

Evaluating the Uncanny Valley Theory Based on Human Attitudes

Louis Laja Uggah¹, Azaini Manaf¹

¹ Faculty of Applied & Creative Arts, UNIMAS, Sarawak, Malaysia.

Abstract

Background The uncanny valley theory is an idea that was proposed by Masahiro Mori in 1970 regarding the psychological effects of lifelike robotics (Mori, 1970). The uncanny valley is a phenomenon that occurs in animation and robotics, wherein things that look extremely similar to the human face, but differ slightly from its natural appearance or from its natural movements and expressions, are perceived to be disturbing, uncanny, and revolting (Mewes & Heloir). This study aims to analyze participants' attitudes towards digital characters in order to understand how the uncanny valley affects audiences. Mori's graph has been criticized on the grounds that familiarity is difficult to define – that it is difficult to determine which emotion accurately represents the opposite of familiarity, and that the word “familiarity” itself may not actually be an accurate description of a positive human response to human-like entities (Ho, MacDorman, & Pramono, 2008). The word “likability” has been proposed as an alternative translation of Mori's original word, because it is claimed by some to be a more accurate representation of the phenomena Mori was describing in his original article (Tinwell, Grimshaw, & Williams, 2011).

Methods This study investigates attitudes toward digital stimuli through employed a quantitative approach based on semantic differential questionnaires. Perceived Humanness and Familiarity indices, based on indices developed by Ho & Macdorman (2011), were used to determine overall perception of human-likeness and familiarity toward all the stimuli, while other subscales, were determined based on the following six factors: hair animation, eye animation, lip sync, lighting, facial expression and the body movement of all the stimuli. The study was conducted in the conference hall of the Yahos Training Centre in Kuching, Malaysia. We applied a systematic sampling method for this study, because we preferred participants with a moderate knowledge of digital characters – either in games or movies. Participants consisted of gamers, university students, moviegoers and creative professionals aged between 18 and 35. Participants were invited via email, on Facebook and over the phone. Participants rated all the stimuli based on the questionnaires.

Conclusion Our findings have concluded that digital Emily has surpassed the uncanny valley in terms of realism and familiarity compared to other stimuli with high ratings in terms of lifelikeness, organicity and familiarity. Animation style and techniques should not focus on avoiding realistic animation but instead on other factors such as target audiences and the animation's genre.

Keywords Digital Animations, Stylized Animations, Lifelike Animation, Psychology

Citation: Louis, L. U., & Azaini, M. (2015). Evaluating the Uncanny Valley Theory Based on Human Attitudes. *Archives of Design Research*, 28(2), 27-41.

<http://dx.doi.org/10.15187/adr.2015.05.28.2.27>

Received : Apr. 03. 2015 ; **Reviewed :** Apr. 06. 2015 ; **Accepted :** Apr. 13. 2015

pISSN 1226-8046 **eISSN** 2288-2987

*Corresponding author:

Louis Laja Uggah

(louislaja83@gmail.com)

This work was done as a part of phd candidacy research in UNIMAS

Copyright : This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted educational and non-commercial use, provided the original work is properly cited.

1. Introduction

What is the uncanny valley?

The uncanny valley theory is an idea that was proposed by Masahiro Mori in 1970 regarding the psychological effects of lifelike robotics (Mori, 1970). The uncanny valley is a phenomenon that occurs in animation and robotics, wherein things that look extremely similar to the human face, but differ slightly from its natural appearance or from its natural movements and expressions, are perceived to be disturbing, uncanny, and revolting (Mewes & Heloir).

The name the uncanny valley refers to a point on a graph that plots the human likeness of a robot or virtual character in relation to its perceived familiarity. At first, familiarity increases as human likeness increases, but at a certain point, when the likeness is perceived as extremely similar and yet not similar enough, the graph takes a swift dive into negative values of familiarity. This dip in the graph is the uncanny valley. The situation, however, does not last long. As the robot's human likeness continues to grow, negative perceptions fade, and once again the robot is perceived as more familiar (Mori, 1970).

Mori's graph has been criticized on the grounds that familiarity is difficult to define – that it is difficult to determine which emotion accurately represents the opposite of familiarity, and that the word familiarity itself may not actually be an accurate description of a positive human response to human-like entities (Ho, MacDorman, & Pramono, 2008). The word likability has been proposed as an alternative translation of Mori's original word, because it is claimed by some to be a more accurate representation of the phenomena Mori was describing in his original article (Tinwell, Grimshaw, & Williams, 2011).

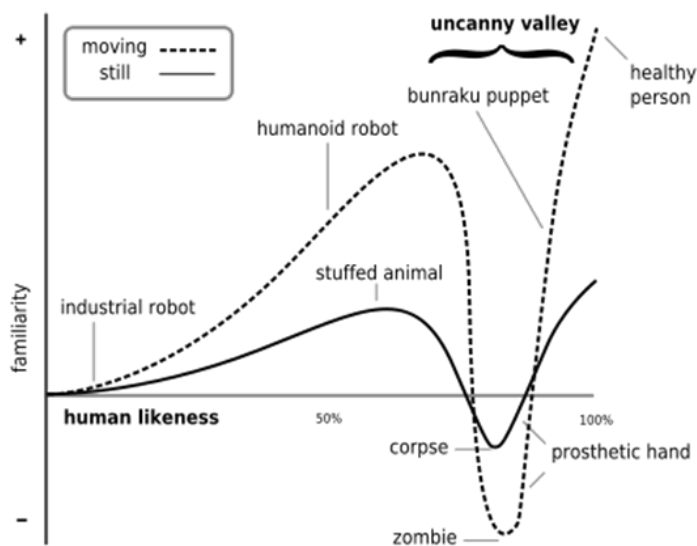


Figure 1 Uncanny Valley Graph (Mori, 1970)

There are several other theories regarding different factors that may influence or cause the uncanny valley phenomenon. One suggestion is that motion increases the uncanniness of the

likeness. As Mori points out, mannequins are lifelike but harmless; however, if they were to move, they would be terrifying (Mori, 1970). This is partially linked to the speed of natural human movement. Laughter, for instance, is a natural human action, but when the changing facial expressions that make up laughter are slowed down they appear to be more like a grimace. If an animated character or an android is designed to laugh but it does it just a little too slowly, the effect will be uncanny (Mori, 1970).

Time is also an important factor. When seen for very short periods of time, human-like androids are not perceived as uncanny – in fact, they are sometimes not even identified as androids at all – but when seen for more than a few seconds they soon begin to look creepy (Mewes & Heloir). Mori further suggests that the reason for the perception of human-like entities as uncanny is related to our instinct for self-preservation: at death we retain our human appearance but subtle changes, such as the face growing pale, make us look uncanny. The fear or disgust felt at the sight of artificial human-like items in the uncanny valley may simply be the same as our natural fearful response to death (Mori, 1970). Exposure to these images also caused them to exhibit fear and feelings of xenophobia (Ho, MacDorman, & Pramono, 2008).

2. Significance of The Research

This research confirmed that the Digital Emily Project developed by Image Metrics is the most realistic digital character perceived by human audiences and familiarity. The methods and techniques applied by Image Metrics to produce this digital character can be referred to by animators as a guideline for their animation production process especially those aiming to produce realistic digital characters in order to avoid the uncanny responses from audiences. Animators and animation studios are recommended not to avoid realistic animations just to avoid the uncanny valley but instead focus on improving their animation techniques such as facial animation, rigging and rendering.

3. Method

In order to understand the effects of the uncanny valley on participants' attitudes towards digitally-animated characters, this study investigates attitudes towards digital stimuli. The study employed a quantitative approach based on semantic differential questionnaires. Perceived Humanness and Familiarity indices, based on indices developed by Ho & Macdorman (2011), were used to determine overall perception of human-likeness and familiarity towards all the stimuli, while other subscales, such as Organic-Mechanical, Lifelike-Fake and Familiar-Eerie, were determined based on the following six factors: hair animation, eye animation, lip sync, lighting, facial expression and the body movement of all the stimuli. The study was conducted in the conference hall of the Yagos Training Centre in Kuching, Malaysia. The venue was selected because it was the most convenient and cost

effective for the organizer. We used a systematic sampling method for this study, because we preferred participants with a moderate knowledge of digital characters – either in games or movies. Participants consisted of gamers, university students, moviegoers and creative professionals aged between 18 and 35. Participants were invited via email, on Facebook and over the phone. Participants rated all the stimuli based on the questionnaires. Responses were saved in an Excel file attached to the e-mail. Of the 300 individuals originally invited to take part, only 229 responded to our invitation. All the participants were Malaysian.

4. Stimuli

The stimuli selected for this study consisted of the following five female digital characters and a female human control stimulus: Ellie from *The Last of Us*, Clementine from *The Walking Dead*, Madison Paige from *Heavy Rain*, Elastigirl from *The Incredibles*, Digital Ellie from Image Metrics Digital Ellie Project, and finally a real human being as the control stimulus.

Female digital characters – most of them the heroines of movies or games – were chosen for the study in order to avoid response bias. Here is brief synopsis of each of the chosen characters.

Ellie (The Last of Us)

Ellie is a 3D digital playable character voiced and motion captured by Ashley Johnson and developed by Naughty Dog for the action-adventure survival horror video game *The Last of Us*. Ellie – a 14-year-old survivor of the apocalypse – and her companion Joel try to survive in the post-apocalyptic environment, where they battle with zombies and fellow humans. It is crucial to the success of the game that players feel an emotional connection with Ellie and Joel, not least because they will be together for the 6 hours of the game. We have included Ellie in this experiment in order to find out how participants rate a highly realistic playable 3D digital character based on their attitudes.

Clementine (The Walking Dead)

Clementine is a playable 2D digital character from *The Walking Dead: Season Two*, developed by Telltale in 2013. As with *The Last of Us*, this game takes place post zombie apocalypse – this time in Georgia. The player decides the outcome of the storyline based on decisions taken throughout the game. Apart from the importance of the storyline, the game also places strong emphasis on character development. Although the main character, Clementine, is only 11 years old, she displays remarkable intelligence and maturity for her age. She was introduced to the game in Season One with two other leading characters Lee and Kenny. After Lee's death at the end of Season One, Clementine was upgraded to a playable character in Season Two. We have included Clementine as one of the stimuli in order to assess participants' attitudes towards a stylized 2D digital character.

Elastigirl (The Incredibles)

Helen Parr, also known as Elastigirl, is the wife of Mr. Incredible in the movie *The Incredibles*. She was developed by Disney as a stylized 3D digital character and played an important role as super heroine and mother at the same time. Elastigirl is included in this study in order to measure participants' attitudes towards a stylized 3D digital character.

Digital Emily (Image Metrics)

Digital Emily was developed in 2008 by USC and Image Metrics using Light Stage 5 technology (Alexander et al., 2009). The non-playable digital character was based on "The Young and The Restless" actress, Emily O'Brien. Digital Emily is currently considered to be one of the most photo-real digital actresses ever created (Alexander et al., 2009). We have included Digital Emily in this experiment in order to measure participants' attitudes towards photo-real digital characters that blur the edges between woman and machine.

Madison Paige (Heavy Rain)

Madison Paige is one of four playable characters (the other three are Ethan Mars, Norman Jayden, and Scott Shelby) and one of the three main protagonists in *Heavy Rain*. Madison is a young journalist living alone in the city. The character's facial features were modeled on those of British model Jacqui Ainsley while her facial movements were performed by American actress Judi Beecher. The Madison Paige character is considered to be highly realistic, and we have included it in our study in order to measure attitudes to a highly realistic playable 3D character.

5. Exploratory Factor Analysis

Principal component analysis was used because it helps to reduce the dimension of the data set by reducing the data into its basic components. The Kaiser-Meyer-Olkin measure of sampling adequacy was .934 which is above the recommended value of .6 while the Bartlett's test of sphericity was significant $p < .05$. The communalities for all the indices were all above 0.3 (Table 1), which concluded that all the items shared some common variance with other items. Based on the Pattern matrix table, all subscales are each loaded into its own factor namely :- Facial, Rig, Hair, Eyes, Lighting, Lip Sync. All the subscales loaded from the range of 0.65 until 0.97

The construct reliability and validity of the measurement model of this study were also calculated using SPSS software to check for its internal consistency. Construct validity test, which is essential to the perceived overall validity of the measurements are divided into two subtypes which are convergent validity and discriminant validity. Convergent validity tests determine whether the factors that are expected to be related are in actually related while discriminant validity test determine whether the factors which are supposed to be unrelated

are actually unrelated. To test the factors convergent validity, we refer to the Patter Matrix which were extracted from SPSS output. Each of the 8 factors achieved an average loading of above 0.7, which indicated that the factors are related. For the discriminant validity tests, the Pattern Matrix indicated that there are no cross loadings among the factors. The Component Correlation Matrix revealed that none of the factors have a correlation of greater than 0.7 which indicated that there are no correlations between the factors. These discriminant validity tests concluded that each factors are unrelated which each other. Cronbach Alpha test is crucial to determine the validity of the study's psychometric test. The reliability analysis showed in the table below (Table 1) revealed that all of the factors achieved an average Cronbach Alpha's of above 0.7 which is considered good in terms of internal consistency; therefore we decided to maintain all the subscale indices for all of the factors.

Table 1 Reliability Statistics

Cronbach's Alpha	N of Items	Factor
.887	3	Lip Sync
.881	3	Lighting
.876	3	Hair
.799	3	Eyes
.849	3	Movement
.758	3	Facial

6. Confirmatory Factor Analysis

This first step in our confirmatory factor analysis is to develop a decent measurement mode (Appendix A) based on the pattern matrix obtained from SPSS. Each subscale is loaded into its own factor, namely: Facial, Rig, Hair, Eyes, Texture and Lip Sync.

The next step in our confirmatory factor analysis is an analysis of the measurement invariance of latent constructs. Rens van de Schoot et al. (2012) stated that analysis of measurement invariance is important in determining whether latent variables are valid across groups. For this study, groups are divided into 3 (realistic, stylized, and all groups) based on the stimuli. Appendix A shows that at least one of the loadings was non-significant. In this study our standard loadings estimates for all factors were greater than 0.7 except for sub-scales Facial-2 which achieved a factor loading of 0.66, which is acceptable.

The final measurement model for exogenous and endogenous was tested by asseing the fit indices. The CMIN/df for this model was 1.62 which indicated a model fit. The comparative fit index (CFI) was 0.993 and goodness of fit index (GFI) was 0.982. The adjusted goodness of fit index (AGFI) was 0.974. The mean square error of approximation (RMSEA) was 0.023. The CFI, GFI, AGFI, RMSEA for this measurement model all met the criteria for a model fit (Hu and Bentler, 1999). We then analyze the average variance of extracted (AVE) values for all items. All the items' AVE ranged from 0.51 to 0.72 which is above the cut-off of 0.5. The CFA analysis confirmed that the data fit the hyphothesize measurement model.

7. Data Analysis

Digital Eyes (Fake-Lifelike)

Based on the Fake-Lifelike scale, the one-way ANOVA revealed that the stimuli are significantly different from each other in terms of Digital Eyes (EYE1), with $f=53.5$ and $p<0.05$. The mean plot for the Digital Eye factor based on the Fake-Lifelike scale shows a significant dip in stimulus 5 and stimulus 1 mean rating was different from the rest of the groups. This speculation could be checked by the Tukey test. The Tukey results show that stimulus 1 is significantly different from all the other groups. Stimuli 2 and 3 are not significantly different from each other, while stimuli 4 and 5 are significantly different from all the other stimuli. The control stimulus achieved the highest ratings in terms of lifelike, with a mean rating of 4.72. Next was stimulus 4, with a mean rating of 4.16. Stimulus 5 was rated the least lifelike, with a mean ratings of 3.44. Stimuli 2 and 3 were not significantly different from each other achieving mean ratings of 3.73 and 3.78 respectively.

Facial Expression (Fake-Lifelike)

In terms of facial expression based on the Fake-Lifelike scale (Facial1), the one-way ANOVA revealed that $f=131.6$ and $p<0.05$. The mean plot for the facial expression factor based on the Fake-Lifelike scale shows a significant dip in stimulus 2. The Tukey results showed that stimulus 2, stimulus 3, and stimulus 5 are significantly different from all the other groups, while stimuli 1 and 4 are not significantly different from each other. The control stimulus and Digital Emily were rated as the most lifelike in terms of facial expression, with mean ratings of 4.74 and 4.73 respectively. Stimulus 2 achieved the lowest mean ratings on the Fake-Lifelike scale for facial expression, with a mean rating of 3.28.

Hair Animation (Fake-Lifelike)

The one-way ANOVA for the Fake-Lifelike scale based on the stimuli's hair animation indicated that $f=216$ and $p<0.05$. The mean plot for the hair animation factor based on the Fake-Lifelike scale shows a significant dip in stimulus 2. The Tukey results showed that stimuli 2, 3, and 5 are significantly different, while stimuli 1 and 4 are not significantly different from each other. The control stimulus and Digital Emily were rated as the most lifelike again in terms of hair animation, with a mean rating of 4.5 and 4.6 respectively. Stimulus 2 achieved the lowest mean ratings for Fake-Lifelike scale in terms of hair animation with a low mean rating of 2.58.

Lip Sync (Fake-Lifelike)

The one-way ANOVA revealed that all the stimuli are significantly different from each other in terms of lip syncing based on the Fake-Lifelike scale (Lip1) with $f=126.43$ and $p<0.05$. The mean plot for the lip sync factor based on the Fake-Lifelike scale again shows a significant dip in stimulus 2. The Tukey results showed that stimuli 1, 2 and 5 are significantly different,

while stimuli 3 and 4 are not significantly different from each other. Again, the control stimulus was rated as the most lifelike in terms of lip syncing, with a mean rating of 4.7 followed by stimuli 3 and 4, achieving 4.17 and 4.18 respectively. Stimulus 2 achieved the lowest mean ratings on the Fake-Lifelike scale in terms of hair animation, with a mean rating of 3.01.

Body movements (Fake-Lifelike)

The one-way ANOVA revealed that all the stimuli are significantly different from each other in terms of body movement on the Fake-Lifelike scale (Hair1) with $f=212.7$ and $p<0.05$. The mean plot for the body movement factor based on the Fake-Lifelike scale again shows a significant dip at stimulus 2. The Tukey results showed that stimuli 2, 3, and 5 are significantly different, while stimuli 1 and 4 are not significantly different from each other. The control stimulus and Digital Emily were again rated as the most lifelike in terms of body movement, with mean ratings of 4.7 and 4.5 respectively. Stimulus 2 achieved the lowest mean ratings on the Fake-Lifelike scale in terms of hair animation, with a low mean rating of 3.03.

Lighting and Rendering (Fake-Lifelike)

The one-way ANOVA revealed that all stimuli are significantly different from each other in terms of lighting and rendering based on the Fake-Lifelike scale (Hair1) with $f=177.7$ and $p<0.05$. The mean plot for the lighting and rendering factor based on the Fake-Lifelike scale shows a significant dip in stimuli 2 and 5. The Tukey results showed that stimuli 2 and 5 are significantly different while stimuli 1, 3, and 4 are not significantly different from each other. Again, the Digital Emily stimulus was rated as the most lifelike in terms of lighting and rendering, with a mean rating of 4.7 followed by stimuli 1 and 3 with mean ratings of 4.67 and 4.64 respectively. Stimulus 2 achieved the lowest mean ratings on the Fake-Lifelike scale in terms of hair animation with a low mean rating of 3.1

Digital Eyes (Mechanical-Organic)

The one-way ANOVA revealed that all the stimuli are significantly different from each other in terms of the digital characters' eyes based on the Mechanical-Organic scale (Eyes2) with $f=93.2$ and $p<0.05$. The mean plot for the digital characters' eyes factor based on the Mechanical-Organic scale shows a significant dip in stimuli 2 and 5. The Tukey results showed that stimuli 1, 3, and 4 are significantly different, while stimuli 2 and 5 are not significantly different from each other. The control stimulus and Digital Emily were rated as the most organic in terms of eye animation, with a high mean rating of 4.73. This is followed by stimuli 3 and 4 which achieved mean ratings of 3.6 and 4.1 respectively. Stimuli 2 and 5 both achieved low mean ratings, with 3.2 and 3.3.

Facial Expression (Mechanical-Organic)

The one-way ANOVA revealed that all the stimuli are significantly different from each other in terms of facial expression based on the Mechanical-Organic scale (Facial2) with $f=45.8$ and $p<0.05$. The mean plot for the eye factor based on the Mechanical-Organic scale shows a significant dip in stimulus 2. The Tukey results showed that stimulus 3 is not significantly different from stimulus 5. Stimuli 1, 2, and 4 are significantly different. The control stimulus and Digital Emily were rated as the most organic in terms of facial expression, both with a high mean rating of 4.74, followed by stimulus 4 with a mean rating of 4.48. Stimuli 3 and 5 achieved mean ratings of 4.24 and 4.29 respectively. Stimulus 2 achieved the lowest mean ratings, with 3.71.

Hair Animation (Mechanical-Organic)

The one-way ANOVA revealed that all the stimuli are significantly different from each other in terms of hair animation based on the Mechanical-Organic scale (Hair2) with $f=230.67$ and $p<0.05$. The mean plot for hair animation factor based on the Mechanical-Organic scale shows a significant dip in stimuli 2 and 5. The Tukey results showed that all the stimuli are significantly different from each other. The control stimulus and Digital Emily were rated as the most organic in terms of hair animation, with a high mean rating of 4.78 and 4.51. This is followed by stimuli 3 and 5, which achieved mean ratings of 4.1 and 3.4 respectively. Stimulus 2 achieved the lowest mean ratings, with 2.5.

Lip Sync (Mechanical-Organic)

The one-way ANOVA revealed that all the stimuli are significantly different from each other in terms of lip syncing based on the Mechanical-Organic scale (Lip2) with $f=137$ and $p<0.05$. The mean plot for the hair animation factor based on the Mechanical-Organic scale shows a significant dip in stimulus 2. The Tukey results showed that stimuli 1, 2, and 5 are significantly different from each other. Stimuli 3 and 4 are not significantly different with each other, with mean ratings of 4.2 and 4.3 respectively. Stimulus 2 was the lowest with a mean rating of 2.98, while the control stimulus achieved the highest mean ratings.

Body Movements (Mechanical-Organic)

The one-way ANOVA revealed that all the stimuli are significantly different from each other in terms of body movement based on the Mechanical-Organic scale (Move2) with $f=212.7$ and $p<0.05$. The mean plot for the body movement factor based on the Mechanical-Organic scale shows a significant dip in stimulus 2. The Tukey results show that stimuli 3, 4 and 5 are not significantly different from each other, while stimuli 1 and 2 are significantly different. Stimulus 3 achieved a mean rating of 4.53, which is not significantly different from stimulus 4 with a mean rating of 4.41. Stimulus 4 is not significantly different with stimulus 5 with mean ratings of 4.41 and 4.3 respectively. Stimulus 2 achieved the lowest mean rating with 3.03.

Lighting & Rendering (Mechanical-Organic)

The one-way ANOVA revealed that all the stimuli are significantly different from each other in terms of lighting and rendering based on the Mechanical-Organic scale (Hair2) with $f=175.8$ and $p<0.05$. The mean plot for the hair animation factor based on the Mechanical-Organic scale shows a significant dip in stimuli 2 and 5. The Tukey results showed that all of the stimuli are significantly different from each other. Stimuli 2, 3, and 5 are significantly different, while stimuli 1 and 4 are not significantly different from each other. Stimuli 1 and 4 achieved high mean ratings of 4.64 and 4.69 respectively.

Digital Eyes (Eerie-Familiar)

The one-way ANOVA revealed that all the stimuli are significantly different from each other in terms of digital eyes based on the Eerie-Familiar scale (Eyes3) with $f=75.7$ and $p<0.05$. The mean plot for the hair animation factor based on the Mechanical-Organic scale shows a significant dip in stimuli 2 and 5. The Tukey results showed that stimuli 1, 3, and 4 are significantly different, while Stimuli 2 and 5 are not significantly different from each other. Stimulus 1 achieved the highest mean rating of 4.75 followed by stimulus 4 with 4.34. Stimulus 5 achieved the lowest mean rating, with 3.53.

Facial Expression (Eerie-Familiar)

The one-way ANOVA revealed that all the stimuli are significantly different from each other in terms of facial expression based on the Eerie-Familiar scale (Facial3) with $f=87.5$ and $p<0.05$. The mean plot for the hair animation factor based on the Mechanical-Organic scale shows a significant dip in stimulus 2. The Tukey results showed that stimuli 1 and 4 are not significantly different from each other. Stimuli 2, 3, and 5 are not significantly different. Stimulus 1 achieved the highest mean rating of 4.74 followed closely by stimulus 4, with 4.73. Stimulus 5 achieved the lowest mean rating with 3.53.

Hair Animation (Eerie-Familiar)

The one-way ANOVA revealed that all the stimuli are significantly different from each other in terms of hair animation based on the Eerie-Familiar scale (Hair3) with $f=240.9$ and $p<0.05$. The mean plot for the hair animation factor based on the Mechanical-Organic scale shows a significant dip in stimulus 2. The Tukey results showed that stimuli 1, 3, and 4 are not significantly different from each other. Stimulus 4 achieved the highest mean rating of 4.85, while stimulus 1 achieved a mean rating of 4.55. Stimulus 5 achieved the lowest mean rating, with 2.81.

Digital Eyes (Eerie—Familiar)

The one-way ANOVA revealed that all the stimuli are significantly different from each other in terms of lip syncing based on the Eerie-Familiar scale (Eyes3) with $f=170.1$ and $p<0.05$. The mean plot for the hair animation factor based on the Mechanical-Organic scale shows a

significant dip at stimuli 2 and 5. The Tukey results showed that stimuli 3, 4, and 5 are not significantly different from each other, while stimuli 1 and 2 are not significantly different from each other. Stimulus 1 achieved the highest mean rating of 4.86 followed by stimulus 3, with 4.17. Stimulus 2 achieved the lowest mean rating, with 2.99.

Body Movements (Eerie-Familiar)

The one-way ANOVA revealed that all stimuli are significantly different from each other in terms of physical movement based on the Eerie-Familiar scale (Move3) with $f=249.4$ and $p<0.05$. The mean plot for the hair animation factor based on the Mechanical-Organic scale shows a significant dip in stimulus 2. The Tukey results showed that stimuli 3 and 5 are not significantly different from each other; neither are stimuli 1 and 4. Stimulus 2 is significantly different from the other stimuli. Stimulus 1 achieved the highest mean rating of 4.77 followed by stimulus 4 with 4.63. Stimulus 2 achieved the lowest mean rating, with 3.08.

Lighting & Rendering (Eerie-Familiar)

The one-way ANOVA revealed that all the stimuli are significantly different from each other in terms of lighting and rendering based on the Eerie-Familiar scale (Txt3) with $f=207.33$ and $p<0.05$. The mean plot for the lighting and rendering factor based on the Mechanical-Organic scale shows a significant dip in stimuli 2 and 5. The Tukey results showed that stimuli 1 and 4 are not significantly different from each other, while stimuli 2, 3, and 5 are significantly different. Stimulus 4 achieved the highest mean rating of 4.79 followed by stimulus 1 with 4.61. Stimulus 2 achieved the lowest mean rating, with 2.92.

8. Discussion

Based on our data for all the other stimuli, Digital Emily's ratings were the most similar to those of the human control. In terms of life-like facial expression, hair animation and body movement, Digital Emily's ratings were not significantly different from the control. Most participants also rated Digital Emily as not significantly different in terms of familiarity in body movement and lighting and rendering techniques. Most participants also rated digital Emily's lighting and rendering techniques "as organic as the control". Furthermore, Digital Emily achieved higher ratings than all of the stimuli, including the control, in terms of lifelike body movement and rendering techniques as well as familiar hair animation and rendering techniques. Most participants also rated digital Emily's lighting and rendering techniques as organic as the control stimuli. Besides that, digital Emily also achieved higher ratings than all of the stimuli including the control stimulus in terms of lifelike body movements and rendering techniques as well as familiar hair animation and rendering techniques. Based on the Mechanical-Organic scale, the lip syncing of stimuli 3 & Digital Emily were not significantly different from each other based on participants' attitudes.

9. Conclusion

Our findings have concluded that digital Emily has surpassed the uncanny valley in terms of realism and familiarity compared to other stimuli with high ratings in terms of lifelike, organic and familiarity. This digital stimulus also achieved very similar ratings with a real human being stimulus. Based on these findings, we recommend animators not to avoid realistic digital characters and opt for stylized digital characters in order to avoid the uncanny valley response from audiences. Animation style and techniques should not focus on avoiding realistic animation but instead on other factors such as target audiences and the animation's genre.

10. Implications of The Research

Research carried out by Macdorman, Green Ho & KochKock (2008), indicated that realistic digital characters were not necessarily eerie. This conflicts with Mori's (1970) theory of the uncanny valley. Seyama and Nagayama, (2007) also stated that audiences are not repulsed by artificiality as it approaches lifelikeness, provided that the level of artificiality is uniform and there are no jarring elements. Our findings support their conclusion, indicating that highly realistic digital characters such as Digital Emily can avoid the uncanny valley and achieve high ratings for in terms of familiarity and realism.

11. Future Research

Future research should focus on identifying the key factors in digitally animated characters, such as hair animation and digital eye movement that cause audiences to feel revulsion and discomfort. By doing so, novice animators will be able to identify which factors have the most significant effects on audiences. We proposed the structural equation model method because this model enables researchers to analyze the regression weights of the factors and draw conclusions based on the structural model.

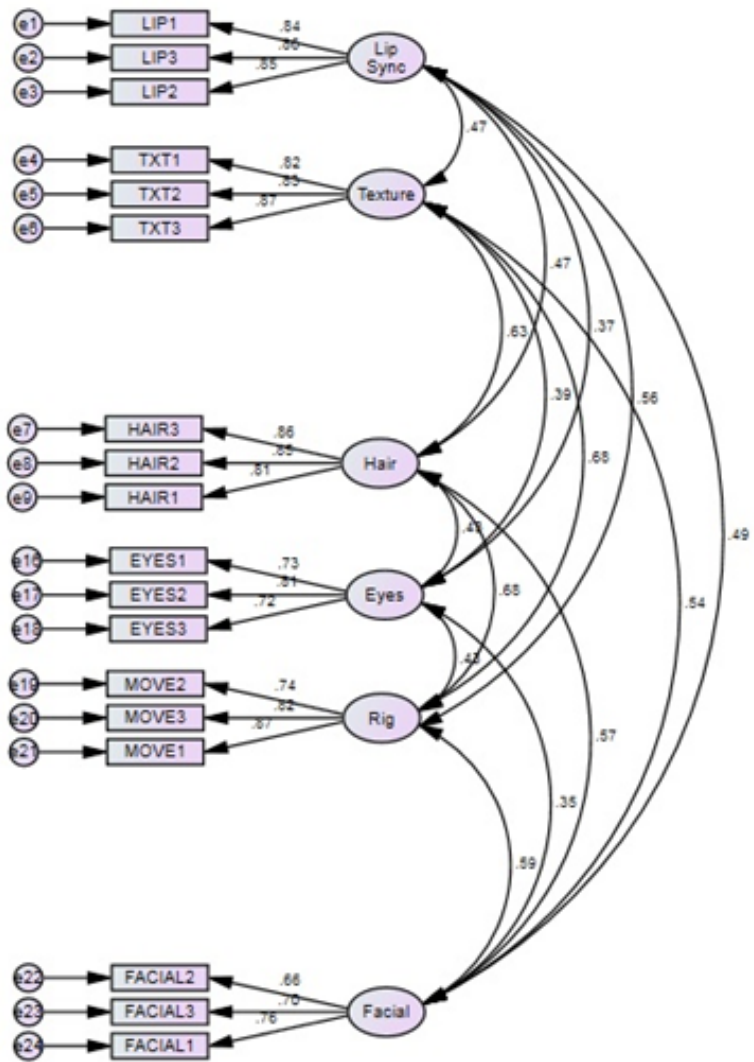


Figure 2 Appendix A

References

- 1 Aldred, J. (2011). From synthespian to avatar: re-framing the digital human in Final fantasy and The polar express. *Mediascape Winter*. Accessed January, 10, 2013.
- 2 Baumgartner, H., & Homburg, C. (1996). Applications of structural equation modeling in marketing and consumer research: A review. *International Journal of Research in Marketing*, 13(2), 139–161.
- 3 Evans, J. S. B., & Over, D. E. (2013). *Rationality and reasoning*. Psychology Press.
- 4 Freedman, Y. (2012). Is it real... or is it motion capture? The battle to redefine animation in the age of digital performance. *The Velvet Light Trap*, (69), 38–49.
- 5 Geller, T. (2008). Overcoming the uncanny valley. *IEEE Computer Graphics and Applications*, 28(4), 11–17.
- 6 Ho, C. C., & MacDorman, K. F. (2010). Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior*, 26(6), 1508–1518.
- 7 Ho, C. C., MacDorman, K. F., & Pramono, Z. D. (2008, March). Human emotion and the uncanny valley: a GLM, MDS, and Isomap analysis of robot video ratings. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction* (pp. 169–176). ACM.
- 8 Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, 6(1), 1–55.
- 9 Keysers, C., & Gazzola, V. (2007). Integrating simulation and theory of mind: from self to social cognition. *Trends in cognitive sciences*, 11(5), 194–196.
- 10 MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological methods*, 1(2), 130.
- 11 MacDorman, K. F., Green, R. D., Ho, C. C., & Koch, C. T. (2009). Too real for comfort? Uncanny responses to computer generated faces. *Computers in human behavior*, 25(3), 695–710.
- 12 Maio, G., & Haddock, G. (2009). *The psychology of attitudes and attitude change*. Sage.
- 13 Mewes, D., & Heloir, A. (2009). The Uncanny Valley.
- 14 Mori, M. (1970). The uncanny valley. *Energy*, 7(4), 33–35.
- 15 Alexander, O., Rogers, M., Lambeth, W., Chiang, M., & Debevec, P. (2009, November). Creating a photoreal digital actor: The digital emily project. In *Visual Media Production, 2009. CVMP'09. Conference for* (pp. 176–187). IEEE.
- 16 Van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, 9(4), 486–492.
- 17 The uncanny valley: Effect of realism on the impression of artificial human faces
- 18 Tinwell, A., Grimshaw, M., & Williams, A. (2011). The uncanny wall. *International Journal of Arts and Technology*, 4(3), 326–341.