SENSITIVITY TO PROPORTIONS IN FACES OF VARYING ANTHROPOMORPHISM

Robert Dale Green

Submitted to the faculty of the School of Informatics
in partial fulfillment of the requirements
for the degree
Master of Science in Human-Computer Interaction
Indiana University

December 2007

Accepted by the Faculty of Indiana University,
in partial fulfillment of the requirements for the degree of Master of Science
in Human-Computer Interaction.

Masters Thesis Committee

____
Karl F. MacDorman, Ph.D., Chair

____
Thomas Busey, Ph.D.

____
Alan C. Roberts, Ph.D.

Dedicated to my wife

I'll be coming to bed soon dear.

Contents

List of Tables

List of Figures

ACKNOWLEDGMENTS

I want to thank Dr. Karl MacDorman, whose guidance, support, comments, questions, encouragement, and goading (when necessary) have helped bring this document to completion. I would also like to thank my committee members, Dr. Thomas Busey and Dr. Alan Roberts, for their comments and support of this work.

Sandosh Vasudevan developed the website `www.theuncannyvalley.org` that hosted the on-line experiment, implemented the proportion controls, and provided the data extraction tool for the results. Chin-Chang Ho provided much guidance in the proper use of statistics and would not allow me to bend the numbers to my will. Their assistance was crucial and greatly appreciated. Heryati Madiapuri recruited participants in Indonesia. Many robotics researcher provided photographs of their work and encouragement in my research. A listing of those researchers and their robots is include in Appendix D.

Finally, I'd like to thank my family for their continual encouragement and support through a very long process.

ABSTRACT

Although it has been said that beauty is in the eye of the beholder, studies in social and medical sciences indicate that certain facial and bodily proportions are perceived to be more attractive across cultures. Additionally, studies indicate that the perception of attractiveness is more hardwired than learned, being present even in infants. Behavioral scientists have found that attractive people are often judged to have more positive character traits. Interface designers must make choices regarding how to represent the human form, whether in animation, virtual reality, or physical robots. An understanding of human preferences, in addition to other developments in the science of perception, can lead to design principles.

This study measured sensitivity to the best proportions, and tolerance for acceptable proportions in people, androids and more mechanical-looking robots, and three-dimensional and two-dimensional computer graphics characters. In an on-line experiment participants set the best point and acceptable range in four facial proportions for eleven characters, and completed a questionnaire rating character attributes such as human likeness. Participants showed greater sensitivity to the best proportions in faces they judged as more attractive and more humanlike. Participants also showed less tolerance for changes in proportion in more attractive faces.

CHAPTER ONE: INTRODUCTION

We live in a world with a growing population of synthetic characters. Animated characters first began appearing in films in the early 20th century. They have been interacting with human actors for many years. Dick Van Dyke danced with cartoon penguins in *Mary Poppins*. A generation later, Michael Jordan played basketball with Bugs Bunny in *Space Jam*. But these movies required a suspension of disbelief. With advances in computer graphics, the line between human and nonhuman actors is being increasingly blurred. Andy Serkis portrayed the character Gollum in Peter Jackson's *Lord of the Rings* trilogy, but movie-goers never really saw him. Computer graphics replaced him on screen with the Gollum movie audiences saw. Digital artistry replaced actor Bill Nighly as Davy Jones in *Pirates of the Caribbean: Dead Man's Chest*. Characters in computer and video games are becoming more realistic, and lifelike avatars lead us through virtual worlds.

Synthetic characters are not confined to the screen. Robots have come out of the laboratory and into our homes. Children's toys now include interactive robotic characters such as Furby and My Real Baby. Sony's Aibo is available for those wanting a no-mess no-fuss pet. Robots are used to grab attention at trade shows and media events (International Robotics, n.d.) and to provide educational lessons for children (Davis, 2000). In Japan and the West, robots are beginning to be marketed to help provide both companionship and eldercare. Robots are going to be with us almost from cradle to grave.

Like their animated counterparts, humanoid robots are becoming more lifelike.

The android Repliee Q1Expo is so realistic it elicits responses like those directed towards another human rather than a more mechanical-looking robots (MacDorman & Ishiguro, 2006). Some older visitors to the 2005 World Expo, in Aichi, Japan, standing next to Repliee Q1Expo, could not find the robot.

The rapidly improving aesthetic of synthetic characters brings many questions to mind. How is this array of synthetic characters perceived? Is Gollum man or beast? Is Repliee Q1Expo a woman or machine? How should synthetic characters be designed? Research indicates that people make judgements of attractiveness in others very quickly and do not change their original assessments. Researchers have proposed that human attractiveness may be based on familiarity, averageness, symmetry, or proportions of facial features. Should we judge synthetic characters by the same standards? Designers of synthetic characters must learn how important imitating the human face can be, and how low the tolerance for error in design is.

This study focuses on human sensitivity to facial proportions in human and synthetic characters such as computer graphics characters and robots. Specifically, it examines how selections of the "best" dimensions and acceptable ranges of facial dimensions of various characters relate to the degree to which they are judged to be human, lifelike, attractive, sexy or eerie.

CHAPTER TWO: LITERATURE REVIEW

It was long believed that human attractiveness was a product of personal taste and culture, and that there can be no universal standard of beauty (Etcoff, 1999). While private taste contributes to ratings of facial attractiveness, shared tastes account for about 50% of judges assessments (Hönekopp, 2006). Likewise, studies seeking evaluations of attractiveness across diverse cultures demonstrate both a cultural bias towards structural differences (e.g., nose size) and the universal appeal other features such as eye height and width, cheekbone prominence, and chin length (Cunningham, Roberts, Barbee, Druen, & Wu, 1995; D. Jones, 1995).

Judgements of attractiveness begin early in life. Infants gaze longer at photographs that adults also rate as attractive (Langlois et al., 1987), and children prefer to play with their more attractive peers (Salvia, Sheare, & Algozzine, 1975). First impressions are also important. Participants rated faces displayed for 13ms (the refresh rate of a computer monitor) on a 10 point scale. Their ratings agreed significantly ($t(9) = 4.90, p < .01$) with pretest assessments (Olson & Marshuetz, 2005). Not only do humans assess attractiveness quickly, but those judgements remain strong after a period of interaction and after details may fade from memory (Goldstein & Papageorge, 1980). This indicates that people remember whether someone was attractive, but not why.

Universal aspects of attractiveness indicate good genes, developmental and hormonal health, and a strong immune system. Males tend to find features indicating fertility attractive in potential mates. As female child-bearing years are limited, a mix of traits indicating sufficient maturity to give birth and sufficient youthfulness to bear many children, constitue the most attractive set of features. Male fertility is not limited in the same way, leading females to seek mates who will be good providers (Grammer & Thornhill, 1994; Scheib, Gangstead, & Thornhill, 1999; Drury, 2000; Alam & Dover, 2001). Females selected more "masculine" faces close to ovulation, when intercourse would most likely result in pregnancy (Johnson, Hagel, Franklin, Fink, & Grammer, 2001). Males selected different faces when asked which of 16 females faces they would prefer to take on a dinner date, prefer for sexual intercourse, and prefer for raising children (Cunningham, 1986). Facial features varied for each group.

Galton (1879) may have been the first to observe the attractiveness of the "averaged" face. In an attempt to isolate cues of criminality he created composites by overlaying photographs of convicts. The resulting composites were found to be more attractive than villainous. Langlois and Roggman (1990) created composites using 2, 4, 8, 16 and 32 faces. Composites made using more faces were rated as more attractive. The attractiveness ratings of the composite faces where higher than the average of the faces that contributed to the composites.

There are many possible explanations for the attractiveness of these composite photographs. The creation of composite photographs tends to smooth out blemishes and irregularities in the skin. B. C. Jones, Little, and Perrett (2004) found a strong positive correlation between the perceived health of facial skin and the attractiveness ratings of male faces. Maintenance of clear facial skin can be an indicator of health (Grammer & Thornhill, 1994). While some models may have distinctive facial features, such as Cindy Crawford's mole, they exhibit otherwise clear skin. Designers would be unlikely to give a character facial blemishes unless it were to communicate an undesirable or stereotypical trait about that character.

Another result of averaging faces is increased symmetry of the resultant face. Symmetry has long been an artistic ideal (Drury, 2000; Alam & Dover, 2001). Faces exhibiting symmetry were judged as more attractive (Rhodes, Proffitt, Grady, & Sumich, 1998). However, people's overt judgments of symmetry are not as good one might think. Scheib et al. (1999) showed participants images of the left or right half of a face only. Participants rated faces that possessed symmetry as more attractive than asymmetric faces, even though half the face was blocked and participants could not observe the symmetry. Maintaining facial symmetry through development indicates strong hormonal health (Grammer & Thornhill, 1994; Scheib et al., 1999), though developmental factors like sleep patterns may influence symmetry (Mealey, Bridgstock, & Townsend, 1999). Writer/director M. Night Shyamalan played up the asymmetry and brokenness of Samuel L. Jackson's comic book loving villain in *Unbreakable* (2000). Jackson's Elijah Prince explains to the hero (played by Bruce Willis) that all good villains are asymmetric and have physical flaws, as Prince himself is asymmetric and lame.

Finally, averaged faces may be deemed more attractive because they remind us of ourselves. Halberstadt (2006) demonstrated a preference for facial prototypes over individual faces. Faces that were considered average were judged as more attractive than more distinctive caricatures of those same faces (Rhodes & Tremewan, 1996; Rhodes, Hickford, & Jeffery, 2000). Functional magnetic resonance imaging (fMRI) studies have demonstrated that same-race faces activate a different portion of the brain than those of other races (Golby, Gabrieli, Chiao, & Eberhardt, 2001). Familiarity can play a role in creating virtual worlds. In an immersive virtual environment (IVE), participants encountering their *virtual selves* (interactive agents with the participants own face) observed less interpersonal distance than participants encountering virtual others (Bailenson, Beall, Blascovich, Raimundo, & Weisbuch, 2001)

There is strong support for the attractiveness of facial symmetry, healthy looking skin and faces that look like us. Averaged faces may be more attractive on average, but they are not the most attractive faces. Indeed, some individuals faces are rated as more attractive than composites Langlois and Roggman (1990). Perrett, May, and Yoshikawa (1994) demonstrated that creating composites of faces deemed more attractive produced faces that were judged as more attractive than an average of the entire group.

Symmetry is one factor of facial attractiveness.. Cunningham (1986) measured 25 facial features of 50 women (27 beauty pageant contestants and 23 college students). From these measurements he compared ratings of attractiveness against 21 facial proportions, finding significant correlations in 12 of the 21 proportions. These proportions had stimulated positive attractiveness ratings in previous studies. A regression analysis indicated eye height, nose area, cheek width and smile width contributed to more than 50% of attractiveness ratings. A study using a larger number of facial landmarks (135 for male and 130 for female stimuli) and proportions (156 for male and 155 for female stimuli) found significant correlations between attractiveness and about 20% of male facial proportions, but only about 10% of female facial proportions (Farkas, 1994).

Using a more limited set of facial proportions, Grammer and Thornhill (1994) found different facial proportions contributed to the perception of traits such as attractive, dominant, sexy and healthy. Prominent eyes and cheekbones contributed most to males' evaluations of females, while jaw width and lower-face proportions contributed most to females' evaluations of males. in another study participants viewed video of changing facial proportions and indicated when the profile was "acceptable" by pressing and releasing a key on a computer. Changes as small as 1mm altered the participant's perceptions (Giddon, Sconzo, Kinchen, & Evans, 1996).

Comparing artwork, from 23 centuries and many cultures, to photographic portraits, Costa and Corassa (2006) found that artists tend to exaggerate the size and roundness of the eyes and lips and make faces longer with more prominent chins. These exaggerations were identified as supernormal stimuli, designed to draw attention and perhaps to simplify neural processing (Latto, 1995). While the extent of exaggeration varies across historical eras, this pattern holds from ancient Egypt to today. A follow-up study demonstrated art students exaggerated these same features in self portraits, even when looking in the mirror.

If our perceptions of human others is based, at least partly, on genetic survival, how do we perceive synthetic characters? Goetz, Kiesler, and Powers (2003) found participants believed humanlike robots were best suited for interactive tasks, while mechanical looking robots were best suited for routine jobs. How do we determine whether a robot is humanlike?. A study of 48 commercial, research and fictional robots indicates that to be considered humanlike a robot should have a distinctively human head shape and a facial area dominated by human features. Most significantly, a nose, eyelids and mouth suggested humanness (DiSalvo, Gemperle, Forlizzi, & Kiesler, 2002).

DiSalvo et al. (2002) recommend exaggerated features and an encasement to hide the mechanics in the head so the robot not only seems humanlike, but also product-like. This product focus is to keep the robot from falling into the uncanny valley. Masahiro Mori proposed that as robots somewhat become more humanlike they appear more familiar up to a point. But just like a human corpse, they risk becoming eerie when they are nearly human, especially when they are discovered to be mechanical through touch or by other means (Mori, 1970). Avoidance of the uncanny valley became a rubric of roboticists as robots were designed to be humanoid, but not humanlike. MacDorman and Ishiguro (2006) plotted the uncanny valley by having participants rate images that morphed between a humanoid robot and an android and then on to the human model for the android. In a similar experiment (Hanson, 2006) demonstrated the uncanny valley could be avoided, not by shunning human likeness, but by careful design. But the uncanny valley is not just an artifact of viewing still images, MacDorman (2006) also found participants respond to the eerieness of certain robotic motion.

Methods used in Previous Studies

A variety of methods have been used to study facial attractiveness. Most studies have involved viewing a series of photographs, slides or images on a computer monitor, and rating each image on a scale, such as a seven-point Likert scale (e.g., Penton-Voak et al., 2001). Attributes such as dominance (Johnson et al., 2001), distinctiveness (Rhodes & Tremewan, 1996), and trustworthiness (Cunningham, 1986) have been assessed along with attractiveness. Some studies asked participants to perform a simple ranking of the images (Udry, 1965).

Studies on composite faces (Langlois & Roggman, 1990; Perrett et al., 1994; Rhodes & Tremewan, 1996) and symmetry (Grammer & Thornhill, 1994; Kowner, 1996; Rhodes et al., 1998) created new images by digitally or mechanically merging two or more images. These studies typically compared the attractiveness of original photographs with the composites created from them. Keating (1985) created stimuli using a Identi-Kit, like those used by police departments. Direct facial measurements (anthropometry) (Farkas, 1994) and measurements of facial features in photographs (indirect anthropometry or photogrammetry, Cunningham, 1986; Cunningham et al., 1995; Grammer & Thornhill, 1994) have also been used in an attempt to quantify facial attractiveness.

Participants were not always limited to rating static images. Evans, Viana, Anderson, and Giddon (2005) cut facial features out from original photographs, repositioned and copied the resultant images to create new stimuli. Participants were asked to indicate whether the resulting stimuli looked acceptable. Giddon (1995, 1996) created videos changing the chin and bite in profile using animation techniques and morphing software. Participants watching videos of the changing facial profiles indicated the range in which the profile looked acceptable. Johnson et al. (2001) created QuickTime videos and instructed participants to use a slider or single step to find frames that satisfy criteria including "most attractive." In another study, participants using keyboard controls increased or decreased the width of photographic portraits with a fixed height in order to select the most attractive configuration (Costa & Corassa, 2006).

Hypotheses

We have seen facial proportion affect how humans perceive one another. There is little understanding of how much leeway designers of synthetic characters have in representing the faces of their creations. Preliterate man had little conception of the type of "other" synthetic characters represent. As literature developed, writers began to imagine other beings, but they became alive only in the mind's eye. Can anyone really say what Jonathan Swift's Houyhnhnms looked like? It has only been in the last century of film and then television that we have been given a shared vision of what an alien or robot might look and act like. And it has only been in the last couple of decades that we have had the opportunity to interact with other creatures through a computer or video game, or directly with a humanoid robot.

We have seen that we humans are sensitive to differing facial proportions in other humans (Cunningham, 1986; Cunningham et al., 1995; Grammer & Thornhill, 1994), and that we have a preference for those who look like us (D. Jones, 1995; Golby et al., 2001). Will we extend this preference to synthetic characters that look human? To examine the relation between human likeness and facial proportion, this study proposes the following hypotheses:

H1: Figures that are subjectively rated as more humanlike will have a narrower range of acceptable facial proportions compared to those that are subjectively rated as less humanlike.

H2: There will be greater intersubjective agreement on what facial proportions are best in more humanlike figures compared to less humanlike figures.

Attractiveness has many costs and benefits. Attractive people are perceived as more intelligent, sociable, healthy, and trustworthy. While attractive people are presumed to be more sexually experienced, it is also believed they are less likely to have affairs. People are more likely to perform acts of altruism, such as helping to move home or donating blood, for an attractive person than an unattractive person. (Cunningham, 1986). Nevertheless, there are specific facial proportions and attributes that are deemed more attractive, which leads to the following hypotheses.

H3: Figures that are subjectively rated as more attractive will have a narrower range of acceptable facial proportions compared to those that are subjectively rated as less attractive.

Studies have indicated that there is agreement not only in what faces are attractive, but also in what faces are unattractive. If it is the possession of certain facial proportions that makes a face attractive, best proportions must be best regardless of how attractive or unattractive the face being evaluated is. Essentially, best is best. The final hypothesis for this study is

H4: Intersubjective agreement on which facial proportions are best in more attractive figures will not be significantly different compared to less attractive figures.

CHAPTER THREE: METHODOLOGY

Participants

Participants in this study were recruited through e-mail and postings to newgroups visited by individuals who might be interested in the results of the experiment. This research may lead to design principles that enhance computer graphics characters. Internet users, like those recruited for this study, are likely to benefit from these enhanced computer graphics characters.

The e-mail and postings provided potential participants with a brief introduction to the study, an estimated time for completion, and the URL for the website hosting the study. The fact participation was voluntary was emphasized in all materials. Additional information on e-mail recruitment tool can be found in Appendix A.

208 participants completed the study. Participants were not required to provide demographic information. Of those who did 61% (n = 126) were male and 39% were female (n = 81). The mean age of participants was 31.9 (SD = 10.27) ranging from 17 to 79. The largest group of participants was those born in the United States (31.3%, n = 65), followed by Indonesia (30.3%, n = 63) and the United Kingdom (12.0%, n = 25). The majority of participants lived in the United States (39.9%, n = 83) followed again by Indonesia (30.3%, n = 63) and the United Kingdom (12.0%, n = 25). Twenty six countries were recorded as country of birth, and 20 as country of residence. 18.6% (n = 29) of participants did not live in their country of birth. Participants averaged 16.1 years of education (representing the completion of a bachelor's degree in the United States) with a standard deviation of 4.61. A complete

discussion of participant demographics is included in Appendix B.

Stimuli

Eleven stimuli were prepared for the study. Stimuli included two photographs of humans (one male and one female), three three-dimensional computer graphics characters (one male human, one female human, and one robot), three computer drawings (one male human, one female human and one robot), and three photographs of robots. Four movies were created using FantaMorph 3 (Abrosoft, 2006) for each stimulus. Each movie warps one facial proportion between the extremes of 10% of the measured dimension. The proportions altered are; cheek width, eye separation, face height, and jaw width.. Cheek width, eye separation, and face height had the strongest correlations to attractiveness for female faces in (Cunningham, 1986; Grammer & Thornhill, 1994), while face height and jaw width had the strongest correlations to attractiveness for male faces in (Grammer & Thornhill, 1994).

Still Images

Photoshop 7.0 was used to convert original photographs and artwork to 400x500 pixel images framing the face on a 50% gray background. Figure 1 depicts the converted images.

A detailed description of how each of the 11 stimuli were acquired and created can be found in Appendix C. A number of researchers provided photographs of their robots and permission to use them in this study. A complete listing of those researchers and their robots is included in Appendix D.

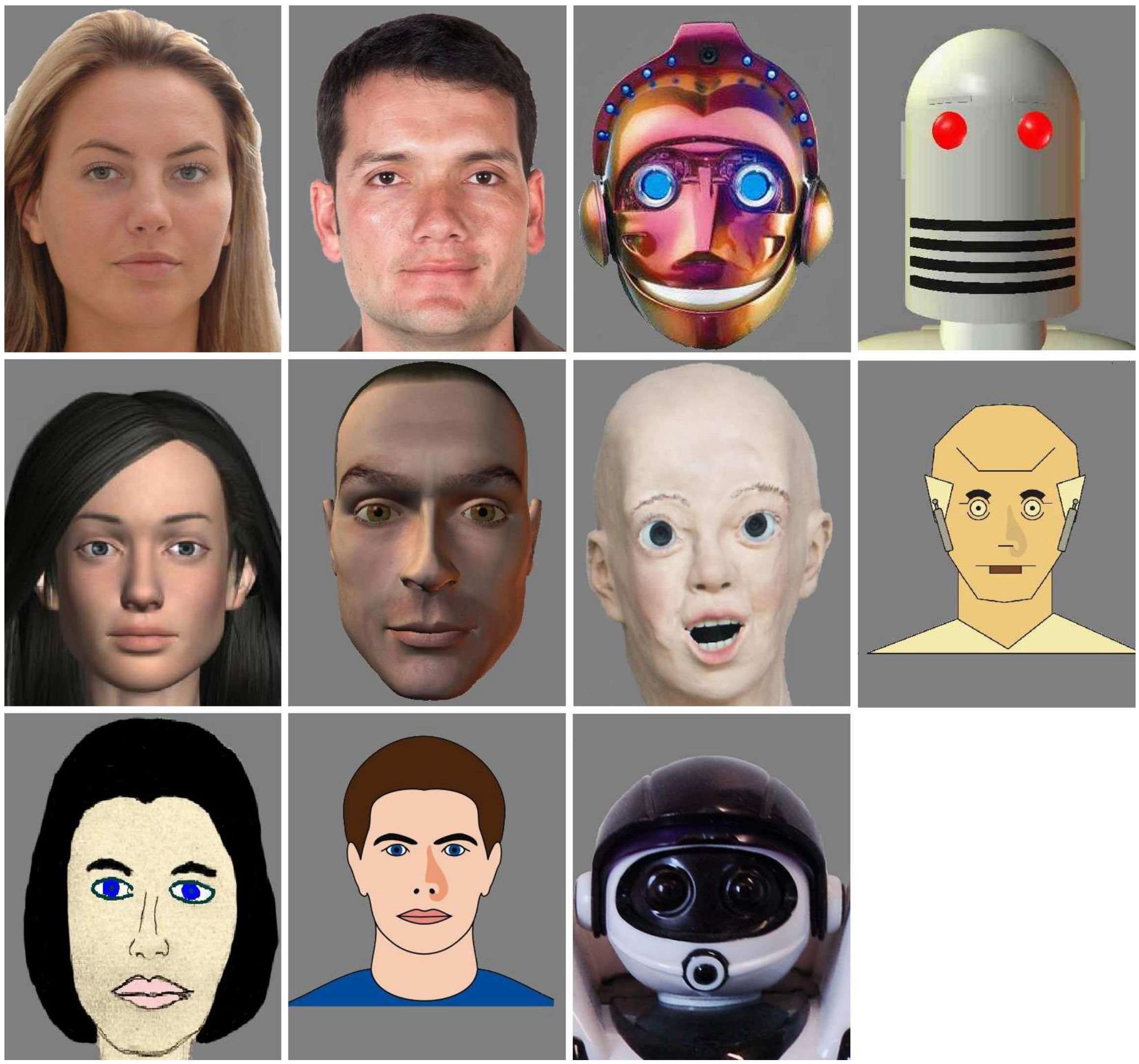

Figure 1: Original stimuli normalized to 400 x 500 pixels. Top row:Female and male photographs, Anhtroboot, 3D robot. Middle row: 3D female, male, Barthoc, Jr., and 2D robot. Bottom row: 2D female, male, and Robosapien.

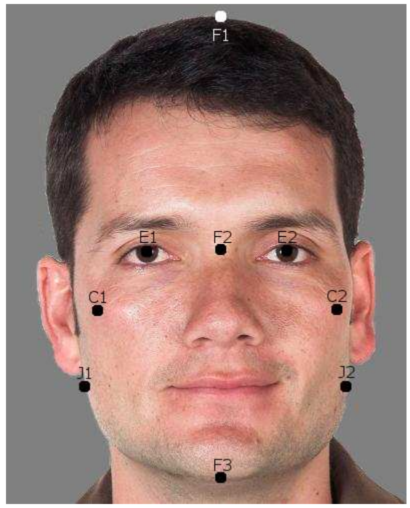

Figure 2: Points used in measuring facial dimensions

Movies

Four movies, one for each proportion, were created using FantaMorph 3 (Abrosoft, 2006) for each stimulus in Figure 1. Each movie warps one proportion between the extremes of ±10% calculated value. The altered proportions were cheek width, eye separation, face height, and jaw width. These dimensions are most frequently associated with attractiveness judgments (e.g., Cunningham, 1986, Grammer & Thornhill, 1994).

Five facial dimensions were used to calculate the proportions Image (Scion, 2006) was used to measure each dimension. The five dimensions are demonstrated in Figure 2 and defined as follows:

E2 - E1: Distance between the center of the pupils

C2 - C1: Distance between the outer edge of the cheekbones at the most prominent point

J2 - J1: Width of the face at the level of the middle of the smile

F3 - F2: Distance between the mid-point of the pupils and the bottom of the chin

F3 - F1: Distance between the top of the head and the bottom of the chin

Four facial proportions were calculated using these dimensions.

Cheek Width: The width of the face at the cheek bones divided by the overall height of the head (C2 - C1) / (F3 - F1)

Eye Separation: The distance between the pupils divided by the width of the face at the cheek bones (E2 - E1) / (C2 - C1)

Face Height: The height of the face between the eyes divided by the overall height of the head (F3 - F2) / (F3 - F1)

Jaw Width: The width of the face at the mouth divided by the distance between the pupils (J2 - J1) / (E2 - E1)

To change the proportions the dimension that served as numerator was increased and decreased to cause a 10% change in proportion. Table 1 shows the ratio and dimension that will be altered for each of the four proportions for all 11 stimuli. All measurements are in pixels.

A FantaMorph (Abrosoft, 2006) movie with no variation was created for each stimulus. This movie was referred to as the base movie. Figure 20 is a screen shot of the base movie for the Robosapien robot. The base movie contains 300 to 400 points outlining the images face and facial features. The base movie consisted of two

Table 1: Measurement of Facial Dimensions

Stimulus	Cheek Width[1]	C2 - C1[5]	Eye Separation[2]	E2 - E1[5]	Face Height[3]	F3 - F2[5]	Jaw Width[4]	J2 - J1[5]
Female photo	0.62	253	0.48	122	0.66	269	1.69	122
Male photo	0.45	206	0.67	138	0.58	265	1.90	262
3-D female	0.51	221	0.56	123	0.49	214	1.87	230
3-D male	0.63	289	0.52	151	0.75	342	1.71	258
2-D female	0.33	161	0.71	115	0.59	252	1.70	196
2-D male	0.74	175	0.47	82	0.73	174	1.76	144
Anthroboot	0.62	246	0.57	139	0.70	248	1.57	218
Barthoc Jr.	0.47	230	0.60	139	0.51	237	1.91	266
Robosapien	0.73	249	0.46	114	0.62	212	1.96	233
3-D robot	0.59	251	0.49	124	0.78	333	2.02	251
2-D robot	0.51	173	0.45	78	0.57	198	1.79	140

[1] Width of the face at the cheek bones / Overall height of head
[2] Distance between pupils / Width of face at the cheek bones
[3] Height of the face between the eyes / Overall height of head
[4] Width of the face at the mouth / Distance between pupils
[5] Pixels

sequences of 11 frames. Each sequence represented a increase or decrease of 10% in the original proportion. The duplicated center frame (last of the first sequence and first of the second sequence) is removed when the movie is exported to an external format.

To create the movie for each dimension the base movie was copied and renamed to indicate the source stimulus and the varied dimension. Additional points were set as needed to assist varying the dimension under consideration. The points of the first image of the first sequence and the second image of the second sequence are altered to achieve the warping effect. The point under consideration was moved in or out by 10% of the measurement listed in Table 1 and neighboring points were moved to provide the smoothest transitions possible. Figure 3 depicts the extremes in placement of points for the eyes of the 3-D female character, and the resulting extreme frames of the movie. For the lower facial length dimension the point identified as F2 in Figure

2 was moved up or down, but the chin (F3) remained fixed, so as not to alter the overall facial length.

Figure 3: Extremes in placements of the eyes of the 3-D female character

There are three exceptions to the above descriptions. In order to avoid artifacts created by warping certain line drawings additional drawings were created for the extremes of the 2-D male and robot jaws and the 2-D robot cheeks. Movies were created morphing between these images, seen in Figure 4.

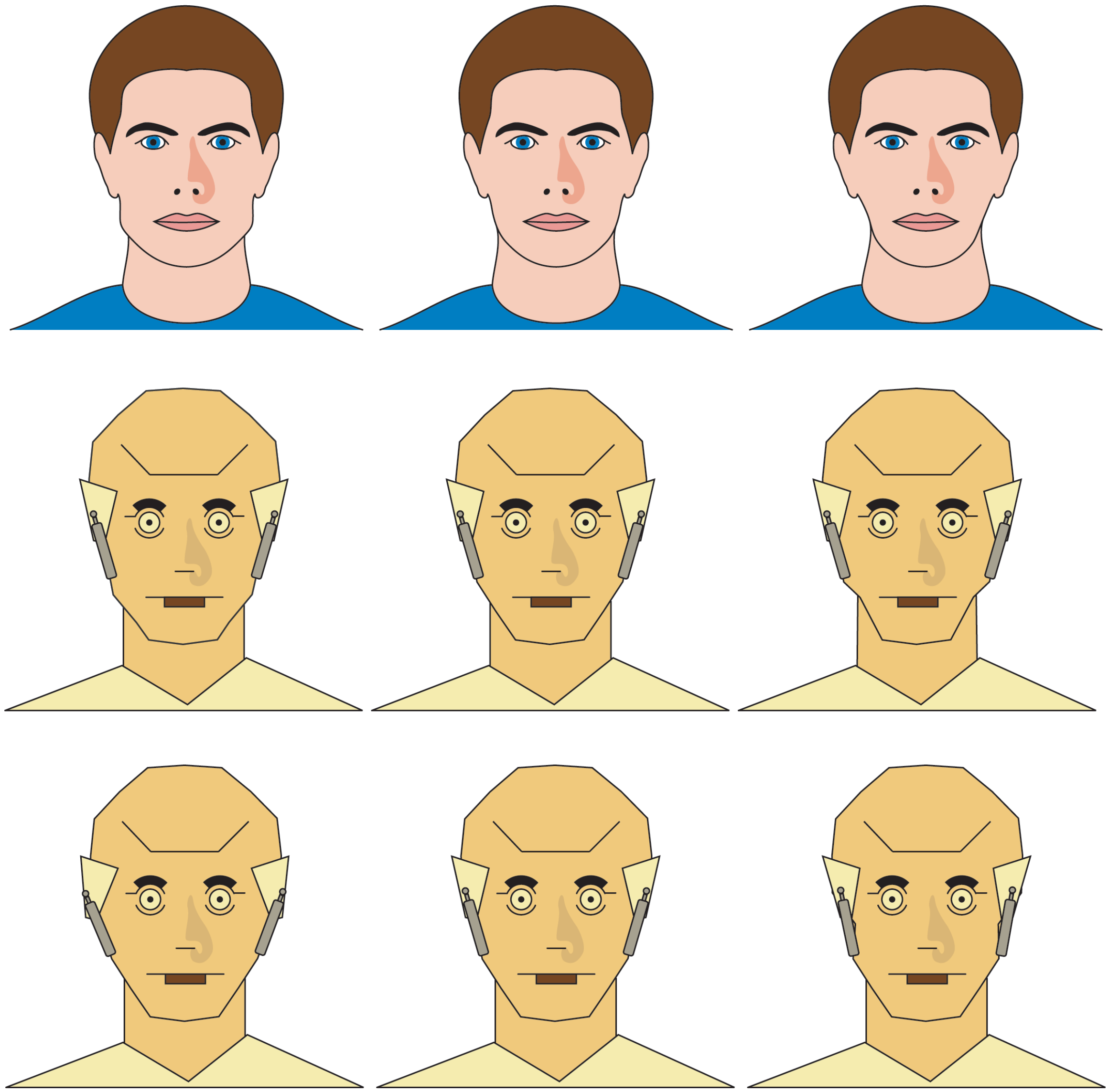

Figure 4: 2-D characters, morphed left to right. Top row is the male character with varied jaw width. Middle row is the robot with varied jaw width. Bottom row is the robot with varied cheek width.

Procedures

The website Exploring the Uncanny Valley (`http://www.theuncannyvalley.org`) was created to host this and other studies related to the uncanny valley. The initial page instructed participants to chose a language. Available languages were English, Japanese, Traditional Chinese, Simplified Chinese, and Bahasa Indonesia. Selecting a language lead the participant through the rest of the website in their chosen language. The homepage invited visitors to register while providing a few pieces of demographic data, or to log in if they are already registered. Participants had the option to skip entering their e-mail address, demographic data or both.

Upon entering the site a listing and description of active studies was displayed. Recruitment materials directed participants to click the participate link in the Perception of Facial Proportions study. There were four tasks to the study. Participants were required to complete the tasks in sequence, though registered participants could leave the website and return to tasks they had not completed.

Task 1

The first page of task 1 was an online consent form. Text of the informed consent form is contained in Appendix E.

Movie stimuli were presented one at a time in random order. A movie frame was selected at random as a starting point. This frame was displayed with positioning buttons that allowed participants to adjust the figure they were viewing. Keyboard arrow keys could also be used to adjust the image. Participants were instructed, to use the arrows to adjust the figure until it looks best. Then click Best (see Figure 5). After clicking Best, the next stimulus was presented and the process repeated for all movie stimuli.

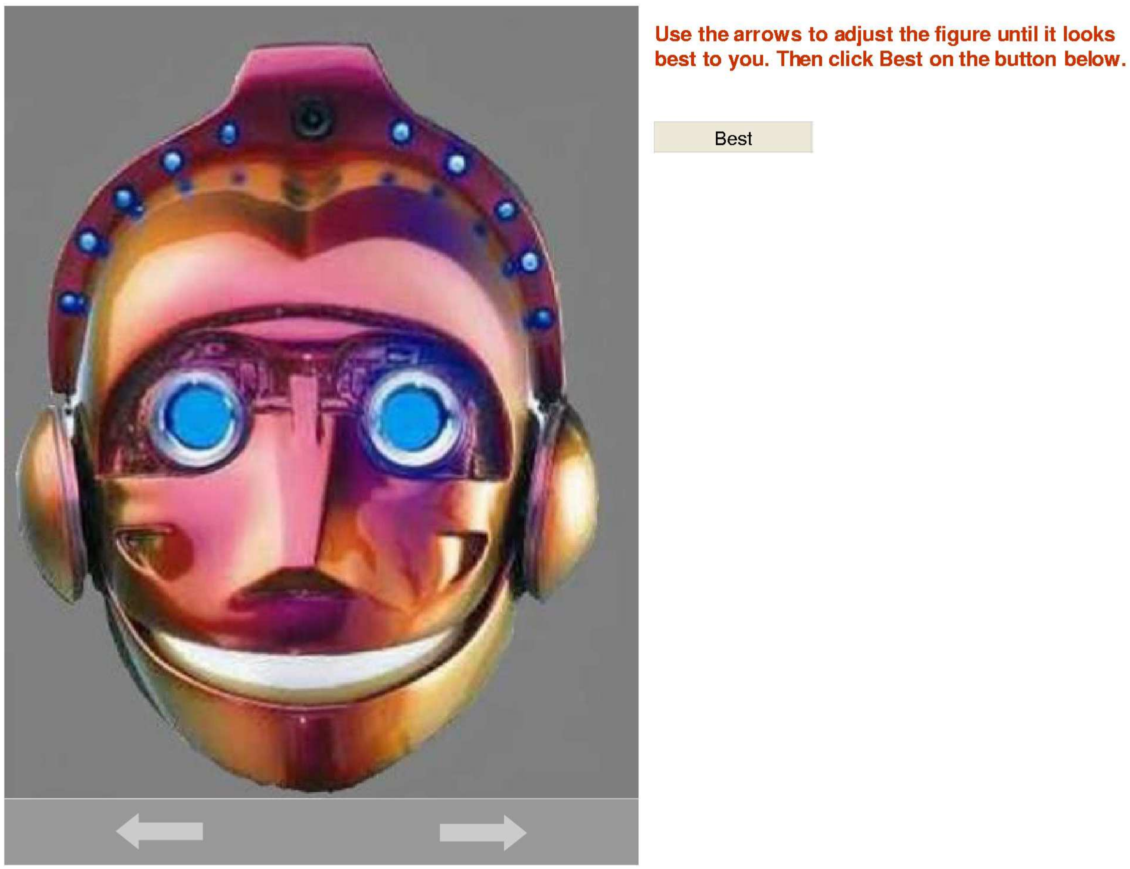

Figure 5: Task 1: Prompting to select the "best" position

Tasks 2 and 3

Tasks 2 and 3 mirrored one another. Movie stimuli were presented one at a time in random order. With a screen similar to Figure 5 the movie frame selected as best in Task 1 is the single frame presented. One direction was locked at random for the task (frames could only be selected to the left or right of the starting point).

Participant was instructed, to use the arrows to find the last point where the figure looks acceptable.

After the participant had selected an acceptable point for the stimulus the next stimulus was presented and the process repeated for all movie stimuli. After all stimuli were rated the participant was returned to the Current Research page.

For Task 3, the question was reversed after the participant selected an acceptable point.

Task 4

Still image stimuli were presented one at a time in random order. Participants were asked a series of questions, to be answered on a seven-point Likert scale. The order of the questions was randomized for each participant, but was presented in the same order for all stimuli in the session. The survey asked the following questions:

- This figure looks female
- This figure looks creepy
- This figure looks sexy
- This figure looks ugly
- This figure looks alive
- This figure looks humanlike

Responses on the scale for all questions were: strongly disagree, moderately disagree, slightly disagree, beutral, slightly agree, moderatly agree, and strongly agree.

Analysis

Method of Data Analysis

Data was collected in a MySQL database and analyzed using a combination of database queries, the statistical functions of Microsoft Excel, and SPSS. Participants choices of best and last acceptable points were converted from frames to the proportions described in the Treatments section. Each participant's distance from the mean best point will be calculated, as will an overall acceptable range. Responses to the survey in Task 4 were converted to a numeric range of -3 for strongly disagree to 3 for strongly agree.

All statistical significance levels reported in this study are two-tailed. The .05 level is the critical level for statistical significance.

CHAPTER FOUR: RESULTS

This chapter looks first at the raw results of participant's assessments of the best point and acceptable range, and ratings for the various attributes. Next, the relations between attributes and sensitivity to the best point, and attributes and tolerance for acceptable range are examined. Figures will be grouped by type (Human or Robot), degree of realism (2D, 3D, or Photograph), and according to observations of multidimensional visualizations of the attribute ratings. These groups are compared and the significant differences reviewed. The relations between attributes and participant-selected best points, as opposed to sensitivity, are evaluated. Finally, the effect of the participant's gender on their responses is compared.

Sensitivity and Tolerance to Facial Proportions

The first task required participants to select the best position on 44 adjustable images (11 characters by 4 facial proportions). Each adjustable image was a 21 frame Flash movie with the 11th frame representing the original image. A change of one frame represents a 1% change in the proportion. The first task required participants to select the frame (viewing them sequentially) that represents the best proportion. Tasks 2 and 3 required participants to indicate the last point at which the figure looked acceptable as the proportion either increased or decreased. The acceptable range of facial proportions was computed as the difference between these two points.

Participant selections were recorded in terms of frames. The data was converted from frames to proportions for comparative purposes. Change from left to right varied between the Flash movies (i.e., some proportions increased left to right and others decreased). However, this difference did not affect range measurements. The standard deviation of the best point and the mean of the range for each figure are detailed in Table 2.

Sensitivity was defined as the standard deviation of the best point. Tolerance of acceptable proportions was defined as the mean of the acceptable range. For Cheek Width (width of the face at the cheek bones divided by the overall height of the head), participants showed the greatest sensitivity to the 2D Female ($SD = 0.0157$) and the least tolerance to the 2D Female ($M = 0.0255$) as well. The 2D Male had the greatest sensitivity ($SD = 0.0212$) in Eye Separation (distance between the pupils divided by the width of the face at the cheek bones), while the Female Photograph had the least tolerance ($M = 0.0290$) in the same proportion. For Face Height (height of the lower face, measured between the pupils to the bottom or the chin, divided by the overall height of the head) participants showed the greatest sensitivity to the Female Photograph ($SD = 0.0190$) and the least tolerance to the Male Photograph ($M = 0.0266$). The 3D Female had the greatest sensitivity ($SD = 0.0545$) and least tolerance ($M = 0.0465$) for Jaw Width (width of the face at the mouth divided by the distance between the pupils).

Table 2: Sensitivity and Tolerance by Figure and Facial Proportion

	Sensitivity[1]				Tolerance[2]			
Figure	Cheek[3]	Eyes[4]	Face[4]	Jaw[4]	Cheek[3]	Eyes[4]	Face[4]	Jaw[4]
Female Photo	0.0171	0.0216	0.0190	0.0587	0.0282	0.0290	0.0283	0.1153
Male Photo	0.0281	0.0299	0.0217	0.0703	0.0448	0.0365	0.0266	0.1319
3D Female	0.0226	0.0259	0.0370	0.0545	0.0480	0.0355	0.0267	0.0948
3D Male	0.0228	0.0249	0.0276	0.0570	0.0304	0.0297	0.0406	0.1000
2D Female	0.0157	0.0367	0.0263	0.0646	0.0255	0.0530	0.0376	0.1239
2D Male	0.0268	0.0212	0.0303	0.0789	0.0447	0.0343	0.0480	0.1416
Anthrobot	0.0326	0.0320	0.0384	0.0697	0.0438	0.0468	0.0527	0.1144
Barthoc, Jr.	0.0273	0.0283	0.0280	0.1020	0.0391	0.0353	0.0323	0.1467
Robosapien	0.0325	0.0264	0.0381	0.0920	0.0577	0.0374	0.0567	0.1679
3D Robot	0.0226	0.0283	0.0369	0.1066	0.0263	0.0454	0.0590	0.1229
2D Robot	0.0309	0.0257	0.0267	0.1114	0.0480	0.0391	0.0414	0.1525

[1] Sensitivity measure as standard deviation from the best proportion.
[2] Tolerance measured as the acceptable range of a proportion.
[3] $n = 194$
[4] $n = 208$

Figure Attributes

Task 4 required participants to rate each figure on six adjectives, or attributes, on a seven-point Likert scale from strongly disagree (-3) to strongly agree (3). The two human figures were rated the most alive (male $M = 2.53$ and female $M = 2.51$) followed by the 3D human figures, the three robots, the 2D human figures, and finally the 3D and 2D robots were rated the least alive ($M = -2.13$). The female figure was rated most sexy ($M = 1.57$), female ($M = 2.79$), and humanlike ($M = 2.87$) as well as least creepy ($M = -2.39$) and ugly ($M = -2.38$). By contrast, the robot Barthoc, Jr. was rated least sexy ($M = -2.85$) and most creepy ($M = 2.43$) and most ugly ($M = 2.30$). The results of Task 4 are summarized in Table 3.

Table 3: Participant Ratings of Figure Attributes

Figure	Humanlike[1]	Alive[2]	Female[2]	Sexy[2]	Creepy[2]	Ugly[2]
Female Photo	2.87	2.51	2.79	1.57	-2.39	-2.38
Male Photo	2.80	2.53	-2.90	-0.42	-1.78	-1.26
3D Female	1.81	0.32	2.63	1.19	-1.91	-2.21
3D Male	1.62	0.15	-2.73	-0.65	-0.90	-1.03
2D Female	0.96	-1.90	1.90	-2.10	-0.50	0.15
2D Male	0.57	-1.80	-2.56	-1.87	-1.32	-0.81
Anthrobot	-2.14	-1.47	-1.63	-2.08	-0.97	-0.95
Barthoc, Jr.	0.62	-1.10	-0.97	-2.85	2.43	2.30
Robosapien	-2.35	-1.40	-1.94	-2.07	-1.82	-1.49
3D Robot	-2.68	-2.09	-2.13	-2.37	-1.25	-0.91
2D Robot	-1.46	-2.13	-2.62	-2.31	-0.90	-0.47

[1] $n = 142$
[2] $n = 208$

Comparison of Attributes and Proportions

To compare sensitivity, each participant's selection of the best point was converted to the difference from the mean best proportion ($X - \bar{X}$). A one-way ANOVA by figure was performed on each of the six attributes and difference and range for each of the four facial proportions. All dependent variables varied significantly ($p \leq .001$). The effect size (ω) was large for all variables except the ranges, which had ω values from .305 (face) to .196 (jaw).

A two-tailed Pearson's correlation was performed using the difference in each facial proportion (sensitivity), the mean range (tolerance), and each attribute. Table 4 presents the correlations between participant's assessments of the six attributes and their sensitivity to best proportions and tolerance for acceptable proportions. The test found the strongest correlations between the attribute humanlike and sensitivity to face height ($r = -.639$) and jaw width ($r = -.538$), both $p < .01$. All correlations between attributes and sensitivity were significant at a level of $p < .01$ except alive–eye separation ($p < .05$) and those between face height and the attributes creepy and ugly, both of which failed to reach a significance level of $p < .05$. Most correlations between attributes and tolerance were significant at either $p < .01$ or $p < .05$. These correlations were generally weaker than those related to sensitivity. The strongest correlation was between alive and face height ($r = -.231,\ p < .01$).

Table 4: Correlation between Selected Attributes and Facial Proportions

	Sensivity[1]				Tolerance[2]			
Attribute	Cheek	Eyes	Face	Jaw	Cheek	Eyes	Face	Jaw
Humanlike[3]	-.448**	.110**	-.639**	-.538**	.015	-.003	-.172**	-.012
Alive[4]	-.203**	-.044*	-.393**	-.391**	-.087**	-.122**	-.231**	-.140**
Female[4]	-.418**	.205**	-.505**	-.425**	-.048*	.053*	-.110**	-.069**
Sexy[4]	-.146**	-.161**	-.428**	-.500**	-.089**	-.133**	-.195**	-.156**
Creepy[4]	-.125**	.184**	-.014	.246**	.090**	.091**	.070**	.145**
Ugly[4]	-.136**	.229**	.018	.290**	.101**	.101**	.080**	.172**

* p < .05
** p < .01

[1] Sensitivity is participant's distance from the mean.
[2] Tolerance is participant's acceptable range.
[3] $n = 1408$ for cheek width and 1562 for other proportions
[4] $n = 2134$ for cheek width and 2288 for other proportions

Participants showed increased sensitivity towards the best point in all proportions as ratings for the attribute humanlike increased, except eye separation, for which they showed slightly decreased sensitivity (all ps $< .01$) . Sensitivity increased in all four proportions as ratings for alive, female, and sexy increased (all ps $< .01$). Correlations between sensitivity and the attributes creepy and ugly were mixed with participants showing decreased sensitivity to the best point for eye separation and jaw width, but greater sensitivity to cheek width as ratings increased (all ps $< .01$).

The relations between tolerance and the attribute humanlike were mixed, with tolerance decreasing in all proportions except cheek width as ratings increased. The relation between face height and humanlike was the only one to reach significance ($p < .01$). Participant's indicated a narrow acceptable range, less tolerance, in all four proportions as ratings for attributes alive, female, and sexy (all ps $< .01$). Tolerance decreased, wider acceptable range, in all proportions as ratings of attributes creepy and ugly increased (all ps $< .01$.

Figures 6 and 7 demonstrate the strongest relations relating to human likeness. These graphs combine the statistics for each figure as reported in Tables 2 and 3.

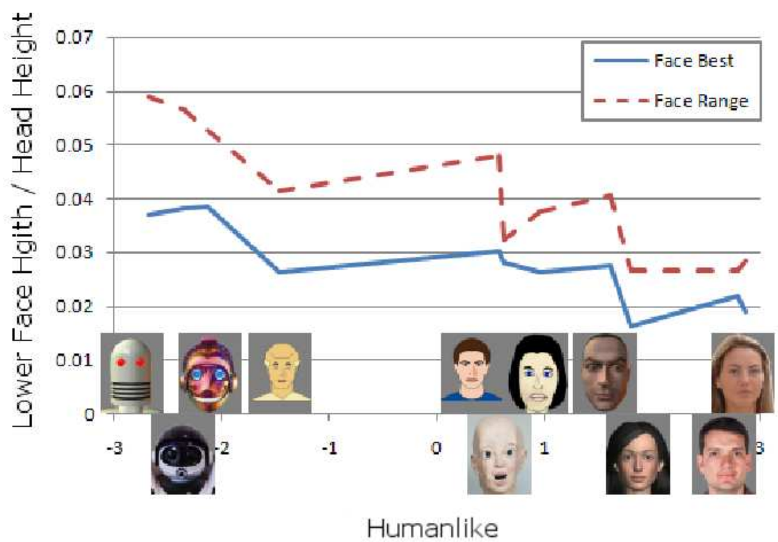

Figure 6: Sensitivity and tolerance in face height sorted by human likeness

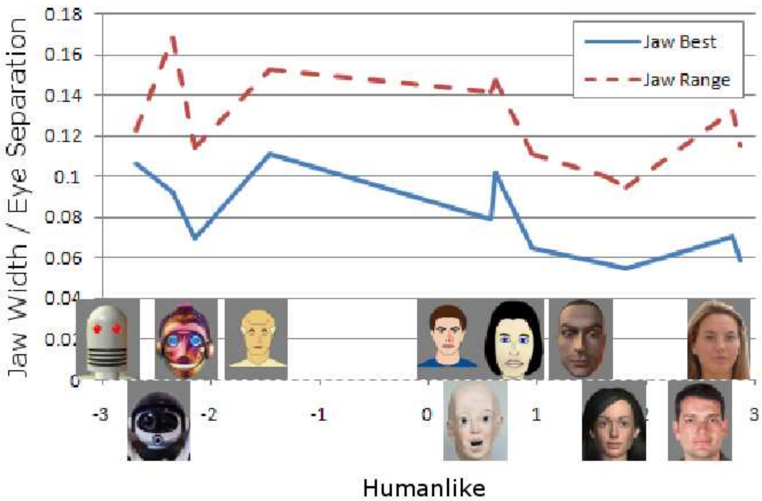

Figure 7: Sensitivity and tolerance in jaw width sorted by human likeness

Comparison of Human and Robot Figures

Reviewing Table 2 indicates participants were generally more sensitive, and less tolerant of the facial proportions of human character than robots. The dataset was divided between those figures known to be human (2D, 3D, and photos of female and male figures) and robot (2D and 3D robots, Anthrobot, Barthoc, Jr., and Robosapien). A one-way ANOVA was performed on the six attributes and range and difference of the four facial proportions. Significant differences ($p < .001$) were detected in all six attributes. The largest effect was observed for humanlike ($F(1,1561) = 1765.621, \omega = .728$), followed by sexy ($F(1,2286) = 783.407$, $\omega = .505$) and alive ($F(1,2286) = 504.471$, $\omega = .425$). All facial proportions had significant differences at $p < .001$ in difference from the best point. Acceptable range had significant differences at $p < .001$, except cheek width ($p < .01$). Difference in jaw width had the largest effect size ($F(1, 2286) = 3223.604,\ \omega = .765$), while range in eye separation had the smallest (significant) effect ($F(1, 2286) = 11.064,\ \omega = .065$).

Two-tailed bivariate correlation tests were performed on the Human and Robot groups separately (see Tables 5 and 6). The Human group had significant increases in sensitivity to the best point, in all four proportions, with attributes humanlike, alive, female, and sexy. It also had significant decreases sensitivity, in three of four proportions, with increased ratings of the attributes creepy and ugly. All ps $< .01$ except alive–eye separation with $p < .05$ and sexy–cheek width and creepy–jaw width, both $p >, 05$. The Robot group produced less predictable results for sensitivity. Correlations between attributes humanlike, creepy, and ugly were significant at $p < .01$ for all four proportions, but the direction of those correlations were mixed, showing increased sensitivity to cheek width and face height, and decreased sensitivity for eye separation and jaw width for each of these three attributes.

Tolerance for the acceptable range increased in the Human group for all four proportions as ratings of humanlike, creepy, and ugly increased, and as ratings of alive, sexy, and female decreased except female–eye range. All ps $< .01$, except humanlike–eye separation with $p < .05$, and humanlike–face height, alive–cheek width, and sexy–cheek width $p > .05$. Correlations for tolerance in the Robot group were weaker, and fewer correlations reached a significance level of $p < .05$. Of those attributes with significant correlations between at least three proportions, tolerance decreased as ratings of alive and sexy decreased.

Table 5: Correlation between Attributes and Facial Proportions of Human Figures

	Sensivity[1]				Tolerance[2]			
Attribute	Cheek	Eyes	Face	Jaw	Cheek	Eyes	Face	Jaw
Humanlike[3]	-.138**	-.053	-.348**	-.177**	.128**	.085*	.003	.102**
Alive[4]	-.064*	-.176**	-.391**	-.197**	-.051	-.135**	-.227**	-.070*
Female[4]	-.436**	.152**	-.639**	-.532**	-.069*	.086**	-.120**	-.076**
Sexy[4]	.051	-.301**	-.481**	-.404**	-.039	-.134**	-.176**	-.118**
Creepy[4]	-.121**	.214**	.220**	.052	.111**	.157**	.208**	.139**
Ugly[4]	-.165**	.306**	.270**	.204**	.094**	.151**	.184**	.157**

* p < .05
** p < .01

[1] Sensitivity is participant's difference from the mean.
[2] Tolerance is participant's acceptable range.
[3] $n = 768$ for cheek width and 852 for other proportions
[4] $n = 1164$ for cheek width and 1248 for other proportions

Table 6: Correlation between Attributes and Facial Proportions of Robotic Figures

	Sensivity[1]				Tolerance[2]			
Attribute	Cheek	Eyes	Face	Jaw	Cheek	Eyes	Face	Jaw
Humanlike[3]	-.328**	.242**	-.547**	.144**	0̇54	.014	.124**	.032
Alive[4]	.061	.168**	-.017	-.084**	-.065*	-.059	-.129**	-.150**
Female[4]	-.042	.315**	-.040	-.111**	.073*	.097**	.054	0̇43
Sexy[4]	.199**	-.050	.163**	-.115**	-.080*	-.091**	-.069*	-.131**
Creepy[4]	-.427**	.329**	-.467**	.104**	.040	0̇08	-.091**	.108**
Ugly[4]	-.416**	.264**	-.470**	.156**	.070*	0̇26	-.068*	.143**

* p < .05
** p < .01

[1] Sensitivity is participant's distance from the mean
[2] Tolerance is participant's acceptable range
[3] $n = 640$ for cheek width and 710 for other proportions
[4] $n = 970$ for cheek width and 1040 for other proportions

Comparison of Figures by Level of Realism

The dataset was divided into three groups based on each figure's level of realism: 2D, 3D, and photograph. Two groups (2D and 3D) contain a female, male, and robot figure. The photo group contains five figures, a female and a male, and three robots. A one-way ANOVA was performed on the six attributes and difference and range. Significant between group differences of attributes were detected at $p < .001$, except in female ($p < .05$), and humanlike which did not reach a level of $p < .05$. Differences between sensitivity were significant for all facial proportions at $p < .001$, except face height which was significant at $p < .01$. Tolerance for each facial proportion showed significant differences, except face height, at various levels: cheek width ($p < .001$), eye separation ($p < .05$), and and jaw width ($p < .001$). Two-tailed bivariate correlations were performed for each group between participant's ratings of attributes and tolerance, and attributes and sensitivity to facial proportions. Tables 7 through 9 provide details.

Participants showed increased sensitivity to the best point all four proportions for stimuli in the Photograph group as the ratings for attributes humanlike, alive, female and sexy increased. Participants also showed a mix of sensitivity for photographic stimuli as ratings of attributes creepy and ugly increased. All correlations were significant at the $p < .01$ level except for ugly–cheek width ($p < .05$) and humanlike–eye separation ($p > .05$). For the 3D group, all correlations were significant at $p < .01$ except humanlike–eye separation, alive–eye separation, and creepy–jaw width (all $p > .05$). Participants showed increased sensitivity to face height and jaw width but decreased sensitivity to cheek width and eye separation as ratings of humanlike, alive, female, and sexy increased. Sensitivity decreased for face height and jaw width, and increased for cheek width and eye separation of ratings of creepy and ugly increased for the 3D group. The 2D group had no attribute, except female, where sensitivity either increased or decreased in more than two proportions as attribute ratings increased. Relations to the attribute alive were weak with face height and jaw width reaching a significance of $p < .05$, and cheek width and eye separation both $p > .05$.

Participants showed decreased tolerance in all facial proportions as ratings of humanlike, alive, female and sexy increased, and as ratings of creepy and ugly decreased for the Photograph group. These relations were not as strong as those shown for sensitivity, and fewer relations reached significance of $p < .05$ or $p < .01$. Correlations between the attributes humanlike, alive, sexy, creepy, and ugly, and tolerance for the 3D group were similar to those of the Photograph group for all proportions except cheek width. The correlations related to humanlike, alive, and sexy were weaker for the 3D group than the Photographic group, while the correlations related to attributes creepy and ugly were slightly stronger. Correlations for 2D group showed an increase in tolerance in all proportions as ratings of humanlike, creepy, and ugly increased, and a decrease in tolerance in all proportions as ratings of alive and sexy increased. The 2D group had weaker correlations for attributes humanlike, alive, and sexy than either the Photograph and 3D groups, and strong correlations for attributes creepy and ugly.

Table 7: Correlation between Attributes and Facial Proportions of Human and Robot Photos

	Sensivity[1]				Tolerance[2]			
Attribute	Cheek	Eyes	Face	Jaw	Cheek	Eyes	Face	Jaw
Humanlike[3]	-.848**	-.002	-.843**	-.296**	-.117**	-.068	-.276**	-.028
Alive[4]	-.586**	-.069*	-.570**	-.428**	-.184**	-.118**	-.300**	-.185**
Female[4]	-.420**	-.632**	-.306**	-.334**	-.146**	-.066*	-.078*	-.054
Sexy[4]	-.463**	-.345**	-.411**	-.555**	-.223**	-.161**	-.236**	-.186**
Creepy[4]	-.100**	.266**	-.081**	.529**	.060	.009	.021	.098**
Ugly[4]	-.077*	.298**	-.087*	.536**	.106**	.065*	.017	.132**

* p < .05
** p < .01

[1] Sensitivity is participant's difference from the mean.
[2] Tolerance is participant's acceptable range.
[3] $n = 640$ for cheek width and 710 for other proportions
[4] $n = 970$ for cheek width and 1040 for other proportions

Table 8: Correlation between Attributes and Facial Proportions of Three-dimensional Human and Robotic Figures

	Sensivity[1]				Tolerance[2]			
Attribute	Cheek	Eyes	Face	Jaw	Cheek	Eyes	Face	Jaw
Humanlike[3]	.548**	.059	-.706**	-.886**	.230**	-.083	-.246**	-.042
Alive[4]	.331**	.005	-.433**	-.536**	.029	-.149**	-.233**	-.122**
Female[4]	.751**	.651**	-.824**	-.364**	.307**	.053	-.188**	.020
Sexy[4]	.626**	.307**	-.698**	-.600**	.190**	-.088*	-.230**	-.089*
Creepy[4]	-.185**	-.204**	.200**	.038	.109**	.107**	.164**	.180**
Ugly[4]	-.273**	-.193**	.343**	.197**	-.016	.101*	.194**	.166**

* p < .05
** p < .01

[1] Sensitivity is participant's difference from the mean.
[2] Tolerance is participant's acceptable range.
[3] $n = 384$ for cheek width and 426 for other proportions
[4] $n = 592$ for cheek width and 624 for other proportions

Table 9: Correlation between Attributes and Facial Proportions of Two-dimensional Human and Robotic Figures

	Sensivity[1]				Tolerance[2]			
Attribute	Cheek	Eyes	Face	Jaw	Cheek	Eyes	Face	Jaw
Humanlike[3]	-.201**	.311**	.224**	-.522**	.057	.233**	.162*	.062
Alive[4]	.016	.012	.090*	-.079*	-.091*	-.026	-.070	-.123**
Female[4]	-.863**	.906**	-.414**	-.659**	-.263**	.249**	-.067	-.069
Sexy[4]	.047	-.020	.138**	-.106**	-.123**	-.075	-.057	-.095*
Creepy[4]	-.215**	.184**	-.189**	-.073	.110**	.159**	.129**	.182**
Ugly[4]	-.234**	.225**	-.193**	-.110**	.159**	.142**	.112**	.196**

* $p < .05$
** $p < .01$

[1] Sensitivity is participant's difference from the mean.
[2] Tolerance is participant's acceptable range.
[3] $n = 384$ for cheek width and 426 for other proportions
[4] $n = 592$ for cheek width and 624 for other proportions

Comparison of Figures Grouped by Attribute Ratings

Multidimensional visualizations of attribute ratings consistently showed the proximity of certain figures. Figure 8 is a 2D Kernel Isometric feature map (Kernel ISOMAP) using the six attributes as dimensions. Principal components analysis (PCA) and multidimensional scaling (MDS) yield similar visualizations.

Based on the visualizations, the following groups were formed; 1) Realistic Females (the 3D female and female photograph), 2) Other Humans (all male figures and the 2D female), 3) Other Robots (all except Barthoc, Jr.), and 4) Barthoc, Jr. which is a clear outlier. The stimuli in Other Humans group did not appear as close to one another as those in the Other Robots, or Realistic Female groups, but were not appear as far away from the other figures as Barthoc, Jr. Table 10 which sorts and displays the figures by the average attribute rating supports this grouping. (Since creepy and ugly are negative judgements, their ratings were negated to create a more positive association.) Each grouping contains a consecutive set of figures.

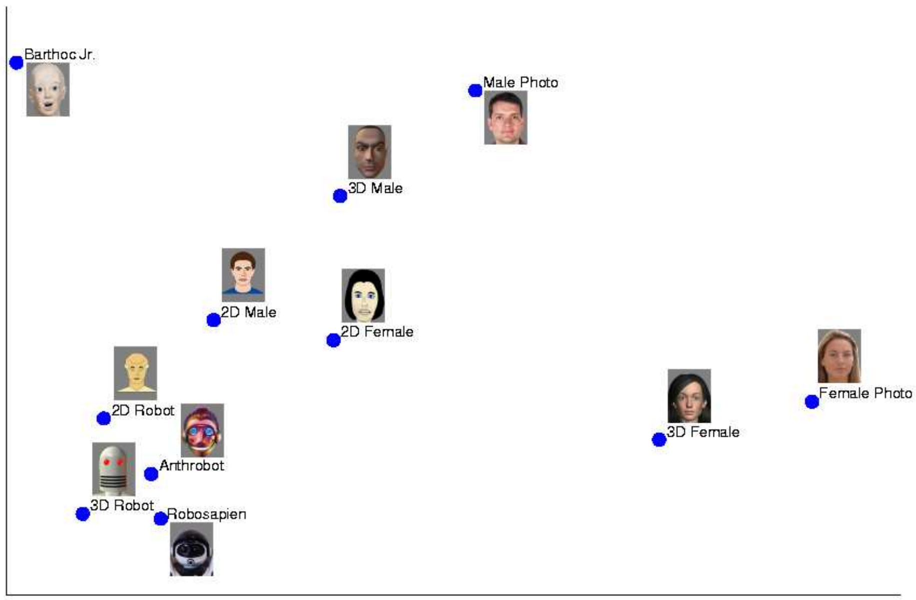

Figure 8: 2D kernel ISOMAP of 11 figures by 6 attributes

Table 10: Participant Attributes about Figures and Average Rating

Figure	Humanlike[1]	Alive[2]	Female[2]	Sexy[2]	Not Creepy[2]	Not Ugly[2]	Mean Rating
Group 1							
Female Photo	2.89	2.50	2.79	1.57	2.40	2.37	2.42
3D Female	1.90	0.31	2.66	1.20	1.93	2.22	1.70
Group 2							
Male Photo	2.80	2.51	-2.90	-0.44	1.71	1.16	0.81
3D Male	1.73	0.12	-2.75	-0.65	0.82	0.97	0.04
2D Female	1.07	-1.92	1.92	-2.10	0.41	-0.25	-0.14
2D Male	0.76	-1.77	-2.55	-1.84	1.30	0.73	-0.56
Group 3							
Anthrobot	-2.36	-1.43	-1.91	-2.07	1.77	1.41	-0.77
Robosapien	-2.08	-1.42	-1.58	-2.07	0.87	0.87	-0.90
3D Robot	-2.66	-2.06	-2.07	-2.36	1.18	0.82	-1.19
2D Robot	-1.46	-2.14	-2.63	-2.31	0.79	0.33	-1.24
Group 4							
Barthoc, Jr.	0.63	-1.28	-1.03	-2.88	-2.46	-2.37	-1.56

Ratings of attributes Creepy and Ugly have been negated to express a positive attribute.

[1] $n = 142$

[2] $n = 208$

These grouping align with the Human and Robot groups described above. The Human group contained the Realistic Females and Other Humans, and the Robot group contained Barthoc, Jr. and Other Robots.

Participants showed increased sensitivity to the cheek width and eye separation of the Realistic Female group as ratings of humanlike, alive, female, and sexy increased, and as ratings of creepy and ugly decreased. All correlations were significant at ps $<$ .01, except ugly–eye separation ($p < .05$) and ugly–face height ($p > .05$). Participants also showed decreased sensitivity to face height as ratings of humanlike, alive, female and sexy increased, and as ratings of creepy and ugly decreased. Sensitivity to jaw width was similar to cheek width and eye separation, although only the relations to attributes female and ugly were significant at $p < .01$. The Realistic Females had few significant correlations between attributes and tolerance. Those relations that were significant at $p < .01$ were alive–cheek width ($r = -.304$), alive–eye separation ($r = -.141$), creepy–cheek width ($r = .220$), and creepy–eye separation ($r = .144$), The relations sexy–cheek width ($r = -.141$), and creepy–jaw width were signficant at $p < .05$. Details for the Realistic Female group are in Table 11.

Sensitivity towards the best proportion was more variable in the Other Humans group, but Tolerance for acceptable range showed a more consistent pattern. These relations are shown in Table 12. Generally the relations for sensitivity are weaker than those of the Realistic Female group. In other groupings the relations between sensitivity and humanlike, alive, and sexy were the opposite of those for creepy and ugly. This was not exhibited in any facial proportion for the Other Humans group. Tolerance increased in all proportions except cheek width as ratings of alive and sexy

Table 11: Correlation between Attributes and Facial Proportions of Realistic Female Figures

	Sensivity[1]				Tolerance[2]			
Attribute	Cheek	Eyes	Face	Jaw	Cheek	Eyes	Face	Jaw
Humanlike[3]	-.560**	-.525**	.533**	-.048	-.037	-.021	.056	.095
Alive[4]	-.572**	-.586**	.571**	.064	-.304**	-.141**	-.045	-.006
Female[4]	-.134**	-.126**	.091	-.132**	.065	.057	.075	.087
Sexy[4]	-.133**	-.145**	.142**	.030	-.116*	-.088	-.026	-.039
Creepy[4]	.204**	.207**	-.196**	.071	.220**	.144**	.094	.097*
Ugly[4]	.083	.115*	-.075	.139**	.068	.036	.038	.014

* p < .05
** p < .01

[1] Sensitivity is participant's difference from the mean.
[2] Tolerance is participant's acceptable range.
[3] $n = 256$ for cheek width and 388 for other proportions
[4] $n = 284$ for cheek width and 416 for other proportions

increased (all ps $< .01$, except alive–jaw width with $p > .05$). Tolerance decreased in all four proportions as ratings of humanlike, creepy, and ugly increased (all ps $< .01$, except creepy–cheek width with $p < .05$ and humalike–face height with $p > .05$). Relations for tolerance were generally stronger for the Other Humans group than for the Realistic Female group.

Table 12: Correlation between Attributes and Facial Proportions of Other Humans

	Sensivity[1]				Tolerance[2]			
Attribute	Cheek	Eyes	Face	Jaw	Cheek	Eyes	Face	Jaw
Humanlike[3]	-.080	. 118**	-.359**	-.026	.189**	.143**	.048	.118**
Alive[4]	.032	.010	-.422**	-.018	.042	-.101**	-.220**	-.082*
Female[4]	-.745**	.735**	-.272**	-.217**	-.186**	.265**	.007	.022
Sexy[4]	.081*	-.090**	-.115**	-.101**	-.054	-.099**	-.104**	-.113**
Creepy[4]	-.181**	.107**	.050	-.173**	.092*	.134**	.184**	.133**
Ugly[4]	-.209**	.186**	-.043	-0.054	.137**	.145**	.141**	.174**

* p < .05
** p < .01

[1] Sensitivity is participant's difference from the mean.
[2] Tolerance is participant's acceptable range.
[3] $n = 512$ for cheek width and 776 for other proportions
[4] $n = 568$ for cheek width and 832 for other proportions

Sensitivity correlation for the Other Robots group were generally weaker than those of the Other Human group, and in turn the Realistic Female group. These correlations are detailed in Table 13. Sensitivity decreased as ratings of humanlike, creepy, and ugly increased, though only a few relations were significant. Sensitivity increased as ratings of alive, female and sexy increased. All relations were significant at $p < .01$ for alive and female. For the attribute sexy, the relation to cheek width was significant at $p < .01$ and jaw width at $p < .05$, but the others were $p > .05$. Tolerance increased with increased ratings of attributes female, creepy, and ugly increased, and as ratings of sexy decreased. All relations were significant at $p < .01$, except sexy–cheek width and creepy–eye separation (both $p < .05$) and creepy–face height and ugly–face height which failed to reach significance at $p < .05$.

Table 13: Correlation between Attributes and Facial Proportions of Other Robots

	Sensivity[1]				Tolerance[2]			
Attribute	Cheek	Eyes	Face	Jaw	Cheek	Eyes	Face	Jaw
Humanlike[3]	-.013	-.071	-.296**	.028	.129**	.068	-.005	.002
Alive[4]	.214**	.107**	.119**	-.139**	.005	-.010	-.060	-.093**
Female[4]	.172**	.242**	.245**	-.226**	.127**	.191**	.178**	.089**
Sexy[4]	.115**	.049	.033	-.078*	-.081*	-.111**	-.111**	-.126**
Creepy[4]	-.151**	.076*	-.066	-.053	.106**	.082*	.035*	.113**
Ugly[4]	-.165**	-.001	-.122**	.041	.132**	.106**	.058	.163**

* p < .05
** p < .01

[1] Sensitivity is participant's difference from the mean.
[2] Tolerance is participant's acceptable range.
[3] $n = 512$ for cheek width and 776 for other proportions
[4] $n = 568$ for cheek width and 832 for other proportions

There are few significant relations between sensitivity and attributes for Barthoc, Jr. However, there are many significant relations between tolerance and attribute ratings. Tolerance decreased as ratings of alive, and sexy increased, and as ratings of creepy and ugly decreased (all ps at least $< .05$ except creepy–cheek width and ugly–eye separation). All relations between tolerance and the attribute humanlike were $p > .05$.

Relation between Attributes and Best Points

All results reported up to this point have been in terms of sensitivity to the best point (standard deviation) or tolerance (acceptable range). Table 14 lists the mean of the best point recorded by figure and facial proportion. The variation between figures is because of the variation between original proportions. Table 1 provided the original proportions.

Table 14: Best Proportions by Figure and Facial Proportion

Figure	Cheek[1]	Eyes[2]	Face[2]	Jaw[2]
Female Photo	0.6020	0.4842	0.6628	1.6453
Male Photo	0.4417	0.6511	0.5910	1.7964
3D Female	0.5392	0.5606	0.4931	1.7577
3D Male	0.6126	0.5061	0.7513	1.6447
2D Female	0.3317	0.7179	0.5901	1.6889
2D Male	0.7306	0.4698	0.7377	1.7437
Anthrobot	0.6078	0.5634	0.7151	1.5485
Barthoc, Jr.	0.4758	0.5854	0.5144	1.9123
Robosapien	0.7297	0.4746	0.6338	1.9242
3D Robot	0.5905	0.4788	0.7532	1.9882
2D Robot	0.4963	0.4548	0.5767	1.8539

[1] $n = 194$
[2] $n = 208$

A two-tailed bivariate correlation between the six attributes and best points for the four facial proportions indicates that a narrower cheek and jaw, and short face height, wider eyes are related positively with humanlike and alive (all ps $< .01$). The attribute sexy only had significant correlations with face height and jaw width (both $p < .01$) aligning with humanlike and alive. Creepy and ugly also had positive correlations with a narrower cheek, short face height and wider eye separation, but jaw width had a negative effect on both attributes. Table 15 details the correlations between attributes and best points.

Table 15: Correlation between Attributes and Best Facial Proportions

Attribute	Cheek	Eyes	Face	Jaw
Humanlike[1]	-.277**	.336**	-.233**	-.270**
Alive[2]	-.077**	.120**	-.090**	-.178**
Female[2]	-.268**	.272**	-.368**	-.230**
Sexy[2]	.037	-.008	-.094**	-.252**
Creepy[2]	-.234**	.167**	-.193**	.126**
Ugly[2]	-.247**	.197**	-.168**	.157**

* $p < .05$
** $p < .01$

[1] $n = 1408$ for cheek width and 1562 for other proportions
[2] $n = 2134$ for cheek width and 2288 for other proportions

Another way of looking at the best point is to compare the best point to the original proportion. The selected frame is used to make this comparison, each frame representing a 1% change. Table 16 indicates the smallest changes from the original proportion are found in the face height of the 2D female (.02%) and cheek width of Robosapien (-0.05%). The largest differences between the original and best proportions were the jaw width of 3D female (-6.00%) and male photograph (-5.45%).

Table 16: Best Frames by Figure and Facial Proportion

	Best Frame				Difference from Original Frame[3]			
Figure	Cheek[1]	Eyes[2]	Face[2]	Jaw[2]	Cheek[1]	Eyes[2]	Face[2]	Jaw[2]
Female Photo	8.10	11.88	11.42	8.36	-2.90	0.88	0.42	-2.64
Male Photo	9.15	8.18	9.10	5.55	-1.85	-2.82	1.90	-5.45
3D Female	12.73	11.10	11.63	5.00	1.73	0.10	0.63	-6.00
3D Male	8.24	8.33	11.17	7.18	-2.76	-2.67	0.17	-3.82
2D Female	11.53	9.89	11.02	10.35	0.53	1.11	0.02	-0.65
2D Male	12.27	11.05	12.05	10.07	-1.27	-0.05	1.05	-0.93
Anthrobot	9.04	9.84	13.16	9.63	-1.96	-1.16	2.16	-1.37
Barthoc, Jr.	9.76	13.44	10.14	10.88	1.24	-2.44	0.86	0.12
Robosapien	10.95	7.82	13.23	9.17	-0.05	3.18	2.23	-1.83
3D Robot	11.09	13.28	14.44	9.43	0.09	2.28	-3.44	-1.57
2D Robot	8.31	9.94	12.18	7.43	-2.69	1.06	1.18	3.57

[1] $n = 194$
[2] $n = 208$
[3] Difference in frames (or percentage) corrected for direction of change in Flash movie

Difference in Response by Gender

An independent samples t-test was performed, grouped by the participant's gender, on the six attributes, four differences from best point, and four acceptable ranges. Female participants rated each attribute higher than male participants, though none of these differences was significant at $p < .05$. Female participants also showed less sensitivity and more tolerance in selecting facial proportions than male participants. The difference in tolerance was significant at $p < .01$ in all four proportions, though the effect sizes were all small with tolerance for check and jaw width the only differences with an $r > .10$. To further understand the effect of participant's gender the stimuli were separated into Female, Male, and Robot groups..

In these groupings female participants continued to show less sensitivity and more tolerance than male participants, with some variation between groups. Male participants rated female figures as more alive ($M = .42$, $SE = .120$) and sexy ($M = .39$, $SE = .110$) than female participants ($M = .17$, $SE = .153$ for alive, and $M = -.02,\ SE = .129$ for sexy). Only the attribute sexy had a significant difference ($t(619) = -2.635,\ p < .05,\ r = .095$) though the effect size was very small. Male participants also rated male figures as more alive ($M = .45,\ SE = .119$) than did female participants ($M = .09,\ SE = .152$), but the difference was not significant ($p > .05$). Female participants rated robots as more humanlike by female participants ($M = -1.38,\ SE = .105$) than by male participants ($M = -1.75,\ SE = .083$). The difference was significant with a small effect size ($t(557) = 2.734,\ p < .01,\ r = .113$). Male participants rated the Robot group slightly more sexy ($M = -.50,\ SE = .083$) than did female participants ($M = .51,\ SE = .103$). The difference in ratings of sexy was significant though the effect size was very small ($t(1033) = -2.650,\ p < .01,\ r < .082$).

Result Summary

Relations between attributes and sensitivity, and attributes and tolerance are generally stronger more human and realistic a stimulus appears. In most instances participants showed increased sensitivity and decreased tolerance as ratings of humanlike, alive, female, and sexy increased, and as ratings of creepy and ugly decreased. However, this pattern had some variance between figure groupings and proportions. Participant's gender can lead to different results, particularly in assessing sexiness, and the human likeness of robots. Female participants also seem to be more tolerant of varying facial proportions.

CHAPTER FIVE: DISCUSSION

Sensitivity and Tolerance Relative to Human Likeness

H1 predicted that the acceptable range will narrow (tolerance will decrease) as ratings for human likeness increases. This hypothesis can not be supported by this study. Tolerance for the range of acceptable facial proportions is not strongly correlated with participant's ratings of human likeness when viewing results for all 11 figures. Only tolerance to face height decreased as human likeness increased, while tolerance of cheek width slightly increased as human likeness increased.

Tolerance decreased as human likeness increased in all proportions in the Photograph group, but only significantly for cheek width, and face height. Other groupings did not have significance at $p < .05$, or significant relations were weak or indicated an increase in tolerance, rather than a decrease.

This study provides good support for H2: there will be greater intersubjective agreement - sensitivity - as human likeness increases. Greater sensitivity is measured by smaller differences from the best point. As ratings of humanlike increased, sensitivity increased in all proportions except eye separation. Face height and jaw width had large effect sizes while viewing all figures. Sensitivity decreased slightly in eye separation as human likeness increased, but the effect size was small.

Sensitivity to the best proportion was better correlated to human likeness in human figures than in robots. Also the more realistic figures had greater sensitivity, larger r, and more proportions had an increase related with an increase in human likeness. Photographic images demonstrated the same pattern of sensitivity and tolerance that was demonstrated for the human group.

There was almost no measurable correlation between human likeness and sensitivity or tolerance for the two-dimensional figures. Our immersion in two-dimensional images might cause this. Whether it is artistic cubism or cartoon characters, we do not expect 2D characters to measure up to three-dimensional humans. Characters that seem to share our space may cause us to evaluate them on our terms.

Sensitivity and Tolerance Relative to Attractiveness

H3 and H4 are both related to ratings of attractiveness. This study asked participants to rate "Ugly" and "Sexy." While these terms may not be exact antonyms (attractive–ugly) or synonyms (attractive–sexy), they will be used to evaluate attractiveness.

The results of this study support hypothesis H3, which predicts decreased tolerance (a narrower acceptable range) as ratings of attractiveness increase. Tolerance decreased significantly for the relation between sexy and eye separation, face height, and jaw width. The relation between sexy and cheek width showed decreased tolerance but the effect size was very small and the relation did not reach significance. The relations between ugly and each facial proportion showed increased tolerance as ugliness increased, and all relations were significant.

The relation between tolerance and sexiness was inverted between the Human and Robot groups. Participants showed decreased tolerance to human figures as sexiness increased but, increased tolerance to robots for the same increase in sexiness. The groupings by Realism displayed the general pattern of an increase in sexiness relating to a decrease in tolerance, and an increase in ugliness relating to an increase in tolerance. As Realism decreases from Photographs to 3D and then to 2D the strength of the relation between tolerance and sexiness decreased while the strength of the relation between tolerance and ugly increased. This pattern, of tolerance increasing with ugliness and decreasing with sexiness, holds in each of the attribute rating groups except Realistic Females. The Realistic Females were rated the most sexy and least ugly of all figures.

Hypothesis H4, sensitivity does not vary based on assessments of attractiveness, is not supported. Participants had greater sensitivity to the best proportions in sexy figures, and less sensitivity in ugly figures than those as the other end of the scale.

There were significant correlations between attributes sexy and ugly, and all four facial proportions. Sensitivity increased as sexiness increased in eye separation, face height, and jaw width. Likewise, sensitivity decreased as ugliness decreased in eye separation, face height, and jaw width. Sensitivity to the best cheek width moved in the opposite direction of the other facial proportions, increasing as ugliness increased and decreasing as sexiness increased.

This effect was observed in both the Human and Robot groups, though the relation between sensitivity and facial height flipped for sexy in the Robot group. As reported with tolerance, the strength of relations decreased as the figure's level of realism decreased. Photographs had strong correlations between sensitivity and sexy, decreasing in all proportions, while sensitivity decreased in eye separation and jaw width as ugliness increased. The effect was even stronger in the 3D group; sensitivity to face height and jaw width increased with an increase in sexiness, and decreased with an increase in ugliness. However, sensitivity had the opposite effect in cheek width and eye separation. Correlations for 2D characters were generally not as strong and a little more sporadic. For the attribute rating groups this effect was strongest within the Other Humans group. Little effect could be observed when isolated in the other groups.

What are “Best” Proportions?

Thus far all discussion has been on sensitivity and tolerance. Do the actual selections of best points reveal anything interesting? Costa and Corassa (2006) found artists exaggerate the size of the eyes and lips, and the length of the face. Table 16 lists the frame selected the as the best point for each figure and facial proportion, and shows the extent of change from the figure’s original proportion. Participants preferred narrower cheeks and jaws and longer faces than the original images. Preferences for eye separation differed between human and robotic figures. Participants preferred narrower set eyes in human figures and wider set eyes in robotic figures.

Another finding from these data is that there is relation between human likeness and the average difference between the original and best frames. The smallest differences between original and best frames seem to be associated with characters with neutral humanlike ratings (see Figure 9). These three figures are the 2D humans and Barthoc, Jr. The 2D humans may have subconsciously been drawn with the type of artistic norms reported by Costa, with participants reacting in the same manner.

Barthoc, Jr. was also *designed* to look human. Did the designers succeed? If so, perhaps that is why the small difference between the best and original points. Another explanation might be that this robot was considered so creepy that the

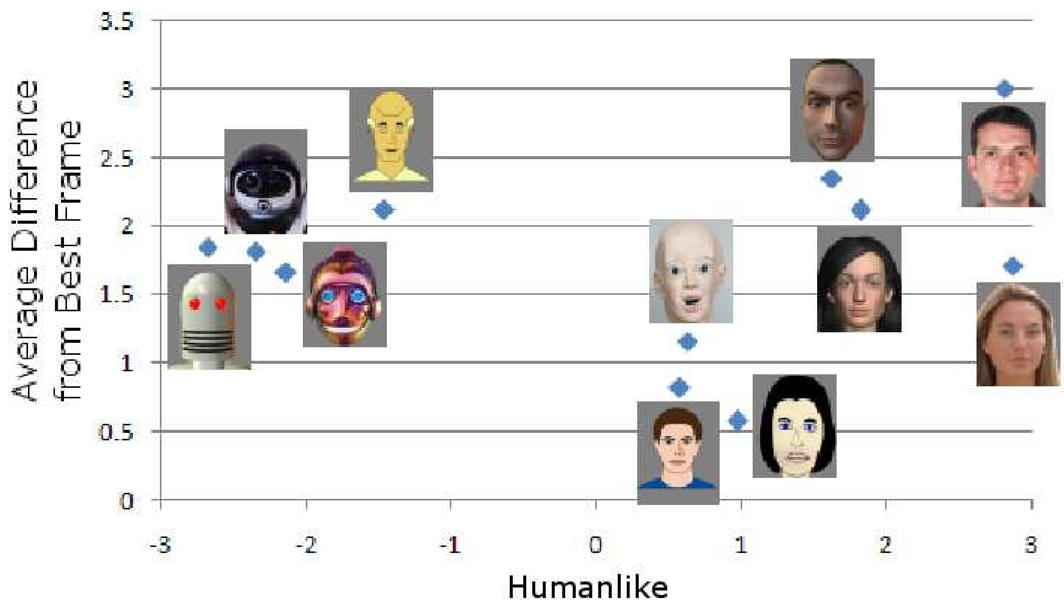

Figure 9: Average difference of best and original frame against humanlike

selected "best" frames were those the alleviated the creepiness to the greatest extent.

It might be surprising that the robots Anthrobot and Robosapien had as much variation in selection of best point for their cheeks, jaws, and face height. Hard surfaces define each of these proportions and changing the proportions would involve bending metal or molding plastics. It also might be surprising that the best point for the 3D robot's cheek and jaw were not the original point, as it has straight sides. Could this deviation from design represent a desire to create things that look like us?

The Creepiness of the Unknown Human

A two-tailed bivariate correlation among attributes shows strong, positive correlations between humanlike and alive, sexy, and female. The relation between humanlike and creepy, and between humanlike and ugly were also significant but weak.

Further analysis reveals an Uncanny Valley. Figure 10 plots the mean value for the attributes alive, sexy, creepy, and ugly when participants rated humanlike according to the values on the x-axis. (Mori, 1970) predicted an Uncanny Valley when a figure is almost perfectly human, but not quite. This study indicates participants sensed heightened creepiness when they rated humanlike as neutral or slightly agree.

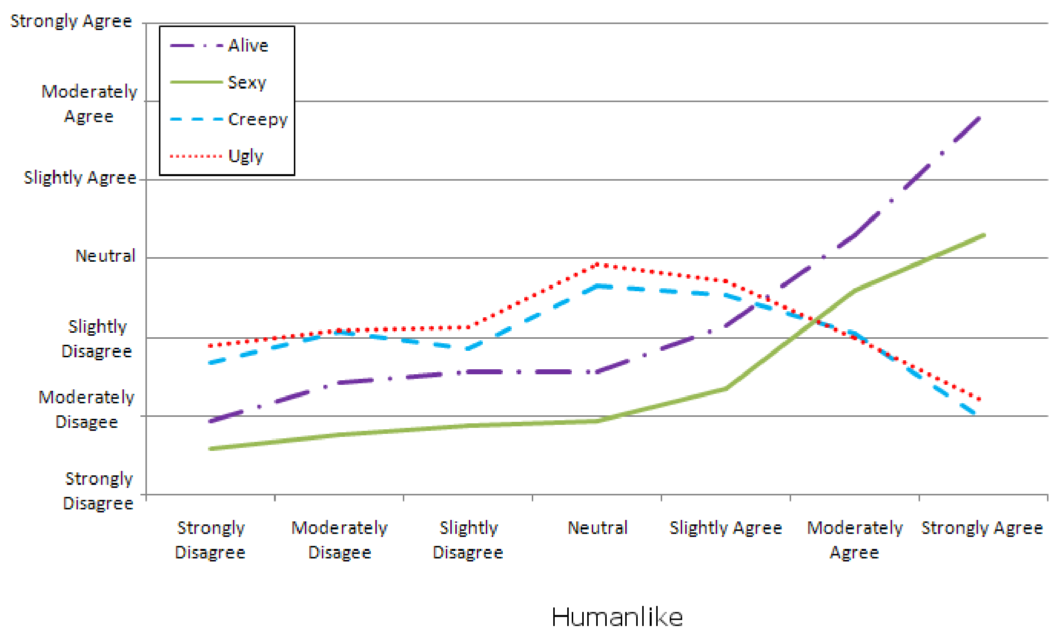

Figure 10: Ratings of humanlike versus mean of other attributes. The rating for the attribute creepy is at its highest when human likeness is indeterminate.

It would be easy to ascribe this phenomenon to the effect of the robot Barthoc, Jr. (Figure 11). Barthoc, Jr. had the highest rating for both creepy and ugly, while the rating for humanlike was very close to neutral. But the Means Plot in Figure 12 indicates Barthoc, Jr. was actually considered least creepy when its human likeness is in question. In fact, it was considered most creepy when participants moderately disagreed or moderately agreed with the human likeness of Barthoc, Jr.

Figure 11: Robot "Barthoc, Jr."

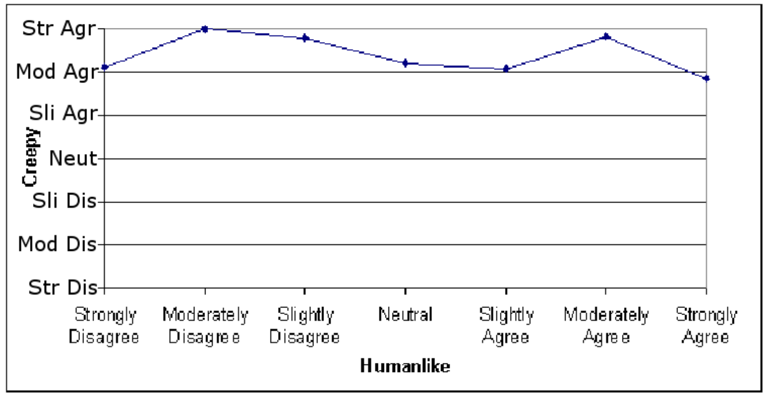

Figure 12: Ratings of humanlike versus mean of creepy for Barthoc, Jr.

Reviewing the relation between human likeness and creepiness in other characters reveals the largest spike in creepiness is associated with the 2D Male (see 13). The 2D Female figure shows a rise in creepiness as ratings of human likeness rise from neutral to moderately agree. Participants who strongly agreed or disagree with the 2D Female being humanlike rated the figure correspondingly creepy. Robosapien was not rated higher than neutral for the attribute humanlike, but that neutral point corresponded with his most creepy rating. In contrast, the 3D Female was not rated lower than neutral for humanlike, but that neutral point also corresponded with her most creepy rating.

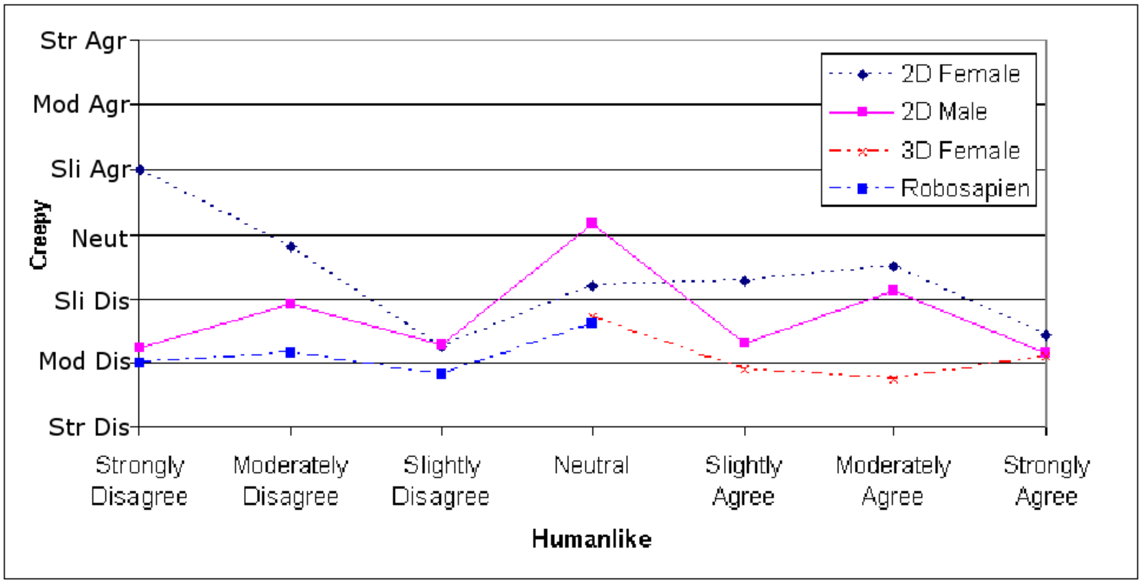

Figure 13: Ratings of humanlike versus mean of creepy for 2D male, 2D female, 3D female, and Robosapien

Gender Differences

There was a concern that male participants would be more reluctant to rate male figures as sexy, than female participants would be for female figures. Indeed, ratings of male figures for the attribute sexy where higher from female participants than from male participants; however, this difference was not statistically significant. What did achieve statistical significance was the difference between ratings of sexy for female figures, with male participants providing higher ratings than their female counterparts. So it seems the concern of gender bias focused on the wrong sex.

Female participants showed greater tolerance in the acceptable range of facial proportions to all stimulus types (Female, Male, and Robot). Female participants also considered the robotic characters to be more humanlike than did male participants. This difference was significant, but the effect size was small. These results may be related. Possible causes could include female's traditional nurturing roles causing increased acceptance, males being more comfortable with mechanical object and having less need to anthropomorphize robots, and males being more critical in evaluating potential female partners. All these are conjecture, and this study does not pretend to address these questions.

Room for Improvement: Issues in Experimental Design

There were several issues that compromised the overall experimental design. The following sections detail some of the major issues.

Figure Selection and Number of Figures

The collection of figures was probably too small. The requirements of the experiment were to find frontal head shots with minimal obstructions from hair, hands or other objects and with neutral facial expressions. Human photographs were collected from a royalty free photographic archive, and photographs of robots were provided by researchers. Finding images that provided this sort of unobstructed head shot proved difficult and limited the number of acceptable figures.

Figures at the same level or realism varied too much in terms of attributes such as sexy and ugly, which may have biased results relating to realism. Additionally, the quality and perceived realism of the 2D and 3D figures varied as they came from different sources and were for the most part collected rather than created for this study. It was also difficult to locate humanoid robots that had enough "face" for the experimental parameters and that was not caricatured to the point of being unusable.

The small set of figures allowed for a reasonable experiment length for volunteer participants and allowed a within-subjects analysis. However, with single figures representing a category, a figure's idiosyncracies could influence results, particularly when broken into smaller groups. For example, the 3D robot had straight sides and a long head. These features, probably, limited the best and acceptable range for cheek and jaw width more than any other relation. Likewise, because it lacked other facial landmarks, face height and eye separation may have been more volatility than other figures.

Using a larger set of characters of each type might lead to an understanding of whether each type of stimulus will tend towards proportions found to be attractive in humans, or if there are a set of proportions that might be best for each particular character type.

Proportion Control

In previous studies on human attractiveness, images were usually standardized either by taking photographs at a fixed distance with standard lighting, of by adjusting images by fixing one proportion such as distance between pupils or height of the head. While the images were created in a standard size (400x500 pixels) the head size and proportions within the image were not controlled. Some heads, such as the 3D male were very large and nearly filled the frame, while others were comparatively small. Even if images had been standardized there were great differences in the design of the synthetic figures, particularly the robots.

Perhaps a better approach would have been to have used the same set of stimuli, humans and robots. Standardized images of these stimuli could be created. Then 2D and 3D images of each of the stimuli could be created with varying degrees of realism.

Another aspect of proportion control is determining what is really affecting participants' judgements. When each dimension is being varied it changes several facial proportions, not just the one we wish to assess. For example, changing the width of the face at the cheek changes the proportion defined as cheek width, but it also changes the proportion eye separation.

Other Figure Factors

The experiment also did not control for other factors that may have affected participants attributes about the figures. If the robot Bathoc Jr. shared the exact facial proportions of the Female Photo it would probably still be considered creepy because of the lack of smoothness and pigment in the skin. The hard edges exhibited by the 3D Robot, Anthrobot and Robospaien may also have artificially limited the points selected by participants.

Attributes

Participant attribute ratings for the figures were measured on a 7-point Likert scale. Values were assigned on a scale of Strongly Disagree (-3) to Strongly Agree (3) with statements such as "The figure is ugly." Responses related to disagreement leave some room for interpretation. For instance, does a participant who strongly disagrees with the statement "The figure is ugly" believe the figure is beautiful or just that it is definitely not ugly? Another example would be ratings for “The figure is female.” All robots were rated from slightly disagree to strongly disagree. Did participants consider the robots as males, or just not females?

A paired word approach could have removed much of this ambiguity, but selecting appropriate word pairs can be a challenge particularly when translating to other languages. Because the attributes were not paired, some terms were inverted so all questions did not seem to have a positive emphasis. This lead to the unfortunate instance of converting attractive to ugly, when “attractive” was a key term in hypothesis H3 and H4.

There were a variety of synthetic figures with varying degrees of realism, beside 2D and 3D. Questions of about realism, strangeness or naturalness might have yielded interesting results.

Implications of Results

These results indicate that we are more sensitive to the best facial proportions as a character increases both in attractiveness and human likeness. These results also indicate that while the range of acceptable facial proportions narrows as a character is considered more attractive, the same relation does not exist for ratings of human likeness. However, it is important to repeat this experiment with greater stimulus control to confirm the results (e.g., by using the same person as a model for three-dimensional, two-dimensional and robotic stimuli).

Original facial proportions varied too much for the resultant best proportions to provide guidance on how characters should be designed for optimal effect.

CHAPTER SIX: CONCLUSION

Interaction between humans and synthetic characters such as androids, robots, and computer graphics characters is a fact of life in the developed world, where it is occurring with increasing frequency. The design of these characters should take the proportions of various facial features into careful consideration.

Limitations

A convenience sample was used and participants were self-selected. This may cause some issues with the generalizability of these results. However, this study had broad international participation, participants came from over twenty different countries. Participants were primarily recruited from individuals most likely to come into contact with synthetic character. The effect of participants' nationality, age, or education has not been evaluated.

Other issues regarding the range of stimuli, experimental design, and experimental control, included in the Discussion chapter, limit the generalizability of these results as well.

Further Research

Many aspects of this study could be repeated while addressing the issues detailed in the Discussion chapter regarding figure selection, proportion control, and other figure factors. A larger but better controlled pool of stimuli might produce more generalizable results and whether more humanlike proportions are preferred for nonhuman characters.

Additional research into the cause of the observed uncanny valley could be performed. Some causes are suggested by the results of this study. The attribute ugly was strongly correlated with creepy; this could suggest that figures rated as creepy violate aesthetics, or trigger disgust. There was also a negative correlation between the attributes alive and creepy, which when plotted showed a shallow peak of creepiness at the point participants slightly agreed that the figure was alive. A larger pool of stimuli also might be able to determine how tightly these two attributes are related. If participants are able to view attractive and unattractive figures of various levels or realism, that would help determine whether ugly is creepy or whether more factors are involved.

Summary

This study examined the relationships between sensitivity to best facial proportions, tolerance for acceptable facial proportions, and several characters attributes. It demonstrated significant correlations between the selection of best proportions to ratings of human likeness and attractiveness. It also demonstrated that while there is a significant correlation between acceptable ranges for facial proportions and ratings of attractiveness, such a relationship can not be established between acceptable range and human likeness.

An uncanny valley was also found when participants considered a face most creepy, not just before faces reached a perfect human resemblance, but when participants were most ambivalent about the human likeness of a face. Participants did not vary the best proportions of faces with indeterminate human likeness as much as they did faces that were perceived to be either humanlike or synthetic.

References

Abrosoft. (2006). Fantamorph 3 (version 3.6.1) [Computer software]. Abrosoft Co.

Adobe. (2006). Illustrator [Computer software]. Adobe Software.

Alam, M., & Dover, J. S. (2001). On beauty: Evolution, psychological considerations and surgical enhancement. *Archives of Dermatology*, *137*, 795-807.

AutoDesk. (2006). Autodesk 3ds max (Version 9.0) [Computer software]. San Rafeal, CA: AutoDesk, Inc.

Bailenson, J. N., Beall, A. C., Blascovich, J., Raimundo, M., & Weisbuch, M. (2001). Intelligent agents who wear your face: (u)sers' reactions to the virtual self. In A. de Antonio, R. Aylett, & D. Ballin (Eds.), *Third internation workshop on intelligent virtual agents* (Vol. 2190, p. 86-99). Berlin: Springer.

Costa, M., & Corassa, L. (2006). Aesthetic phenomena as supernormal stimuli: The case of eye, lip and lower-face size and roundness in artistic portraits. *Perception*, *35*, 229-246.

Cunningham, M. R. (1986). Measuring the physical in physical attractiveness: Quasi-experiments on the sociobiology of female facial beauty. *Journal of Personality and Social Psychology*, *50*(5), 25-935.

Cunningham, M. R., Roberts, A. R., Barbee, A. P., Druen, P. B., & Wu, C.-H. (1995). "Their ideas of beauty are on the whole the same as ours?" Consistency and variability in cross-cultural perception of female physical attractiveness. *Journal of Personality and Social Psychology*, *68*(2), 261-279.

Davis, K. (2000). *Multi-media robots as scientifically validated educational tools in formal school settings: The Million Dollar Machine®life skills enrichment program for US elementary school students.* (Paper presented at Workshop on Edutainment Robots 2000, Sankt Augustin, Germany)

DiSalvo, C. F., Gemperle, F., Forlizzi, J., & Kiesler, S. (2002). All robots are not created equal: The design and perception of humanoid robot heads. In *Proceedings of the Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques* (p. 321-328). London.

Drury, N. E. (2000). Beauty is only skin deep. *Journal of the Royal Society of Medicine*, *93*, 89-92.

Etcoff, N. (1999). *Survival of the prettiest: The science of beauty.* New York: Random House.

Evans, C. A., Viana, G., Anderson, N. K., & Giddon, D. B. (2005). Tolerance of deviations in eye and mouth position. *Orthodontics and Craniofacial Research*, *8*(2), 75-84.

Farkas, L. G. (Ed.). (1994). *Anthropometry of the head and face.* New York: Raven Press.

Galton, F. (1879). Composite portraits, made by combining those of many different persons into a single resultant figure. *The Journal of the Anthropological Institute of Great Britain and Ireland*, *8*, 132-144.

Giddon, D. B., Sconzo, R., Kinchen, J. A., & Evans, C. A. (1996). Quantative comparison of computerized discrete and animated profile preferences. *The Angle Orthodontist*, *66*(6), 441-448.

Goetz, J., Kiesler, S., & Powers, A. (2003). Matching robot appearance and behavior to tasks to improve human-robot cooperation. In *Robot and Human Interactive Communications. Proceedings of ROMAN 2003* (p. 55-60).

Golby, A. J., Gabrieli, J. D. E., Chiao, J. Y., & Eberhardt, J. L. (2001). Differential responses in the fusiform region to same-race and other-race faces. *Nature Neuroscience*, *4*(8), 845-850.

Goldstein, A. C., & Papageorge, J. (1980). Judgements of physical attractiveness in the absence of eye movements. *Bulletin of the Psychonomic Society*, *15*, 269-270.

Grammer, K., & Thornhill, R. (1994). Human (*Homo sapiens*) facial attractivenss and sexual selection: The role of symmetry and averageness. *Journal of Comparative Psychology*, *108*(3), 233-243.

Halberstadt, J. (2006). The generality and ultimate origins of the attractiveness of prototypes. *Personality and Social Psychology Review*, *10*(2), 166-183.

Hanson, D. (2006). Exploring the aesthetic range of humanoid robots [electronic resource]. In *Toward social mechanisms of android science, the 28th annual conference of the cognitive science society long seminar* (p. 16-20).

Hönekopp, J. (2006). Once more: Is beauty in the eye of the beholder? Relative contributions of private and shared taste to judgements of facial attractiveness [electronic version]. *Journal of Experimental Psychology Human Perception and Performance*, *12*(2), 199-209.

International Robotics. (n.d.). *The field of communication robots and robot-to-people behavioral psychology.* [Brochure]. New York.

Johnson, V. S., Hagel, R., Franklin, M., Fink, B., & Grammer, K. (2001). Male facial attractiveness: Evidence for hormone-mediated adaptive design. *Evolution and Human Behavior*, *22*, 251-267.

Jones, B. C., Little, A. C., & Perrett, D. I. (2004). When facial attractiveness is only skin deep. *Perception*, *33*, 569-576.

Jones, D. (1995). Sexual selection, physical attractiveness, and facial neoteny: Cross-cultural evidence and implications. *Current Anthropology*, *36*(5), 723-748.

Keating, C. F. (1985). Gender and the physiogomy of gender and attractiveness. *Social Psychology Quarterly*, *48*(1), 61-70.

Kowner, R. (1996). Facial assymetry and attractiveness judgement in developmental perspective. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(3), 662-675.

Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average [electronic version]. *Psychological Science*, *1*, 214-220.

Langlois, J. H., Roggman, L. A., Casey, R. J., Ritter, J. M., Rieser-Danner, L. A., & Jenkins, V. Y. (1987). Infant preferences for attractive faces: Rudiments of a stereotype [electronic version]. *Developmental Psychology*, *23*(3), 363-369.

Latto, R. (1995). The brain of the beholder. In R. Gregory, J. Harris, P. Heard, & D. Rose (Eds.), *The Artful Eye* (p. 60-94). Oxford: Oxford University Press.

MacDorman, K. F. (2006). Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley [electronic resource]. In *ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science.* Vancouver, Canada. (http://www.macdorman.com/kfm/writings/pubs/MacDorman2006SubjectiveRatings.pdf)

MacDorman, K. F., & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research [electronic version]. *Interactive Studes*, *7*(3), 297-337.

Mealey, L., Bridgstock, R., & Townsend, G. C. (1999). Symmetry and perceived facial attractiveness: A monzygotic co-twin comparison. *Journal of Personality and Social Psychology*, *76*(1), 151-158.

Mori, M. (1970). Bukimi no tani [the uncanny valley]. *Energy*, *7*, 33-35.

Olson, I. R., & Marshuetz, C. (2005). Facial attractiveness is appraised in a glance. *Emotion*, *5*(4), 498-502.

Penton-Voak, I. S., Jones, B. C., Little, A. C., Baker, S., Tiddeman, B., Burt, D. M., et al. (2001). Symmetry, sexual dimorphism in facial proportions and male facial attractiveness. *Proceedings Biological Sciences*, *268*(1476), 1617-1623.

Perrett, D. I., May, K. A., & Yoshikawa, S. (1994). Facial shapes and judgements of female attractiveness. *Nature*, *368*, 239-242.

Rhodes, G., Hickford, C., & Jeffery, L. (2000). Sex-typicality and attractiveness: Are supermale and superfemale faces super-attractive? *British Journal of Psychology*, *91*, 125-140.

Rhodes, G., Proffitt, F., Grady, J. M., & Sumich, A. (1998). Facial symmetry and the perception of beauty. *Psychonomic Bulletin & Review*, *5*(4), 659-669.

Rhodes, G., & Tremewan, T. (1996). Averageness, exageration and facial attractiveness. *Psychological Science*, *7*(2), 105-110.

Salvia, J., Sheare, J. B., & Algozzine, B. (1975). Facial attractiveness and personal-social development. *Journal of Abnormal Child Psychology*, *3*(3), 171-178.

Scheib, J. E., Gangstead, S. W., & Thornhill, R. (1999). Facital attractiveness, symmetry and cues of good genes. *Proceedings of the Royal Society of London. Biological sciences*, *266*, 1913-1917.

Scion. (2006). Scion image (Version 4.0.3) [Computer software]. Federick, MD: Scion Corporation.

Udry, J. R. (1965). Structural correlates of feminine beauty preferences in Britain and the United States: A comparison. *Sociology and Social Research*, *49*, 330-342.

Weinberg, L. (2004). Poser 5 (Version 5.0.4) [Computer software]. Santa Cruz, CA: Curious Labs.

APPENDIX A: Participant Recruitment e-mail

An e-mail was prepared and sent to individuals in the researcher's personal address book, as well as the address books of several colleagues. Additionally, this message was sent to mailings lists of IUPUI Informatics and New Media graduate students, New Media undergraduate students, and the Association for Computing Machinery's Computer Human Interaction mail list.

The text of the mail message is as follows:

> Greetings,
>
> You are invited to participate in a web-based study of perception of facial proportions. Please follow the link below, provide a small amount of information about yourself, and click on the participate link for the study titled Perception of facial proportions. There are four tasks in the study that should take a total of about 50 minutes to complete. You do not have to complete all four tasks in one sitting. You may wish to find a quiet place where you may complete the study without interruption.
>
> `http://www.theuncannyvalley.com`
>
> While you are there feel free to participate in any of the experiments on the site. Please forward this message to others who might be interested in participating.
>
> Sincerely,
>
> Robert Green

Graduate Student

School of Informatics @ Indianapolis

Indiana University

APPENDIX B: Participant Demographics

208 participants completed the study. Participants were asked to provide their gender, year of birth, number of years of eduction, country of birth and current country of residence. Participants were allowed to complete the study without providing demographic information if they were concerned about privacy.

207 participants reported gender. Of those 60.6% ($n = 126$) were male and 38.9% were female ($n = 81$).

The mean age of participants was 32.0 with a standard deviation of 10.27. The minimum and maximum ages were 17 and 79. Age is reported as the year 2007 minus the year of birth, their may be an error of 1 year based on the participant's birth date and the date they completed the study. Figure 14 is a histogram demonstrating the participant's ages.

Participants averaged 16.1 years of education (representing the completion of a bachelor's degree in the United States) with a standard deviation of 4.62. Reported years of eduction ranged from 2 to 33. It is possible that many participants reported the number of years of post-secondary eduction, but it is not possible to confirm that assumption. Figure 15 is a histogram demonstrating the participant's reported years of education.

The largest group of participants were born in the United States (31.3%, $n = 65$), followed by Indonesia (30.3%, $n = 63$) and Great Britain (12.0%, $n = 25$). A larger proportion of participants lived in the United States (39.9%, n = 83) followed again by Indonesia (30.3%, $n = 63$) and Great Britain (12.0%, $n = 25$). Twenty six

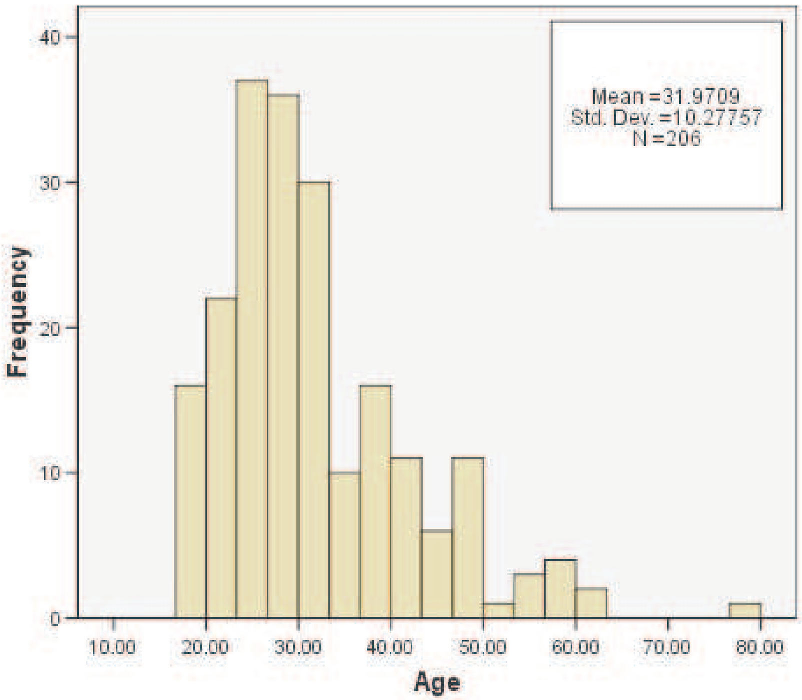

Figure 14: Participant ages

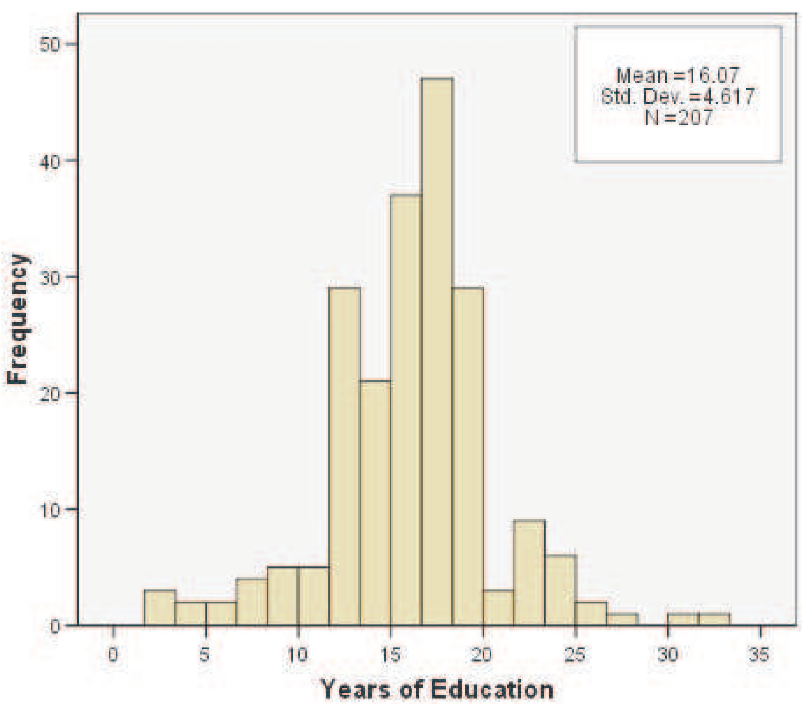

Figure 15: Participant years of education

countries were recorded as country of birth, and 20 as country of residence. 18.6% (n = 29) of participants did not live in their country of birth. A complete summary of participants by country of birth and residence is included in Table 17.

Table 17: Participant Countries of Birth and Residence

	Country of Birth		Country of Residence	
Country[1]	n	%	n	%
Albania	1	0.5	1	0.5
Austria	1	0.5	1	0.5
Bosnia and Herzegovina	1	0.5	–	–
Brazil	3	1.4	2	1.0
Canada	4	1.9	6	2.9
China	4	1.9	–	–
Finland	2	1.0	1	0.5
France	–	–	1	0.5
Germany	7	3.4	5	2.4
Great Britain	25	12.0	25	12.0
Greece	1	0.5		
India	1	0.5	–	–
Indonesia	63	30.3	63	30.3
Italy	4	1.9	3	1.4
Japan	5	2.4	4	1.9
Luxembourg	1	0.5	1	0.5
Malaysia	1	0.5	–	–
Mexico	1	0.5	–	–
Netherlands			1	0.5
New Zeeland	1	0.5	–	–
Phillipines	1	0.5	–	–
South Africa	1	0.5	1	0.5
Spain	3	1.4	3	1.4
Sweeden	2	1.0	3	1.4
Syria	–	–	1	0.5
Taiwan	7	3.4	1	0.5
Tunisia	1	0.5	–	–
Turkey	1	0.5	1	0.5
United States	65	31.3	83	39.9

[1] $n = 208$

APPENDIX C: Sources of Still Images

Original photographs for this study had to present a straight forward pose with the subject looking directly at the camera with nothing obscuring the subject's facial features. Many potential subjects were either looking away from the camera or had their facial partially obscured by hands, hair or props of some sort. Original artwork had to be created in a similar pose.

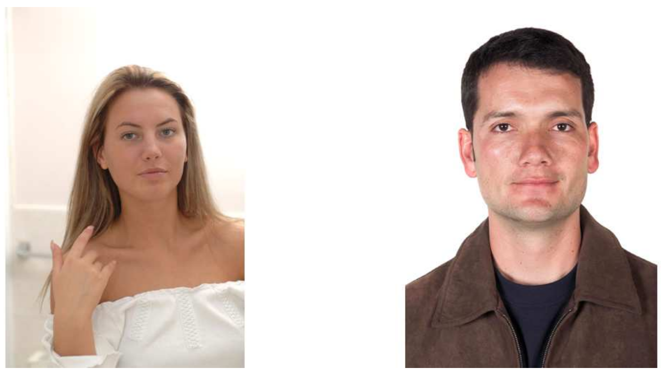

Figure 16: Original images: Human photographs

The two human photographs, Figures 16, were purchased from the website 123 Royalty Free, www.123rf.com. The female photograph was entitled "Model in mirror" and was photographed by EML Photography. The male photograph was entitled "Casual man portrait over white" and was photographed by Andres Rodriguez.

123 Royalty Free contains thousands of stock photographs in varying image resolutions. Print resolution photographs were purchased to allow a large enough image to edit and still provide a sharp image of the face only. Purchase of photographs from this site provides a non-exclusive license not only to use images, but to modify them before publication. Use of purchased photographs from this site also relieved

the researcher of the need to obtain model releases, and reduced costs as models did not have to be paid.

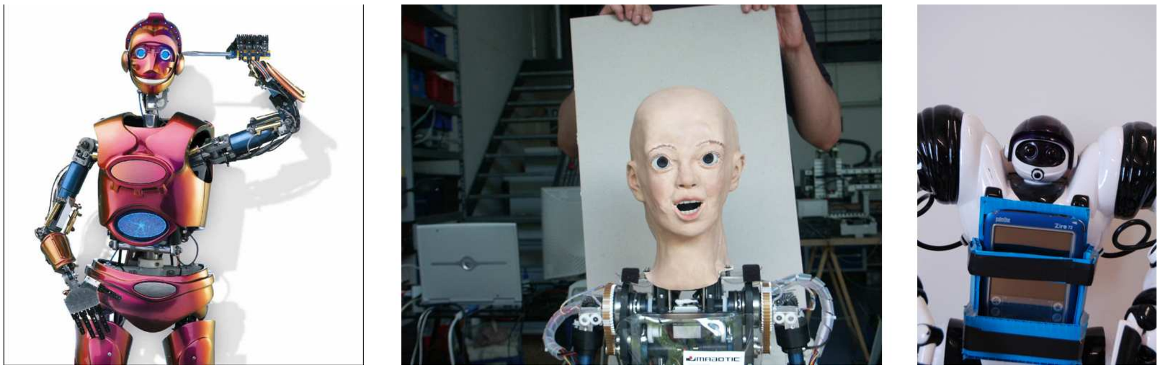

Figure 17: Original images: Robot photographs. (a) “Anthrobot”, (b) Barthoc, Jr., (c) Robosapien.

The search for robotic stimuli involved locating humanoid robots that had an adequate facial structure to manipulate in the second part of the study. The photograph of Anthrobot, Figure 17(a), was provided by Robert Doornick, CEO of International Robotics. Matthias Hackel of Mabotic Robotics and Automation provided the photograph of the rubber faced robot, Figure 17(b), this robot was called “Sue” during its development, but now is known as “Barthoc, Jr.” at the University of Bielefeld. Sara McGrath, of the University of Manitoba, provided photographs of their rendition of Robosapien, Figure 17(c). A complete listing of companies and researchers providing photographs and permission to use them in this study is detailed in Appendix D: Listing of Robots and Researchers.

The 2-D female character in Figure 18, was hand drawn by Taffy Green, and was scanned and enhanced using Photoshop. The 2-D male and robot characters in Figure 18 were created by Karl MacDorman, Ph.D. using Illustrator (Adobe, 2006).

The female 3-D character in Figure 19(a), named Masha, was developed for 3D

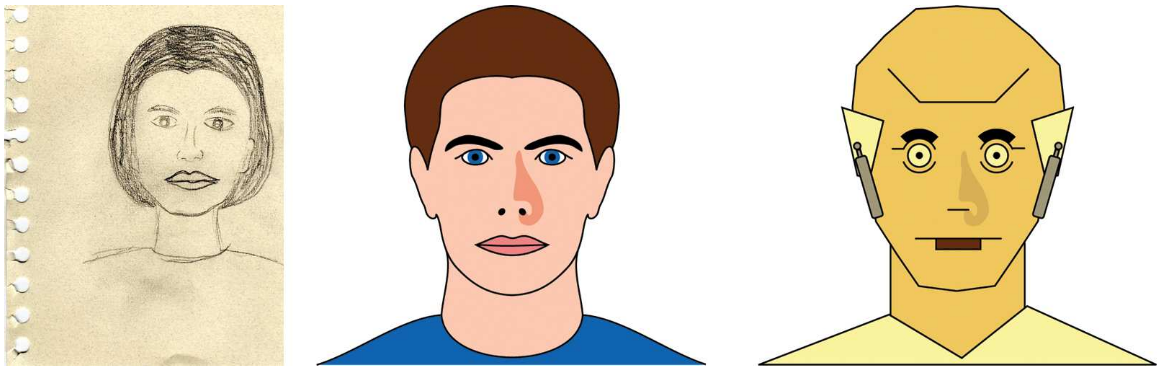

Figure 18: Original images: 2D characters

Figure 19: Original images: 3-D characters. (a) Masha, (b) Don, (c) Rustbot

Studio Max (AutoDesk, 2006) by Andrey Kravchenko, and was acquired through www.turbosquid.com. The male 3-D character in Figure 19(b) is a standard character with the Poser 5 (Weinberg, 2004) software package. The 3-D robot in Figure 19(c) was designed for Poser by Holger Hinzberg of Atomic Knights and was acquired through www.contentparadise.com.

APPENDIX D: Listing of Robots and Researchers

The following is a listing of the many robotics researchers and institutions that offered photographs of their robots, permission to use them in this research and encouragement to the researcher.

Atlas Robotics Jules Bamberger granted permission to use images of robots developed by Atlas Robotics in this study.

ATR Dr. Takahiro Miyashita provided photographs of Robovie III and permission to use them in this study.

University of Eindhoven Hu Jun provided images of the robot Tony, and granted permission to use them in this study.

Entertainment Technologie Michael H. Mosieur granted permission to use images of his robot located on the www.androidworkd.com.

Florida Robotics Fay Martin granted permission to use images from the Florida Robotics website in this study.

University of Frieburg Sven Benhke provided several photographs of Fritz, and granted permission to used them in this study.

International Robotics Robert Doornick provided several photographs and informational material about the robots at International Robots, as well as permission to use them in this study. Mr. Doornick offered much encouragement to the researcher. Antrobot was used as a stimulus.

Mabotic - Robotics & Automation Matthias Hackel provided many photographs of robots developed at Mabotic and granted permission to use them is this study. One of Mabotic's robots served as a stimulus in this study.

University of Manitoba Jacky Baltes granted permission to use images of research robots for use in this study. Sara McGrath generously photographed many robots in their lab. The stimulus Robosapien robot in the study was provided by Ms. McGrath.

Robomedia, Inc. Kent Davis provided a number of images and information on the robots at Robomedia, and granted permission to use them is this study.

Sarcos Jon Price provided images of Sarcoman and granted permission to use them in this study.

Tech Works Studios Valek X. Sykes provided images of robots under development at Tech Works Studios and granted permission to use them in this study.

APPENDIX E: Informed Consent

The following is the text of the on-line informed consent form approved by the Institutional Review Board and included as the first step of the research study. The IRB application was approved 12/02/2006 and expires 12/01/2007.

IUPUI and Clarian Study Information Sheet

This document describes a research study entitled Perception of Facial Proportions. You are invited to participate in this study. Please read this document and, if you want to participate, print it for your records, and click the consent link at the bottom of the page.

1. Participation: Participation is voluntary. You may refuse to participate at any time. No disadvantage will arise from refusing. Incomplete results are retained.

2. Purpose: The purpose of this study is to determine whether human beings vary in their sensitivity to the facial proportions of people, computer graphics characters, and robots.

3. Procedure: When you first sign up to participate, you register your email address and enter your gender, birth year, years of education, nationality at birth, and current country of residence. Once you register, you may sign in with the same email address without reentering this information. You will be shown images of people, robots, and 2D and 3D characters. For each image, you will adjust the image along one of four facial dimensions. You will then select the

acceptable range of values along that dimension. Finally, you will be shown the unmodified image and asked to rate statements about it. There is no wrong or right answer.

4. Time required: The procedure will take about 20 to 30 minutes to complete.

5. Age restriction: You must be at least 18 years old to participate.

6. Dissemination of results: Results may be reported in talks, documents, and publications of the principal investigator, experimenter, and their co-authors.

7. Confidentiality: Your personal information will not be identified or shared or used for another purpose. Reported results will not contain information that may be used to identify you. SSL encryption protects the privacy of your Internet session.

8. Risks: While we do not anticipate any risks from participating, you must stop participating and notify the principle investigator if at any time you feel your mental or physical well-being, personal values, or dignity is being harmed.

9. Benefits: The data gathered may contribute to an understanding of implicit associations about robots.

10. Compensation: You will not be paid for participating.

11. Questions: If you have any questions or concerns about the study, feel free to contact the principal investigator, Karl F. MacDorman, Associate Professor,

School of Informatics, Indiana University, 535 West Michigan Street, Indianapolis, IN 46202, or email him at the address given at macdorman.com ¿ self ¿ contact.. If you have any questions about your rights as a research participant, or unresolved problems, complaints, or concerns about a study, contact the IUPUI/Clarian Research Compliance Administration office at +1 317 278-3458 or +1 800 696-2949.

By clicking I consent you indicate the following: I am at least 18 years old; I understand and agree to the above conditions; my questions have been answered satisfactorily; and I have printed a copy of this study information sheet for my records.

I consent

I do not consent

Approval date: 02 Dec 2006 Expiry data: 01 Dec 2007

Figure 20 show the baseline movie created for the stimulus Robosapien. Each point can be individually positioned and the resulting movie will warp the image between the corresponding points in the two images.

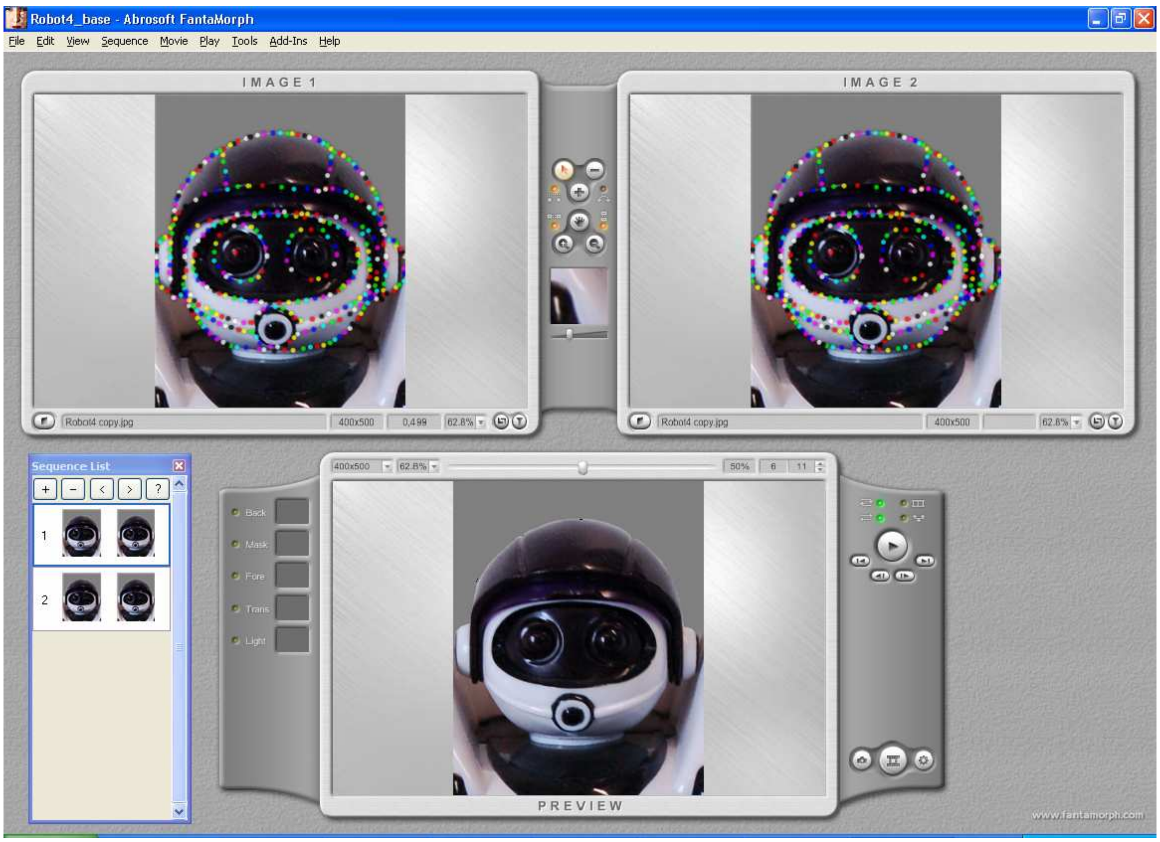

Figure 20: Base FantaMorph movie for robot Robosapien

VITA

Robert Dale Green

6337 Brokenhurst Road
Indianapolis, IN 46220
(317) 251-6649
rogreeniupui.edu

Education

Masters of Science, Human Computer Interaction, Indiana University Purdue University at Indianapolis, May 2008.

Bachelors of Arts, Computer Science, Indiana University, Bloomington, IN, May 1982.

Professional

11/2006-Present, ADP, Inc., Indianapolis, IN
Senior Software Developer.

12/2003 - 11/2006, Computer Horizons Corp., Indianapolis, IN
Software Consultant.

04/1999 - 11/2003, Kinexis Data Migration Services, Indianapolis, IN
Manager of Application Development.

02/1992 - 03/1999, Logicon, Inc., Indianapolis, IN
Senior Software Developer.

04/1987 - 01/1992, Computer Sciences Raytheon, Patrick AFB, FL
Senior Software Engineer.

04/1985 - 03/1987, Computer Sciences Corporation, Cape Canaveral AFS, FL
Software Analyst.

06/1982 - 03/1985, RCA International Services Corporation, Patrick AFB, FL
Programmer Analyst.

Presentations

Green, R.D., MacDorman, K.F., Huehls, P., Koch, C.T. (June 2007). The Uncanny Valley. Panel at New Media Consortium Summer Conference, Indianapolis, IN.