Supervised Hybrid Expression Control Framework

for a Lifelike Affective Avatar

BY

SANGYOON LEE
B.S., Yonsei University, Republic of Korea, 1997
M.E., Yonsei University, Republic of Korea, 1999
M.F.A., University of Illinois at Chicago, Chicago, 2006

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Computer Science
in the Graduate College of the
University of Illinois at Chicago, 2013

Chicago, Illinois

Defense Committee:

Andrew E. Johnson, Chair and Advisor
Jason Leigh
Barbara Di Eugenio
Luc Renambot
Steve Jones, Communication

ACKNOWLEDGEMENTS

First and foremost, I would like to express my appreciation to my great advisor, Andrew Johnson, who trained me from a MFA art student to a Computer Science researcher. During my stay at EVL, he has put tremendous amount of effort to guide me through each step of my Ph.D. program. Always, he gives me lots of helpful advices from working on a simple project or programming code to large-scaled research problems. In real life, he and his wife, Julieta, have been a wonderful family to me and to my wife. Andy, without your guidance and trust, this dissertation would be impossible. Thank you so much!

I wish to express “thank you” for my committee, Jason Leigh, Barbara Di Eugenio, Luc Renambot, and Steve Jones for their support and advice on my research. I owe them my heartfelt appreciation. Especially, I thank Jason for his careful encouragement and helpful comments on my future work as well as this dissertation. He has taught me a lot of things about doing research and so much beyond that. Luc has been such a best friend in my stay at Chicago and always gave me enough help on my research as well as shared inspiring ideas with me. Also, I would like to appreciate Steve and Barbara for valuable discussion from the beginning to the end of this dissertation work.

I am grateful for the technical staffs and research staffs in EVL, Maxine Brown, Alan Verlo, Jonas Talandis, Dana Plepys, and Lance Long. Also, no words can express my appreciation to students who attended in my user study for this dissertation and then shared their enough knowledge and ideas with me. To all my friends at EVL, thank you for making my professional education so much more than that. EVL is the first place where I started my studying in Unites States and is the one place where I spent the most time in my entire life. EVL meant so much to

ACKNOWLEDGEMENTS (CONTINUED)

me emotionally as well as intellectually. I would like to cite Maxine's saying, *"No one leaves EVL."* EVL is engraved in my deepest heart and I will proudly live with it.

I am thankful for people outside EVL for various help on my research. Special thanks to people at University of Central Florida, Avelino Gonzalez, and Ronald F. DeMara. While working with them, the project Lifelike had a big challenge step and I could capture the idea for this dissertation.

Last, but not least, my deepest appreciation goes to my family. Without them, this dissertation would have much less value. I would like to thank my parents for their infinite love and unconditional support. To my brother, Sangheon, and his family, I couldn't have finished this Ph.D. without our parent's care and understanding. Finally, my deepest gratitude goes to my wife Jungmin, who provided the emotional and social support I needed to get through this. Her love and encouragement allowed me to finish this long journey. Not only did she keep me sane during this time, but she played a substantial role in this work. I also would like to thank to my parents-in-law, sister-in-law, and brother-in-law for their support and countless trust that make this dissertation possible.

This dissertation is based on work supported in part by the National Science Foundation (NSF), awards CNS-0703916 (LifeLike), CNS-0420477 (LambdaVision), OCI-0225642 (OptIPuter), and OCI-0943559 (SAGE). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the funding agencies.

SL

TABLE OF CONTENTS

TABLE OF CONTENTS (CONTINUED)

TABLE OF CONTENTS (CONTINUED)

LIST OF FIGURES

FIGURE **PAGE**

LIST OF FIGURES (CONTINUED)

LIST OF FIGURES (CONTINUED)

LIST OF FIGURES (CONTINUED)

LIST OF FIGURES (CONTINUED)

LIST OF FIGURES (CONTINUED)

LIST OF TABLES

LIST OF ABBREVIATIONS

API	Application Program Interface
AU	Action Unit
CRF	Conditional Random Field
ECA	Embodied Conversational Agent
FACS	Facial Action Coding System
FFM	big Five Factor Model
FPS	Frames Per Second
GPU	Graphics Processing Unit
GUI	Graphical User Interface
HD	High-definition
LCD	Liquid Crystal Display
LRAF	Lifelike Responsive Avatar Framework
NSF	National Science Foundation
NVB	Nonverbal Behavior
PAD	Pleasure-Dominance-Arousal
SAGE	Scalable Adaptive Graphics Environment
SHECF	Supervised Hybrid Expression Control Framework
VH	Virtual Human
VP	Virtual Patient

SUMMARY

The use of an avatar-enabled applications have been rapidly growing over the last decade as they promise more natural computer interaction with advanced technologies in various domains. Furthermore, recent research efforts towards natural and affective avatar capabilities have become more prevalent in the field. However, developing such an application still remains a very difficult and time-consuming task. This is mainly because a believable avatar model intuitively aims to mimic a real human, including realistic appearance and a wide spectrum of complex behaviors. Even though we have to approach this problem as a whole, most previous studies have focused on only a small part of the model due to its complexity.

Recently, state-of-the-art research in computer graphics has presented realistic renderings of the human face with a programmable commodity graphics processing unit. However, these results are mostly focused on a static model or non-interactive character animations, whereas avatar researchers tend to use very limited visualization capabilities. Merging these two research efforts will provide a better chance to surpass Mori's Uncanny Valley and achieve a more natural experience interacting with an avatar. Another issue found in the literature is the use of two distinct methodologies in modeling human behavior: rule-based model and data-driven model. The model of the rule-based system offers highly coherent behavior based on psychological theories, but it lacks in subconscious or unconscious behavior. The data-driven method relies on a large amount of rich data to extract the uncertain nature of human beings to mimic it on an avatar model, however it lacks the depth of knowledge required to understand progressive causalities in our face-to-face communications.

This thesis presents a high quality visualization method and a behavior-modeling framework that can enhance the user experience with an autonomous avatar and eventually achieve the goal of an avatar as a natural lifelike computer interface. A hybrid behavior modeling technique sets the middle ground to combine both rule-based and data-driven models. Highly realistic avatar visuals with an emotionally expressive behavior model offer better congruency and naturalness at the same time to increase avatar believability. It promotes simple design and development for an affective avatar-enabled application to overcome current limitations. This thesis contributes to the avatar research domain in computer science by promoting the synergy between widely available technologies to create a natural and believable avatar control framework. A user study is conducted to evaluate the perceived naturalness of the framework's behavior model within an autonomous expressive storytelling context. As we establish a better tractable model for an avatar as a more natural alternative computer interface, it will broaden the possibilities of our computational needs where we suffer from limited resources.

1. INTRODUCTION

An avatar is typically a virtual embodiment of a real or fictional entity used to participate in a computer generated virtual environment. Avatars have been used since the early 1990s to provide representational embodiment in virtual reality (VR) applications including military training environments, scientific visualization, games, and applications in education and art. The concept of an avatar has often appeared in movies, games, and on television even before realtime computer graphics technology became capable of realizing an avatar in virtual environments. The term “avatar” for the on-screen representation of the user was coined in 1985 by Lucasfilm’s online role-playing game, Habitat. Individual gamers had a third-person perspective of themselves, avatars. The iconic debut of an avatar in public media was “Max Headroom”, a British artificial intelligence in 1984. Although “Max Headroom” was conceptually designed as an avatar, it was portrayed by a human actor because the computer technology at that time was not mature enough to create it virtually. During 1990s, many instances of computer-generated avatars were used in VR application in a primitive form. After this early development in limited venues, as VR systems were expensive research-oriented resources, more popular and accessible avatars started to emerge in multi-player online environments such as World of Warcraft (2004) and Second Life (2003).

Recently there has been a resurgence of interest in avatar research driven by the availability of low-cost computing and the popularity of computer and video games. The term ‘avatar’ is often used interchangeably with Virtual Human (VH) or Embodied Conversational Agent (ECA) in the literature. A VH or ECA is a self-contained computerized human with artificial intelligence as opposed to the classical definitely of an avatar as a visual representation of a user/gamer. We envision that an avatar is a superior natural computer interface when compared to other computer interfaces. As we establish a better tractable model for an avatar as a more natural alternative computer interface, it will broaden the possibilities of our computation needs where we suffer from limited resources. For example, people with autism may use avatar-mediated intervention to learn how to read and reproduce facial expressions on daily basis instead of visiting a clinic only once a week. Virtual tour guides can provide personal help to museum patrons, eliminating the need to wait in line for help from museum staff at information kiosks. Students can spend unlimited time with a virtual tutor who expresses excitement over his or her improvements.

Avatar research on this thesis started with a National Science Foundation (NSF) funded collaborative research project, *Towards LifeLike Computer Interfaces that Learn,* in 2007. Researchers from the Electronic Visualization Laboratory (EVL) and a team from the University of Central Florida joined to develop a knowledge-based dialog system to implement a lifelike avatar framework. The first prototype system was desgined to capture a retired NSF program manager, Dr. Alexander Schwarzkopf, in both his knowledge and visual representation so that we can preserve his expertise via his digital clone (Figure 1). Throughout the project, the work focused on how we can create the most realistic avatar visuals and how the avatar can accurately reproduce his facial expressions. With the development of the LifeLike Responsive Avatar Framework (LRAF) and its method [Lee '10c], we successfully implemented several applications to prove our approach.

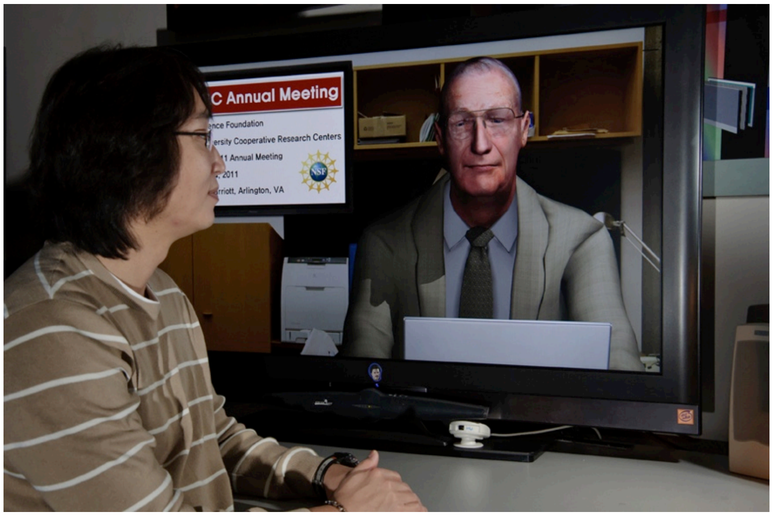


Figure 1. The first prototype application of LifeLike Responsive Avatar Framework; user can interact with Alex's avatar via microphone to ask questions about Industry & University Cooperative Research Center (I/UCRC) program.

The use of an avatar-enabled application has been rapidly growing over the last decade as it promises more natural computer interaction with advanced technologies in various domains [Yee '07] such as training, education, simulation, design, entertainment, and so on. For example, a virtual guide is often used in museum exhibitions [Bickmore '08b; Gebhard '09; Jan '09; Swartout '10; Yuan '05] and in the medical domain including healthcare educational avatars [Dickerson '05; Kenny '07; Rossen '10], counseling [Bickmore '08a], virtual therapist / nurse [Bickmore '09; Pontier '08], and tele-home healthcare [Lisetti '03]. In education, scientists developed software to treat children with autism [Konstantinidis '09], to teach emotional recognition [Finkelstein '09], and to provide general interactive tutors [Woolf '07]. Sometimes the use of an interactive avatar in the commercial domain is found such as salesclerk [Mumme '09] and sport game announcers [Strauss '08]. Interactive storytelling [Bee '10], sign language visualization [Pan '11] and representation of affect via instant messaging in virtual environment [Neviarouskaya '10] are also promising applications. Avatar-mediated technologies are not limited to those examples but are widely applicable to where its usage can enhance our experience with more natural and personalized interactions. It can also broaden our limited availability of resources that is critical to serve people.

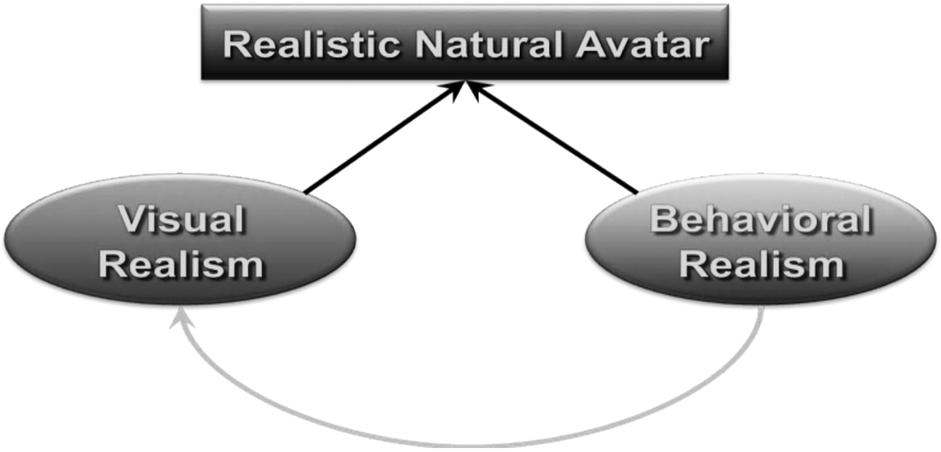


Figure 2. What makes an avatar more natural is twofold; visual realism and behavioral realism. Visual appearance often determines our first impression for an avatar and behavioral naturalness enhances our experience in longer interaction with an avatar. Behavior also often directs avatar visualization such as facial expression.

The presented Supervised Hybrid Expression Control Framework (SHECF) combines highly realistic avatar visuals with an emotionally expressive behavior model to achieve a more natural and lifelike avatar experience. Recent research efforts towards natural and affective avatar capabilities have become more prevalent in the field [Konstantinidis '09; Lee '10a; Smith '10; Zhang '10]. Researchers confirmed that a naturally behaving avatar being capable of interaction becomes more natural and effective. Burgoon and Hoobler pointed that more than 50% of social meaning in our communication is carried via nonverbal cues [Burgoon '02], for instance, natural eye-gaze is important in storytelling [Bee '10], and people refer to emotional expressions on decision making [de Melo '10]. However, developing such an application still remains a very hard and time-consuming task. This is because a believable avatar model intuitively aims to mimic a real human, including realistic appearance and wide spectrum of our complex behavior [Vinayagamoorthy '05] (Figure 2). Even though we have to approach this problem as a whole, most studies have focused on only a small part of the model due to its complexity. SHECF in this dissertation supports both high fidelity avatar visualization and a natural behavior model to overcome current limitations.

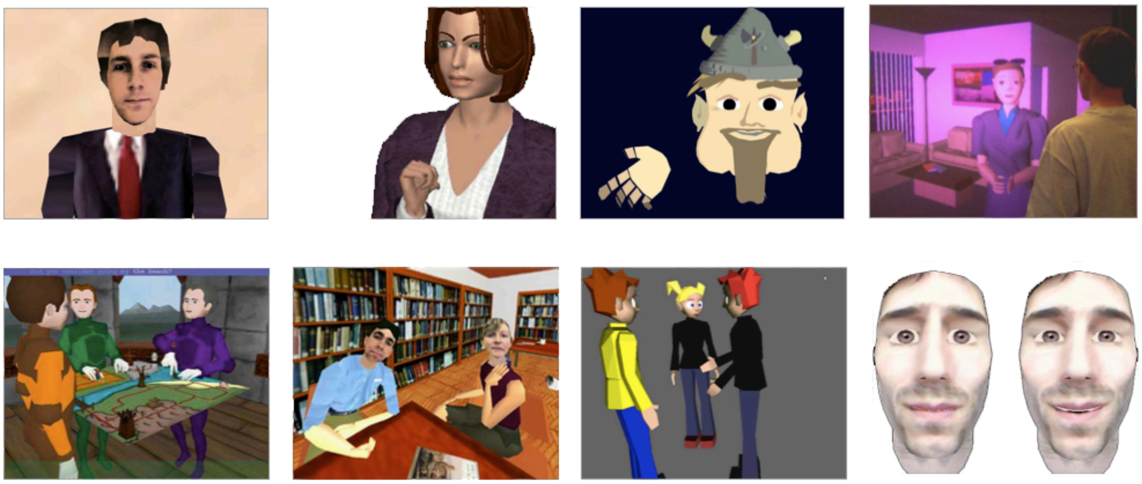


Figure 3. Avatars used in many previous studies are very limited in their visualization capabilities. Some of them even look like a cartoon or very flat illustrations (shown works were selected from 1999~2004 [Vinayagamoorthy '06]).

The LRAF visualization method in this thesis offers a low-burst efficient avatar design and development process to help researchers and developers create a visually compelling avatar application within a reasonable timeframe. State-of-art research in recent computer graphics presents realistic renderings of the human face with a programmable commodity Graphics Processing Unit (GPU) [Alexander '10; Beeler '10; Beeler '11; Jimenez '10]. However, these results are mostly focused on a static model or non-interactive character animations, whereas avatar researchers tend to use very limited visualization capabilities as shown in Figure 3. Research efforts to adopt advanced graphics techniques in avatar research will provide a better chance to surpass Mori’s Uncanny Valley [Mori '70] and achieve a more natural experience with an avatar.

SHECF presented in this thesis establishes a middle ground between the two most popular behavior modeling methodologies and offers better congruency and naturalness, which increases avatar believability. The challenge found in the literature is the separation of two distinct methodologies of understanding human behavior; a rule-based model and a data-driven model. The model of a rule-based system offers highly coherent behavior based on psychological theories such as the Five Factor Traits model [McCrae '92] and the Pleasure-Arousal-Dominance space model [Mehrabian '96b]. However, it lacks subconscious or unconscious gestural behavior. It is clear that all of our behaviors do not have a direct connection to certain rules. The latter method relies on a large amount of rich data to extract the uncertain nature of human characteristics using machine learning techniques such as the Hidden Markov Model (HMM) [Rabiner '89] and the Conditional Random Field (CRF) [Lafferty '01] (i.e. [Bailly '10; Levine '10]). However, it suffers from a limited depth of knowledge required to understand progressive causalities in our face-to-face communications. Common belief in a recently popular data-driven approach is that a personalized model can be built upon that specific person’s behavior, and advanced available technologies today make it easy and less distracting to capture such data [Picard '10]. The framework in this thesis combines both methods to support an avatar with more natural behavior

1.1 Summary of Contributions

This thesis presents a novel avatar visualization and behavior-modeling framework that achieves better visual quality and natural perception for an interactive autonomous avatar. It is developed to support an easy design production pipeline with advanced graphics techniques; a hybrid behavior modeling method; and emotion control framework to realize dynamic facial expressions.

A high quality avatar design and visualization method is presented in this thesis. The design process describes a comprehensive set of modeling details for interactive avatar development. The visualization method presents iterative enhancement with advanced graphics techniques that are highly re-usable. This design and visualization method demonstrates better visual qualities compared to many recent avatar visualization examples and shows its reusability and efficiency through different types of avatar-enabled applications developed with this method.

A methodology to model more believable avatar nonverbal behavior (NVB) is presented in three parts. First, the framework extracts and analyzes emotion eliciting affect values from avatar speech utterances via shallow speech parsing and affect retrieval from the dictionary. Second, a PAD space modeling method is used to compute the avatar emotional state. Third, the calculated emotional state is converted to control avatar facial expression and head gesture behavior. The hybrid head behavior model combines both the rule-based system and the data-driven system to achieve better congruency and naturalness. The avatar's portrayed emotional state is particularly designed to augment avatar NVB to enable more dynamic and diverse avatar behavior.

Finally, the avatar behavior-modeling framework is evaluated in a user study designed to simulate an autonomous storyteller for two emotional stories. The results demonstrate that the perceived naturalness of avatar behavior by the presented framework is significantly higher than the monotonous fixed control model.

1.2 Document Structure

Chapter 2 gives previous literature in avatar visualization methodologies followed by behavior modeling models with respect to avatar emotion and head gesture. Chapter 3 describes background of the presented framework in an aspect of the overall LRAF system architecture and its deployed applications. An overview of the realistic visualization method used in this dissertation is presented in Chapter 4. The details of rendering techniques, facial expression framework, and animation model follow. In Chapter 5 the avatar nonverbal behavior evaluation and modeling framework are presented. The evaluation of the avatar behavior model is given in Chapter 6. Finally, the summary of the framework and future research direction is presented in Chapters 7 and 8.

2. RELATED WORK

The avatar research area deals with wide range of aspects to produce natural behavior similar to what real humans show, and the research topics are very broad. This chapter primarily focuses on two of most important aspects: avatar visual realism and its behavioral realism. The reason behind this discussion is the fact that most of the research is separated in this regard and a natural believable avatar cannot be realized without the balance of such efforts. In Chapter 2.1, the relationship between visual and behavioral realism is presented. Chapter 2.2 describes the general trend in avatar visualization examples and their limitations in current literature. The behavioral modeling part in Chapter 2.3 will be divided into subsections to cover two dominant approaches in the field. First, a rule-based behavior modeling method is introduced in Chapter 2.3.1 followed by the recently emerging data-driven method in Chapter 2.3.2.

2.1 <u>Visual Realism vs. Behavior Realism</u>

The importance of a balanced development in visual realism and behavioral realism to achieve a more natural avatar is not only confirmed by our intuition but by previous research. Vinayagamoorthy et al analyzed prior avatar studies to examine the relationship between high visual fidelity and behavioral authenticity with respect to the effectiveness of virtual characters [Vinayagamoorthy '05]. They summarized results in three main parts:

> *(1) Increasing either the visual realism or the behavioral complexity of characters is not sufficient to enhance user responses*
> *(2) It is important to maintain consistent levels of behavioral fidelity with increasing levels of visual realism. There is empirical evidence to suggest that increasing behavioral realism for characters of low visual quality might lead to a degradation of user responses*

> *(3) The behavioral expressivity of a character must be modeled in correlation to the context within which the character is placed* [Vinayagamoorthy '05].

The first issue reported in the summary suggests that both visual and behavioral realism should come together to achieve our ultimate goal of improving users' acceptance of an avatar and efficacy of an avatar in achieving its aim. The second challenge presented by the summary reinforces the argument that the role of realism in the visual representation of an avatar is as significant as realism in its behavior. The evidence used in their analysis was a comparison study between eye-gaze behavior models with varying degrees of visual realism conducted by [Garau '03]. Garau et al found that inferred eye simulations, a more realistic behavior model with a higher quality of visual representation, improves a users' response to the avatar because the communication is perceived as being closer to real face-to-face interaction. This student also presented the significant finding that lower quality visuals with the random eye-gaze model performed significantly better than mixed conditions, such as higher visual realism with the random gaze model or lower visual realism with inferred gaze model. Although their study is outdated, due to limited visual capability (Figure 4) from the current graphics' point of view, its findings are still sound and intact with respect to a well-known Mori's Uncanny Valley [Mori '70].

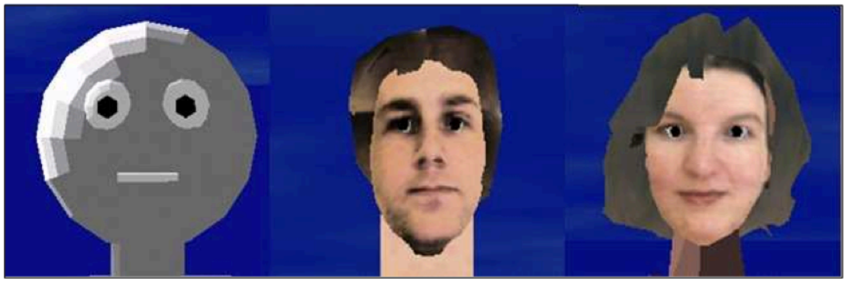


Figure 4. Virtual character models used in the eye-gaze behavior study [Garau '03]. Left one is genderless lower visual realism model and two higher visual realism models are shown in center and right.

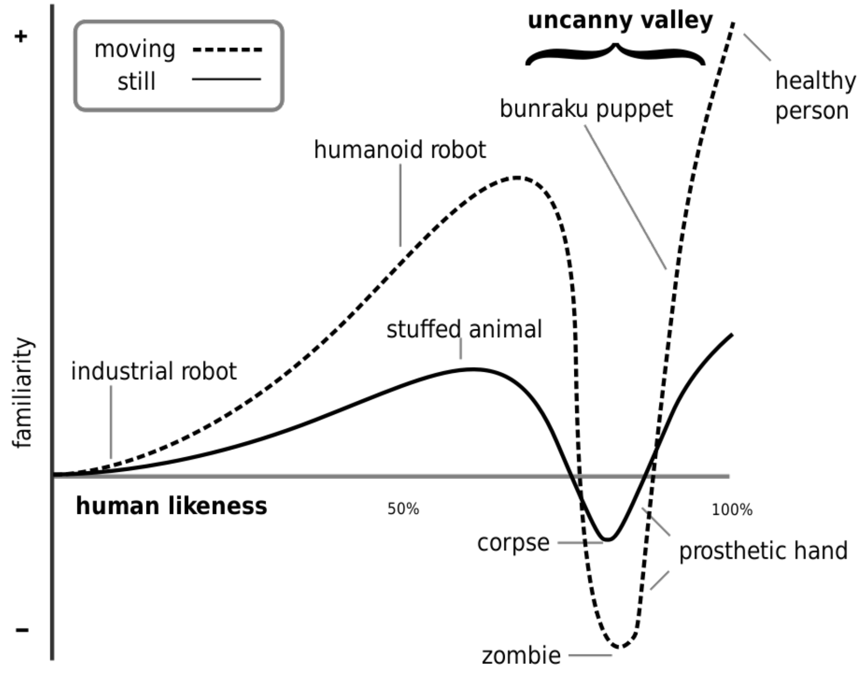


Figure 5. Mori's Uncanny Valley. X-axis human likeness refers to how the appearance of a subject is close to real human and y-axis familiarity defines how familiar it is to human. (Image source: Wikipedia “Uncanny Valley” page)

The Uncanny Valley, *a valley of familiarity*, is a hypothesis presented by Masahiro Mori in 1970. His familiarity graph is illustrated in Figure 5. This descriptive graph was originally proposed to compare a robot or an android to a real human being. For example, an industrial robot in the graph has very low resemblance in appearance (likeness) and it is not familiar to us whereas a healthy person has 100% likeness and the highest familiarity. Between those two end points, various types of robots or artificial objects in the form of a human are located based on their characteristics.

The Uncanny Valley in the graph occurs largely due to our sensation of strangeness. For example, a well-designed prosthetic hand that has a very close resemblance to a real hand presents very high visual likeness. However, when we shake the prosthetic hand, we immediately notice its fakeness from lack of soft tissues, cold temperature, and so on, which suddenly surprises us. This strangeness causes negative familiarity. An object that moves, but fails to move naturally seems even stranger, because motion itself builds an expectation of naturalness which when unmet leads to the perception of strangeness. In other words, our expectation of a given object primarily comes from visual appearance, or how realistic it looks at first glance. Then, factors from our other sensory or perceptual cues determine the overall impression. Mori suggested a few design guidelines in his original manuscript:

> *(1) Designers take the first peak as the goal in building robots rather than the second. Although the second peak is higher, there is a far greater risk of falling into the uncanny valley;*
> *(2) Designers may consider a non-human-like design to produce a safe familiarity. A good example is glasses. Glasses do not resemble the real eyeball, but this design is adequate and can make the eyes more charming. So we should follow this principle when we design prosthetic eyes. (*[Mori '70]. Translated by translated by Karl F. MacDorman and Takashi Minato)

However, since a natural avatar interface is highly sought after, and current research in advanced graphics offers us a means to create a visually compelling avatar, we cannot keep designing avatars in non-human-like form. An improved behavior model will help avoid falling back into Uncanny Valley. In other words, we need to endeavor to develop a more natural avatar behavior model to resolve *conflicting perceptual queues*.

2.2 Avatar Visual Realism

Computer graphics research has a long history of realizing natural phenomena in computerized virtual environments. Even though scientists discovered a precise model to explain natural phenomena, it was not trivial to realize in a reasonable amount of time until recently. However, as enormous amounts of computing power are available on Graphics Processing Unit

(GPU) today, and one can create realistic visualizations even in real-time. Human related graphics is not an exception. For example, Beeler et al proposed a method to capture a high fidelity geometry model for a human head with a single set of images from multiple conventional DSLR cameras or a one stereo camera [Beeler '10]. Jimenez et al developed realistic translucent surface rendering model for facial flesh using sub-surface-scattering algorithms [Jimenez '10], and the Emily project successfully recreated a human actor using virtual face [Alexander '10]. Nevertheless, such achievements are designed for static mesh construction, rendering or offline animation, whereas avatar research requires real-time interactive animation with high quality renderings.

Realistic rendering techniques in computer graphics are not easily adapted to avatar related research since it requires a fair amount of work by itself. As a consequence, most of the avatar visuals found in prior studies show very limited capabilities in this regard. Figure 6 presents avatars used in the most recent studies from 2010 to 2011. More or less, their quality of renderings is primitive. Some of them used 2D cartoon like characters and others employed a flat shading model without common rendering features such as shadows. The examples of avatar visualization in Figure 6 is summarized as following:

(a) De Melo et al studied how an avatar’s display of emotions affects how people make decisions in a negotiation, such as negotiating a price between a seller subject and a buyer avatar. They found that the difference between offer and counteroffer is lower in case where the avatar expressed joy than anger. Their avatar visualization featured with blushing and wrinkles [de Melo '12]

(b) Bickmore et al studied reusability of their previously developed health counseling dialogue management framework by applying it to diet promotion cases and reported 98%

reuse of the abstract model. The same visualization method was found in their prior work [Bickmore '05] in the context of relational agents for healthcare domain that improve user engagement by using empathy, social dialog, and other relational behavior. [Bickmore '11b]

(c) Niewiadomski et al conducted a user study to measure the effect of smiling as a listener's behavior. A subject asked to read a comic story then tell it to an agent. An operator controlled the agent behavior. Results suggested that smiling while listening is more positive than no smiling but found no significance between random smile and smile synchronized with a speaker's smile. [Niewiadomski '10]

(d) This avatar model was used to collect SEMAINE video corpus database: dialogues between an operator and a subject. The avatar was a visual representation of the operator. The operator directly led all conversations including voice. Substantial amounts of annotations were conducted to build a rich corpus data for future research. [McKeown '11]

(e) Neviarouskaya et al developed a web-based instant message system that supports automatic affect analysis for a message and shows visual representation of processed affect. A user study results indicated that their automatic affect analysis method is comparable to manual evaluation (a user selects proper affect). [Neviarouskaya '10]

(f) Hoque and Picard conducted a user study to measure difference between acted vs. natural emotional expression for frustration and delight. 2D image of avatar was used to ask a subject to tell a frustrating and delight experience. For natural emotion a subject were asked very difficult questions or showed a funny video to invoke desired natural expression. The use of an avatar in this study is rather limited without any interactive features. [Hoque '11]

(g) Poppe el al used the avatar to evaluate a listener's behavior modeling strategy using nod, blink, and short vocalizations (e.g. "uhhuh"). They replaced one participant in a dialog video corpus with the avatar animation. Results suggested that higher frequency of gesture and timing copied from a real listener was perceived better. [Poppe '10]

(h) Čereković and Pandžić developed RealActor avatar framework which supports various avatar behaviors including speech, motion, facial expression, and other nonverbal behavior. RealActor is highly capable system but relies on external Behavior Markup Language (BML) [Kopp '06; Vilhjálmsson '07] that carries all necessary behavior details to visualize an avatar. [Čereković '11]

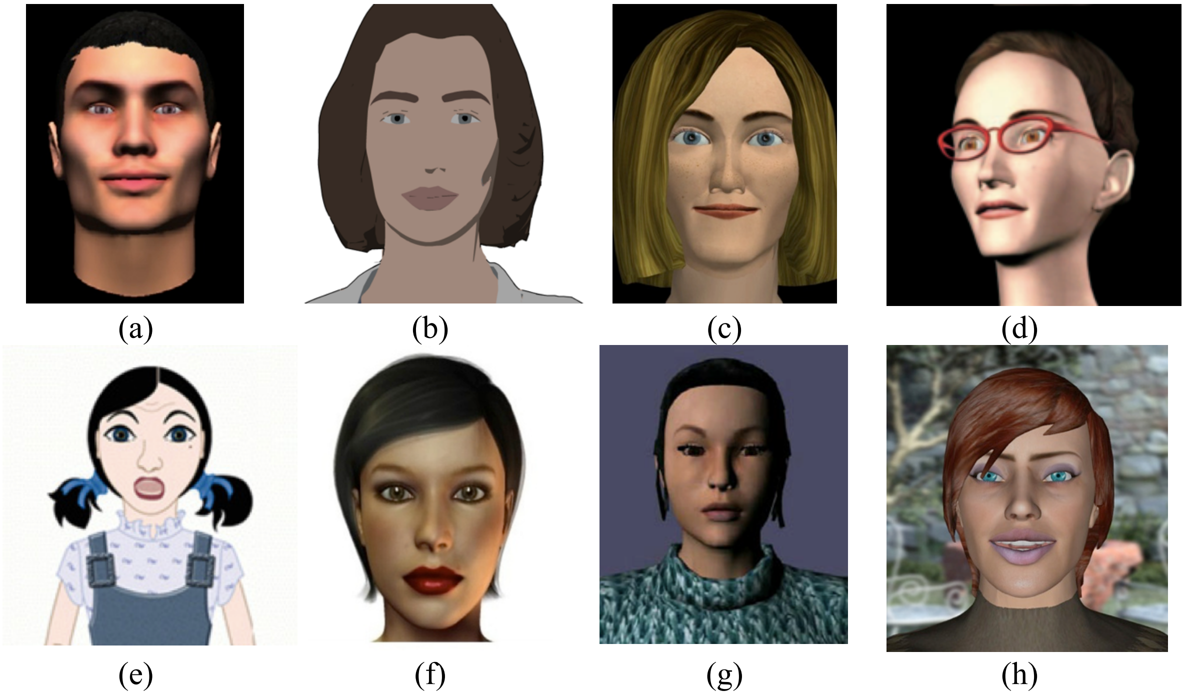


Figure 6. Avatar examples from studies in 2010, 2011, 2012; from (a) to (g), [de Melo '12], [Bickmore '11b], [Niewiadomski '10], [McKeown '11], [Neviarouskaya '10], [Hoque '11], [Poppe '10], and (h) [Čereković '11].

The eight examples in Figure 6 can be categorized into four cases based on the purpose of avatar usage in their study: emotion related case (a), (c), (e); behavior related case (b), (g); data collect related case (d) (f); framework development case (h).

(1) Emotion related case: studies in this category focus on understanding emotional behavior. Researchers often evaluate the emotional behavior's affect on task performance or a subject's reaction. However, the quality of emotional behavior is not often included in evaluations. It is more common to vary conditions whether the chosen behavior is presented or not. Depending on the context, an iconic representation of expression is used to convey the meaning efficiently. However, the complex nature of human behavior cannot be realized in this way if we wish to have more natural experience with a lifelike avatar.

(2) Behavior related case: avatar behavior studies focus on nonverbal gestures to enhance user experience such as subjective likeness or perceived naturalness of behavior. They rarely discuss the quality of visualization to this end. It would be interesting to perform a study to examine the relationship between natural behavior and visualization quality.

(3) Data collection related case: two studies in this category did not use interactive features of an avatar. The context is given by textural instruction or controlled by a human operator in Wizard-of-Oz (WoZ) scheme. It does not necessarily evaluate avatar behavior. Rather, it assumes that a human subject is capable of imagining a given context and showing natural behavior for researchers to collect desired data. However, this method may not be sufficient to extract natural spontaneous behavior. It is very possible for us to show posed behavioral characteristics or stereotypical behaviors.

(4) Framework development case: development of a framework to realize a lifelike avatar is important with respect to its features and capabilities. However, it is required to evaluate the efficiency and fidelity of the framework as well as to verify its practicality by more application development examples.

There have been few research efforts to improve the visual quality of an avatar in the literature. However, two interesting studies addressed how enhanced graphical features affected users' perceived naturalness of an avatar in 2009 [Courgeon '09; de Melo '09]. The enhanced graphical features used in those studies are auxiliary facial features (e.g. wrinkles, tears, and blushing) appearing occasionally while an avatar shows certain emotional expressions.

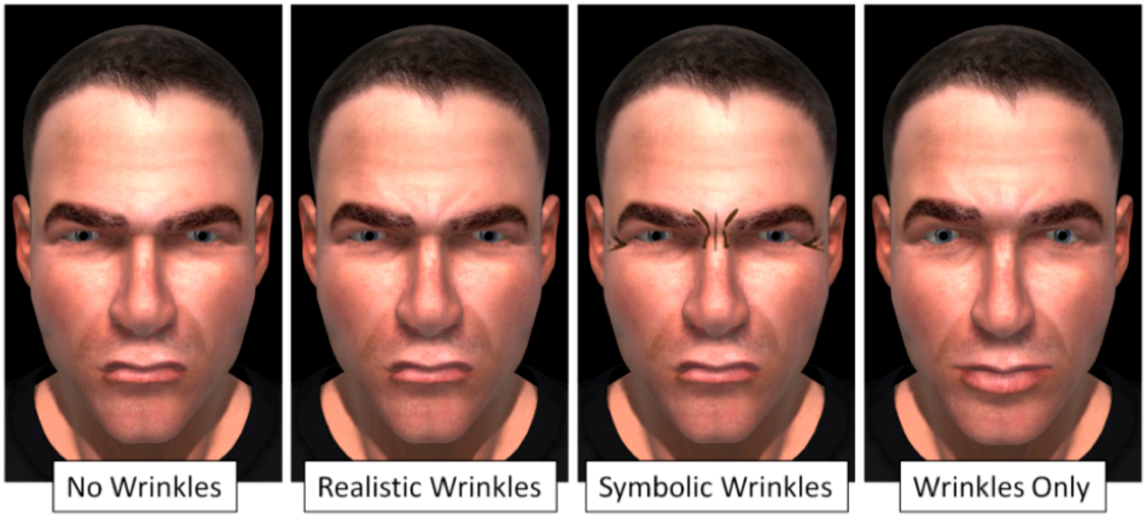


Figure 7. Anger expression using 4 wrinkle modes in [Courgeon '09]

Courgeon et al conducted a study to evaluate the impact of different rendering modes of facial wrinkles by measuring users' perception as well as their recognition accuracy for expressed emotions: four basic emotions (joy, anger, fear, and surprise) and four complex emotions (interest, contempt, guilt, and fascination) [Courgeon '09]. Figure 7 shows four different rendering modes for the anger expression. They created a series of videos with an avatar showing one emotion and images of the emotional expression at the maximum intensity for each condition. A video stimulus was used to measure users' recognition accuracy. A pair of images were used for a user to rank rendering modes according to their expressivity level and to choose his/her favorite ones. The study results suggested that *realistic wrinkles increase avatar's* *expressivity and user's preference, but not the recognition of emotion category*. The user's preference over wrinkle rendering mode indicates that realistic wrinkle mode is significantly better than the others; no wrinkle mode is preferred to symbolic and wrinkle only mode; no significance was found between symbolic and wrinkle only conditions. Although this study was not able to confirm objective performance improvement (emotion recognition accuracy), the use of extra graphical facial features led to better subjective perception that could improve the quality of user experience for a longer interaction with an avatar.

Another interesting study addressed by De Melo et al engaged more facial features such as tears, blushing, and sweat to assess the influence of them on the perception of basic emotions (surprise, sadness, anger, shame, pride, and fear) shown in Figure 8 [de Melo '09].

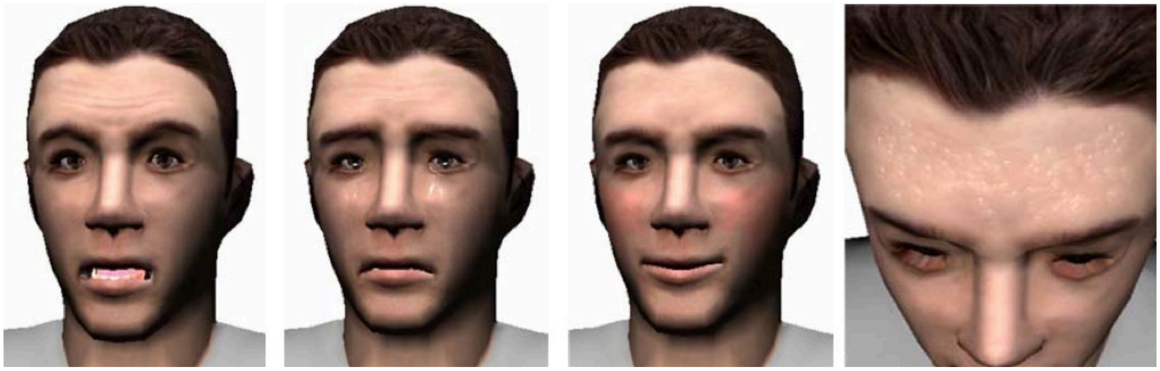
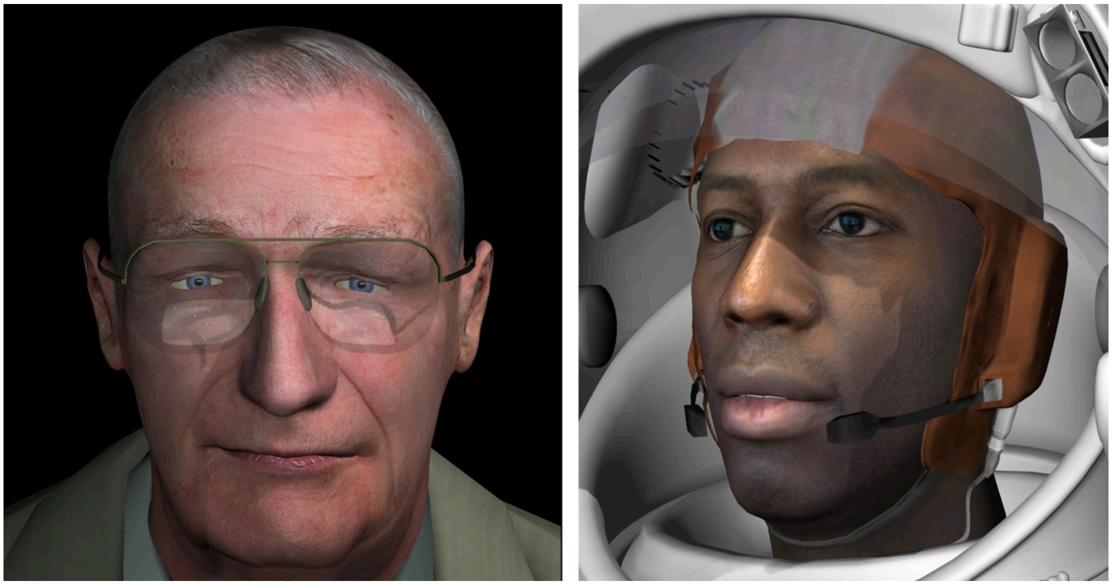


Figure 8. Enhanced avatar graphical facial features. Avatar visualization examples showing wrinkles, tears, blushing, and sweat when it expresses surprise, sadness, pride, and fear in [de Melo '09]

The authors of this study implemented the simulation model using the Graphics Processing Unit (GPU) to visualize facial features. Wrinkles are realized by adopting extra wrinkle normal maps. The model controls wrinkle parameters to select maps and compute wrinkle features when an avatar shows facial expressions associated with wrinkles. Blushing is implemented by adding auxiliary mask information for each vertex in the face mesh to set vertex color (tinting). Similar to wrinkle realization, the model controls which mask should be applied based on avatar expression. Tears and sweat are also supported by normal map techniques. However, tears also require dynamic simulation to mimic dropping (flow from eyes to cheek and downwards). The evaluation study was designed to compare the avatar expression with and without this implemented feature. A subject was shown a pair of rendered images and asked to classify whether the avatar expresses the emotion on 1-10 scale. The study results suggested that anger, sadness, shame, fear, surprise can improve user's perception of expressed emotion. The limitation of this study is that it assessed the perceived expressivity of emotion and it does not indicate the user's preference in such facial features, whereas [Courgeon '09] discussed both expressivity and preference together. However, the findings in this study also serve to support the idea that more graphical features can increase avatar expressivity, which is consistent with Courgeon et al's work.

The aforementioned studies present how we may improve user's perceived preference or expressivity of a lifelike believable avatar with enhanced graphical features. However, their evaluations are limited to their own design using simple configurations: the avatar only showed one emotion without any context; they are only short animations or even rendered images that are not likely to show a more persistent user experience. Wrinkles on a face is a more universal feature for a senior model. Blushing, tears, and sweat seldom appear in our daily communications. However, they offer advantages when we need to show an extreme situation or want to exaggerate the intensity of emotional expression. Nevertheless, the overall visual realism of current examples found in literature is not very close to our ultimate goal to accomplish a lifelike believable avatar.

In contrast, the Lifelike Responsive Avatar Framework (LRAF) provides photo-realistic rendering results with a proven graphic asset pipeline shown in Figure 9. Our visual realism is far superior to those used in others' work. It uses high-resolution realistic textures from real photos to enhance rendering qualities that are especially important on large-scale displays or high-resolution desktop monitors, features skin details with pores and wrinkles using normal mapping, and generates realistic soft shadow-casting. It is expected that supplementing the current state-of-art research in the field with realistic rendering techniques will clearly push our standards one step closer to the most believable lifelike avatar interfaces, which is our ultimate goal of this research.



(a) Dr. Alex Avatar (b) Mike Astronaut Avatar

Figure 9. Avatar examples used in LRAF application; In (a), fine-grained skin details were preserved to represent a senior target person. In (b), a mid-aged male was visualized as an astronaut tour guide avatar.

2.3 Avatar Behavior Modeling

A rule-based system has long been sought in conjunction with relevant literature such as psychology, cognitive science, and neuroscience. Engineers seek for a good model to explain the characteristics of human behavior and design a computational model to re-construct them. As opposed to the rule-based system, the data-driven method has appeared relatively recently. As technology advances further, the availability of meaningful data has improved and some efforts towards standardized data acquisition methods enable us to collect data easier than before. The following two chapters introduce those two distinct methodologies in detail.

2.3.1 Rule-based Model

A large amount of research effort to model avatar behavior has been devoted to building literature-based models. Researchers believe that we can build a better model by understanding profound human nature for a natural avatar or an intelligent agent. However, it is often not tractable to find a good foundation in literature to build such a computational model. To explain our emotional state at a certain moment, we have to consider many different aspects surrounding us. Personality takes an important role in cognition and processing external events; mood also matters; accumulated emotions may change our behavior too. Our internal emotional state can alter many behavior aspects. There are other factors that may have an affect on our nonverbal behavior. For example, when we are talking we often show head gestures to better convey certain meanings of utterances. In this chapter, we introduce some of successful studies: (1) personality and an emotional state model; (2) gestural behavior rules models.

Personality and Emotional State Model

There have been two well-established personality models in recent studies. The first one is the big Five Factor Model (FFM) [Goldberg '90] and the second one is the Pleasure-Arousal-Dominance (PAD) space model [Mehrabian '96b]. While the PAD space model uses three factors with polarity and associated degree (intensity), the FFM model describes traits with five factors; openness, consciousness, extraversion, agreeableness, and neuroticism. Especially the PAD space model provides well-formed computational foundations that can be utilized to depict one's characteristics. Mehrabian also showed how the conventional FFM model could be mapped into PAD space in his following study [Mehrabian '96a]. He conducted a user study to gather relationship information between FFM model and PAD model to build a conversion equation. Findings from this study were:

(1) Extravert is primarily dominant and secondarily pleasant.
(2) Agreeableness resembled with pleasant, arousable, and submissive characteristics.
(3) Conscientiousness included equal degrees of pleasant and dominant qualities.
(4) Neuroticism involved almost equal degrees of pleasant and unarousable.
(5) Openness was weighted primarily by dominant and secondarily by arousable.

PAD to FFM conversion (1)		
Extraversion	=	0.23 P + 0.12 A + 0.82 D
Agreeableness	=	0.83 P + 0.19 A – 0.21 D
Consciousness	=	0.32 P + 0.30 D
Neuroticism	=	0.57 P – 0.65 A
Openness	=	+ 0.33 A + 0.67 D
FFM to PAD conversion (2)		
Pleasure	=	0.59 Agree + 0.19 Neuro + 0.21 Extra
Arousal	=	-0.57 Neuro + 0.30 Agree + 0.15 Open
Dominance	=	-0.60 Extra - 0.32 Agree + 0.25 Open + 0.17 Conc

Table I. Mehrabian's mapping equation for PAD and FFM: (1) Estimation of true relationship between the FFM and PAD scales; (2) Prediction of PAD space values using the FFM scales [Mehrabian '96a]

Mehrabian's findings were further analyzed to map his PAD model to FFM and vice versa shown in Table I. For example, personality traits (Openness: 0.4, Consciousness: 0.8, Extraversion: 0.6, Agreeableness: 0.3, Neuroticism: 0.4) are mapped to PAD vector (0.38, -0.08, 0.50) that is a slightly relaxed state in PAD space model.

Gebhard proposed a layered approach to explain how a human processes his/her emotional state based on its temporal aspect [Gebhard '05]. Personality Traits are long-term characteristics that lasts years or even a lifetime, Mood is medium-term effects showing gradual changes over time, and Emotion is likely a short-term response to immediate stimuli that decays shortly after. In his hierarchical model, Trait is a fixed given factor, Mood changes upon accumulated emotions, and Emotion is resonating with short-term events. Therefore, he could compute Mood by projecting all stimuli to PAD space with decay and intensity parameters (Figure 10). This provides a well-structured computational model for avatar's emotional state.

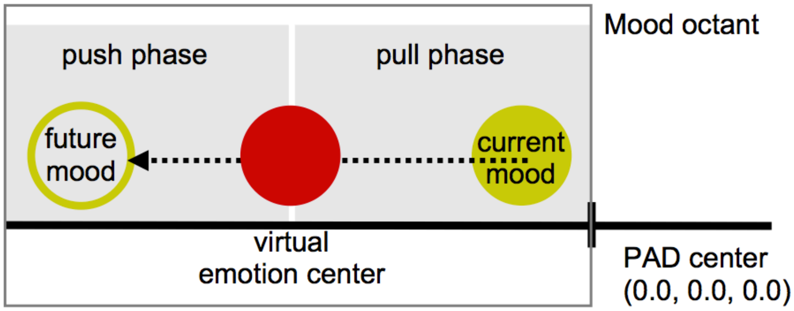


Figure 10. Pull and push mood change function. In PAD space, if there are active Emotions, first its center is computed by summing all their PAD vectors (virtual emotion center). If the current mood locates between the PAD origin and the virtual emotion center, the current mood is attracted towards the virtual emotion center (pull phase). Otherwise, the current mood is pushed away (push phase).

Gebhard utilized computed mood information to control avatar behaviors: the wording and phrases; the use of dialogue strategies; trigger idle gestures; change the characteristics of conversational gestures and postures; and control facial expression. However, the intensity of emotion is not influenced by the mood. In other words, the eliciting emotion always has the same intensity regardless of the current mood, which is not natural behavior since a real person may alter his/her emotional stimulus based on his/her current mood. For example, a sad person will become much sadder upon encountering an additional sad stimulus in the future.

Gestural Behavior Rules Models

There has been enough evidence that utterances have a certain relationship between their underlying meaning and our nonverbal behavior. For example, we extensively use head gestures such as nods and shakes to express affirmation and negation. We often use a nodding gesture while listening to others to express conversational grounding. This dissertation focuses on head gestures tightly-coupled with utterances. Hand and body gestures are also often correlated with our daily conversation. However, they are beyond the scope of the work. Therefore, we will review a few important prior works related to head gestures with respect to speech utterances.

To this end, one of the most influential works is [McClave '00]. McClave conducted a microanalysis of videotaped conversations between native speakers of American English to categorize head movements co-occurring with speech. She confirmed that there are linguistic functions of head movements in the context of speech. Her findings were summarized in three categories such as semantics, discourse, and interactive functions. Semantic function is particularly tightly coupled with one word that can be analyzed by a shallow parsing scheme used in this thesis. The sematic function in her study is composed of:

> *(1) Inclusivity: the lateral sweep that co-occurs with the concepts of inclusivity such as the words 'everyone' and 'everything'. Gesture occurs within one word.*
> *(2) Intensification: lateral movements of the head often co-occur with lexical choices such as 'very', 'a lot', 'great', 'really', 'exactly', and the like. Gesture occurs mostly within one word span. However, it covers a whole sentence occasionally.*
> *(3) Uncertainty: affirmative statements are often marked verbally as uncertain by 'I guess', 'I think', 'whatever', 'whoever', and similar expression. Gesture occurs mostly within a whole sentence or phrases.*

Lateral sweep is a single head movement from one side to the other. There three linguistic functions have been most used in literature when an avatar head gesture is based on shallow parsing, which does not involve deeper understanding of sentences.

Even though a rule-based computational model for avatar behavior provides highly coherent behavior, it does not offer a detailed model for subconscious or unconscious behavior occurring without strong correlation with utterance. These pitfalls can be supplemented by other methods in this work.

2.3.2 Data-driven Model

Data-driven computational modeling, more specifically machine learning techniques, is well established in various areas such as natural language processing, pattern recognition, medical diagnosis, and many more. However, its adoption in avatar research is relatively new. A trained machine can predict an uncertain state based on known empirical data. For instance, when we want to diagnose a patient's disease, we can train a machine with a large set of symptoms and correlated diseases, and then feed the patient's abnormal symptoms into the system to predict a causing disease with probabilities. Recent attention to this method in avatar related research is rooted in the fact that modeling human behavior requires vast amount of knowledge in several domains [Lee '10b], and it is very hard in many occasions. Therefore, it may be possible to train machines for behavior prediction from selected features we feed into the system.

Most recent studies in this method were applied to body gesture generation from speech prosody [Levine '10] and utterance text [Michael '08], to a listener's backchanneling upon incoming utterances [Huang '10; Lee '10a; Morency '08], and to head motion prediction while an avatar is speaking [Gunes '10]. Often used techniques include the Hidden Markov Model (HMM) [Rabiner '89], the Support Vector Machine (SVM) for classification [Cortes '95], and the Conditional Random Field (CRF) model [Lafferty '01]. Selected features used to train the models are lexical information of utterance, affection of surface text, dialog act upon avatar goals and appraisal, and direct audio signals.

The drawback of current studies in data-driven approaches is that they support only a few gestural behaviors such as head movements and body gestures. Another is the fact that they use only short-term events or input to predict behavior. In other words, they do not account for accumulated or progressive change of emotional state affecting avatar behavior over time. For example, if an avatar receives repetitive annoyance from a user, it should express far more anxious or irritated behavior. However, the data-driven model does not accommodate it very well. The length sequential causality in the machine learning algorithm requires an exponential amount of data to predict such chain effects.

In this chapter, we reviewed some relevant works including visual realism and behavior modeling for an interactive avatar. As compared to the visual qualities in prior studies, which are fairly limited to represent a real person in virtual space, presented examples in LRAF demonstrated far superior quality with photo-realistic renderings. As far as behavioral modeling concerns, since the two distinct methods each have its own drawbacks, a hybrid approach is desired to obtain advantages from both techniques supplementing each other’s shortcomings. More details will be presented in Chapters 3 and 4.

3. LIFELIKE RESPONSIVE AVATAR FRAMEWORK

The Lifelike Responsive Avatar Framework (LRAF) is the outcome of collaborative research efforts to reproduce a real human as an avatar. This prior project aimed to develop a flexible framework to support capturing a particular subject's visual appearance as well as personalized behavioral mannerisms. The framework is mostly focused on an avatar's upper body model to best utilize a display monitor in landscape mode. An avatar fullbody behavior is beyond of the scope of the LRAF other than some prefabricated body motions. The LRAF helps researchers and developers overcome steep initial burdens to create an avatar-enabled application by offering a solid low-burst design and production pipeline for a lifelike avatar. One can imagine learning about a Turing Machine from an Alan Turing avatar instead of reading a textbook or talking to family members who have passed away. With LRAF, several applications were successfully developed to prove our approach.

An overview of LRAF presented in the thesis is described in this chapter. The LRAF system is fully capable of all fundamentals to create a lifelike avatar and to implement more believable avatar behavior models in Supervised Hybrid Expression Control Framework (SHECF). Chapter 3.1 discusses the overall LRAF architecture including input, output, and internal processors. The deployed application examples developed with the LRAF system are explained in Chapter 3.2.

3.1 Overall Architecture

LRAF is comprised of multiple modules to accommodate various features of a lifelike avatar. Figure 11 shows the high-level framework architecture. There are two external inputs, the scene manager, the GUI manager, the expression synthesizer, and graphic/audio rendering modules. The separation of two input modules enables the main framework to be more flexible so that an application developer can easily add more features later.

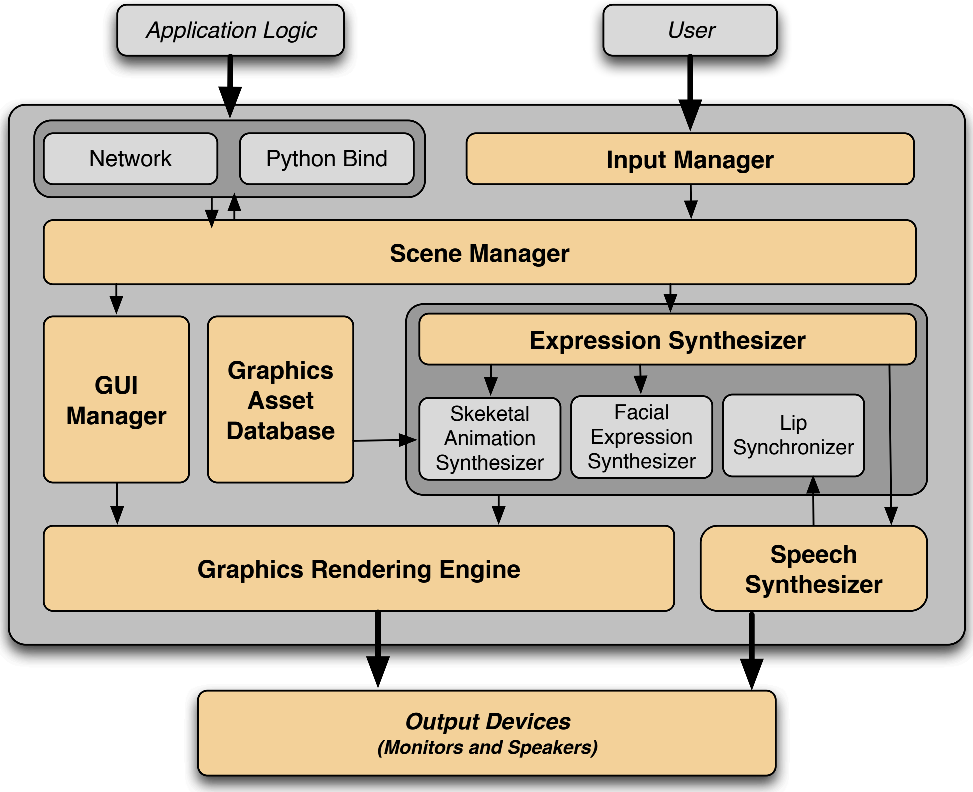


Figure 11. Functional architecture of LRAF system comprised of 8 main modules; Input Manager, Scene Manager, Expression Synthesizer, Graphic Asset Database, GUI Manager, Graphics Rendering Engine, and Speech Synthesizer, and Output Devices such as monitors and speakers.

LRAF can communicate with an external dialog manager to drive avatar interaction logic via a network or receive direct commands through Python bindings. User inputs are microphone signals, video stream from a webcam, and other types of sensor networks to detect a user and environmental changes to control avatar visualizations. Once the system receives all necessary information, it passes those signals to the Scene Manager to produce an appropriate outcome, and it is transferred to other internal modules to realize an avatar - GUI components show auxiliary information, and the Expression Synthesizer composes facial expression and body motion. Finally, processed data will be passed to the underlying rendering routine and audio generator.

LRAF visualization relies on the Object-Oriented Graphics Rendering Engine (OGRE) library for most of its low-level graphics modules [OGRE '13]. OGRE is an open-source platform independent rendering engine that supports modern graphics features such as skeletal / shape animation, shader languages, and a flexible plug-in architecture. The audio rendering module for an avatar speech supports two different methods. One is a computer synthesized voice, Microsoft Text-To-Speech (TTS) Engine [Microsoft '13], and the other is a recorded voice. Both modules feed their outcome to the Lip Synchronizer to propel the avatar’s lip animation on the fly.

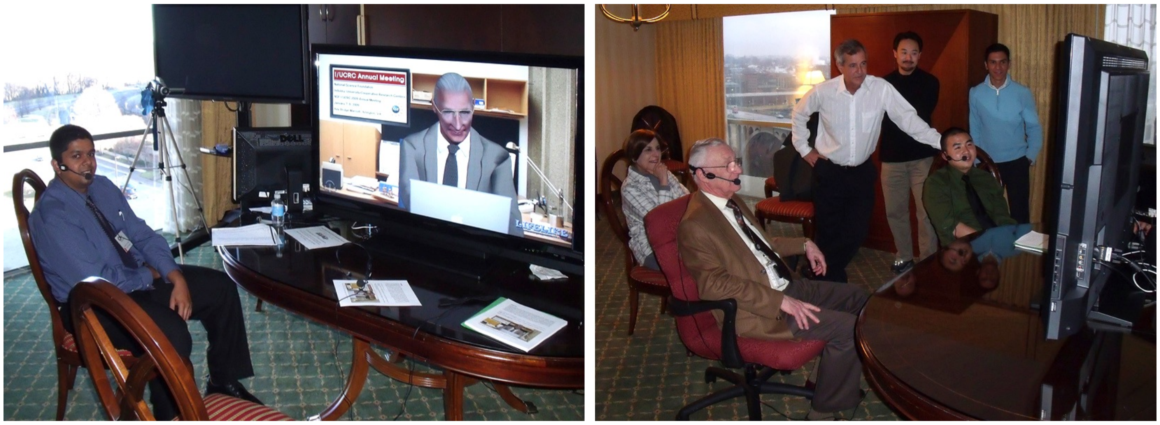
In the Expression Synthesizer, the system applies two different approaches to implement avatar nonverbal behavior. For body motion, the Skeletal Animation Synthesizer uses a prefabricated motion database that is tagged with action-based commands such as idle, speak and pointing. The Facial Expression Synthesizer controls facial feature animations in a procedural manner without a predefined database. Some of the personalized unconscious behaviors such as blinking and subtle mouth movements are defined in an external template xml file and controlled by the Facial Expression Synthesizer. Finally, processed data is sent to output devices such as HD monitors for visualization and stereo speakers for audio rendition.

3.2 LRAF Applications

In this chapter, three types of example applications are described. First, an application that tries to preserve a particular person's knowledge and visual appearance is introduced. Modeling a particular person as an avatar is challenging because we have to recreate his/her visual appearance as close to the target person as possible as well as to capture behavioral characteristics of the person. Second, a museum tour guide avatar and an educational exhibit installation for a more broad audience in a public space are presented. An avatar-enabled application in public space needs more engaging interactions and requires supporting a flexible framework to integrate with other educational materials or learning tools. Last, an experimental virtual patient application for a medical / healthcare application is explained. Applications in this domain inevitably impose more theory proven approaches because the context used in this domain is more serious than other public use case. For example, an emotional facial expression to convey painful feeling has to be very representative and simple to vary in diverse conditions.

3.2.1 Preservation of a Real Person

The first prototype application designed with LRAF system was aimed to preserve NSF program manager, Dr. Alex Schwarzkopf. The application was supported by an external dialog management system that supports speaker-independent continuous speech recognition with a context-based dialog in addition to the LRAF visualization framework. Interaction between a user and an avatar is handled by unsupervised natural speech shown in Figure 12. An example subset of his knowledge of NSF protocols was encoded in a grammar-based speech interpretation system and a context-based reasoning system.



(a) Alex Avatar Application (b) Dr. Alex using application

Figure 12. Alex avatar application at I/UCRC 2009 annual meeting. In (a), the system used a conventional 50” TV screen to realize a lifesize avatar and one moderator wore a headset microphone to interact with Alex avatar. In (b), the real Dr. Alex was interacting with his digital version.

During the demonstration at the I/UCRC annual meeting, most attendees were able to recognize the Alex avatar at first glance and to interact with him to retrieve information. This example application supports multi-party interaction between a user, a moderator, and an avatar. The system takes inputs via two microphones and generates answers for questions via synthesize voice. The Alex avatar is also capable of engaging eye-gaze and head turns between a user and a moderator with respect to its interaction context. An avatar starts interaction with a greeting; introduces a moderator and itself to a user; and asks the user’s name to identify him/her from the meeting register to offer more customized information such as: showing acquaintance with the user by identifying his/her affiliation; recalling the last interaction; sending an email for more detailed information; or arranging a meeting with someone who he/she knows from meeting participants. A user can further interact with Alex avatar by asking questions including a planning grant proposal, the deadlines, a reason for rejected proposals, and so on. More detailed information can be found in two previous papers [DeMara '08; Lee '10c]. Developed Alex avatar's facial expressions showed similar recognition rate and pattern compared to real Alex's expressions from a large-scale online survey (*n=1,744*) reported in [Lee '10c].

The main goal of this application was to create an avatar as close to a particular person as possible. Therefore, it greatly improved LRAF visualization framework to accommodate advanced graphics techniques to resemble a real person's visual details in iterative design and development process later on. This helped us establish a solid design and production pipeline, and, as a consequence, we were able to create other types of avatars and applications within a very reasonable amount of time. More details of visualization techniques are presented in Chapter 4.

One more lesson learned from this prototype application was how we can further customize an avatar for a specific target person. Other than visual resemblance, the behavioral aspect of an avatar was also very important. The concept of *Behavioral Personification* was incorporated to mimic Alex's specific characteristics in the later stage of the application. For example, his eye-blinking rate is far more frequent than average person's blinking rate, he types keyboard only with two index fingers, and he often uses very specific utterances such as *"Keep the peace."* All these observed characteristics were specified in the avatar specification template so that LRAF system could recreate his diverse mannerism. Many users who knew Alex responded very positively to these additions from an informal conversation after several demonstrations.

3.2.2 Museum Oriented Context-Aware Application

Astronaut Application for Chicago Adler Planetarium

After the first successful prototype application, we tried to reuse LRAF system in another domain with more publicly engaging scenarios. EVL have collaborated with many other disciplines over decades, and Chicago’s Adler Planetarium is one of those collaborative venues where EVL often transfers research works in the public domain. Since a museum-based avatar installation is one good application area to enhance user experience along with computerized or even traditional exhibition, we designed a context-aware avatar application. This is a more engaging method to augment existing installations with an avatar-mediated approach instead of setting a standalone avatar as a generic museum kiosk type of installation. Figure 13 shows the developed application running on a high-resolution large display. A user can ask questions such as *“tell me more about this photo”, “what is this?”, “take me on a tour”*, and so on via a microphone.

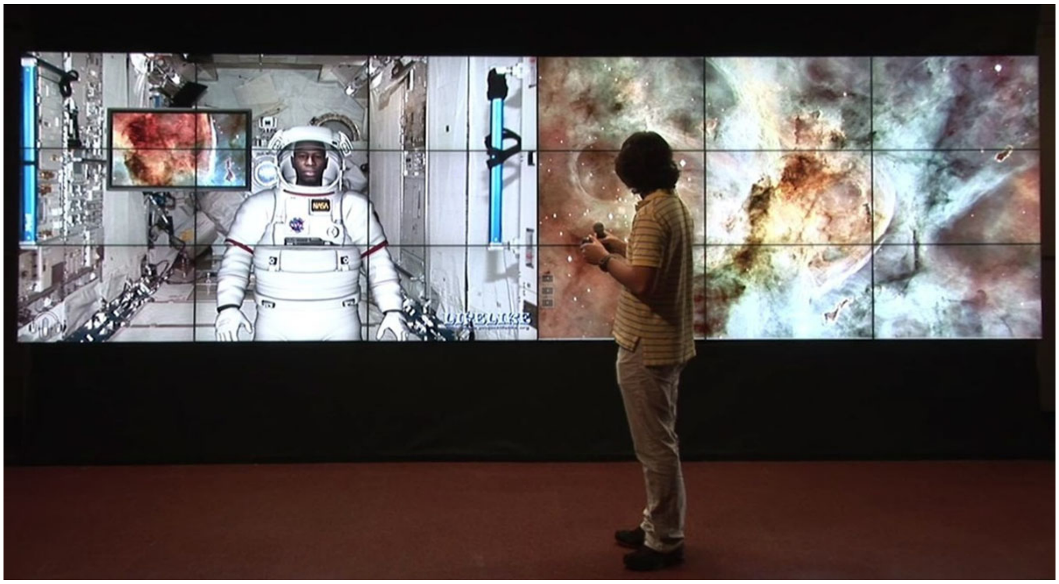


Figure 13. A tour guide astronaut avatar describes features of the high-resolution Carina Nebula image on a tile display system.

Communication between the avatar framework and accompanying application, a high resolution seamless image viewer showing nebula images, was implemented using a Python script with its binding to the LRAF main modules. Since there are many other possible types of applications that can take advantage of an avatar interface, LRAF was extended to support a scripting module to offer more flexible architecture in integration. It helps developing application interaction logic with scripts rather than writing custom functions in the LRAF main framework. Additionally a custom knowledge encoding tool and a dialog designer (hierarchical finite state machine authoring tool) were implemented so that domain experts can easily add more content to the application.

The most significant innovation in this application is an extension to support indirect or a context-sensitive input from a third-party application and a hybrid control scheme. The system analyzes a user's speech input via a microphone as well as monitors his/her interaction with other application. For example, when a user navigates images using a joystick and asks a question, *"What is this?"* the astronaut avatar performs three steps to answer the question: (1) Compute the current coordinate of the viewport in the image viewer; (2) Search knowledge database to identify information entry that is directly associated with the coordinate or within reasonable boundary of it; (3) If matching knowledge is found, then explain it. Otherwise, find the nearest one and responds with *"I don't see anything interesting there. The nearest one I can explain is XXX. Do you want me to take you there."*

The hybrid control scheme in this application offers a primary control of the image viewer to a user and also takes the control over when he/she asks directive command. For example, when a user asks, *"Tell me more about Carina Nebular"*, the avatar answers in three steps: (1) smoothly adjusts current view on the image viewer to center the questioned region; (2) zoom in to show more details in the region; (3) then, explains the region. The astronaut tour guide also utilizes a multi-purpose whiteboard to show auxiliary information such as an image in different color spectrum, text information, and videos when it explains an image or a region. To the best of our knowledge, this is the very first and a novel approach to enhance user's experience with an avatar in great extent.

This application was pilot tested in Adler's Space Visualization Laboratory. Results from a small-scale informal pilot user study indicated that people highly preferred this context-aware avatar-enabled exhibit compared to the standalone interactive installation, an interactive image viewer. It is planned to improve the current application further and add more similar applications in the near future. For example, we can design an avatar, Apollo astronaut Jim Lovell, to integrate with Adler's Moonwall exhibit in order for visitors to interact with him while navigating the moon and hearing his explanation about the moon and various landing sites.

Orlando Science Center Exhibit

There is another upcoming museum-oriented application developed using the LRAF system for Orland Science Center (opening expected in the end of summer 2013). This application was implemented for the theme of Artificial Intelligence exhibition with a team of researchers from University of Central Florida (UCF). A historical figure, Alan Turing, was restored as an interactive avatar (Figure 14) and followed by additional 11 diverse avatars.

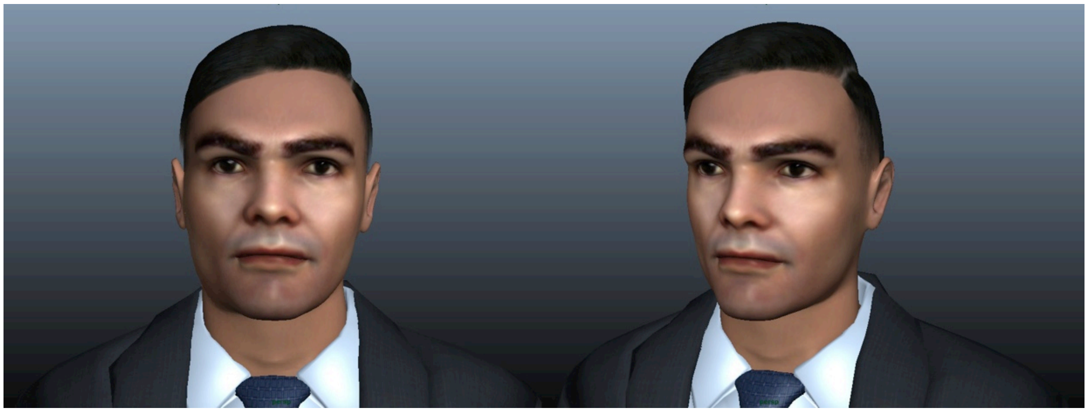


Figure 14. Alan Turing avatar was developed for Orlando Science Center Exhibition to invite middle school students to get interested in Computer Science / Engineering, and to teach them about Turing and the Turing test.

Visitors will be able to talk to a virtual Alan Turing to learn about Artificial Intelligence and to select an avatar of their own choice to interact with. The installation will be composed of two touch-enabled 24” monitors driven by one desktop computer and a microphone so that people can experience more natural multi-modal interaction.

Similar to the first application, Alex avatar, this application is driven by the external dialog manager developed by UCF. However, LRAF framework utilizes the python script module to communicate with the dialog manager. This method reduces overhead to change main framework to accommodate additional communication protocol between LRAF and the dialog manager. There are several more changes that the research team made throughout the course of this project:

(1) Multi-agent interaction support: The application provides more than one avatar at the same time so that a user can interact with multiple-agents. When a user selects their choice of an avatar model, the moderator avatar, Alan Turing, will introduce it to a user and continue multi-agent interaction.

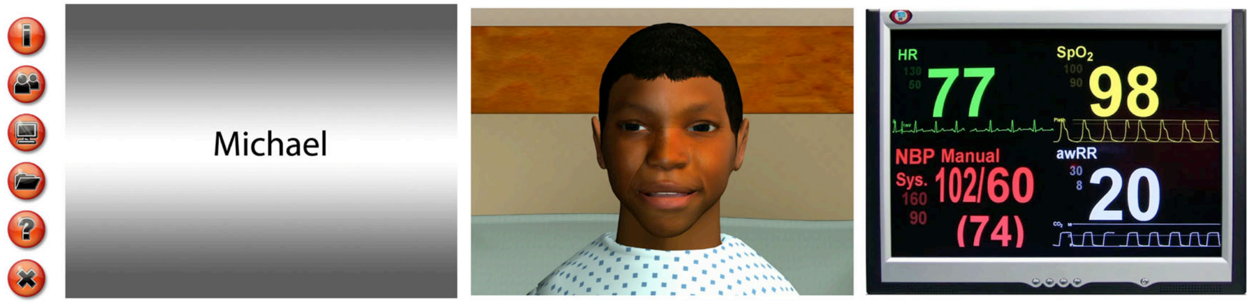
(2) Enhancement of an avatar body motion control: This application engages more body motions than previous examples as each avatar moves/walks in the scene. Spatial context for a body motion is added to diversify avatar behavior. For example, when the avatar walks in the scene, the Skeletal Animation Synthesizer computes the destination point and determines a proper motion type such as a slow walking sideways for a short distance and a normal walking cycle for a long distance.

As this exhibit is close to opening, the author expects more details regarding how the newly developed features enhance user interaction and how effective the system is at improving users' learning experience. The collaborative research team will report detailed results after collecting data in the future publications.

3.2.3 Virtual Patient

The use of an avatar in healthcare and clinical settings has been actively studied in literature. The flexibility and practicality of LRAF system allow us to implement such an application without much effort. A Virtual Patient (VP) Vignette application for Pediatric Intensive Care Unit (PICU) nurses was developed in collaboration with the UIC nursing department and is published in [Lee '13] (to appear).

The goal of this project was to re-design one of classic Vignette studies using VP techniques and to measure nurses' recognition of the facial expression and consistency of VP Vignette with their professional experiences. The chosen scenario was derived from the Pain Beliefs and Practices Questionnaire (PBPQ) case studies. Figure 15 presents a developed VP application.



(a) Main screen (b) Virtual patient (c) Vital Monitor

Figure 15. PBPQ case presents a 10-year-old boy who had abdominal surgery a day before, and have a self-assessed pain rating of 8/10 with stable vital signs. In (a), nurses can select 6 menus; patient information, observes patient, vital signs, medical records, pain rating, and exit. (b) shows the 10-years-old boy VP and (c) shows simulated vital sign.

The preliminary results from this study indicated that LRAF system is also capable of developing a VP application. 98.5% of nurses were able to recognize given expressions (smile and grimacing) and 87% validated the similarity of the vignettes to their experiences with patients. More findings and lessons learned from this VP application are:

(1) Possibility of avoiding subjective bias in Vignette studies: A classic Vignette study commonly relies on a written scenario and it is likely to introduce biased subjective interpretation because a subject has to imagine the given situation by itself. However, the VP Vignette study can help reduce such bias by portraying the scenarios with more realistic visual presentation.

(2) Needs a more systematic approach to create believable facial expression: The directed facial expression to realize the VP Vignette requires widely acceptable representation that is different from mimicking a particular person. Ekman’s Facial Action Coding System (FACS) provides a profound analysis of human facial feature movements with respect to our categorical basic emotions [Ekman '78]. The Facial Expression Synthesizer is enhanced to support FACS based expression realization and was able to create highly recognizable facial expression even for grimacing one that is not in six basic emotion category.

(3) Development of children avatar model: LRAF is used to model an adult avatar in various applications. A child avatar model in this case was our first trial. Although the basic mesh structure of a face model remains similar, LRAF was able to create a 10-year-old boy avatar model without altering design and development process.

In this chapter, an overall architecture of LRAF system is explained in conjunction with several example applications that were implemented with LRAF system. LRAF is flexible to adopt many different scenarios and is capable of accommodating various application needs. Although developing a well-defined and efficient framework is a very important role of research, it is also critical to prove its usability and significance as well as to transfer its research outcome to the broad audience and public domain.

4. AVATAR VISUALIZATION

An overview of the avatar visualization framework as a part of the Lifelike Responsive Avatar Framework (LRAF) system presented in the thesis is described. Chapter 4.1 discusses the visualization method used to realize believable high quality avatar appearance with advanced graphics techniques. The design and implementation details of avatar facial expression are explained in Chapter 4.2 followed by the avatar body motion acquisition and control method to create non-repeatable natural human gestures in Chapter 4.3.

4.1 Realistic Avatar Visualization

At the beginning of the NSF funded project that formed this thesis framework, the goal of the research was aimed to design a prototype avatar for a particular real person. Therefor, the primary objective in visualization was focused on how we can make an avatar most visually similar to a given target person. The whole design and development process was divided into several components such as head generation, body modeling, full body skeletal rigging, and animations. Then, further graphical enhancements with advanced graphics techniques followed. Many commodity software packages were used together with customization techniques to reproduce Dr. Schwarzkopf model until a fully established workflow pipeline was set to improve design productivity.

Firstly, the base model of the avatar head was created using Singular FaceGen Modeler [FaceGen '13] from a few photographs of Dr. Schwarzkopf. Then, this base model was exported to Autodesk Maya Modeling software [Autodesk '13]. Next, the avatar full body model was also designed in Autodesk Maya software. Since we use a skinned mesh to animate the avatar body, we built a body skeleton structure with about 70 bones including all finger bones, and rigged it to the body mesh. Finally, the head model was attached to the body.

While the first design process was an efficient way to begin with, the generated base mesh model and its texture was often not detailed enough or not similar enough to the target model. In particular, we noticed that the accuracy of resemblance was lower if the target person is elderly and has many skin wrinkles in the photos. The default texture resolution from a head generation tool was also not enough to fully realize the detailed facial features in high-resolution display. Fine tuned adjustment of facial proportion and texture reconstruction were necessary to achieve the best resemblance. Figure 16 illustrates this realism enhancement by applying a high-resolution photo-based texture projection.

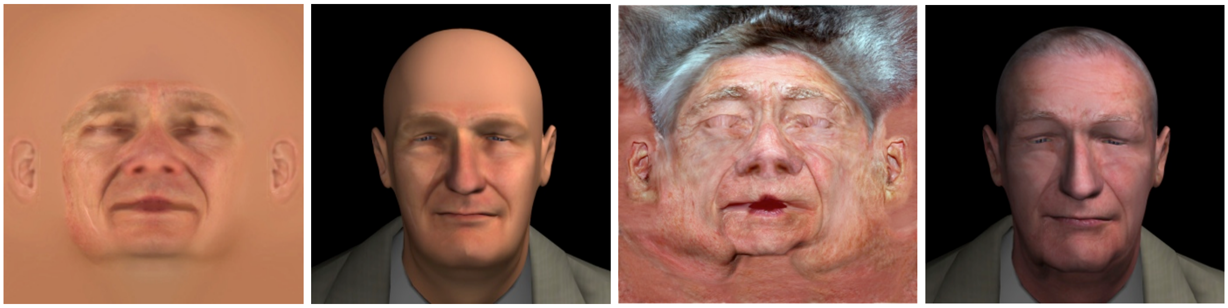


Figure 16. Skin Texture Enhancement; left two images used default texture (512x512 resolution) from FaceGen and show its rendering result, right two images adopted a high-resolution skin texture (4096x4096 resolution) obtained by projecting high quality photos of a target person onto a 3D mesh model.

Although a higher resolution color texture map enhanced rendering quality significantly, the shading on the skin was still a bit too plain and too smooth. To realize human skin characteristics further, it was necessary to add more subtle features such as pores and wrinkles. The most widely adopted method to accommodate those details in graphics is to use a tangent space normal map [Cohen '98]. This technique utilizes the vast computation power of graphics hardware in pixel space without losing rendering frame rate compared to a high-density polygonal mesh model approach. A normal map texture was generated from the color texture based normal extraction method that was used in [de Melo '09]. The final rendering result of a normal map is shown in Figure 17.

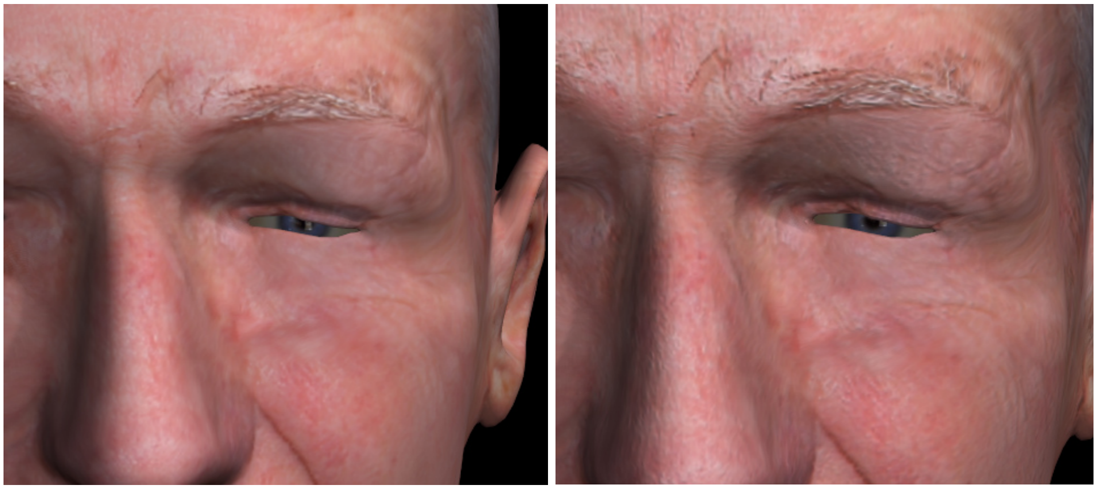
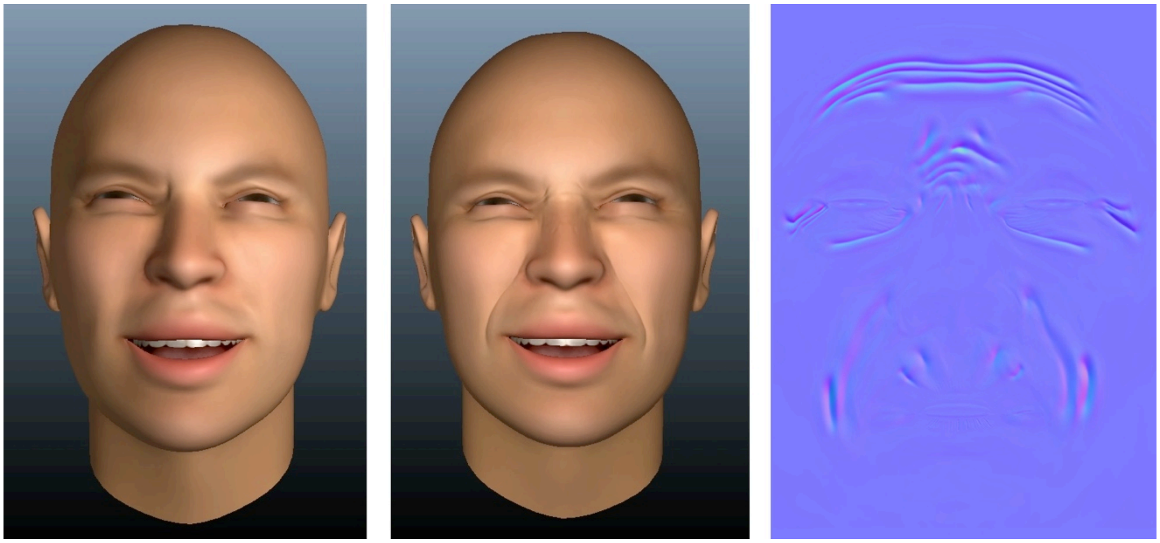


Figure 17. Effect of Normal Map for Skin Details; Left image rendered with color map only and right image used the normal map technique.

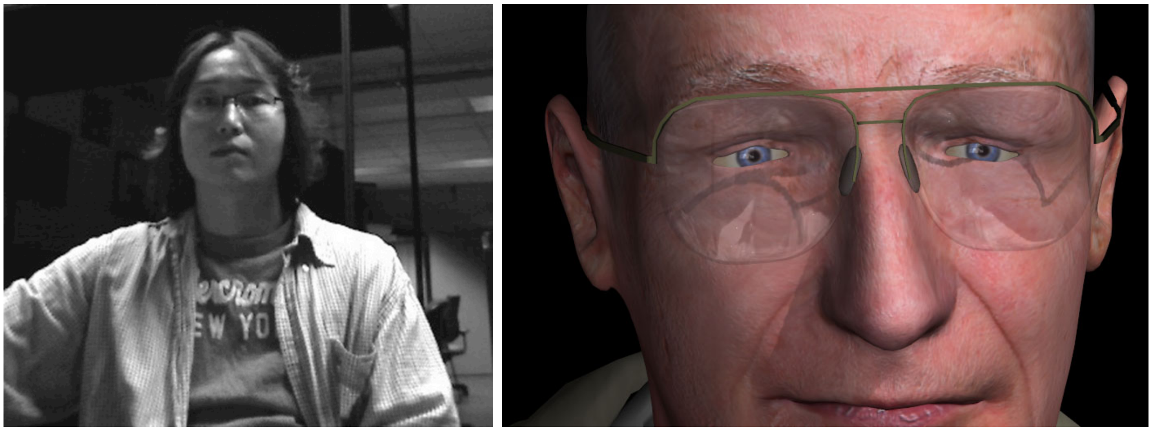
The aforementioned normal map techniques to preserve skin details for an elderly subject were mostly a static facial feature. However, when one moves facial muscles, it often triggers other dynamic features such as deep wrinkles caused by muscle contraction. This dynamic feature can greatly improve perceived naturalness of avatar facial expression [Courgeon '09]. The author’s previous study also confirmed the efficiency of dynamic wrinkles to this regard [Lee '13] (to appear). As opposed to the static normal map approach, a dynamic wrinkle generation requires a different control scheme to simulate muscle contraction model. One method is to calculate distance from each vertex on the face so that wrinkles appear when the distance is reduced. However, this is rather expensive operations and not easily portable to Graphics Processing Unit (GPU), which causes overhead in GPU computation. Therefore, a region-based masking and facial expression-mapping method was adopted. Similar approach was presented in [Oat '07]. For example, when the Facial Expression Synthesizer triggers a certain expression (i.e. wide open smile), a dynamic wrinkle generator sends a wrinkle-enabler flag and expression intensity to the custom wrinkle shader on the GPU to turn on a low-frequency masking map and to compute an appropriate lighting model to generate wrinkles like a mouth furrow for wide open smile. Figure 18 presents the result of the dynamic wrinkle shading.



(a) Rendering without wrinkle (b) Renderings with wrinkles (c) Wrinkle normal map

Figure 18. A dynamic wrinkle generation example for grimacing; In (a), an avatar shows grimacing without wrinkle. In (b), a dynamic wrinkle technique is applied to show enhanced expression. (c) shows a normal map texture used in this example.

To improve the user experience and embodiment further in interaction with the avatar, a graphical gadget was implemented to reflect user's presence in virtual environment. This possibly fills the gap between the virtual and real worlds. This rendering technique recreated a user's existence on reflective material in the scene by using a live video feed shown in Figure 19. During our internal review, most users noted higher engagement and exhibited excitement with this feature; *"Wow, I can see myself there," "It looks pretty cool," "That's very realistic."*



(a) Webcam Live Video Stream (b) Reflection on Glasses

Figure 19. Dynamic texturing on a reflective surface; In (a), a live video stream via a webcam is converted in black and white image. In (b), streamed image from a webcam is blended on to the reflective surface of glasses.

4.2 Avatar Facial Expression

The avatar head itself is one of the most complicated pieces in the system. It should be able to express all emotional facial features and lip motion to realize speech utterances as well as the subtle emotional state of an avatar. The head model generated in the previous chapter is capable of 38 facial feature blendshapes including 6 emotions, 16 modifiers and 16 phonemes shown in Figure 20. A modifier is similar to Ekman's Facial Action Coding System (FACS) Action Unit (AU) [Ekman '78] that controls small parts of the face such as a blink and eyebrow up / down. However, the base model only supports a partial set of FACS AUs. More comprehensive modifier channels are added to LRAF to enhance flexibility of avatar facial expression later.

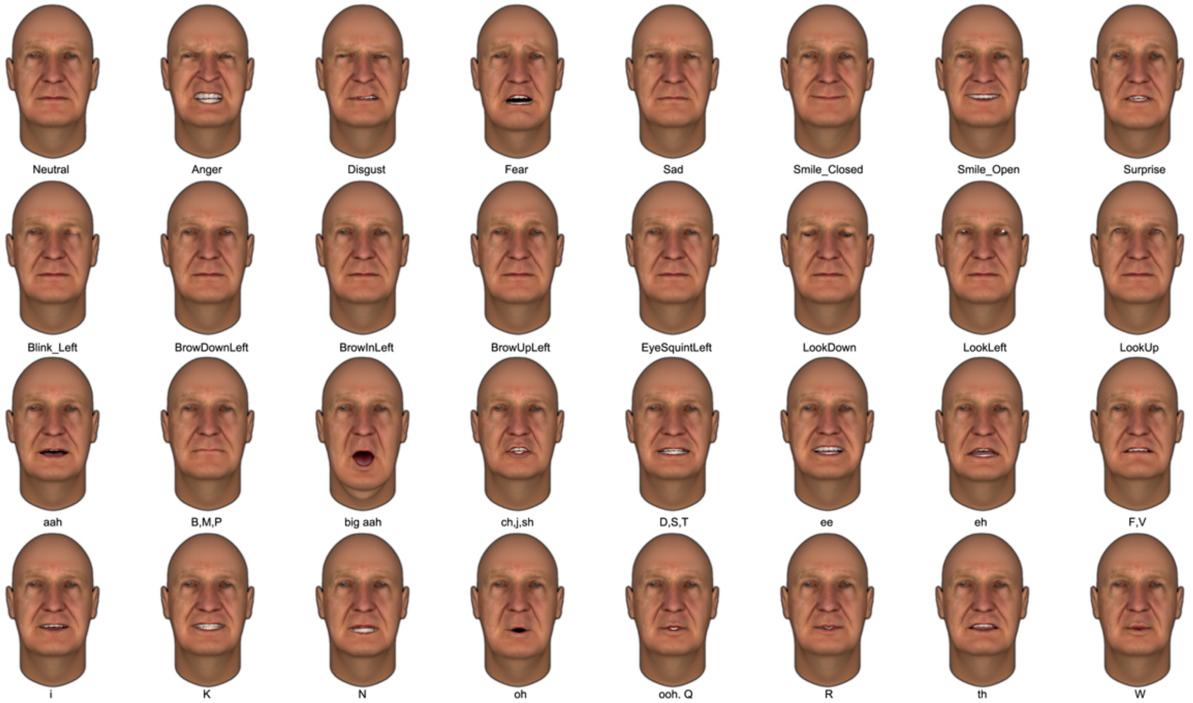


Figure 20. The set of blendshapes for facial expression in LRAF. A blendshape is a static mesh model that Facial Expression Synthesizer controls to realize expression transition. The synthesizer gradually changes a weight value for a target shape to compute an intermediate vertex position from neutral mesh to a target mesh.

Even though the base head model has various facial features and basic emotional expression, it does not guarantee that the model is valid to recreate a target person's emotional state with respect to the user's perception of such emotions. Therefore, a large-scale online user study was conducted to verify this method by comparing implemented result with a real person's emotional facial expression [Lee '10c].

The goals of the study were twofold: determine whether the specific avatar being developed was capable of conveying emotional states; determine, more generally, whether realistic avatars are a good means for conveying emotional states accompanying spoken information. The study used still renderings of an avatar and photos of a human to determine whether users identified the emotional states comparably between the avatar and the human upon which it was based.

The human model of the avatar, Dr. Alex Schwarzkopf, was chosen for photographing. Our work here was based on Ekman's approach to expressing emotions. Photographs were taken of Alex exhibiting six classic emotional states: anger, fear, disgust, happiness, sadness, and surprise. Three photos of each emotional state were selected based on how well they corresponded to the elements of Ekman's emotional characteristics. Images of the avatar were rendered to mimic the photos of Alex as closely as possible by manipulating key facial variables. Subjects were directed to an online survey tool and asked to identify which of the six emotional states shown by the face. Sample cases and result are displayed in Figure 21.

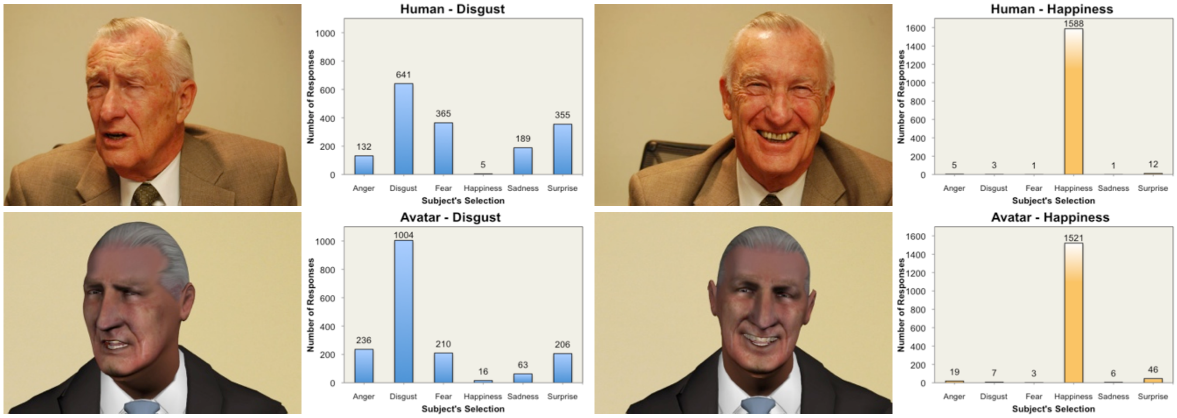
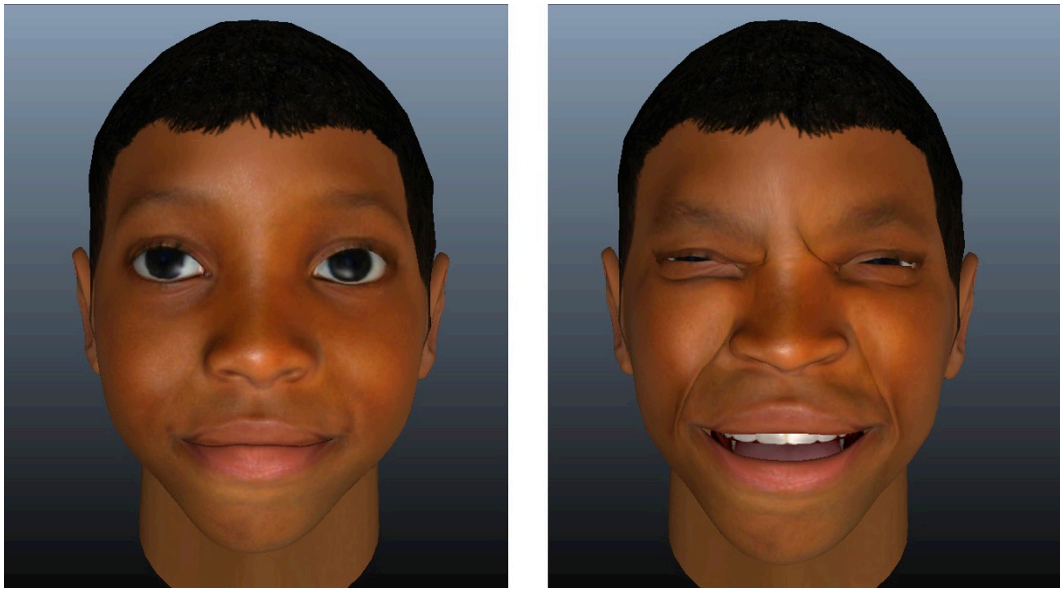


Figure 21. Two sample cases (Disgust and Happiness) from the avatar expression validation study: results indicated that two emotions, happiness and sadness, were successfully identified with high degree of accuracy in both real person's and avatar's. While the avatar did not successfully display the other four emotions, the human photos did not achieve reliable levels of emotional indication either.

Although a large-scale online survey validated the LRAF model's capability to reproduce human emotional expression extremely similar to a real person did, the implementation process to depict a real human expression mostly relied on a designer's manual control of available blendshapes in the LRAF base head model. Therefore, it was necessary to develop more systematic control method to better support an interactive application without much manual intervention. Ekman's FACS offers a well-defined analysis for facial features and its expression to this end. The LRAF base head model was then extended to include more FACS AUs to improve Facial Expression Synthesizer based on the latest version of FACS manual [Ekman '02].

This extension of facial expression blendshapes was deployed in the author's VP study [Lee '13] (to appear). A set of FACS AUs blendshapes from the extended avatar head model was applied to depict more complex facial expressions, such as grimacing, than categorical basic emotions, such as happy, sad, angry, disgust, fear, and surprise. The realized grimacing expression is shown in Figure 22.



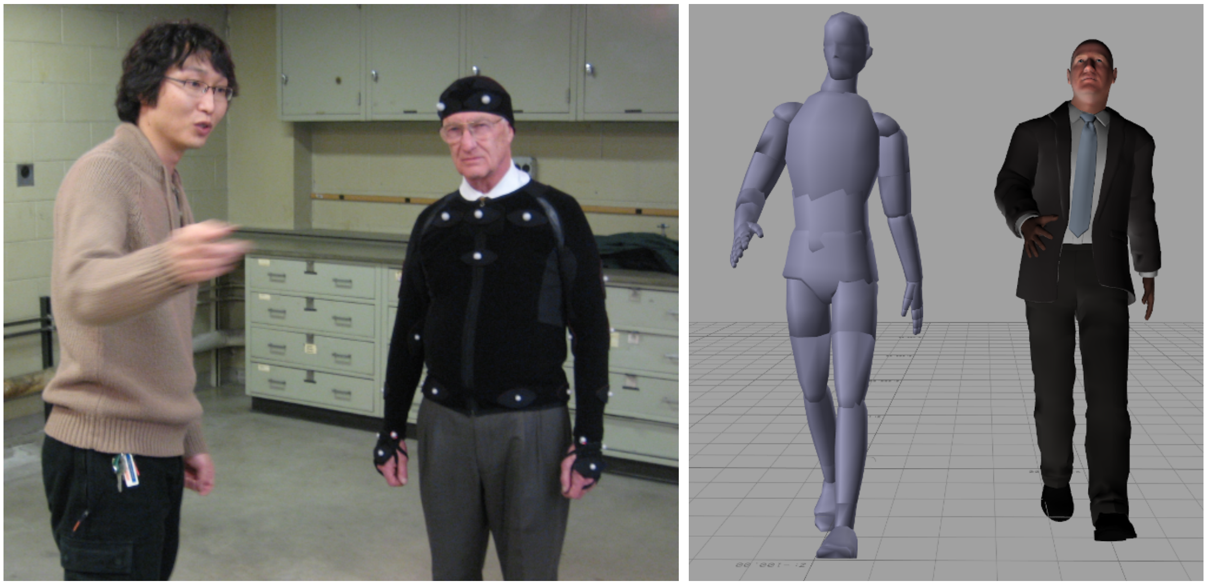
(a) Neutral Face (b) Grimacing Face

Figure 22. Virtual Patient facial expression implemented based on FACS AUs in the extended LRAF avatar head model. In (a), a boy model shows neutral face. In (b), grimacing expression is composed of brow lowerer, cheek raiser, lip tightener, lip stretcher, and lip part. These AUs were selected from [Wilkie '95] and an iterative review process with a certified FACS coder.

4.3 Body Motion Acquisition

As we digitize a person, all behaviors and gestures are based on a target person. We utilized an optical motion capture system, Vicon MX-F40, to acquire full body motion and mannerisms.

Then, a segmentation of the individual motion clip was processed to construct the internal motion database. Each motion clip corresponds to a unit action such as look left / right, pointing, idle, and other actions. The animation module in LRAF selects proper motion clips and blends them in real time. This prefabricated motion database approach is widely used in literature to accomplish realistic body motion along with speech. Figure 23 presents one instance of our motion acquisition process. Later, this data is integrated into the LRAF motion database and used to drive an avatar in virtual space.



(a) Dr. Alex in Mocap Studio

(b) Motion Retarget on Alex Avatar

Figure 23. Performance Capture. In (a), Dr. Schwarzkopf in the motion capture studio was directed to perform a set of gestures that were required to build the motion database in the application. In (b), acquired motion data was processed and retargeted to his avatar model to reconstruct his real body gestures and mannerism.

LRAF avatar animation is a combination of a full body skeletal animation and shape-based facial animation as shown in the LRAF system architecture diagram (Figure 11). The Facial Expression Synthesizer controls verbal and nonverbal expressions together. On the other hand, the Skeletal Animation Synthesizer manages full body animation. Avatars intended to mimic human behavior need to behave somewhat non-deterministically; else they will appear unnatural and mechanistic. To accomplish this, the concept of a Semi-Deterministric Hierarchical Finite State Machine (SDHFSM) was devised, which is a hierarchical finite state machine where a sub-state is chosen either based on a set of constraints or randomly given multiple possible options (Figure 24).

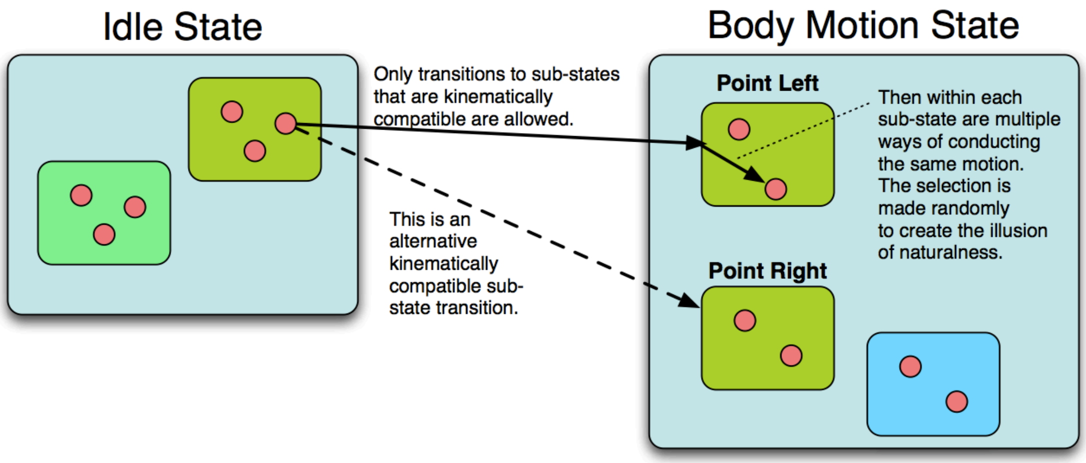

Idle State
Body Motion State
Only transitions to sub-states that are kinematically compatible are allowed.
Point Left
Then within each sub-state are multiple ways of conducting the same motion. The selection is made randomly to create the illusion of naturalness.
This is an alternative kinematically compatible sub-state transition.
Point Right

Figure 24. Semi-Deterministic Hierarchical Finite State Machine (SDHFSM) for motion synthesis. When an avatar changes its action state (i.e. idle) to new state (i.e. point left), SDHFSM picks destination state group from motion database that includes multiple variation of desired action. Then, SDHFSM randomly selects one from this group to best avoid repetitive animation.

In this chapter, we reviewed the avatar visualization details including a modeling process, basic rendering features and advanced techniques, facial expression enhancement, and full body animation for avatar body gestures. Realizing the high quality of avatar visual appearance is a very time consuming task and needs much effort, but a well-defined design process and advanced techniques can help us reduce overall time to develop an avatar-enabled application. Nonetheless, many different avatar models were created in a reasonable amount of time over the course of the LRAF project and the thesis study.

5. AVATAR BEHAVIOR MODELING

An avatar's nonverbal behavior in the thesis is comprised of autonomous emotional facial expressions and head movements such as head nodding and shaking. The facial expression is based on Ekman's FACS and the underlying emotion process is driven by PAD space psychological model. Head gestures that are loosely coupled with internal emotion or less coherent gestures are handled by a data-driven method with a machine learning algorithm.

Supervised Hybrid Expression Control Framework (SHECF) is an extension of the previous work, Lifelike Responsive Avatar Framework (LRAF), to accommodate more natural emotional expression and head movements by a behavior modeling process using rule-based emotion modeling and data-driven behavior prediction. A dialog management module or system is beyond of the scope of this thesis. SHECF assumes that there is an external dialog management module to execute the presented avatar behavior processor.

5.1 Overview of Behavior Modeling Process

The avatar behavior modeling process consists of three pre-processing components: Affect Analyzer; Gesture Predictor; NVB Encoder, and another two realization components; Emotion Processor; Gesture Processor; and Behavior Modifier. Figure 25 represents a function workflow of the modeling process. Pre-processing components in SHECF handle avatar speech utterances to detect an eliciting affect and to generate avatar gesture behavior whereas realization components drive avatar behavior based on processed behavior information through the Gesture

Processor and Emotion Processor. There is an optional probing tool for a designer or an application developer to modify automatically generated avatar expressions and behavior.

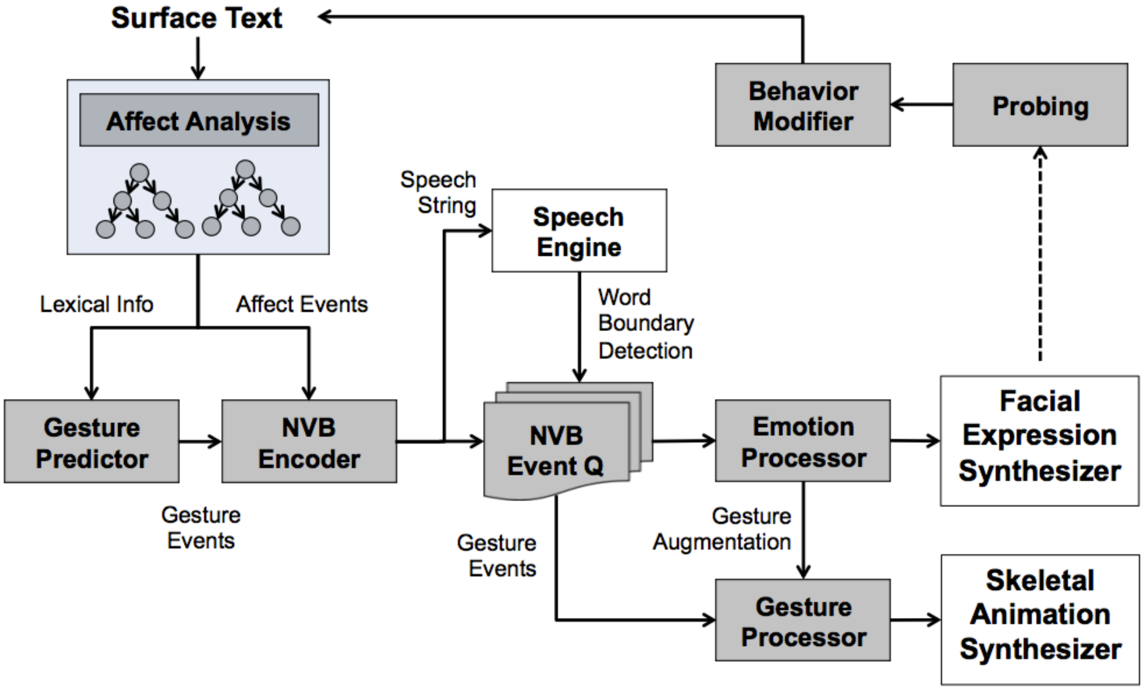


Figure 25. Avatar behavior processor workflow model. Surface text is processed through Affect Analysis module to extract lexical information and raw affect values, and then behavior encoder passes this information to NVB Event Queue to process it synchronously with speech. An event popped from the queue is sent to Emotion and Gesture Processor to augment behaviors. Finally, Facial Expression Synthesizer and Skeletal Animation Synthesizer realize it.

The goal of this processing model is to compute the dynamic avatar emotional state and then to generate emotionally augmented facial expressions and head gesture behavior. At the beginning of application execution, the mood of an avatar is derived from an avatar trait. The processing model assumes that there is no pre-existing emotional state apart from the traits. More details of each individual module will be discussed in the following chapters.

5.2 Affect Analyzer

Affect Analyzer is the first component that receives external stimulus, so called surface text, to extract baseline affect value to control avatar nonverbal behavior. This module is composed of three stages: Part-Of-Speech (POS) tagging, word level affect extraction, and word dependency analysis. Affect Analyzer utilizes a shallow parsing scheme to minimize complexity of understanding utterances. Therefore, it does not require much external knowledge that needs extensive human intervention. A shallow parsing result serves the following two tasks: prepare for information that Gesture Predictor uses to classify head gesture behavior in a later process, and supplement word level affect analysis with associated dependency relationship.

5.2.1 POS tagging

A shallow parsing of utterances in Affect Analyzer starts with POS tagging. The classification of POS tag is based on Penn Treebank notation [Marcus '93]. Affect Analyzer uses the Java implementations of the Stanford Parser package for POS tagging [Stanford '13]. Usages of this library are fairly straightforward. However, calling a function via JNI is relatively slow to get the results back, which is not an ideal situation for a real time application. Therefore, all sub-routines associated with this parsing process are implemented in a separate thread not to delay the main rendering loop.

The results of POS tagging are a string composed of a series of word and its tag associated POS. For example, the tagger processes a sentence, *"My dog also likes eating sausage,"* and the return string is *"[My/PRP$, dog/NN, also/RB, likes/VBZ, eating/VBG, sausage/NN, ./.]"* Then, we tokenize this string to build a pair with word and POS tag for the following analysis modules: Affect Extractor (Chapter 5.2.2) and Gesture Predictor (Chapter 5.3).

5.2.2 Affect Extraction for Words

Affect value for each word in a sentence is processed through a pre-defined affect dictionary, WordNet-Affect [Strapparava '04]. WordNet-Affect is an extension of WordNet Domain. This dictionary offers hierarchical information to associate a-label Emotion, which is a subset of word synsets with affective concepts. *Synsets is a synonym set: a set of words that are interchangeable in some context without changing the truth value of the preposition in which they are embedded* (WordNet glossary). 535 noun synsets are defined in WordNet-Affect 1.1 (refer to Table II). The dictionary size is already relatively limited, it can be expanded by searching synsets group or adding more custom synsets.

	#Nouns	#Adjectives	#Verbs	#Adverbs	Total
Synsets	535	557	200	22	1,314
Words	1,336	1,472	592	40	3,340

Table II. Number of affective synsets and words, grouped by part of speech, in WordNet-Affect [Valitutti '04]

Surface text fed into the Affect Extraction module is decomposed to a series of words first. Then, the search function runs through each word to find associated emotion category in the affect dictionary to build a pair of word and affect category.

For a given word, the search function first retrieves word index id from the dictionary including morph words. Then, it iteratively searches matching Emotion in the WordNet-Affect database for all associated synsets with the word up to the first 5 synsets. This synsets search expansion supplements limited vocabulary in the affect database. Once the search function discovers an affect category for the word then, it assigns a baseline affect value. The last step in the search function is to consider direct augmentation for a certain POS tag such as comparative and superlative for adjective and adverb. In case of a comparative form, 10% increment is applied and 20% is used in a superlative form.

The Affect Extraction module is tested with two emotional stories chosen for a user study in Chapter 6. While reviewing the stories, a total of 25 desired affect words or moments were selected. An initial affect analysis with the developed module successfully detected 12 emotional words. Then, 12 custom synsets were added in the WordNet-Affect dictionary. These custom words are most likely to carry emotional sense but are not found in the dictionary. Further refinement was conducted to achieve the desired goal by injecting 10 more manual emotion entries during Probing process in SHECF. The manually inserted annotations are for: altering detected emotion; add new affect for a word; and add an affect based on context. More details of Supervision control and Probing process is presented in Chapter 5.7.

5.2.3 Word Dependency Analysis

As mentioned, affect values for words are purely based on the meaning of each individual word, but that meaning can be altered by other parts of a sentence. For example, it may mean the opposite affect: “I am happy” vs. “I am not happy”, or may mean greater degree of affect than the simple case: “I am happy” vs. “I am really happy.” Therefore, it is necessary to examine relations between words in a sentence if any affect word exists.

The Stanford Parser package supports POS tagging as well as words dependency graph [De Marneffe '06] shown in Figure 26. For the efficiency of implementation, the Dependency Analysis module also uses the same parser package to retrieve dependency information for a given utterances.

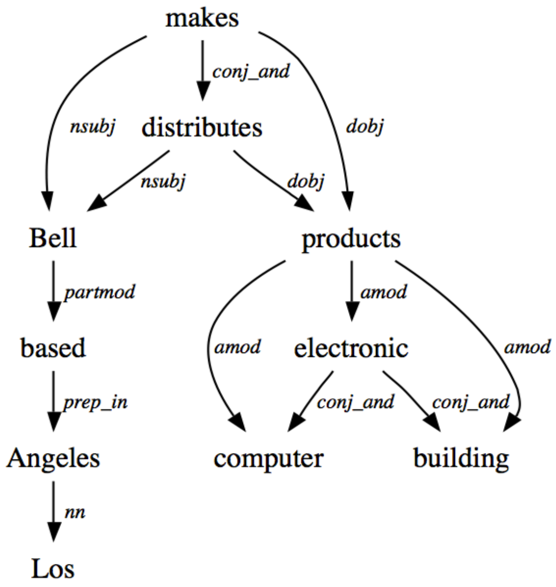


Figure 26. Graphical representation of the Stanford Dependencies for the sentence: *"Bell, based in Los Angeles, makes and distributes electronic, computer and building products."* [De Marneffe '08]

Stanford dependencies are represented as triplets: name of the relation, governor and dependent. For example, a dependency for a sentence, *"I am not happy"*, consists of *"(nsubj, happy, I), (cop, happy, am), (neg, happy, not), (root, ROOT, happy)"*. Especially, the negation relation, *neg*, in this dependency example is critical information to reverse the meaning of word, *"happy"*. More details of definitions of dependency relation can be found in Stanford dependency manual [De Marneffe '08]. Table III summarizes dependency relation rules used in Word Dependency Analyzer.

Relation	Rules
nsubj	**nominal subject**: if the subject is a second person pronoun, decrease by 20% if the subject is a third person, decrease by 60%
neg	**negation modifier**: if governor has affect (sad or happy), then reverse it (sad ⇄ happy) if governor has other affect, remove affect
prepc_without	**prepositional clausal modifier**: remove affect of dependent word
amod	**adjectival modifier**: if dependent depicts greater degree, increase affect value. else, decrease affect value.
advmod	**adverbial modifier**: similar to amod relation case
advcl	**adverbial clause/clausal complement/open clausal complement**: trace back all relation with governor word to see if there is any negation relation, then modify affect in the same way as *neg* relation
ccomp	
xcomp	
conj_and conj_or	**conjunct and/or**: reserve this information for head gesture behavior predictor

Table III. Summary of Word Dependency Analyzer rules. These rules are applied only if either of words in relation has associated affect value from initial Affect Extraction process.

5.3 Gesture Predictor

The objective of the Gesture Predictor in SHECF is to determine whether an avatar shows head gestures while it speaks a given utterances or not. Gesture Predictor is composed of two modules: the Head Gesture Tagger using machine learning techniques; and the Head Gesture Rules engine applying literature-based correlation between gesture and functional elements in speech utterances. This helps a behavior model improve gestural fluency by a hybrid approach that combines two distinct methods. More details of each module are presented in the following sub-chapters.

5.3.1 Head Gesture Tagger

The Head Gesture Tagger uses a machine learning algorithm to predict runtime head gesture behavior with given speech utterances. The scope of head gestures in the thesis is comprised of head nodding and head shaking. A machine-learning algorithm requires well-defined data with feature sets and labels to train a model and then it can compute probabilities for given labels on input data.

The Conditional Random Fields (CRFs) modeling method [Lafferty '01] is chosen to train a model and to classify head gesture labels. CRFs can take contextual information to calculate probabilities. In other words, it can predict sequences of labels for orders of input samples. For example, sequences of input samples are words in the speech utterances and output sequences are labels for each word. C++ CRFsuite library [Okazaki '07] is used to implement the Head Gesture Tagger.

There are two public data sources available to train a head gesture model: SEMAINE video corpus [McKeown '11] and AMI meeting corpus [Carletta '07]. Both database had been transcribed and extensively annotated with various aspects of speaker and listener’s behavior. AMI meeting corpus is a multi-modal meeting corpus including 100 meeting hours that European funded multi-discipline researchers had formed. The SEMAINE database is a large audiovisual database obtained by simulating an avatar interacting with a real user in various situations for over 900 conversations. In this thesis, the SEMAINE database is used because it was collected in the context of a dialogue between two persons, whereas AMI meeting corpus is a multi-party conversation. Conversation between two agents is better suited to model one-to-one interaction between a user and an avatar.

SEMAINE dialog corpus has video sessions, transcripts, some personality information, gestures, and affect. However, it does not provide complete information for all these features.

For example, the total number of head nods and shakes annotation is about 250 from 48 sessions. Initial review found that there were more instances of such behavior than annotated in transcripts. Therefore, two raters re-annotated all dialogues of the 48 sessions for head nodding and shaking behavior for each word level and Inter-rater reliability was 0.812 (Cohen's kappa [Cohen '60]).

Among 48 sessions with complete transcripts and annotation in the pre-processed dialog corpus, 40 sessions were used as training data and the remaining 8 sessions were reserved for model evaluation. For each label, a separate machine was implemented: one for head nodding and another for head shaking.

Selected features to train a machine were POS tag, Phrase type, and Lexical Function type with a trigram model. POS tag is Penn Treebank tag; Phrase type is a phrase that a word belongs to such as noun phrase (NP), verb phrase (VP), and adjective phrase (ADJP); and Lexical Function type is a special function type that the word carries intensifier (INT) and inclusion (INC). The Lexical Function type is inspired by [McClave '00]: a lateral sweep movements co-occurring with intensifying or inclusive words. More details of their work can be found in Chapter 2.3.1. The selected features are likely associated with head gesture behavior in general so that a machine can build internal probability rule sets to classify nodding/shaking label for words in each sentence.

After training nodding behavior and shaking behavior classification models, the remaining 8 session data were used to evaluate their performance. The results showed that nodding classification (*precision=0.605, recall=0.622, F-score=0.612*) performed better than shaking case (*precision=0.289, recall=0.565, F-score=0.382*). These results are also comparable to prior work by Lee and Marsella's nodding model (*precision=0.249, recall=0.378, F-score=0.300)*

[Lee '12]. Since they used different corpus data (AMI meeting corpus) with more features, it is required to investigate further and to fine-tune the model to validate its performance measure in future work.

5.3.2 Head Gesture Rules

Head Gesture Rules rely on widely accepted literature that reveals a stronger relation between a certain functional aspect of word or phrase and our nonverbal behavior. For example, when we express strong affirmative phrases, we likely nod our head in most occasions. Similarly we use head shaking gesture to express a negation. Two more linguistic functions, inclusivity and intensification, that covers mostly one word from [McClave '00] are adopted to use in the Head Gesture Rules. The selected linguistic functions are consistent with features used in the Head Gesture Tagger and can be analyzed by a shallow parsing scheme used in SHECF. Although there are some instances in intensification and additional uncertainty function invoking head gesture behavior covering across a sentence or a longer phrase, those cases are excluded in the thesis due to the limited power of a sentence level analysis. Lastly, a short nodding gesture for listing is added to give the slight intensification for each object in the longer listing phrase based on the observation from SEMAINE video corpus. A total of 5 basic rules of head gesture behavior were used in the Head Gesture Rule tagger. The classification of each linguistic function is extracted from POS tag information, dependency relation, and a custom dictionary. Table IV describes how these supplemental rules assign to head gesture behavior.

Linguistic Function	Rules
Negation	Shaking for three consecutive words (-1, 0, 1)
Affirmative	Nodding for three consecutive words (0, 1, 2)
Intensifier	Shaking (lateral sweep) for two consecutive words (0, 1)
Inclusive	Shaking (lateral sweep) for two consecutive words (0, 1)
Listing	Nodding behavior for one word (0)

Table IV. Linguistic functions and associated head gesture behaviors in Head Gesture Rules module. The index of consecutive words in the rule was chosen to make smooth transition between a NVB head animation and a full body motion. Smaller number of consecutive words triggers shorter animation.

Rule-based head gesture tagging can make up for incompleteness of the data-driven probability model. In particular, when the trained model suffers from low recall rate, the rule-based tagger can increase recall by adding more gestures showing higher causality.

5.4 NVB Encoder

NVB Encoder generates a list of events that describes pre-processed avatar behaviors: emotion and head gestures. As SHECF receives a surface text for an avatar to speak, pre-processing modules evaluate all necessary information before an avatar actually speaks the given utterances. Therefore, the processed information, NVB events, has to be sent to intermediate storage to synchronize all behaviors with audio rendition, voice synthesis in this case. This intermediate storage is the NVB Event Queue in SHECF visualization framework.

Each NVB event has associated word position information so that it is synchronously processed when the visualizer detects word boundary on the fly while TTS renders audio for the utterance. This boundary detection is implemented by catching a speech event from TTS engine. When NVB Event Queue is requested to pop all events for word rank *n*, the events are popped until the word position of next element is greater than *n*. Depending on the types of event, it is sent to either the Emotion Process or the Gesture Process to realize avatar behavior.

5.5 Emotion Processor

Affect computation in the Emotion Processor takes three stages similar to the layered approach proposed by [Gebhard '05] to produce momentary emotional state (short-term) for expression and to accommodate accumulative mood (medium-term). Both emotion and mood are also affected by an individual's trait (long-term). The computational model of the Emotion Processor is based on Mehrabian's PAD space [Mehrabian '96b].

Figure 27 explains components and its flow model of the Emotion Processor. First, external stimulus, pre-process affect and its value, is passed to the appraisal module to convert it into a PAD vector. Next, it applies traits and current mood to augment received stimulus. Finally, the Emotion Processor outputs the avatar's emotion as well as the avatar's mood that feed back to accumulate its short-term affect.

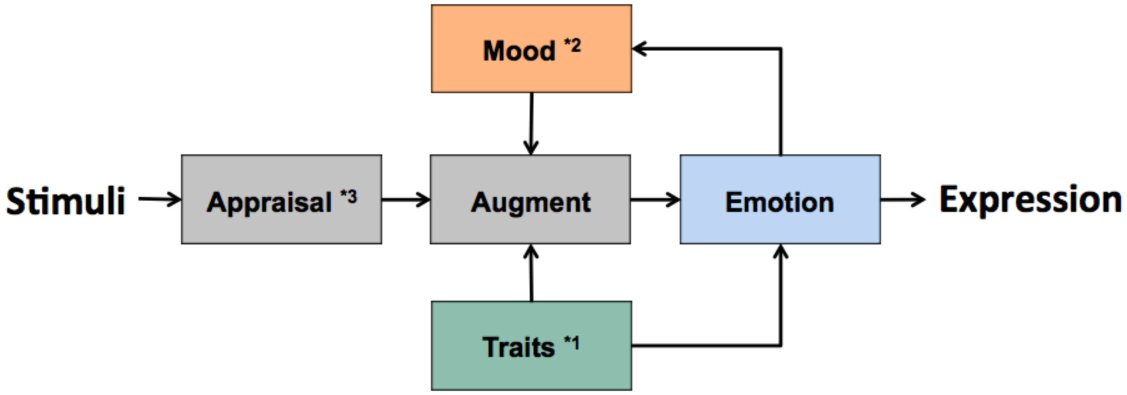


Figure 27. Components of Emotion Processor: *1 is big Five Factor Model of traits converted in PAD space; *2 is PAD space mood computation module; and *3 is an initial appraisal of event in PAD space vector.

Table V shows the PAD vector for six basic emotion categories. These values are adopted from prior works [Becker-Asano '09; Gebhard '06; Zhang '10] and are used to compute the initial PAD vector in the Appraisal module.

Category	Pleasure	Arousal	Dominance	Mood Octant
Angry	-0.8	0.8	1.0	-P+A+D
Disgust	-0.4	0.1	0.1	-P+A+D
Fear	-0.8	0.8	-1.0	-P+A+D
Happiness	0.8	0.8	0.2	+P+A+D
Sadness	-0.5	-0.2	-0.3	-P-A-D
Surprise	0.2	0.8	0.1	+P+A+D

Table V. PAD appraisal values used to convert affect value to 3D PAD vector for basic six emotion categories.

For consistency in the computation model, the FFM traits model specified in the avatar XML specification file is converted to PAD space by Equation (1) [Mehrabian '96a]. Since traits are long-term factors that are not likely change over application lifetime, this conversion initializes an avatar's emotional state in PAD space at the beginning of application execution.

$$\begin{aligned} P &= 0.21 \times Extraversion + 0.59 \times Agreeableness + 0.19 \times Neuroticism \\ A &= 0.15 \times Openness + 0.30 \times Agreeableness - 0.57 \times Neuroticism \\ D &= 0.25 \times Openness + 0.17 \times Consciousness + 0.60 \times Extraversion \\ &\quad - 0.32 \times Agreeableness \end{aligned} \quad \ldots(1)$$

Once an avatar is set with initial conditions for traits and mood, an appraised stimulus is calculated by considering those two factors together. Equation (2) describes how an initial stimulus PAD vector, e, is evaluated. With traits effect constant factor c_t, trait T vector will be applied to the length of e, $\|e\|$, to compute the effect of the trait. This works as percentage of e to augment upon *T*. Then, the accumulated mood vector that is a sum of all active emotions, E_{active}, is added with the mood effect constant factor c_a.

$$E_{evaluated} = e + c_t \times \|e\| \times T + c_a \sum E_{active} \qquad \ldots(2)$$

Then, the Emotion Processor updates current mood, $M_{updated}$, with the traits vector, the sum of active emotions, and the newly evaluated stimulus vector as described in Equation (3).

$$M_{updated} = T + E_{evaluated} + \sum E_{active} \qquad \ldots(3)$$

Each emotion in the Emotion Processor has decay and intensity parameters so that momentary emotional state changes over time to simulate natural decay of our elicited feelings. If there is no further external stimulus changing avatar's feeling state, its emotional state will eventually go back to initial condition of mood as specified in traits model.

Finally, the intensity for an evaluated emotion, $E_{intensity}$, is calculated by Equation (4). $e_{intensity}$ denotes the initial intensity value from Affect Analyzer. $\|E_{evaluated}\|$, the ratio of the length of the evaluated emotion to the length of initial emotion, $\|e\|$, is an augmentation factor for the intensity. This ratio is applied to $e_{intensity}$ to reflect the change caused by current mood. Then, the last term in Equation (4) is the auxiliary effect to account for repeated eliciting emotion defined as $R(e)$. The function $R(e)$ traces the last three emotions in the mood accumulator and compares them with newly received emotion, *e*. If there is the same emotion (same category of emotion) found in these three, especially the immediate adjacent one, $R(e)$ generates positive value to boost the intensity. Otherwise, $R(e)$ does not change the intensity.

$$E_{intensity} = e_{intensity} \times \frac{\left\|E_{evaluated}\right\|}{\left\|e\right\|} + R(e) \qquad \ldots(4)$$

Each evaluated emotion, $E_{evaluated}$, and its associated intensity value, $E_{intensity}$, is sent to the Facial Expression Synthesizer to visualize avatar facial animation.

SHECF uses FACS facial expression convention to realize received affect from the Emotion Processor. Figure 28 presents one example deployed in SHECF expression database. Since we modulate the expression database template, one can add more custom FACS templates for basic emotions to achieve better fluency and diversity in facial expression.

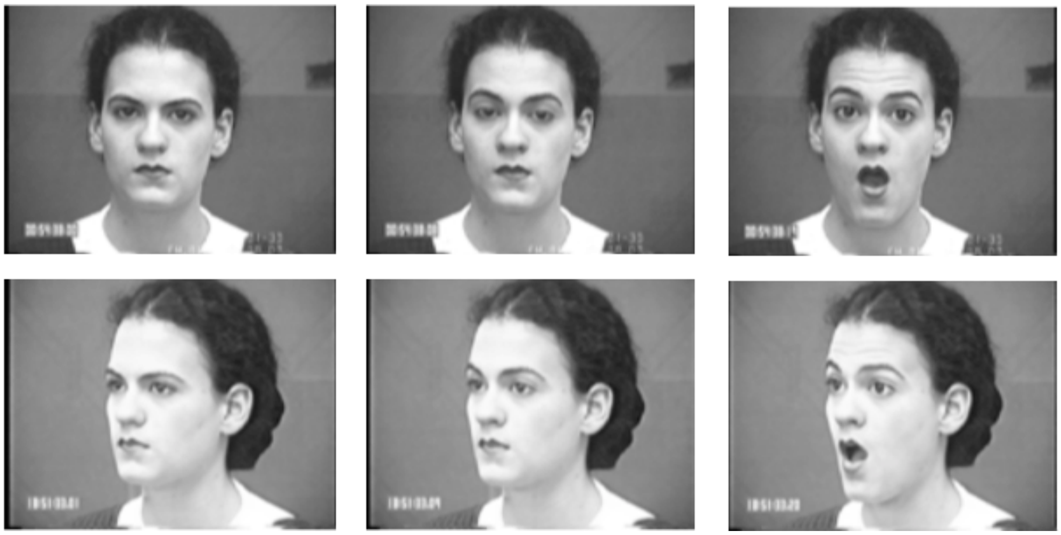


Figure 28. Surprise Expression: frontal and side view of surprise expression from neutral to peak level. FACS AUs are composed of 1+2+5+27: Cohn-Kanade DB includes expression by over 97 posers [Kanade '00]. SHECF expression DB has 327 variations for 7 emotion categories extracted from Cohn-Kanade DB.

5.6 Gesture Processor

The Gesture Processor receives a request for gesture generation from the NVB Event Queue when it detects a word boundary event from the speech engine and there is a head gesture related event in the queue. This request includes information with a type of head gesture, its pre-processed intensity value, and the number of consequent words associated with this gesture behavior. Then, the Gesture Processor calculates emotionally augmented intensity for the gesture.

The notion of an emotionally augmented head gesture comes from psychological understating of human behavior with respect to his/her emotional state. For example, when someone feels very depressed or sad, he/she may not show much head movement. Instead, he/she expresses very slight motions while he/she is speaking or interacting with others. This emotional aspect of head gesture was well described by Cowie et al [Cowie '10]. My thesis used their same corpus database to analyze head gesture, nod and shake, behavior and to provide a basis for future research. Their review results indicated the relationship among the energy of gesture, the intensity of head movements, and the arousal value from observed person's mental state. Equation (5) describes how the arousal value from current avatar mood changes the intensity of head gestures.

$$
\begin{aligned}
&;\text{for negative arousal}\\
&\quad E_{gesture} = M_{Arousal} \times 0.25 + 0.5\\
&;\text{for posive arousal}\\
&\quad E_{gesture} = M_{Arousal} \times 1.5 + 0.5 \qquad \ldots(5)\\
&;\text{ given intensity of gesture augmented by } E_{gesture}\\
&\quad I_{gesture} = I_{gesture} \times E_{gesture}
\end{aligned}
$$

When $M_{Arousal}$, the current Arousal value in Mood is negative, then the energy affecting gestures, $E_{gesture}$, becomes 25%~50% upon Arousal value ranging from -1.0 to 0.0. Then, it augments the intensity of gesture value, $I_{gesture}$, by multiplying computed gesture energy. In case of positive Arousal value, gesture energy factor, $E_{gesture}$, will increase the intensity of gesture up to 200% at most. Constant factors in Equation (5) were determined during preliminary testing and pilot study.

Emotionally augmented head gesture type and its intensity value will be sent to the Skeletal Animation Synthesizer in SHECF to merge together with full body animation. Each NVB head gesture animation was extracted from some motion data collected over the course of the LRAF project. Those animation clips include only head motion data. Therefore, the Skeletal Animation Synthesizer merges motion data from the NVB head animation DB with existing and currently playing full body motion by blending two separate piece of data. The intensity of head gesture defines maximum blending factor, which is the weight value for the animation track, for the NVB head animation when this blending occurs. The current NVB head motion database has 6 different types: two different behaviors with three different length: short, medium, and long continues. Each type of motion data has multiple instances so that SHECF can avoid unnatural repetitive motion synthesis.

5.7 Behavior Modifier: Supervision Control

Supervision Control in SHECF is designed to supplement possible imperfect results from an automated model. There are three example scenarios here: (1) a user defined personality trait may not show clear distinction compared to what a designer expected, (2) computed emotion may not be desirable in some cases, or (3) the system may create the wrong interpretation for a given utterances due to limited capabilities of the current language processing algorithm or database.

Supervision Control takes place between the SHECF affection/behavior processing modules and avatar realization routines. An application designer/developer can simulate avatar speech along with expression, and then customize its result by adding annotation to manipulate or to override automated result. Such custom annotation will be encoded inside speech utterances as a simplified xml tag so that the Affect Analyzer properly detects and will override the evaluated result on the fly in later execution. Since the personality trait parameters are a fixed template, any changes made on trait factors will be stored in avatar template file. In other words, as different from common annotations affecting temporal behavior, traits model is applied to avatar's global behavior. Therefore, it is recommended to review the changes made by traits carefully not to overly affect on other results that are undesirable.

The current Behavior Modifier supports three types of custom annotation for affect computation: inserting, neutralization, and modification. Each case is explained below.

S1 Insertion *(false negative error correction)*

If the SHECF Emotion Processor fails to extract a certain affect from a surface text, a designer can add affect annotation directly into speech text. For example, Emotion Processor does not extract any affect from a given utterance, *"I am puzzled by this"*, because there is no affect-sensitive word found in the WordNet-Affect dictionary for any word in the sentence. However, "puzzled" may have some surprise feeling associated with. Then, we can easily add new affect in front of that word as following.

"I am <emo id="SURPRISE" value="0.2"/>puzzled by this."

S2 Neutralization *(false positive error correction)*

If the SHECF Emotion Processor reports incorrect affect due to incomplete analysis power or misinterpretation of a sentence structure, a developer can annotate the associated word with zero valued affect tag to neutralize this affect. In other words, this neutralization will get rid of computed affect in the Affect Analyzer. For example, a speech text, *"I am not going to be poor"*, is analyzed with Sad affect for word *"poor"* in Affect Analyzer. Although that word has sadness, the meaning of the sentence does not carry much of sadness. It may be either neutral or slight happiness for future expectation. This false positive error occurs due to the limited power of context analysis in the current model. Then, we can annotate the word with neutralization tag.

"I am not going to be <emo id="NONE"/>poor."

S3 Modification

This case is similar to *S2 Neutralization* except *S3 Modification* is aimed to change computed affect with a given desired affect and its intensity. For the previous example in S2 Neutralization case, if the surrounding story context suggests that affect state more leans toward happiness, we may want to modify it as following.

"I am not going to be <emo id="HAPPY" value="0.3"/>poor."

Although current SHECF Supervision Control provides a facility to manipulate affect only, its simple implementation scheme can be easily extended to other behavior like head gestures or even body gestures. Even though its usage is simple and easy, the thesis found that more intuitive tools apply those controls effectively and seemed desirable. For example, it is feasible to implement an easy to use GUI toolset such as trait sliders, drag and drop style expression icons, and emotion adjustment sliders. The future research is planned to extend the current Supervision Control framework to include such tools.

In this chapter, we reviewed the details of the avatar behavior modeling method including an affect analysis process, a gesture prediction model, emotionally augmented affect and head gesture generation techniques, and a facility to probe and modify realized behaviors. Designing a computational model to explain complex human behavior requires enormous amount of knowledge and consideration in many aspects, and it may even be an intractable task. However, research efforts to maneuver such a complex problem can contribute to improve current limitation and to get us close to the ultimate goal. The behavior modeling method in SHECF has been implemented and demonstrated a new novel approach to accommodate our emotional affect in avatar nonverbal behavior for both facial expression and head gestures. Design approaches and implementation details studied in this thesis can be used as a guideline for the future research.

6. EVALUATION

How do people express their emotion? Do we always show the same amount of expression regardless of our feelings? How should we control the strength of emotional expression for an autonomous avatar? Does different intensity of expression conveys one's mental state to us better than a simple fixed factor model?

This chapter provides some answers to these questions by presenting the results of a user study where people rate various avatar behavior models that are composed of the presented dynamic avatar behavior control scheme and fixed factor models with respect to its perceived naturalness in both facial expression and head gesture behavior.

6.1 Introduction

In order to evaluate whether the presented model can increase perceived naturalness of an avatar behavior or whether people prefer a dynamic modeling method to a fixed behavior realization scheme, an empirical study in which subjects were asked to review 7 different avatar models that are identical in graphical representation but are varied by their behavior was conducted. Each avatar told an emotional story and subjects were asked to rate their perceived naturalness of each model. For the purpose of the experiment, three stories were selected from psychological literature (therapeutic healing story) and web-blogs (personal stories). One story was used as a training set and the other two were used to collect data. Approval for this user study was obtained from the Institutional Review Boards (UIC Research Protocol # 2013-0521. Refer to the approval letter in Appendix).

The hypotheses for this empirical evaluation are *(H1)* the presented model can simulate an emotional state similar to what people depict based on a given story, *(H2)* the dynamic control of the intensity of emotional facial expressions is preferred to a fixed control condition, *(H3)* the dynamic control of the intensity of the head gesture model receives better naturalness than the control case, and *(H4)* combined dynamic facial expression and head gesture model is superior to the other conditions.

6.2 Experimental Method

The experiment followed a repeated-measures design with seven combined conditions for each story that conveys a primary emotion. Each of the seven conditions are described as follows.

- Mx-Mg: Model Expression & Model Head Gesture
- Mx-Cg05: Model Expression & Control Head Gesture intensity 0.5
- Mx-Cg01: Model Expression & Control Head Gesture intensity 0.1
- Cx05-Mg: Control Expression intensity 0.5 & Model Head Gesture
- Cx05-Cg05: Control Expression intensity 0.5 & Control Head Gesture intensity 0.5
- Cx05-Cg01: Control Expression intensity 0.5 & Control Head Gesture intensity 0.1
- Cx05a-Cg05: Control Expression intensity 0.5 abrupt transition & Control Head Gesture intensity 0.5

6.2.1 Data Preparation

For the purpose of experiment, three stories were selected from a psychological literature and web-blogs. Each story is a personal experience from the past and has engaging emotions. The original stories were slightly adopted to be first-person narrative so that attached emotional feelings can be conveyed to subjects without much effort to imagine how others' feelings would project to a speaker, an avatar in this case. Two stories used as the test set in the study are described below.

Story I: "My Minxy Mouse Encounter: DISNEY WORLD"

> "Growing up in Texas, as a family of 6, we would drive everywhere for our family vacations. Florida was a popular destination as we could go to the beach or Disney World after a couple of days driving. I'll be honest with you; I was a big fan of both of these vacation destinations. Florida is still one of my favorite places because of my childhood memories and fondness for the state. This favorite travel memory is about me finally meeting Minnie Mouse. Even though we went to Disney as a family on several occasions, I somehow, after multiple visits, never seemed to be able to find her, shake her paw, and get that precious photography opportunity. It was an inside joke within our family and an obsession of mine. I'd see her in the distance running off with Mickey and not be able to catch up. Or she'd be swarmed by a crowd and need to take a break from the heat of the Florida sun just as I was about to have my turn. It was ridiculous the luck I had with that sassy mouse!
>
> Many years later, my mother suggests we have a girl's weekend at Disney. My mom, my sister, and me. OK, sign me up. Here's my chance again! Going back to Disney as an adult was great. It brought back a flood of happy childhood memories and how magical the place can actually be. I found myself teary on many occasions. And I no longer had to chase that minx of a mouse all over the park. At this point, Disney had wised up and had the popular characters on continuous rotation over at the Toontown Fair. I waited patiently in line over 30 minutes at least. They let in groups of 10, and then you go one at a time to get your photography opportunity. I was so excited. I would be the first in our group of 10. It was finally my moment and unscripted I said, "I've been waiting all my life to meet you!" What? This is a teenage person inside a mouse costume. I said to my mother "you better not mess this photo up!" My mom broke down laughing and said that the whole experience of me and my mouse encounter made the trip worth the expense. I proceeded to leave the meet and greet tent and buy myself a princess hat because, after all, I felt like one. The rest of the day little girls all over the park asked for my autograph!"

Story II: "Poor No More"

> "I grow up in grinding poverty. It is very painful for me and I hate it. Around the age of 13, it really gets to me and I swear to myself that when I grow up, I am not going to be poor. Twenty years later, I am made my vow come true. I have fifty million dollars in the bank, I am making money hand over fist, I have got yachts, cars, houses on the Riviera, and more. But I am not happy. I am puzzled by this. I think to myself, "What's wrong here?" It can't be that I haven't reached my goal. I swore I would not be poor when I grew up, and I'm not. I've escaped poverty and put fifty million dollars distance between it and me. It can't be that I don't have the things that money can buy, since I have everything I ever dreamed of and more. Why am I not happier? Why am I not getting the sort of satisfaction out of all this that I think I should be getting?
>
> One day the answer comes to me. I realize that my entire life has always been, and continues to be, centered around not being poor. That where I have always been coming from is escaping from that hated, painful condition of my childhood. And, coming from there, the best that I can ever do is to be poor no more by avoiding that awful condition. Even if I succeed entirely, what I succeed at is avoiding that negative thing, not anything positive. I can never be a success. I can only avoid failure. I can never be rich. I can only be poor no more. If this is the significance I attach to things, the primary emotions that I can experience can never be the elation or satisfaction that come with a positive achievement. They can only be the relief that comes when I escape and put ever-increasing distance between poverty and myself. I am in essence spending my life not being some way and running from that way, and that is a losing game in which the best I can ever do is break even. I can never succeed." [Ossorio '76]

The first story is mostly composed of happy memories. The speaker recalls wishes to meet Minnie Mouse at Disney World and finally encountered her full of joy although there were many unsuccessful visits before then. We can easily determine speaker's gender from the context itself with some distinct words such as "*girl's weekend*." Therefore, A female avatar model was decided to use in the experiment.

The second story invokes negative feelings such as depressed and sad. I chose this one as sad story to contrast the first one. Although this story does not have any strong gender specific words or context, the same female avatar model was used for the second story to avoid any model specific bias in the experiment.

Then, the selected stories were processed with the presented affect analysis model and head gesture generation algorithm described in Chapter 5. This whole process can be done on the fly because the framework is designed to drive an interactive autonomous avatar application. However, data sets used in the experiment were a series of offline videos that were captured from a real-time application via Fraps software [beepa '13] at 60 FPS with a medium close-up shot. The audio track was excluded intentionally to avoid any emotional bias caused by audio cues while subjects reviewed it. Instead, sub-titles are provided in the bottom of each video so that a subject can keep track of story progress. Videos for the seven conditions were made separately and merged together in a video-editing tool to create one large video including all 7 variations. The order of each condition in the final video was randomized to prevent any order bias in the experiment. Figure 29 shows this composite video.

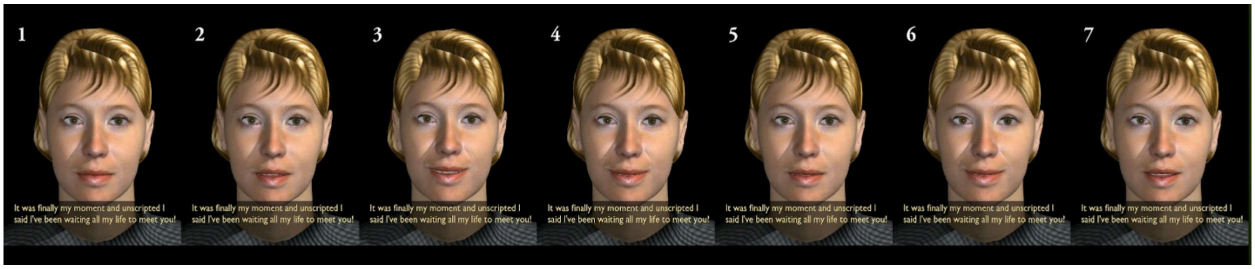

1
It was finally my moment and unscripted I said I've been waiting all my life to meet you!
2
It was finally my moment and unscripted I said I've been waiting all my life to meet you!
3
It was finally my moment and unscripted I said I've been waiting all my life to meet you!
4
It was finally my moment and unscripted I said I've been waiting all my life to meet you!
5
It was finally my moment and unscripted I said I've been waiting all my life to meet you!
6
It was finally my moment and unscripted I said I've been waiting all my life to meet you!
7
It was finally my moment and unscripted I said I've been waiting all my life to meet you!

Figure 29. Seven conditions of avatar videos used in the experiment. To preserve best details of avatar visualization each video was captured at 960x1080 resolution. Then, merged into a single video at 4,096x768 resolution.

The length of the avatar video for the first story was 2:52 including 5 seconds lead-in before it started speaking and 5 seconds lead-out after it finished speaking. The presented affect analysis model detected 14 emotion eliciting words or phrases (12 Happy and 2 Sad) and the NVB generator parsed 32 head gestures (10 Nods and 22 Shakes). The PAD value was calculated based on the Mood Equation described in Chapter 5.5 and the result is shown in Figure 30.

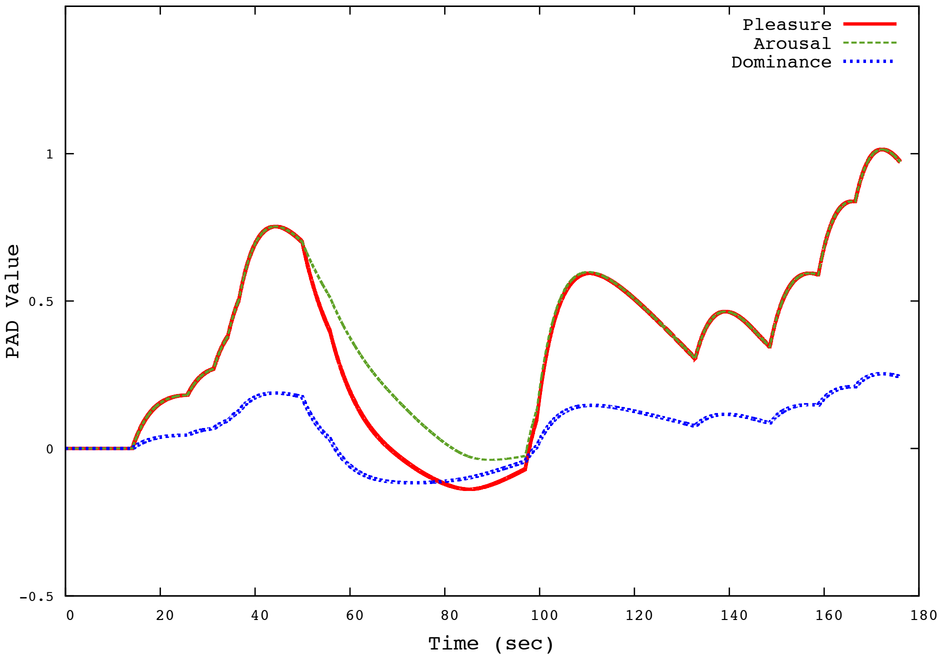


Figure 30. *Story I* PAD graph. 12 *Happy* emotion eliciting stimulus had Pleasure and Arousal value increase throughout the story. 2 *Sad* vectors appeared around 49 and 55 seconds and caused a steep decline in Please and Arousal values. The second decreasing point was between 110 to 132 seconds when all active PAD vectors started to diminish.

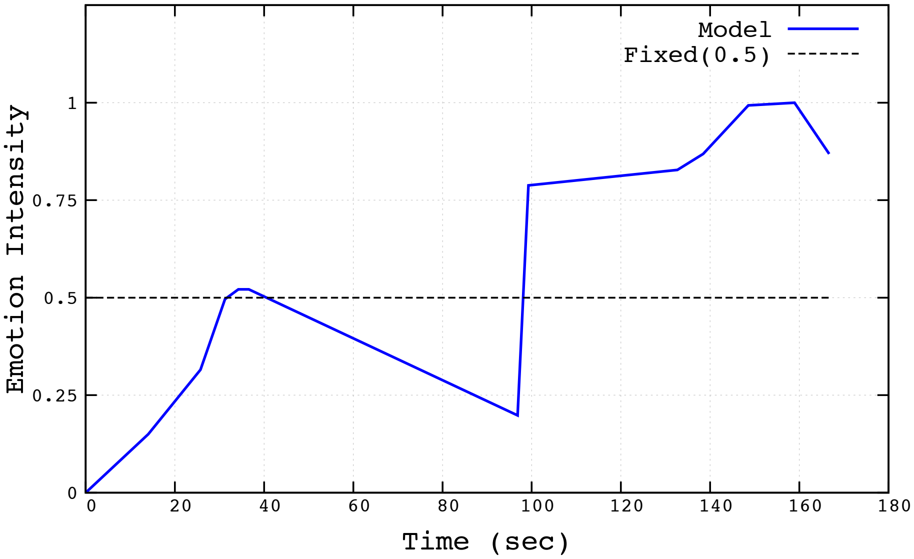


Figure 31. Happy Emotion Intensity graph for *Story I*. Model denotes a dynamic peak value of *Happy* emotion calculated by the presented affect model whereas Fixed shows a control case with intensity 0.5. Decreasing pattern found in time line from 40 to 100 seconds was caused by 2 *Sad* stimuli (mood and repeated emotion factor decreased).

The mood computed from accumulated PAD factors also changed the intensity for an avatar head gesture as described in Chapter 5.3.2; the Arousal value of the mood determines the augmenting factor for the gesture. As observed in Figure 30, *Story I* had higher positive Arousal value except a small part of the story and its computed augmenting factor ranged from 0.5 to 2.25 (Figure 32).

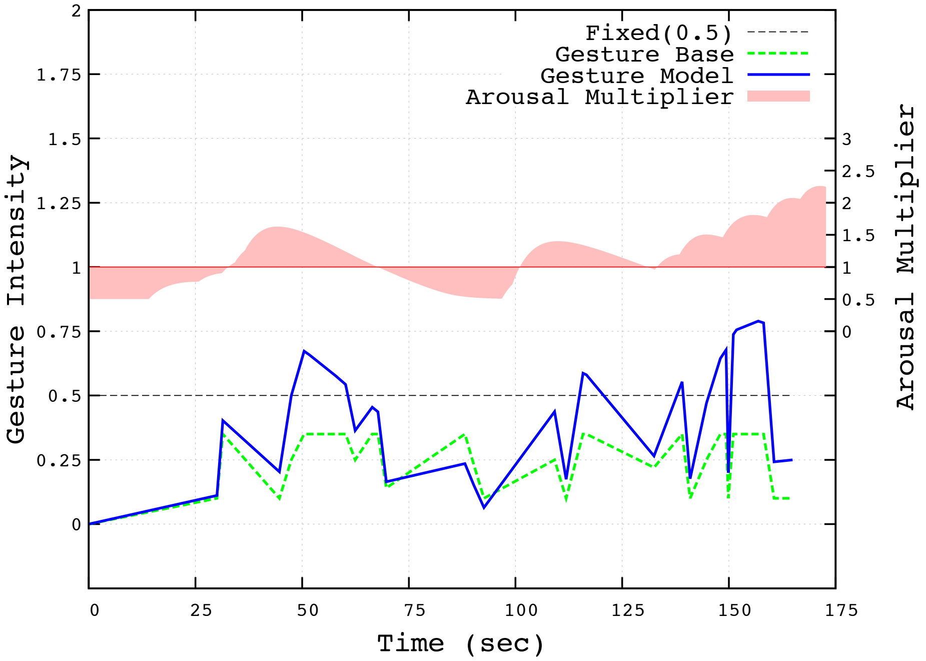


Figure 32. Head gesture intensity graph for *Story I*. Base value denotes the intensity calculated from the context of utterances and associated modifiers. Arousal multiplier indicates the PAD augmenting factor. At 50 seconds point, base intensity (0.35) was multiplied by Arousal factor (~1.5), and then became model intensity (~0.525).

The length of the avatar video for the second story was 2:29 including 5 seconds lead-in before it started speaking and 5 seconds lead-out after it finished speaking. The affect analysis model detected total 12 emotion eliciting words or phrases (10 Sad, 1 Fear, and 1 Surprise) and the NVB generator parsed total 44 head gestures (23 Nods and 21 Shakes). The PAD value for the second story is shown in Figure 33, the intensity of the primary emotion in Figure 34, and the intensity of head gesture in Figure 35.

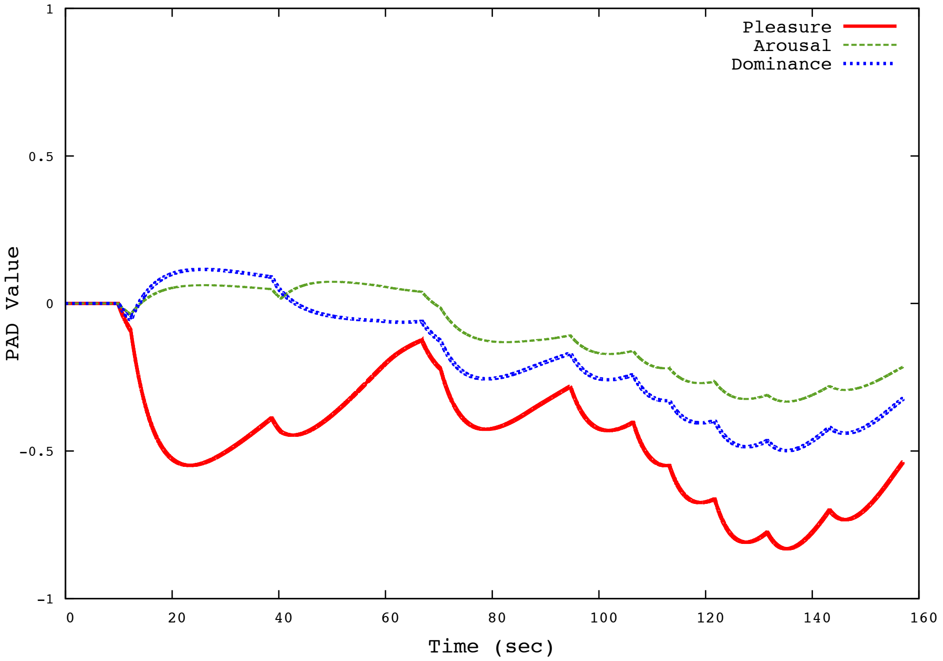


Figure 33. *Story II* PAD value graph. 10 *Sad* emotion eliciting stimulus had Pleasure value decrease throughout the story. The *Fear* emotion is different from *Sad.* However, its PAD vector has similar to *Sad* and didn't affect much on an avatar mood other than slight increase in Arousal (occurred at 12 seconds point). *Surprise* stimulus at 40 seconds pushed Arousal value up again.

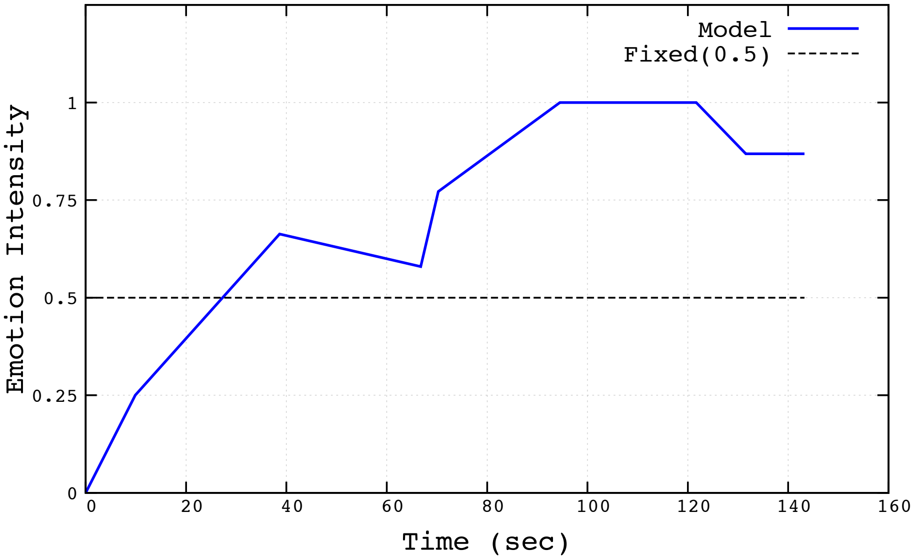


Figure 34. Sad emotion intensity graph for *Story II*. Model denotes a dynamic peak value of happy emotion calculated by the affect model whereas Fixed shows a control case with intensity 0.5. Decreasing pattern from 40 to 65 seconds was caused by *Surprise* stimuli.

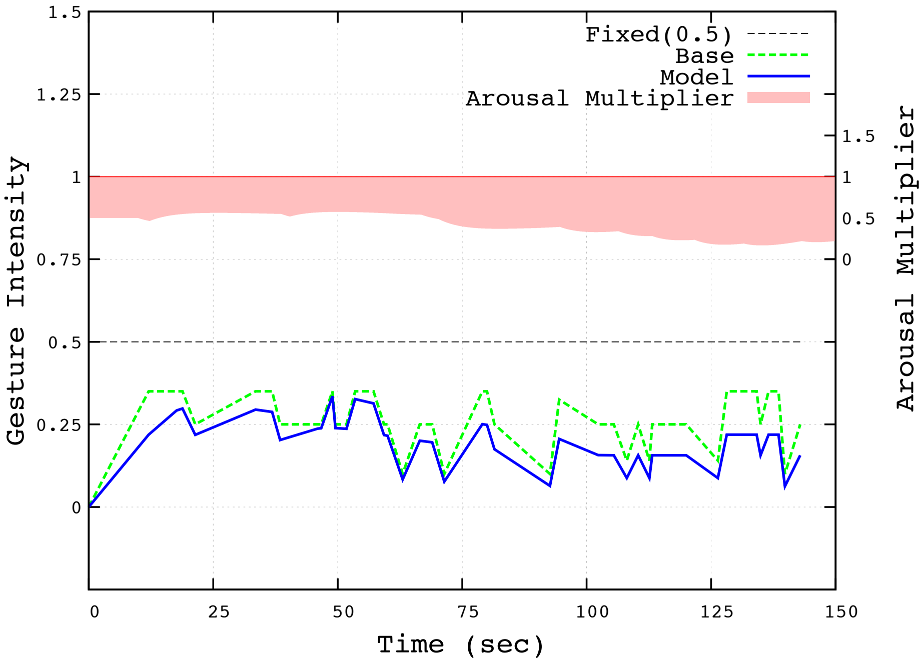


Figure 35. Head gesture intensity graph for *Story II*. Base value denotes the intensity calculated from the context of utterances and associated modifiers. Arousal multiplier indicates the PAD augmenting factor. In *Story II*, the avatar had very low Arousal due to many *Sad* emotions arose in the story. As consequence, the intensity of gesture became lower than the base line value.

6.2.2 Apparatus

The large-scale virtual reality environment called the CAVE2™ system [CAVE2 '13] and graphics middleware, Scalable Adaptive Graphics Environment (SAGE) [Jeong '10], was deployed for this empirical evaluation. To properly review very subtle facial expression and changes in avatar behavior, it is important to preserve all detailed information and to provide avatar visualization in a reasonable size on a display. Conventional computer monitors or TV screens are too small to show seven avatars on a single screen. A projection screen can display larger image than monitors but is limited to its resolution. CAVE2™ offers much larger scale seamless screen space with high resolution that benefits the review process of subjects in the experiment.

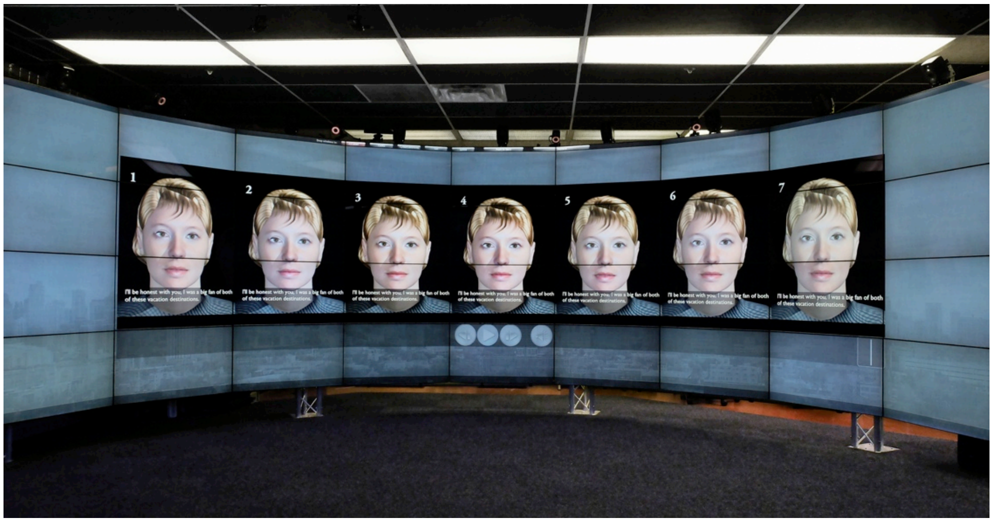


Figure 36. CAVE2™ system showing avatar video. Each avatar is displayed on a single column to minimize interference from the bezel between LCD panels. A distance from the center of CAVE2™ system to the screen is approximately 12 feet.

CAVE2™ is composed of 72 tiled LCD panels (4 rows by 18 columns) forming cylindrical space. Seven conditions of the avatar video were presented within seven columns of display array to fit one avatar in one column of displays. Figure 36 shows CAVE2™ presenting an avatar video for the experiment.

The position of avatar videos was captured via the SAGE session saving feature in advance so that the video appears in the same position for the entire experiment. Subjects were allowed to freely move around during their review process either to see as many avatars as possible at one point or to examine details of each individual avatar from a close distance (Figure 37).

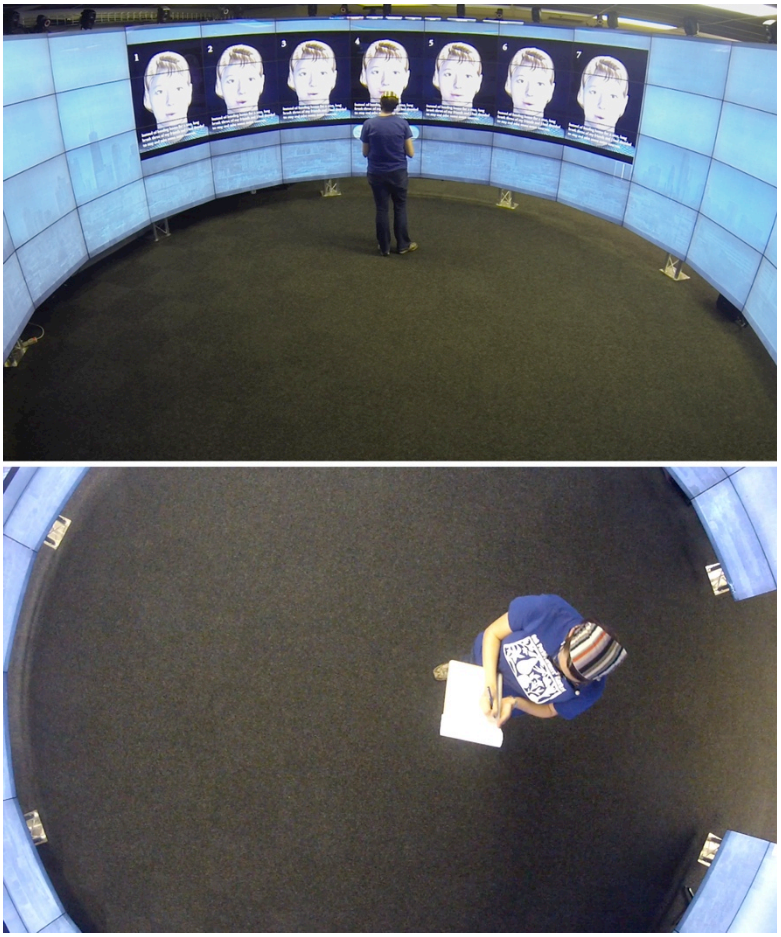


Figure 37. A subject is reviewing an avatar video in CAVE2™. Within 24 feet diameter of CAVE2™ system, a subject moves freely to compare some or all conditions, or to examine details of one condition.

6.2.3 Procedure

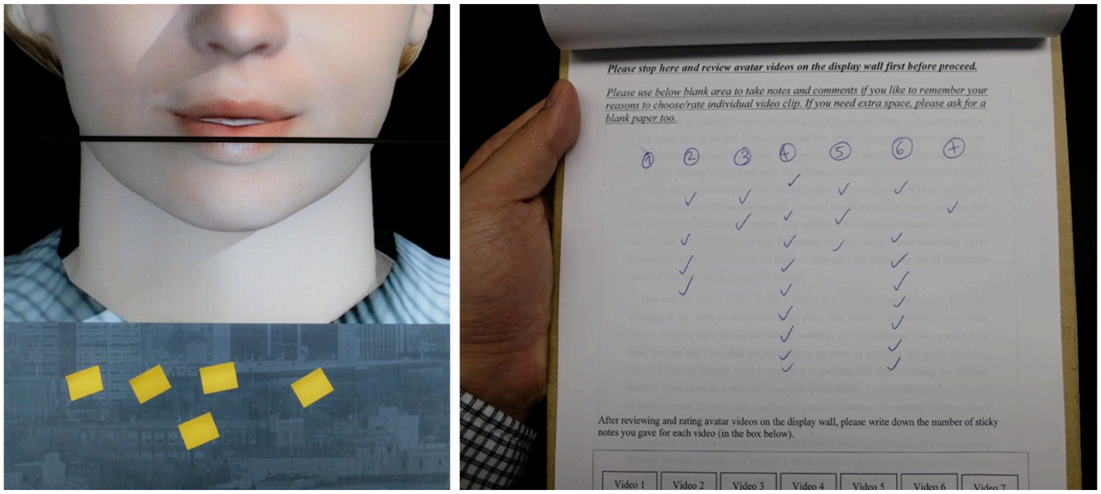
Subjects were recruited from UIC computer science major students (both graduate and senior undergraduate students). A Total of 24 subjects participated in the study and the average age was 32.61 (*SD 11.25*). Among the 24 subjects in the study, 75% were male, 66.67% identified their primary language as English, and 33.33% non-native speakers.

After a subject signed a consent form to participate in the study, a brief procedure of the experiment was explained. The experiment was composed of three sessions. The first session was a training set for the individual subject to get familiar with the study and the following two sessions were test data sets. These two sessions were randomized to avoid any order effect in the experiment. Each session had two parts, a story review part and an avatar evaluation part. Then, a semi-structured interview followed at the end of each session.

This first part of a session was conducted outside of CAVE2™. During the first part of each session, subjects were given a written story to read and asked to select a primary emotion that they might feel if the subjects were telling the story themselves. Then, they were directed to draw a graph depicting the intensity of primary emotion over the story time line.

The purpose of an emotion graph is a two-fold. First, the pattern of changes in the intensity can be used to measure how well the presented framework can simulate an avatar emotional state in terms of its intensity. Second, it can help subjects pre-visualize their expectation for an autonomous storyteller better. An idea was derived by Raij's social perspective-taking method [Raij '09] while there was only one player in the experiment.

In the second part of each session, subjects were told that they would review seven different avatars telling the story that they had read and would rate each avatar with respect to their perceived naturalness of avatar behavior throughout the story. Which specific behaviors are different from each other was not explained until the end of entire evaluation. Subjects were directed to review a video once at the beginning without rating to obtain an overview of the story and avatar behavior. Two rating methods were proposed and subjects were told to choose either way for their convenience (Figure 38). Both methods were not required to set discrete rankings for all seven conditions.



Please stop here and review avatar videos on the display wall first before proceed.

Please use below blank area to take notes and comments if you like to remember your reasons to choose/rate individual video clip. If you need extra space, please ask for a blank paper too.

After reviewing and rating avatar videos on the display wall, please write down the number of sticky notes you gave for each video (in the box below).

(a) Sticky Notes Rating

(b) Notes on a Survey Form

Figure 38. Two rating methods used during avatar video evaluation. In (a), subjects posted sticky notes on each avatar on the display. In (b), the second method was to take a note on the survey while reviewing an avatar video. At the end of each session, subjects counted total number of sticky notes or rating counts.

6.2.4 Measures

In the first part of the experiment session, the basic emotion category chose to describe the primary emotion of the story was measured to determine whether subjects were able to evaluate the story with the same emotion as the presented framework did. Secondly, graphs that subjects drew to depict the intensity of the chosen primary emotion were collected.

Subjective evaluations of the perceived naturalness for the seven different conditions were measured by scores that subjects gave to each avatar (the total number of sticky notes on each case or count noted on the survey form). These raw scores values were used to compute the rankings for each condition. The total time spent for the study and each session were also recorded to validate the results. During an avatar review process (part II of each session), the positions of subjects in CAVE2™ were tracked and recorded via a motion tracking system. Even though this position date was not analyzed in the experiment, it is reserved to perform further analysis in the future. In addition to quantitative measures, the entire experiments were videotaped by two cameras mounted on the ceiling of CAVE2™ and near the top of the entrance to observe subjects' review process as a reference. Lastly, a semi-structured interview after each session was recorded via an audio recorder to perform a qualitative analysis to measure the overall experience, the difficulty of the rating task, the adopted rating strategy, the correspondence of avatar behavior to subjects' expectation, and so on.

6.3 Results

Most of participants were able to finish the study in about an hour including preparation, training, evaluation, and debriefing. Average study time was 1:06:31 (*SD 18:56*). The shortest time was 39:37 and the longest one took about two hours to complete the task.

6.3.1 Quantitative Measures in Part I

Average survey time for the first part of *Story I* was 3:42 (*SD 1:42*) and 4:25 (*SD 3:34*) was for *Story II*. 95.83% of subjects chose *Happy* as its primary emotion for *Story I*, whereas one participant chose *Surprise*. In *Story II*, subjects reported mixed results. 62.5% felt *Sad*, 16.7% was *Disgust*, 12.5% noted *Angry*, and the remaining 8.3% of subjects picked *Surprise*. 4 subjects wanted to pick two emotions such as *Sad/Pain*, *Sad/Surprise* and *Sad/Angry*, which implied it was a non-trivial case to choose only one primary emotion. These 4 subjects were included in the *Sad* group because one of their two choices was *Sad*.

For the graphs depicting the intensity of the primary emotion, several pattern-matching schemes were applied. First of all, the graph itself does not necessary represent very accurate intensity of emotion. It was a rather rough sketch that subjects felt from the stories. Three features (patterns) used in this graph analysis were (1) overall trend (increase, decrease, no change), (2) local (a part of the story time line) increase / decrease points, and (3) the intensity of the beginning and the ending. With these simplified notations, all graphs were classified into a few groups (four major groups for each stories).

In *Story I*, total 91.7% of graphs were fit to one of the four patterns shown in Figure 39 and the remaining 8.3% were either flat line or decrease at the end of the story. Most of subjects (95.8%) described the general trend of the primary emotion as ***overall increasing pattern*** (all subjects included in Figure 39 and one more subject who drew a slight decrease at the end of story time line). 45.8% out of this 95.8% explained that there was ***one or two local decreases*** in the early part of the story, which was caused by the fact that the person could not meet Minnie Mouse during previous visits (Figure 39a and Figure 39d).

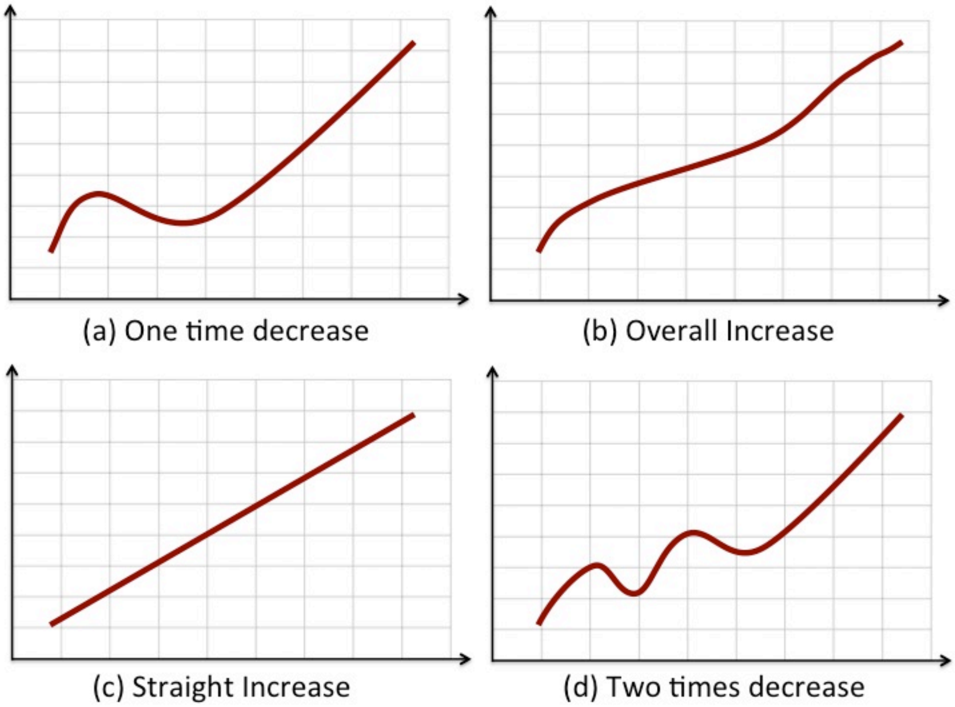


Figure 39. Primary emotion graph described by subjects in *Story I*. X-axis denotes story time line and y-axis measures the intensity of the primary emotion. In (a), there is only one decreasing point found (37.5%). In (b), graph illustrates overall increasing pattern (33.3%). In (c), subjects drew constant increase (12.5%). In (d), there are two decrease points (8.3%).

These results indicate that a majority of subjects identified the primary emotion as *Happy* and found the given story (*Story I*) increased in happiness throughout the story with a few relieves during the first half of story. This description is similar to what we reviewed in the simulated avatar emotion intensity graph (Figure 31).

In *Story II*, 66.7% of subjects described the intensity of primary emotion as ***overall increasing pattern***. This group is composed of Figure 40a, Figure 40b, and two more subjects who noted a slight decline at the end. The group in Figure 40a (41.7%) also found ***one local decrease*** at the beginning of the story when the person made lots of money and assets to avoid poverty (Figure 40a). The results in the second story resemble the graph pattern the presented framework simulated in Chapter 6.2.1 (Figure 34) although the number of subjects who identified similar patterns are not as many as in the first story (66.7% vs. 95.8%).

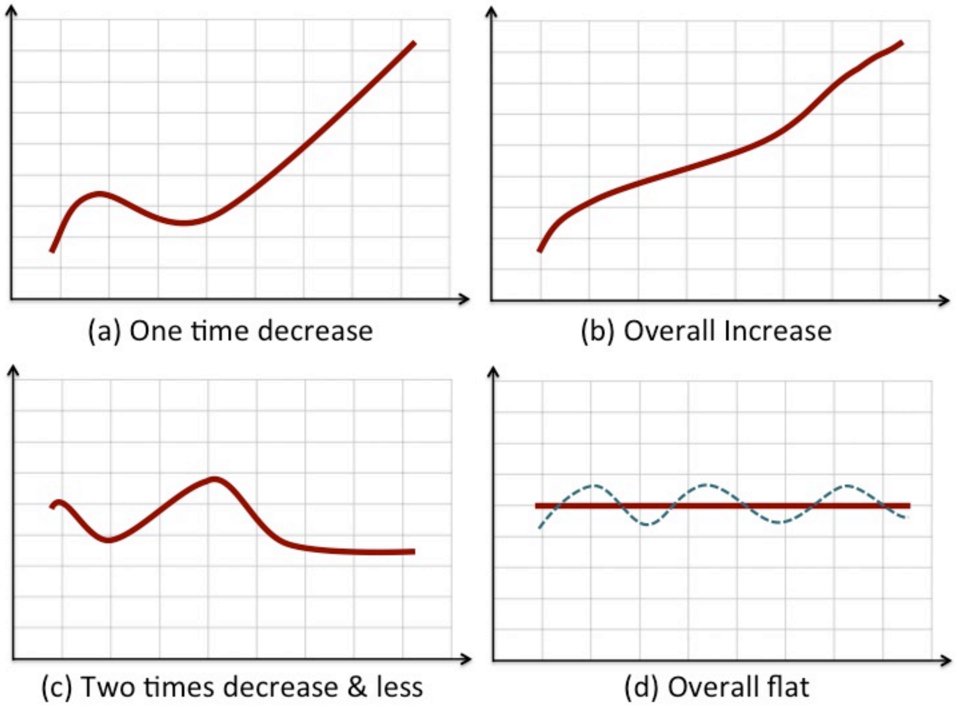


Figure 40. Primary emotion graph described by subjects in *Story II*. In (a), there is only one decreasing point (41.7%). In (b), graph illustrates overall increasing pattern (16.7%). In (c), subjects drew two decrease points and graph ends with less intense emotion (16.7%). (d) shows a flat line or three times fluctuation where its overall intensity stays similar range (12.5%).

Evaluation of Hypothesis H1: Emotion graph depicts similar patterns in the simulated one

Overall, hypothesis H1 received strong support; graphs in the first story mostly depicted similar patterns found in the simulated results (95.8% overall increase and 45.8% local decrease pattern at the beginning of the story). The majority of graphs in the second story described overall increasing patters and one local decrease that are close to the simulation results.

6.3.2 Quantitative Measures in Part II

In the second part of the experiment, total review time and avatar naturalness perceived in its behavior were measured. The average review time for the first story was 9:07 (*SD 5:08*) that is equivalent to 3.18 times (*SD 1.79*) video review and the average raw total score count was 22.33 (*SD 37.92*). In the second story, subjects spent average 8:15 (*SD 5:18*) to review avatar behavior that is 3.32 times (*SD 2.14*) of a single video playback time and average raw total score count was 20.50 (*SD 33.79*).

Among 24 subjects in the study, 4 cases in each story were invalidated because those subjects reviewed avatar behavior less than twice to rate their preferred behavior. Those data could be greatly affected by subjects' initial impression that might miss a short moment of avatar behavior occurring at a certain point and lose the overall progression of emotional changes presented. Therefore, those cases were excluded from this analysis. The effective number of subjects is 20 for each case. The normalized mean score for each condition is presented in Table VI.

Facial Expression		Head Gesture		
		Model	Control 0.5	Control 0.1
Story I	Model	*0.237 (0.102)*	*0.226 (0.097)*	***0.269 (0.139)***
	Control 0.5	*0.107 (0.090)*	*0.066 (0.078)*	*0.053 (0.056)*
	Control 0.5a	*n/a*	*0.042 (0.051)*	*n/a*
Story II	Model	***0.220 (0.145)***	*0.180 (0.145)*	*0.199 (0.124)*
	Control 0.5	*0.092 (0.102)*	*0.133 (0.133)*	*0.073 (0.103)*
	Control 0.5a	*n/a*	*0.114 (0.138)*	*n/a*

Table VI. Normalized mean score for 7 conditions. The combination of a model expression and a fixed gesture with the intensity 0.1 received the first rank in *Story I* and a model expression with a model gesture took the first place in *Story II*. Mean values in bold represent the first rank condition.

For further analysis, the Shapiro-Wilk test [Shapiro '65] was applied to assess the normality of the data in each condition. The Shapiro-Wilk test is widely adopted to evaluate the normality of a sample data whose size is less than 2000. The results in *Story I* showed that control conditions Cx05-Cg05 (*W*(20)=0.826, *p*<0.05), Cx05-Cg01 (*W*(20)=0.814, *p*<0.05) and Cx05a-Cg05 (*W*(20)=0.797, *p*<0.05) were significantly non-normal. Similarly, the results in *Story II* showed that control conditions Cx05-Cg01 (*W*(20)=0.748, *p*<0.05) and Cx05a-Cg05 (*W*(20)=0.793, *p*<0.05) were significantly non-normal. Therefore, the Wilcoxon signed-rank test [Wilcoxon '45] was applied to compare ranks between variables (combined cross-variable analysis over model vs. control group in both expression and head gesture). The combined mean score for emotional facial expression consists of Model Combined, Control Combined, Control 0.5, and Control 0.5a. Similarly, the recomposed head gesture score is composed of Model combined, Control Combined, Control 0.5, and Control 0.1.

In the first story, Model Combined, in both facial expression and head gesture, received the highest score (0.244 and 0.172) shown in Figure 41. Table VII presents the Wilcoxon signed-rank test for this story. In both behaviors, the Model Combined condition was most preferred by subjects and its differences compared to other control conditions were statistically significant (*p<0.05*) except Model Combined vs. Control 0.1 case in head gesture. The difference in this case (0.172 vs. 0.161) is meaningful but not significant (*p>0.05*).

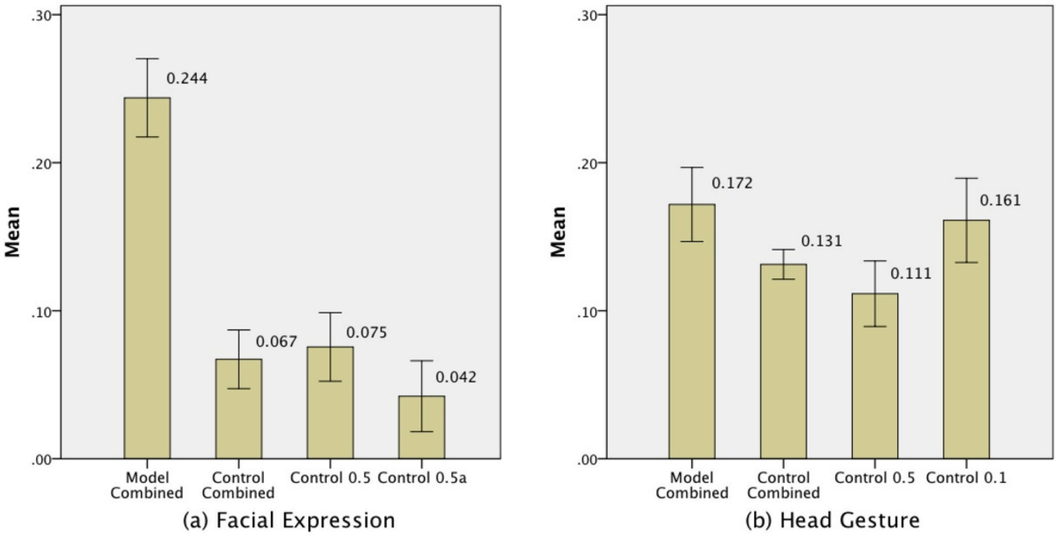


Figure 41. Combined facial expression and head gesture mean ratings for *Story I*. In (a), Model Combined took the first place with great difference with other conditions. In (b), Model Combined ranked first. Error bar indicates 95% confidence interval.

Variables		*N*			*Mean Rank*		*Z*	*Sig.*
		Neg.	Pos.	Ties	Neg.	Pos.		*2-sd*
Expression	**Model vs. Control**	1	19	0	1.00	11.00	-3.884	0.000
	Model vs. 0.5	1	18	1	1.00	10.50	-3.785	0.000
	Model vs. 0.5a	0	20	0	0.00	10.50	-3.922	0.000
	0.5 vs. 0.5a	4	13	3	8.50	9.15	-2.013	0.044
Gesture	**Model vs. Control**	4	15	1	8.88	10.30	-2.395	0.017
	Model vs. 0.5	6	14	0	5.83	12.50	-2.613	0.009
	Model vs. 0.1	7	10	3	9.00	9.00	-0.639	0.523
	0.5 vs. 0.1	14	6	0	11.43	8.33	-2.053	0.040

Table VII. *Wilcoxon* Signed-Rank Test for *Story I*: statistics for difference in mean ranks between the presented model and control conditions for both facial expression and head gesture. Bold conditions indicate statistically significant difference (*p < 0.05*)

The Model Combined condition in the second story, for both facial expression and head gesture, also received the highest score (0.200 and 0.156 in Figure 42). The facial expression model was most highly rated with significant difference (*p<0.05*) from control groups whereas the Model Combined head gesture condition was not statistically significant (*p>0.05*) compared to control conditions (Table VIII). Another finding in the second story as opposed to the first story was that there was no significant difference found between Expression Control 0.5 and Expression Control 0.5a (*p>0.05*). This means that subjects did not show any strong preference toward either a fixed facial expression 0.5 or a fixed facial expression 0.5 with a fast transition.

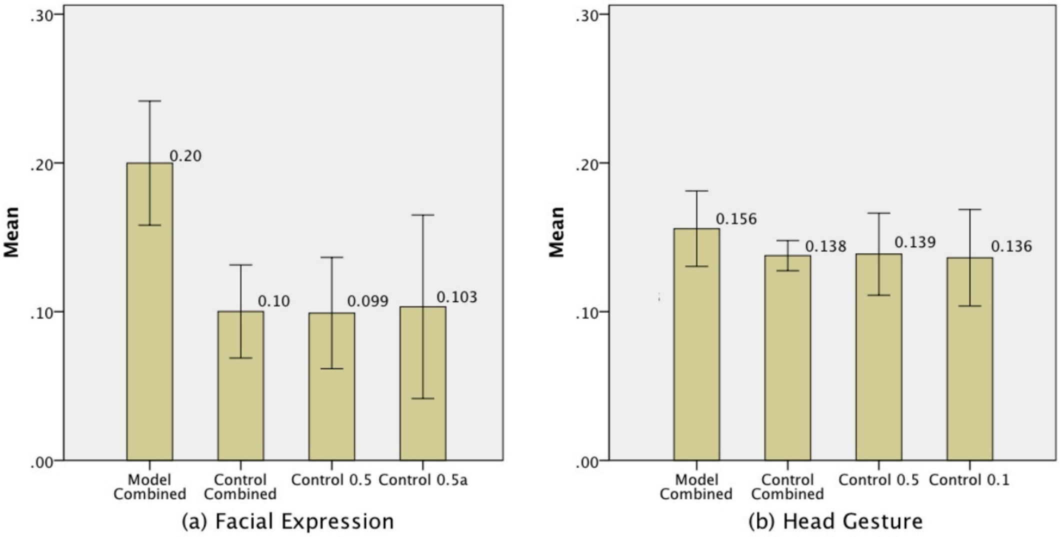


Figure 42. Combined facial expression and head gesture mean ratings for *Story II*. In (a), Model Combined took the first place with great difference with other conditions. In (b), Model Combined ranked first. Error bar indicates 95% confidence interval.

Variables		*N*			*Mean Rank*		*Z*	*Sig.*
		Neg.	Pos.	Ties	Neg.	Pos.		*2-sd*
Expression	**Model vs. Control**	5	15	0	8.40	11.20	-2.352	0.019
	Model vs. 0.5	4	15	1	10.00	10.00	-2.214	0.027
	Model vs. 0.5a	6	14	0	7.92	11.61	-2.147	0.032
	0.5 vs. 0.5a	7	12	1	11.36	9.21	-0.624	0.533
Gesture	Model vs. Control	11	8	1	7.64	13.25	-0.443	0.658
	Model vs. 0.5	7	9	4	7.57	9.22	-0.776	0.438
	Model vs. 0.1	6	10	4	8.17	8.70	-0.982	0.326
	0.5 vs. 0.1	8	8	4	8.69	8.31	-0.078	0.938

Table VIII. *Wilcoxon* Signed-Rank Test for *Story II*: statistics for difference in mean ranks between the presented model and control conditions for both facial expression and head gesture. Bold conditions indicate statistically significant difference (*p < 0.05*)

Evaluation of Hypothesis H2: Model Expression is more natural than others

Overall, hypothesis H2 received strong support; subjects rated higher naturalness of avatar emotional expression than all control conditions (Control combined, Control 0.5, and Control 0.5a) in both *Story I* and *II*. The Wilcoxon signed-rank test indicated that there were significant differences in the rank between Model and Control conditions ($p<0.05$).

Evaluation of Hypothesis H3: Model Head Gesture is more natural than others

There was partial support for hypothesis H3; subjects' preference for the Model head gesture was significantly higher than Control conditions (Control 0.5 and Control 0.1) in *Story I* ($p<0.05$) whereas in *Story II* it did not receive statistical significance. However, the data suggested some support of hypothesis H3; the Model condition in the second story obtained moderate-or-better ratings than those in control conditions.

Evaluation of Hypothesis H4: Combined Model is more natural than others

Hypothesis H4 received partial support; mean score for the Model Expression combined with Model Gesture (Mx-Mg) ranked in second place for *Story I* and in first place for *Story II* as shown in Figure 43 and Figure 44. Overall, Combined Model for both facial expression and head gesture received higher score than most of other conditions. However, its difference compared to conditions used Model Expression is not significant.

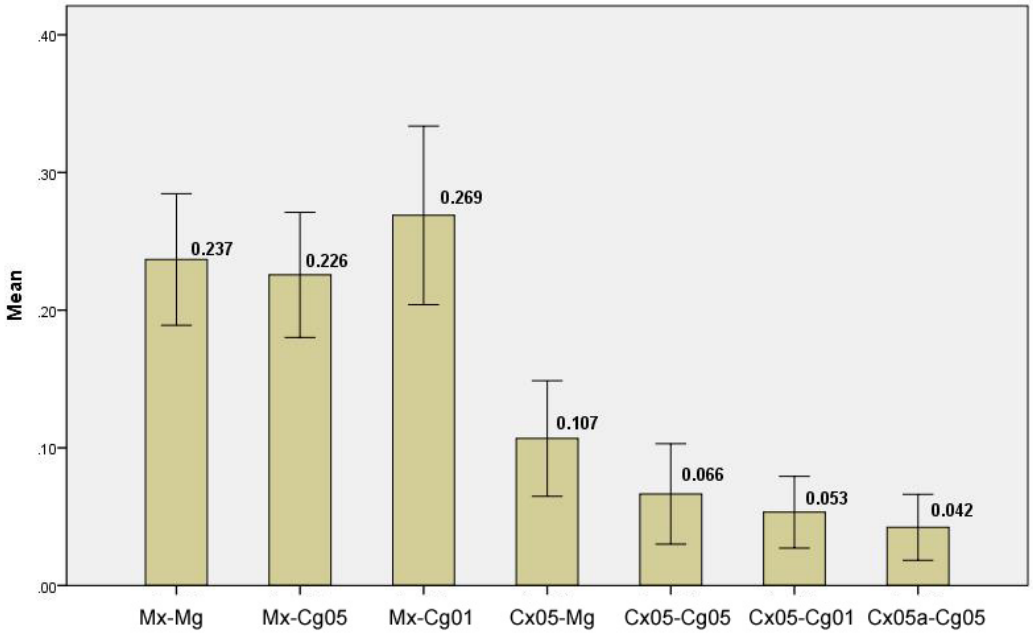


Figure 43. Mean score for 7 conditions in *Story I*. Model Expression with Model Gesture condition (Mx-Mg) ranked in the second place among 7 conditions (0.237). Error bar indicates 95% confidence interval

Among 6 conditions (Mx-Cg05, Mx-Cg01, Cx05-Mg, Cx05-Cg05, Cx05-Cg01, and Cx05a-Cg05) in *Story I*, four conditions (Cx05-Mg, Cx05-Cg05, Cx05-Cg01, and Cx05a-Cg05) showed significantly lower score than the combined model (*p<0.05*), whereas Mx-Mg condition was significantly better than three conditions (Cx05-Mg, Cx05-Cg01, and Cx05a-Cg05) in *Story II.*

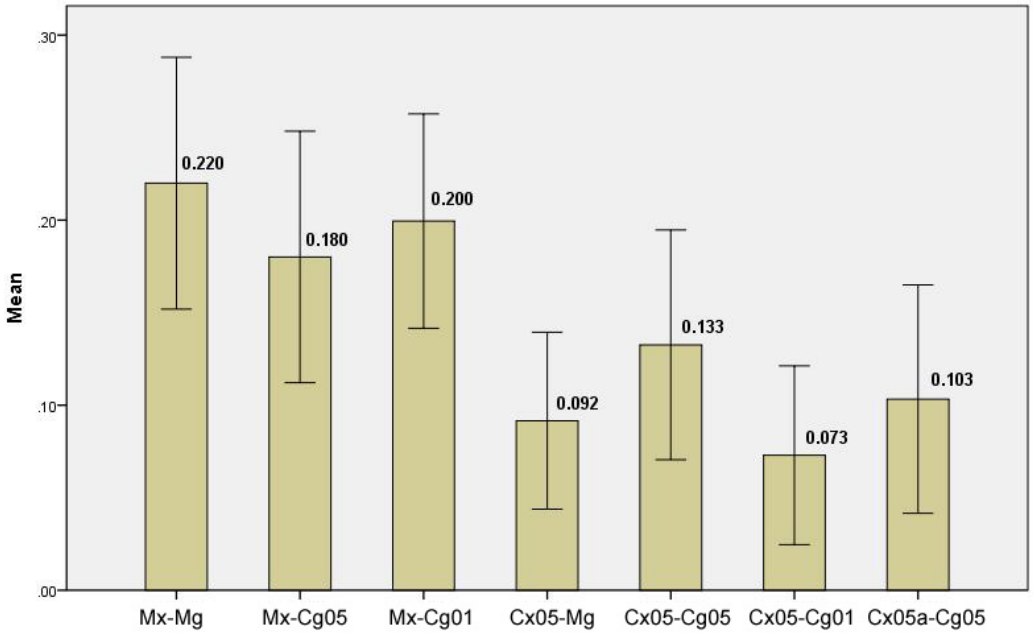


Figure 44. Mean score for 7 conditions in *Story II*. Model Expression and Model Gesture condition (Mx-Mg) ranked in the first place among 7 conditions (0.220). Error bar indicates 95% confidence interval

6.3.3 Qualitative Results

Post-session interviews were all transcribed and analyzed in several categories such as the difficulty of evaluation, subjects' evaluation strategy, what features they looked at, and so forth.

Difficulty of evaluation

This empirical study was designed to examine very subtle differences between the presented emotion and behavior modeling method and the conventional control method. As a consequence, most subjects confirmed that the evaluation task was non-trivial and very hard. A few subjects also noted that reading subtitle as well as reviewing avatar behavior at the same time made it more difficult:

"Pretty hard to see any differences, to be honest there..."

> *"It was very difficult to read and watch them at the same time."*
> *"It's difficult. They're scarily similar at the same time, once you start."*
> *"It's really hard. Even when I am looking at the key part, it's still hard to see difference between faces."*

Evaluation strategy

During the avatar video evaluation process, subjects adopted a several different strategies to discern each condition. A majority of subjects started with viewing all, or as many as possible, avatars to obtain an overall impression, and then divided them into smaller groups (2, 3, or half) to compare each other or went through them one by one. Some subjects compared several instances that they liked across the screen. One suggested method was to separate 7 conditions and to let a subject rate them by moving each video in their preferred order (ranking):

> *"I tried to be aware of the best I can see all over them. So, before I go up the place note, I tried to make sure to analyze one through four then four through seven, and then I put down the notes."*
> *"I'm looking… kind of possibly all of them equally and then also like really focusing on during certain part… focused on what the response was."*
> *"I tried to mostly focus on one because it's hard to try to compare. ... I was looking at one by one basis. Occasionally, reference the one next to it."*
> *"... I switched to the other side... focused on number X and see Y and Z…"*
> *"I was thinking that optimally it will be the best… if you separate 7 videos, then you may just arrange to create some sort of ranking…."*

Features focused

Features how each avatar behavior was different were not given to subjects during the evaluation process, but most of subjects were able to pick those feature sets during the evaluation process. In *Story I*, subjects tended to mostly focus on mouth, eyes, and head movement whereas eyebrows, eyes, and wrinkles were more important in *Story II*. Those features are well matched to how a current avatar model expresses happy (mouth, mouth furrow, and eyes) and sad (lower eyebrows and wrinkles) facial expression. Head gesture was also noted as a part of their evaluation criteria. A few subjects were able to notice the difference in head movement between *Story I* and *II*:

> *"I guess the most noticeable thing that I can see is the mouth." (Happy)*
> *"Mostly, the eyes and the lips and the nodding gestures." (Happy)*
> *"So, I feel that the head movement is not that obvious." (Sad)*
> *"When I'm feeling sadness, the eyebrows change." (Sad)*
> *"I definitely think eyes made huge differences. Wrinkles and eyes…" (Sad)*
> *"I definitely noticed the eyes. Eyebrows jumped to me." (Sad)*
> *"The head shaking, head movement is another thing that I think pretty important." (Sad)*
> *"She was one. When reading the subtitle at the same time, it matched up really closely sad like her eyebrows and the way her head movement." (Sad)*
> *"With the previous one, I found myself finding more emotion in the mouth. This one, I found a lot more eyes. Narrowing their eyes…bringing the eyebrows together." (Sad)*

Evaluation of emotion

Overall, design of emotional expression and the simulated model was corresponded well with subjects' expectation from the given story:

> *"Basically, I feel like it's the same as my expected."*

One of most interesting comments from subjects with respect to emotion was that the avatar should not express a certain emotion or not much of it at least if the avatar is overwhelmed by opposite or different emotional state:

> *"There wasn't a facial expression had happy because I found overall she's always sad."*
> *"I would think that there never be that neutral for her or happy or anything."*
> *"I don't like sadness in the middle. This particular story, I don't see any sadness. It's little bit of diminishing the happiness."*
> *"I thought the expression is overall happy except when she was describing when she could not get the autograph. I wouldn't say she was sad."*

A few subjects also noted that increasing pattern of emotion should produce more intense expression. These are important features that the presented PAD model is capable of:

> *"The emotion's really really great, but it doesn't smile so much. It's very very happy story. So, I think there should be more than that."*
> *"I can see which one is I preferred... at the very end. Because I can see more happiness at the end and there's more head movement."*

Evaluation of head gesture

Most subjects had hard time to evaluate the head gesture model in terms of detecting differences between models. However, many subjects confirmed that head movement was an important factor to enhance naturalness of avatar behavior:

> *"Head movement… I didn't really even noticed difference at all through any of them."*
> *"Whenever a head shaking, it really made sense because head emotions make sense. People move the head when they talk."*
> *"The other thing is that I see when people shaking the head, that go with the context and become more like."*

Some subjects addressed that a head gesture should show differently based on avatar's emotional state. For example, if it feels sad or depressed (lower arousal) it moves very subtle:

> *"This case is sad… depressed… Then, I don't move much. If I move much, that's kind of weird or something kind of opposite."*
> *"Between two stories, I expected less change of reactions because this is much more sustained one and that's what I saw this one. It's more subtle changes, but it's what I expected from."*
> *"I don't expect too much movement. I expect there will be much more subtle movement and a lot more sadness…"*

6.4 <u>Discussion</u>

Across both stories (happy and sad cases), there was strong support for hypothesis H1; most subjects described the avatar's primary emotion changes very similar to the results computed by the presented framework. Hypothesis H2 also received strong support from the study; subjects in both stories recognized better naturalness of emotional expression than other control conditions that used a fixed amount of intensity level regardless of the story context. However, there was only partial support for the other two hypotheses H3 and H4, although the data suggested that the presented model received higher ratings for both the head gesture model and the combined expression and gesture model than other conditions.

6.4.1 Accuracy of Emotion Analysis

Hypothesis H1 was the most significant one with respect to the goals of this thesis. The understanding of a story to be told and its appropriate emotional analysis are a key to implement a successful interactive expressive avatar. Most subjects in both interventions (two stories) described its emotional curve as increasing pattern with a few declining points in the first half of story line with respect to its chosen primary emotion. Even though graph descriptions were not scientifically defined (asked to draw rough sketch to depict progressive avatar emotion), those were closely related to the simulated results from the presented framework. Therefore, the avatar affect evaluation model was able to achieve a desired goal.

There was another very important finding to this end. When subjects described a decreasing moment in the primary emotion, some of them clearly mentioned that it was not necessarily the case that avatar should show the other conflicting emotion or not much of that emotion at least. For example, when an avatar is in high mood (very happy) a disappointment or sadness should not be realized in great extent. This is how the presented affect evaluation model works based on the current mood (PAD model). In the first story, our avatar had increasing happiness throughout the story and when it received the first SAD eliciting utterance with base line intensity 0.6, it was reduced down to 0.35 after evaluation due to high pleasure value in PAD model. Therefore, the presented PAD emotion model was capable of relieving conflicting emotions.

6.4.2 Perceived Naturalness of Emotion

As analyzed in Chapter 6.3.2, the combined emotion model for facial expressions received very strong support in both stories. This combined emotion model includes different types of head gesture conditions (model gesture, control with intensity 0.5, control with intensity 0.1). The model's signed rank difference was statistically significant compared to the combined control conditions. This result indicates that the PAD model driven emotion expression is better than all control groups (fixed intensity 0.5 and 0.5 faster transition). As far as emotion concerns, this is a consistent result compared to people's expectation in the aforementioned emotion graph comparison. Therefore, the presented avatar emotion evaluation and its realization method are superior to a conventional method that uses a certain fixed intensity value.

There were some other suggestions from subjects such as use of exaggerated emotional expression in great degree when an avatar is at the extreme end of one emotion. For example, when she feels much happier due to accumulated happy stimulus or eventual realization of long wanted wish, she may open eyes widely, raise up brow higher, and some auxiliary utterances (i.e. "wow" and "yay"). However, this secondary augmentation requires deeper understanding of dialogue and story than shallow parsing that is used in the current model. There are several possible approaches to achieve this end. Carefully designed dialog act of utterance may be used to trigger the secondary nonverbal behavior. Otherwise, we can also use the supervision control facility in the presented framework to add a nonverbal behavior event within speech utterances.

6.4.3 Perceived Naturalness of Head Gesture

Even though only one story (happy story) received a significantly higher rating for the head gesture model than the combined control group, the data suggests moderately higher scores in the sad story, too. The lack of significant differences in the sad story may be due to several factors. First, the influence of head gesture behavior might not be as great as emotional expression on subjects' evaluation. Second, the underlying simulation model might not resonate with subjects' expectation.

Although some subjects noted that head gesture was an important factor in their evaluations, others utilized facial expression more that head gestures to determine their subjective likeness for each conditions. Also, some subjects were not able to distinguish the difference in head gesture behavior. Furthermore, there was one subject who preferred more gestures regardless of one's mood (*"I like more gestures, so, I am looking for broader head nod"*). This might have negatively impacted on their ratings in the sad story because the model simulated less aroused mood causing the avatar to present less intense gestures as a result. This is a somewhat controversial case since psychological literature suggests similar rules that the current model adopted and some subjects addressed strong distinction in head gesture varied by the mood. A further study is required to examine this case in the future.

While the majority of subjects selected Sad as primary emotion in the second story, there were some cases whose choice was not Sad. Ones who chose either Surprise or Angry showed especially interesting correlation (total 4 subjects). Assuming that either of two emotions was the primary emotion, then the head gesture model would calculate more intense head gestures because Surprise and Angry have positive arousal value in PAD model whereas Sad has negative arousal. Interestingly, those subjects rated control 0.5 higher than the model in the experiment (mean score 0.175 vs. 0.112), which means that they gave more score to intense head gestures condition (fixed 0.5 vs. varied from 0.05 to 0.26). This is one possible explanation. However, statistical analysis cannot be applied to this phenomenon due to small sample size.

6.4.4 Perceived Naturalness of Emotion and Head Gesture

There was moderate partial support for the last hypothesis H4. Despite the combined expression model achieving a significantly higher score, the combination of model expression and model head gesture (Mx-Mg) did not obtain the same significance when it was compared with 2 other conditions that used expression model (Mx-Cg05 and Mx-Cg01). However, Mx-Mg condition was significantly better than control expression conditions (Cx05-Mg, Cx05-Cg05, Cx05-Cg01, and Cx05a-Cg05). This result is consistent with the discussion of hypothesis H1 and H2. Facial expression was more likely the main factor to distinguish each condition. Because there were three conditions with the same facial expression model (Mx-Mg, Mx-Cg05, and Mx-Gg01) in the video, if there were some subjects who were not able to recognize differences in head gestures, it would be difficult to choose the most preferred one among those three variations.

6.4.5 Difficulty of Evaluation

Even though they watched the avatar videos multiple times, the fact that most subjects found the evaluation very difficult indicates several lessons. First, the empirical study was successful in terms of unbiased stimulus design. If the task was too easy, it could mean that variables between model and control conditions were too obvious and findings in the study might not be that valuable. Second, the evaluation method may need to be revised to help subjects accomplish the task easier without much technical difficulty. This may appear to contradict the first lesson, however this is not about a stimulus design but about a methodology.

The seven different conditions in a single video are hard to review by itself. Without an efficient filtering methodology, it may take much more time to accurately evaluate subtle differences in behavior models. Some suggestions received during interviews were follows: Reducing the number of conditions to watch at the same time would be helpful. Otherwise, ability to re-order 7 conditions while reviewing may help subjects effectively filter undesirable conditions. In fact, those two suggestions were considered at the very beginning of study design. De Melo and Gratch conducted a study to examine the influence of some advanced facial features on the perception of basic emotions [de Melo '09]. They used a pair of static image of avatar renderings with different features and asked subject to rate perceived expressiveness. In this case, only one pair of images was given at one time, which make comparison much easier. Another similar example is [Courgeon '09]. Courgeon et al. performed a user study to evaluate the effect of their expressive wrinkle renderings with respect to users' perception of facial expression. They presented sequence of short video stimulus successively in one session and showed 4 rendering images at a time in the second session. However, the study of this thesis has some difference from theirs in terms of characteristics of material used. First, static images cannot be used to measure the effect of avatar expressiveness that employs a progressive dynamic model. Second, dividing variables into small sets can drastically increase the review time since avatar video runs 2 to 3 minutes for only one review. Finally, ability to move each video in preferred order while reviewing was not feasible without developing an application only for that purpose because multiple instances of video player with synchronization facility is not available and comparing multiple videos that are out-of-sync is much harder to identify difference. However, this last suggestion is the most doable among others and it is expected to be a very effective way of filtering. A future study may utilize this method with a custom video review application.

7. CONCLUSION

An avatar-enabled application promises more natural computer interaction with advanced technologies, and recent research efforts towards emotional avatar capabilities have become more prevalent in the field. This believable avatar model intuitively aims to mimic a real human including realistic appearance and a wide spectrum of complex behavior. However, developing such an application still remains a very difficult and time-consuming task. State-of-art computer graphics techniques to create realistic renderings have to be adopted in the interactive avatar research domain. Designing a computational model to express complex human behavior requires an enormous amount of knowledge and even that may be an intractable task. Nonetheless, research efforts to maneuver such a complex problem can contribute to improve current limitation and to get us close to our ultimate goal.

A high quality avatar design and visualization method is presented in this dissertation. The design process describes a comprehensive set of modeling details for interactive avatar development and the visualization method presents an iterative enhancement with advanced graphics techniques that are highly re-usable. This design and visualization method demonstrates a more visually compelling avatar than many recent avatar visualization examples. The method also shows its reusability and efficiency by presenting different types of avatar-enabled applications developed with this method.

In this dissertation, a novel method to accommodate emotional affect in avatar nonverbal behavior for both facial expressions and head gestures is also presented. Unlike traditional two distinct behavior-modeling approaches, the framework facilitates a hybrid approach to merge both rule-based and data-driven methods to improve believability of avatar behavior. The presented affect control framework evaluates avatar emotional states as well as utilizes its results to augment avatar nonverbal behavior, so called the *emotionally augmented behavior modeling* method. This emotional augmentation is obtained by computing the dynamic avatar emotion state in PAD space model where momentary avatar mental state is changing upon incoming stimulus.

A user study was conducted to evaluate the naturalness of the presented behavior model in the context of an autonomous expressive storytelling avatar. The empirical study results show that the presented hybrid modeling method receives higher ratings with respect to the perceived naturalness of avatar behavior for both emotional facial expressions and head gestures over a fixed factor model. Facial expression driven by the Supervised Hybrid Expression Control Framework (SHECF) model is recognized as significantly better than control conditions. The result of the head gesture behavior model was not as significant as facial expression case. However, it gives meaningful support for the presented framework and suggests interesting future research directions to investigate further.

High quality avatar visuals with an emotionally expressive behavior model in this dissertation offer better congruency and naturalness to increase avatar believability and to help us eventually achieve the goal of an avatar as a means of a natural lifelike computer interface. *Increasing either the visual realism or the behavioral complexity of an avatar is not sufficient to enhance our experience.* The presented balanced approach will contribute to resolve our *conflicting perceptual queue* against a biased unfamiliar avatar to best avoid Mori's Uncanny Valley.

LRAF and SHECF promote easy design and development for an avatar-enabled application by the synergy between widely available technologies to create a natural and believable avatar control framework. As we establish a better tractable model for an avatar as a more natural alternative computer interface, it will broaden the possibilities of our computation needs where we suffer from limited resources.

8. FUTURE RESEARCH DIRECTION

There are many productive and advanced directions for the future in this dissertation. Currently, the framework uses undirected flat surface text to extract and process avatar affect to control its nonverbal behavior. However, there is another important aspect that possibly changes avatar behavior drastically. It is intention or goal of dialogue known as *Dialog Act*. In fact, this feature requires either a significant amount of human intervention to provide such information in pre-defined dialog or a highly intelligent dialog management system to extract it on the fly. Nevertheless, it enables more engaging interaction by providing another layer to control avatar behavior. To accommodate this feature, the current framework can be modified to diversify its emotion augmentation method. For example, when the Behavior Processor receives a dialog act that involves an emotional aspect or a certain functional element, it can refer to the intention of the dialog instead of its mood to augment behavior.

There are also many interesting research problems discovered during this dissertation. One of the most interesting challenges is cultural difference in our behavior. For example, it is a well-known fact that people use head nod/shake gestures differently to express negation/affirmative action in a certain culture. Even more fine-grained with respect to this study, is the possibility of a different reaction or augmentation to an emotional state. One interesting comment from a subject during pilot study was: *"If I feel so sad or depressed, I would show more intense gesture. I think it's the way that female would do. Or, maybe because I am not from US? … I don't know. I just feel like that,"* This cannot be generalized by any means without careful investigation about cultural or gender diversity in behavioral gestures. However, it signals that there are needs to study it. Cowie [Cowie '10] noted this possibility when they observed people's behavior in

SEMAINE corpus. To address this case best, the current framework needs to be revised so that logical assignments for augmentation factors, recipients of augmentation, and associated parameters can be altered. This can be done by separating that kind of information and specifying them in each individual avatar descriptor, an avatar specification file. This will help the revised framework be more acceptable in diverse culture, and even more specific instances of avatar realization would become possible.

Furthermore, a study needs to be conducted to examine task-oriented performance in addition to the current perception based measures. The subjective preference or scoring result may not draw task performance of desired objective users even though it will clearly help their likeness of an avatar. Prior work showed that a preferred relationship between a user and an avatar results in better performance or willingness of participation improving the goal in turn [Bickmore '11a; Bickmore '05]. Therefore, it will be interesting to design a task-oriented avatar application and to measure its performance to this extent.

CITED LITERATURE

[Alexander '10] Alexander, O., Rogers, M., Lambeth, W., Chiang, J.-Y., Ma, W.-C., Wang, C.-C., and Debevec, P., 2010. The Digital Emily Project: Achieving a Photorealistic Digital Actor. *Computer Graphics and Applications, IEEE 30*, 4, 20-31.

[Autodesk '13] Autodesk. 2013. Autodesk Maya - Comprehensive 3D Animation software 2013 [cited 7/1 2013]. Available from http://www.autodesk.com/products/autodesk-maya/overview.

[Bailly '10] Bailly, G., Raidt, S., and Elisei, F., 2010. Gaze, conversational agents and face-to-face communication. *Speech Communication 52*, 6 (Jul 01).

[Becker-Asano '09] Becker-Asano, C. and Wachsmuth, I., 2009. Affective computing with primary and secondary emotions in a virtual human. *Autonomous Agents and Multi-Agent Systems 20*, 1 (Jun 10), 32-49.

[Bee '10] Bee, N., Wagner, J., André, E., Vogt, T., Charles, F., Pizzi, D., and Cavazza, M., 2010. Discovering eye gaze behavior during human-agent conversation in an interactive storytelling application. *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction*, 1-8.

[Beeler '10] Beeler, T., Bickel, B., Beardsley, P., Sumner, B., and Gross, M., 2010. High-quality single-shot capture of facial geometry. *SIGGRAPH &apos;10: SIGGRAPH 2010 papers*(Jul).

[Beeler '11] Beeler, T., Hahn, F., Bradley, D., Bickel, B., Beardsley, P., Gotsman, C., Sumner, R.W., and Gross, M., 2011. High-quality passive facial performance capture using anchor frames. *ACM Trans. Graph. 30*(Aug), 75:71-75:10.

[beepa '13] beepa. 2013. Fraps: Real-time video capture and benchmarking 2013 [cited 7/1 2013]. Available from http://www.fraps.com.

[Bickmore '08a] Bickmore, T.W., Mauer, D., Crespo, F., and Brown, T., 2008a. Negotiating task interruptions with virtual agents for health behavior change. *AAMAS '08: Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems 3*(Jun 01).

[Bickmore '09] Bickmore, T.W., Pfeifer, L., and Jack, B., 2009. Taking the time to care: empowering low health literacy hospital patients with virtual nurse agents. *Proceedings of the 27th international conference on Human factors in computing systems*.

[Bickmore '11a] Bickmore, T.W., Pfeifer, L., and Schulman, D., 2011a. Relational Agents Improve Engagement and Learning in Science Museum Visitors, H. VILHJÁLMSSON, S. KOPP, S. MARSELLA and K. THÓRISSON Eds. Springer Berlin / Heidelberg, 55-67.

[Bickmore '08b] Bickmore, T.W., Pfeifer, L., Schulman, D., Perera, S., Senanayake, C., and Nazmi, I., 2008b. Public displays of affect: deploying relational agents in public spaces. *CHI '08 extended abstracts on Human factors in computing systems*.

[Bickmore '05] Bickmore, T.W. and Picard, R., 2005. Establishing and maintaining long-term human-computer relationships. *ACM Trans. Comput.-Hum. Interact. 12*, 2, 293-327.

[Bickmore '11b] Bickmore, T.W., Schulman, D., and Sidner, C.L., 2011b. A reusable framework for health counseling dialogue systems based on a behavioral medicine ontology. *Journal of biomedical informatics 44*, 2 (May), 183-197.

[Burgoon '02] Burgoon, J.K. and Hoobler, G.D., 2002. Nonverbal signals, M.L. KNAPP and J.A. DALY Eds. SAGE Publications, Inc., Thousand Oaks, CA, 240-299.

[Carletta '07] Carletta, J., 2007. Unleashing the killer corpus: experiences in creating the multi-everything AMI Meeting Corpus. *Language Resources and Evaluation 41*, 181-190.

[CAVE2 '13] CAVE2. 2013. CAVE2: Next-Generation Virtual-Reality and Visualization Hybrid Environment for Immersive Simulation and Information Analysis. Electronic Visualization Laboratory 2013 [cited 7/1 2013]. Available from http://www.evl.uic.edu/core.php?mod=4&type=1&indi=424.

[Čereković '11] Čereković, A. and Pandžić, I., 2011. Multimodal behavior realization for embodied conversational agents. *Multimedia Tools and Applications 54*, 143-164.

[Cohen '60] Cohen, J., 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement 20*, 1, 37-46.

[Cohen '98] Cohen, J., Olano, M., and Manocha, D., 1998. Appearance-preserving simplification. In *Proceedings of the Proceedings of the 25th annual conference on Computer graphics and interactive techniques* (1998), ACM, 280832, 115-122.

[Cortes '95] Cortes, C. and Vapnik, V., 1995. Support-vector networks. *Machine learning 20*, 3, 273-297.

[Courgeon '09] Courgeon, M., Buisine, S., and Martin, J.-C., 2009. Impact of Expressive Wrinkles on Perception of a Virtual Character's Facial Expressions of Emotions, Z. RUTTKAY, M. KIPP, A. NIJHOLT and H. VILHJÁLMSSON Eds. Springer Berlin / Heidelberg, 201-214.

[Cowie '10] Cowie, R., Gunes, H., McKeown, G., Vaclau-Schneider, L., Armstrong, J., and Douglas-Cowie, E., 2010. The emotional and communicative significance of head nods and shakes in a naturalistic database. In *Proc. of LREC Int. Workshop on Emotion*, 42-46.

[De Marneffe '06] De Marneffe, M.-C., MacCartney, B., and Manning, C.D., 2006. Generating typed dependency parses from phrase structure parses. In *Proceedings of LREC*, 449-454.

[De Marneffe '08] De Marneffe, M.-C. and Manning, C.D., 2008. Stanford typed dependencies manual. *URL* http://nlp.stanford.edu/software/dependencies_manual.pdf.

[de Melo '10] de Melo, C., Carnevale, P., and Gratch, J., 2010. The Influence of Emotions in Embodied Agents on Human Decision-Making, J. ALLBECK, N. BADLER, T.W. BICKMORE, C. PELACHAUD and A. SAFONOVA Eds. Springer Berlin / Heidelberg, 357-370.

[de Melo '12] de Melo, C., Carnevale, P., and Gratch, J., 2012. The Effect of Virtual Agents' Emotion Displays and Appraisals on People's Decision Making in Negotiation, Y. NAKANO, M. NEFF, A. PAIVA and M. WALKER Eds. Springer Berlin / Heidelberg, 53-66.

[de Melo '09] de Melo, C. and Gratch, J., 2009. Expression of Emotions Using Wrinkles, Blushing, Sweating and Tears, Z. RUTTKAY, M. KIPP, A. NIJHOLT and H. VILHJÁLMSSON Eds. Springer Berlin / Heidelberg, 188-200.

[DeMara '08] DeMara, R.F., Gonzalez, A.J., Hung, V., Leon-Barth, C., Dookhoo, R.A., Jones, S., Johnson, A., Leigh, J., Renambot, L., Lee, S., and Carlson, G., 2008. Towards Interactive Training with an Avatar-based Human-Computer Interface. *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC) 2008*(Dec 03).

[Dickerson '05] Dickerson, R., Johnsen, K., Raij, A., Lok, B., Stevens, A., Bernard, T., and Lind, D.S., 2005. Virtual patients: assessment of synthesized versus recorded speech. *Studies in Health Technology and Informatics*.

[Ekman '78] Ekman, P. and Friesen, W., 1978. *Facial Action Coding System: A Technique for the Measurement of Facial Movement.* Consulting Psychologists Press, Palo Alto.

[Ekman '02] Ekman, P., Friesen, W., and Hager, J., 2002. FACS Manual. *A Human Face*.

[FaceGen '13] FaceGen. 2013. FaceGen -3D Human Faces. Singular Inversions 2013 [cited 7/1 2013]. Available from http://www.facegen.com.

[Finkelstein '09] Finkelstein, S.L., Nickel, A., Harrison, L., Suma, E.A., and Barnes, T., 2009. cMotion: A New Game Design to Teach Emotion Recognition and Programming Logic to Children using Virtual Humans. *Virtual Reality Conference, 2009. VR 2009. IEEE*(Apr 14), 249-250.

[Garau '03] Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A., and Sasse, M.A., 2003. The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *Proceedings of the Proceedings of the SIGCHI conference on Human factors in computing systems* (Ft. Lauderdale, Florida, USA2003), ACM, 642703, 529-536.

[Gebhard '05] Gebhard, P., 2005. ALMA: a layered model of affect. In *Proceedings of the Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems* (New York, NY, USA2005), ACM, 29-36.

[Gebhard '09] Gebhard, P. and Karsten, S., 2009. On-Site Evaluation of the Interactive COHIBIT Museum Exhibit, Z. RUTTKAY, M. KIPP, A. NIJHOLT and H. VILHJÁLMSSON Eds. Springer Berlin / Heidelberg, 174-180.

[Gebhard '06] Gebhard, P. and Kipp, K., 2006. Are Computer-Generated Emotions and Moods Plausible to Humans?, J. GRATCH, M. YOUNG, R. AYLETT, D. BALLIN and P. OLIVIER Eds. Springer Berlin / Heidelberg, 343-356.

[Goldberg '90] Goldberg, L.R., 1990. An Alternative "Description of Personality": The Big-Five Factor Structure. *Journal of Personality and Social Psychology 59*, 6 (Dec), 1216-1229.

[Gunes '10] Gunes, H. and Pantic, M., 2010. Dimensional Emotion Prediction from Spontaneous Head Gestures for Interaction with Sensitive Artificial Listeners, J. ALLBECK, N. BADLER, T.W. BICKMORE, C. PELACHAUD and A. SAFONOVA Eds. Springer Berlin / Heidelberg, 371-377.

[Hoque '11] Hoque, M.E. and Picard, R.W., 2011. Acted vs. natural frustration and delight: Many people smile in natural frustration. In *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition* (Santa Barbara, Apr 22 2011).

[Huang '10] Huang, L., Morency, L.-P., and Gratch, J., 2010. Learning Backchannel Prediction Model from Parasocial Consensus Sampling: A Subjective Evaluation, J. ALLBECK, N. BADLER, T.W. BICKMORE, C. PELACHAUD and A. SAFONOVA Eds. Springer Berlin / Heidelberg, 159-172.

[Jan '09] Jan, D., Roque, A., Leuski, A., Morie, J., and Traum, D., 2009. A Virtual Tour Guide for Virtual Worlds, Z. RUTTKAY, M. KIPP, A. NIJHOLT and H. VILHJÁLMSSON Eds. Springer Berlin / Heidelberg, 372-378.

[Jeong '10] Jeong, B., Leigh, J., Johnson, A., Renambot, L., Brown, M.D., Jagodic, R., Nam, S., and Hur, H., 2010. Ultrascale Collaborative Visualization Using a Display-Rich Global Cyberinfrastructure. *Computer Graphics and Applications, IEEE 30*, 3, 71-83.

[Jimenez '10] Jimenez, J., Whelan, D., Sundstedt, V., and Gutierrez, D., 2010. Real-Time Realistic Skin Translucency. *Computer Graphics and Applications, IEEE 30*, 4, 32-41.

[Kanade '00] Kanade, T., Cohn, J.F., and Tian, Y., 2000. Comprehensive database for facial expression analysis. In *Proceedings of the Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on* (2000), 46-53.

[Kenny '07] Kenny, P., Parsons, T., Gratch, J., Leuski, A., and Rizzo, A., 2007. Virtual Patients for Clinical Therapist Skills Training, C. PELACHAUD, J.-C. MARTIN, E. ANDRÉ, G. CHOLLET, K. KARPOUZIS and D. PELÉ Eds. Springer Berlin / Heidelberg, 197-210.

[Konstantinidis '09] Konstantinidis, E., Hitoglou-Antoniadou, M., Luneski, A., Bamidis, P., and Nikolaidou, M., 2009. Using affective avatars and rich multimedia content for education of children with autism. *PETRA '09: Proceedings of the 2nd International Conference on PErvasive Technologies Related to Assistive Environments*(Jul 01).

[Kopp '06] Kopp, S., Krenn, B., Marsella, S., Marshall, A., Pelachaud, C., Pirker, H., Thórisson, K., and Vilhjálmsson, H., 2006. Towards a Common Framework for Multimodal Generation: The Behavior Markup Language, J. GRATCH, M. YOUNG, R. AYLETT, D. BALLIN and P. OLIVIER Eds. Springer Berlin / Heidelberg, 205-217.

[Lafferty '01] Lafferty, J., McCallum, A., and Pereira, F.C., 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data.

[Lee '12] Lee, J. and Marsella, S., 2012. Modeling Speaker Behavior: A Comparison of Two Approaches, Y. NAKANO, M. NEFF, A. PAIVA and M. WALKER Eds. Springer Berlin / Heidelberg, 161-174.

[Lee '10a] Lee, J. and Marsella, S.C., 2010a. Predicting Speaker Head Nods and the Effects of Affective Information. *Multimedia, IEEE Transactions on 12*, 6 (Oct), 552-562.

[Lee '10b] Lee, J., Wang, Z., and Marsella, S., 2010b. Evaluating models of speaker head nods for virtual agents. In *Proceedings of the Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1 - Volume 1* (Richland, SC2010b), International Foundation for Autonomous Agents and Multiagent Systems, 1257-1264.

[Lee '10c] Lee, S., Carlson, G., Jones, S., Johnson, A., Leigh, J., and Renambot, L., 2010c. Designing an Expressive Avatar of a Real Person. In *Proceedings of the International Symposium on Intelligent Virtual Agent 2010* (2010c), Springer Berlin / Heidelberg, 64-76.

[Lee '13] Lee, S., LaFond, C.M., Johnson, A., Vincent, C., Leigh, J., and Renambot, L., 2013. A Virtual Patient to Assess Pediatric Intensive Care Unit (PICU) Nurses' Pain Assessment and Intervention Practices. In *Proceedings of the International Symposium on Intelligent Virtual Agent 2013* (Edinburgh, UK2013).

[Levine '10] Levine, S., Krähenbühl, P., Thrun, S., and Koltun, V., 2010. Gesture controllers. *SIGGRAPH '10: SIGGRAPH 2010 papers*(Jul 01).

[Lisetti '03] Lisetti, C., Nasoz, F., LeRouge, C., Ozyer, O., and Alvarez, K., 2003. Developing multimodal intelligent affective interfaces for tele-home health care. *Int. J. Hum.-Comput. Stud. 59*, 1-2, 245-255.

[Marcus '93] Marcus, M.P., Marcinkiewicz, M.A., and Santorini, B., 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational linguistics 19*, 2, 313-330.

[McClave '00] McClave, E.Z., 2000. Linguistic functions of head movements in the context of speech. *Journal of Pragmatics 32*, 7 (Jul 01), 855-878.

[McCrae '92] McCrae, R.R. and John, O.P., 1992. An introduction to the five-factor model and its applications. *Journal of personality 60*, 2 (Jul), 175-215.

[McKeown '11] McKeown, G., Valstar, M., Cowie, R., Pantic, M., and SCHRODER, M., 2011. The SEMAINE Database: Annotated Multimodal Records of Emotionally Coloured Conversations between a Person and a Limited Agent. *Affective Computing, IEEE Transactions on*, 99, 1.

[Mehrabian '96a] Mehrabian, A., 1996a. Analysis of the Big-five Personality Factors in Terms of the PAD Temperament Model. *Australian Journal of Psychology 48*, 2, 86-92.

[Mehrabian '96b] Mehrabian, A., 1996b. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in Temperament. *Current Psychology 14*, 261-292.

[Michael '08] Michael, N., Michael, K., Irene, A., and Hans-Peter, S., 2008. Gesture modeling and animation based on a probabilistic re-creation of speaker style. *ACM Trans. Graph. 27*, 1, 1-24.

[Microsoft '13] Microsoft. 2013. Microsoft Speech API (SAPI) 5.3 2013 [cited 7/1 2013]. Available from http://msdn.microsoft.com/en-us/library/ms723627(v=vs.85).aspx.

[Morency '08] Morency, L.-P., de Kok, I., and Gratch, J., 2008. Predicting Listener Backchannels: A Probabilistic Multimodal Approach, H. PRENDINGER, J. LESTER and M. ISHIZUKA Eds. Springer Berlin / Heidelberg, Berlin, Heidelberg, 176-190.

[Mori '70] Mori, M., 1970. Bukimi no tani [The uncanny valley]. *Energy 7*, 4, 33-35.

[Mumme '09] Mumme, C., Pinkwart, N., and Loll, F., 2009. Design and Implementation of a Virtual Salesclerk, Z. RUTTKAY, M. KIPP, A. NIJHOLT and H. VILHJÁLMSSON Eds. Springer Berlin / Heidelberg, 379-385.

[Neviarouskaya '10] Neviarouskaya, A., Prendinger, H., and Ishizuka, M., 2010. User study on AffectIM, an avatar-based Instant Messaging system employing rule-based affect sensing from text. *International Journal of Human-Computer Studies 68*, 7 (Jul 01).

[Niewiadomski '10] Niewiadomski, R., Prepin, K., Bevacqua, E., Ochs, M., and Pelachaud, C., 2010. Towards a smiling ECA: studies on mimicry, timing and types of smiles. *Proceedings of the 2nd international workshop on Social signal processing*, 65-70.

[Oat '07] Oat, C., 2007. Animated wrinkle maps. In *Proceedings of the ACM SIGGRAPH 2007 courses* (New York, NY, USA2007), ACM, 33-37.

[OGRE '13] OGRE. 2013. OGRE - Open Source 3D Graphics Engine 2013 [cited 7/1 2013]. Available from http://www.ogre3d.org.

[Okazaki '07] Okazaki, N., 2007. *CRFsuite: a fast implementation of Conditional Random Fields (CRFs).*

[Ossorio '76] Ossorio, P.G., 1976. *Clinical topics: A seminar in Descriptive Psychology.* Linguistic Research Institute.

[Pan '11] Pan, J., Wang, J.-m., Cao, S.-t., and Luo, X.-n., 2011. Interactive sign language synthesis based on adaptive display resolution visibility for ubiquitous devices. *Computer Animation and Virtual Worlds 22*, 2-3, 213-220.

[Picard '10] Picard, R.W., 2010. Emotion Research by the People, for the People. *Emotion Review 2*, 3 (Jul 09), 250-254.

[Pontier '08] Pontier, M. and Siddiqui, G., 2008. A Virtual Therapist That Responds Empathically to Your Answers, H. PRENDINGER, J. LESTER and M. ISHIZUKA Eds. Springer Berlin / Heidelberg, 417-425.

[Poppe '10] Poppe, R., Truong, K., Reidsma, D., and Heylen, D., 2010. Backchannel Strategies for Artificial Listeners, J. ALLBECK, N. BADLER, T.W. BICKMORE, C. PELACHAUD and A. SAFONOVA Eds. Springer Berlin / Heidelberg, 146-158.

[Rabiner '89] Rabiner, L., 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE 77*, 2, 257-286.

[Raij '09] Raij, A.B., Kotranza, A., Lind, D.S., and Lok, B.C., 2009. Virtual Experiences for Social Perspective-Taking. *Virtual Reality Conference, 2009. VR 2009. IEEE*(Apr 14), 99-102.

[Rossen '10] Rossen, B., Cendan, J., and Lok, B., 2010. Using Virtual Humans to Bootstrap the Creation of Other Virtual Humans, J. ALLBECK, N. BADLER, T.W. BICKMORE, C. PELACHAUD and A. SAFONOVA Eds. Springer Berlin / Heidelberg, 392-398.

[Shapiro '65] Shapiro, S.S. and Wilk, M.B., 1965. An analysis of variance test for normality (complete samples). *Biometrika*.

[Smith '10] Smith, C., Crook, N., Boye, J., Charlton, D., Dobnik, S., Pizzi, D., Cavazza, M., Pulman, S., de la Camara, R., and Turunen, M., 2010. Interaction Strategies for an Affective Conversational Agent, J. ALLBECK, N. BADLER, T.W. BICKMORE, C. PELACHAUD and A. SAFONOVA Eds. Springer Berlin / Heidelberg, 301-314.

[Stanford '13] Stanford. 2013. The Stanford NLP Parser. The Stanford NLP Group 2013 [cited 7/1 2013]. Available from http://nlp.stanford.edu/software/lex-parser.shtml.

[Strapparava '04] Strapparava, C. and Valitutti, A., 2004. WordNet-Affect: an Affective Extension of WordNet. In *Proceedings of the In Proceedings of the 4th International Conference on Language Resources and Evaluation* (2004), 1083-1086.

[Strauss '08] Strauss, M. and Kipp, M., 2008. ERIC: a generic rule-based framework for an affective embodied commentary agent. In *Proceedings of the Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems - Volume 1* (Richland, SC2008), International Foundation for Autonomous Agents and Multiagent Systems, 97-104.

[Swartout '10] Swartout, W., Traum, D., Artstein, R., Noren, D., Debevec, P., Bronnenkant, K., Williams, J., Leuski, A., Narayanan, S., Piepol, D., Lane, C., Morie, J., Aggarwal, P., Liewer, M., Chiang, J.-Y., Gerten, J., Chu, S., and White, K., 2010. Ada and Grace: Toward Realistic and Engaging Virtual Museum Guides, J. ALLBECK, N. BADLER, T.W. BICKMORE, C. PELACHAUD and A. SAFONOVA Eds. Springer Berlin / Heidelberg, 286-300.

[Valitutti '04] Valitutti, A., Strapparava, C., and Stock, O., 2004. Developing Affective Lexical Resources. *PsychNology Journal 2*, 1, 61-83.

[Vilhjálmsson '07] Vilhjálmsson, H., Cantelmo, N., Cassell, J., E Chafai, N., Kipp, M., Kopp, S., Mancini, M., Marsella, S., Marshall, A., Pelachaud, C., Ruttkay, Z., Thórisson, K., van Welbergen, H., and van der Werf, R., 2007. The Behavior Markup Language: Recent Developments and Challenges, C. PELACHAUD, J.-C. MARTIN, E. ANDRÉ, G. CHOLLET, K. KARPOUZIS and D. PELÉ Eds. Springer Berlin / Heidelberg, 99-111.

[Vinayagamoorthy '06] Vinayagamoorthy, V., Gillies, M., Steed, A., Tanguy, E., Pan, X., Loscos, C., and Slater, M., 2006. Building Expression into Virtual Characters. *Eurogrpahics State of The Art Report*, 21-61.

[Vinayagamoorthy '05] Vinayagamoorthy, V., Steed, A., and Slater, M., 2005. Building characters: Lessons drawn from virtual environments. *Proceedings of Toward Social Mechanisms of Android Science: A CogSci 2005 Workshop*, 119126.

[Wilcoxon '45] Wilcoxon, F., 1945. Individual comparisons by ranking methods. *Biometrics bulletin*.

[Wilkie '95] Wilkie, D., 1995. Facial expressions of pain in lung cancer. *Analgesia 1*, 2, 91-99.

[Woolf '07] Woolf, B.P., 2007. Building intelligent interactive tutors. *management-projet.org*.

[Yee '07] Yee, N., Bailenson, J.N., and Rickertsen, K., 2007. A meta-analysis of the impact of the inclusion and realism of human-like faces on user experiences in interfaces. *Proceedings of the SIGCHI conference on Human factors in computing systems*.

[Yuan '05] Yuan, X. and Chee, Y., 2005. Design and evaluation of Elva: an embodied tour guide in an interactive virtual art gallery: Research Articles. *Computer Animation and Virtual Worlds 16*, 2, 109-119.

[Zhang '10] Zhang, S., Wu, Z., Meng, H.M., and Cai, L., 2010. Facial Expression Synthesis Based on Emotion Dimensions for Affective Talking Avatar. *Modeling Machine Emotions for Realizing Intelligence*.

APPENDIX

UNIVERSITY OF ILLINOIS
AT CHICAGO

Office for the Protection of Research Subjects (OPRS)
Office of the Vice Chancellor for Research (MC 672)
203 Administrative Office Building
1737 West Polk Street
Chicago, Illinois 60612-7227

Exemption Granted

May 29, 2013

Sangyoon Lee, MS, MFA
Computer Science
842 W Taylor Street
M/C 152
Chicago, IL 60607
Phone: (312) 996-9768 / Fax: (312) 413-0024

RE: Research Protocol # 2013-0521
"Supervised Hybrid Expression Control Framework for a Lifelike Affective Avatar"
Sponsors: None

Dear Sangyoon Lee:

Your Claim of Exemption was reviewed on May 29, 2013 and it was determined that your research protocol meets the criteria for exemption as defined in the U. S. Department of Health and Human Services Regulations for the Protection of Human Subjects [(45 CFR 46.101(b)]. You may now begin your research.

<u>**Exemption Period:**</u> **May 29, 2013 – May 29, 2016**

Performance Site(s):	UIC
Subject Population:	Adult (18+ years) subjects only
Number of Subjects:	30

The specific exemption category under 45 CFR 46.101(b) is:

(2) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior, unless: (i) information obtained is recorded in such a manner that human subjects can be identified, directly or through identifiers linked to the subjects; and (ii) any disclosure of the human subjects' responses outside the research could reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, or reputation.

You are reminded that investigators whose research involving human subjects is determined to be exempt from the federal regulations for the protection of human subjects still have responsibilities for the ethical conduct of the research under state law and UIC policy. Please be aware of the following UIC policies and responsibilities for investigators:

1. Amendments You are responsible for reporting any amendments to your research protocol that may affect the determination of the exemption and may result in your research no longer being eligible for the exemption that has been granted.

2. Record Keeping You are responsible for maintaining a copy all research related records in a secure location in the event future verification is necessary, at a minimum these documents include: the research protocol, the claim of exemption application, all questionnaires, survey instruments, interview questions and/or data collection instruments associated with this research protocol, recruiting or advertising materials, any consent forms or information sheets given to subjects, or any other pertinent documents.

3. Final Report When you have completed work on your research protocol, you should submit a final report to the Office for Protection of Research Subjects (OPRS).

4. Information for Human Subjects UIC Policy requires investigators to provide information about the research protocol to subjects and to obtain their permission prior to their participating in the research. The information about the research protocol should be presented to subjects in writing or orally from a written script. When appropriate, the following information must be provided to all research subjects participating in exempt studies:
   a. The researchers affiliation; UIC, JBVMAC or other institutions,
   b. The purpose of the research,
   c. The extent of the subject’s involvement and an explanation of the procedures to be followed,

d. Whether the information being collected will be used for any purposes other than the proposed research,
e. A description of the procedures to protect the privacy of subjects and the confidentiality of the research information and data,
f. Description of any reasonable foreseeable risks,
g. Description of anticipated benefit,
h. A statement that participation is voluntary and subjects can refuse to participate or can stop at any time,
i. A statement that the researcher is available to answer any questions that the subject may have and which includes the name and phone number of the investigator(s).
j. A statement that the UIC IRB/OPRS or JBVMAC Patient Advocate Office is available if there are questions about subject's rights, which includes the appropriate phone numbers.

Please be sure to:

→Use your research protocol number (listed above) on any documents or correspondence with the IRB concerning your research protocol.

We wish you the best as you conduct your research. If you have any questions or need further help, please contact me at (312) 355-2908 or the OPRS office at (312) 996-1711. Please send any correspondence about this protocol to OPRS at 203 AOB, M/C 672.

Sincerely,

Charles W. Hoehne

Assistant Director

Office for the Protection of Research Subjects

cc: Peter C. Nelson, Computer Science, M/C 152

Andrew Johnson, Computer Science, M/C 152

VITA

EDUCATION

Ph.D., Electronic Visualization Laboratory (EVL), Computer Science Department, University of Illinois at Chicago (UIC) [Aug, 2006 – Aug, 2013]

M.F.A., Electronic Visualization (EV), School of Art and Design, UIC [Aug, 2003 – Aug, 2006]

M.E., Architectural Engineering Department,

Yonsei University, Seoul, South Korea [Sep, 1997 – Jun, 1999]

B.S., Architectural Engineering Department,

Yonsei University, Seoul, South Korea [Mar, 1993 – Feb, 1996]

PROFESSIONAL EXPERIENCE

Research Assistant, EVL, UIC [2006 – 2013]

Adjunct Assistant Professor, School of Art and Design, UIC [2010 – 2011]

Teaching Assistant, School of Art and Design, UIC [2003 – 2006]

Engineering Staff, Electronics & Telecommunications Research Institute, Daejeon, South Korea [2001 – 2002]

Researcher, Research & Development Department, Korea Game Development and Promotion Institute, Seoul, South Korea [1999 – 2000]

Teaching Assistant, Department of Architectural Engineering, Yonsei University, Seoul, South Korea [1997 – 1999]

PROFESSIONAL ACTIVITIES

Demonstrations:

- National Science Foundation (NSF) I/UCRC 2010 Annual Meeting, Washington DC (1/13-15/2010). Demonstrated Lifelike Avatar project
- National Science Foundation (NSF) I/UCRC 2009 Annual Meeting, Washington DC (1/7-9/2009). Lifelike Avatar project demo

Presentations:

- International Symposium on Parallel and Distributed Computing 2011 (ISPDC'11), Cluj-Napoca, Romania. Presented the paper "Performance Evaluation of Incremental Vector Clocks".
- International Virtual Agent 2010 (IVA'10), Philadelphia, PA, USA. Presented the paper "Designing an Expressive Avatar of a Real Person".

Guest Lecture & Hands-on Workshops:

- Guest lectures and hands-on workshops, "EVL Motion Capture System – Mocap Boot camp & Data Processing", for Computer Science courses (Human Augmentics 03/2013, Virtual Reality 09/2010, Computer Game Design 4/2009, Computer Animation 04/2008) and Art & Design courses (Advanced Electronic Visualization Critique 09/2008, Advanced 3D Modeling and Animation 02/2009)
- Guest lecture on "Advanced Character Design, Modeling, and Animation" for Advanced Interactive 3D class (11/2008)
- Guest lectures, "Stereoscopic rendering, character modeling, and 3D prototyping," for Biomedical Visualization class (02/2013, 02/2012)

Volunteer:

- Chicago Adler Planetarium: Deploying avatar-enabled museum application (2011–2012)

JOURNAL PUBLICATIONS

1. Gonzalez, A., DeMara, R., Hung, V., Leon-Barth, C., Elvir, M., Hollister, J., Kobosko, S., Leigh, J., Johnson, A., Jones, S., Carlson, G., **Lee, S.**, Renambot, L., Brown, M., "Passing an Enhanced Turing Test – Interacting with Lifelike Computer Representations of Specific Individuals," Journal of Intelligent Systems (JISYS), (accepted on 17 Apr, 2013)

CONFERENCE PUBLICATIONS

2. **Lee, S.**, LaFond, C., Johnson, A., Vincent, C., Leigh, J., Renambot, L., "“A Virtual Patient to Assess Pediatric Intensive Care Unit (PICU) Nurses' Pain Assessment and Intervention Practices," International Virtual Agent (IVA) 2013, Edinburgh, UK, August 2013
3. LaFond, C., **Lee, S.**, Johnson, A., Vincent, C., "Virtual Human Vignettes for Evaluating Pediatric Nurses' Pain Assessment and Intervention Choices," Midwest Nursing Research Society 37th Annual Research Conference, Chicago, IL, March 2013
4. **Lee**, **S.,** Kshemkalyani, A.D., Shen, M., "Performance Evaluation of Incremental Vector Clocks," International Symposium on Parallel and Distributed Computing, Cluj-Napoca, Romania, July 2011.
5. **Lee, S.**, Carlson, G., Jones, S., Johnson, A., Leigh, J., Renambot, L., "Designing an Expressive Avatar of a Real Person," International Virtual Agent (IVA) 2010, Philadelphia, PA, September 2010
6. Sun, Y., Leigh, J., Johnson, A., **Lee, S.**, "Articulate: A Semi-automated Model for Translating Natural Language Queries into Meaningful Visualizations," *Proceedings of 10th International Symposium on Smart Graphics, Banff, Canada, June 201*0
7. Chen, Y., Hur, H., **Lee, S.**, Leigh, J., Johnson, A., Renambot, L., "Case Study - Designing An Advanced Visualization System for Geological Core Drilling Expeditions," Proceedings of the ACM Conference on Human Factors in Computing Systems 2010 (CHI 2010), Atlanta, GA, April 2010
8. Chen, Y., **Lee, S.**, Hur, H., Leigh, J., Johnson, A., Renambot, L., "Design an Interactive Visualization System for Core Drilling Expeditions Using Immersive Empathic Method," Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Boston, MA, April 2009
9. DeMara, R., Gonzalez, A., Hung, V., Leon-Barth, C., Dookhoo, R., Jones, S., Johnson, A., Leigh, J., Renambot, L., **Lee, S.**, Carlson, G., "Towards Interactive Training with an Avatar-based Human-Computer Interface," Industry Training, Simulation & Education Conference (I/ITSEC) 2008, Orlando, Florida, December 2008
10. Rao, A., Chen, Y., **Lee, S.**, Leigh, J., Johnson, A., Renambot, L., "Corelyzer: Scalable Geologic Core Visualization using OSX, Java and OpenGL", Apple's Worldwide Developers Conference 2006, July 2006

MEDIA PUBLICATIONS

11. Work featured on NOVA scienceNOW "Can We Live Forever?" January 2011 <http://www.pbs.org/wgbh/nova/tech/jason-leigh-avatars.html>
12. Work featured on Discovery Science channel "Future of: Immortal Avatars," September 2009 <http://science.discovery.com/videos/popscis-future-of-immortal-avatars.html>

EXHIBITIONS

The NEW ECO series, Gosia Koscielak Studio & Gallery, Chicago, Jan 2007

"PAN – A Life Force", the first solo exhibition in the NEW ECO international art curated series examining new ecology and media phenomena.

Drawing Current – EV M.F.A. Thesis Show, Great Space, UIC, Chicago, May 2006

"PAN", Electronic Visualization M.F.A. Thesis Group Exhibition.

Year End Show of EV, Center for Virtual Reality Arts, UIC, Chicago, May 2005

"3.D.UO.PAD" and "Bouncing Space", Annual exhibition of EV program.

2005 Indiana IDEAS Festival, Bloomington, Indiana, April 2005

"3.D.UO.PAD", Best work award in Virtual Reality Session hosted by Indiana University.

EV Show, Center for Virtual Reality Arts, UIC, Chicago, May 2004

"The Creation – From Eight Signs of Divination to the World", a part of a portal application of Virtual Reality Grid.

AWARDS

- UIC Image of Research Awards 2010 for "Avatar – A Virtual Human", 1st place winner
- Keio University's Keio Research Institute Digital Art Awards 2006, "PAN", Honorable Mention in Interactive category
- Indiana IDEAS Festival 2005, "3.D.UO.PAD", Best work in Virtual Reality Session

SKILLS

Operating Systems: Microsoft Windows, Mac OSX, and Linux
Programming Languages: C/C++, Java, Python. Also Multi-threading, Network Programming
Modeling and Animation: Vicon Motion Capture, MotionBuilder, Maya, ZBrush, Blender
Computer Graphics: OpenGL, DirectX programming for real-time application; Advanced Shader programming, General Purpose GPU (GPGPU) programming

PROFESSIONAL MEMBERSHIP

Member, Association for Computer Machinery (ACM)
Member, Institute of Electrical and Electronics Engineers (IEEE)