THE AUDIO/VISUAL MISMATCH AND THE UNCANNY VALLEY:
AN INVESTIGATION USING A MISMATCH IN THE HUMAN REALISM OF FACIAL AND VOCAL ASPECTS OF STIMULI

Kevin A. Szerszen Sr.

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

December 2010

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

Master's Thesis
Committee

Karl F. MacDorman, Ph.D., Chair

Anthony Faiola, Ph.D., M.F.A.

Davide Bolchini, Ph.D.

Amy Shirong Lu, Ph.D.

Dedicated to all of those who never give up, take personal responsibility, and believe they can accomplish anything they set their mind to.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I extend my sincerest thanks to my committee members for their time and feedback during this process: My committee chair, Dr. Karl MacDorman, for his wealth of knowledge, invaluable persistence, guidance that enabled me to develop an understanding of the subject, Dr. Anthony Faiola who supported me with his wisdom and patience whilst allowing me to work independently, Dr. Davide Bolchini for his positive encouragement and never-ending supply of ideas and suggestions, and Dr. Amy Shirong Lu for lending her expertise in ensuring the proper methods of analysis were applied to the appropriate research design. Their assistance in more ways than I could list has been invaluable. I would also like to thank the other members of Dr. MacDorman's research group: Chin-Chang Ho for his patience and assistance in analyzing the vast amount of data collected into interpretable results, Himalaya Patel for his assistance with establishing the experiment website and providing copious amounts of information at a moment's notice, and to all others who have been involved in both criticizing and improving this research. I cannot overstate my gratitude to Wade Mitchell, without whom I would never have been able to finish. I have not only acquired knowledge throughout this endeavor, but also a true friend.   I am also grateful for my supportive family, friends, and coworkers who put up with my shortcomings during this study. Most of all, I am expressly grateful to my wife and biggest supporter, Linda Szerszen for her infinite devotion and love.

ABSTRACT

Kevin A. Szerszen Sr.

THE AUDIO/VISUAL MISMATCH AND THE UNCANNY VALLEY:
AN INVESTIGATION USING A MISMATCH IN THE HUMAN REALISM OF
FACIAL AND VOCAL ASPECTS OF STIMULI

Empirical research on the uncanny valley has primarily been concerned with visual elements. The current study is intended to show how manipulating auditory variables of the stimuli affect participant's ratings. The focus of research is to investigate whether an uncanny valley effect occurs when humans are exposed to stimuli that have an incongruity between auditory and visual aspects. Participants were exposed to sets of stimuli which are both congruent and incongruent in their levels of audio/visual humanness. Explicit measures were used to explore if a mismatch in the human realism of facial and vocal aspects produces an uncanny valley effect and attempt to explain a possible cause of this effect. Results indicate that an uncanny valley effect occurs when humans are exposed to stimuli that have an incongruity between auditory and visual aspects.

CHAPTER ONE: INTRODUCTION

1.1 Introduction

Voice interface technology provides users with a familiar and intuitive interaction style. Auditory interaction has become commonplace for interaction with several technologies including smart house systems, robots, navigation systems, and any technology where a traditional interface such as a keyboard and mouse would be impractical or where the visual modality is otherwise occupied. As auditory interaction grows in popularity, it is important to understand the various social and psychological aspects of auditory interactions. While research has been conducted to understand aspects of interactions with technology incorporating a visual representation of the system such as an android or a computer-animated agent (MacDorman, 2005a; MacDorman, 2006; MacDorman, Green, et al., 2009), little has been done to apply this understanding to interactions that include voice.

1.2 Importance Understanding Auditory Interactions

Speech is a natural and familiar communication method that can be a powerful form of interaction in the design of computer information systems (Allen et al., 2001). It is important to understand the interaction of auditory communications from both a social and psychological point of view. When synthetic speech is combined with a synthetic agent, the mixed modalities of audio and vision can complicate perception. Androids and humanlike synthetic agents that use speech for communication could benefit from a better understanding of the effects of combining visual and auditory modalities. Persons who have undergone medical treatments such as a laryngectomy (partial or full removal of the voice box) as well as those who suffer debilitating illnesses like Parkinson's disease rely on synthesized speech aids to communicate. Technologist's increased understanding of the combination of the modalities could enhance social interaction for people with vocal impairments (Clements et al. 1997). Designers should understand the various aspects and implications for design. Failure to identify unease and discomfort could result in the development of a system that humans dislike interacting with. Furthermore, increasing vocal realism and personalization in speech synthesis devices could enhance social interaction for people with vocal impairments. A phenomenon known as the uncanny valley refers to the fact that a higher degree of human realism can elicit either a feeling of great comfort or a feeling of aversion. Therefore, understanding the uncanny valley becomes increasingly important as it applies to the creation of interactive systems that contain a human voice.

1.3 Intention of the Study

As technology improves, the humanness of both the visual and auditory aspects of synthetic agents also improves. An incongruity in these levels of humanness may reduce the human likeness of the synthetic agent despite increases in humanness of either the visual or auditory aspects. For example, an increase in the human likeness of the facial features of an android without a corresponding increase in the human likeness of the voice would result in incongruent levels of humanness for the android. This may cause the android to appear eerie and, therefore, less humanlike than it was before. This study intends to investigate if an uncanny valley effect occurs when the auditory and visual aspects of a stimulus are incongruent.

CHAPTER TWO: LITERATURE REVIEW

2.1 The Uncanny Valley

The uncanny valley, first proposed by Mori in 1970, posits that as a robot more closely resembles a human, the more familiar it becomes. However, as the robot's human likeness further increases, a point is reached at which subtle deviations from genuine human will cause a dip in familiarity and the robot will appear eerie or creepy (Mori, 1970). The uncanny valley has been examined to some extent in the context of robotics and computer animation. For example, research following Mori's premise focuses on visual aspects such as the look and movements of a robot or an animated character and the human response to deviations from what is expected of other humans in these aspects (Green et al., 2008; Ho et al., 2008; MacDorman, 2006; MacDorman, Green, et al., 2009). Several theories of the cause of the uncanny valley have been posited since Mori's paper 40 years ago. They include terror management theory and mortality salience, mate selection, pathogen avoidance, the violation of human norms, and sorites paradox.

Terror management theory and mortality salience (MacDorman, 2005b; MacDorman & Ishiguro, 2006) propose that an uncanny robot or computer generated (CG) character may elicit a fear of death for several possible reasons. Jerkiness of movement or movement in a manner impossible for a human might elicit fear of loss of bodily control. Realistic androids which are shown with pieces missing may elicit a fear of loss of limb. When reminded of their mortality, humans will initiate terror management which uses self-esteem and cultural world view as a method of protection from the emotional effects of reflecting on their inevitable death (Pyszczynski et al., 1999).

The hypothesis that mate selection may produce an uncanny valley effect is linked to biological aspects which might trigger a person to accept or reject a potential mate. Norms of beauty that apply to physical appearance and movement, acting as markers for fitness, fertility, and health, may be difficult to replicate in a synthetic agent. As a result, a synthetic agent may be subconsciously rejected based on human norms concerning mate selection (Green et al., 2008; Rhodes & Zebrowitz, 2002). This effect is more pronounced for synthetic agents designed as the opposite sex of the observer.

Pathogen avoidance has been examined through Rozin's theory of disgust (MacDorman, 2006; Green, et al., 2009; Rhodes & Zebrowitz, 2002). This theory involves men and women's use of disgust as a mechanism to protection themselves from others who are closely related and appear to be diseased. Rozin and Fallon (1987) claimed that disgust evolved as a cognitive mechanism to protect human beings from infection.

The hypothesis that the violation of human norms causes an uncanny valley effect is based on very humanlike synthetic agents such as androids or CG characters initially evoking similar expectations as those another human would elicit (MacDorman & Ishiguro, 2006). When the synthetic agent fails to satisfy these expectations, in an interaction context for example, this violation may induce an uncanny valley effect.

Sorites paradox as a cause of the uncanny valley effect is based on the hypothesis that the linking of two qualitatively different categories, for example, human and robot, creates a paradox which calls into question one or both of the original categories (Ramey, 2005). Cognitive dissonance is the term for the uncomfortable feeling that is caused by holding two contradictory ideas simultaneously (Festinger, 1957). Festinger (1956) first proposed the concept of cognitive dissonance in *When Prophecy Fails*, a study of a doomsday cult and their reaction to discovering that the world had in fact not ended. Contrary to the author's assumptions about the cult following the realization that their prophecy was incorrect, many members recruited new members rather than admit that their doctrine was flawed. Cognitive dissonance theory states that in situations of cognitive dissonance, people will experience negative arousal that they will then seek to reduce. Festinger and Carlsmith (1959) demonstrated the need to reduce cognitive dissonance in an experiment where participants were asked to complete a boring task. Some participants were paid one amount for completing the tasks and a second group was paid a much higher amount. The group that was paid less had to reconcile doing a boring task for less money. As theorized, the participants who were paid less concluded that they must have enjoyed the task more; thereby reducing their cognitive dissonance. Cognitive dissonance has since become one of the most widely-held theories of social psychology. In summary, an observer is exposed to stimuli that violate their expectations. Competing theories regarding the categorization of the stimuli are created, resulting in cognitive dissonance and a negative feeling. It is this negative arousal that supports the theory of cognitive dissonance as a cause of the uncanny valley effect.

Ramey (2005) posited that an uncanny valley is caused when two incongruent conceptual categories (e.g. human and robot) come together. Other authors have posited this type of paradox to explain the uncanny valley concerning computer generated characters in films and games (Plantec, 2007; Tinwell & Grimshaw, 2009). The expected congruence between audio and visual stimuli can set the stage for expectation violation, and when the unexpected occurs, it could create cognitive dissonance in the observer.

Some research has been done to examine ways to overcome the uncanny valley. Hanson and colleagues (2005) argue that the uncanny valley is more of an arbitrary concept, claiming that technological advancements, social aesthetics, and art will take android sciences past any such valley. Others have tried to determine where that line is in hopes of finding ways to improve interactions without approaching the valley (Newell, 2005).

The consequences of dipping into the uncanny valley can be significant for both designers and users. Science fiction and horror genres have both capitalized on the uncanny valley and the fear that is instilled by something that is close to, but not quite, human. It has been theorized that the uncanny valley triggers a neural "emergency alarm" of sorts that instinctually repulses humans (Hanson, 2003).

2.2 Audio and the Uncanny Valley

Although audio stimuli have not been directly investigated in the context of the uncanny valley, several studies have shown evidence of clear preferences among participants of certain voice aspects. In a study of intelligibility and the use of synthetic voice over telephone, Drager and colleagues (2004) found that participants preferred listening to, and assigned more positive attitudes towards, natural speech over synthetic speech. Even cutting-edge speech synthesis is still often difficult to understand. The Blizzard Challenge of 2008, a challenge to create synthetic voices, tested the intelligibility of the voice using the English word corpus with 438 participants. Results showed that of the voices created from the English corpus, only 34 participants were able to understand all the words, while 197 were able to "usually able to understand most of the words", and 170 participants found it "very hard to understand the words" (Karaiskos et al., 2008, table 46).

The naturalness of the synthetic voice is an important quality. Stevens and colleagues (2005) presented evidence that suggests there is a correlation between the preferences of listeners for a voice synthesis system and its perceived naturalness. Evidence has also shown that the naturalness of the voice, or lack thereof, may be understood in light of speech-accommodation theory (Street & Giles, 1982). The theory states that when talking to another person, it is natural to accommodate one's own speech to the situation, and with whom one is talking. The basis for this theory is that the behavior of a conversation partner is determined in part by the perception of the expression of values, attitudes, and intentions in the voice of the conversational other. For example, if one partner desires the social approval of the other or wishes to display solidarity, there will be a convergence between the speech styles. If one partner diverges from the other or maintains their original speech pattern, then the opposite is considered true. There is evidence that this type of behavior extends to human-computer interaction. Using an unnatural voice, the robotic style of the output is a mismatch that interferes in the expectation of represented humanness.

In a study of spoken enquiries over a telephone line giving location directions, Moore and Morris (1992) discovered that when an agent on the telephone used the agent's own voice either modified to sound robotic or without modification, the participants used much more concise language, a much smaller number of words, and fewer interactional turns when they thought they were speaking to a machine as opposed to a human. Bell and colleagues (2003) learned that participants adapt their rate of speech to that of the conversational other, even when that other is thought to be a machine using a synthesized voice. Both Coulston and colleagues (2002) and Darves and Oviatt (2002) found that when interacting with computer animated characters with synthesized speech voices as part of an educational software tool, children adapted the amplitude and duration of their speech to that produced by the animated character. Creak (1999) differentiated the purposes of communication between a human and a machine and a human and another human as follows: when interacting with a computer, the purpose is to find out or provide information. Therefore the language used is more compact so that the machine understands and extracts what is relevant. Interaction with a human is different in that the communication is at a more abstract level; expressing ideas that are extrapolated and understood by the other human being. As such, an unnatural voice could be an obstacle to having a human-human interaction style. Stern and colleagues (1999) observed that participants preferred natural speech over synthetic speech, perceiving the speaker with the synthetic voice as less truthful, knowledgeable, and involved in the interaction. Higher quality synthetic voices were rated closer to the natural voice stimuli, providing evidence that the quality of the synthetic voice is correlated to the positivity of a listener's attitudes towards the agent (Stern, 2008; Stern et al., 2006).

2.3 Auditory/Visual Mismatch

Humans are very good at quickly categorizing elements of their environment. This is especially true when applied to other humans. Infants as young as 3.5 months have been shown to be able to discern matching faces and voices (Brookes et al., 2001). There is evidence that infants only a few hours old are capable of detecting matching audio-visual events (Morrongiello et al., 1998). This ability at such a young age is important for identifying caregivers. The human mind is constantly formulating new hypotheses based on the perception of the environment (Slater, 2002). Based on factors such as past experience and expectations, humans have at least a basic idea of what to expect out of the various elements of their environment.

A humanlike appearance should cause observers to categorize the agent as human but following this categorization, if the observed agent were to violate human norms, this would cause a mismatch and thus, cognitive dissonance. Evidence of this has been provided in previous research concerning motion, facial proportions, and image quality (Green et al., 2008; MacDorman, 2006; MacDorman et al., 2009; Matsui et al., 2005). Little research has been done on the mismatch between voice and human likeness. Research has shown that auditory stimuli in the context of ambient sound can contribute to uncanniness in film (Anderson, 1996; Spadoni, 2003; Spadoni, 2007) and in video games (Grimshaw, 2008; Kromand, 2008; Tinwell & Grimshaw, 2010). What has not been adequately addressed is the mismatch between the visual and auditory modalities as it relates to an agent, more specifically, research that can answer if the combination of the human photorealism and voice quality of an agent can create a mismatch, which elicits an uncanny valley effect.

It is reasonable to consider that the more humanlike the voice of the agent, the more anthropomorphized the agent is. While this can help to increase social presence, when expectations are violated, cognitive dissonance would cause the agent to enter the uncanny valley. Speech is the most common of human interaction media. People create models of others based on what is most familiar (Gordon, 1986), and speech interaction is very familiar. In vocal interactions, it is likely that the user's mental model of the synthetic participant will be anthropomorphic because human-human vocal interaction is most familiar. Grimshaw (2009) proposed that the uncanny valley was just as applicable to voice as it is to the look and movement of the synthetic other. There has been little direct research conducted to validate this hypothesis. Tinwell and Grimshaw (2010) provide evidence that when an agent's voice does not match expectations brought about by visual elements, participants find the agents strange. As part of their study, Tinwell and colleagues (2010) asked participants to rate the human likeness of the agent's voice and to choose one of five statements that best described the voice. These statements were (1) the voice is slow, (2) the voice is monotone, (3) the voice is of the wrong pitch/intonation, (4) the voice sounds as if it belongs to the character, or (5) none of the above. Participants were also asked to rate the quality of voice synchronization. Unfortunately, this study did not directly examine a mismatch in the human likeness of the voice and visual aspects of the agent, and the use of the five statements concerning the voice appears to be leading by restricting the responses of the participants. However, studies examining similar phenomena may be helpful in theorizing possible results.

One example of speech causing an uncanny valley effect comes from speech dubbed into another language, often seen in television and feature films. In these cases, a human actor is speaking, moving, and gesturing, yet the speech being heard by the audience is in another language and therefore incongruent. No direct research has been done on the uncanny valley and dubbed speech, but evidence from a 1988 survey of British television viewers showed that 51% of viewers preferred dubbed speech and 32% preferred subtitling. Of those that preferred subtitling, 42% simply dislike dubbed programs, 26% felt subtitling was easy to follow, and 20% liked hearing the original soundtrack (Kilborn, 1993). Koolstra and colleagues (2002) provide evidence that young children exposed to both a subtitled and dubbed version of the same recording preferred the subtitled version even though they did not have sufficient reading skills to process the subtitles as quickly as necessary.

Incongruent auditory and visual signals are not limited to dubbing. Humans use many cues to identify congruent voice and speakers including location in space (Broadbent, 1954), frequency range (Bregman & Pinker, 1978), and voice pitch and gender (Broadbent, 1952). From a very early age, children adopt audiovisual cues to help distinguish audio streams. Even though infants have little ability to lip-read (Massaro, 1987), they are adept at noticing synchronous correlations between sight and sound (Bahrick, 1987; Lewkowicz, 1986; Lewkowicz & Lickliter, 1994; Meltzoff & Borton, 1979). In an experiment with 30 infants aged 7.5 months, Hollich and colleagues (2005) showed video of a female speaking a passage designed to garner the attention of infants. They also played a male voice of the same amplitude speaking the same passage. Results of this study demonstrated that the infants focused more of their attention toward the congruent source (e.g., female speaker/female voice) regardless of gender of the participant or previous gender familiarization.

Some studies have demonstrated the importance of congruency between auditory and visual aspects of synthetic agents by providing evidence that the attitude of an interaction partner is influenced by this congruency. Crabtree and colleagues (1990) provide evidence that participants had a preference for gender-appropriate and age-appropriate voices. O'Keefe and colleagues (1998) showed that intelligent and socially appropriate voices were preferred by participants.

Multiple studies of cross-modality between vision and other modalities have repeatedly exhibited a strong effect of visual signals on the other modalities; consistent with the commonsense notion that humans are primarily a vision-dominated animal (Stein et al., 1996). Although the visual modality is often dominant, other modalities can affect vision. For example, the addition of an auditory cue has been shown to increase the intensity of a visual cue (Shimojo & Shams, 2001). Studies of communication using both visual and auditory cues indicated that the visual channel has the most impact on affect (Ambady & Rosenthal, 1992; Mehrabian, 1972). A phenomenon known as the McGurk Effect also provides evidence of the importance of the visual channel. McGurk and MacDonald (1976) observed that when presented with video of a face making the motions for one sound accompanied by audio of a second sound, participants would hear either the sound the face was making (with visual cuing overriding auditory cuing) or a third sound which was a combination of the visual and auditory stimuli. For example, when auditorily presented the syllable /ba/ and visually presented a videotaped face articulating /ga/, the auditory perception of the participants heard either /da/ or /ga/ (McGurk and MacDonald, 1976; Skipper et al., 2007). This illustrates the impact that visual aspects of stimuli have on participants. It could be predicted that levels of humanness will be affected in a similar manner. As a result, visual elements would take precedence over audio elements in stimuli that are both visual and auditory.

If both vocal and visual elements create expectations in the mind of the user, a violation of either expectation should produce cognitive dissonance. Along with this cognitive dissonance, there should be a feeling of unease which could be an uncanny valley effect. If these assumptions are correct, this effect should be measurable using self-report measures guided by previous uncanny valley studies. This could only be discerned if the level of humanness of the stimuli is also predictable.

2.4 Presence

The human mind's quick categorization can lead to false identifications and therefore, false expectations. When events occur that do not match our hypothesis, they can trigger a switch of hypothesis that Slater (2002) terms a break in presence (BIP). When experiencing conditions that can elicit an uncanny valley, the brain is choosing between two different hypotheses: *this is a human being* versus *this is not a human being*. Existing methods of studying and measuring presence (Slater & Steed, 2000; Slater, Brogni & Steed, 2003) and how they relate to cognitive dissonance theory might therefore be relevant to understanding the uncanny valley.

Social presence is integral to the idea of breaking presence. Social presence, as defined by Short and colleagues (1976), is "the degree of salience of the other person in the interaction and the consequent salience of the interpersonal relationships" (p. 65). Social presence has been used in previous research as the underpinning of human interaction through telecommunication media (Rice, 1993; Short et al., 1976; Walther, 1996). When that telecommunication medium allows the interactions to take on a more humanlike style, humans are prone to *automatic social responsiveness*. This is the tendency for people to automatically respond socially to entities that look or behave like humans (Reeves & Nass, 1996).

Anthropomorphism has been shown to play a role in the amount and type of presence in human-computer interactions. The more humanlike a virtual agent is the more users will treat the agent as a social other. Research supports the possibility that this increase in social presence is due to users' belief that a more anthropomorphized agent is more human and therefore, more of a social other (Koda, 1996; Slater & Steed, 2000; Taylor, 2002; Turkle, 1995; Wexelblat, 1997). Much of the interactive agent design work has been done with a goal towards increasing human likeness (Isla & Blumberg, 2002). The unintended side effect of increased human likeness could be that this sets higher expectations from the user which, if not met, could result in unintentional transition into the uncanny (Hindmarsh, et al., 2001; Slater & Steed, 2000).

2.5 Self-Report Measures

When measuring the uncanny valley effect, it becomes clear that there are two issues to measure, affect and humanness. If the predictions of the uncanny valley are correct, as the humanness of the stimuli increases, the affect should be more positive until a sudden drop in both humanness and affect occurs, followed by a rise in both humanness and affect. To measure the humanness of the stimuli, this study uses the humanness scale of Ho and MacDorman (2010) which showed high internal reliability (Cronbach's $\alpha = .92$).

Social psychology research has consistently shown warmth and competence as two universal dimensions of human social cognition for the initial social perception of positive and negative affect (Fiske et al., 2002). Using a series of semantic differentials denoting traits, Asch (1946) found "striking and consistent" differences between affect when using *warm* versus *cold* as descriptive terms while studying how personality impressions are formed. The two dimensions discovered by Asch using semantic differentials became a starting point for many other researchers, eventually taking the forms we know today as *warmth* and *competence*. Rosenberg and colleagues (1968) built on the work of Asch and also found two primary dimensions when forming impressions which they labeled *social* and *intellectual*. The social dimension shares many traits with *warmth* and in fact, socially desirable words clustered around *warm* and socially undesirable words clustered around *cold* when plotted on an axis. Though both of these dimensions emerge at the time of first exposure to a social other, research has shown that judgments of *warmth* are primary, emerging first and carrying more weight in affective and behavioral reactions (Fiske et al., 2002). Wojciszke and colleagues (1998) provide evidence of *warmth* and *competency* accounting for 82% of the variance of perceptions of everyday social behaviors. The original measures of *warmth* and *competence* were not designed as semantic differentials. Recently these measures have been converted to semantic differentials and show high internal reliability (Mitchell et al., in press).

Finally, a measurement of the uncanny valley must include a measure of eeriness. As stimuli cross the threshold of the uncanny valley, affect (*warmth* and *competence*) should decrease while *eeriness* should increase. This study again turns to Ho and MacDorman (2010) for their *eeriness* scale which through a series of experimental refinement showed a relatively high internal reliability (Cronbach's $\alpha = .74$). This scale has been reversed and renamed *reassurance* for the purposes of this study.

2.6 Research Question and Hypotheses

RQ: Does an uncanny valley occur when humans are exposed to stimuli that have an incongruity between auditory and visual aspects?

Previous work which has provided evidence for the dominance of the visual modality predicts that the level of humanness with be reliant first on visual elements and then on audio elements. Thus, H1 can predict the order of humanness of the stimuli necessary to produce an uncanny valley effect (Ambady & Rosenthal, 1992; McGurk & MacDonald 1976; Mehrabian, 1968; Mehrabian, 2007; Shimojo & Shams, 2001; Stein et al., 1996). The Hypotheses are as follows:

H1: Because of the predominance of the visual aspects of the stimuli, the *humanness* measure will increase in the following order: *Robot Figure* - *Synthetic Voice*, *Robot Figure* - *Human Voice*, *Human Figure* - *Synthetic Voice*, and *Human Figure* - *Human Voice*.

If an uncanny valley effect does exist, then the results of the *reassurance* measure should produce similar results to previous uncanny valley studies. Thus, the following prediction can be made.

H2: The *reassurance* measure will be higher for congruent auditory-visual stimuli (*Human Figure* - *Human Voice* and *Robot Figure* - *Synthetic Voice*) compared to incongruent auditory-visual stimuli (*Human Figure* - *Synthetic Voice* and *Robot Figure* - *Human Voice*).

CHAPTER THREE: METHODOLOGY

3.1 Participants

Participants were recruited from a random selection of undergraduate students of a Midwestern university via email on April 14, 2010. Of 48 participants, 28 (58.3%) were female, 20 (41.7%) were male, aged 19 to 44 years with a mean age of 21.19 years (*SD* = 3.67), 48 (100%) listed their country of birth as the United States, 48 (100%) were current residents of the United States, and the mean education level was 14.52 years (*SD* = 2.51). Data from participants that did not complete the questionnaire was not retained. There were no significant differences among the data reported in this paper by gender or age.

3.2 Materials

Each participant was invited to take part in the experiment by completing a task on a website[1]. The task was to observe a series of videos while providing ratings of explicit measures. A total of four possible stimulus conditions were constructed: *Human Figure* - *Human Voice*, *Human Figure* - *Synthetic Voice*, *Robot Figure* - *Human Voice*, and *Robot Figure* - *Synthetic Voice*. A single video clip, 480 pixels wide by 360 pixels high with an aspect ratio of 4:3, was designed to represent each of the four stimuli conditions. Each video contained a figure reciting several four neutral, two-word phrases (Table 3.1) that were 14 seconds in length. The four phrases were derived from a list of words in a previous study that were rated neutral in pleasantness by male and female college students (Mitchell et al., in press).

Table 3.1: Four neutral, two-word phrases for audio stimuli

Vacuum Cleaner
Picket Fence
Cardboard Box
Television Set

[1] http://sandbox.informatics.iupui.edu/~keszersz/

The four separate stimulus conditions that were developed are represented in Table 3.2.

Table 3.2: Descriptions of the four stimuli conditions

Figure	Voice	Description
Human	Human	Human male actor with his real voice.
Human	Synthetic	Human male actor with a synthetic audio track created with CoolSpeech 5.0.
Robot	Human	Robot actor with an audio track from the Human Figure Human Voice condition.
Robot	Synthetic	Robot actor with a synthetic audio track created with CoolSpeech 5.0.

A series of affective scales developed and refined in previous studies were used to measure the existence of the uncanny valley effect. *Warmth* and *competence* scales were modified from stereotype research (Fiske et al., 2002; Mitchell et al., in press) and the *humanness* and *reassurance* scales demonstrate high internal reliability (Ho & MacDorman, K. F., 2010). The scales were presented to participants as a series of semantic differential items ranging in value from -3 to 3 with 0 as neutral (Table 4.1).

3.3 Procedures

The experiment had a within subjects repeated measures research design with four different conditions. Participants were exposed to the four video stimuli conditions (Table 3.2) in random order. Following each video condition and before moving to the next, participants completed a series of semantic differential scales designed to measure feelings toward the figures in the videos using the *warmth*, *competence*, *humanness*, and *reassurance* scales (Table 4.1). Each video was presented in a continuous loop, with a brief pause between two-word phrases, while participants rated the figure featured in each video clip. Additionally each figure recited each of the four two-word phrases in the same order. In order to counterbalance ordering affects, the semantic differential items were also presented in random order as well as having the anchors of the differential items randomly transposed.

3.4 Data Analysis

The independent variables are the stimulus conditions. The first independent variable is human photorealism, which is defined as either a real human figure or a robot figure. The second independent variable is voice quality, which is defined as either human (i.e., natural) or synthetic. The dependent variables are the participants' ratings of each condition during the explicit measures task.

The data collected from the *warmth*, *competence*, *humanness*, and *reassurance* scales were analyzed by finding the mean and standard deviation of responses for each video stimulus. This was calculated by averaging the results from each semantic differential scale with respect to the current stimulus condition. The data was also tested for internal reliability using Cronbach's α. A two-way repeated measures analysis of variance (ANOVA) was also conducted to compare the effect of the stimulus condition by gender and condition.

CHAPTER FOUR: RESULTS

The results are reported in three main sections. Section 4.1 displays the internal reliability of the semantic differential scales used during the experiment. Section 4.2 presents the two-way repeated measures ANOVA results of the different stimuli conditions selected for analysis and discusses the participants preference collected from the self-reporting measures. Section 4.3 discusses the correlations among the semantic differentials scales for each stimuli condition selected for analysis.

4.1 Internal Reliability

Table 4.1: Semantic differential scales and Reliability (Cronbach's α)

Condition	HFHV	HFSV	RFHV	RFSV
Humanness Scale	α = .83	α = .81	α = .70	α = .77
Human-made–Human				
Artificial–Natural				
Without Definite Lifespan–Mortal				
Mechanical Movement–Biological Movement				
Inanimate–Living				
Warmth Scale	α = .88	α = .82	α = .72	α = .83
Cold-hearted–Warm-hearted				
Hostile–Friendly				
Spiteful–Well-intentioned				
Ill-tempered–Good-natured				
Grumpy–Cheerful				
Competence Scale	α = .85	α = .80	α = .79	α = .80
Incompetent–Competent				
Helpless–Capable				
Inefficient–Efficient				
Stupid–Intelligent				
Clumsy–Skillful				
Reassurance Scale	α = .80	α = .79	α = .79	α = .70
Usual–Eerie				
Numbing–Freaky				
Ordinary–Supernatural				
Uninspiring–Spine-tingling				
Boring–Shocking				

Predictable–Thrilling
Bland–Uncanny
Unemotional–Hair-raising

4.2 Assumption of Normality

A two-way repeated measures analysis of variance (ANOVA) was conducted to examine how participants perceived each condition. To verify the normality of the data histograms and Q-Q plots of the residuals for each of the semantic differential scales were plotted. Each histogram and normal probability plot of the residuals demonstrates a good approximation to normality by the data points reasonably following a normal curve (Fig. 4.1) and following the line (Fig. 4.2) respectively.

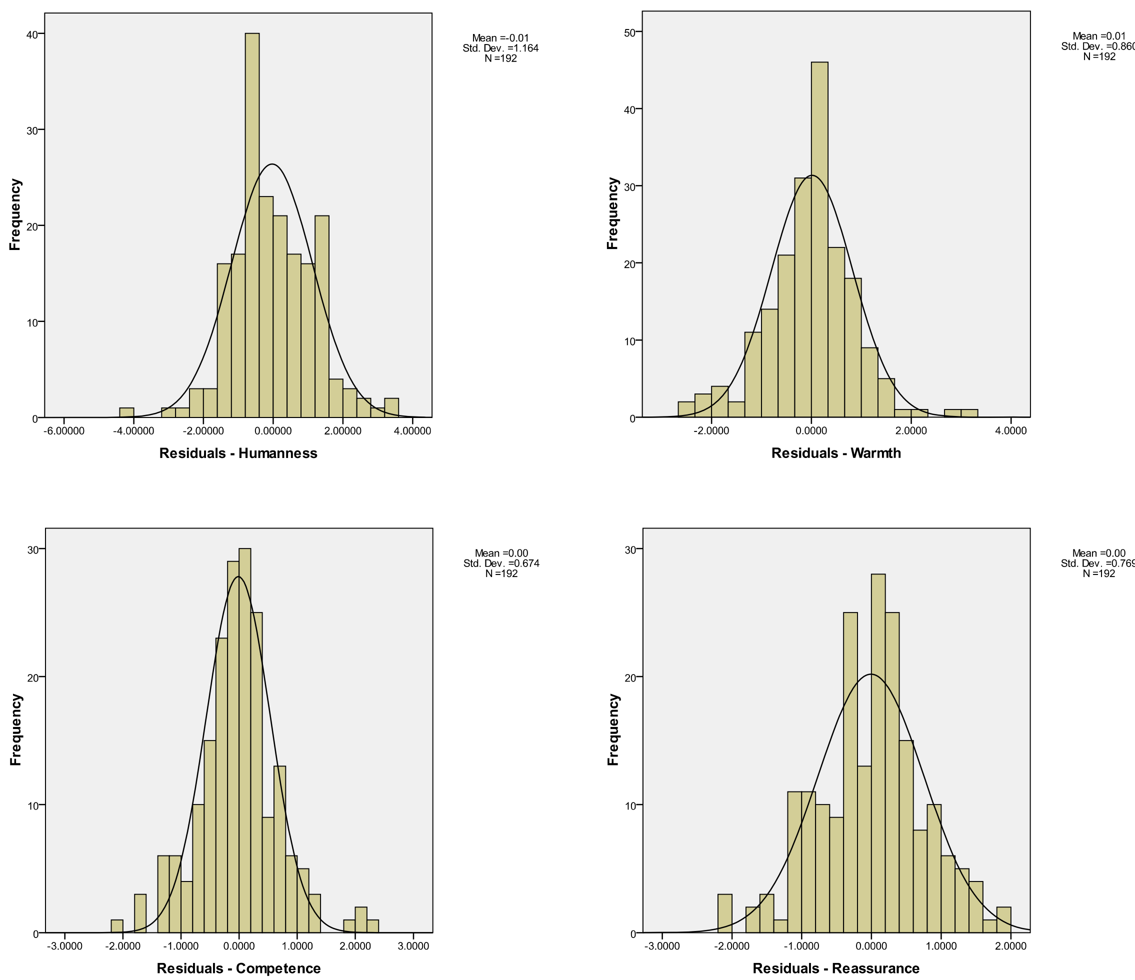

Figure 4.1. Mean values of *humanness*, *warmth*, *competence*, and *reassurance* residuals.

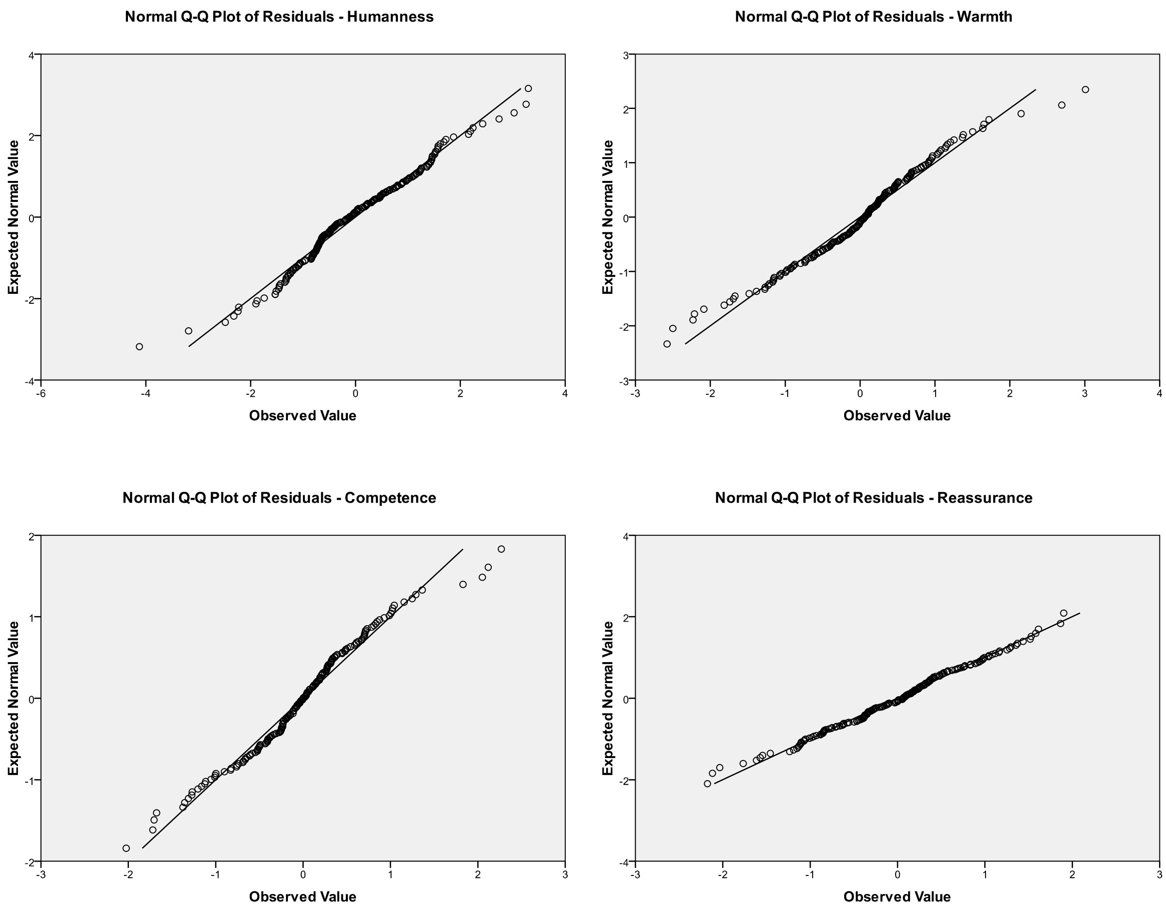

Figure 4.2. Q-Q plots of estimated normal values versus observed values of *humanness*, *warmth*, *competence*, and *reassurance* residuals.

4.3 Assumption of Homogeneity of Variance

To verify the homogeneity of variance of the data box plots of the residuals for each of the semantic differential scales were plotted. Each box plot of the residuals demonstrates that there were no statistically significant differences between group means.

4.4 Assumption of Sphericity

This experiment employs a 2 x 2 factorial design where there are two independent variables and each variable has two levels. At least three factors are needed for sphericity to be an issue; therefore since there are only two levels within each factor, the assumption of sphericity is met.

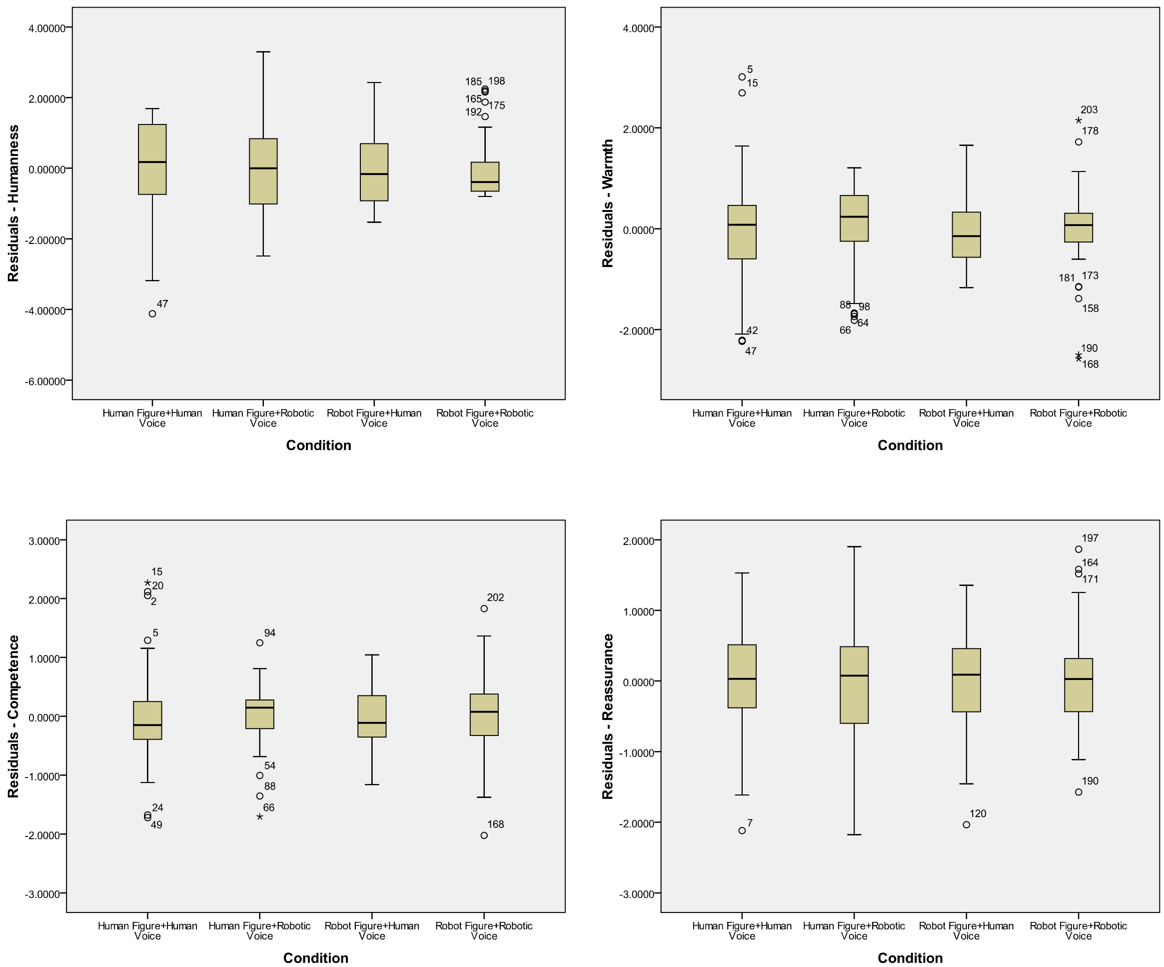

Figure 4.3. Box plots which represent the mean values of *humanness*, *warmth*, *competence*, and *reassurance* residuals.

4.5 Two-Way Repeated Measures ANOVA

Two-way repeated measures ANOVA tests were conducted to analyze the data from the semantic differentials (Table 4.1). Two-way repeated measures ANOVA tests were also conducted for grouped differences in the stimuli conditions by factors such as human photorealism (*Human Figure*, *Robot Figure*), voice quality (*Human Voice*, *Synthetic Voice*) *Matched* (*Human Figure – Human Voice* and *Robot Figure – Synthetic Voice*) and *Mismatched* (*Robot Figure – Human Voice* and *Human Figure – Synthetic Voice*) conditions. There were no significant differences among participants grouped by gender or age.

There was a significant main effect for the voice quality factor with respect to *humanness*, $F(1, 47) = 110.15$, $p < .001$. The effect size is .07. Conditions which contained a human voice were rated higher than conditions that had a synthetic voice. There was also a significant main effect for the human photorealism factor $F(1, 47) = 75.94$, $p < .001$. The effect size is .06. Conditions which contained a human figure were perceived as more *human* than conditions with the robot figure. There was also a significant interaction (voice*figure), $F(1, 47) = 18.65$, $p < .001$ (Figure 4.4), such that the *Human Figure – Human Voice* condition was rated as most *human* ($M = 1.39$, $SD = 1.39$), the *Human Figure – Synthetic Voice* condition next ($M = -.48$, $SD = 1.30$), followed by the *Robot Figure – Human Voice* condition ($M = -1.34$, $SD = .94$), then the *Robot Figure – Synthetic Voice* condition as least *human* ($M = -2.09$, $SD = .90$) (Figure 4.5). The effect size is .03.

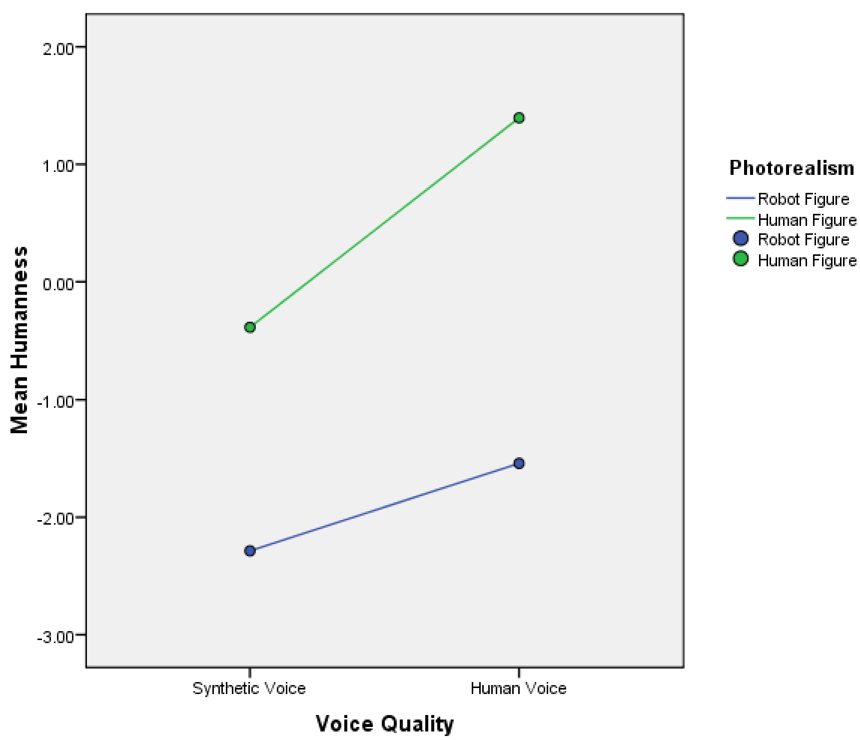

Figure 4.4. Main effect of human photorealism and voice quality factors in *humanness.*

There was a significant main effect for the voice quality factor with respect to *warmth*, $F(1, 47) = 27.62$, $p < .001$. The effect size is .37. Again, conditions that contained a human voice were rated higher than conditions that had a synthetic voice. There was also a significant main effect for the human photorealism factor $F(1, 47) = 75.94$, $p < .001$. The effect size is .20. Contrary to the *humanness* ratings, conditions that contained a robot figure were perceived as more *warm* than conditions containing the human figure. There was not a significant interaction (Figure 4.6). As Figure 4.5 indicates, the *Robot Figure – Human Voice* condition was rated the

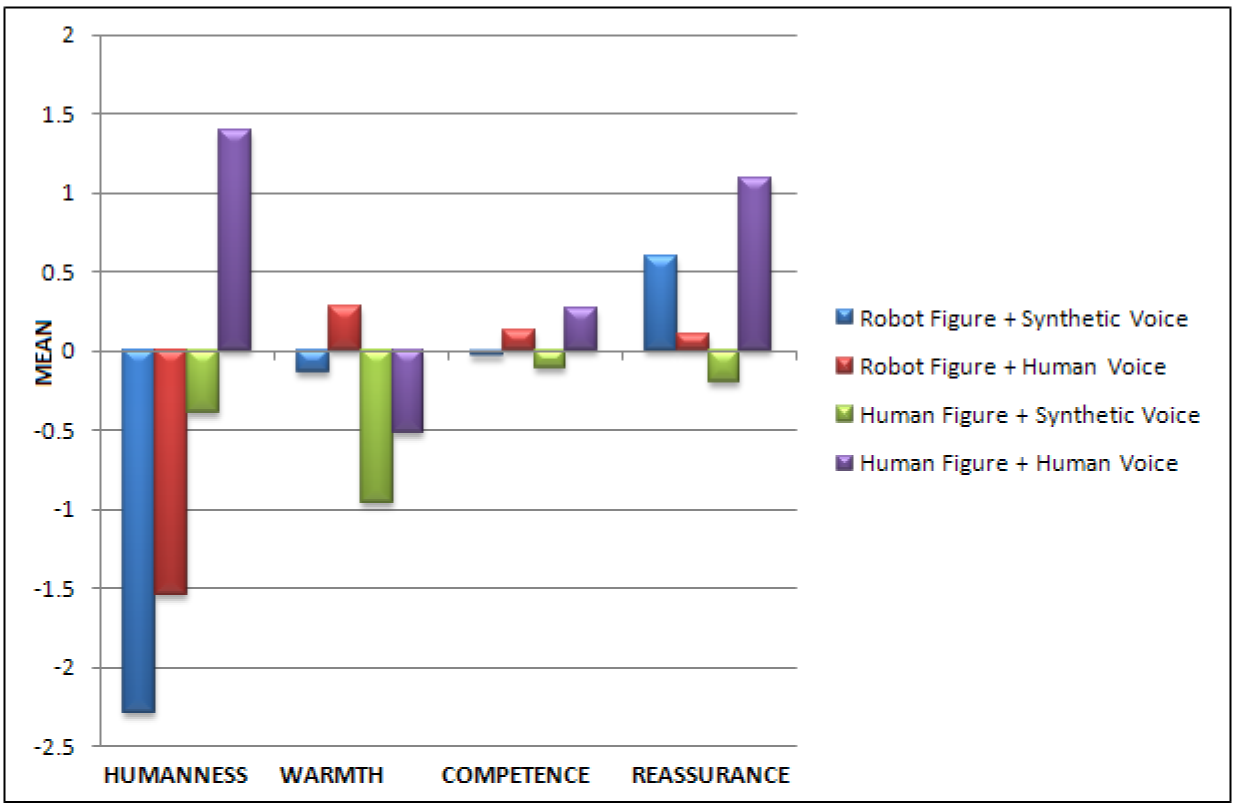

Figure 4.5. Mean values of *humanness*, *warmth*, *competence*, and *reassurance* differentials per condition (*Robot Figure – Synthetic Voice*, *Robot Figure – Human Voice*, *Human Figure – Synthetic Voice*, *Human Figure – Human Voice*).

most *warm* (M = .28, SD = .75), the *Robot Figure – Synthetic Voice* condition next (M = -.14, SD = .91), followed by the *Human Figure – Human Voice* condition (M = -.51, SD = 1.07), and then the *Human Figure – Synthetic Voice* condition as least *warm* (M = -.96, SD = .91).

There was not a significant main effect for the voice quality factor with respect to *competence*. However there was a significant main effect for the human photorealism factor $F(1, 47) = 4.46$, $p < .05$. The effect size is .09. There was not a significant interaction. The *Human Figure – Human Voice* condition was rated the most *competent* ($M = .27$, $SD = 1.06$), the *Robot Figure – Human Voice* condition next ($M = .13$, $SD = .98$), followed by the *Robot Figure – Synthetic Voice* condition ($M = -.03$, $SD = 1.18$), and then the *Human Figure – Synthetic Voice* condition as least *competent* ($M = -.12$, $SD = .98$).

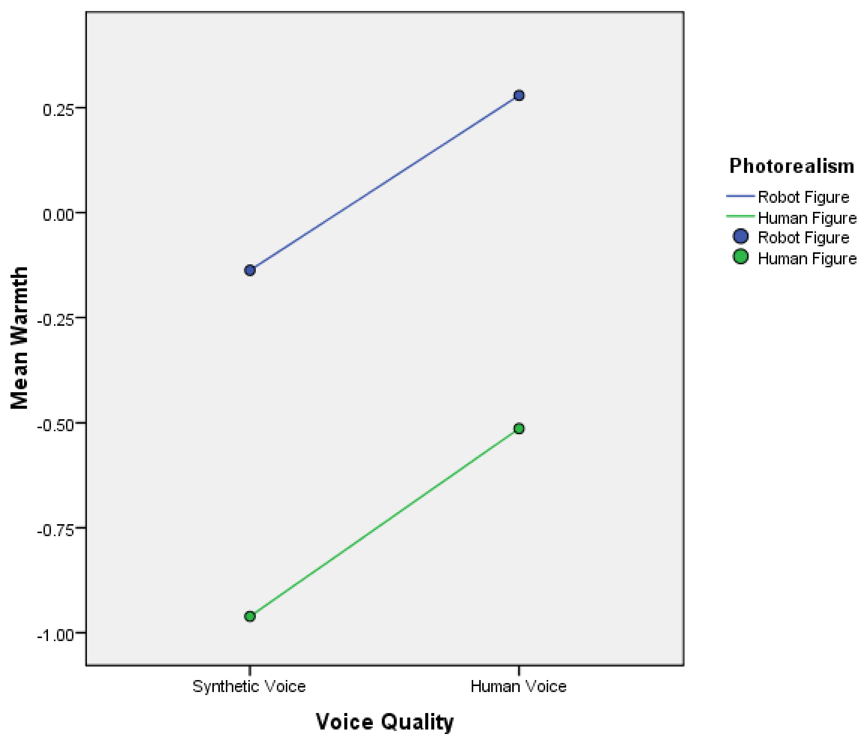

Figure 4.6. Main effect of human photorealism and voice quality factors in *warmth.*

There was not a significant main effect for the voice quality factor with respect to *reassurance*. However there was a significant main effect for the human photorealism factor $F(1, 47) = 13.28$, $p < .01$. The effect size is .22. There was also a significant interaction (voice*figure), $F(1, 47) = 36.51$, $p < .001$ (Figure 4.7), such that the *Human Figure – Human Voice* condition was rated the most *reassuring* ($M = 1.10$, $SD = .98$), the *Robot Figure –*

Synthetic Voice condition next (*M* = .60, *SD* = .91), followed by the *Robot Figure – Human Voice* condition (*M* = .10, *SD* = 1.02), and then the *Human Figure – Synthetic Voice* condition as least *reassuring* (*M* = -.19, *SD* = 1.12). The effect size is .44.

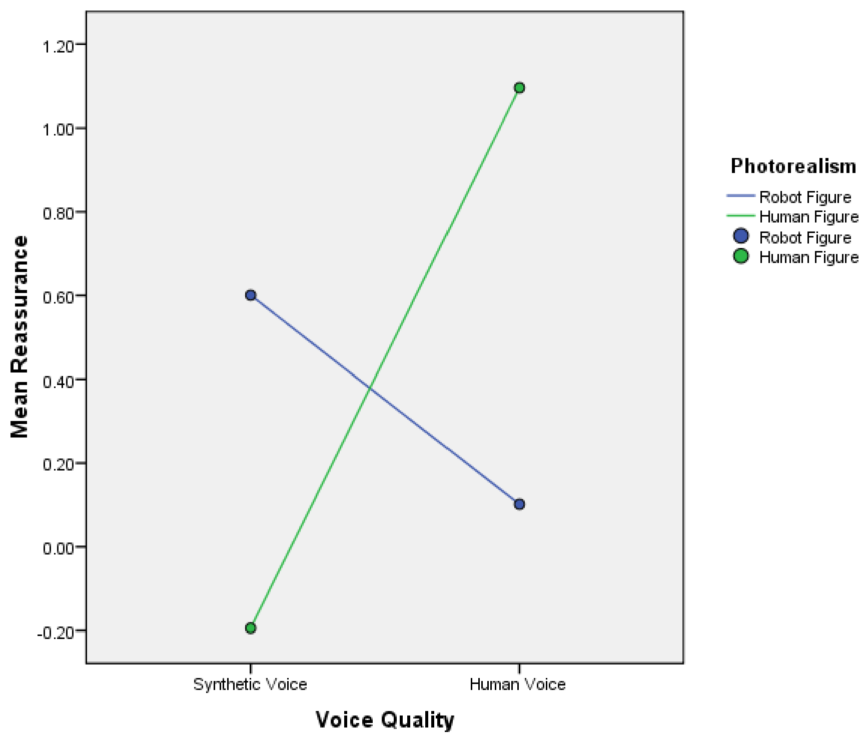

Figure 4.7. Main effect of human photorealism and voice quality factors in *reassurance.*

When grouped by visual elements, further analysis indicates that the conditions which contain the *Robot Figure* are rated less *human* (*M* = -1.91, *SD* = .83) than the *Human Figure* conditions (*M* = .50, *SD* = 1.20) and there is a significant main effect $F(1, 47) = 110.15$, $p < .001$. The effect size is .70 (Figure 4.8). Participants rated the *Robot Figure* conditions more *warm* (*M* = .07, *SD* = .70) than the *Human Figure* conditions (*M* = -.74, *SD* = .79). There is also a significant main effect with respect to *warmth* $F(1, 47) = 27.62$, $p < .001$. The effect size is .37. The *Human Figure* condition's *competence* and *reassurance* were rated (*M* = .08, *SD* = .84) and (*M* = .45, *SD* = .82) respectively. The *Robot Figure* conditions *competence* and *reassurance* were rated (M = .05, SD = .97) and (M = .35, SD = .75) respectively. Neither *competence* nor *reassurance* were significant when comparing the *Human Figure* versus *Robot Figure* conditions.

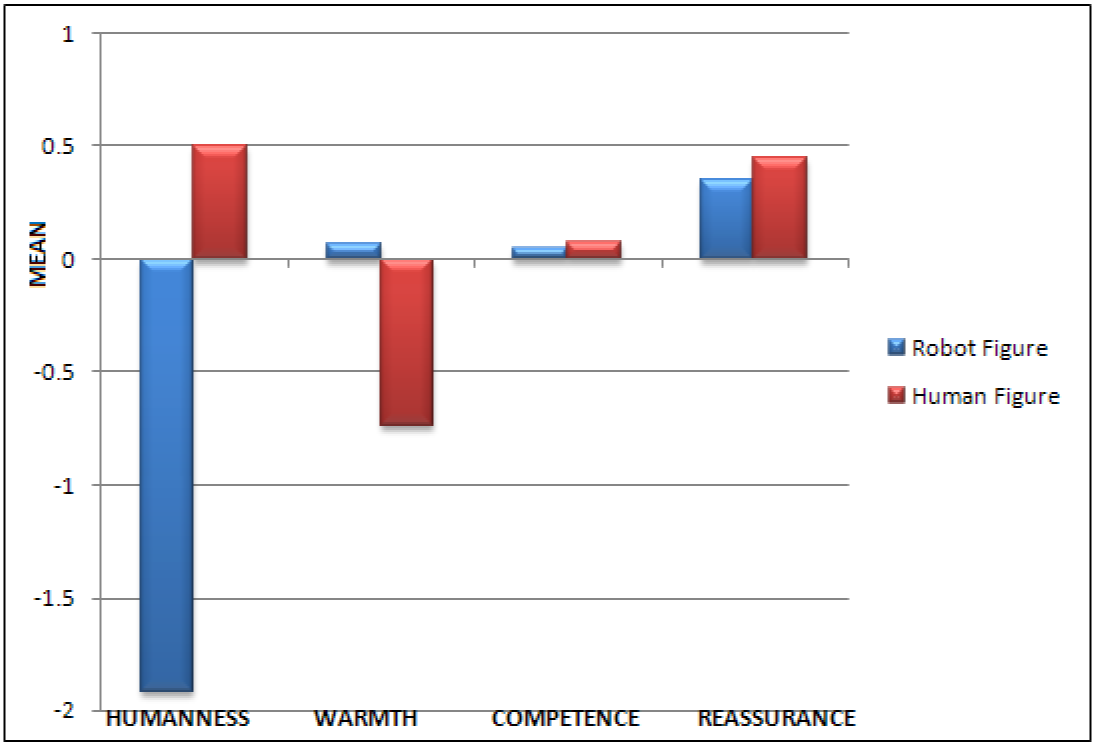

Figure 4.8. Mean values of *humanness*, *warmth*, *competence*, and *reassurance* scales categorized by human photorealism factors (i.e. *Human Figure* or *Robot Figure*).

When grouped by auditory elements of the stimuli, (Figure 4.9) conditions which contain the *Synthetic Voice* component were rated less *human* (M = .11, SD = .88). The conditions which contain the *Human Voice* are rated more *human* than the conditions with synthetic voices (M = .75, SD = .86). Unlike Figure 4.8 where the *Human Figure* conditions are rated as less warm, the conditions which contain the *Synthetic Voice* component are rated less warm (M = -.55, SD = .78) than the conditions containing the *Human Voice* component (M = -.12, SD = .59). Both *humanness* $F(1, 47) = 23.92$, $p < .001$ and *warmth* $F(1, 47) = 11.15$, $p = .002$ are significant. The effect sizes are .34 and .19 respectively. The *Human Voice* conditions were also rated more *competent* (M = .20, SD = .88) and more *reassuring* (M = .60, SD = .79) than the synthetic voice conditions. There is also a significant effect for both *comptence* $F(1, 47) = 4.46$, $p = .04$ and *reassurance* $F(1, 47) = 13.28$, $p = .001$. The effect sizes are .09 and .22 respectively.

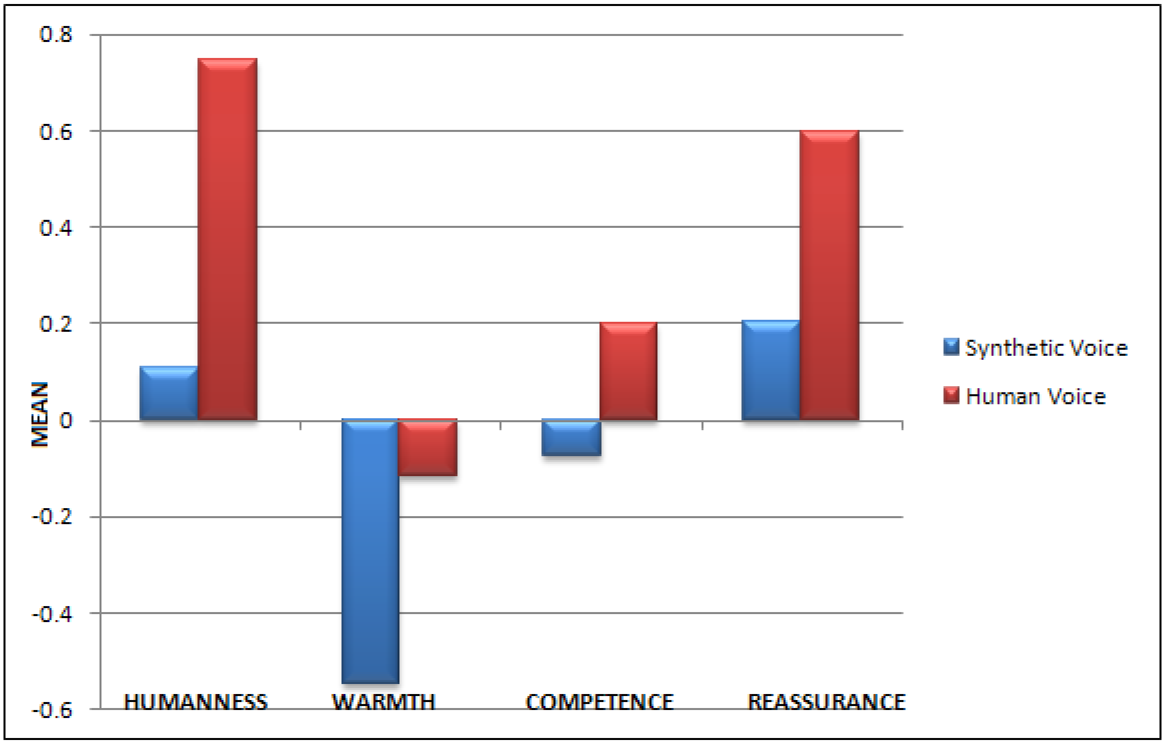

Figure 4.9. Mean values of *humanness*, *warmth*, *competence*, and *reassurance* scales categorized by voice quality factors (i.e. *Human Voice* or *Synthetic Voice*).

The *Mismatched* and *Matched* conditions (Figure 4.10) appear to elicit a greater disparity in ratings for both *humanness* and *reassurance*. The *Mismatched* conditions are rated less *human* (M =.96, SD = .85) and less *reassuring* (M = -.05, SD = .98) than the *Matched* conditions. Both *humanness* $F(1, 47) = 126.61$, $p < .001$ and *reassurance* $F(1, 47) = 24.33$, $p < .001$ are significant. The effect sizes are .28 for *humanness* and .44 for *reassurance*. The *Matched* conditions are also rated as more *warm* and more *competent* than the *Mismatched* conditions, however there is not a significant effect.

4.2 Correlations

A Pearson correlation coefficient was computed to assess the relationships between the overall rating of semantic differential items detected for each stimuli condition or group selected for analysis. First we will focus on the four original stimuli categories as displayed in Table 4.1.

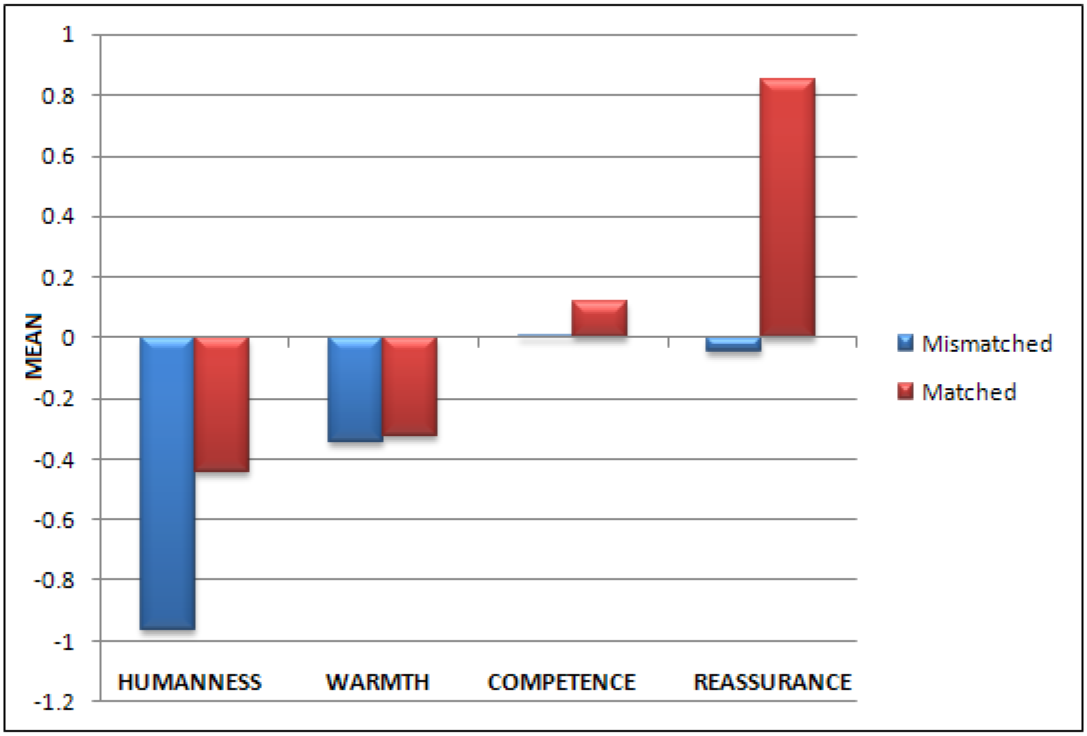

Figure 4.10. Mean values of humanness, warmth, competence, and reassurance scales categorized by stimulus condition which consisted of a *Matched* or *Mismatched* condition based on Figure and Voice attributes.

In the *Human Figure – Human Voice, Human Figure – Synthetic Voice, Robot Figure – Human Voice,* and *Robot Figure – Synthetic Voice* conditions *humanness – warmth* are all positively correlated. However, only the *Human Figure – Human Voice* (r = .33, n = 48, p = .022) condition is significant. *Humanness – competence* are positively correlated in all conditions, but is highly significant in only the *Human Figure – Human Voice* condition (r = .57, n = 48, $p <$ 0.001). In the *Human Figure – Human Voice* (r = .42, n = 48, p = .003) and *Human Figure – Synthetic Voice* conditions (r = .33, n = 48, p = .023) *humanness – reassurance* are both positively and significantly correlated. *Warmth – competence* are all positively and significantly correlated for all conditions although the *Human Figure – Synthetic Voice* is the least significant (*r* = .32, n = 48, *p* = .027). We now move our focus toward the correlated scales rather than the stimuli conditions.

Table 4.1: Correlation between the humanness, warmth, competence, and reassurance scales arranged by Human Figure – Human Voice condition, Human Figure – Synthetic Voice condition, Robot Figure – Human Voice condition, Robot Figure – Synthetic Voice condition, All but Human Figure - Human Voice, and *All but Robot Figure - Synthetic Voice*.

Correlated Scales	*Conditions*	HFHV	HFSV	RFHV	RFSV	All But HFHV	All But RFSV
Humanness –	Warmth	.33**	.24	.23	.20	.33*	.47**
Humanness –	Competence	.57**	.11	.15	-.05	.02	.37**
Warmth –	Competence	.41**	.32*	.44**	.42**	.41**	.40**
Humanness –	Reassurance	.42*	.33*	-.13	-.17	.14	.23
Warmth –	Reassurance	.10	.04	.03	.13	.10	-.11
Competence –	Reassurance	.03	-.20	-.32*	-.16	-.24	-.25

* $p < .05$ (2-tailed).
** $p < .01$ (2-tailed).

For *humanness* and *reassurance*, the *Robot Figure – Human Voice* and *Robot Figure – Synthetic Voice* conditions are both negatively correlated, but not significant. Nevertheless, *humanness – reassurance* are significantly negatively correlated in the *Human Figure – Human Voice* (*r* = .42, n = 48, *p* = .003), *Human Figure – Synthetic Voice* (*r* = -.33, n = 48, *p* = .023), *Human Voice* (*r* = .68, n = 48, $p < .001$), *Synthetic Voice* (*r* = .60, n = 48, $p < 0.001$), and *Human Figure* (*r* = .34, n = 48, *p* = .019) conditions (Table 4.1 and Table 4.2). Each of these conditions includes some sort of a human aspect whether it is the human figure or the human voice or a combination thereof.

Competence and *reassurance* are negatively correlated in all cases except the *Human Figure – Human Voice* condition yet are negatively correlated and both significant in the *Robot Figure – Human Voice* (*r* = -.32, n = 48, *p* = .017), and *Mismatched* conditions (*r* = -.29, n = 48, *p* = .05). This indicates that as ratings of *competence* increase in these conditions *reassurance* decreases. Conversely, as the stimuli in these conditions are rated more *reassuring*, they also become less *competent*. Again, these are for conditions which include a synthetic component. It is interesting that although *warmth – competence* are positively and significantly correlated in the *Robot Figure – Human Voice* and *Mismatched* conditions they are also negatively and significantly correlated with respect to *competence* and *reassurance*.

Table 4.2: Correlation between the humanness, warmth, competence, and reassurance scales arranged by Human Voice, *Synthetic Voice*, *Human Figure*, *Robot Figure*, *Matched*, and *Mismatched factors.*

Correlated Scales	*Conditions*	Human Voice	Synthetic Voice	Human Figure	Robot Figure	Matched	Mismatched
Humanness –	Warmth	.07	.31*	.35*	.18	.46**	.37**
Humanness –	Competence	.25	-.04	.41**	.05	.34*	.11
Warmth –	Competence	.56**	.37**	.31*	.47**	.32*	.34*
Humanness –	Reassurance	.68**	.60**	.34*	-.15	.09	.23
Warmth –	Reassurance	-.21	.15	.01	.05	.15	.05
Competence –	Reassurance	-.20	-.17	-.15	-.25	-.14	-.29*

* $p < .05$ (2-tailed).
** $p < .01$ (2-tailed).

All conditions indicate that *warmth – competence* are all correlated both positively and are significant. *Human Voice* correlations of *humanness – competence* (r = .32, n = 48, p = .002), *warmth – competence* (r = .35, n = 48, p < .001), and *humanness – reassurance* (r = .37, n = 48, p < .001) are positively and significantly correlated. The conditions which contain a human component indicate that as *competence* rises, *humanness* rises, when *competence* increases *warmth* increases, and when *reassurance* increases *humanness* increases. In contrast, *warmth* and *reassurance* are not significantly correlated for any condition.

The correlations for the *Matched* conditions closely resemble the *Human Figure – Human Voice* condition. *Humanness – competence, warmth – competence, and humanness – warmth* correlations for the *Matched* conditions are all positively and significantly correlated. H*umanness – competence* in the *Matched* conditions are positively and significantly correlated (r = .34, n = 48, p = .02), however they are not in the *Mismatched* conditions. In contrast, *competence – reassurance* is negatively and significantly correlated in the *Mismatched* conditions (r = -.29, n = 48, p = .05), but not in the *Matched* conditions.

CHAPTER 5: DISCUSSION

H1 predicted that the *humanness* measure will increase in the following order: *Robot Figure – Synthetic Voice, Robot Figure – Human Voice, Human Figure – Synthetic Voice, Human Figure – Human Voice*. This prediction is based on the importance of visual aspects of stimuli taken from previous studies (Ambady & Rosenthal, 1992; McGurk & MacDonald 1976; Mehrabian, 1968; Mehrabian, 2007; Shimojo & Shams, 2001; Stein et al., 1996), using visual elements of the stimulus as the primary factor and auditory elements as the secondary factor. The results of the *humanness* measure show a linear increase following the predicted order of *humanness*. This provides evidence in support of H1. Further predictions were made based on the ability to predict the level of *humanness* of the stimuli. A known ranking of perceived *humanness* of the stimuli allows for investigation of an uncanny valley effect.

H2 predicted that the *reassurance* measure will be higher for congruent auditory-visual stimuli (*Human Figure – Human Voice* and *Robot Figure – Synthetic Voice*) compared to incongruent auditory-visual stimuli (*Human Figure – Synthetic Voice* and *Robot Figure – Human Voice*). The results of the *reassurance* scale show evidence that the *Human Figure – Human Voice* condition was most reassuring while the *Robot Figure – Synthetic Voice* condition was second in *reassurance* rating. Again, the *Human Figure – Synthetic Voice* condition had the lowest rating while the *Robot Figure – Human Voice* was rated the second lowest. These results support H2 and provide support for the *humanness* order predicted in H1.

A key aspect of the uncanny valley is the relationship between *humanness* and *reassurance*. A closer look at the results reveals that for the *Human Figure – Human Voice* and *Human Figure – Synthetic Voice* conditions *humanness* and *reassurance* are significantly positively correlated. They are not significantly correlated in the *Robot Figure – Synthetic Voice* and *Robot Figure – Human Voice* conditions. This is also reflected in the correlations of *humanness* and *reassurance* when the conditions are grouped by visual aspects. The *Human Figure* condition's ratings for *humanness* and *reassurance* are significantly positively correlated while they are significantly negatively correlated in the *Robot Figure* conditions. Thus, for the *Human Figure* conditions when *humanness* increases, *reassurance* increases and for the *Robot Figure* conditions when *humanness* increases, *reassurance* decreases. When removing the *Robot Figure – Synthetic Voice* condition, *humanness* and *reassurance* are significantly positively correlated but there is no significant correlation when removing the *Human Figure – Human Voice* condition. From the perspective of the uncanny valley these results indicate that visual aspects of the stimuli are important to ratings of both *humanness* and *reassurance,* but what does the inclusion of the voice indicate?

When grouping the conditions by voice, *humanness* and *reassurance* are significantly positively correlated in the *Human Voice* conditions but not significantly correlated (though slightly negatively correlated) in the *Synthetic Voice* conditions. When grouping the conditions into *Matched* and *Mismatched* categories, *warmth* and *competence* are significantly positively correlated in the *Matched* conditions but not significantly correlated in the *Mismatched* conditions. Additionally the results of the *reassurance* scale for the *Matched* and *Mismatched* conditions show the *Matched* conditions are significantly more *reassuring* than the *Mismatched* conditions. These results provide evidence that the voice contributes to the *humanness* of the stimuli but more importantly, when the auditory and visual expectations match, the stimuli are much more *reassuring*, whether they are human or robot.

CHAPTER SIX: CONCLUSION

The mean ratings of *humanness* and *reassurance* and their correlations indicate that a mismatch of visual and auditory stimuli may be causing an uncanny valley effect. Of the theorized causes of the uncanny valley *mate selection* does not appear to be valid in this study as there was no significant difference between male and female participants, which would be expected if *mate selection* were a factor. *Pathogen avoidance* also does not appear to be a reasonable explanation as the stimuli showed no overt signs of illness. Similarly, the stimuli did not appear to remind participants of death, removing *mortality salience* as a perspective cause. This leaves the most likely candidates for the uncanny valley effect to be *violation of human norms* and *sorites paradox*. It is possible that participants were unable to determine a sufficient dividing line between human and robot, thus bringing about a *sorites paradox*, but the limited number of stimuli and the clearly human or synthetic aspects make this a less likely scenario. This leaves *violation of human norms* as the most likely valid reason for the uncanny valley effect in this study. Participants created expectations based on the visual aspects of the stimuli which were then violated by the auditory aspects in the mismatched conditions.

6.1 Limitations of the Study

The experiment was conducted via website. While this lends itself as producing higher ecological validity to the study, there is less experimental control. Outside influences beyond our purview could have influenced our results. However of more important concern is external validity. This study was conducted with only one human actor and one robot actor with their corresponding voices. Increasing the number of available stimuli adds several more data points to draw inferences from, however this could lead to participant fatigue. Participants were required to provide a rating for 23 separate semantic differential pairs for each condition totaling

92 individual ratings. There was a concern of participant fatigue and boredom during this study due to the large number of ratings. After internal reliability was confirmed, attention was given to removing semantic differentials to shorten future experiments. Additionally, this study was conducted with only one human actor and one robot actor and their corresponding voices. A greater diversity of stimuli would increase the generalizability of the study. This could be addressed in future experiments.

6.2 Future Research

The semantic differential scale for *humanness* can be further reduced to three semantic differential pairs and sustain sufficient internal reliability (Cronbach $\alpha = .64$). The semantic differential pairs *Mechanical Movement – Biological Movement*, and *Inanimate – Living* can be removed in further studies. The semantic differential scale for *warmth* can also be further reduced to three semantic differential pairs and still maintain good internal reliability (Cronbach $\alpha = .67$). The semantic differential pairs *Cold-hearted – Warm-hearted*, and *Ill-tempered – Good-natured* can be removed in further studies. The semantic differential scale for *competence* can only be further reduced to four semantic differential pairs and still maintain high internal reliability (Cronbach $\alpha = .74$). The semantic differential pair *Stupid-Intelligent* can be removed in further studies. No further attempts will be made to reduce the semantic differential pairs in the *reassurance* scale. As previously mentioned, future studies should also incorporate both male and female human actors and more than one robot actor to increase external validity. Having multiple actors and genders will increase the generalizability of the results and remove possible doubts of which factors may have caused the uncanny valley effect.

References

Allen, J., Byron, D., Dzikovska, M., Ferguson, G., Galescu, L., & Stent, A. (2001). Towards conversational human-computer interaction. *AI Magazine, 22*, 27–37.

Ambady, N., & Rosenthal, R. 1992. Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, 111: 256-274.

Anderson, J.D. 1996. *The Reality of Illusion: An Ecological Approach to Cognitive Film Theory*. Carbondale and Edwardsville: Southern Illinois University Press.

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology, 41(3),* 259–290.

Bahrick, L. E. (1987). Infants' intermodal perception of two levels of temporal structure in natural events. *Infant Behavior and Development, 10,* 387 – 416.

Bell, L., Gustafson, J., and Heldner, M. (2003) Prosodic adaptation in human-computer interaction. In Proceedings of ICPhS, pages 2453–2456. Barcelona, Spain.

Bregman, A. S., & Pinker, S. (1978). Auditory streaming and the building of timbre. *Canadian Journal of Psychology, 32,* 19 – 31.

Broadbent, D. E. (1952). Listening to one of two synchronous messages. *Journal of Experimental Psychology, 44*, 51 – 55.

Broadbent, D. E. (1954). The role of auditory localization in attention and memory span. *Journal of Experimental Psychology, 47*, 191 – 196.

Brookes, H., Slater, A., Quinn, P. C., Lewkowicz, D. J., Hayes, R., & Brown, E. (2001). Three-month-old infants learn arbitrary auditory–visual pairings between voices and faces. *Infant and Child Development*, 10, 75–82.

Clements K, Rassekh C, Seikaly H, Hokanson J, Calhoun K. Communication after laryngectomy: an assessment of patient satisfaction. *Otolaryngol Head Neck Surg.* 1997;(123)5:493–496.

Coulston, R., Oviatt, S., and Darves, C. (2002) Amplitude convergence in children's conversational speech with animated personas. In Proceedings of ICSLP, pages 2689–2692. Denver: CO, USA.

Crabtree, M., Mirenda, P., and Beukelman, D. (1990) Age and gender preferences for synthetic and natural speech. *Augmentative and Alternative Communication*, 6(4):256–261.

Creak, G. (1999) When HCI should be HHI. *Information Technology and Disabilities*, 6.

Darves, C. and Oviatt, S. (2002) Adaptation of users' spoken dialogue patterns in a conversational interface. In Proceedings of ICSLP, volume 1, pages 561–564,. Denver: CO, USA.

Drager, K. D., Hustad, K. C., & Gable, K. L. (2004). Telephone communication: Synthetic and dysarthric speech intelligibility and listener preferences. *Augmentative and Alternative Communication, 20*(2), 103–112.

Festinger, L. (1956). *When Prophecy Fails: A Social and Psychological Study of A Modern Group that Predicted the Destruction of the World*, by Leon Festinger, Henry Riecken, and Stanley Schachter. Harper-Torchbooks, Jan. 1956.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.

Festinger, L. & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology*, 58, 203 – 210.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from status and competition. *Journal of Personality and Social Psychology, 82(6),* 878–902.

Gordon, R. M. (1986). Folk psychology as simulation. *Mind and Language, 1,* 158–171.

Green, R. D., MacDorman, K. F., Ho, C.-C., & Vasudevan, S. K. (2008). Sensitivity to the proportions of faces that vary in human likeness. *Computers in Human Behavior, 24(5),* 2456–2474.

Grimshaw, M. (2008). 'Sound and Immersion in the First-Person Shooter', *International Journal of Intelligent Games & Simulation*, 5:1, pp. 2–8.

Grimshaw, M. (2009). 'The Audio Uncanny Valley: Sound, Fear and the Horror Game', in *Audio Mostly: 4th Conference on Interaction with Sound*, Glasgow, 2–3 September, http://digitalcommons.bolton.ac.uk/gcct_conferencepr/9/. Accessed 5 December 2009.

D. Hanson, A. Olney, I. A. Pereira, and M. Zielke, "Upending the uncanny valley," in *National Conference on Artificial Intelligence (AAAI'05)*, Pittsburgh, PA, 2005.

Hindmarsh, J., Fraser, M., Heath, C., & Benford, S. (2001). Virtually missing the point: Configuring CVEs for object-focused interaction. In E. F. Churchill, D. N. Snowdon, & A. J. Munro (Eds.). *Collaborative virtual environments: Digital places and spaces for interaction* (pp. 115–142). London Berlin Heidelberg: Springer-Verlag.

Ho, C.-C., MacDorman, K. F., & Pramono, Z. A. D. (2008). Human emotion and the uncanny valley: A GLM, MDS, and ISOMAP analysis of robot video ratings. *Proceedings of the Third ACM/IEEE International Conference on Human-Robot Interaction* (pp. 169–176). March 11–14. Amsterdam.

Ho, C.-C., & MacDorman, K. F. (2010). Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior*, 26(6), 1508-1518.

Hollich, G., Newman, R. S., & Jusczyk, P.W. (2005). Infants' use of synchronized visual information to separate streams of speech. *Child Development*, 76, 598–613.

Isla, D., & Blumberg, B. (2002, March). *New challenges for character-based Al for games.* Paper presented at the AAAI Spring Symposium on AI and Interactive Entertainment, Palo Alto, CA.

Karaiskos, V., King, S., Clark, R., and Mayo, C. (2008) The Blizzard challenge 2008. In Proceedings of the Blizzard Challenge Workshop, 2008. Brisbane, Australia.

Kilborn, R. (1993). 'speak my language': Current attitudes to television subtitling and dubbing. *Media Culture Society, 15*(4), 641-660.

Koda, T. (1996). *Agents with faces: A study on the effects of personi.cation of software agents.* Unpublished master's thesis, Massachusetts Institute of Technology.

Koolstra, C. M., Peeters, A. L., & Spinhof, H. (2002). The pros and cons of dubbing and subtitling. *European Journal of Communication, 17*(3), 325-354.

Kromand, D. (2008), ‘Sound and the Diegesis in Survival-Horror Games’, in Audio Mostly 2008, Sweden, 22–23 October, http://www.audiomostly.se/images/stories/proceeding08/proceedings_am08_low.pdf. Accessed 10 August 2010.

Lewkowicz, D. J. (1986). Developmental changes in infants’ bisensory response to synchronous durations. *Infant Behavior and Development*, 9, 335 – 353.

Lewkowicz, D. J., & Lickliter, R. (1994). *The development of intersensory perception: Comparative perspectives*. Hillsdale, NJ: Erlbaum.

MacDorman, K. F. (2005a). Androids as an experimental apparatus: why is there an uncanny valley and can we exploit it? In *Proceedings of the CogSci 2005 workshop: toward social mechanisms of android science* (pp. 106–118), Stresa, Italy (2003).

MacDorman, K. F. (2005b). Mortality salience and the uncanny valley. *IEEE-RAS International Conference on Humanoid Robots* (pp. 339–405). December 5–7, 2005. Tsukuba, Japan.

MacDorman, K. F. (2006). Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley. *ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science* (pp. 26–29). July 26, 2005. Vancouver, Canada.

MacDorman, K. F. & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive science research. *Interaction Studies, 7*(3), 297-337.

MacDorman, K. F., Green, R. D., Ho, C.-C., & Koch, C. (2009). Too real for comfort: Uncanny responses to computer generated faces. *Computers in Human Behavior, 25,* 695-710.

MacDorman, K. F., Vasudevan, S. K., & Ho, C.-C. (2009). Does Japan really have robot mania? Comparing attitudes by implicit and explicit measures. *AI & Society, 23(4),* 485-510.

Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ: Erlbaum.

Matsui, D., Minato, T., MacDorman, K. F., & Ishiguro, H. (2005). Generating natural motion in an android by mapping human motion. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 1089–1096). August 2–6, 2005. Edmonton, Canada.

McGurk, H & MacDonald, J (1976); "Hearing lips and seeing voices," *Nature*, Vol 264(5588), pp. 746–748.

Mehrabian, A. (1972). *Nonverbal communication.* Chicago: Aldine.

Meltzoff, A. N., & Borton, R. W. (1979). Intermodal matching by human neonates. *Nature, 282,* 403 – 404.

Mitchell, W. J., Ho, C.-C., Patel, H., & MacDorman, K. F. (in press). Does social desirability bias favor humans? Explicit–implicit evaluations of synthesized speech support a new HCI model of impression management. *Computers in Human Behavior*.

Moore, R. and Morris, A. (1992) Experiences collecting genuine spoken enquiries using WOZ techniques. In Proceedings of the 5th DARPA Workshop on Speech and Natural Language, pages 61–63. New York: NY, USA.

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

Morrongiello, B. A., Fenwick, K., & Nutley, T. (1998). Developmental changes in associations between auditory–visual events. *Infant Behavior and Development*, 21, 613–626.

Newell, C. (2005). Achieving a Lively Voice. *Thesis for University of York* "prosody." *Merriam-Webster Online Dictionary*. 2007. http://www.merriam-webster.com (8 Sept. 2007).

O'Keefe, B., Brown, L., and Schuller, R. (1998) Identification and rankings of communication aid features by five groups. Augmentative and Alternative Communication, 14(1):37–50.

Plantec.P. 2007. The Digital Eye: Crossing the Great Uncanny Valley. (Dec. 19, 2007). DOI= http://vfxworld.com/?atype=articles&id=3494

Pyszczynski, T., Greenberg, J., & Solomon, S. (1999). A dual-process model of defense against conscious and unconscious death-related thoughts*:* An extension of terror management theory*. Psychological Review, 106* (4), 835–845.

Ramey, C.H. (2005). The uncanny valley of similarities concerning abortion, baldness, heaps of sand, and humanlike robots. In Proceedings of the Views of the Uncanny Valley Workshop, IEEE-RAS International Conference on Humanoid Robots.

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places.* Stanford, CA: CSLI Publications.

Rhodes, G. & Zebrowitz, L. A. (eds) (2002). Facial Attractiveness: Evolutionary, Cognitive, and Social Perspectives, Ablex Publishing.

Rice, R. (1993). Media appropriateness: Using social presence theory to compare traditional and new organizational media. *Human Communication Research,* 19(4), 451–484.

Rosenberg, S., Nelson, C., & Vivekananthan, P. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology, 9(4),* 283–294.

Rozin, P. & Fallon, A. E. (1987) A perspective on disgust. *Psychological Review*, 94, 23–41.

Shimojo, S., & Shams, L. (2001). Sensory modalities are not separate modalities: plasticity and interactions. *Current Opinion In Neurobiology*, Vol. 11 (4), pp. 505-509.

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications.* London: John Wiley & Sons.

Skipper, J.I., van Wassenhove, V., Nusbaum, H.C. & Small, S.L. (2007). Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex*, 17, 2387–2399.

Slater, M. (2002). Presence and the Sixth sense, in Presence: Teleoperators and Virtual Environments, MIT Press 11(4) pp. 435-439.

Slater, M., Brogni, A., & Steed, A. (2003). Physiological responses to breaks in presence: A pilot study. In Presence: *The 6th Annual International Workshop on Presence*.

Slater, M. and Steed, A. (2000). A Virtual Presence Counter, Presence: Teleoperators and Virtual Environments 9(5), pp. 413-434.

Spadoni, R. (2003). The Uncanny Body of Early Sound Film. *The Velvet Light Trap*, 51, 4-16.

Spadoni, R., (2007). *Uncanny Bodies*. Berkeley: University of California Press.

Stern, S., Mullennix, J., Dyson, C., and Wilson, S. (1999) The persuasiveness of synthetic speech versus human speech. *Human Factors*, 41:588–595.

Stern, S., Mullennix, J., and Wilson, S. (2002) Effects of perceived disability on persuasiveness of computer synthesised speech. Journal of Applied Psychology, 87:411–417.

Stern, S., Mullennix, J., and Yaroslavsky, I. (2006) Persuasion and social perception of human vs. synthetic voice across person as source and computer as source conditions. *International Journal of Human-Computer Studies*, 64:43–52.

Stern, S. (2008) Computer synthesised speech and perceptions of the social influence of disabled users. Journal of Language and Social Psychology, 27(3):254–265.

Stevens, C., Lee, N., Vonwiller, J. and Burnham, D. (2005) On-line experimental methods to evaluate text-to-speech (TTS) synthesis: effects of voice gender and signal quality on intelligibility, naturalness and preference. Computer Speech and Language, 19:129–146.

Street, R. and Giles, H. (1982) Speech accommodation theory: a social cognitive approach to language and speech. In M. Roloff and C. R. Berger, editors, Social Cognition and Communication, pages 193–226. Beverly Hills: Sage.

Taylor, T. L. (2002). Living digitally: Embodiment in virtual worlds. In R. Schroeder (Ed.), *The social life of avatars: Presence and interaction in shared virtual environments* (pp. 40–62). London: Springer-Verlag.

Tinwell, A. & Grimshaw, M., *Bridging the uncanny: An impossible traverse?*, Mindtrek, Tampere, (2009).

Tinwell, A., Grimshaw, M., and Williams, A. "Uncanny behaviour in survival horror games." (2010). *Games Computing and Creative Technologies: Journal Articles* (Peer-Reviewed). Paper 13.

Turkle, S. (1995). *Life on the screen: Identity in the age of the Internet.* New York: Simon & Schuster.

Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication Research, 23,* 3–43.

Wexelblat, A. (1997). *Don't make that face: A report on anthropomorphizing an interface.* Available: http://wex.www.media.mit.edu/people/wex/anthro-expt-paper/Anthro-r.htm.

Wojciszke, B., Bazinska, R., & Jaworski, M. (1998). On the dominance of moral categories in impression formation. *Personality and Social Psychology Bulletin, 24(12),* 1245–1257.

CURRICULUM VITA

KEVIN ANDREW SZERSZEN SR

6828 Blackwell Circle ·Indianapolis, IN 46237 · (317) 644-2190 Email: keszersz@iupui.edu

EXPERIENCE

L-3 STRATIS

NETWORK ADMINISTRATOR III – TEAM LEAD (JUNE 2010 –PRESENT)

Responsibilities.

- Supervise 13 team members by evaluating their progress, achievements, training needs, escalations and other performance related attributes. Document employee performance reviews and appraisals.
- Identify efficiencies; present and recommend improvements to processes and procedures to generally improve the customer experience.
- Maintain a clear and strong understanding of contractual requirements, workload and workflow and personnel performances against key management performance indicators.
- Work with Project Manager and other team leads to determine metrics and measurements and create and present reports on their areas against those matrixes.
- Submit weekly reports to the Project Manager which documents progress on approved initiatives, projects, technical research, or any operational or other issues.
- Participate in and lead high visibility processes and/or national projects as assigned by the Project Manager.
- Responsible for identifying, documenting, posting and maintaining key performance activities, milestones, achievements and other relevant documentation.
- Evaluate, schedule, assign and ensure coverage and backfill requirements are met within the region. Review and approve time cards.

LOCKHEED MARTIN CORPORATION

NETWORK ENGINEER SR (MARCH 2002 –JUNE 2010)

Responsibilities and Accomplishments.

- Project co-lead responsible for the installation and configuration of workstations, printer and networks at 85 facilities in the central United States.
- Project lead for data center relocation which required the relocation of approximately 30 servers and associated networking equipment.
- Project lead for relocation of the Federal Aviation Administration's Springfield System Support Center's entire local area network. Supervised the move and installation of wiring, switches, servers, printers, and phone system.
- Project lead for the design, development, and implementation of a web-based application for safety inspections, and monthly management scorecard reporting utilizing ColdFusionMX, SQL Server, and Dreamweaver MX technologies.
- Installed, configured, maintained and troubleshot Cisco routers and switches. Identified and analyzed IT problems using structured problem resolution approaches while maintaining network hardware and software.
- Performed network administration on computer networks which supported over 1000 users and client workstations in both Novell and Active Directory environments.
- Provided security and vulnerability protection for the Local Area Network against intrusion and virus attacks using McAfee ePolicy Orchestrator and McAfee Virus Scan which included over 1000 workstations. Wrote technical reports and status reports relative to implementation status and technical evaluations for a variety of hardware, operating systems and software applications.

UNITED STATES MARINE CORPS
INFORMATION SYSTEMS ANALYST (JANUARY 1999 –MARCH 2002)
Responsibilities.

- Analyzed, designed, tested, and evaluated network systems, such as local area networks, wide area networks, Internet, intranet, and other data communications systems.
- Provided technical assistance and resolved difficult Information Technology support requests to customers involving a wide variety of different platforms, operating systems, applications, and desktop configurations.
- Developed and maintained the local organizations intranet website and knowledge management initiative. Supervised and instructed network support specialists.

UNITED STATES MARINE CORPS
DATA / TELECOMMUNICATIONS MAINTENANCE TECHNICIAN (JANUARY 1999 –MARCH 2002)
Responsibilities.

- Utilized Lotus Notes databases, Marine Corps Integrated Maintenance Management software, and IBM's Host On Demand software, to coordinate equipment repair with other communication and electronic sections, evacuate irreparable components, complete requisitions, compile deadline reports, and other required documentation.
- Specialized in working with various types of communication hardware, data terminals, cryptographic devices, and a wide range of test equipment and calibration devices.
- Supervised maintenance activities and instructed subordinates in the use and repair of radio-related equipment and systems.

UNITED STATES MARINE CORPS
SATELLITE COMMUNICATIONS ENGINEER (JULY 1992 –DECEMBER 1999)
Responsibilities.

- Provided the supported unit global and complex telecommunication networks via satellite based systems configured for C, KU, and X band with fiber optic LAN/WAN capability.
- Specialized in Systems Planning, Management, and Engineering of networks, using Unix Workstations, satellite tracking systems and global positioning devices.
- Appointed as a UNIX system administrator of JDISS and Warrior.
- Responsible for instructing junior officers on network implementation.

EDUCATION

INDIANA UNIVERSITY-PURDUE UNIVERSITY INDIANAPOLIS
535 W. MICHIGAN ST. INDIANAPOLIS, IN 46202
January 2007 – December 2010
M.S. – Informatics – Human Computer Interaction
Honors, related activities, accomplishments.
Golden Key International Honour Society

UNIVERSITY OF PHOENIX
3157 E. ELWOOD ST. PHOENIX, AZ 85034
November 2003 – April 2006
B.S. – Information Technology with Distinction

CERTIFICATIONS

ITIL V3 FOUNDATION	MAY 2010
COMPTIA NETWORK+	OCTOBER 2003
COMPTIA A+ HARDWARE	OCTOBER 2003
COMPTIA A+ SOFTWARE	OCTOBER 2003