film 6th Edition.* New Jersey, USA: Pearson, 2013. Print.

—. "Psychoanalytic Film Theory and the Problem of the Missing Spectator". Bordwell, David and Noel Carroll. *Post Theory - Reconstructing Film Studies*. Madison, Wisconsin: The University of Wisconsin Press, 1996. p71-86. Print.

—. *True Lies : Perceptual Realism, Digital Images and Film Theory*. Film Quarterly, Vol 49, No. 3, 1996: 27 - 37. Print.

Raga, Elisabeth and Stuart Sumida. *Anatomical Considerations in Facial Motion Capture.* ACM Siggraph Computer Graphics - Building Bridges- Science, the Arts and Technology Vol 43 Issue 2, 2009 May. Print.

Riley, Jenelle. *Brave New World*. 10 February 2010. Web. 12 January 2014 www.backstage.com.

Robertson, Barbara. *Facing the Future*. Volume 36, Issue 6. Web. 18 January 2014. www.cgworld.com.

—. *Motion Capture Mania* - Computer Graphics World Magazine Issue: Volume 35 Issue 3. Web. 28 June 2013.

Ross, Dalton. *Entertainment Weekly: Inside Movies.* 3 February 2012. Web. 19 October 2013.

Sagar, Mark. *Facial Performance Capture and Expressive Translation for King Kong.* Performance-Driven Facial Animation – SIGGRAPH, July 2006. Print.

Scott, Remington. *Remington Scott.com.* n.d. Web. 12 December 2013 www.remingtonscott.com..

Scot Rubin. "Demonstration of the Simulcam invented for the filming of James Cameron's Avatar." *Youtube*. Youtube, 21 October 2010. Web. 14 January 2014.

Serpell, James. *Anthropomorphism and Anthropomorphic Selection—Beyond the "Cute Response".* Society & Animals 11:1, 2003: 83 - 100. Print.

Seymour, Mike and Bill Dawes. *The Curious Case of Aging Visual Effects.* 1 January 2009. Web. 15 January 2013.

Seymour, Mike. *Art of Optical Flow*. 28 February 2006. Web.18 January 2013. www.fxguide.com.

Shaviro, Steven. *Emotion Capture Affect in Digital Film.* Projections - Berghahn Journals - Volume 1, Issue 2 Winter 2007: 63-82. Print.

Steckenfinger, Shawn and Asif Ghazanfar. *Monkey visual behavior falls into the uncanny valley*. Research Paper (Neuroscience). Princeton: Princeton University, 2009. Print.

Stranahan, Lee. "The Amazing Special Effects of Benjamin Button". *Youtube*. Youtube, 2009. Web. 14 December 2014.

Sweeney, Kevin. "Medium". Livingston, Paisley and Plantinga, Carl. *The Routledge Companion to Philosophy and Film*. Oxon, UK: Routledge, 2009: 173 - 183. Print.

Sydell, Laura. *NPR - Building The Curious Faces Of 'Benjamin Button'.* 16 February 2009. Web. 14 August 2012.

Tanenbaum, Jim. *Avatar Revisted: (One of) the Production Mixer's Cut.* 27 April 2011. Web. 14 January 2014. www.soundandpicture.com .

Tierney, Dolores. *Digital Filmmaking, Realism and the Documentary Mode in Recent Latin American Films.* Paper. Brighton: University of Sussex, UK, 2007. Print.

Tobon, Ricardo. *The Mocap Book: A Practical Guide to the Art Of Motion Capture.* Orlando: Foris Force, 2010. Print.

Ulbrich, Ed. “TedTalk - Ed Ulbrich: The Curious Case of Benjamin Button”. *Youtube*. Youtube, February 2009.Web. October 2013.

Unknown. 3D World. *LA noire- the future of facial capture.* Web. 10 September 2012.

—. KubrickCorner. Web. 5 January 2014. www.brickfilms.tripod.com.

Wolfe, Josh. *James Cameron: Blockbuster Businessman*. Forbes. Wolfe Emerging Tech Report. November 2011: 1 - 8. Web. 11 December 2013.

Appendices

Appendix 1

Date: 18 February 2011

FXGuide.com- fxpodcast: Dr. Mark Sagar

Mark Sagar **(SG),** Lead Facial Motion Capture Developer at Weta Digital - New Zealand, speaks to Mike Seymour **(MS)** detailing the FACS Solver and other aspects of his work on facial +motion capture on *Monster House* (2006), *King Kong* (2005) and *Avatar* (2009)

Mark Sagar specializes in facial motion capture, animation and rendering technologies and is currently focusing on bio-mechanical simulation of the face. Mark was a Post-Doctoral Fellow at the Massachusetts Institute of Technology and holds a Ph.D. in Engineering from The University of Auckland, New Zealand, where he worked on Virtual Reality Surgical Simulation and Anatomic Modelling with Peter Hunter's Bio-Engineering group
Source: fxguide.com

Fig. 15. Dr. Mark Sagar. Picture courtesy of Zorpette, Glenn, www.spectrum.ieee.org

Full Interview: 57mins52secs
Transcription: Beginning: 18 mins02secs - Ending: 44mins20sec

This transcript had been edited for content pertinent to the topics in this report. It is an abridged version of the podcast that can be accessed at:
www.fxguide.com/fxpodcasts/dr-mark-sagar/

Beginning of transcript 20 mins15secs

MS Monster House. Even though the film didn't come out for a year or two [after it was completed in production] is a really pivotal point in the development in facial animation because … that's the first time, - we spoke about it a moment ago it - one to one correlation between the first iteration of trying to performance capture a face and trying to track that from the markers on the face to the target face but now at Monster House you actually go to the Faces Action Coding System, the actual idea of having an intermediate kind of performance space. Can you explain that?

SG Years ago at M.I.T, I think, in the mid-90s a guy name E Esser had developed a computer vision technique for trying to calculate the FACS of video images. I'll just explain what FACS or Facial Action Coding System is. It's like a facial alphabet. It's a way of breaking down facial expressions into groups of muscle activations. It's basically the most efficient way to describe the information in a face. The stuff that was done at MIT wasn't applicable to a production environment. With LifeFX, we were experimenting with building faces, especially when we got into the internet business, we started looking at what an efficient way to stream that data. We were looking at PCA and things like that. After LifeFX, before I started at Sony, I started thinking how could we efficiently represent something editable that could work for a face, so that you could break down that motion and have something editable. Then I started going back to the FACS stuff that I became aware of at MIT. Then I started experimenting with mathematical techniques to break down that data and make it useable. The big problem with facial motion is that it is very non-liner. [Interpretation: Expressions on the face never just go straight from one expression to the other eg. 'I am happy' to 'I am Sad'. It goes through stages and arrives at sad.] I started to work with that. When I started at Sony, I suggested this as a potential approach but they had already committed to another approach for Polar Express. Anyhow, after I had worked on Spiderman, the Monster House project came up. Damien Gordon motion capture supervisor and I had discussed the FACS system with him during the year, about how they could solve a bunch of the problems they were having. Damien was happy to try it out and we did a test for Monster House, using these techniques. It was successful and Sony decided to go with this technique [for Monster House].

The thing that was useful about it is that you had very stylized cartoon faces for the characters in Monster House which was very different from the actors. So the was no easy way to take the motion of the skin of the actor because the faces are very different and they are very stylized. So the FACS was perfect in this case, as you could break down what they were doing into expressions. All you have to do is build a corresponding expression in the stylized character and as long as you can calculate what that expression is, you now have a 1 to 1 mapping. You can drive any face with the expression data. That was the key thing that worked with Monster House. You've broken the geometric connection between the motion capture data and the actual character

MS: But it's not just topology, is it? The idea of going into expression space is the interim step between the first and the second, gives you a reduction in noise, the ability to art direct. Gives you a whole lot of benefits.

SG: Yes. 'Noise' [interference] is one of the things that Damien had been dealing with on *Polar Express* because the problem is they many actors in a very large volume and so there was a lot of noise in the face. If you apply that to a face, you get all kinds of swimming motion on the actual model. So the only way to fix that is with filtering and you end up throwing out the baby with the bath water. You're killing what motion capture is giving you, which is this wonderful high frequency information. What [FACS] is doing is its increasing the signal to the noise. It's breaking down the motion capture data into its fundamental information content. If I made an analogy with music, you've recorded a symphony and now you're working out what notes are being used because you want to apply that same piece of music to a completely different orchestra.

MS: **Using that music analogy, it would seem that if you were going from one expression to another.. as a straight 'I am happy. I am now sad' then that solves a lot of problems but what if I was going from .. happy to a bit sad and a bit perplexed..? I am therefore having to go for multiple instruments in the orchestra whose notes are a bit different**

SG: Exactly. That is where the FACS system solves it, not only does it work out what expression, say in the case of happy to sad, it's actually going from Action Unit 12 plus Action Unit 6, its breaking it down into the individual muscles. So then when you go from one expression to the other it will be changing the muscles accordingly. Exactly what would be happening in your face as your brain is changing the signals that its sending to your muscles. It's really calculating the nerve signals that are sent to the muscles. That’s really what's happening. So when an actor does change emotional state, then you're measuring exactly that. You're getting that exact information content. Rather than the actor have their own face driven by these nerve signals, it’s a computer generated face. What we're trying to do is calculate those nerve signals really.

MS: **So Weta is heavy into subsurface scattering. That’s really dealing with a muscle that moves beneath the skin, independent. The initial skin movement is one thing, and when you're modelling to the level of realism that Weta has been doing, you need to have the sea of muscles underneath the skin, then this approach is perfect for that. You're not just moving the topology at the top of the model. You're actually driving the muscles. That’s exactly what you need, to *drive* the muscles, is that correct?**

SG: It is. You can do it at multiple levels. You can drive an actual muscle, which drives skin on top. Or you can drive what would this face look like if this muscle was active.. which is what we did for King Kong. So.. King Kong wasn't a muscle model, *but* it could do all the FACS expressions. And they actually, with the Gollum face, that was done by building all the FACS shapes. When I moved to Weta, they had already approached faces in a way which was amenable to the motion capture things that I was calculating. It was an easy fit. Weta, more than any other company, has a real drive for physical realism. A lot of stuff done at Weta is making creatures which exist in the real physical world. It's not stylized, like at Pixar, where you can have a face stretch in any way that you want. With Weta you want to really feel its living flesh and bone, it's been the specialty with Weta over the years.

MS: **Lets discuss that. With Kong, you had indirect eye tracking. It was a high frequency problem. You were trying to get actual subtly with eye motion trying to capture eye movement**.

SG: With Monster House we were limited to the number of markers on the face, so we couldn't get a lot of the fine detail. There were 40 or so markers and they had to be seen. They were big markers, we couldn't get a lot of the really fine detail, only big expressions could be captured. (See Fig. 8.) Whereas with Kong, we could put 150 markers [on Andy Serkis'] face and really zoom in the cameras and capture the absolute subtlety of motion. And one of the things you can do is if you have a couple of markers on the upper eyelid and lower eyelid, that is enough to give you the changing contour of the eyelid as the eye looks left and right. When the eye moves up and down, because you're also solving for which muscles are active, you can differentiate eye motion from the contraction of the eyelids due to expressions. You can actually end up *solving* for eye motion. The important thing about that is that you can choreograph with the motion in the character. If the character is blinking or looking around, those eyes are going to be in the right position. One of the hardest things to do is to get the eyes right. So we're getting that data for it but in an indirect way. For Avatar it was much easier because we had a helmet camera that was looking directly at the face so we could see the eye, we could see *exactly* where they were for *every* frame whereas for King Kong it was still a cloud of dots. That kind of thing really helped. Now because it is represented as a FACS component, if we needed to change the gaze of the eyes, say he's looking at somewhere slightly different, you could just scale that whole channel or move that whole channel [in the proprietary programme] to change where he's looking without destroying all the saccadic motion.

MS: Kong was a good candidate because you were re-targetting to a non-human face. There a lot of muscles that are similar between a human and a gorilla. I was interested to read in one of your earlier papers that Andy who was doing the performance. But what happens when a gorilla does things that we just can't do?

SG: Some of the lip motions. They have much more mobile lips. Anything that needed complex and can stretch in ways we just can't do, so that part would be key frame animation. We could capture whoops but if it was something complicated that doesn't fit into FACS framework, then we mixed in 6 key expressions a gorilla would do. We made this 'gorilla' FACS but you fill in the holes where you can with animation.

MS: Did the Na'vi have an expression that a human can't mimic on Avatar?

SG: It's pretty much one to one. Maybe they have more extreme snarls.. or their eyes would fold back in different expressions but that’s it. But James Cameron wanted the bottom half of the face to be one to one with humans, so a lot of these Na'vi characters when they do snarls, they are more extreme. They do things with their ears, in certain expressions their ears fold back. Those things are all extras. A lot of it is one to one [with the actor-Na've ratio] I think James Cameron deliberately wanted the bottom half of the face to be almost identical. A lot of the mouths and chins of the Na'vi characters are really close to the actors. But the upper part of the face is wider but it's still very close to the actors. Its just they have a wide rigid nose.

MS: **So you did an initial helmet test in 2006. Was that a Cameron initiative? Or was it from your team**?

SG: **(33mins20secs)** He'd already been playing around with these helmet cameras. He was quite determined to use them because he had direct feedback *right* from the actor. He was really looking at using these. So, Joe Letteri [Senior visual effects supervisor] said James Cameron is looking at this. What can we do for that [with the footage]…? I did a test experimenting to see if I could calculate the FACS from a single camera in front of my face and decided to drive the Gollum puppet with it. That ended up working quite well and we ended up going down that path. We knew we could do it from a single camera. One of the things is that, even though we only have 2Dimensional data that we're dealing with, as an image plane, but because the face is a constrained system. There is still enough information in there to know what is happening in 3Dimensions. So, for example, if you pucker your lips even though the lips are coming forward out of the plane, because we're actually calculating what muscles are active, then on 3dimensional model of course the lips would come out of the plane, even though the 2D data is being completely flattened. And there would have been no way to do that using traditional motion capture mapping techniques. So, breaking them down into these Action Units was essential for the approach we took on *Avatar*.

MS: **While that was a gift in terms of having a fixed camera to face relationship, the other curveball was that you had to have a real-time solution. Not a final solution but certainly some kind of real time implementation of what you'd done previously.**

SG: Well because we were using computer vision techniques to track this, Shane Kemp [Motion Capture Artist] had been working on some computer vision techniques to track the face and Luka Fascione [Rendering Research Lead] had come up with a real-time version of the Solver, we thought: we can actually do this in real time now and have a direct live system. Joe Letteri kept that under wraps then revealed when Cameron came round that lo and behold, not only is the camera working but it's also working in real time. We stayed up til 2am trying to get it working, the night before. Things have a tendency to not work until the very last minute…

When Cameron came in we had Shane Rangi [Motion Capture performer] and we had him driving a model of Andy Serkis, so when James Cameron looked through the virtual camera, he could see a virtual Andy Serkis performing and doing facial expressions even though it was Shane Rangi. I think that is one of the things that helped sell what we were doing.

MS: **Over the next few years as the film got completed there must have been more improvements and refinements and rightly so, this is the pinnacle of this kind of performance system. Were there many innovations that came through during production or did you have to have it bedded down.?**

SG: **(38mins15secs)** Oh no. Ian Nephews had written some really robust eye tracking software. It helped a lot because you get the eyes exactly right. When we first started getting the production data, it was noisy because we had the actors running and the cameras bouncing around. So dealing with this real data actually ended up being quite a big deal. Dejan Momcilovic was the Motion Capture Supervisor, he came up with some good ideas to stabilize the helmet camera and we implemented that into the real-time solver. And he also set up a [Autodesk] Maya control system driving the character going into film box, so we had this pipeline system that Luca [Fascione] engineered. Even though the body [motion capture data] was coming from Giant [Studios in California] we were doing the face. We had a way to tweak it on the fly in Maya. This was the real time system. What we decided was to do was have two packs. We had a real time system which was very rough and ready, then we had a very accurate off-line system. And because we were getting this video data and recording it, later on for the actual shot we would process *that* a sub-pixel level for the tracking. John Millers had written some software that would really help in automatically calculating where these markers were and tracking them across the frames. We had really good computer vision and tracking software that would handle a lot of the tracking very accurately and then we would get the data. Then we had a motion editing team who would go in and really tune the motion capture solver to exactly hit the performance, then that would get sent to animating and they would tweak what they had to do, then you would get the final product. We had quite a well oiled pipeline for that.

Appendix 2

Interview with Peter Busch (PB), Vice President of Business Development - Faceware Technologies®, a spin-off of Image Metrics

Place: **Image Metrics Headquarters**
(Marina Del Rey, Santa Monica, Los Angeles)
Date: 20 August 2012
Breakdown Part 1: 48 minutes (break for demonstration)
Part 2: 20 minutes

Fig. 16. Peter Busch at Image Metrics Office in Santa Monica, California (Personal photograph. 20 August 2012)

(Note: This is a faithful transcription of the interview. For the purposes of this research report, only the questions and answers pertaining to the process and research on facial capture, the *uncanny valley*, realistic integration of CG characters into film and Peter Busch's involvement with Digital Emily and Benjamin Button are recorded in this transcription. This accounts for the gaps in times and, perhaps, the interrupted flow of the interview as parts of the interview that did not involve the research being carried out in this report are omitted)

PART 1:

Busch begins by explaining the company history, where they were as a high-end service based company in Manchester, England and the limitations of that, in spite of their impressive list of clients. He explains about their successes, then their decision to shift to software and how much wider their market has become due to the accessibility and portability of being software based. New markets are now opening up, especially in the emerging economies.

[6.30mins into interview]

CL: Peter, can I ask about the Uncanny Valley. When did it become so important? When you worked on Digital Emily...?

PB: Yes *[nodding]*

(I explain that I have been to the Institute of Creative Labs at University of Southern California (USC) on 20 August 2012, met Dr Paul Debevec and enjoyed a demonstration on the Lightstage X's very impressive

facial scanning technology. The Lightstage was used to obtain the very high resolution geometry, texture maps and lighting information utilised on the digital versions of the actors in 'Avatar' and 'The Curious Case Of Benjamin Button' to name a few...)

[6.55mins]

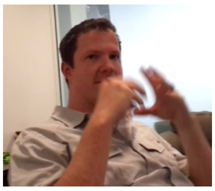
CL: ... as a production manager (at the time) what was your involvement [with Digital Emily]?

PB: The *uncanny valley* has been around for years and you can find research back in the mid to late 90s. The idea of creating a digital human that people don't know is artificial...and the closer you get to that, there's this...drop in believability and it looked freakish. That's what the *uncanny valley* is. Where you don't buy it. Really, you can only point to a few examples, even today where [realism] has actually been achieved. What's funny, is when you know it has been achieved, you won't know. There could be successful examples in film where you're not even aware of it...

CL: Can you think, off the top of your head, of any examples in the last year..?

PB: There are probably only about 10 to 15 action films that have face replacement. They had a digital double or an actor. You know, there are action shots that the actual professional talent didn't want to go through. You look at something that was much more prominent, like Tron: Legacy (2010). There were very convincing shots, even in Beowulf, but not the full body of work. That's the challenge and we're not there yet ... There are definitely better [facial motion capture] approaches.

So, how Emily came to be is, we were involved with Digital Domain on The Curious Case of Benjamin Button, but we couldn't talk about it. Non-disclosure agreement. We didn't want to, and we were excited because alot of our work to date had been in games...and this was one of our first big feature clients. We decided that we will create our own project to really showcase the quality of work that our technology and software is capable of. And we didn't have to talk about Benjamin Button because it was still a year out from being public domain.



CL: It took that long?

PB: Benjamin Button took three years to get the pipeline and the agreement on how they [*Digital Domain*] were going to do it. They went through extensive testing and...that was their pipeline and they just decided to use our tool. On Emily, the idea was that we were collaborating with Paul Debevec's team and what they were doing with [facial] scanning. You can create a digital still, a still image that is photo-real, looks exactly like a human and you cannot tell the difference, but you have to get it to *move* right, to get it to move like a

human does. That's where the combination of ICT [Institute of Creative Technologies] and Image Metrics, at the time, came together. And that's where Emily came from.
Let's recreate a human, a live human, at HD quality, photo-real to basically showcase that our technology is capable of producing animation at that fidelity. And Paul can basically highlight the creation of Emily as a character.

It was a four month project. Lots of collaborating back and forth. What shapes to scan, how do we shoot the performance? Who do we shoot? Emily O'Brien is her name. She's an established actress... she has a pretty face, it's some of the hardest things to do in recreating... Benjamin Button is very old and has lots of wrinkles, so it's really forgiving. You have a bit more freedom. With Emily, she's flawless, beautiful great complexion, an even skin tone... to recreate her was a bit more of a challenge. That's where Paul came in with the amount of fidelity and resolution when they captured [Emily] in the Lightstage...Each scan had an immense amount of geometry data, physical data and we also had the actual texture maps. Creating the character is what took four months. We have all this data of Emily so, how do we actually build a character rig? A rig that can be animated by an animator to look real...?

(I admit to still learning about the intricacies of facial rigging at that point in time)

(12.05mins into interview)

PB: The rigging process is the entire key to the uncanny valley. Not the *entire* key but one of the biggest elements. Our technology works *and* it works on any character. The idea is that it's only as good as two things. One is the asset that you're putting into it -the rig. The digital asset. The second thing is the end user -the animator. The technology is meant to embrace the animator so that they have the ability to manipulate the character in a fast way that is very easy and intuitive. There are alot of approaches that you see in the past, either with *Polar Express, Beowulf* or some of the WETA or ILM (Industrial Light and Magic) stuff, some of the high-end feature work where they over-engineer the rig. Where you have alot of different approaches as to how you create this asset. There are studies and documentation on it but ultimately, at the end of the day, what creates that final quality is the artist. So, that's where our technology is a little different than other approaches. An animator will actually use Emily to decide how Emily will look digitally...and they create a pose on the rig. When Emily smiles, how does her digital version smile? And it seems like it's just one-to-one, you just make her smile exactly how she normally smiles. There's an interpretation there because a digital representation can't do what physical Emily can do. So that was the idea behind Paul and the ICT team ... was building an asset that was as faithful to her as possible but has the ability to be pushed or augmented to create that convincing performance...and it's the subtle things, the very little things where if I showed you with something on and then something off (*Peter indicates the digital model on the computer screen in front of him*)... you wouldn't be able to tell. It's how you use the maps, eye twitches, etc. All kinds of things... it's beyond my expertise. The process was mostly spent on creating this character, and only two weeks was spent in animation. What we learnt from Digital Emily is that we didn't focus on that [as company PR].

CL: Two weeks in the whole process?

Two weeks in the entire process. And really, only a week of that was spent doing all the animation. That's how powerful our software was. That's really what we should have focused on. And the animators, they were less than three years out of school. They weren't super expensive, twenty year veterans. They were good animators and had been working with for a while, they knew how to use out tool. And we had this great asset from Paul, so in two weeks we were able to get almost all the animation complete, to the level that it is. One of the reasons I think Emily was a success is... -and this is where you see in films and other content where they fall short, where the *uncanny valley* creeps right back up- that we had full access to Emily. What we tried to recreate was something she actually said. We basically put a mask over her and the mask matched her faithfully, one-to-one. So you're watching the Emily video at the end where it reveals her, all the way back to her shin... I'm not taking away from the work that we did, but we had a very faithful reference of what we were trying to create..she *said* what we made her say. In entertainment we take that one step further and make her say something she never said.

(15.25mins)

So, out of Emily we got these requests from people having rights to likenesses and deceased actors -from all over the world- they want to bring them back to life and have them say something they never said. James Dean, Marilyn Monroe...and have them talk about iPads or something they couldn't possibly have access to. and that's one user case and everyone always looks at Emily as a crowning success...

We were successful because we created something that already existed. She was there and she said the two minutes of dialogue we had her say and we re-created that exactly. So, it's kind of a cheat, if you will. What we should have done was have someone else say something and have *that* video drive the performance. We only basically replaced this *(Busch indicates just the front of the face*) and we had so many people looking at the final

Emily frame and commenting on the hair and the cloth. "That doesn't look real" and "that doesn't look real" It *was* real! We only did the face mask, that's what the reveal shows. People are always going to find things to nitpick at. But so much of crossing that *uncanny valley* is beyond the movement of the face. We have skin shading, you have...the head position. The movement of the face is one thing. Matchmoving, how [the digital character] fits on the face. All that needs to work, you have to get all the stars to align. It's *a lot* of work.

[...] So [Emily] is a great example that we can look at but there are some recent pieces with some of our clients that are using our software that have crossed the uncanny valley.

(Regrettably, Peter cannot reveal the names of the projects at the time of the interview but admits that one project done by Californian visual effects company, Atomic Fiction was a digital de-aging of a character. I ask if it was more successful than the Jeff Bridges character in Tron, and he agreed that from what he had been shown, the animators had succeeded more than any other de-aged character he'd seen before)

(20.14mins)

CL: *Mova uses a phosphorescent spray that they spray onto the actor to use as tracking points. Phase Space (San Francisco based company) uses 6 very high resolution cameras that use the actual pores on your face as tracking points and that data is transferred to their software. What is your approach, with your software*

PB: Well, our approach is to embrace the artist. (There are) so many different approaches in trying to cross the uncanny valley, whether it's in games or film. They try to automate the process and remove the end user. Remove the artist, remove the animators and have all this done automatically... like Mova® and Depth Analysis® (*who captured for the game, LA Noire*[37]*, released 2012*) basically capturing the movement of the face... similar approach to Mova in that you capture a moving mesh. An animated texture that can be applied back to that and its exactly faithful, one-to-one, pore level detail... so, you create that instance of time. Someone is standing there and you are capturing exactly what they are doing. Again, it's like Beowulf and some of these other projects that you've seen with (attempts at) photoreal human. There were shots that were very convincing. In LA Noire, there were moments in the game that were just spot on and there are alot where it just lacked. The idea behind our process is to embrace the artist, they always are responsible

for the final quality of the content. They should be involved as early in the process as possible. If you have something like Mova or Depth Analysis, where you get data and you have to edit on top of it. You're fixing things. If the actor wasn't looking where you needed them to, that's the big thing...it's the *eyes.(Busch stresses this point by indicating eye movement*) With Mova, you can't put anything *on* the eyes. And then you have something like Depth Analysis where you are scanning the eyes, but you are only getting the eye direction of where they looked at the time they were captured. If you need to adjust that, you might...if the actor you're looking at is taller or in another position, you need to be able to edit that data. And change that. So... we've had success in basically embracing the artist in how that data is applied to the character and they can always edit. There is nothing 'locked'. It's a very intuitive workflow.

[37] Research revealed that LA Noire was not facial motion capture per say. They used motion capture for the bodies. What Busch is referring to in this segment is that the technique is essentially streamed animated textures, probably one per frame. This amounts to a 3D recording-and-playback system.

(23.10mins) We also embrace the performer, like Emily. It just works on video, that's separate from what happens (after) in 3D. Actors can do what actors do best, so they're going to deliver a more convincing performance rather than having something applied to their face or put markers, phosphorescent paint, work under intense lighting conditions...(here) they just act on camera....this is where our facial recognition comes in, because all of our maps are built off the key features on the face and the study is sampling basically all the pixels that make up the face. But you can still control that to a degree. That's why we developed our head mounted camera, so that actors' facial performance can be recorded in sync with their body performance. It's the Avatar full performance adoption. We've been doing full performance capture since 2004.

So basically it's capturing the facial performance. A key thing, again, is the eyes. So if you have 2 or 3 actors playing off each other and they are wearing a facial camera, their eye lines are in sync. Wherever they look, that's how the animation is going to be. You need to capture that performance faithfully as it needs to be represented in whatever you're doing. If you try to capture an actor and they are performing like this *(Busch sits up very stiffly in the chair)* and they can't really move, to try to have that be a convincing performance, and apply that to a body performance, it's a Frankenstein. You're applying this face to a moving body. We see this all the time. This is why you don't get good quality. Because your actor is looking dead pan, straight ahead and in the content, he's supposed to be looking around or talk to someone. That's where the performance needs to come from. So if you capture the face, body, audio and hands all together, that translates to digitally all together, you're in a better starting place in creating a photoreal human. ..so, the artists are basically free to act and perform as they will. So you deliver a better performance which comes across better in the end result. And when we involve the animators with our software, and they interpret how the performance captured on video applies to the character digitally.

(29.12mins)
On the facial animation of Benjamin Button, our team had 6 animators here and 3 animators at Digital Domain. Just for the animation. Their rendering and shading pipeline was huge, and that's another thing, but we won't get into that- but for the core animation, that's the power of our software.

(On how advancement in facial motion capture technology has improved pipelines in recent years, Peter elaborates..)

My background is in feature animation production management. My job was just to manage artists and I had to make sure that we hit the schedule and everything was delivered on time. So, I started seeing our throughput, the numbers that our team of animators was producing each week. If you're managing a project, you have to produce a 100 minutes in, say, 10 weeks, so that's 10 minutes a week. So when I was generating our production reports and seeing how much we were producing with a team of 12 animators with less than 3 years experience, I enquired if this was correct,

One of the big things about Emily is the way we showcased it, was at SIGGRAPH. Very public and it became viral. It was also around the time that Youtube® was really starting to become mainstream and people were getting used to it.. The idea of sharing links. I wasn't a part of anything like that before Emily... What had been done previously is alot of it was done internally with R&D (Research and Development) by these big studios. Disney in 2001 and 2002, some dating back to the 90s, a Zemekis project that was very highly funded, high profile. Who do you bring in? You bring in engineers, programmers, the brains to try to combat this issue *(Busch is referring to the uncanny valley)* It's the common challenge in any CG environment ... you have the technical expertise and you need the artistic expertise. They need to be able to work together hand-in-hand to produce convincing content. And if you approach it in a very technical manner, and I use the word over-engineered all the time... placing sensors around your eye sockets to see where you're looking is an inane approach whereas if you're shooting video to see where you're looking , you can use pixel tracking. There are lots of pixel tracking software out there, and it's more natural.
If you look at Polar Express , and you see all the other motion capture techniques, you see the markers on Tom Hanks' (face)...

CL: Is that 'old school'?

I wouldn't say it's old school, just a very engineered approach to producing a digital character. It's a very technical way of thinking. We need to capture -motion capture..we need to capture *exactly* what happened and then that should translate exactly into CG. Our approach is, we capture what happened, and when we translate that into CG, we involve the artist because they are going to be involved, at the end of the day, making sure that that look is polished. And that is where we've found alot of success. Besides Benjamin Button, there are more recent projects, 2 in India, that were very good.

(45.18mins)

(Busch elucidates after a shift to general conversation about games companies they are involved with in creating content -especially with the story driven game narratives driving the push to raise the bar in search of better graphics)

There are alot of things that need to come together to cross the uncanny valley. I'm very critical in challenging that...

I think what was so successful with Emily, is it was done by a no-name company. Nobody really knew who we were. It was not a big budget studio, we were a small technology company... It gives me alot of hope but there are alot of things that have to come together to cross the uncanny valley... We get projects now where people want to recreate this person and those people are *alive*. Why don't they just use the real person. It's cheaper to shoot someone, than have that person (digitally) recreated. Why would you invest this much time and money..? For something like '*EkTha Tiger'(2012)* (Indian project with Bollywood star, Salman Khan), it was stunt work and the actor didn't want to do the stunt, they had to replace his face over the stunt double. Things like that make alot of sense. Or for Benjamin Button, where Brad *wasn't* 75 years old and they needed to recreate him. That's why *that* was a success. So, you get content where they had to cross the *uncanny valley*.

(50.22mins)

(Busch ends this part of the interview explaining some of the considerations that have to be taken into account when thinking of producing a photoreal digital human)

When it comes to taking on this, there's less risk in using somebody like us, than some of these new things coming out of research. We've been around, we know our products work. We can help adopt things like scanning, adopting Lightstage data into the pipeline. Out technology is really easy to use, but..ok...what does that mean? How are you creating your characters? Where are you recording your performances? If you really want to cross this *uncanny valley* and create convincing performances, you need to do certain things. Not just use our technology..but where are you capturing your performance. Are you capturing your body , your face, your hands, your eyes all together. Yes/no? If you're not, you're not going to end up there.
Capture everything, full performance capture. Start there!
Whether you're using Phasespace® for the body, if you're using Lightstage for capturing the scanning and the actual person. Be faithful to that true performer. Then, what are you doing in the process? Use us for moving the face but what are you doing to the body? How are you moving your hands? What are you doing with the shading. All of these questions you will be asked... Now you have all this data, put it all together. Now, who are you going to use to finish it? What can your engine support? What else do you have to factor in to what you're doing? How are you rendering it for your film...? How are you shooting your plates? Thousands of questions you have to ask. Adopting our software and motion capture is a small part of that...

(Break for Demonstration)

PART 2:

PB: We'll start with performance capture, that's the first step where you're recording your performance. So this idea of FULL performance in capturing the face of these two actors (video shows faces of two actors - close up on face-cam views). Where we're capturing two actors playing off each other because they're going to naturally give a better performance, they can kind of react to what the other is doing..rather than if you capture one actor at a time and they have audio in their headphones, they're going to try to sort of react. The idea is 'theatre in the round'- basically capturing everybody ALL together and that's what you START with.

Marker data is capturing in real time, which is nice, but you had to put markers on the face. Two big things in convincing facial performance are the eyes and lip-synch. You can't put markers on the inside of the lips, you can't put markers on the eyes. Now you seeing all these marker based companies trying to develop new pixel based things that let you see the eyes and inside of the lips and whatnot.

CL: **...a combination of the two?**

Yes. Viacom® is heavily invested in the last 15 years of building motion capture cameras and then improving their facial capture to capture faces on set. Full performance.
This was shot at a company called Centroid in London. What (the actors)are saying isn't very interesting, but how they're doing it, if you just watch the eyes. ..
So, we scripted this scene very specifically to do certain things. With facial motion capture you have to currently capture in front of an array of cameras and you're limited to what you can do. With head mounted capture, your actors can go as far as the volume can. They can *look down*, which is the big thing ... because with markers when you look down, you have occlusion unless you have cameras down there capturing them..you can't see them. With this, it's just capturing video. Very simple. (It's a) very simple way to capture this. This is (*referring to video of two actors with head mounted cameras having a conversation*)...basic dialogue

CL: **What was this for?**

It was just a test, just to show full performance capture.

(*Showing video similar to video on the Faceware Technologies® website*)

One thing to study is..if you watch the video and watch what the actors are actually doing, they're just standing there, isn't it? Dialogue, back and forth, pointing a gun, but watch how much their eyes move. All these little subtle things. Head turns are the most interesting thing

(video continues...)

Yeah, they canned it and these are professional actors. Goes through a range of emotion and they go away from each other, come back. This is kind of what they came up with. The body, the audio..it's a good performance. This is not capturing hands but you could if

you wanted to. With a cyber-glove or something like that, if you wanted to take it one step further. This is also capturing what we call *tethered.* You see those cords coming out from each actor, we do wireless capture as well. we did it on this shoot... you can wirelessly transmit the video. Two options: One you can record the video on your own body, there's a tiny digital video recording device. Or you can wirelessly transmit it onto a receiver and its recorded onto disc somewhere off to the side.
(4'27 mins) That's where our pipeline is now. We shot this about two years ago. We didn't need to capture wirelessly. But it was just to demonstrate another element of complexity. The idea was to show how capturing two actors at once creates a better performance. So, we start there.

(Starts another video sample)

So... here's another sample that we show basically on how..this sample on the Olympic Games. We did a test for the Olympics. They were making a game for the Olympics and they wanted to have better facial performances of all their athletes, so we captured their facial performance with a head mounted camera. This was shot out in our parking lot but the idea was to show how stable our 'headcams' are if you are running at full speed. So this video..he's getting ready to get into the blocks...
He's actually wearing a backpack, that's what's recording the video in.. This is 2010, so it's a little bigger than (the packs) we have now but the test was to show how stable the head cams are.

(demonstrates a video of an actor running)

He's running in a straight line, but if he had to do the javelin or something, this would have been a little more interesting. So, this angle on face, at that stage, we'd probably never seen that before on camera. Ever. Of someone running -and capturing their face from a face point of view- running a 100metre dash. No one has probably done this.

CL: And this was a few years ago..?

We didn't really show this before but now it's to show that the headcam is stable without having to physically bolt it to someone's skull. Have you heard of Xsens®..? We used their suits in partnership.

End of Interview.

Appendix 3

Date: **9 January 2010**

FXGuide podcast: **Avatar/Weta Interview with Joe Letteri**

Joe Letteri **(JL),** senior Visual Effects Supervisor for Weta Digital - New Zealand, speaks to Mike Seymour **(MS)** detailing their facial capture progression. This includes: the decision to use a single head-mounted camera, the evolution of their motion capture pipeline, the FACS-muscle solver and other aspects of their work on *Avatar* (2009)

Full Interview: 44mins54sec
Transcription: 10mins40sec - 19mins15sec

This transcript had been edited for content pertinent to the topics in this report. As such, it is an abridged version of the podcast. The full podcast can be found at: http://www.fxguide.com/modules/fxpodcast/files/fxg-100108-avatarweta.mp3 or www.fxguide.com/fxpodcasts/Avatar_Weta_Digital/

(Beginning of transcription: 10mins40sec)

MS: **In those early tests [of facial capture], did you have the helmet [camera]or was that not part of it yet?**

JL: Yes, [the camera] was on the helmet. We did our *very* early test with markered [sensors]. We used markered [sensors] for Kong. The thing about markered [sensors] is that you're limited in the amount of space and performance etc but the data is absolutely accurate in 3D space. So you have 150 markers on Andy Serkis' face for Kong, you have dead-on accurate data but you are limited to a small volume. When we decided to go with the single camera, you have to still recreate that 3-Dimensional (3D) data, but you only have one 2-Dimensional (2D) perspective. So, we had another problem to solve there, which is: How do we know from the 2D data that we have accurate data in 3D? That was another step there.

MS: **Why didn't you get two cameras with some kind of stereo split on the head rig. Not because you wanted to capture stereo because the film is stereo, but because it would give you the triangulation you would need for 3-Dimensional data space**

JL: Yes, we thought of that. We’re considering putting three cameras and even four cameras. However the more cameras you put on, the more heavy and unbalanced the rig becomes. It becomes more cumbersome for the actors because they have to carry two data recorders that you now have to keep synchronized to get the data through. It becomes a much more complex problem from a production point of view. Even just keeping two cameras registered is a real problem. So we decided it would be easier to solve the problem from a single camera. A single camera was easier for everyone to work with and in [Cameron's] mind that was he wanted. He thought it would be the ideal system - with just the single camera. So we lived with the limitations and figured out a way to solve the 3D out of the one camera.

MS: **So you have ability to scan an actor's face and not withstanding performance because the actual shape of someone's face or head structure doesn't change, I'm wondering at this stage when you're still capturing the information off them, is the fidelity to get to the musculature really all the way there? It seems to me that it's a problem of not just working out what the points are moving in 3D space, but you said yourself, from Kong it's all about working out what the muscles were doing to provide that movement that happens in 3D space.**

JL: (agreeing) Is it an automatic process? No. It's not there yet, it's not a push-button process. We used as much tracking-and-solving technology as we could to help us along but in the end, it's actually a team of artists looking at this data, trying to figure it out and make the call on the interpretation. But we built the system, so it is basically a learning-based system. If we saw a new shape that the system hadn't encountered before, we could input it to the learning data base and add that to our solve. The next time it encountered that shape, it would be able to do that. You have to remember that when you're looking at a camera based system, you don't see just the shape of the face. You also see all the lighting changes that are happening on set. The shadows are moving around. So the same expression may look different if the [actor] had someone obstructing them and blocking the light versus someone not blocking the light. The camera doesn't know that it is the same expression. We spent a lot of time training it.

MS: **…you say "an" [facial] expression? But if someone like Sigourney [Weaver] is acting a scene, there's more than one expression across her face - and they overlap.**

JL: They do overlap. That's right. They do overlap but what we're able to do is, frame-by-frame, analyze the position of those markers and use that to derive the main muscle groups and understand how they're working together. So you get the overlap by doing that frame-to-frame. But, you're right. The expressions overlap. So sometimes, to a tracking-and-solving system like that, it's not going to know the difference between some of these intermediate poses. And again, that's where we rely on the artist to build up a repertoire of what these things actually mean.

So the whole solving system is really like an expansible animation system. In fact, its built to be flexible enough so you can go from the solve straight into key-frame animation if you need to edit anything. And that is all based on going back to what we did for Gollum and [King] Kong. The idea that we used a FACS system to solve for the muscles.

(Mike Seymour agrees that Leterri should describe the FACS system and how it was used on *Avatar*)

FACS is a facial action coding system that was developed by Paul Eckman. It does exactly what it says - coding expressions in the face. So when you have a particular expression, you can give weights to certain muscles and certain muscle groups in the face and you can describe how they are activated. To what strength and in what combination to create the expression that you see. What we're doing when we solve, is that we are looking at all the points on the face and we are trying to derive from those points, what are the muscles that must be engaged here - and to what degree. And we use that [information] to drive the muscles on the character. That really *is* the key to the performance. What we are trying to do is try to understand what the emotion is by reading what muscles are engaged.
So you're not taking the geometric data from one face and targeting it to another face. You're actually solving for the underlying muscle activations and then activating those same muscles to the same degree on the new character

MS: In the case of Kong, you said in an interview that they were the same fundamental muscles but not exactly the same as in a human. Does the digital actor in Avatar have the same musculature in Avatar and the same relationship with muscles as a human face or is it slightly different as it was in the case of Kong

JL: We treated it the same way. We used the same muscles you would have the human face. Obviously the layout and the attachment points are slightly different because of the jaw is more elongated, the nose is wider. We wanted to get some of the characteristics of for example, what a lion would do with its nose. We made the activations different. Some of the muscles could pull more strongly in certain directions than it would on a human, but we used the same basic human layout to drive everything.

MS: Because there was that cat influence...

JL: Yes absolutely. The Na'vi have that [feline] influence. Lion nose... the yellow eyes were *really* inspired by lions. They have a cat-like ears and tail.

MS: One last question on the camera. Did you at any time actually map the feed you were getting from the helmet onto the face?

JL: That's a good point. We didn't do that in real-time but that was only for editorial. The camera feed was mapped directly onto what we called the 'kubuki mask' and that was always done so that [Cameron] could see the actor's performances and that you could always go back to look at the reference tracking and you saw that it was the same thing in sync. For the most part, that's actually what we relied on for the mapped video. The times when we used the real-time data, what really helped out was when you had big action movements, for instance the character would growl or something really big where you saw the jaw movement or could see it at a distance. Because with a mapped video you don't get that. That changes the profile of the face and you could read that pretty clearly. So Cameron would pick which way he wanted to go. We always had the real time on [the motion capture] stage, and he would pick whether or not he wanted to use that or some of the shots for when he was building his templates.

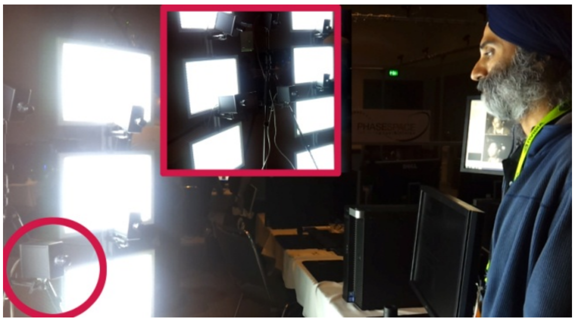
(End of segment: 19mins15sec)

Appendix 4:

Demonstration and Interview with Kan Anant, Product Manager/ Systems Engineer at PhaseSpace Inc (San Francisco, California)

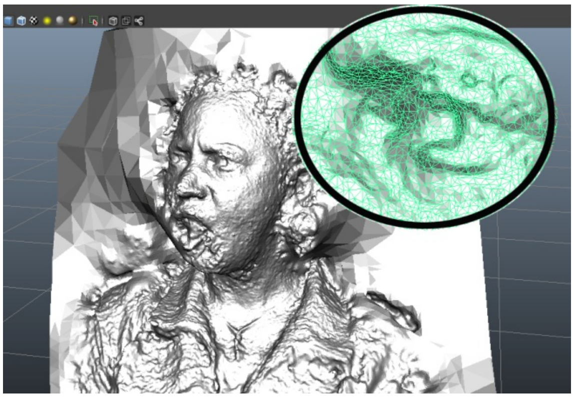
Place: **SIGGRAPH2012 - Los Angeles Convention Centre** (Los Angeles)

Date: 20 August 2012



a) Kan Anant demonstrating the optimum position to stand before six high resolution cameras of the Phasespace Motion Capture System.

b) Images from the six cameras



c) Resultant 3D model after the images are processed through photogrammetry software to produce a high resolution model.