Evaluating Digital Avatars - A Systematic Approach to Quantify the Uncanny Valley Effect by Using Real Life Samples

Hakan Arda
Faculty of Computer Science and Business Information Systems
Technical University of Applied Sciences Würzburg-Schweinfurt
Würzburg, Germany
Hakan.Arda@thws.de

Andreas Henneberger
Faculty of Computer Science and Business Information Systems
Technical University of Applied Sciences Würzburg-Schweinfurt
Würzburg, Germany
Andreas.Henneberger@study.thws.de

Karsten Huffstadt
Faculty of Computer Science and Business Information Systems
Technical University of Applied Sciences Würzburg-Schweinfurt
Würzburg, Germany
Karsten.Huffstadt@thws.de

Nicholas Müller
Faculty of Computer Science and Business Information Systems
Technical University of Applied Sciences Würzburg-Schweinfurt
Würzburg, Germany
Nicholas.Mueller@thws.de

***Abstract*—Virtual reality has seen significant advancements in recent years, particularly in rendering nearly perfect replicas of real-world objects. These improvements in environmental realism enhance user immersion. However, increasing the realism of human-like avatars seems to have the opposite effect, often leading to negative emotional reactions and breaking immersion. This is explained by the uncanny valley curve, where subtle deviations in lifelike avatars can evoke discomfort. In this study, we developed a method to evaluate human-like avatars based on the uncanny valley curve, aiming to pinpoint where this discomfort originates. A database of over 200 avatar images was created, and studies were used to identify key characteristics that make these avatars resemble humans. Additionally, we conducted a follow-up experiment with 30 participants, where real avatars were developed based on our findings and tested in live, direct conversations. This allowed us to validate our approach and offer a more precise method for future research on avatar evaluation and uncanny valley effects. Our results open up new avenues for refining avatar design, potentially mitigating negative reactions and increasing overall immersion in virtual reality experiences.**

Keywords—Uncanny valley; human-like avatars; human-likeness; database; virtual reality.

I. Introduction

Based on the findings of the previous study [1], in this journal article we want to delve deeper into the Uncanny Valley and show which criteria are particularly relevant to overcome it. The Uncanny Valley effect is a psychological phenomenon that describes the unease or discomfort people experience when encountering human-like entities that are almost, but not quite, convincingly realistic. The term "Uncanny Valley" was coined by robotics professor Masahiro Mori in 1970 [2]. The concept suggests that as the appearance or behavior of humanoid entities becomes increasingly close to human-like, there is a point at which they elicit a strong negative emotional response before eventually be-coming indistinguishable from real humans.

There have already been many far-reaching attempts to investigate the effects of the human likeness of robots on people's emotional reaction [3–9]. One example of this is the work of Kim et al. [7] who used the open-source Anthropomorphic RoBOT (ABOT) database to analyze the human similarity of 251 robots. They asked a group of 150 participants to rate images of robots from the ABOT database according to their human likeness and uncanny valley factor. With the results of this survey, they have found evidence of Mori's Uncanny Valley [2]. This valley was evident in participants' perceived uncanniness of 251 robots that varied widely in terms of the range and characteristics of human likeness. They also found evidence of another, second valley of uncanniness in robots that showed a moderately weak resemblance to humans.

The developers of the ABOT [4] database took a similar approach in their study, providing a basis for research in this area. The researchers found that the human-like appearance of robots can be divided into three dimensions of human-like appearance: the robots' surface features (e.g., skin, hair, clothing), the main components of the robots' body manipulators (e.g., torso, arms, legs) and the robots' facial features (e.g., eyes, mouth, face) [7]. These results suggest that the overall perception of the physical human-likeness of robots and its relationship to emotional reactions to the robots can be explained by different constellations of the three human-like appearance dimensions. If the hypothesized uncanny valley phenomenon could be understood at the level of specific human-like appearances, this could also lead to the improvement of virtual avatars. A deeper understanding of the uncanny valley could help developers fine-tune the design of avatars to evoke more positive emotional responses from users, enhancing user interaction and engagement. from users.

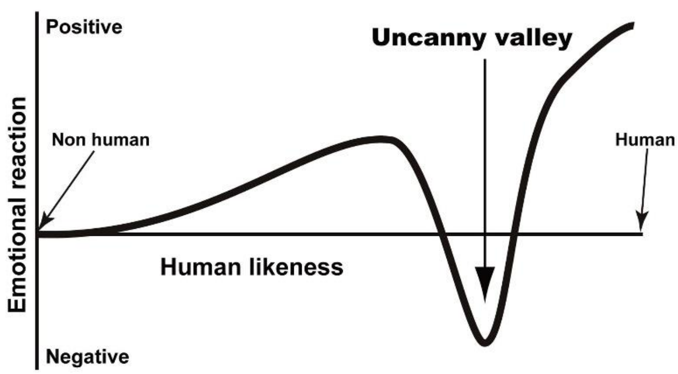

Fig. 1. Graphical representation of the Uncanny Valley. Retrieved from [10].

This approach of using a large database such as ABOT to test different human-like robots for their Uncanny Valley factors does not exist in relation to digital avatars currently. One reason for this could be the lack of such a database and the associated basis for the resulting research. Another reason is that there is currently no systematic, evidence-based approach for categorizing avatars into a continuum of perceived human-likeness. Consequently, researchers and designers are usually forced to rely on heuristics and intuitions when it comes to selecting human-like avatars for studies or developing human-like features in avatar design. This approach faces several problems. Firstly, there is currently no quantitative system to describe the degree of human likeness in different avatars, which makes it difficult to compare research results between different studies. Therefore, a precise scale is needed to compare different avatars on a common scale and allow researchers to replicate their results with their avatars.

Secondly, even when researchers manage to quantitatively assess the impression people have of an avatar's appearance, they usually treat the concept of "human likeness" as one-dimensional. However, human likeness can be expressed in different ways. For example, through gestures and facial expressions or, more generally, through the mere presence of arms and legs. Humanity has many different characteristics and therefore also different features that need to be considered.

Thirdly, the effects of the appearance of avatars on different types of avatars must be considered in the investigations. While it may be of practical advantage to limit yourself to a certain type of avatar, such as simplified human likenesses like the ones Meta uses in her Horizon. However, these restrictions can mean that certain small differences between avatars are lost. And the psychological conclusions of this work would not be applicable to a variety of other avatars.

In addition, previous studies have extensively explored methods to overcome the uncanny valley in character design. For example, the work by Schwind et al. examined how atypicality's (strong deviations from the human norm) for high levels of realism cause negative sensations in humans and animals [11, 12]. The negative effects of atypical features, such as unnaturally large eyes or human emotions in realistic animals, are stronger for more realistic characters, than for characters with reduced realism. Consistently rendered realism can reduce the negative effects of atypicality and increase affinity, as shown by the first peak in the uncanny valley. Therefore, it is important to avoid combining realistic renderings and detailed textures of skin or eyes with non-human-like features. At high levels of realism, atypical features can cause the uncanny valley. Another possibility is to avoid so-called "dead eyes". A virtual character's eyes are crucial in determining its realism. The work of Schwind et al., which used eye tracking, found that users first fixate on the eyes before assessing other features. This is consistent with previous research showing that the realism and inconsistencies of human characters are primarily judged based on the realism of their eyes [13]. This also explains why skin makeup does not affect animacy, unlike atypical eyes or the eyes of a deceased person. The symptom of 'dead eyes' can make artificial characters feel eerie and strange. The eyes communicate intentions, behavior, and well-being, which are essential for assessing and creating affinity for the depiction.

To address these issues, and to provide a way to conduct more systematic, general, and repeatable research on virtual human-like avatars, this thesis has developed an avatar database based on the findings of Phillips et al. thus continuing the research on the uncanny valley. Furthermore, the results of this study were used to create specific categories for the creation of real avatars. In a follow-up experiment, the newly created avatars were evaluated using the same questions and criteria as in the original study, allowing for a direct comparison between the real avatars and the baseline data.

II. METHOD

A. The Avatar Database

Similar to the ABOT database by Phillips et al., the avatar database pursues three goals. Firstly, it offers an overview of the broad landscape of different human-like avatars. Secondly, the avatar database provides standardized images of human-like avatars and a growing dataset of people's perceptions of these avatars, both of which will be made public for further research in the future. Thirdly, the Avatar Database can help us better understand what makes an avatar appear human. To begin, we will discuss the development of the database. Then, two empirical studies will identify various dimensions of avatar appearance and determine which of these dimensions contribute to the perception of overall human-likeness of avatars. Subsequently, a further empirical study on the validation of the database is presented and a possibility for future research is introduced.

B. The Development of the Avatar Database

To create a comprehensive database of human-like avatar images, we searched for as many avatars as possible with the required human-like appearance characteristics. The avatars were identified from various sources, such as game characters (e.g., Fortnite, Final Fantasy, Skyrim), technology-oriented media (e.g., online magazines, newsletters, social media), companies and university websites, online communities and discussion forums, and general Google searches. To identify avatars that had not yet received significant media attention, we also created our own avatars based on real humans and fictional characters. Between January 2024 and April 2024, we created an initial collection of over 200 images of human-like avatars, each with one or more different characteristics.

Next, we reviewed the collection of avatars and removed the ones that are already represented in the same or similar enough form. In addition, avatars that are similar to animals have also been removed because this study explicitly looks at human avatars only and uses mixed avatars later as a safety checkup. Each image also had to fulfil a certain standard to be included in the collection: No obstacles, no motion blur, no groups pictures, in color, and the entire body (With feet, hands, and head). In addition, all pictures were taken one more time in a close-up of the face. This allowed the participants to examine the entire body and then explicitly the face for the individual features. Any image that did not fulfil this standard was either not included in the collection or removed accordingly. Based on this approach we have created a collection of 200 avatars, with the corresponding source for downloading the avatars.

It was important to us that we cover as many different types as possible with the large number of avatars, such as cartoonish, stylized, realistic or minimalist. In addition, it should be possible to find an avatar that is reduced to the desired characteristics with the help of filters. This is a similar concept to searching for robots in the ABOT database. With the difference that here the avatars can be downloaded with the help of the source and used for further purposes. Some of the characters were also tested with the help of the unity engine and put to the test for suitable images. This was especially the case with avatars that we developed ourselves for this collection.

The images in the database were sorted and edited to ensure uniform recognition. This was done to ensure consistency in the images. Avatars were photographed in a frontal and neutral position with a neutral facial expression whenever possible. For avatars where this was not possible, the model was rendered again in the unity engine and photographed accordingly. Finally, the images were cropped to just the avatars with a white background using photoshop and tagged with different tags for better analysis. Here, for example, attention was paid to the recognizable gender, potential age, art style and source.

C. Measuring the Human-Likeness

In order to accurately determine the degree of humanity in avatars, the individual avatars must be evaluated according to clearly defined characteristics. Because we are dealing here with human-like avatars and not anthropomorphic robots, we are unfortunately unable to use the results of the work by Phillip et al. [4], but we can use a similar approach to determine the characteristics that we will use later. We rely on a bottom-up, feature-based approach and base our expectations on the results of the work of Phillip et al. Our goal with this approach was to define the individual features that constituted humanity in avatars and then bundle them together.

To determine the appearance characteristics of our avatars, we have developed a collection of possible characteristics that speak for humanity. We then checked all the images for the respective characteristics and deleted any that did not fit. This included features that were rare, repeated, or confusing. As a result, we defined 16 characteristics that we used for our further procedure. In addition, like Phillip et al., we decided to contribute definitions for the features. We started with relevant definitions from the Oxford English dictionary and adapted them according to our application. For example, we were able to retain the biological functions for features such as "mouth" because our test objects are not robots but human-like avatars. This resulted in a table of features and their definitions. These definitions served as a way for our participants to focus on certain characteristics when evaluating the avatars. Since they are human-like avatars and not robots, all characteristics are always present in some form. They only differ in their design. However, there are also avatars that do not have smooth white skin and are instead green with lots of dots. Therefore, the question here is not whether the respective characteristic is present, but rather to what degree it stands out.

TABLE 1. COLLECTION OF APPEARANCE FEATURES AND ASSOCIATED DEFINITIONS

Feature	***Definition***
Arm	The upper limb of the human body, or the part of the upper limb between the shoulder and the wrist.
Eye	The organ of sight. Either of the paired globular organs of sight in the head of humans.
Eyebrow	The (usually arched) line of short fine hair along the upper edge of each of a person’s eye sockets.
Eyelashes	The line of hairs fringing each edge of an eyelid, serving to help keep the eye free of dust or other extraneous matter.
Face	The front part of the head, from the forehead to the chin, and containing the eyes, nose, and mouth.
Finger	Each of the five slender jointed parts attached to either hand.
Genderedness	Features of appearance that cand indicate biological sex, or the social categories of being male or female.
Hand	The terminal part of an arm, typically connected to the arm by a wrist.
Head	The uppermost part of a body, typically connected to the torso by a neck. The head may contain facial features such as the mouth, eyes, or nose.
Head hair	A collection of threadlike filaments on the head.
Leg	The lower limb of the human body, or the part of the lower limb between the hip and the ankle.
Mouth	The orifice in the head of a human or other vertebrate through which food is ingested and vocal sounds emitted.
Nose	The part of the head or face in humans which lies above the mouth and contains the nostrils.
Skin	The layer of tissue forming the external covering of the body.

Fig. 2. All 200 human-like vr avatars in the database.

D. Measuring the Uncanniness

To address the question of how uncanny the avatars appeared, we followed the methodology outlined in the study by Kim et al. [7]. Their research involved a large-scale investigation utilizing the ABOT database, building on the foundational findings of Phillip et al. [4]. Kim et al. conducted a follow-up study that specifically compared robots based on their levels of human-likeness and perceived uncanniness. Drawing inspiration from their approach, we adopted the same definition of uncanniness to ensure consistency and standardization in our evaluation process. To help participants understand and apply a uniform criterion during the evaluation, we provided them with a clear definition of uncanniness. Specifically, they were informed that uncanniness refers to "the characteristic of seeming mysterious, weird, uncomfortably strange, or unfamiliar." This definition was derived directly from the Oxford English Dictionary's explanation of the term, ensuring its alignment with established and conceptual standards. By using this definition, we aimed to create a common framework to assess the avatars' uncanniness systematically.

E. Real Life Avatar Samples

For the creation, certain criteria were considered that were based on the results of the first study. These categories are mentioned in the results section.

There are numerous methods for the creation of a lifelike or stylized digital persona. One method is to scan the user in their entirety in order to create a lifelike avatar (Fig. 3, far right). The process of creating virtual avatars through scanning involves capturing the physical features of real individuals and translating them into digital representations. This process is primarily reliant on the utilization of cameras and bespoke software in order to achieve accurate and realistic virtual avatars. In the initial stage of the process, an iPhone 13 Pro camera and the Polycam application were employed. The camera of the iPhone 13 Pro is equipped with the requisite technical capabilities for the capture of 3D scans. In the data capture phase, this camera is employed to record information from the subject. The cameras of the iPhone 13 Pro range from conventional RGB (red, green, blue) cameras to advanced depth-sensing cameras such as LiDAR (light detection and ranging) or structured light cameras. Their function is to capture a range of characteristics of the individual, including their appearance, shape, and texture.

Depth-sensing cameras are of particular significance, as they provide data regarding the spatial distance between the camera and various points on the individual's body. This depth data serves as the basis for the construction of a three-dimensional representation of the individual. Concurrently, RGB cameras capture color data, which is indispensable for texturing the 3D model and endowing it with visually realistic details. To ensure accuracy, the cameras are calibrated and aligned. Calibration ensures that the cameras are correctly configured and calibrated, whereas alignment determines the precise positions and orientations of the cameras in relation to one another. Such meticulous calibration is of paramount importance for ensuring the seamless integration of data from disparate camera perspectives. The subsequent phase is to utilize the data obtained from the camera in the Polycam application. The application utilizes the data to generate a "point cloud," which is a set of 3D points in space that collectively delineate the surface of the person.

Fig. 3. Real life avatar samples. From left to right; the generic, - stylized, - mixed, - and lifelike avatar.

This point cloud represents the initial representation of the person's physical form. Subsequently, the point cloud is transformed into a "mesh", which is a network of interconnected triangles that collectively create a coherent three-dimensional surface. The resulting mesh provides a more detailed and accurate representation of the person's body. In order to imbue the virtual avatar with realistic visual details, texture information derived from the RGB images is projected onto the mesh, thereby imparting the avatar with characteristics such as skin color, clothing patterns, and other visual attributes. In order to optimize the final result, the mesh was subjected to a cleaning and refinement process. The objective of these steps is to eliminate imperfections, smooth surfaces, and eliminate any undesired artefacts that may have emerged during the earlier stages.

In order to facilitate dynamic poses and movements, it is necessary to provide the virtual avatar with a skeletal structure. The process of rigging involves the creation of a digital skeletal structure with joints and bones, which emulates the human anatomy. This stage was completed using the Mixamo platform. This software is a free product from Adobe and offers a plethora of animations, avatars, and possibilities for rigging. In rigging, the software is employed to construct a skeleton or rig of bones or joints that determines the manner in which the constituent parts of a mesh can be moved. The resulting skeleton is then connected to the avatar, thus enabling the latter to perform animations and poses that mirror real-world actions. Once the avatar has been rigged, it becomes possible to animate it through various methods, including the utilization of motion capture data or virtual reality headsets. This animation process imbues the avatar with a lifelike quality, enabling it to replicate the movements and expressions of the real person on whom it was based. The subsequent stage is to incorporate the avatar into the Unity environment.

In this instance, the avatar is furnished with an animation controller. The controller oversees the transitions between disparate animations, thereby enabling the user to regulate the animations in accordance with parameters such as movement, gestures, and interactions. In order for the headset to transmit these parameters, the head mounted display (HMD) must be integrated into Unity with the assistance of a VR software development kit (SDK). A variety of HMD manufacturers, including Meta, HTC Vive, and Steam, provide their own SDKs. However, there is also an open-source variant, OpenXR, which promises compatibility across all headsets. Integration of OpenXR into existing projects is a straightforward process, requiring only configuration for the specific headset in question. Accordingly, OpenXR was employed in this study. The SDK serves as an interface between the hardware and the software, routing all parameters to the requisite scripts of the animation controller and the HMD. Ultimately, the individual body parts of the avatar must be connected to the corresponding hardware controllers. With slight adjustments, the digital persona can be readily exported and imported between various scenes in Unity.

F. Stylized Avatar

An additional method for creating a more stylized avatar (Fig. 3, second from left) is through the utilization of the tool Ready Player Me. The process of creating an avatar using the Ready Player Me tool is relatively straightforward [14]. To commence, open your web browser and navigate to the relevant website. On the website, a button or link will be provided, which will prompt the user to proceed with the creation of their avatar. Once the process has been started, the initial step is to specify the gender of the avatar. A series of options will then be presented, including the choices of male, female, and other gender options. After the gender has been selected, the next step is to proceed with the customization of the avatar. This entails making adjustments to a number of features in order to ensure that the avatar's appearance aligns with your personal preferences. The available customization options may include the selection of a hairstyle, hair color, modification of facial features, choice of clothing items, and potentially other features. By selecting the requisite customization option, the desired adjustments can be made to create an avatar that reflects the user's envisioned appearance. Once the avatar has been meticulously customized, the finalization stage will be reached. At this point, the user will encounter an option labelled "Finish" or "Export." This is the phase during which the system generates a distinctive avatar based on the user's selections. Upon completion of the aforementioned steps, the avatar will be available for download. The format in which the avatar is provided will depend on the intended use.

One advantage of Ready Player Me is that no additional hardware is necessary, with the exception of a device with web browsing capabilities. The avatar is equipped with all the requisite elements, including mesh, textures, and a skeleton, immediately following the download. It can be operated directly with the HMD. However, the lack of photographic input limits the degree of realism achievable with the avatar. In contrast, the process here is based on the user's perception of themselves. It is possible to create an avatar based on a photograph. In order to utilize this functionality, it is necessary to capture an image of the user's face. Subsequently, the program identifies the most salient features of the face and searches the database of potential "bricks" to determine the optimal match. Such characteristics include hair color and style, eye color, eyebrows, nose, mouth and general facial shape. Subsequently, the user is afforded the opportunity to personalize the avatar they have created, selecting from a range of physiques, clothing and accessories.

G. Mixed Avatar

To offset this issue, a third method for creating avatars has been developed, which combines the benefits of lifelike and stylized avatars to create a hybrid (Fig. 3, third from left) that encompasses both approaches. In the case of avatars intended to closely resemble a real person, the face is likely to be of paramount importance. The face is a reflection of the individual's behavior, empathy, and perception [15, 16]. The website Avaturn employs this characteristic, offering avatars that are created expeditiously yet remain highly realistic [14]. To achieve this, the software employs the same methodology as that used by Polycam for scanning individuals, but restricts its application to the user's face. This necessitates the capture of three biometrically-verified images of the user's face. Three photographs are required, taken from the front, the left, and the right side, respectively. The photographs serve as the basis for the creation of a digital mesh and texture for the face. Subsequently, the remaining body parts, including hair, the torso, the arms, and the legs, are created using a character creator tool that is similar to the Ready Player Me tool. At the outset, the user is prompted to select their gender. Subsequently, the user is afforded the opportunity to modify the height and stature with the assistance of sliders. Afterwards, the user may select from a range of hairstyles, clothing, and accessories. Additionally, the software provides the option to incorporate animations and other design elements, thus allowing users to customize their avatar to a considerable extent. Furthermore, as the body is constructed from a predefined model, the avatar can be imported directly into Unity and linked to the animation controller, in a manner analogous to that observed in the Ready Player Me avatar. The result is an avatar that, upon initial observation, appears to be a near-identical replication of the original. However, upon closer examination, notable discrepancies emerge. This is not, however, an issue in itself, as an avatar that is too similar to humans can be subject to the uncanny valley effect, which can result in a reduction in immersion [17].

Furthermore, an avatar devoid of any connection to the human subject was included as a control group for the tests. The character model in question is that of Space Robot Kyle (Fig. 3, far left). This robot, which is devoid of any distinctive characteristics, provides all the necessary features to create a VR avatar, without specifying a gender or body shape. As the robot does not aspire to emulate the human form, the potential for evoking the Uncanny Valley effect is minimal [2]. Consequently, it is feasible to ascertain whether there are discernible discrepancies between realistic and generic avatars. The four types of avatars, namely lifelike, stylized, mixed and generic, constitute the fundamental variable upon which the experimental design of the second case study is based.

H. Participants Study 1

For the first study, we recruited a total of 160 participants via Prolific crowdsourcing website. Data collected via crowdsourcing websites such as Prolific is currently very much in vogue. This is mainly due to the fact that the data can be collected very easily and quickly and there are already studies showing that the data collected here can keep up with traditional methods in terms of quality [18]. Nevertheless, the data should also be checked for quality [19]. We have therefore decided to incorporate various quality checks into the data collection process. Firstly, all data sets with incorrect answers to six or more "catch trials" are removed. Secondly, we considered a lack of variation in ratings between participants as an indicator of inattention. Therefore, we removed data from participants whose ratings had a standard deviation of less than 10 ($SD < 10$) on a scale of 0 - 100. Finally, we compared each participant's ratings with the average of the remaining judgements in their group (between participants) by calculating the correlation between the individual judgements and the remaining judgements in their group. If this correlation between the individual participant's ratings and the group mean was less than 30, the participant's data were discarded as these individuals may have been performing a different assessment task to the group. After this quality check, the total number of participants, from around the world, was 143 (M Age = 20, SD Age = 10, 104 Male, 41 Female, 2 No Responses). This means that each avatar had a rating of 15 - 20 participants.

I. Design and Procedure Study 1

The 200 avatars were divided into four groups of 50 avatars each. Each group was also provided with 10 catch trials. This meant that each group had 60 avatars, which were rated by 20 participants. Because two different questions were asked in this study, one asking, "how human-like is the avatar?" and the other "how uncanny is the avatar?", we asked each group twice. This gave us a total of 4 groups of 60 avatars per question.

The participant begins the survey with a brief introduction to the topic and a short briefing on how to complete the survey. This was followed by an example task on how the participant should rate the avatars. The same example avatar was used for each block. The participant sees two pictures of an avatar on their screen. On the left the entire body and on the right the profile picture with the face in focus. Below the pictures is the definition of the respective question. For the question about human likeness, the participant sees the various characteristics that make an avatar like humans and a slider from 0 to 100. Above the slider is the question "How similar to humans do you think this avatar is?" (0 - not at all like humans and 100 - very similar to humans). We used a similar method for the question of how uncanny you think the avatars are. You can see the same pictures and again a slider from 0 to 100 but this time with the definition about uncanniness and the question "How uncanny do you think this avatar is?" (0 - not uncanny at all and 100 - very uncanny). The participants are randomized into one of the respective groups and are only allowed to answer one question type. This is to prevent the questions from influencing each other. The catch trials are pictures of real people or objects that have also been randomized into the respective groups.

After half of the questions, the participants were given a 10-second break during which their attention was drawn to the definition and characteristics again. After judging all the images, participants were asked to complete a demographic questionnaire in which they were asked to indicate their age, gender, native language, level of education, previous knowledge of robotics and experience with virtual avatars. The entire study took approximately 5 minutes to complete, and participants received $1 as compensation for their participation.

J. Participants Study 2

A total of 30 participants took part in the second study. The experimental group is comprised of 29 male and one female participant. The participants ranged in age from 16 to 50, with the majority falling between the ages of 20 and 28. The probands themselves rated their experience with VR/AR on a scale of 1 (no experience) to 5 (high experience). The mean score was 2.20. The average test time was 57 minutes and 10 seconds. This figure includes the time taken to conduct the interviews and complete the survey. The majority of participants had a background in education. The study group comprised 15 bachelor's and master's students, 4 computer scientists, and 6 individuals from other non-informatics professions. It is reasonable to assume that a computer scientist or prospective computer scientist in college will have some experience with VR or AR. However, it is not possible to derive any concrete findings from this, given that each course of study has a different focus and the topic of VR and AR was not included in the subjects' basic studies. Accordingly, the experience values assigned to the probands represent an approximate estimation of their own interest and experience. In contrast to the initial study, the participants were not sourced through crowdfunding platforms such as Prolific. Instead, they were recruited through the university, in person, through advertising, or through personal connections. Except for a few small items such as sweets or pens, there was no monetary compensation for the subjects of the second study.

K. Design and Procedure Study 2

The second study consists of three steps. First, the participants are informed about the upcoming events. It is explained that this is a role play in which the respondent takes the role of a politician and the interviewer takes the role of various journalists from different newspapers. The two parties then meet in a generic virtual space provided by the university. The respondent has a generic avatar with no particular resemblance. His counterpart is one of four avatars: lifelike, stylized, mixed, or generic.

The second step is the beginning of the discussion. The journalist asks if the politician knows the last generation. The last generation is a group of climate activists who have gained attention for their direct-action protests aimed at raising awareness about climate change. Depending on the answer, either a short explanation based on an example of the last generation or a direct question about the politician's opinion on the topic follows. Based on the politician's opinion, an argument for or against the last generation follows. The politician can respond to this argument. The journalist then proposes a possible solution to the problem and discusses whether it is within the politician's possibilities. This discussion continues until a compromise is reached or the attempt to find one is declared a failure by both parties. After the discussion, there is a short break during which the interviewee can answer the first two questions regarding the uncanniness and the human likenesses. The same questions and definitions were used here as in the first study. The interview takes place four times in total. Each time, the avatar's appearance and newspaper will change. After each round, the participant has time to answer the questions. After the fourth round, the discussion ends and the final phase begins.

In the last phase, the respondent is informed about the topic and asked to fill in the voluntary feedback questions about the avatars. After that, there is a farewell and the test is successfully completed.

III. Results

The findings of the initial and subsequent studies are presented individually in the following sections, as the latter was conducted post-hoc.

A. Results Study 1

For the data analysis, all the results of the individual surveys were added together and an average for human likeness and uncanniness was calculated for each avatar. We then inserted these results into Microsoft Excel to generate various graphs. Looking at the first graph (Fig. 3 a), no uncanny valley can be recognized. Instead, there is a linear gradient between the two

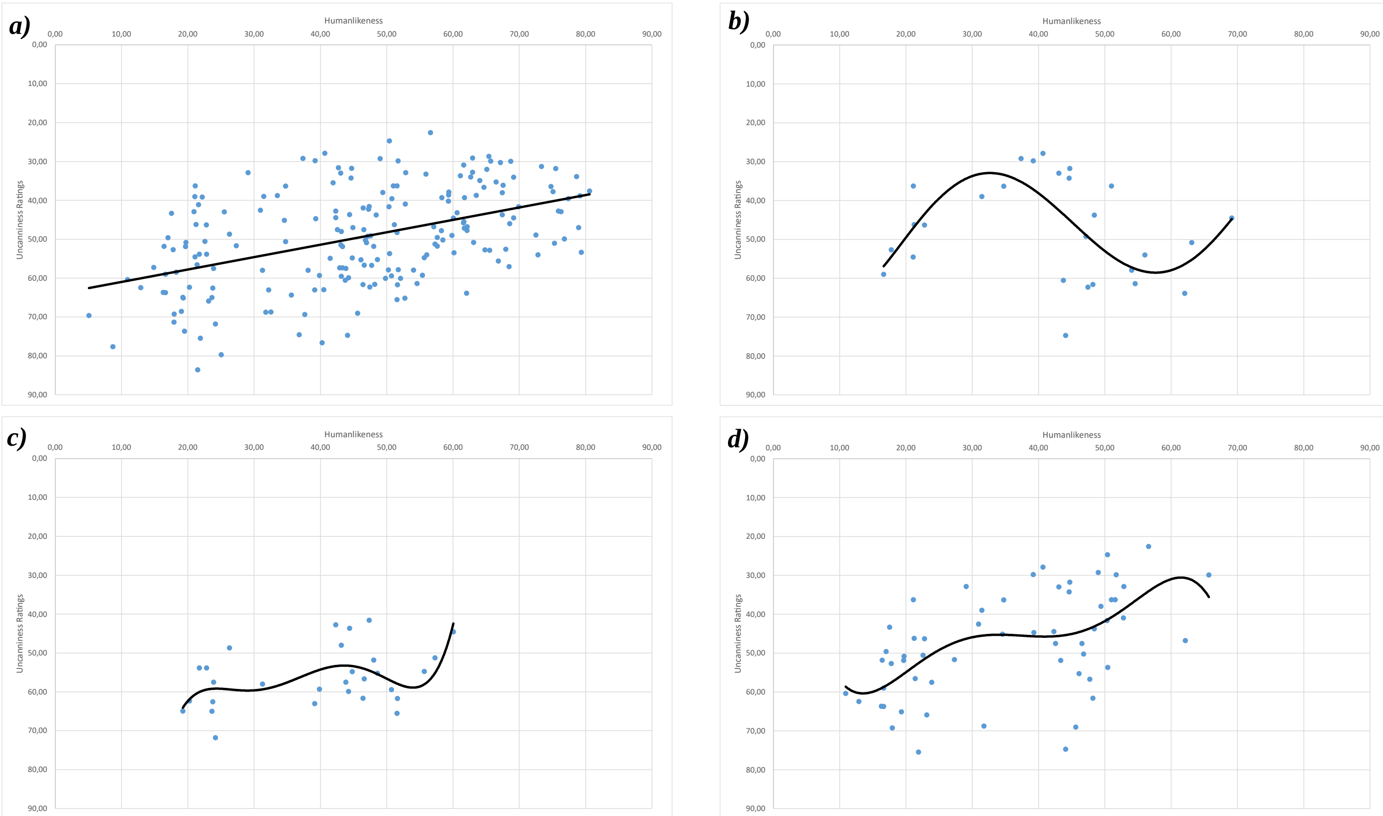

Fig. 4. Four scatterplots a) Total scope of all avatars b) Only avatars representing children c) Avatars caricaturing a real-life person d) Avatars representing a cartoonized style. This Y axis has been inverted to represent the uncanny valley.

factors uncanniness and human likeness, with uncanniness decreasing as human likeness increases. However, if one does not look at the entire amount of data and instead only at certain categories, such as only avatars that represent children, an uncanny valley is clearly recognizable. Just as in Mori's uncanny valley hypothesis [2], a large valley can be recognized between the moderately realistic and realistic avatars. When looking at other avatar categories, a slight uncanny valley can also be recognized. The other avatar categories, such as avatars that are based on a real person and represent them as a caricature, also have a slightly uncanny valley. The same applies to avatars that are not based on a real person but are depicted as a cartoon. based on a real person and represent them as a caricature, also have a slightly uncanny valley. To further investigate these results, we performed a polynomial mixed fit for the three different categories of. We determined the different coefficients of determination = r^2 for different polynomial mixed effects 3rd, 4th, and 5th models. In addition, based on the results of Kim et. al. [7] we also assumed that if there are one or more valleys here, then these are recognized in the 4th or 5th polynomial model.

B. Results Study 2

As with the initial study, the findings of the individual surveys were aggregated and imported into Microsoft Excel. Subsequently, based on the preceding graphs, structural graphics were generated in this manner as well.

The new graph (Fig. 5) illustrates the ratings of the four new avatars (generic, stylized, mixed, and life-like) in comparison to the other avatars based on their similarity to humans and

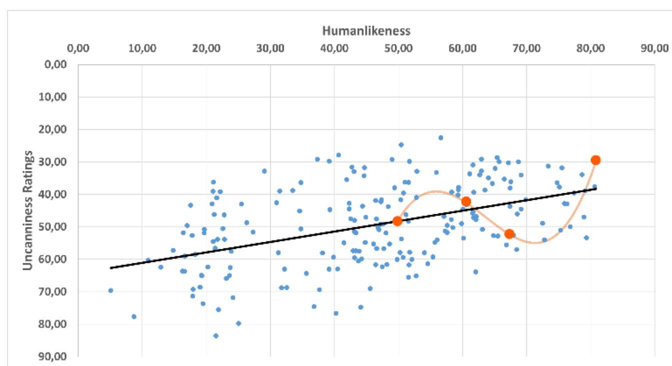

Fig. 5. Scatterplot of all avatars including new generic, stylized, mixed, and life-like avatars in order from left to right.

perceived uncanniness. For enhanced readability, the four new avatars were marked in red and 200% larger than the other points on the graph. The generic avatar (Fig. 5, far left) scored 49.87 points in human likeness and 48.20 in uncanniness. This places the avatar in a relatively intermediate position in comparison to other avatars. The stylized avatar (Fig. 5, second from left) is more human-like than the generic avatar, with a higher score of 60.47 on the human likeness and a slightly lower score of 42.50 on the uncanniness scale. When the data set is considered as a whole, the curve demonstrates that the degree of uncanniness decreases as the degree of human likeness increases. This finding is comparable to the results observed in Fig. 4 a). The aforementioned increase is, however, interrupted by the mixed avatar (Fig. 5, third from left), which, despite its greater similarity to humans with 67.33 points, was rated as more uncanny with 52.30 points. Furthermore, the mixed avatar exhibits a significantly lower mean than the overall distribution of all avatar in the same human likeness category. The life-like avatar (Fig. 5, far right) is deemed the most human-like, with an average score of 80.81 on the human likeness scale and 29.81 on the uncanniness scale. This makes the avatar the most human-like and least uncanny of the new avatars, and one of the best performing overall. As a result, an uncanny valley curve emerges between the new avatars, although this is of limited significance due to the small sample size of four avatars. Accordingly, the r^2 values were not included in the analysis. The following chapters will present a discussion of the results, together with a comparison to the findings of the initial study.

IV. DISCUSSION

Using a database of 200 different VR avatars, we were able to find evidence for the uncanny valley phenomenon of Mori et al. [2]. Contrary to expectations, however, this was not the case for the entire sample, but only when we looked more closely at different sub-categories of avatars. This means that when many different avatar categories, such as different age classes or different styles, are analyzed together, the individual graphs overlap and thus close the uncanny valley. The valley can only be created if the data is sorted precisely.

Particularly noticeable here were avatars representing children. We found that when trying to make this type of avatar more realistic, some avatars were perceived much more negatively than avatars that did not try. Avatars that received a lower human-likeness score of under 40 out of 100 points for uncanniness were significantly better than those with a higher humanlike score and over 60 out of 100 points for uncanniness. This phenomenon cannot be replicated in the other age groups. We assume this is due to the proportions of the avatars. Because the unrealistic avatars in particular have a significantly larger head than the realistic avatars, which have normal proportions here. We were able to make this observation with the caricatures of real personalities such as former presidents of the USA. An uncanny valley can also be recognized here and, similar to the avatars representing children, these are mainly avatars with unusual proportions. This could be due to the fact that an attempt was made here to depict real people and by increasing the similarity the avatars fall into uncanny again.

By focusing solely on the uncanny valley phenomenon in each curve, it is now possible to hypothesize certain characteristics that the avatars possess in common with one another. For example, all the avatars represented on the uncanny valley curve exhibit the "dead eye" syndrome [13]. As the face and eyes are the most prominent features, it is unsurprising that these avatars are perceived as more uncanny. It is likely that this is also the reason why the mixed avatar performed significantly worse than the other avatars (Fig. 5). This is because, when the avatar's eyes are generated, the actual textures of the eyes are not taken from the photograph; instead, only an approximation is generated by the computer, which results in the eyes appearing lifeless (Fig. 3, third from left). Furthermore, a significant number of the avatars identified as uncanny exhibit disproportionate limb sizes. In most cases, the head is the area of the avatar that is perceived as disproportionate. However, the shoulders and hands are also often observed to be out of alignment with the rest of the body. This may explain why the children and caricature avatars from Fig. 4. b) and c) have been classified in such a diverse manner. It is challenging to find the appropriate proportions for avatars representing children without appearing unnatural or unsettling. To "cute up" a human usually means to adjust the proportions so that the head is larger than the rest of the body. Similarly, the avatars caricature famous personalities. Here, too, larger heads were used to make the person behind the personality appear cute or amusing. Another feature is the skin of the avatars. Particularly striking were the avatars that, although they were humanoid in stature, had little or no skin similar to that of humans.

To illustrate, the avatar with the highest degree of uncanniness among the simple avatars was the "Mannequin" avatar (Fig. 6). Furthermore, this avatar exhibits the highest degree of uncanniness when evaluated across the entire database.

Fig. 6. The image of the mannequin, which has the highest score for uncanniness.

Furthermore, the aspect of poor skin quality also serves to differentiate the avatars, which exhibited a similarly elevated level of uncanniness. Mori had previously referenced this category in his uncanny valley curve, situating these types of avatars at the lowest point on the curve [2]. The aspects that have been identified as the most significant contributors to the perception of uncanniness can be distilled into three primary criteria: eyes, body proportions, and skin. By incorporating these three criteria into the design of avatars, it is possible to create representations of individuals that are not perceived as uncanny, as shown by the stylized (Fig. 3, second from left) or lifelike avatar (Fig. 3, far right) in the second study (Fig. 5).

V. CONCLUSION

With the drastic development of virtual reality and the constantly growing environment and possibilities it offers us, human-like avatars are also becoming an important topic that will affect us in the coming years. Even now, avatars from different areas are being rated according to their appearance and the term uncanny valley is being used more and more frequently. Based on the results of this study, we were able to find out that the uncanny valley is not an all-encompassing phenomenon in relation to VR avatars. Instead, the uncanny valley can only be found when taking a closer look at the individual subcategories of avatars. For example, if you take all the avatars in this database, there is an increasingly linear development between human-likeness and uncanniness, with the uncanny factor decreasing as human-likeness increases. However, if you look at certain subcategories, you can see a valley. This observation can also be observed in other categories, which leads us to assume that an overlap between the individual categories means that the uncanny valley is closed and thus balanced out by different avatars.

A more detailed examination of the uncanny valley revealed three primary factors: eyes, body proportions, and skin. These factors, among others, exert a significant influence on the perception of uncanniness. These factors were used to create actual example avatars, thereby enabling the generation of realistic representations of individuals who exhibited minimal or no uncanniness. Consequently, it can be reasonably inferred that with the assistance of enhanced scanning options and adherence to specific criteria, it is feasible to create avatars that accurately reflect the human form. In order to confirm this assumption, further and possibly even larger-scale studies than this one are needed. And by continuing to develop this database, we want to make this possible for everyone.

Acknowledgment

This work was supported by the Würzburg-Schweinfurt University of Applied Sciences. We would like to acknowledge the work of Kim et al. and the developers of the ABOT database. Their work can be found in [4] and [7]. Furthermore, we would like to thank everyone who participated in the study, especially those who participated in the second study, which took a relatively long time to conduct.

References

[1] Arda H, Henneberger A. Evaluating Digital Avatars in VR - A Systematic Approach to Quantify the Uncanny Valley Effect. In: Stigberg S, Karlsen J, Müller NH, editors. ACHI 2024: The Seventeenth International Conference on Advances in Computer-Human Interactions : May 26-30, 2024, Barcelona, Spain. Wilmington, DE: IARIA; 2024. p. 1–6.

[2] Mori M, MacDorman K, Kageki N. The Uncanny Valley [From the Field]. IEEE Robot. Automat. Mag. 2012;19:98–100. doi:10.1109/MRA.2012.2192811.

[3] Verplank B, Sutcliffe A, Mackay W, Amowitz J, Gaver W, editors. Proceedings of the 4th conference on Designing interactive systems: processes, practices, methods, and techniques. New York, USA; 2002.

[4] Phillips E, Zhao X, Ullman D, Malle BF. What is Human-like? In: Kanda T, Ŝabanović S, Hoffman G, Tapus A, editors. HRI '18: ACM/IEEE International Conference on Human-Robot Interaction; 05 03 2018 08 03 2018; Chicago IL USA. New York, NY, USA; 2018. p. 105–113. doi:10.1145/3171221.3171268.

[5] Mathur MB, Reichling DB. Navigating a social world with robot partners: A quantitative cartography of the Uncanny Valley. Cognition. 2016;146:22–32. doi:10.1016/j.cognition.2015.09.008.

[6] MacDorman KF, Ishiguro H. The uncanny advantage of using androids in cognitive and social science research. 2006;7:297–337. doi:10.1075/is.7.3.03mac.

[7] Kim B, Bruce M, Brown L, Visser E de, Phillips E. A Comprehensive Approach to Validating the Uncanny Valley using the Anthropomorphic RoBOT (ABOT) Database. In: 2020 Systems and Information Engineering Design Symposium (SIEDS); 24.04.2020 - 24.04.2020; Charlottesville, VA, USA: IEEE; 2020. p. 1–6. doi:10.1109/SIEDS49339.2020.9106675.

[8] DiSalvo CF, Gemperle F, Forlizzi J, Kiesler S. All robots are not created equal. In: Verplank B, Sutcliffe A, Mackay W, Amowitz J, Gaver W, editors. DIS02: Designing Interactive Systems 2002; 25 06 2002 28 06 2002; London England. New York, NY, USA: ACM; 2002. p. 321–326. doi:10.1145/778712.778756.

[9] Bartneck C, Kanda T, Ishiguro H, Hagita N. Is The Uncanny Valley An Uncanny Cliff? In: RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication; 26.08.2007 - 29.08.2007; South Korea: IEEE; 2007. p. 368–373. doi:10.1109/ROMAN.2007.4415111.

[10] Sasaki K, Ihaya K, Yamada Y. Avoidance of Novelty Contributes to the Uncanny Valley. Front Psychol. 2017;8:1792. doi:10.3389/fpsyg.2017.01792.

[11] Chattopadhyay D, MacDorman KF. Familiar faces rendered strange: Why inconsistent realism drives characters into the uncanny valley. J Vis. 2016;16:7. doi:10.1167/16.11.7.

[12] Schwind V, Wolf K, Henze N. Avoiding the uncanny valley in virtual character design. Interactions. 2018;25:45–9. doi:10.1145/3236673.

[13] Moore RK. A Bayesian explanation of the 'Uncanny Valley' effect and related psychological phenomena. Sci Rep. 2012;2:864. doi:10.1038/srep00864.

[14] Ready Player Me. Integrate an avatar creator into your game in days - Ready Player Me. 13.09.2024. https://readyplayer.me/de. Accessed 13 Sep 2024.

[15] Forsythe SM. Effect of Applicant's Clothing on Interviewer's Decision to Hire. J Applied Social Pyschol. 1990;20:1579–95. doi:10.1111/j.1559-1816.1990.tb01494.x.

[16] Zebrowitz LA, Montepare JM. Social Psychological Face Perception: Why Appearance Matters. Soc Personal Psychol Compass. 2008;2:1497. doi:10.1111/j.1751-9004.2008.00109.x.

[17] Ding LI, Moon H-S. Uncanny Valley Effect in the Animation Character Design - focusing on Avoiding or Utilizing the Uncanny Valley Effect. Cartoon and Animation Studies. 2016;43:321–42. doi:10.7230/KOSCAS.2016.43.321.

[18] Mortensen K, Hughes TL. Comparing Amazon's Mechanical Turk Platform to Conventional Data Collection Methods in the Health and Medical Research Literature. J Gen Intern Med. 2018;33:533–8. doi:10.1007/s11606-017-4246-0.

[19] Ahler DJ, Roush CE, Sood G. The micro-task market for lemons: data quality on Amazon's Mechanical Turk. PSRM. 2021:1–20. doi:10.1017/psrm.2021.57.