

Dartmouth College

Dartmouth Digital Commons

Dartmouth College Undergraduate Theses

Theses and Dissertations

5-1-2019

Is Augmented Reality in Denial of the Convention? Examining the Presence of Uncanny Valley in Augmented Reality

Sung Jun Park
Dartmouth College

Follow this and additional works at: https://digitalcommons.dartmouth.edu/senior_theses



Part of the [Computer Sciences Commons](#)

Recommended Citation

Park, Sung Jun, "Is Augmented Reality in Denial of the Convention? Examining the Presence of Uncanny Valley in Augmented Reality" (2019). *Dartmouth College Undergraduate Theses*. 143.
https://digitalcommons.dartmouth.edu/senior_theses/143

This Thesis (Undergraduate) is brought to you for free and open access by the Theses and Dissertations at Dartmouth Digital Commons. It has been accepted for inclusion in Dartmouth College Undergraduate Theses by an authorized administrator of Dartmouth Digital Commons. For more information, please contact dartmouthdigitalcommons@groups.dartmouth.edu.

Is Augmented Reality in Denial of the Convention? Examining the Presence of Uncanny Valley in Augmented Reality

Sung Jun Park

Advisor: Xia Zhou

Abstract—Uncanny valley is a theorized psychological phenomenon, which captures a non-monotonic relationship between an entity's anthropomorphic level and the *shinwakan* (affinity) its viewers feel toward the entity [1]. According to the theory, viewers feel a stronger affinity to an anthropomorphic entity as its level of human likeness increases until it reaches a certain point where that affinity is brought to a sudden drop. This valley, although frequently observed, still remains not well understood or explained. That said, most studies purport to present an explanation to the valley in context of robotics or computer-generated images portrayed on 2D surfaces, but it is unclear whether these explanations are applicable to different platforms that offer a completely different user experience. Hence, this study attempts to explore the uncanny valley in context of the augmented reality. AR overlays texts and images on top of our visual feed to interactively blend the real and the virtual world. With its immersive and interactive nature, AR may potentially offer us a new perspective in understanding uncanny valley. The rules that we believed governed the uncanny valley in robotics and previous 2D platforms may no longer be applicable to AR. Consequently, this study observes a relational trend between entities uncanniness and human likeness using Microsoft HoloLens via the user survey and biometrics data to offer viable explanations to how we should interpret the uncanny valley in AR setting.

1 INTRODUCTION

As Figure 1 suggests, uncanny valley portrays a relationship between the anthropomorphic level of a figure and the emotional responses the figure evokes from its viewers; current holding theory asserts that viewers feel increasingly uncomfortable as the portrayed object, whether it be a robot or a 3-D modeled character, parallels and resembles its real-life counterpart to a certain level of imperfection. Where the valley exactly lies remains in question and, more importantly, differs across all platforms.

However, in most cases, consensus seems to assimilate to 70% - 85% mark; in other words, the theory holds that viewers feel the most disturbed when the portrayed figure reaches 70% - 85% human-like resemblance.

It should be acknowledged, however, that human likeness is a tricky quality to be measured with a consistent metrics. Most studies rely on users' self-reported rating on anthropomorphic figures to assess their human likeness, and as people have varying standards and viewpoints, it always poses a big challenge to conduct a proper assessment. That said, it is hard to affirm whether the prior studies point to the same "human likeness", where they claim to have observed the uncanny valley.

Along with addressing these problems in evaluating these subjective qualities that define uncanny valley, this study purports to suggest a way to measure them. And using these metrics, we intend to discern whether uncanny valley exists in the world of AR and if so, where it lies.

1.1 Motive

One of the areas where uncanny valley comes to a great importance is the world of interactive application and gam-

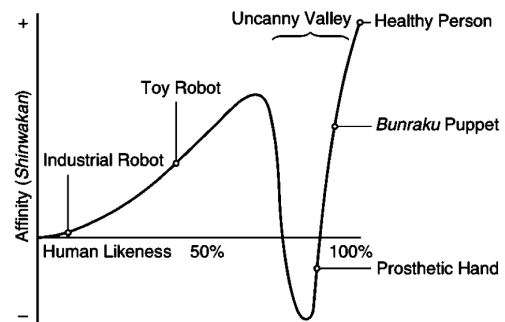


Figure 1: Realism vs. Emotional Discomfort:

This graph denotes Mori's original proposition with his suggested examples to mark each differing level of human likeness.

ing serves as a great epitome of that. In order for games to be appealing to the public, the game characters have to be visually approachable. Game characters serve as the only form of identity users have in the given virtual world and thus some users tend to form strong associations with their virtual avatar. Therefore it is not an overstatement to say that the visual believability and acceptability of the game characters is one of the most important game properties that determine the level of virtual immersion users experience. In fact there are already various anecdotal evidence of users reporting on how creeped out they were after interacting with characters that did not seem quite acceptable. Given that the most prominent benefits of augmented reality for users is its ability to offer immersive and interactive environment, AR has and will continue to impact the industry heavily. Therefore research into uncanny valley, especially

concerning anthropomorphic avatar will be paramount in determining future roadmaps for the industry as a whole.

And augmented reality HMDs offer novel opportunities to analyze where the uncanny valley lies. What sets them apart from everyday 2D platform experience is the level of interactivity and immersivity the AR platform can offer to users. These new apparatus, though with their limitations, are becoming more and more sophisticated in terms of blending the real 3D world around us with artificial virtual images and animations. Given the rate of technological advance, it won't be long until AR technology is mature that even the most shrewd users will have a hard time determining what is real and what is augmented. This is where research into uncanny valley will truly shine. Insights from this research will be able to approximate how the unique experience AR has to offer can effect the uncanny valley and a plausible explanation of why.

1.2 Study Briefing

The general idea is to examine how users react to differing level of realism a character exhibits in the AR world. The study participants were shown 7 different types of avatars, each with a distinct facial and bodily anthropomorphism levels. Then, the participants responses were collected via biometric sensors (GSR and EEG) and post experiment survey. The participants' survey responses were analyzed to ascertain the relationship 1) between the avatars' induced uncanniness and their human likeness, 2) between the avatars' induced uncanniness and their facial anthropomorphic level, and 3) between the avatars' induced uncanniness and their body proportion level. In supplementation with the survey analysis, frontal alpha asymmetries were computed from the collected EEG data in representation of the participants' emotional status while viewing the avatars [2]. Moreover, participants' skin conductance responses (SCR) gathered from the GSR sensor were used to determine how much emotional arousal each avatar triggers from the users.

The result did not align with the conventional expectation of the uncanny valley theory. Normally, we would expect to see a character on a higher end of human likeness level to be the least liked, but both the survey and the sensor results indicated that a character on a lower end of the human likeness level was the "least liked"¹ with the characters becoming more "likeable" as they approach the higher end. This study attempts to find the reason in what makes augmented reality unique: the blending of the real and the virtual. That is, the realism inconsistency between a character and its surrounding environment affects how viewers perceive the character and has caused the uncanny valley to shift.

2 BACKGROUND

2.1 Its Origin and Implication

Term uncanny valley originates from a Japanese robotics professor named Masahiro Mori from 1970; Mori hypothesized that in contrast to how most "phenomena of everyday

1. characters' likeability is a high-level view on the characters' uncanny index determined from the survey result and alpha asymmetries determined from EEG data

life" can be explained with monotonically increasing functions, "a persons response to a humanlike robot [abruptly shifts] from empathy to revulsion as it approach[s], but fail[s] to attain, a lifelike appearance" [1]. That is, a robots increasing resemblance to its human counterpart leads to an increased liking until it reaches a certain level of human-likeness, where that liking is brought to an immediate drop. Mori compares machinery robotic arms to realistic prosthetic hands to describe the phenomenon; he claims that prosthetic hand's rather realistic appearance combined with its subtle artificialism can bring negative affinities in its observers unlike the bluntly artificial robot arms.

Moreover, as Figure A and B exhibits, he claims that movements (body animacy and facial deformation) extremizes the relational magnitude. In other words, one is more likely to feel a closer affinity to a moving human being than a still one and a stronger revulsion to an animate humanoid than an inanimate one.

2.2 How To Measure the Uncanny Valley Effect

2.2.1 Godspeed Indices

Christoph Bartneck suggests a series of questionnaires to measure humans' perception of robots, which, in most cases, are directly applicable to most forms of uncanny valley studies [3]. His indices involve:

- anthromorphism
- animacy
- likeability
- perceived intelligence
- perceived safety

Anthropomorphism describes the "attribution of a human form, human characteristics, or human behavior to nonhuman things such as robots, computers, and animals" [3]. It often refers to physical attributes we would often relate human quality with.

Likeability, as the name hints, describes the tone and quality, whether it be positive or negative, of the subjects' first impressions.

Animacy embodies the concept of "moving of one's own accord"; the concept of animacy has to be, therefore, strongly inducive to the notion of the being alive [3].

Perceived intelligence is a formulation of human behavior in understandable extent; in order to be considered to have a perceived intelligence, the subject would have to behave and react in accordance to human understood patterns and

Finally, perceived safety is the amount of threat the subject poses. And he called these indices the Godspeed Indices.

2.2.2 MacDorman's Indices

However, Ho and MacDorman questions the validity of Bartneck's Indices via conducting a survey based user study on the semantics of the listed indices. They claim that the indices have high correlation with one another, and, hence, are incapable of target-analyzing different dimensions responsible for uncanny valley. For instance, the correlation between "anthromorphism" and "animacy" were 0.89, from which, they believed, can be arguably stated that the 2

indices are “measuring same concept instead of measuring distinct concepts” [4].

Furthermore, they propose a new set of indices to measure uncanny valley effect. The indices involve:

- attractiveness
- eeriness
- humanness
- warmth

They conducted the same user study on the amended indices and concluded that the correlations between the indices were not statistically significant, allowing them to argue that these indices are capable to measuring distinct concepts contributing to the uncanny valley [5].

Moreover, Ho and MacDorman conducted series of 3 experiments in their succeeding study to further validate the appropriateness of their newly suggested indices: 1) card categorization experiment, 2) adjective scale evaluation survey, and 3) new scale validation survey. [5] Card categorization experiment was used to assess peoples standard of categorization when they are viewed robots, 3d animated characters, and humans. Adjective scale evaluation was used to determine what adjectives people relate robot-related, animation-related, and human-related subjects with. And the new scale validation survey was used to test the effectiveness of their new sub-categorization adjectives in evaluating uncanny valley effect.

Consequently, it can be stated that Ho and MacDorman has provided a guideline for future uncanny valley studies with the attributable adjectives (Inanimate - Living, Synthetic - Real, Mechanical Movement - Biological Movement, Dull - Freaky, Predictable - Eerie, Messy - Sleeky, ... etc.) that could be used to describe ascribable dimensions of featured characters in surveys.

In this study, I will be using a simplified version of MacDorman’s indices to evaluate the users’ perception of the avatars they will be viewed in AR setting.

2.3 Possible Causes of Uncanny Valley Effect

2.3.1 Categorical Inconsistency vs. Realism Inconsistency

The concept of categorical inconsistency origins from Jentsch from 1906; Jentsch argued that “eerie feelings are most reliably elicited by uncertainty about whether it is non human or human” and that “[c]ategory uncertainty occurs whenever an entity transitions from one category to another.” [6] And this explanation seems to easily align with the uncanny valley phenomenon; if artificial anthropomorphic character is real enough to trick one into initially perceiving it as a human, the viewer will unconsciously categorize the character ‘human’, but as soon as the viewer starts recognizing unnatural details and is triggered to believe what was thought to be a human is not a human, the uncanny feeling is evoked.

However, MacDorman and Chattopadhyay argues against the applicability of categorical inconsistency to explaining uncanny valley effect and suggests realism inconsistency theory as its replacement [6]. Realism inconsistency is a concept that differing level of realism in one entity can cause an internal conflict within its viewers as to discerning whether the entity is real or not. And this internal conflict

leads to a prediction error which triggers a formation of negative impression.

To validate their viewpoints, they conducted an experiment, where they blended a photograph of real entities and their 3D replicas in differing ratio and asked the users to rate the eeriness of the blended image to observe the relationship between the level of realism inconsistency and level of eeriness.

2.3.2 Mind Perception in Anthropomorphic Subjects

Gray and Wegner purports to find the causes of uncanny valley effect in humans psychological tendency to “perceive mind in [anthropomorphic subjects]” [7]. According to Gray and Wegner, mind is perceived through recognition of 1) agency and 2) experience. They define agency as “the capacity to do, to plan, and exert self-control”, and experience as “the capacity to feel and to sense” [7]. Essentially, if a robot fails in achieving in one of the two essential qualities of mind, it can bring unnerving sensation to its observers.

Are perceptions of experience and agency equally contributive to the uncanny valley effect or is one significantly more inductive than the other? Gray and Wegener conducted three series of user experiments to find out [7]. For the first experiment, the participants were viewed two different sets of videos of a lifelike robot named Kaspar: 1) videos featuring its human-like appearance and 2) videos featuring its machine hardware (wiring and electrical components). Then, they were asked to rate each set of videos for its level of eeriness, visible agency, and experience. The result showed that the attributed rating of agency between the two sets were not of significant difference, but that the humanlike set were attributed a greater experience and eeriness than the mechanical.

For the second experiment, participants were provided one of three questionnaires: 1) a control version with set of questions on a normal powerful supercomputer, 2) the with experience condition version with set of questions on a supercomputer with the capacitative ability to feel emotions, and 3) with agency condition version with set of questions on a supercomputer with the capability of independently planning and executing actions. The result showed that people found computers with the ability to feel emotions unsettling and suggested that the uncanny valley effect may not only be related to the superficial appearances but also to the perception of experience in the viewed subjects.

The third experiment attempts to unveil whether the findings from the second study also applies to human subjects. The participants were viewed a picture of a man that was described as being either normal, lacking agency, or lacking experience [7]. The result showed that participants were noticeably more unsettled with the man lacking experience.

Consequently, their experiments may suggest a possible explanation to why we are subject to uncanny valley in viewing anthropomorphic figures; human like physical attributes often guide us to believe in the unconscious level that the figure has the ability to be its own agent (capable of perceiving and thinking) but their imperfect facial animation leaves us the impression that it is incapable of feeling emotions, thereby, triggering an uncanny feeling in us.

2.3.3 Effect of Animacy in Uncanny Valley Effect

Lukasz and his team tested out Mori's prediction that motions accentuate and thereby increases the magnitude of uncanny valley effect [8]. They used seven avatars:

- 1) battle robot
- 2) toy robot
- 3) mannequin
- 4) skeleton
- 5) zombie
- 6) low quality man
- 7) high quality man

and motion captured animations to test the effect animacy has on the avatars' acceptability. Moreover, they distorted the motion captured animation to different extents to measure how artificial and unnatural animations' effect differ in comparison to natural animations. They conducted a user study ($n = 40$), where they showed the user a static picture of the 7 avatars (followed by a survey) and a video of the animated ones (also followed by a subsequent survey). The result showed that natural motions, in fact, allowed the subjected avatars to embody higher acceptability and trigger a stronger affinity within the observers [8]. And understandably, avatars with distorted motion were deemed less acceptable than ones with natural motions.

3 STUDY DESIGN

The study can largely be divided into two components: 1) User Interactive Application and 2) Post-scene Survey. The avatar application was developed with Unity Game Engine (2018.3.1f) and was specifically designed to provide the study participants with a controlled immersive / interactive experience featuring 7 characters of differing realism level. After being viewed the AR scene, users are asked to fill out a survey rating for each character based on its human-likeness, attractiveness, and eeriness with subsidiary questions for more specific feedback for a more accurate and fine-grained analysis.

3.1 Hololens Application Design

The app was developed specifically for Microsoft Hololens devices. Microsoft Hololens was chosen over its competitors for its compatibility with the EEG sensor, computing / rendering power and detailed SDK documentation. Non-intrusive AR goggles like Google Lens and Vuzix Blade do not have neither an enough horsepower to render the

avatar characters nor a sufficient field of view to provide immersive experience. On the other hand, powerful goggles like Magic Leap One or Meta 2 are highly intrusive and do not allow the EEG sensor to properly sit on the participants' foreheads to make accurate readings.

As a lengthy study can bore the participants and adversely affect their impression on the avatars that are presented towards the end, thereby compromising the study result, the Hololens app was designed to take no longer than 5 minutes in total. In essence, the app features 7 scenes (Figure 3a), in which a unique character of differing realism level interact with the user under a set animation sequence.

Also, it was expected that the effect the character can have on users' emotions can differ depending on the order the characters are presented; the first character the user sees will most likely leave the strongest impression on users as the user will eventually grow expectations on how the characters will behave for the succeeding characters and be less affected by them. To minimize this potential bias, the app randomizes the order in which the characters are presented.

The app design takes into consideration of what the AR platform can offer that other platforms cannot.

The app uses HoloToolkit's Spatial Mapping and Spatial Understanding prefabs [9] to properly place the virtual characters in real, physical environment. The app uses the devices region scanning capability to map the nearby spatial features as floors and walls so to apply appropriate physics to the AR character allowing them to stand and walk properly on a floor surface: not floating in the air or sinking into the ground.

It is understood that making the avatars interactive and animate can introduce multiple uncontrolled variables to study, but envisioning most use cases of anthropomorphic avatars in AR setting (virtual assistant, virtual conference, games, etc.), the characters had to be designed to be interactive and animate.

The characters were given the ability to interact with the user by 1) speech and 2) motion. Using Windows Diction Recognition system [9], the app allows users to freely call the character at idle state to trigger responses. And using Unity Engine's Inverse Kinematics and vector delta calculation, the characters were scripted so that they can be triggered to stare at and follow the users around.

As the Figure 4 shows, the characters are bound to three layers of animation sequences in total: 1) body animation, 2) facial animation, and 3) blink animation. At the start of each

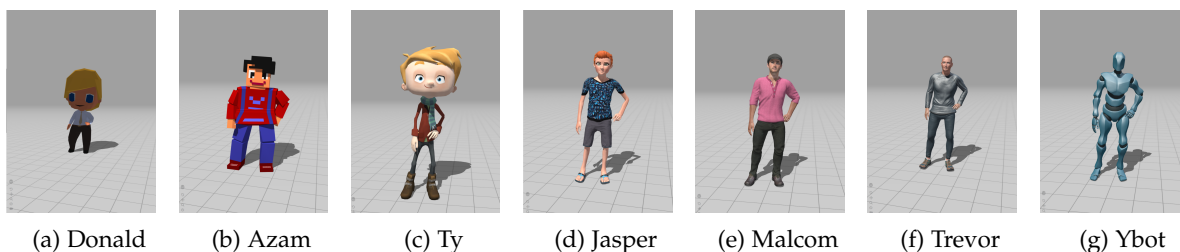


Figure 2: These are the seven 3D modeled CGI characters used for our study. Each avatar was evaluated a distinct facial anthropomorphism and body proportion score (1-7).

(a) The models are originally from Adobe Mixamo and CGTraders and were further modified with Maya for facial animations.

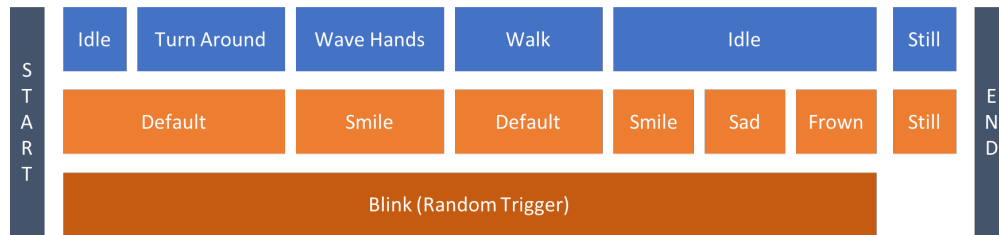


Figure 4: The animation hierarchy shows how 3 different layers of animations that were interwoven together to allow the characters to be more realistically making actions and faces.

scene, the character will be spawned 5 meters away from where user is with its back facing the user. Then, once the user calls the character with his / her voice, the animation sequences are initiated. The character will turn around and greet the user waving its hand and smiling. And it will walk toward where the user is and stop to make various facial expressions as smiling, frowning, and glooming. Eventually, the scene comes to an end with the character losing its animacy and closing its eyes.

The app also holds a Monobehavior script attached to each character's game object that stores the animation state at each epoch.milliseconds in a separate csv file so that the users emotion status can be analyzed in finer grain and in relation to the animation type by aligning EEG and GSR sensor readings with the recorded animation states.

3.2 2D Avatars App Design

To compare the depth of the uncanny valleys in AR platform and regular 2D platforms, an abbreviated 2D version of the Hololens App was developed for Windows PC. The Windows App features the same 7 characters but with a shorter animation window to keep the user from feeling bored from over-repetition as there is a risk that boredom can adversely affect the users' impression of the characters. It is understood that the Windows App is not perfectly comparable to the Hololens App with the scenes and animations not being perfectly identical, but it should suffice the purpose of exposing the user to the same set of avatars in two different setting, allowing us to assess how the setting differences affect the uncanny valley.

3.3 Survey Design

The survey, developed and deployed via Qualtrics, is composed of three bodies of questionnaires. The first body includes questions to assess the users' perception of the characters they observed in AR scenes. Each character was assessed of its human likeness (1-5), attractiveness (1-5) and eeriness (1-7) from participants' self-reported rating. The second body includes questions to assess uncontrolled variables that can affect users' answers to the first body of questionnaires: users' prior experiences with mixed reality, users' level of gaming experience, and users' age. The third body includes questions to observe the differences between AR and 2D platform in perceiving the characters' uncanniness.

The survey is designed to take no longer than 10 minutes per participants, thereby making the total duration of the study approximately 15 minutes in total. After viewing both

the Hololens and Windows App, the users were asked to answer following questions:

- 1) What is your age? (1 - 3)
- 2) Please rate your level of gaming experience (1 - 3)
- 3) Do you have any previous VR/AR experience? (Yes / No)
- 4) Please rate each character by its human-likeness (1 - 5)
- 5) Why did you find it to be the most human-like?
 - a) Facial Anthropomorphism
 - b) Body Proportion
 - c) Others
- 6) Why did you find it to be the least human-like? (Multiple Choice)
 - a) Facial Anthropomorphism
 - b) Body Proportion
 - c) Others
- 7) Please rate each character by its attractiveness (1 - 5)
- 8) Please rank each character by its eeriness (1-7)
- 9) Was watching the characters on a computer monitor any different from watching them on a AR device?
 - In which platform did you find the characters to be more uncanny?

4 PROCEDURE

The study was conducted at a same site in same setting: a room with dim ambient light and no obstacles lying in front of the participants' seat.

It is to be noted that there were a few compromises that had to be made to the study procedure; although the Hololens app was designed so that the participants can freely move around to interact with the characters, biometric sensors used for the study were highly sensitive and fragile, causing the sensor readings to be heavily influenced by motion. To ensure more reliable and accurate sensor readings, the participants were asked to stay seated while wearing the sensors.

All participant ($n = 50$) were asked to wear two biometric sensors: 1) EEG sensor and 2) GSR sensor. EEG sensor named 'Muse' was worn over the participants' forehead and GSR sensors named 'eSense Skin Response' were wrapped around the participants' index and middle fingers. And the sensor data stream was checked in real-time to ensure that there is no sensor reading irregularity.

Then, the participants were asked to wear the Hololens device. The Hololens device's position was adjusted until it was ensured that the participants had a full field of view the device offers. Once Hololens view was adjusted for optimal sight, the participants were asked to follow instruction displayed on the AR display and interact with the portrayed character onscreen.

After viewing all 7 AR scenes, the participants were asked to watch the 7 scenes (abbreviated) on a computer monitor. And once done, the participants were given survey questionnaires to answer, where they were also encouraged to ask clarification questions. And each participant was compensated with 5 dollars in cash at the end of his/her survey.

5 METHODS

5.1 Independent variables

As Mori's original graph (Figure 1) suggests, human likeness ought to be considered "the" independent variable of this study. However, it is hard to really quantify human likeness, and, moreover, there are multiple attributes necessary in describing human likeness.

In a preliminary stage of designing the study, an AWS survey was conducted to assess peoples' standards in identifying anthropomorphic figures. Hundreds of users were viewed a set of static images of 3D generated avatars with varying human qualities, and were asked to evaluate their human likeness. To our surprise, 40% of the survey participants rated a faceless, but bodily well-proportioned avatar (Ybot in Figure 3a) with a high human likeness score. Using this early survey result, we decided to assess avatars' human likeness using two objectively evaluable attributes:

- 1) facial anthropomorphic level
- 2) body proportion level

Both the facial anthropomorphic score and body proportion score were determined ordinally in a relative scale not an absolute scale. This is because although avatars can be objectively compared of their facial anthropomorphic and body proportion levels, their absolute magnitudes cannot be accurately quantified given a set of avatars with widely varying facial and bodily features.

Moreover, users' subjective human likeness rating of each avatar was also used to observe the correlation between the users' self-reported human likeness scores on the avatars and their facial anthropomorphic score / body proportion score. This allows the study to embody both a subjective and an objective scale in measuring human likeness.

5.2 Dependent Variable

As discussed earlier, 'uncanny' can be an ambiguous term with no uniform semantics; as MacDorman suggests, uncanniness should be assessed in multiple dimensions using subsidiary, descriptive adjectives to keep the definition of the word consistent; in this study, I used *attractiveness* and *eeriness*, two of MacDorman Indices, to evaluate the *uncanniness* of each avatar. It is to be acknowledged that this method is founded on top of an assumption that *uncanniness* share a strong positive correlation with *eeriness* and a strong negative correlation with *attractiveness*. That said, *uncanny index* was evaluated by scaling and merging each avatar's *attractiveness* and *eeriness* scores from the survey:

$$1 + a \left(\frac{\text{Eeriness Score}}{\text{Number of Avatars}} \right) - b \left(\frac{\text{Attractiveness Score}}{\text{Rate Range}} \right) \quad (1)$$

The characters' eerinesses were measured ordinally as there is no quantifiable metric to denote a subtle difference

in perceived eeriness of each character. Given that the goal of this study is to understand the relational trend between anthropomorphic realism of an avatar and its uncanniness in AR platform, using ordinal score would not pose any problem. To observe peaks and troughs in a graph to determine where the uncanny valley resides only requires a relative measure for determining whether an arbitrary avatar A induces more uncanny feeling than another arbitrary avatar B, which can be assessed without using an absolute metric. Moreover, using a normal scale to measure eeriness bears a significant risk of inconsistent result as having the participants rate each character in accordance to their subjective standard will likely result in a large variance and high inconsistency as each user's ground level or threshold for assessing eeriness in what they see can be vastly different. Most importantly, given that the uncanny valley effect captures a very subtle change in emotional state, using a rating questionnaire may result in avatars with indistinguishable eeriness scores.

5.3 Uncontrolled Variables

It is understood that there are uncontrolled variables that can hinder us from accurately assessing the relationship between the independent and the dependent variables. In light of our uncanny valley study, the participants' prior game and AR/VR experiences are significant variables that can affect the participants' standards in perceiving anthropomorphic characters in an immersive environment.

Therefore, ANOVA² was tested to exhibit relationships 1) between participants' prior AR/VR experience and their self-reported characters' human likeness score and 2) between participants' level of gaming experience and their self-reported characters' human likeness score. The F-Statistic and p value from ANOVA will be used to verify whether these variables ought to be considered statistically significant in this relationship or not [10].

Moreover, accounting for the possibility that ANOVA result was compromised by the relatively small sample size of this study, a relational trend graph was drawn using seaborn library in Python [11] to get a visual intuition of the effect these variables might have on the relationship.

5.4 Sensor Data

Users' EEG and GSR data were collected along with the survey questionnaires. Using the timestamped Hololens data, EEG and GSR data points were aligned with avatar's animation and facial expressions states at each recorded milliseconds.

Although controversial, prior researches suggest that, to some level, frontal EEG alpha asymmetry can represent frontal cortex activity, which is known to be the control tower for emotional processing [2]. Positive and negative alpha asymmetry supposedly suggests positive and negative mood respectively. Aligning with these previous findings, our study will use alpha asymmetry to measure users' event related mood to each avatar.

EEG data were recorded with a sampling rate of 256 Hz [12] with a Muse Headband. Each EEG data session

2. analysis of variance

was labeled with the avatar name, animation state, user rated human likeness level, objective facial anthropomorphic level, and body proportion level. The data were scaled across different users by configuring each user’s alpha asymmetry ground level and mapping the differences. After a elementary calibration, the minimum alpha asymmetry was recorded in order to capture one defining negative moment for each avatar and each animation segment.

Galvanic Skin Response sensor measures the changes in skin conductivity, which, prior studies suggest, is an indicator of emotional arousal [13]. High skin conductance response (SCR) suggests high emotional arousal while low response suggests low emotional arousal.

For our study, GSR data were recorded with a sampling rate of 10 Hz [14] with eSense Mindfield. Each GSR data session was labeled with the avatar name, animation state, user rated human likeness level, objective facial anthropomorphic level, and body proportion level. The SCR peaks were measured for each avatar / animation segment to determine what avatar or what animation is most likely to trigger a SCR peak. However, SCR cannot capture the polarity between positive and negative emotion, which is why we need both EEG and GSR data to assess users’ mood in our study.

5.5 Resources

The collected data were statistically analyzed, tested, and calibrated using *scipy*, *statsmodels*, *sklearn* and *pandas* Python libraries. And trimmed data were visualized and represented using *pyplot* and *seaborn* libraries [11], [15], [16], [17], [18].

Moreover, Gaussian function was used to draw non-linear regression to better understand the trend.

6 RESULTS

6.1 Using User Reported Human Likeness Level

To observe the effect each character’s subject human likeness level has on its uncanniness, a relational trend plot was

Human Likeness	Uncanny Index		
	mean	median	Std(λ)
1	0.568	0.571	0.179
2	0.575	0.614	0.209
3	0.464	0.486	0.195
4	0.357	0.329	0.174
5	0.315	0.271	0.206

TABLE 1

drawn for uncanny index and subjective human likeness level acquired from the survey. If AR follows the uncanny valley convention, we should see a peak in uncanny index towards the higher end of the subjective human likeness level.

However, as Figure 5b shows, the uncanniness peak was found was at low subjective human likeness level (2), and both the mean trend graph and the Gaussian fit falls monotonically after that early peak of uncanny index. This result can be interpreted in different ways.

First, it could be argued that the uncanniness peak at human likeness level of 2 is the actual uncanny valley in AR. An AR property, which I will be discussing in section 7, could have underscored and amplified the uncanniness of the characters with low human likeness level, causing the valley to be observed much earlier than conventionally expected.

Second, it could be argued that none of the avatars used for the study were humanlike enough to push its viewers down the valley. That is, the monotonic fall in the uncanny index after the peak aligns with the first part of the uncanny valley graph before the dip, where the characters become more “likeable” along with their rising human likeness.

Either way, the uncanny valley in AR seems to be in denial of the convention. The valley is observed either earlier or later than we would normally expect. I suspect that realism inconsistency has a role in this phenomenon,

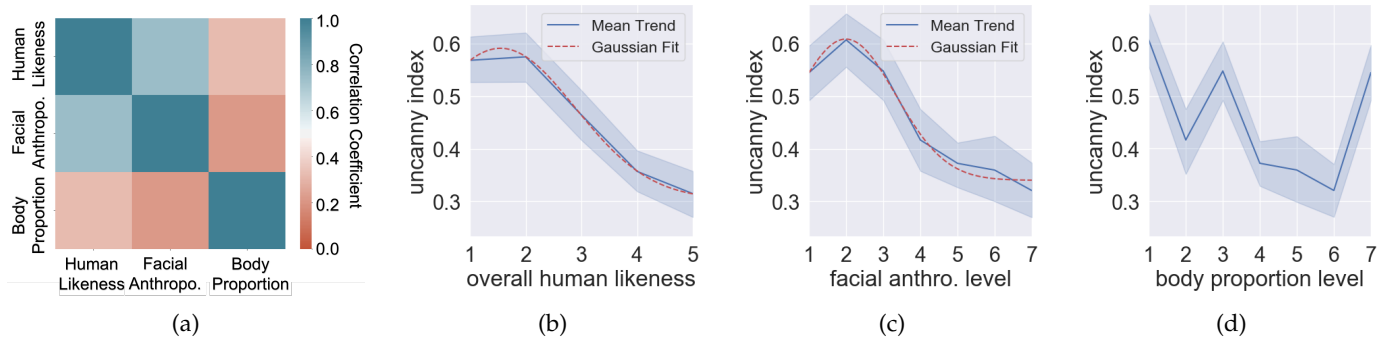


Figure 5:

- (a) Heat map in representation of the **correlation coefficient** among the avatars’ human likeness, facial anthropomorphic level, and body proportion level.
- (b) A relational plot between the avatars’ **subjective human likeness level** and uncanny index with Gaussian function used to fit a non-linear regression.
- (c) A relational plot between the avatars’ **facial anthropomorphic level** and uncanny index with Gaussian function used to fit a non-linear regression.
- (d) A relational plot between the avatars’ **body proportion level** and uncanny index.

which I will be discussing in section .

Before moving on, we have to note the not so small standard deviation of the uncanny index values for each level of human likeness. Given that the uncanny index is an interval scale between 0 and 1, the standard deviation, which ranges from 0.174 to 0.209 can be argued to be substantially large. This indicates users' nonuniform responses to the avatars they were viewed, which highly suggests a larger sampling size for future studies.

6.2 Using Objective Scales

6.2.1 Facial Anthropomorphic Level

Facial Anthropomorphism	Uncanny Index		
	mean	median	Std(λ)
1	0.545	0.557	0.188
2	0.607	0.643	0.188
3	0.548	0.614	0.209
4	0.417	0.414	0.214
5	0.372	0.379	0.151
6	0.359	0.293	0.225
7	0.321	0.271	0.188

TABLE 2

As Figure 5a suggests, high correlation was observed between subjective human likeness level and objective facial anthropomorphic level. Hence, it is expected that the relational plotting of uncanny index and facial anthropomorphic level would show a similar trend with that of uncanny index and subjective human likeness level.

The relational plot in Figure 5c shows that the uncanniness peak was observed at facial anthropomorphic level of 2 followed by a continuous down trend afterward, which is consistent with the results from the subjective human likeness plot.

The high correlation and the very similar trend graphs seems to indicate that the uncanny valley effect is aroused specifically by the characters' facial anthropomorphism, but the standard deviations across user responses are substantially large, which prevents us from further insisting this viewpoint until we gather a larger sample size to minimize the effect of outliers for a more reliable and accountable conclusion.

6.2.2 Body Proportion Level

Although 40% of the participants claimed to relate human like attributes with the figure's body proportion rather than with its facial anthropomorphism in the survey, the correlation (0.32) between the participants' subjective human likeness rating and objective proportion level is not in support with their responses.

Moreover, the relational plot (Figure 5d) of uncanny index and body proportion level show that there is no real visible trend and hint that uncanny valley effect may not be related to characters' body proportion level.

6.3 Uncontrolled Variables

6.3.1 Gaming Experience and Its Effect on Uncanny Valley

A one way ANOVA was conducted for the 5 levels of human likeness to compare the effect of gaming experience on uncanny index in 3 differing levels of gaming experience level: low, moderate, and high. (The mean uncanny index value for each level of gaming experience will be furthered referred to as μ)

The null hypothesis to be tested is $\mu_{low} = \mu_{medium} = \mu_{high}$: that is, the group means for samples with low, moderate, and high level of gaming experiences are not different. A p-value below the alpha level of 0.15 should allow us to revoke this null hypothesis.

Three metrics can be acquired from an ANOVA test to assess the statistical significance in greater depth: 1) degree of freedom, 2) F-Statistic, and 3) p-value.

First, degree of freedom (Df) can be calculated from the sample size and the number of groups that are being tested for, which is the level of gaming experience in our case. Df_1 is the degree of freedom for group means: the number of group means that can vary to produce the grand mean³. In other words, with n groups, there can at most be $n - 1$ group means that are free to vary, from which can be deduced that $Df_1 = n - 1$. Df_2 , on the other hand, refers to the degree of freedom for individual cell data. And given the total number of individual cell data (N) and the total degrees of freedom lost, we can get Df_2 by subtracting $n - 1$ from N . Second, F-Statistic compare the variance between group means (v_b) and the variance within the groups (v_w) and can be computed by dividing v_b from v_w . A significantly larger v_b in

3. the mean value of the entire sample

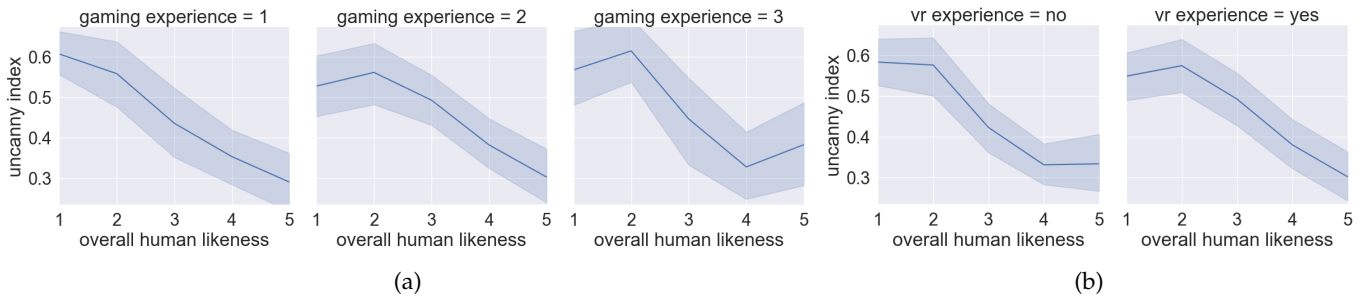


Figure 6:

(a) Series of relational trend observed in 2 different groups: participants **with** and **without** prior AR/VR experiences.

(b) Series of relational trend observed in 3 different groups respectively: participants with **minimal**, **moderate**, **maximal** gaming experiences.

comparison to v_w (in other word, a large F-Statistic value) suggests that the variation between the group means did not happen by chance, allowing us more evidences to revoke the null hypothesis. This conversely means that we lack ground to revoke the null hypothesis when the F-Statistic value is low. Third, p value represents the area to the right of F-Statistic under F-distribution curve, representing the probability of observing different group means if the null hypothesis were true.

Human Likeness	(Df ₁ , Df ₂)	F-Statistic	p
1	(2, 65)	1.45×10^1	0.241
2	(2, 67)	4.62×10^{-1}	0.632
3	(2, 63)	6.01×10^{-1}	0.551
4	(2, 67)	5.10×10^{-1}	0.603
5	(2, 73)	1.12×10^1	0.332

TABLE 3

However, for all avatars, $p > 0.15$ and low F-Statistic values suggests no significant effect of users' gaming experiences on the avatar's uncanny index was observed with $p > 0.15$ for the three conditions.

However, this could be due to the large variance within the sample groups. As described earlier, F-Statistic value is computed by dividing the variance between the group means from the variance within the groups. If the data within each group varied significantly, F-Statistic can be compromised. That said, it would be helpful to review the relational trend graphs (human likeness level vs. uncanny index) for each group as illustrated in Figure 6a to assess the potential effect users' gaming experiences can have on the avatars' induced uncanniness that were potentially overlooked by the ANOVA test.

6.3.2 AR/VR Experience and Its Effect on Uncanny Valley

A one way ANOVA was conducted for the 5 levels of human likeness to compare the effect of AR/VR experience on uncanny index in 2 conditions: with and without prior VR/AR experiences.(The mean uncanny index value for

each level of gaming experience will be furthered referred to as μ)

The null hypothesis to be tested is $\mu_{no} = \mu_{yes}$: that is, the group means for samples with and without prior VR/AR experiences not different. A p-value below the alpha level of 0.15 and a high F-Statistic value should allow us to revoke this null hypothesis.

Human Likeness	(Df ₁ , Df ₂)	F-Statistic	p
1	(1, 66)	1.25×10^1	0.267
2	(1, 68)	6.84×10^{-1}	0.685
3	(1, 64)	7.04×10^{-2}	0.792
4	(1, 68)	1.15×10^{-1}	0.736
5	(1, 74)	1.74×10^1	0.191

TABLE 4

For all avatars, no significant effect of prior VR/AR experiences on uncanny index was observed with $p > 0.15$ for the two conditions. However, as described in 6.3.1, the low F-Statistic values could have been caused by the large variances within each group.

6.4 EEG Sensor Results

Alpha asymmetry values were computed from Muse raw data in order to determine what mood the study participants were in when viewing the avatars. It is acknowledged that the ground level asymmetry can vary across different people; so, in order to equalize each participants' asymmetry values, the asymmetry data were subtracted from a mean asymmetry value individually calculated for each participant. Then, for each avatar, the alpha asymmetry data were divided into 4 distinct segments defined by the avatar's animation state, from which a minimum alpha asymmetry was extracted to represent each segment's defining negative moment. The trend graphs in Figure 7 was obtained by plotting these "defining" asymmetry values for each avatar, each participant, and each animation segments.

The animation segments can largely be categorized into two types: 1) body animation and 2) facial animation. Figure

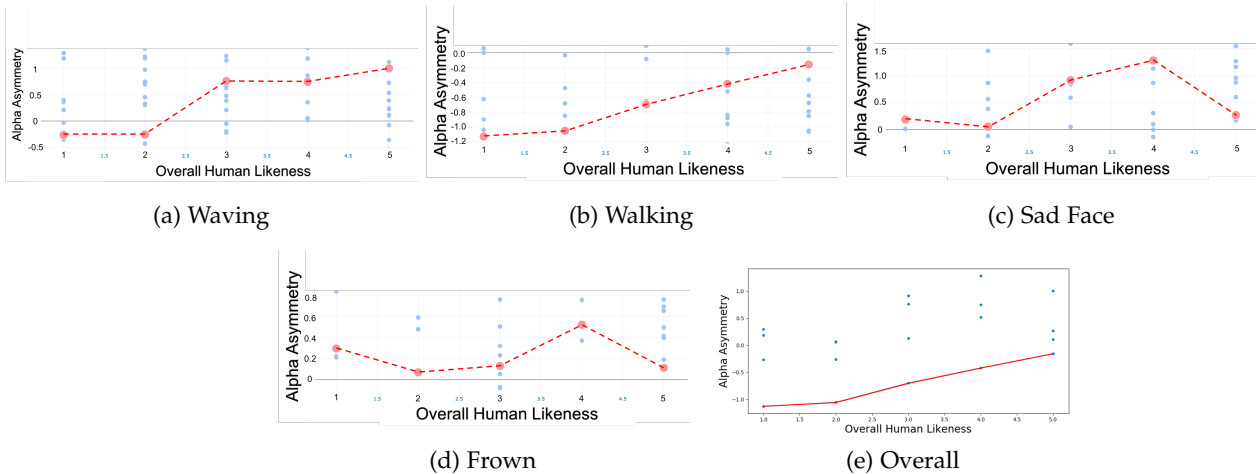


Figure 7: Frontal Alpha Asymmetry trend graph for animation segments: (a) **Waving**, (b) **Walking**, (c) **Sad face**, and (d) **Frowning**. Lastly, (e) the mean asymmetry for the 4 segments were accrued and plotted to show the **overall** trend.



Figure 8: Skin conductance response peak index graph for each animation sequence: (a) **Waving**, (b) **Walking**, (c) **Sad face**, and (d) **Frowning**

7a and 7b are trend graphs for body animation segments and figure 7c and 7d are trend graphs for facial animation segments. And there are noticeable differences in graph trend between the two sets of animation segments that I believe is worth discussing.

The body animation segments (Figure 7b and 7a) captures a monotonic rise in the frontal alpha asymmetry along with the increases in avatars' human likeness. This suggests that the more human like avatars triggered more positive emotional feedback from their viewers than the less human like avatars. This relationship is consistent with our observation from the survey results and poses no conflict to our finding.

However, the facial animation segments (Figure 7c and 7d) illustrates a very different trend, where the monotonic rise in the frontal alpha asymmetry is brought to a sudden drop at human likeness level of 5. This indicates that there is something that makes facial animation highly compelling in highly human like avatars; as MacDorman suggests, this may be due to the realism inconsistency between the characters' physical resemblance to an actual human and their imperfect facial animation [6]. This is analogous to how a new dirt spot really stands out on a clean shirt; the new dirt spot would have not been so noticeable if it was on a dirty shirt full of previous dirt spots but it is because the shirt is so clean, the discrepancy becomes highly noticeable, bringing the new dirt spot to attention. Likewise, it could be because the hyper-realistic model highly resembles human, its imperfect facial animation stands out, causing its viewers to feel unsettled by it.

6.5 GSR Data Result

GSR was measured in hopes of finding out what avatar triggers the most emotional arousal. The participants' GSR data were divided into 4 segments based on the avatars' animation state, for each of which was determined if there was a skin conductance peak. The presence of a skin conductance peak in an animation segment would indicate that the animation event raised an emotional arousal. So, for each avatar and their segments, the number of skin conductance peaks were tallied to produce an interval value (0-1) in representation of the likelihood of an emotional arousal for each animation segment. An interval value of

1 would indicate that the animation segment for avatar A had evoked an emotional arousal from all the participants, and an interval value of 0 would indicate that it failed to evoke an emotional arousal from all the participants.

For all figures, we observe a general V-shape trend. This may seem inconsistent with our survey and EEG results, but we have to take into account that GSR is incapable of capturing whether the arousal is for a positive or a negative emotion. That is, our figures 8a, 8b, 8d, and 8c have to be supplemented with our EEG result to produce a meaningful index for the uncanny valley.

From our alpha asymmetry analysis, we have learned that more positive emotions are evoked from the observers when viewing more human like avatars than when viewing less human like avatars. We could align these findings with our GSR results and infer that the first part of the V (the declining slope) represents negative emotional arousal and that the second part of the v (the inclining slope) represents positive emotional arousal.

6.5.1 AR Platform vs. 2D Monitor

To get a higher-level understanding of how people's perception of anthropomorphic entities in AR and non-AR platforms differ, the study participants were viewed the same set of characters on both the HoloLens device and a computer monitor and were asked on whether they perceived any palpable difference between the two viewing platforms. The survey result showed that 96% of study participants thought viewing the characters in AR was more uncanny than viewing them on 2D monitors. And the remaining 4% of the participants claimed that there was no discernible difference.

7 DISCUSSION

The results seem to be in denial of the expected uncanny valley effect: there was no observable spike in uncanny index with the increase of human likeness level or facial anthropomorphic level. Simply stated, the results indicate that in AR setting, the more human like the characters are, the less uncanny they are perceived. Then, is uncanny valley not present in AR? Or is it located somewhere else? How should we make sense of the resulting data?

7.1 Study Limitations and Their Effect on the Study Result

Maybe the inherent limitations of this study may not have allowed us to reach the uncanny valley at all. As previously discussed, Microsoft HoloLens is a standalone AR device with limited processing / graphic capability. Therefore, there is a limit to number of polygon counts the device can render, which prevented us from using hyper-realistic model with high polygon counts. So, we cannot exclude the possibility that the uncanny valley simply could not be reached with the level of human likeness that was presentable with Microsoft HoloLens. [19]

However, the result looked fairly different in a VR version of the uncanny valley study that shares almost the same selection of the 3D characters used in this study. Participants in the VR study perceived the most human like character to be the most uncanny out of all. Unlike the trend observed in the AR study, where the peak in uncanny index was observed at a low level of human likeness followed by a monotonic fall, the trend observed in the VR study similes Mori's originally suggested graph, where the peak is observed at a higher level of human likeness. So what could be differentiating the results of the two studies?

7.2 Realism Inconsistency in Augmented Reality

There is a possibility that some AR specific components have shifted the uncanny valley to the left: to a lower level of human likeness. That is, we may have observed uncanny valley effect in this study after all but at a much lower level of human likeness level than expected. And realism inconsistency may have had a role in shifting the uncanny valley when viewing anthropomorphic figures in AR setting. Although MacDorman's take on realism inconsistency discusses the effect differing level of realism within an entity can have on the entity's uncanniness [6], this concept maybe can be extended to the differing levels of realism between an entity and its surrounding environment. Unlike most conventional viewing platforms, AR technology blends a virtual entity with a real physical environment; the AR device renders a computer generated images on top of users' visual perception of the physical environment, and allows virtual entities to break away from its virtual scenes to enter the users' reality. This may cause a cognitive dissonance in viewers to some level.

The viewers likely adjust their expectation on the level of realism of entities they expect to encounter based on what the entities' surrounding environments are. For instance, in a virtual 3D scenes, users will be inclined to expect a virtual 3D entities while in a real physical environment, they will be inclined to expect real entities not virtual ones. That said, this expectation dissonance may be causing the uncanny valley effect to appear at a lower level of realism in AR. The characters with lower human likeness level are likely to cause more dissonance with the real physical spaces it is overlaid on top of than the characters with higher human likeness level. And their higher realism inconsistency leads to a larger expectation let down, invoking a negative impression.

This is also in line with the users' responses to the survey question on viewing the characters in AR vs. 2D monitor.

The realism inconsistency could be accentuating and amplifying the perceived uncanniness of the characters with low human likeness level in AR, making the participants to believe that viewing the characters in AR scenes was more uncanny than viewing them in 2D monitor.

7.3 Alternate Effect of Realism Inconsistency

Or maybe realism inconsistency between entities and their surrounding environment do not trigger an adverse impression on the entities. Realism inconsistency may be raising users' standards in evaluating the 'realness' of the entities, pushing the uncanny valley further to the right. That is, the characters that were perceived real in other platforms may no longer pass the 'test' and be perceived real in AR. Uncanny valley effect describes a trend where entities' uncanniness continues to decrease until they achieves a certain level of human likeness, where it rebounds, and if this entire graph has been shifted to the right because of the viewers' altered standards in AR setting, what we observed in this study can be the first part of the uncanny valley trend before reaching that rebound point.

7.4 Limitations and Possible Resolutions

7.4.1 Hardware Limitation

Augmented reality devices are still early in their development phase and currently available devices have with them significant technical limitations that get in the way of an immersive experience. Among all, narrow field of view (FOV) and restrained opacity were the two major obstacles to conducting this study.

Field of View Field of view represents the range of vision sight, to which an image overlay can be applied. The narrower FOV is, the less area the overlay can be drawn on top of. Currently, most commercial AR devices have FOV of $30^\circ \times 17.5^\circ$, which covers a very minute portion of humans FOV, which is conventionally known to be $180^\circ \times 110^\circ$. Most users testify that their AR experience similes that of watching through a narrowly framed window.

Such limitation makes a crucial problem to the research as it becomes difficult for the users to get a full view of the avatars. The avatars used for the study are designed to be in life-size, and therefore will not fit into the devices FOV when the avatar is positioned close to the study participant.

This potentially could be resolved by using the 2nd generation HoloLens device, which offers a twice the FOV ($43^\circ \times 29^\circ$) the first generation device offered [20]. However, this is still very small in comparison to human FOV and will still not provide us with a full view of the avatar.

An alternate solution would be to replicate/mimic the AR experience using a VR device by appending a stereoscopic camera in front for a live visual feed of the external physical environment. There actually is a commercial stereo camera called Zed Mini that can be attached to a VR device [21]; the camera offers a video pass-through with FOV of $85^\circ \times 54^\circ$ in 720p resolution. The low resolution of this device could get in the way of a fully immersive experience, but this option will be able to provide a full frontal view of the avatars.

Opacity The second major limitation is the opacity of the augmented portrayal. Currently most commercial AR

devices project beams onto layers of holographic lenses, allowing the users to see the beamed images on top of their vision. However, holographic projection is extremely vulnerable to external light source, meaning that if the device was worn at a bright environment with lots of external light source, the opacity of holographic projection can be seriously diminished, thereby compromising the immersivity of the AR experience.

This problem was addressed in this study by viewing the avatars in a low ambient light environment, but the lack of opacity can be further remedied by using a VR device with stereo camera attachment as mentioned earlier.

7.4.2 Study Design Limitations

The biggest challenge in studying uncanny valley is the difficulty in assessing and quantifying the uncanniness and human likeness of the characters. This study attempted at resolving this difficulty by using both a subjective self-reported human likeness score and objectively evaluable facial anthropomorphism and body proportion scores to assess the characters' human likeness and using both an ordinal and an interval score to rate the characters' uncanniness. However, this method also introduces new problems and challenges. This method requires a larger set of avatars to represent all the different combinations of varying facial anthropomorphic and body proportion level. In case of our study, as there are total 7 levels of facial anthropomorphism and body proportion, total of 49 avatars would be necessary to fully represent the plane. However, this study has 7 avatars in total, representing only 14% of the total possible combinations, which can result in a skewed representation of the relationship between a character's facial anthropomorphic level vs. the uncanniness and its body proportion level vs. the uncanniness.

This brings us to the second problem: the characters used for this study were vastly different in their looks (facial features, hair shape, color, etc.). This essentially introduces too many unknown variables to assess. Let's say that A and B are two 3D rendered anthropomorphic characters, A is more facially anthropomorphic than B but they are also different in multiple other attributes. Then, if character A was perceived to be more uncanny than character B, what assures that this is due to A's higher facial anthropomorphism? It becomes harder and harder to assess the cause and effect relationship when there are too many variables to handle and take into consideration.

That said, the result would be more statistically significant if the study uses a single 3D character and modify that one character's facial anthropomorphic level and body proportion level uniformly.

8 CONCLUSION

As the title suggests, this study examines the uncanny valley in augmented reality. A Microsoft HoloLens application was designed and built to present 7 3D CGI characters with unique facial anthropomorphic and body proportion levels to examine peoples' reactions to avatars with varying human likeness. The participants' responses were collected in both a survey form and biometric sensor data (EEG, GSR). The result shows that uncanny valley in augmented reality

is, in fact, against the convention: the survey results and frontal alpha asymmetry results suggest that characters on lower-end of the human likeness level are the least liked and the most uncanny while the characters on the higher-end are deemed more likeable and less uncanny in AR setting. That is, the uncanny valley either happens either much earlier or much later than where it would normally happen in other non-AR platforms. I suspect that the cause lies in AR's capability of blending the virtual and the real. As virtual entities are brought to a real physical space, they likely arouse an expectation dissonance within the viewers. That said, the less human like characters will have a greater realism inconsistency with the physical environment than the more human like ones, causing the uncanny valley to appear at a low-end of human likeness level. Alternatively, the realism inconsistency between the virtual entities and the physical environment might raise the viewers' standards in evaluating entities' human likeness, and a virtual entity that was deemed human like in a virtual space may no longer be deemed human like in the real space, causing the uncanny valley to appear at a much higher level of human likeness.

9 ACKNOWLEDGEMENT

There is an overdue appreciation I owe to Professor Zhou and Professor Loeb for their persistent guidance throughout the entire study. Having a weekly discussion with them allowed me a great chance to learn how to view matters in a widely different perspective. And I thank Professor Mahoney for insightful advices and suggestions!

Moreover, I would like to acknowledge Ting Yan for his wonderful work with the bio-metrics sensors!

I deem this to be one of the greatest culminating experiences. And honestly, the studies were a novel joy with daily interaction with different students. They brought me insights and passion.

REFERENCES

- [1] M. Mori, K. Macdorman, and N. Kageki, "The uncanny valley [from the field]," *IEEE Robotics Automation Magazine*, vol. 19, no. 2, pp. 98–100, 2012.
- [2] T. Suo, L. Liu, C. Chen, and E. Zhang, "The functional role of individual-alpha based frontal asymmetry in the evaluation of emotional pictures: Evidence from event-related potentials," *Frontiers in Psychiatry*, vol. 8, Sep 2017.
- [3] C. Bartneck, D. Kulic, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *International Journal of Social Robotics*, vol. 1, no. 1, pp. 71–81, 2009. [Online]. Available: <http://www.bartneck.de/publications/2009/measurementInstrumentsRobots/>
- [4] C.-C. Ho and K. F. Macdorman, "Revisiting the uncanny valley theory: Developing and validating an alternative to the godspeed indices," *Computers in Human Behavior*, vol. 26, no. 6, pp. 1508–1518, Jun 2010.
- [5] —, "Measuring the uncanny valley effect," *International Journal of Social Robotics*, vol. 9, no. 1, pp. 129–139, Sep 2016.
- [6] K. F. Macdorman and D. Chattopadhyay, "Reducing consistency in human realism increases the uncanny valley effect; increasing category uncertainty does not," *Cognition*, vol. 146, pp. 190–205, Oct 2015.
- [7] K. Gray and D. M. Wegner, "Feeling robots and human zombies: Mind perception and the uncanny valley," *Cognition*, vol. 125, no. 1, pp. 125–130, Jul 2012.
- [8] L. Piwek, L. S. Mckay, and F. E. Pollick, "Empirical evaluation of the uncanny valley hypothesis fails to confirm the predicted effect of motion," *Cognition*, vol. 130, no. 3, pp. 271–277, Nov 2013.
- [9] Keveleigh, "Mr spatial 230 - spatial mapping - mixed reality." [Online]. Available: <https://docs.microsoft.com/en-us/windows/mixed-reality/holograms-230>
- [10] R. G. Brereton, "Introduction to analysis of variance," *Journal of Chemometrics*, vol. 33, no. 1, Jan 2019.
- [11] "statistical data visualization." [Online]. Available: <https://seaborn.pydata.org/>
- [12] Muse, "Hardware specifications." [Online]. Available: <http://developer.choosemuse.com/hardware-firmware/hardware-specifications>
- [13] M. Liu, D. Fan, X. Zhang, and X. Gong, "Human emotion recognition based on galvanic skin response signal feature selection and svm," *2016 International Conference on Smart City and Systems Engineering (ICSCSE)*, Nov 2016.
- [14] Hkosney, "Mindfield esense skin response." [Online]. Available: <https://www.mindfield.de/en/Biofeedback/Products/Mindfield-eSense-Skin-Response.html>
- [15] "Pyplot tutorial." [Online]. Available: <https://matplotlib.org/3.1.0/tutorials/introductory/pyplot.html>
- [16] "Documentation of scikit-learn 0.21.2." [Online]. Available: <https://scikit-learn.org/stable/documentation.html>
- [17]
- [18] "Welcome to statsmodels's documentation." [Online]. Available: <https://www.statsmodels.org/stable/index.html>
- [19] A. Tinwell, M. Grimshaw, D. A. Nabi, and A. Williams, "Facial expression of emotion and perception of the uncanny valley in virtual characters," *Computers in Human Behavior*, vol. 27, no. 2, pp. 741–749, Nov 2010.
- [20] D. Heaney, "Hololens 2's real field of view revealed," Feb 2019. [Online]. Available: <https://uploadvr.com/hololens-2-field-of-view/>
- [21] "Getting started - introduction." [Online]. Available: <https://www.stereolabs.com/docs/getting-started/>
- [22] T. J. Burleigh, J. R. Schoenherr, and G. L. Lacroix, "Does the uncanny valley exist? : an empirical test of the relationship between eeriness and the human likeness of digitally created faces," *Computers in Human Behavior*, vol. 29, pp. 759–771, Jan 2013.
- [23] M. Cheetham, "The uncanny valley hypothesis and beyond," *Frontiers In Psychology*, vol. 8, Oct 2017.
- [24] J.-P. Stein and P. Ohler, "Venturing into the uncanny valley of mindthe influence of mind attribution on the acceptance of human-like characters in a virtual reality setting," *Cognition*, vol. 160, pp. 43–50, Dec 2016.
- [25] "3d model store." [Online]. Available: <https://www.cgtrader.com/>
- [26] "Mixamo." [Online]. Available: <https://www.mixamo.com/>
- [27] MSFT, "Hololens hardware details - mixed reality." [Online]. Available: <https://docs.microsoft.com/en-us/windows/mixed-reality/hololens-hardware-details>