Virtual Reality & Intelligent Hardware

• **Article** •

October 2022 Vol. 4 No. 5: 393–405
10.1016/j.vrih.2022.08.003

Uncanny valley for interactive social agents: an experimental study

Nidhi MISHRA[1], Manoj RAMANATHAN[2*], Gauri TULSULKAR[1],
Nadia Magneat THALMANN[3,4]

1. *Institute for Media Innovation, Nanyang Technological University,* 637553, *Singapore;*
2. *Rehabilitation Research Institute of Singapore, Nanyang Technological University,* 308232, *Singapore;*
3. *Institute for Media Innovation, Nanyang Technological University,* 637553, *Singapore;*
4. *MIRALab, Battelle, Building A 7, Route de Drize, CH-*1227 *Carouge, Geneva, Switzerland*

Received 10 June 2022; **Revised** 22 July 2022; **Accepted** 18 August 2022

Abstract: Background The uncanny valley hypothesis states that users may experience discomfort when interacting with almost human-like artificial characters. Advancements in artificial intelligence, robotics, and computer graphics have led to the development of life-like virtual humans and humanoid robots. Revisiting this hypothesis is necessary to check whether they positively or negatively affect the current population, who are highly accustomed to the latest technologies. **Methods** In this study, we present a unique evaluation of the uncanny valley hypothesis by allowing participants to interact live with four humanoid robots that have varying levels of human-likeness. Each participant completed a survey questionnaire to evaluate the affinity of each robot. Additionally, we used deep learning methods to quantify the participants' emotional states using multimodal cues, including visual, audio, and text cues, by recording the participant–robot interactions. **Results** Multi-modal analysis and surveys provided interesting results and insights into the uncanny valley hypothesis.

Keywords: Uncanny valley hypothesis; Human robot interaction; Interactive robots; Humanoid robots; Virtual humans

Supported by the National Research Foundation, Singapore under its International Research Centers in Singapore Funding Initiative; and Institute for Media Innovation, Nanyang Technological University (IMI-NTU).

Citation: Nidhi MISHRA, Manoj RAMANATHAN, Gauri TULSULKAR, Nadia Magneat THALMANN. Uncanny valley for interactive social agents: an experimental study. Virtual Reality & Intelligent Hardware, 2022, 4(5): 393–405

1 Introduction

In the last decade, significant progress has been made in the science of robotics and artificial intelligence (AI). This has led to the development of humanoid robots or virtual humans with a human-like appearance, intelligence, and behavior (verbal and non-verbal), such as Nadine[1], Erica[2], and Sophia. Despite being validated in real-life applications such as banking[3], newscasting[2], therapy, and other roles[4–7], a main concern for humanoid research and study is the uncanny valley hypothesis[8]. A previous study[8] hypothesized that a

***Corresponding author**, mramanathan@ntu.edu.sg

person′s emotional or affective response varies depending on the appearance of the robot with which they interact. A person may feel uneasy and unnerved with a more human-like robot. This hypothesis is considered primarily when designing human-like robots, although other studies[9,10] have found inconsistent empirical evidence supporting it. The original hypothesis generalizes the definition of a humanoid robot into a single data point, which may be inaccurate. For example, Eaton provided a comprehensive taxonomy of different types of possible humanoids[11]. With the development of human-sized robots, such as Nadine[1], Erica[2], and Ishiguro[12,13] that can be placed in social scenarios to interact directly with users, it is crucial to reassess the uncanny valley hypothesis for its relevance.

A more detailed examination of this hypothesis for interactive, social humanoid robots is required to evaluate present-day opinions of such agents. No specific study has examined interactive robots for the uncanny valley problem in action. Most studies involved no interactions[14–16] in their assessments. They instead showed videos of the robots or virtual avatars to participants and then asked the participants questions to gauge the impact of the robots' appearances. In this study, we evaluate the uncanny valley theory through live human interactions with four human-like entities:

1. Maya: simple voice assistant (only a human voice).
2. Nao[17]: child-sized programmable humanoid robot with articulated limbs but without skin or hair.
3. Nicole: virtual human with a complete virtual human-like embodiment.
4. Nadine[1]: complete life-sized humanoid robot with skin, articulated hands, and other human-like features.

Our study intends to answer the following research questions.

・ **Uncanny valley for Interactive Humanoid Robots**: Exploring this theory to provide an in-depth look into how people's emotions and perceptions vary for different types of human-like interactive robots.

Examining how the Uncanny valley affects the current human generation, which is more accustomed to advanced technologies and may be more open to human-like entities.

・**Using AI for Uncanny valley quantification**: Quantifying the emotional responses of participants using surveys and multimodal emotion and sentiment analyses (using visual, audio, and text).

This remainder of this paper is organized as follows:

・ Section 2 reviews previous research in the field of the uncanny valley hypothesis.

・ Section 3 addresses our proposed experimental setup and the details of each humanoid robot used in the current research.

・ Section 4 provides details of the data collection procedures and emotion analysis methods employed in this study.

・ Section 5 details the results obtained and provides insights from visual, audio, text, and survey data analyses.

・ Finally, Section 6 provides the conclusions and discussions.

2 Related work

According to the uncanny valley hypothesis[8], users may experience negative affective emotions or a state of eeriness when interacting with near-human entities or agents. Owing to this, increasing the agent's anthropomorphic realism would have a counterproductive effect on the users′ subjective experiences[10]. Thus, the hypothesis remains a guiding principle in robot design and cross-modal technologies, such as virtual character design[18], video games, and animations[19].

Since the proposal of the hypothesis, several studies have recreated or visualized the effect[20], tested its validity[9,10] and used perceptual analyses, cognitive analyses[21–23], and other procedures to investigate this effect. A previous study[10] noted that the original hypothesis was not validated by any empirical tests. Studies examining the validity of the hypothesis[9] have uncovered no empirical evidence to support it and have reported inconsistent findings with different conceptualizations. A few other studies[9,10] primarily perceived uncanny valleys owing to the perceptual mismatch, categorical ambiguities, and other factors. For example, unusual physical attributes or inconsistent human-like realism can lead to negative emotions.

People also dislike human-like robots making moral decisions compared to the same choices made by humans or non-human robots[24]. These studies have also noted that an uncanny effect is not generalizable across different individuals, stimuli, situations, tasks, and time. A nuanced understanding is required to precisely know when and under what conditions the negative emotions are observed.

A major drawback of the original hypothesis is it does not provide the exact definition or standards of human-likeness, affinity, eeriness, and means of quantifying these[9,25], which causes methodological circularity[25]. In recent years, researchers have expanded their investigations to examine the possibility of observing the same uncanny valley in zoomorphic robots[26] or virtual animals[27]. Like humanoid robots, zoomorphic robots that combine realistic and nonrealistic features are less preferred[26]. After recent technological advancements, the development of AI, and realistic looking social humanoid robots, such as Nadine[1], Erica[2], and Ishiguro[12,13], it has become essential to define the type of morphological traits that can cause uncanny valley effects. Additionally, several studies have insufficiently controlled the variation in human likeness portrayed in stimulus images, i.e., the nature of the stimuli that elicit the uncanny valley is not well defined or quantified[25]. Therefore, analyzing users' affective states when interacting with robots with different levels of human likeness is necessary. In this study, we consider the robots or agents as mentioned above, which have varying degrees of human likeness in their appearance.

In the past, this hypothesis was primarily validated by allowing participants to view non-morphed[28–30] or morphed images of robots[20,31], video clips of agents performing tasks[16,32,33], or computer-generated models[34]. Because the facial features of a human are an essential characteristic that lend to the realism of the agents, many studies have focused on simply showing virtual faces[31,35] and facial images[30,36] to participants. However, these methods consider only the faces and avoid other possible causes of the uncanny valley, such as the movement of robots. The study reported in [37] showed how attributes, such as skin and body movements, are essential to how humans perceive such agents. Most robots are designed to interact with humans and their environment to accomplish tasks. With social humanoids, the interaction capabilities of robots have become essential. Assessing or studying the uncanny valley would be difficult by viewing only images or video clips, as people don't interact with these robots or use them for any purpose. In contrast, in this study, we let participants interact live with four different robots of varying human likeness. The participants communicated with the robots directly via Zoom calls as no in-person communication was possible, owing to COVID-19 restrictions. We analyzed human-robot interactions to observe any uncanny valley effects on the participants.

Another critical aspect of this hypothesis is how user-related affinity and eeriness caused by a robot's appearance are quantified or measured. Several studies have attempted to examine the psychological aspect of the uncanny valley hypothesis[21–23,38] using functional MRI (fMRI) to characterize human behavior and observe the uncanny effect on participants. However, these studies hooked subjects up to bulky MRI machines. While methods such as those in [9,20] scrutinized past studies to find empirical evidence of the uncanny effect, studies reported in [39,40] provided a Bayesian explanation of the observed phenomena. The study reported in [39] concluded that human-looking robots have a huge potential to improve social interactions in individuals with autism. A defined protocol for determining and validating the affective or emotional state of a participant is unavailable. Most of these studies[30] use ad-hoc self-rating scales[10]. Few studies[14,41] have considered valid psychometric and behavioral evaluation methods to study how the human mind perceives human and nonhuman entities. Because we allowed subjects to directly interact with the robots, we recorded interaction via videos, audio, and text (via Zoom with consent from participants). We analyzed these various modalities using state-of-the-art deep learning video, audio, and text emotion and sentiment analysis methods to determine if any negative affective traits were visible during these interactions. In addition to these modalities, participants completed a Godspeed questionnaire[42] for each robot. Using both the questionnaire and multimodal deep learning emotion and sentiment analysis, we quantified the likeability of each robot and scrutinized the uncanny valley phenomena for socially interactive humanoid robots (each robot with a varying level of humanness).

3 Experiment setup

The experiment setup is intended to study the uncanny valley hypothesis for interactive humanoid robots and determine whether we can observe the presence of any uncanny effects from the participants. Unlike previous studies, we allowed participants to interact with the robots (or agents).

We conducted our experiment online over Zoom video calls because of COVID-19 pandemic-related restrictions. On the calls, we first introduced all robots to the participants and explained the nature of the interactions. This information was provided to eliminate any categorical ambiguity that could result in uncanny valley effects. Consent was obtained for audio and video recording during the initial explanation. At the end of the human-robot interaction, the participants were asked to fill out a questionnaire.

Participants interacted with the following four types of robots (or agents) in the same scenario:

1. Maya, the voice assistant;
2. Nao, the child-sized humanoid robot;
3. Nicole, the virtual human with human-like appearance;
4. Nadine, the complete life-size humanoid social robot with skin, articulated hands, and other human-like features.

Figure 1 shows these human-like agents/entities. All robots share the same architecture. Please refer to [1] for more information on the architecture and how the input information is processed. The robots varied only in their physical appearances, as indicated.

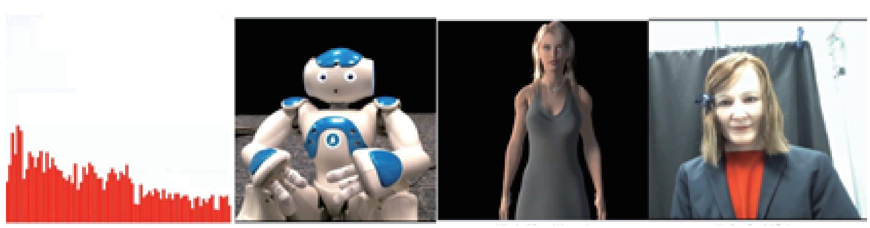

Figure 1 The four different robots used in the experiment.

The audio and video of user interactions were recorded for analysis. Despite the lack of physical presence-related constraints during the conversation sessions, we ensured fluid dialogues between the participants and robots. During these interactions, participants were free to interact with the robots on a one-to-one basis in any form, which covered both verbal and non-verbal aspects. They could ask and discuss any topic of their choice. The humanoid robots, Nao and Nadine, were capable of greeting and waving at the participants owing to their physical capabilities. Nadine's controller generated physical gestures, whereas Nicole's controller generated animated gestures. Nadine and Nicole could both emote expressions with lip synchronization, gaze at the participants, and generate gestures based on the context of the conversations. Because Maya was a voice assistant, physical or virtual nonverbal forms of communication were nonexistent.

The robots were placed in an isolated room, where external noise was excluded, and a black background was used to enhance the participants′ attention. All conversations were conducted in English. At the end of the study, five sets of questionnaires were completed by the participants to obtain their overall feedback and experience about their one-to-one sessions with the robots.

4 Data collection and analysis

To obtain a holistic idea of the effect of each of the robot (or agent) interaction sessions with the participants, we recorded every session and analyzed the videos of 77 participants. The objective tools are based on natural language processing and computer vision techniques and use state-of-the-art deep neural networks (DNNs) to automatically evaluate the mental states of the participants during conversations. The subjective tools consisted of five questionnaires for each robot, including a generic survey targeted at the participants. Furthermore, we performed a statistical analysis to draw meaningful comparisons. In our analysis, we aimed to identify the uncanny valley by measuring the participants' eerie or creepy experiences. Studies such as [14,43] have experimentally shown that negative emotions such as fear, shock, disgust, anxiety, and nervousness are associated with eeriness. The fMRI samples shown in [14] have been proven to show correlations between the uncanny valley and such emotions. In the same manner, we used these objective and subjective analysis tools to identify negative emotions during human-robot interactions and used them as evidence of the uncanny valley.

4.1 Video analysis

Non-verbal cues, such as expressions and gestures, and speech cues are equally important in determining the engagement in conversation[44]. To recognize the facial emotions of participants in the video, we first detected faces using a technique based on a convolutional neural network (CNN) with Dlib[1]. Using ResNet-50[45] as the backbone, we trained an emotion recognizer with eight expression classes: neutral, happy, sad, surprise, fear, disgust, anger, and contempt. We then used the recognizer for facial expression detection on the video frames. Our emotion recognizer was trained on the largest in-the-wild facial expression dataset called AffectNet[46] (with approximately 320000 images excluding none, uncertain, and non-face categories) until the network converged. For the analysis, the model provided confidence levels for each emotion observed on every detected face in each video frame. Therefore, we defined the average emotions displayed by a participant during interaction in a complete video stream as follows:

$$\langle emotion \rangle = \frac{1}{L}\sum_{l=1}^{L} emotion_l \tag{1}$$

where $emotion_l \in [0, 1]$ denotes the probability of a participant's detected facial emotion in the l-th frame belonging to each of the eight classes which were estimated by our emotion recognizer. Figure 2 shows some of the emotions detected in the frames. We used statistical methods to evaluate the differences in interactions with different agents (or robots) and compared them.

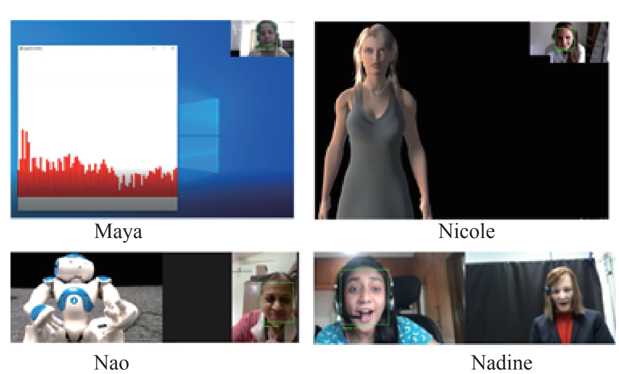

Figure 2 Different emotions detected in the video analysis.

4.2 Audio analysis

A subject's audio or speech pattern is another direct indication of their emotional state. Audio data can contain several explicit and subtle cues that reflect a subject′s mental state. Although numerous audio

[1]http://dlib.net/

emotion recognition investigations have been conducted, we encountered two issues when implementing them. First, several studies had no pretrained models[47,48]. Second, such studies did not focus on a set of emotions that could be useful in completely studying and evaluating the uncanny valley. For instance, the study reported in[49] classified only five emotions and identified gender. Therefore, we used an implementation[2] to train the audio emotion classifier. An RNN-based deep learning model was adopted and trained using four different datasets: RAVDESS[50], TESS[51,52], EmoDB, and a custom dataset. Our model included two RNN layers and two dense layers each having 128 units. We ran a train-test-validation cycle on these datasets to obtain the final model. The model considered nine different emotions: "neutral", "calm", "happy", "sad", "angry", "fear", "disgust", "ps" (pleasant surprise), and "boredom". These varied emotions were selected because they allowed us to better gauge the uncanny valley effect. An entire audio recording was analyzed, and the intensity of each emotion observed was provided by the model.

4.3 Text analysis

For text analysis, we converted the obtained audio files into text using Google Speech-to-Text[3]. To identify emotions in context during conversation between participants and agents, we used the emotion categorization model in SenticNet[53]. This API uses the hourglass of emotions[54], a biologically inspired and psychologically motivated emotion categorization model for sentiment analysis, in conjunction with SenticNet and deep learning to extract emotion labels from texts. The input of this API is a piece of text (a sentence or paragraph), and the output is a list of emotion labels. Dominant emotions were determined using statistical analysis based on the data collected for each agent.

4.4 Questionnaire

The Godspeed questionnaires defined in[42] were used to assess participants′ perceptions of the four robots. The Godspeed questionnaire measures five factors: perceived anthropomorphism, animacy, likeability, intelligence, and safety[4]. There were 24 semantic differential items for these five indices. Participants were required to rate their impressions of each robot according to these 24 semantic traits, which made this a comprehensive survey. We used the collected survey data to evaluate participants′ impression of each agent and the agent′s effect on the five indices. The survey was conducted online using Survey Monkey. In total, there were five different surveys–four surveys for each of the robots and a generic questionnaire to collect demographic information about the participants, such as age, education status, and prior experience with robots. We formulated a 100-point scale for comprehensive subjective scoring. The scores assigned by participants for every item were used to validate the relationships or comparisons captured using the video, audio, and text data. Additionally, our survey included general questions regarding all robots (Table 1).

Table 1 General Questionnaire

Survey Questionnaire
Have you interacted with Robots before?
Have you interacted with virtual characters?
Have you interacted with Voice assistants before?
Which robot/agent do you think is most human-like?
Which robot/agent did you like most?
Did the robot/agent's human-like appearance affect your interaction positively?
Did the robot/agent's human-like voice affect your interaction positively?

5 Results

In this section, we examine the results for each analysis method: video, audio, text, and survey questionnaire analyses. Based on these evaluations, we provide insights into what each modality reveals regarding parti-

[2]https://github.com/x4nth055/emotion-recognition-using-speech
[3]https://cloud.google.com/speech-to-text/
[4]https://www.bartneck.de/2008/03/11/the-godspeedquestionnaire-series/

cipants′ emotions and affective states during their interactions with our agents. The collected data for each modality were statistically analyzed to determine the emotions observed during interactions with each agent.

5.1 Video analysis results

Figure 3 shows the average emotions expressed by participants during their encounters with each robot. We performed repeated measures analysis of variance (ANOVA) tests, to determine the presence or absence of significant differences between emotions regarding the robots on each of the variables. Because the same facial expressions were tracked for each of the four robots, the scores for the emotions expressed for each of them were considered as independent variables.

When variables violated the assumptions of sphericity, the Greenhouse-Geisser modification[55] for degrees of freedom was used. Table 2 presents the ANOVA results.

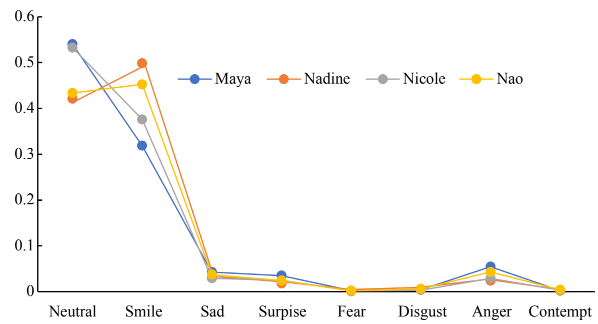

Figure 3 Average levels of each emotion observed from facial expressions over the ensemble of experiments for each of the robots.

Table 2 Results of repeated measures ANOVA tests for all emotions; video-based facial expressions analysis

Emotion	Degree of Freedom (df)	F statistic (F)	Value of test (p)
Neutral	(2.547,152.824)	7.655	0.000
Smile	(2.429,145.715)	15.882	0.000
Sad	(1.882,112.910)	2.612	0.081
Surprise	(2.625,157.494)	2.752	0.052
Fear	(2.201,132.045)	0.791	0.466
Disgust	(2.402,144.146)	0.419	0.695
Anger	(2.286,137.133)	2.944	0.049
Contempt	(1.351,81.085)	0.711	0.442

5.2 Audio analysis results

Figure 4 shows the average emotions of participants for each robot. Like the video modality, repeated measures ANOVA tests were conducted for the audio modality and the Greenhouse-Geisser modification[55] was applied as necessary. Table 3 lists the results from these ANOVA tests and shows significant differences in all emotions between the robots. Thus, post-hoc analyses were conducted for all outcomes to determine the specific differences in the emotions elicited by the robots.

- The post-hoc test for anger showed that it was the highest for interactions with Maya, significantly different from the rest (p<0.001 for all three pairs). The other robots did not differ in this measure (p-values ranging from 0.530 to 0.962).
- The post-hoc test for boredom showed that it was also the highest for interactions with Maya, which was significantly different from the other three (p-values ranging from 0.001 to 0.002). Nicole differed marginally from Nao (p=0.046) but did not differ from Nadine (p=0.380). Nao and Nadine showed no significant differences (p=0.160).
- The post-hoc test for calmness showed that Maya incited the lowest levels, which were significantly different from the other three (p-values

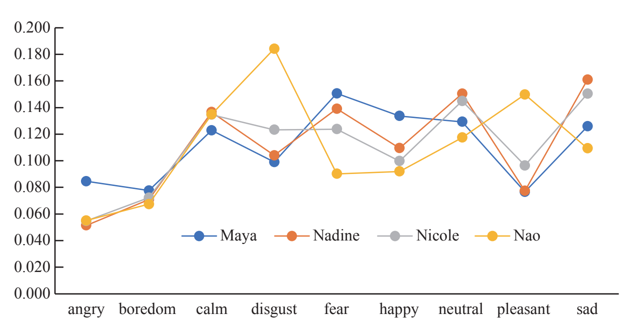

Figure 4 Different emotions derived from audio recordings.

Table 3 Results of repeated measures ANOVA tests for all emotions; audio recordings analysis

Emotion	Degree of freedom (df)	*F* statistic (*F*)	Value of test (*p*)
Angry	(2.636,168.711)	11.498	0.000
Bored	(2.186,139.885)	7.244	0.001
Calm	(3, 192)	6.224	0.000
Disgust	(3, 192)	21.969	0.000
Fear	(2.599,166.348)	18.895	0.000
Happy	(2.683,171.731)	16.043	0.000
Neutral	(3, 192)	20.514	0.000
ps	(3, 192)	24.299	0.000
Sad	(3, 192)	41.499	0.000

ranging from 0.001 to 0.003). The other three robots showed no significant differences (p-values ranging from 0.553 to 0.943).

・ The post-hoc test for disgust showed that it was expressed the most towards Nao (p<0.001). The other three robots had no significant differences in this emotion (p-values ranging from 0.072 to 0.628).

・The post-hoc tests for fear showed that Nao's scores were the lowest, statistically different from the remaining robots (p<0.001 for all pairs). The next lowest was Nicole, marginally different from Nadine (p=0 .049) and significantly different from Maya (p=0.012). Nadine and Maya did not differ significantly (p=0.210).

・ The post-hoc tests for happiness showed that Maya incited the most of this emotion, significantly more than the others (p<0.001 for all pairs). Nadine was significantly higher than Nao (p=0.002) but not higher than Nicole (p=0.082). Nicole and Nao did not significantly differ (p=0.217).

5.3 Text analysis results

Figure 5 presents the sentiment analysis results. The most prevalent emotions were ecstasy, enthusiasm, and delight. There were numerical differences in the prevalence of emotions towards the various robots for each emotion, but these differences could not be tested statistically because of the nature of the data collection. However, they were useful in observing the ranking of the robots for each emotion scrutinized.

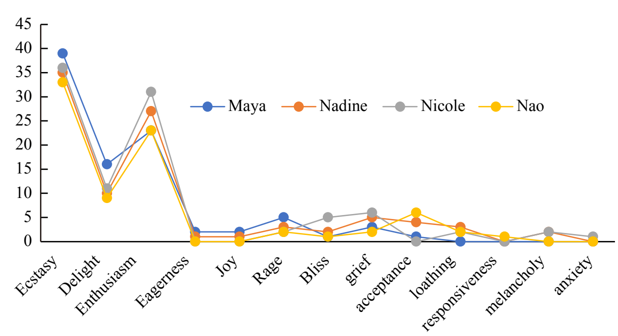

Figure 5 Number of detections of each emotion towards each robot in the spoken text.

Maya elicited the most ecstasy and delight, followed by Nicole, Nadine, and Nao. In terms of enthusiasm, Nicole had the highest scores, followed by Nadine. The third highest were Nao and Maya with the same number of occurrences. Rage was shown most towards Maya, with the other three agents incurring equal amounts. Bliss was mostly directed in Nicole's regard, with less of it directed towards the others. Grief was primarily shown in interactions with Nicole and Nadine and not so much with Maya and Nao. Acceptance was the highest for Nao and lowest for the rest. Eagerness, joy, loathing, responsiveness, melancholy, and anxiety were rarely displayed (i.e., a total of five times for all robots combined).

5.4 Questionnaire results

Figure 6 shows the participants' average scores on the five scales for the four robots. To establish the presence or absence of significant differences among the robots on each scale, five ANOVAs and a post-hoc test were conducted to determine whether specific pairs of robots differed among themselves. As listed in Table 4, all five multiple measure ANOVAs were significant at the <0.01 level, and all but the perceived safety scale were significant at the <0.001 level.

This indicates that the differences between the robots are significant for each scale. To determine which robots elicited the differences, we conducted post hoc analyses and obtained the following results:

・ The post-hoc analyses for anthropomorphism showed that Nadine had the highest scores, which were significantly different from those of the other robots (all $p<0.001$). Nicole′s scores were the second-highest, considerably different from those of Maya ($p=0.004$) and marginally indifferent from Nao′s score ($p=0.061$). Nao and Maya′s scores were insignificantly different ($p=0.314$).

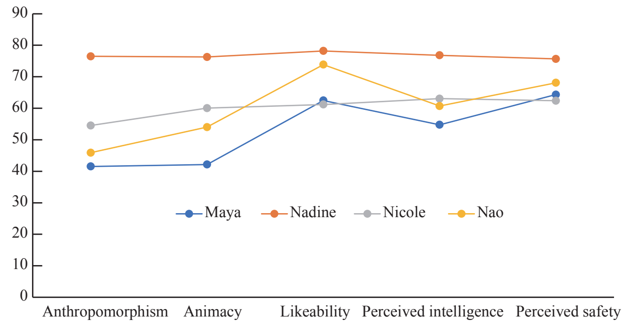

Figure 6 Average scores on each of the five scales for each of the four robots.

Table 4 Multiple ANOVA results for the five scales

Emotion	Degree of freedom(df)	*F* statistic (*F*)	Value of test (*p*)
Anthropomorphism	(3, 161)	25.295	0.000
Animacy	(3, 161)	18.513	0.000
Likeability	(3, 164)	6.701	0.000
Perceived intelligence	(3, 160)	12.108	0.000
Perceived safety	(3, 164)	5.218	0.002

・ The post-hoc analyses for animacy were also higher for Nadine than for the other robots ($p<0.001$). Nicole′s animacy scores were significantly higher than those of Maya ($p<0.001$) but not from those of Nao ($p=0.215$). Finally, Nao′s score was significantly higher than that of Maya′s ($p<0.001$).

・ The post-hoc analyses for likeability showed that Nadine′s scores were significantly higher than those of Maya and Nicole ($p<0.001$) but not higher than those of Nao ($p=0.351$). Nao also had significantly different scores than Maya and Nicole′s ($p=0.013$ and 0.008, respectively). The scores of Maya and Nicole did not differ significantly ($p=0.779$).

・ The post-hoc tests for perceived intelligence showed that Nadine′s scores were significantly higher than those of the others ($p<0.001$). Nicole′s scores were significantly higher than those of Maya ($p=0.029$) but not higher than those of Nao ($p=0.551$). There were no significant differences between the scores of Nao and Maya ($p=0.116$).

・ The post-hoc tests for perceived safety showed that Nadine had the highest scores that were significantly different from those of the other robots (p-values ranging from 0.001 to 0.04). The remaining robots did not differ in this measure (p-values ranging from 0.125 to 0.587).

The participants also completed a general questionnaire regarding their past experiences with robots; 45% of them had previous experience with robots, 60% had previous experience with virtual characters, and 95% had previous experience with voice assistants. They selected Nadine as the most human-like robot, as shown in Figure 7a. For the “most liked robot”, most participants also gave Nadine the highest likeability. However, as shown in Figure 7b, the differences between all robots were not substantial.

Finally, the participants rated how the dimensions of human-like appearances, gestures, voice, and facial expressions impacted the quality of their interactions. As shown in Figure 8, participants rated

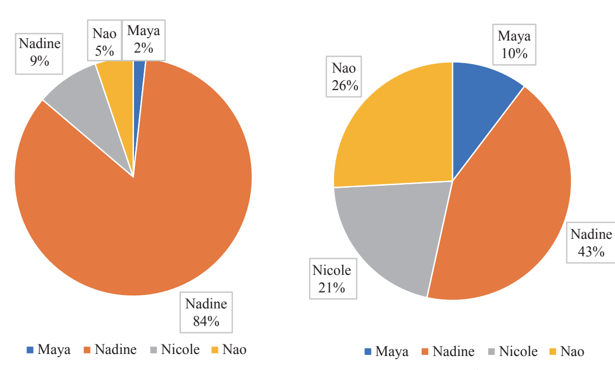

Figure 7 (a) Most human-like robot based on participant perception. (b) Most liked robot.

all features as important.

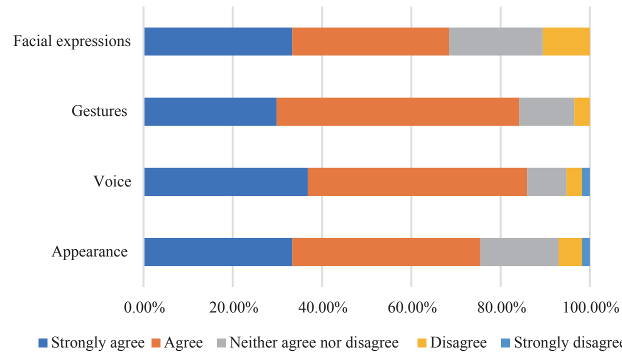

Figure 8 Importance of the various human-like characteristics of the robots.

6 Conclusion and discussion

The results show that all four robots were perceived in various manners and that the emotions expressed varied in their regard. As shown in Figure 6, Nadine was chosen as the "favorite" robot; she was seen as the most anthropomorphic, animatic, intelligent, and safe. The only characteristic that she failed to top was likeability, where she tied with Nao. Furthermore, Maya was seen as the least anthropomorphic and animate, which is predictable given that she is only a voice assistant. Nao and Nicole scored the same on anthropomorphism and animacy. However, Nao was significantly more likeable, which indicated that a physical body may elicit a higher degree of likeability. Findings from the multi-modal analysis and surveys do not clearly show any effect of the "uncanny valley"[56]. Nadine, the most humanoid robot, was seen as the most likable robot, with similar findings obtained through the sentiment and facial expression analyses. Nao, the humanoid-but-toyish robot, was also seen as likable and provoked the highest positive surprise. However, it incurred high disgust, as determined through the analyses of audio data, without evoking many emotions. However, while being the most likable, Nadine generated the most sadness and was the second most feared. This fear is possibly indicative of the "uncanny valley" effect. Nonetheless, with the "uncanny valley," eeriness would be expected to increase with the increasing degree of anthropomorphism, which was not the case. The highest fear was expressed towards Maya, the voice assistant without visual characteristics. Furthermore, owing to the human-like appearance of Nadine, most participants were tentative and fearful as they wanted to impress her. This fear can be considered good, as it indicates that people want to connect with her. Extending our multimodal analysis to other cues such as pose estimation[57,58], body language cues[59–61] could be considered for detecting the evidence of an "uncanny valley".

While Nao had a toy-like, childish appeal, Nadine had a human-resembling body, which was clearly sufficient for increasing her likability compared to a bodyless or virtual agent. However, all robots were well-liked. The lowest-ranked robot scored above 60/100, which indicated that no "uncanny valley" effect was determined in this study. Furthermore, the lack of a correlation between being more anthropomorphic and less likeable and provoking more negative emotions provides evidence against the hypothesis[10]. Compared to previous research, the robots used in this investigation might not have provoked the uncanny valley effect, as they have been carefully and coherently designed and constructed. As shown in [9,10], the uncanny valley is triggered by notable characteristic traits or aspects (for example, the non-human characteristics of agents with a human appearance and vice-versa).

The results showed that the robots incited different emotions, but the most anthropomorphic robots were the most liked. Furthermore, all robots were well-liked, and there was no correlation between the anthropomorphism of the robots and the negative emotions they provoked. Therefore, this study, like previous studies[9,10,25], did not observe or substantiate the uncanny valley hypothesis. This could be because of the characteristics of these specific robots or the specifics of today's world in which both humanoid and non-humanoid robots are becoming increasingly prevalent and people are accustomed to them. Regardless, the future design of interactive robots should be open to creating anthropomorphic robots, while ensuring a coherent design.

Declaration of competing interest

We declare that we have no conflict of interest.

Acknowledgement

Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of National Research Foundation, Singapore. We thank our colleague Tham Yiep Soon for his strong technical support during the study.

References

1 Ramanathan M, Mishra N, Thalmann N M. Nadine humanoid social robotics platform. In: Advances in Computer Graphics. Springer International Publishing, 2019, 490–496
DOI: 10.1007/978-3-030-22514-8_49
2 Pranav D. A robot called erica set to become news anchor in Japan. 2018
3 Mishra N, Ramanathan M, Satapathy R, Cambria E, Magnenat-Thalmann N. Can a humanoid robot be part of the organizational workforce? A user study leveraging sentiment analysis. In: 2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). 2019
DOI:10.1109/ro-man46459.2019.8956349
4 Mishra N, Ramanathan M, Satapathy R, Cambria E, Magnenat-Thalmann N. Can a humanoid robot be part of the organizational workforce? A User Study Leveraging Sentiment Analysis. 2019
5 Mishra N, Baka E, Magnenat Thalmann N. Exploring potential and acceptance of socially intelligent robot. In: Intelligent Scene Modeling and Human-Computer Interaction. Springer International Publishing, 2021, 259–282
DOI:10.1007/978-3-030-71002-6_15
6 Tulsulkar G, Mishra N, Thalmann N M, Lim H E, Lee M P, Cheng S K. Can a humanoid social robot stimulate the interactivity of cognitively impaired elderly? A thorough study based on computer vision methods. The Visual Computer, 2021, 37(12): 3019–3038
DOI: 10.1007/s00371-021-02242-y
7 Mishra N, Tulsulkar G, Li H H, Thalmann N M. Does elderly enjoy playing bingo with a robot? A case study with the humanoid robot Nadine. 2021
8 Mori M. Bukimi no tani [the uncanny valley]. Energy, 1970, 7: 33–35
9 Kätsyri J, Förger K, Mäkäräinen M, Takala T. A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. Frontiers in Psychology, 2015, 6390
DOI: 10.3389/fpsyg.2015.00390
10 Cheetham M. Editorial: the uncanny valley hypothesis and beyond. Frontiers in Psychology, 2017, 81738
DOI: 10.3389/fpsyg.2017.01738
11 Eaton M. Humanoid robots, their simulators, and the reality gap. In: Evolutionary Humanoid Robotics. Part of SpringerBriefs in Intelligent Systems (Artificial Intelligence, Multiagent Systems, and Cognitive Robotics). 2015
12 Minato T, Shimada M, Ishiguro H, Itakura S. Development of an android robot for studying human-robot interaction. Innovations in Applied Artificial Intelligence, 2004
13 Guizzo E. Hiroshi ishiguro: The man who made a copy of himself, 2010
14 Ho C C, MacDorman K F. Revisiting the uncanny valley theory: developing and validating an alternative to the Godspeed indices. Computers in Human Behavior, 2010, 26(6): 1508–1518
DOI: 10.1016/j.chb.2010.05.015
15 Seyama J, Nagayama R S. The uncanny valley: effect of realism on the impression of artificial human faces. Presence: Teleoperators and Virtual Environments, 2007, 16(4): 337–351
DOI: 10.1162/pres.16.4.337
16 Ho C C, MacDorman K F. Measuring the uncanny valley effect. International Journal of Social Robotics, 2017, 9(1): 129–139
DOI: 10.1007/s12369-016-0380-9
17 Softbank Robotics. Nao the humanoid and programmable robot, 2021
18 Tinwell A, Grimshaw M, Nabi D A, Williams A. Facial expression of emotion and perception of the Uncanny Valley in virtual characters. Computers in Human Behavior, 2011, 27(2): 741–749
DOI: 10.1016/j.chb.2010.10.018
19 Tinwell A. The uncanny valley in games and animation. Boca Raton: CRC Press, 2014
20 Palomäki J, Kunnari A, Drosinou M, Koverola M, Lehtonen N, Halonen J, Repo M, Laakasuo M. Evaluating the replicability of the uncanny valley effect. Heliyon, 2018, 4(11): e00939
DOI: 10.1016/j.heliyon.2018.e00939
21 Pollick F. Rearch of the Uncanny valley. 2009, 40(12): 69–78

22 Cheetham M, Suter P, Jäncke L. The human likeness dimension of the "uncanny valley hypothesis": Behavioral and functional MRI findings. Frontiers in Human Neuroscience, 2011, 5126
DOI: 10.3389/fnhum.2011.00126
23 Saygin A P, Chaminade T, Ishiguro H, Driver J, Frith C. The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. Social Cognitive and Affective Neuroscience, 2012, 7(4): 413–422
DOI: 10.1093/scan/nsr025
24 Laakasuo M, Palomäki J, Köbis N. Moral uncanny valley: a robot's appearance moderates how its decisions are judged. International Journal of Social Robotics, 2021, 13(7): 1679–1688
DOI: 10.1007/s12369-020-00738-6
25 Lay S, Brace N, Pike G, Pollick F. Circling around the uncanny valley: design principles for research into the relation between human likeness and eeriness. i-Perception, 2016, 7(6): 204166951668130
DOI: 10.1177/2041669516681309
26 Löffler D, Dörrenbächer J, Hassenzahl M. The uncanny valley effect in zoomorphic robots: the U-shaped relation between animal likeness and likeability. Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction. Cambridge, United Kingdom. New York, ACM, 2020, 261–270
DOI:10.1145/3319502.3374788
27 Schwind V, Leicht K, Jäger S, Wolf K, Henze N. Is there an uncanny valley of virtual animals? A quantitative and qualitative investigation. International Journal of Human-Computer Studies, 2018, 11149–11161
DOI: 10.1016/j.ijhcs.2017.11.003
28 Bartneck C, Kanda T, Ishiguro H, Hagita N. Is the uncanny valley an uncanny cliff? 2007, 9: 368–373
29 MacDorman K F, Ishiguro H. The uncanny advantage of using androids in cognitive and social science research. Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems, 2006, 7(3): 297–337
DOI: 10.1075/is.7.3.03mac
30 Looser C E, Wheatley T. The tipping point of animacy. Psychological Science, 2010, 21(12): 1854–1862
DOI: 10.1177/0956797610388044
31 MacDorman K F, Green R D, Ho C C, Koch C T. Too real for comfort? Uncanny responses to computer generated faces. Computers in Human Behavior, 2009, 25(3): 695–710
DOI: 10.1016/j.chb.2008.12.026
32 MacDorman K F. Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: an exploration of the uncanny valley. In: ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science. 2006
33 Walters M L, Syrdal D S, Dautenhahn K, te Boekhorst R, Koay K L. Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion. Autonomous Robots, 2008, 24(2): 159–178
DOI: 10.1007/s10514-007-9058-3
34 Łupkowski P, Rybka M, Dziedzic D, Włodarczyk W. Human-likeness assessment for the uncanny valley hypothesis. Bio-Algorithms and Med-Systems, 2017, 13(3): 125
DOI: 10.1515/bams-2017-0008
35 McDonnell R, Breidt M. Face reality: investigating the uncanny valley for virtual faces. ACM SIGGRAPH ASIA 2010 Sketches. Seoul, Republic of Korea. New York, ACM, 2010, 1–2
DOI: 10.1145/1899950.1899991
36 Stephanie Claire Lay. The uncanny valley effect. The Open University, 2015
37 Silvera-Tawil D, Garbutt M. The far side of the uncanny valley: 'healthy persons', androids, and radical uncertainty. In: 2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), 2015
38 Rosenthal-von der Pütten A M, Krämer N C, Maderwald S, Brand M, Grabenhorst F. Neural mechanisms for accepting and rejecting artificial social partners in the uncanny valley. The Journal of Neuroscience, 2019, 39(33): 6555–6570
DOI: 10.1523/jneurosci.2956-18.2019
39 Ueyama Y. A Bayesian model of the uncanny valley effect for explaining the effects of therapeutic robots in autism spectrum disorder. PLOS ONE, 2015, 10(9): e0138642
DOI: 10.1371/journal.pone.0138642
40 Moore R K. A Bayesian explanation of the 'Uncanny Valley' effect and related psychological phenomena. Scientific Reports, 2012, 2(1): 864
DOI: 10.1038/srep00864
41 Cheetham M, Wu L, Pauli P, Jancke L. Arousal, valence, and the uncanny valley: psychophysiological and self-report findings. Frontiers in Psychology, 2015, 6981
DOI: 10.3389/fpsyg.2015.00981
42 Bartneck C, Kulić D, Croft E, Zoghbi S. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. International Journal of Social Robotics, 2009, 1(1): 71–81
DOI: 10.1007/s12369-008-0001-3

43 Ho C C, MacDorman K F, Pramono Z A D D. Human emotion and the uncanny valley: a GLM, MDS, and Isomap analysis of robot video ratings. In: Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction. Amsterdam, The Netherlands. New York, ACM, 2008, 169–176
DOI: 10.1145/1349822.1349845
44 Ramanathan M, Satapathy R, Magnenat Thalmann N. Survey of speechless interaction techniques in social robotics. In: Intelligent Scene Modeling and Human-Computer Interaction. Springer International Publishing, 2021, 241–257
DOI: 10.1007/978-3-030-71002-6_14
45 He K M, Zhang X Y, Ren S Q, Sun J. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA, IEEE, 770–778
DOI: 10.1109/cvpr.2016.90
46 Mollahosseini A, Hasani B, Mahoor M H. AffectNet: a database for facial expression, valence, and arousal computing in the wild. IEEE Transactions on Affective Computing, 2019, 10(1): 18–31
DOI: 10.1109/taffc.2017.2740923
47 Aouani H, Ayed Y B. Speech emotion recognition with deep learning. Procedia Computer Science, 2020, 176251–176260
DOI: 10.1016/j.procs.2020.08.027
48 Cunningham S, Ridley H, Weinel J, Picking R. Supervised machine learning for audio emotion recognition. Personal and Ubiquitous Computing, 2021, 25(4): 637–650
DOI: 10.1007/s00779-020-01389-0
49 Trigeorgis G, Ringeval F, Brueckner R, Marchi E, Nicolaou M A, Schuller B, Zafeiriou S. Adieu features? End-to-end speech emotion recognition using a deep convolutional recurrent network. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing. Shanghai, China, IEEE, 5200–5204
DOI: 10.1109/icassp.2016.7472669
50 Livingstone S R, Russo F A. The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): a dynamic, multimodal set of facial and vocal expressions in North American English. PLOS ONE, 2018, 13(5): e0196391
DOI: 10.1371/journal.pone.0196391
51 Dupuis K, Pichora-Fuller M. Toronto emotional speech set (TESS)-Younger talker_Neutral, 2010
52 Pichora-Fuller M K, Dupuis K. Toronto emotional speech set (TESS). Scholars Portal Dataverse, 2020
53 Susanto Y, Livingstone A G, Ng B C, Cambria E. The hourglass model revisited. IEEE Intelligent Systems, 2020, 35(5): 96–102
DOI: 10.1109/MIS.2020.2992799
54 Cambria E, Livingstone A, Hussain A. The hourglass of emotions. In: Cognitive behavioural systems, 2012
55 Greenhouse S W, Geisser S. On methods in the analysis of profile data. Psychometrika, 1959, 24(2): 95–112
DOI: 10.1007/bf02289823
56 Mori M, MacDorman K F, Kageki N. The uncanny valley. IEEE Robotics & Automation Magazine, 2012, 19(2): 98–100
57 Kamel A, Sheng B, Li P, Kim J, Feng D D. Hybrid refinement-correction heatmaps for human pose estimation. IEEE Transactions on Multimedia, 2021, 231330–231342
DOI: 10.1109/tmm.2020.2999181
58 Ramanathan M, Yau W Y, Teoh E K. Improving human body part detection using deep learning and motion consistency. In: 2016 14th International Conference on Control, Automation, Robotics and Vision (ICARCV). Phuket, Thailand, IEEE, 2016, 1–5
DOI: 10.1109/icarcv.2016.7838651
59 Ramanathan M, Yau W, Teoh E. Human posture detection using H-ELM body part and whole person detectors for human-robot interaction. 2016
60 Aouaidjia K, Sheng B, Li P, Kim J, Feng D D. Efficient body motion quantification and similarity evaluation using 3-D joints skeleton coordinates. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2021, 51(5): 2774–2788
DOI: 10.1109/TSMC.2019.2916896
61 Ertugrul E, Li P, Sheng B. On attaining user-friendly hand gesture interfaces to control existing GUIs. Virtual Reality & Intelligent Hardware, 2020, 2(2): 153–161
DOI: 10.1016/j.vrih.2020.02.001