The Exploration of the Uncanny Valley from the Viewpoint of the Robot's Nonverbal Behaviour

Chidchanok Thepsoonthorn[1] · Ken-ichiro Ogawa[1] · Yoshihiro Miyake[1]

Abstract

Many studies have been conducted to find approaches to overcome the Uncanny Valley. However, the focus on the influence of the robot's appearance leaves a big missing part: the influence of the robot's nonverbal behaviour. This impedes the complete exploration of the Uncanny Valley. In this study, we explored the Uncanny Valley from the viewpoint of the robot's nonverbal behaviour in regard to the Uncanny Valley hypothesis. We observed a relationship between the participants' ratings on human-likeness of the robot's nonverbal behavior and affinity toward the robot's nonverbal behavior, and define the point where the affinity toward the robot's nonverbal behavior significantly drops down as the Uncanny Valley. In this study, an experiment of human–robot interaction was conducted. The participants were asked to interact with a robot with different nonverbal behaviours, ranging from 0 (no nonverbal behavior, speaking only) to 3 (gaze, head nodding, and gestures) combinations and to rate the perceived human-likeness and affinity toward the robot's nonverbal behavior by using a questionnaire. Additionally, the participants' fixation duration was measured during the experiment. The result showed a biphasic relationship between human-likeness and affinity rating results. A curve resembling the Uncanny Valley is found. The result was also supported by participants' fixation duration. It showed that the participants had the longest fixation at the robot when the robot expressed the nonverbal behaviours that fall into the Uncanny Valley. This exploratory study provides evidence suggesting the existence of the Uncanny Valley from the viewpoint of the robot's nonverbal behaviour.

Keywords Uncanny Valley · Robot's nonverbal behaviours · Human-likeness · Affinity · Fixation duration

1 Introduction

In 1970, Prof. Masahiro Mori introduced the term "Uncanny Valley" to scientific research societies [1]. He described the "Uncanny Valley" as the phenomenon where human's positive feeling significantly drops down and turns to negative response as human-likeness increases. Later, the studies of MacDorman and Ishiguro further emphasized the impacts of the Uncanny Valley in human–robot interaction [2, 3]. The awareness of the Uncanny Valley effect has been increasingly arisen in these passing decades, especially in social robot and computer-generated character fields. Many studies

✉ Chidchanok Thepsoonthorn
  chidchanok.thepsoonthorn@gmail.com

1 Department of Computer Science, Tokyo Institute of Technology, Yokohama, Japan

started to investigate the influences of robot/character features to overcome or to avoid the Uncanny Valley effect while endeavoring to find approaches to increase likability or positive feelings toward robots. Various aspects have been taken into the account in previous studies, for example, human-likeness [4–6], familiarity [2], affinity/likability [7, 8], anthropomorphism [9], and eeriness [10–12]. And the uncanny feelings have been interpreted in several dimensions like low rating in familiarity and affinity [6, 11, 13, 14]. These previous studies aimed to find approaches to achieve the highest rating in affinity and human-likeness and observe the causalities of the Uncanny Valley phenomena in order to enhance more positive feelings toward robots and pleasantness in human–robot interaction.

This study is inspired by the fact that the majority of the researches in robotics and human–robot interaction paid main attention to the investigation on the Uncanny Valley from the viewpoint of the robot's appearance. However, apart from the robot's appearance, robot's nonverbal behaviours may play a very important role in human–robot interaction as

well. According to the studies in human–human interaction, there is some evidence indicating that a large part of interaction is expressed nonverbally [15–17], for instance, via gaze [18, 19], head nodding [19, 20], and gestures [21]. Many studies also supported that nonverbal behaviours have a major impact on human's impression [22] and interaction engagement [23, 24]. Most of the studies in human–robot interaction adopt guidelines from the studies in human–human interaction as references to investigate whether the phenomena that are observed in human–human interaction are also found in human–robot interaction. Therefore, the focus on the influence of the robot's appearance leaves a big missing part: the influence of the robot's nonverbal behaviour. This impedes the complete exploration of the Uncanny Valley in human–robot interaction.

In this study, we aimed to investigate and explore the Uncanny Valley from the viewpoint of the robot's nonverbal behaviour in regard to the Uncanny Valley hypothesis. Many previous studies of the robot's appearance observed the relationship between the participants' ratings on human-likeness and affinity toward the robot and indicated the Uncanny Valley as the point where the affinity rating significantly decreases [8, 9, 14, 25]. The previous studies also addressed that the violation of expectation (e.g. conflicting cues and perceptual mismatch) is one of the influence factors leading to the Uncanny Valley phenomena from the viewpoint of the robot's appearance as well [26–28]. And it is intriguing to investigate whether such expectation violation theories are also applicable to the viewpoint of the robot's nonverbal behavior or not.

In order to investigate and explore the Uncanny Valley from the viewpoint of the robot's nonverbal behaviour, we conducted a human–robot interaction experiment. The participants were asked to rate the perceived human-likeness and affinity toward the robot's nonverbal behaviour after the interaction with a robot using a questionnaire based on Ho and MacDorman's 2010 and 2017 studies. We implemented 15 robot's behaviour conditions with the combinations of the following behaviours: speaking only (no nonverbal behaviour), face tracking, gaze, head nodding responding, head nodding, and gestures. The robot's behaviour conditions range from 0 nonverbal behavior (speaking only) to 3 combinations of the 5 robot's nonverbal behaviours. The number of nonverbal behavior combination level (level 0 to 4) was set as the discrete parameter in order to obtain different levels of the perceived human-likeness and affinity ratings from humans and to investigate the causality of the Uncanny Valley phenomena from the robot's nonverbal behaviour or the combination of robot's nonverbal behaviour. We hypothesized that the Uncanny Valley also exists from the viewpoint of the robot's nonverbal behaviour in which the significant decrease in affinity rating is observed.

2 Methodology

In this experiment, the robot gave a TED talk to each participant for 15 trials. For each trial, one of the 15 robot's behaviour conditions is employed in random order.

2.1 Robot's Nonverbal Behaviours and Its Behaviour Conditions

NAO robot was used in this study. The NAO robot is a humanoid robot developed by Aldebaran Robotics (France). It is equipped with sensors, gyroscopes, accelerometer, microphone, speaker, and camera. Additionally, the NAO robot is equipped with a software suite, which allows the developers to fully program and control the NAO robot platform (SDK package with NAOqi API). The reason that we selected the NAO robot for this study is due to the study of the quantitative of cartography of the Uncanny Valley [5]. The study indicated that the NAO robot's appearance is rated in the middle range of human-likeness based on the participants' rating. It can be inferred that the NAO robot's appearance is neutral, neither too non-human-like nor too human-like. Therefore, its appearance has low potential to cause uncanny feelings and we could simply ask the participants to overlook the NAO robot's appearance and pay more attention to the robot's behaviours instead. We programmed 5 main robot's nonverbal behaviours, which are face tracking, gaze shifting, nodding initiating, nodding responding, and gestures, using Python language.

In giving a talk context, the behavioural cues between the speaker and the listener are expressed in form of backchanneling behaviours [17] and turn-taking behaviours [29] in which gaze, head nodding, and gesture behaviours are included. In case of gaze, it is asserted to be the means for expressing not only attention and interest [30] but also positive feelings like affection [31] to each other. During the interaction, the robot tracked the participants' face to create eye contact and averted its gaze from the participant from time to time to break eye contact, which is similar to human–human interaction and complied with the findings of the momentary fixation at their partner in face-to-face communication and collaborative task in the previous studies [24, 32] For head nodding, it is asserted to be one of the obvious and prominent nonverbal behaviours that the interactional partner can easily perceive, representing attentiveness and acceptance to the interactional partner, [33, 34]. Furthermore, humans also initiate their own nodding [15] and respond to interactional partner's head nodding as backchanneling behaviour as well [17]. In this study, the robot will initiate its head nodding from time to time with random frequencies when no head nodding is detected. It will also nod with random frequencies as a backchanneling response to the detected human head nodding as well. These initiating and

responding behaviours are similar to human–human interaction in speaker–listener context. For gestures, it is served as a compensatory role for speech [35] and is evinced that gestures and speech together form an integrated communication system in the speaker–listener context [21, 36, 37]. To allow the robot to express gestures while speaking (including hand gestures and body language), the ALAnimatedSpeech module from NAOqi API was used. We applied the contextual mode and word tagging feature, which maps the content that the robot spoke with the related gesture and body language animations in order to minimize the mismatch effect between the speech content and the robot's gestures, for instance, wave its hand while saying "Hi, I'm Nao", open palm and bring it to the front while saying "You", open palm and touch the chest when saying "I or Me", and open both palms and bring them to the front while saying "For example" (Supplementary Note 1).

Seven robot's behaviour features (Speaking, face tracking, gaze, head nodding responding, head nodding, gestures, and Kinect) were implemented as described in Table 1. With the combinations of 7 robot's behaviour features regarding the number of nonverbal behavior combination level and the obviousness of the nonverbal behaviour, we generated 15 robot's behaviour conditions. Figure 1 illustrates the robot's nonverbal behaviour combination level, ranging from 0 to 4. Level 0 is speaking only level (no other nonverbal behaviour). Level 1 is the fundamental level with 5 components. Each component consists of 1 fundamental nonverbal behavior (face tracking, gaze, head nodding responding, head nodding, and gestures). Level 2 is a 2-combination level with 6 components. Each component consists of 2 combinations from the 5 fundamental nonverbal behaviours. Level 3 is a 3-combination level with 2 components. Each component consists of 3 combinations from the 5 fundamental nonverbal behaviours. And Level 4 is the Kinect level. It consists of all 5 fundamental nonverbal behaviours, controlled by Kinect. Here, we defined Condition 1 (speaking only) as a baseline condition (hereinafter referred to as baseline condition) and Kinect condition (Condition 15) as the highest level in our experiment. Here, we implemented the Kinect condition to ensure that it is the most human-like condition and referred to as the most human-like nonverbal behaviours in this study. For other conditions, we used the NAO robot's nonverbal behaviour capabilities itself to gain the evaluation on its actual nonverbal behaviour performance.

The 15 conditions were divided into three groups from the viewpoint of the robot's nonverbal behaviour, gaze-based, head nodding-based, and gesture-based. The first group (gaze-based) includes the baseline condition and the conditions that consist of face tracking and gaze behaviour only (Conditions 1–3). The second group (head nodding-based) includes the conditions that consist of head nodding behaviours with no gestures (Conditions 4, 5, 7, 8). In particular, Conditions 4 and 5 are comprised of head nodding behaviours only. These fundamental conditions are hereinafter called pure head nodding-based group. The third group (gesture-based) includes the conditions that consist of gestures (Conditions 6, 9–15). In particular, Condition 6 is comprised of gestures only. This fundamental condition is hereinafter called pure gesture-based condition.

For the controlling method in Kinect condition, we employed the Wizard-of-Oz technique. The Microsoft Kinect was used to detect the experimenter's movements and create the experimenter's skeleton data. In order to control the NAO robot's movements according to the experimenter's movements, we converted the Kinect's axis system to the NAO robot's Euler angles using a framework and conversion matrix from the previous studies [38, 39] and mapped the coordinates to move the NAO robot's joint angles accordingly.

2.2 Experimental Design

The experiment task was face-to-face, human–robot interaction. The experiment was conducted in between participants-within condition design. The NAO robot gave a TED talk to each participant. The talk was divided into 15 parts. Each part corresponded to each trial (15 trials in total). In 15 trials, the robot expressed random order of the 15 robot's behaviour conditions in order to avoid the order effect. The duration of each trial was approximately 2 min. Twenty international participants participated in this experiment. The age range of the participants was from 23 to 35 years old (M = 27.35, SD = 2.54), 11 males and 9 females. All participants had prior experience of interacting with robots and studied in technology fields. To exclude language difficulty effect, all participants were required to be native English speakers or pass standard English proficiency test (TOEFL PBT > 525, TOEFL iBT > 71, TOEIC > 780, IELTS > 6 or SAT). Informed consent was obtained from all participants before participating in the experiment.

Figure 2 illustrates the experimental setup of this experiment. All participants were not informed and no one noticed about using Kinect during the experiment at all. There was a partition between the interaction area and the control area. In the control area, the experimenter could see the participant's behaviors via a monitor. All robot control procedures were done behind the partition (with/without using Kinect). In the interaction area, there was a video camera set behind the robot to capture the participant's behaviours. The participant and the robot were faced directly to each other, 90 cm apart. The robot took a standing position to perform gestures while the participant took a sitting position. To minimize the external factors that might affect the participant's attention and judgment (e.g. robot's appearance and verbal behaviour), we asked the participant to pay attention to the robot's non-

**Table 1** Descriptions of robot's behaviour features

Robot behaviour features	Description
1. Speaking	Common behaviour in every condition (baseline)
2. Face tracking	Track human's face to make eye contact
3. Gaze (face tracking and gaze shifting)	Both make and break gaze with human similar to human–human interaction [58, 59]
4. Head nodding responding	Nod as a response when human nods
5. Head nodding (head nodding responding and head nodding initiating)	Both respond to human's nodding and initiate its own head nodding similar to human–human interaction [15, 17]
6. Gestures	Express gestures while speaking (hand gestures and body language)
7. Kinect	The experimenter uses Wizard-of-Oz technique to control robot's nonverbal behaviours (including gaze, head nodding, and gestures) using Microsoft Kinect 2 where the robot mimics the experimenter's nonverbal behaviours. The experimenter perceives the participants' nonverbal behaviours via monitor screen and interacts with them via the robot, behind the partition
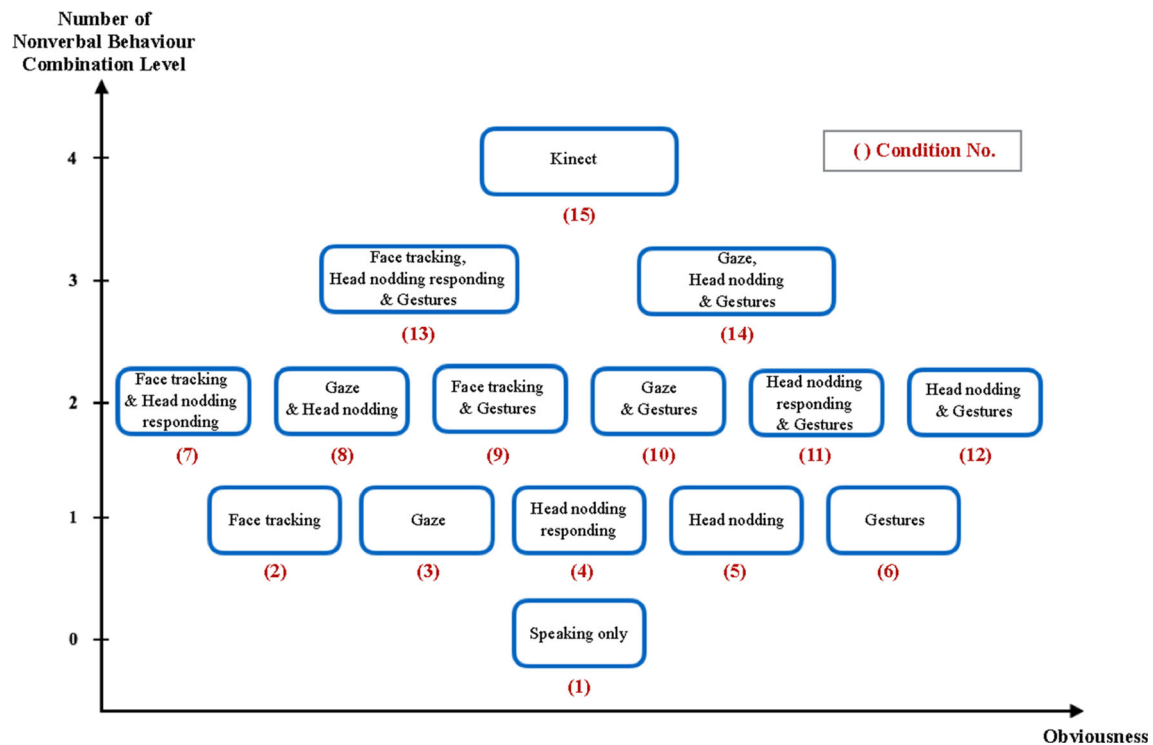
Fig. 1 Robot's behaviour conditions

verbal behaviours and informed them to ignore the robot's appearance and speech variables, for example, pronunciation, speed, spacing, and tone. For minimal saliency effects, there were neither experimenter nor movable objects presented in the interaction area during the experiment.

This experiment consisted of three sessions: pre-interaction, interaction, and evaluation. In the pre-interaction

session, the participant was asked to complete a paper-based questionnaire about his/her demographic and experience with robots. In the interaction session, the participant interacted with the NAO robot for 2 min. Each trial corresponded to each part of the robot's speech content and the random order of the robot's behaviour conditions. For the robot's speech content details, please see Supplementary

Fig. 2 Experimental setup
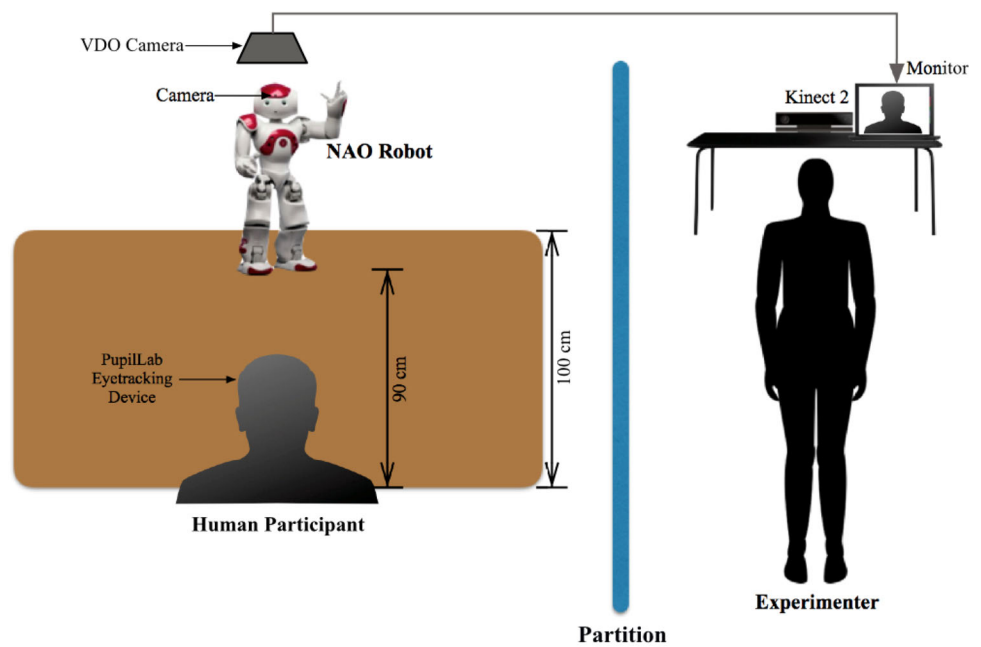
In the evaluation session, the participant was asked to complete a paper-based questionnaire after each trial. The questionnaire was designed to assess the participant's feelings about the robot's nonverbal behaviours in terms of perceived human-likeness and affinity toward the robot using Ho and MacDorman's 2010 and 2017 studies. The sequence of interaction sessions and evaluation sessions repeated for 15 times to complete the experiment. For the highest likability, as adopted from the previous studies in the robot's appearance [2, 11], a positive correlation between human-likeness and affinity ratings was required (highest human-likeness and affinity ratings).

Additionally, as the human-likeness and affinity ratings were both dependent and subjective variables in this study, we decided to measure human's fixation duration at the robot as an implicit measure to support and affirm the effect of the Uncanny Valley and the influence of robot's nonverbal behaviour. According to the previous studies, the correlations between likability and duration of eye fixations have been reported [40–43]. These previous studies asserted that the human's gaze or duration of eye fixations on the robot can be used as a measurement to evaluate the robot's perceived human-likeness and affinity toward the robot. In this study, the PupilLab gaze tracking device and software were used to capture and measure the participants' gaze fixation at the robot during the interaction.

3 Experimental Results

In this section, we analyzed the questionnaire results of the human-likeness and affinity ratings from 20 participants and the fixation duration of the participants to support the questionnaire results.

The questionnaire for evaluating the robot's nonverbal behaviour consisted of four questions (7-scale rating) for human-likeness and another four questions (7-scale rating) for affinity. We categorized the participants' rating scores into the two aspects accordingly. The participants' rating scores were calculated in percentage to normalize the data across the two aspects. We also calculated the mean values of each behaviour condition. Once we obtained the rating tendencies of both human-likeness and affinity aspects, the relationship between the two aspects was examined in order to fit into the Uncanny Valley hypothesis. We found a biphasic relationship between human-likeness and affinity of the robot's nonverbal behaviour, which demonstrates a curve resembling the Uncanny Valley (Fig. 3). The smoothing curve is created from JMP Statistical Data Analysis Software's cubic spline function with a default lambda of 0.05 and standardized human-likeness and likability rating values. The result supports the classification of the three groups from the viewpoint of the robot's nonverbal behaviour, as defined in Sect. 2.1. It can be classified based on the result of the Wilcoxon Signed-rank test of the human-likeness rating score as well. The first group (gaze-based group) has no significant difference from the baseline condition, and is represented in blue. The second group (head nodding-based group) has a significant difference from the baseline condition, and represented in red. The third group (gesture-based group) has a significant difference from those baseline condition and the second group, and is represented in green.

According to the result, it shows that the head nodding-based group (Condition 4, 5, 7, 8) gained less affinity than the
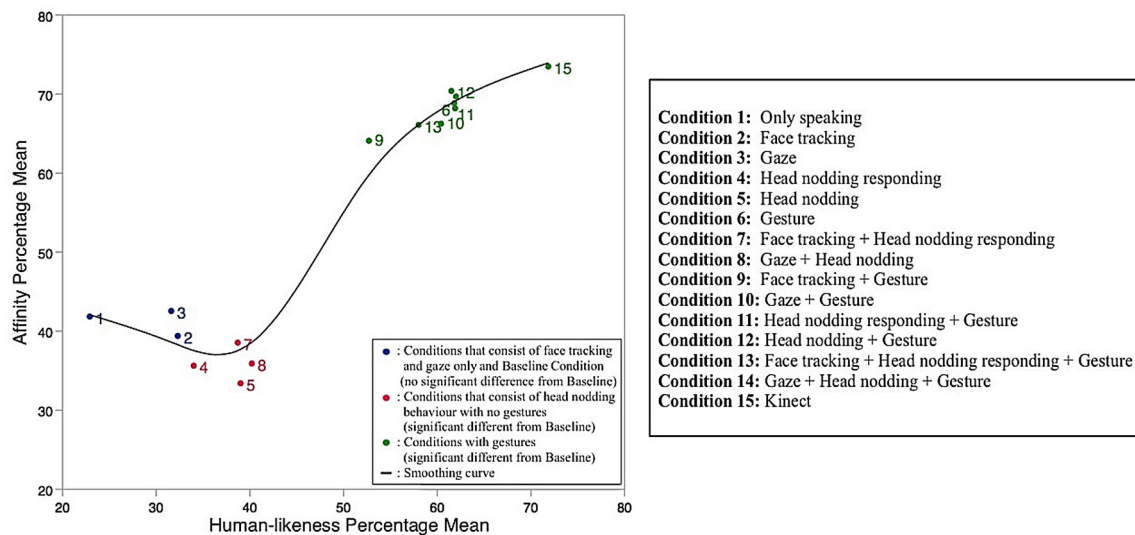
Fig. 3 Result of the evaluation on human-likeness and affinity of the robot's nonverbal behaviours

baseline (Condition 1) though it is perceived as more human-like. On the other hand, the gesture-based group (Condition 6, 9–15) is highly rated in both human-likeness and affinity aspects. We further conducted a Wilcoxon Signed-Ranks test to affirm the statistical significance of the human-likeness and affinity ratings between the head nodding-based group (Condition 4, 5, 7, 8), and the baseline condition (Condition 1) and the condition next to the Uncanny Valley(Condition 9), hereinafter referred to as the condition next to the Uncanny Valley.

For human-likeness aspect, the Wilcoxon Signed-Ranks test indicates that the head nodding-based group (Condition 4, 5, 7, 8) are perceived as significantly more human-like than the baseline condition (Condition 4 and Condition 1: $Z = -2.69$, $p = 0.0072$; Condition 5 and Condition 1: $Z = -3.25$, $p = 0.0012$; Condition 7 and Condition 1: $Z = -2.98$, $p = 0.0029$; Condition 8 and Condition 1: $Z = -3.12$, $p = 0.0018$) and significantly less human-like than the condition next to the Uncanny Valley (Condition 4 and Condition 9: $Z = 3.42$, $p = 0.0006$; Condition 5 and Condition 9: $Z = 2.61$, $p = 0.0090$; Condition 7 and Condition 9: $Z = 2.7$, $p = 0.0069$; Condition 8 and Condition 9: $Z = 2.48$, $p = 0.0129$). Furthermore, there is a significant difference between the pure head nodding-based group (Conditions 4 and 5) and the pure gesture-based condition (Condition 6) (Condition 4 and 6: $Z = 4.05$, $p < 0.0001$; Condition 5 and 6: $Z = 3.32$, $p = 0.0009$).

In case of affinity aspect, the Wilcoxon Signed-Ranks test indicates that pure head nodding-based group (Condition 4, 5) gains significantly less affinity than the baseline condition (Condition 4 and Condition 1: $Z = 2.20$, $p = 0.0275$; Condition 5 and Condition 1: $Z = 2.71$, $p = 0.0068$). Also, the affinity rating of the head nodding-based group (Condition

4, 5, 7, 8) is significantly less than the affinity rating of the condition next to the Uncanny Valley (Condition 4 and Condition 9: $Z = 4.41$, $p < 0.0001$; Condition 5 and Condition 9: $Z = 4.68$, $p < 0.0001$; Condition 7 and Condition 9: $Z = 3.81$, $p = 0.0001$; Condition 8 and Condition 9: $Z = 4.18$, $p = < 0.0001$). In addition, there is a significant difference between the pure head nodding-based group (Conditions 4 and 5) and the pure gesture-based group (Condition 6) (Condition 4 and 6: $Z = 5.06$, $p < 0.0001$; Condition 5 and 6: $Z = 5.21$, $p < 0.0001$). On the contrary, no significant difference is found within the same group (except Condition 15: Kinect condition). Table 2 shows the details of the Wilcoxon Signed-Ranks test results.

For the result of fixation duration, an implicit measure to support the questionnaire results, we found that the head nodding-based group (Condition 4, 5, 7, 8) gained the longest fixation from the participants while the gesture-based group, except Condition 9 and 14, (Condition 6, 10–13, 15) gained shorter fixation than the baseline condition. See Fig. 4. A Wilcoxon Signed-Ranks test indicates that the fixation duration of the head nodding-based group (Condition 4, 5, 7, 8) is significantly longer than the fixation duration of the baseline condition (Condition 4 and Condition 1: $Z = -2.66$, $p = 0.0077$; Condition 5 and Condition 1: $Z = -2.64$, $p = 0.0084$; Condition 7 and Condition 1: $Z = -1.99$, $p = 0.0468$; Condition 8 and Condition 1: $Z = -2.79$, $p = 0.0053$). It also indicates that the fixation duration of Condition 4, 5, 8 are significantly longer than the fixation duration of the condition next to the peak (Condition 4 and Condition 9: $Z = -1.96$, $p = 0.0499$; Condition 5 and Condition 9: $Z = -2.04$, $p = 0.0411$; Condition 8 and Condition 9: $Z = -2.12$, $p = 0.0337$). No significant difference is found within the
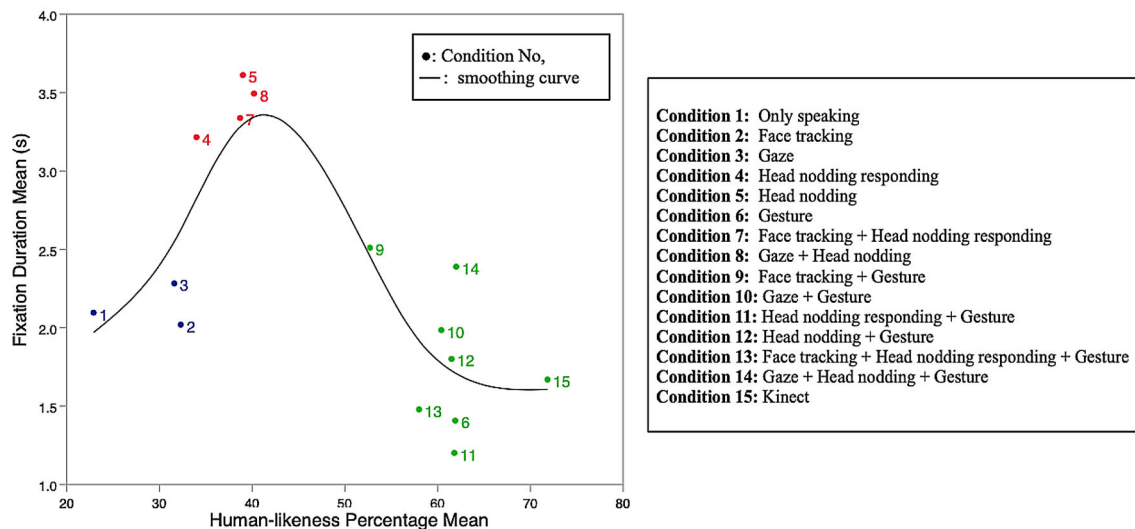
Table 2 Wilcoxon signed-ranks test and Bonferroni correction results

Condition		Face-to-face Interaction		
		Human-likeness	Affinity	Fixation Duration
1■	4↓	$Z = -2.69, p = 0.0072$** Bonferroni-corrected $p = 0.00036$ Median of difference $= 10.71$	$Z = 2.20, p = 0.0275$* Bonferroni-corrected $p = 0.001375$ Median of difference $= -7.14$	$Z = -2.66, p = 0.0077$ ** Bonferroni-corrected $p = 0.000385$ Median of difference $= 1.57$
1■	5↓	$Z = -3.25, p = 0.0012$** Bonferroni-corrected $p = 0.00006$ Median of difference $= 14.29$	$Z = 2.71, p = 0.0068$** Bonferroni-corrected $p = 0.00034$ Median of difference $= -7.14$	$Z = -2.64, p = 0.0084$** Bonferroni-corrected $p = 0.00042$ Median of difference $= 1.65$
1■	7↓	$Z = -2.98, p = 0.0029$** Bonferroni-corrected $p = 0.000145$ Median of difference $= 14.29$	$Z = 0.38, p = 0.7$ Bonferroni-corrected $p = 0.035$ Median of difference $= 0.00$	$Z = -1.99, p = 0.0468$* Bonferroni-corrected $p = 0.00234$ Median of difference $= 0.84$
1■	8↓	$Z = -3.12, p = 0.0018$** Bonferroni-corrected $p = 0.00009$ Median of difference $= 17.86$	$Z = 1.2, p = 0.23$ Bonferroni-corrected $p = 0.0115$ Median of difference $= -3.57$	$Z = -2.79, p = 0.0053$ ** Bonferroni-corrected $p = 0.000265$ Median of difference $= -2.79$
9★	4↓	$Z = 3.42, p = 0.0006$** Bonferroni-corrected $p = 0.00003$ Median of difference $= 21.43$	$Z = 4.41, p < 0.0001$** Bonferroni-corrected $p < 0.000005$ Median of difference $= 28.57$	$Z = -1.96, p = 0.0499$* Bonferroni-corrected $p = 0.002495$ Median of difference $= -1.29$
9★	5↓	$Z = 2.61, p = 0.0090$** Bonferroni-corrected $p = 0.00045$ Median of difference $= 17.86$	$Z = 4.68, p < 0.0001$** Bonferroni-corrected $p < 0.000005$ Median of difference $= 32.41$	$Z = -2.04, p = 0.0411$* Bonferroni-corrected $p = 0.002055$ Median of difference $= -1.19$
9★	7↓	$Z = 2.7, p = 0.0069$** Bonferroni-corrected $p = 0.000345$ Median of difference $= 17.86$	$Z = 3.81, p = 0.0001$ ** Bonferroni-corrected $p = 0.000005$ Median of difference $= 25.00$	$Z = -1.5, p = 0.1333$ Bonferroni-corrected $p = 0.006665$ Median of difference $= -0.64$
9★	8↓	$Z = 2.48, p = 0.0129$* Bonferroni-corrected $p = 0.000645$ Median of difference $= 17.86$	$Z = 4.18, p < 0.0001$** Bonferroni-corrected $p < 0.000005$ Median of difference $= 32.14$	$Z = -2.12, p = 0.0337$* Bonferroni-corrected $p = 0.001685$ Median of difference $= -1.14$

■Baseline Condition ↓Conditions at the bottom of Uncanny Valley or peak ★Condition next to the Uncanny Valley or peak
The original test significance level of p value is < 0.05
The Bonferroni-corrected significance level of p value is < 0.0025 [original p value (0.05)/number of test (20)]

Condition 1: Only speaking
Condition 2: Face tracking
Condition 3: Gaze
Condition 4: Head nodding responding
Condition 5: Head nodding
Condition 6: Gesture
Condition 7: Face tracking + Head nodding responding
Condition 8: Gaze + Head nodding
Condition 9: Face tracking + Gesture
Condition 10: Gaze + Gesture
Condition 11: Head nodding responding + Gesture
Condition 12: Head nodding + Gesture
Condition 13: Face tracking + Head nodding responding + Gesture
Condition 14: Gaze + Head nodding + Gesture
Condition 15: Kinect

Fig. 4 Result of the participants' fixation duration at the robot

same group (except Condition 15: Kinect condition). Table 2 shows the details of the Wilcoxon Signed-Ranks test results.

In order to avoid the Type I error, which is the rejection of a true null hypothesis (false positive finding), the Bonferroni correction was also calculated on the human-likeness, affinity, and fixation duration results. Table 2 shows the Bonferroni correction results.

4 Discussion

4.1 Relationship Between the Uncanny Valley and Human's Fixation Duration

Figure 5 illustrates the comparison of the 2 tendencies obtained from the analysis results, which are the Uncanny Valley tendency from the questionnaire result and the Bell-shaped tendency of the fixation duration result. According to the results, the Uncanny Valley is found when the robot expressed head nodding behaviour with no gestures. Furthermore, the influence of the robot's head nodding behaviour with no gesture is also supported by the result of fixation duration as shown in Fig. 5. The conditions that consist of head nodding behaviour with no gesture, which fall into Uncanny Valley (low affinity rating), are corresponding to the peak of the bell-shaped tendency and gained the longest fixation duration from the participants. A possible reason is that the participants were curious and tried to understand the robot's intention of expressing its behaviour in a certain way when they interacted with the robot in the conditions that fall into Uncanny Valley. This can result in a longer fixation from the participants. This circumstance is also supported by the previous studies indicating that the incongruous signals or the difficulty of the stimulus influence and increase curiosity and attention, which causes longer fixation duration from the observer [44–46].

4.2 Possible Strategy to Overcome the Uncanny Valley from the Viewpoint of the Robot's Nonverbal Behaviour

In this study, we found a strong positive influence of robot's gestures. The finding reveals that when the robot expressed gestures, it is highly rated in both human-likeness and affinity, which can be considered as a possible strategy to overcome the Uncanny Valley from the viewpoint of the robot's nonverbal behaviour. Many previous studies also supported that gestures are very important to complete speech components and these two modalities have a strong connection to the principles of gesture-speech integration [21, 35–37, 47]. McNeill's study revealed that gestures are found alongside of speech in every spoken language and gestures are a robust component and pervasive for human when talking.

For instance, humans still express gestures while talking on the phone where no observer is present [21]. Based on these previous studies, it suggests that in giving a talk context, gestures may fulfill human's expectation and familiarity on gestures-speech relationship similar to human–human interaction. It evinces that in the context that the robot gives a talk, no other nonverbal behaviour is as important as gestures. Thus, to overcome the Uncanny Valley from a viewpoint of the robot's nonverbal behaviour, especially in giving a talk context, gestures should be considered as a fundamental element of the robot's nonverbal behaviour in order to gain more likability from humans.

4.3 Existence of the Uncanny Valley from the Viewpoint of the Robot's Nonverbal Behaviour

According to the result, it demonstrates that a curve resembling the Uncanny Valley when fitting the human-likeness and affinity ratings on the robot's nonverbal behaviours into the Uncanny Valley hypothesis. Based on the Uncanny Valley hypothesis, the Uncanny Valley can be examined by observing the relationship between human-likeness and affinity toward the robot and is defined as the point where the affinity rating significantly drops down [7, 8, 14, 25]. The results of this study provide evidence suggesting that the Uncanny Valley also exists from the viewpoint of the robot's nonverbal behaviour. However, the Uncanny Valley found in this study seems to be small compared to the previous studies from the viewpoint of the robot's appearance [3, 6].

The main reason might lie in the fact that the appearance and human-like attributes of the NAO robot are still neutral, which are neither too non-human-like nor too human-like. This leads to a low expectation of its nonverbal behaviours and no extreme evaluation on its nonverbal behaviour is observed yet. Saygin et al.'s study revealed that an agent with a very human-like appearance may cause the users to have expectations for more human-like movement than what the agent is capable of exhibiting, and this can then be detected as contributing to the Uncanny Valley effect [26]. Prakash and Rogers's study also supported that humans may ascribe human-like attributes and expectations on the robot when its appearance is human-like, which influences their acceptance and behaviour toward the robot [48]. Moreover, in Paepcke and Takayama's study, they found that the influence of the robot's appearance can set human's expectation on its behaviours and capabilities [49]. These previous studies suggested that when interacting with a human-like robot, humans also have high expectation from the robot in terms of its abilities to act or behave like a human, in other words, the more human-like appearance, the more expectation of human-like behaviour. If the robot with human-like appearance cannot keep up with humans' expectation, it might lead to a stronger
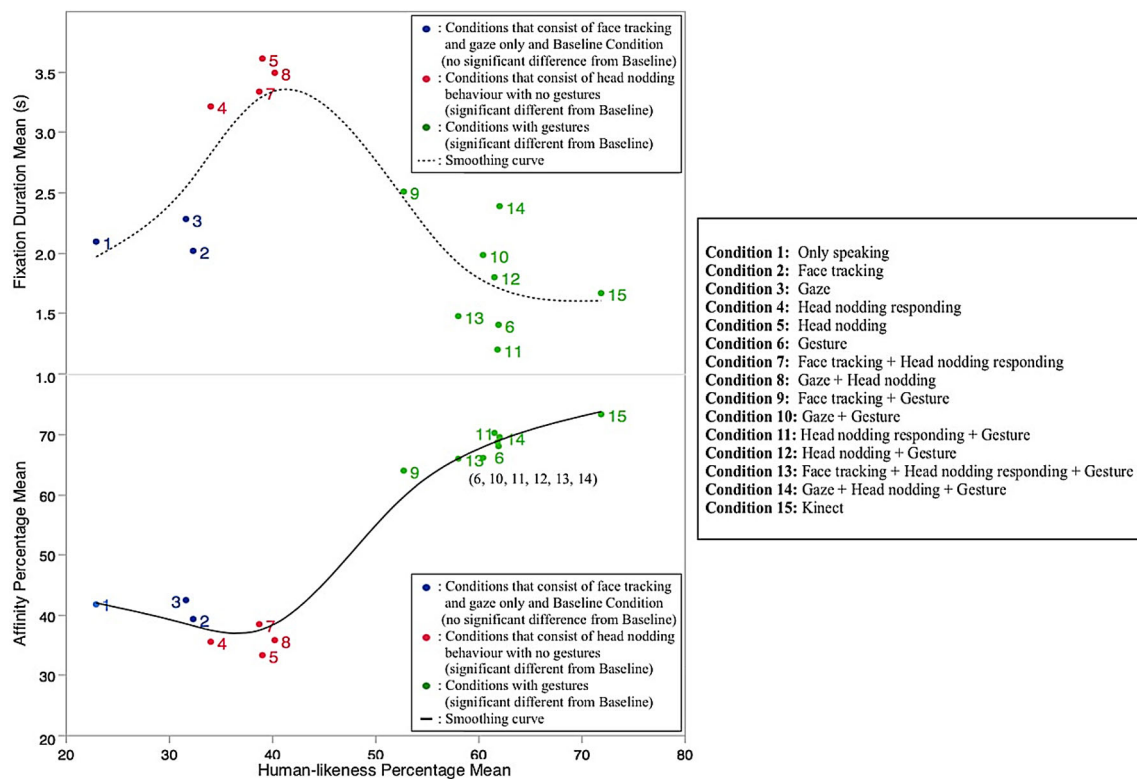
Fig. 5 Comparison of the 2 tendencies obtained from the analysis results. 1. The Bell-shaped tendency from the fixation duration result (Upper). 2. The Uncanny Valley tendency from the evaluation result (Lower)

rejection or negative responses. With the supporting evidence from the previous studies, we strongly believe that when deploying a robot with a more human-like appearance in this experiment, more significance and deeper Uncanny Valley will be detected due to the increase in humans' expectation. All in all, even though the Uncanny Valley found in this study is small, this study can be considered as the first finding of the Uncanny Valley from the viewpoint of the robot's nonverbal behaviour.

4.4 Possible Influence Factors Leading to the Uncanny Valley from the Viewpoint of the Robot's Nonverbal Behaviour

According to the Uncanny Valley found in this study, the robot's head nodding behaviour with no gesture is located at the bottom of the Uncanny Valley where the least affinity rating is found. The influence factors leading to the Uncanny Valley from the viewpoint of the robot's nonverbal behaviour may lie in the consequences of the violation of expectation effect (perceptual mismatch, and conflicting cues) [12, 26–28] and the double bind effects [50].

The violation of the expectation effect is firstly proposed by Prof. Mori. He proposed that the eerie feelings are caused by the mismatch between visual and tactical information

obtained from the prosthetic hand [1]. Additionally, Moore's study provided a mathematical model (Bayesian model) of categorical perception that indicates how stimuli conflicting cues can cause a perceptual tension at category boundaries that leads to the uncanny phenomena [28]. Other studies also demonstrated the variety of cross-modal mismatches that can cause the uncanny feeling, for instance, face-voice mismatch [12], robot's appearance-action mismatch [26], and perceptual mismatch between individual features like artificial eyes on a human-like face [27]. However, these studies are observed regarding the viewpoint of the robot's appearance. Although we examined limited combinations of robot's nonverbal behaviours in this study, the results also present the effect of conflicting cues and align with the previous studies.

Another possible influence factor is the double bind effect. Bateson et al.'s study describes the double bind effect as an emotionally distressing situation in communication where conflicting messages are received, and one message is against another. The messages can be conveyed via words, tone of voice, or body language [50]. In our study, when the robot expressed head nodding behaviour with no gesture, the conflicting messages between the robot's role and its expressed nonverbal behaviour was occurred, which may result in confusion and lead to the least likability from humans.

With these supporting studies, we can infer that the Uncanny Valley in the robot's nonverbal behaviour is context-dependent. In giving a talk context, the robot's head nodding behaviour with no gesture might present context-behaviour and perceptual mismatch, which violates humans' expectation and lead to the double bind situation during the interaction. Consequently, it is prone to fall into the Uncanny Valley. Here, it suggests that the Uncanny Valley in the robot's nonverbal behaviour is context-dependent.

According to the previous studies, together with the results in this study, it can be inferred that the uncanny stimuli are produced by the activation of the expectation violation. These influence factors may cause eerie feelings or negative attitudes toward robots and lead to the Uncanny Valley phenomena.

4.5 Limitations and Future Works

In this study, there are some plausible limitations that still remain to be further investigated in future works in order to gain an insight understanding on the influence factors and causalities of the Uncanny Valley from the viewpoint of the robot's nonverbal behaviour.

Firstly, the robot used in this study (NAO robot) is still not human-like enough to be able to express a full range of nonverbal behaviours similar to human's nonverbal behaviours. Thus, the results in this study still do not cover all nonverbal behaviour range yet. Though only a partial range of robot's behaviours was investigated in this study, this study does provide significant results and can be considered as a first step of proving the existence of the Uncanny Valley from the viewpoint of the robot's nonverbal behaviour.

Secondly, the human-likeness level of the robot's nonverbal behaviours could not be directly defined beforehand unlike what had been done in the previous studies on robot's appearance where they could pinpoint the human-likeness of the robot's appearance and observed the Uncanny Valley tendency directly. In this study, we then used the number of nonverbal behaviour combination levels as the core to investigate the tendency and later used the analysis results of human-likeness from the evaluation to further investigate the tendency of the Uncanny Valley instead. Therefore, the Uncanny Valley tendency presented in this study is limited to the context and conditions used in this study since the human-likeness and affinity ratings are both dependent variables.

Thirdly, in order to thoroughly explore the Uncanny Valley from the viewpoint of the robot's nonverbal behaviour, further investigations with more varieties of the robot's nonverbal behaviours and the robot's appearances ranging from non-human-like to very human-like should be conducted. With more studies varied in robot appearance and complete range of robot's nonverbal behaviours, we can overlay the Uncanny Valley tendencies obtained from these studies with our study and achieve the overall Uncanny Valley effect from the viewpoint of the robot's nonverbal behaviour in order to complete the exploration of the Uncanny Valley from the viewpoint of the robot's nonverbal behaviour.

Fourthly, the investigations of the Uncanny Valley mechanism from the viewpoint of the robot's nonverbal behavior should be further conducted under the experimental setup based on the previous studies, for instance, the effect of conflicting cues [28] and perceptual mismatch [26, 49], in order to examine and affirm the mechanism and causality of the Uncanny Valley from the viewpoint of the robot's nonverbal behaviour.

Fifthly, possible effects of the robot's verbal behaviour to the Uncanny Valley effect have not been fully examined in this study. This is because the paper concerns the robot giving a TED talk, and the basic starting point has thus been to study the effect of non-verbal behaviour including the accompanying speech. The condition of the robot performing nonverbal behaviours without speaking is considered out of scope of the current goal of the experiments.

Finally, further investigations of the Uncanny Valley based on individual differences should be specifically conducted from the viewpoint of the robot's nonverbal behaviour since the individual differences are also widely considered as the influence factors that can lead to the Uncanny Valley phenomenon as well. For the culture aspect, Shibata et al.'s study analyzed the subjective evaluation on a seal robot, Paro, in Japan, UK, Sweden, and Italy. They found that culture influences human's attitudes and expectations toward the robot [51]. Bartneck et al.'s study also supported that culture has a significant influence on attitudes toward robots (observed in Chinese, Dutch, and Japanese participants) [52]. Furthermore, previous studies revealed that the higher level of exposure to robots in daily life and robot-related experiences can lead to less negative attitudes toward robots, for instance, the Japanese could be more aware of robot's abilities and their shortcomings more than other countries [53, 54]. In case of the robot's nonverbal behaviours, it can be implied that with a lower level of exposure or unfamiliarity with certain robot's nonverbal behaviours in daily life, it can also possibly result in lower affinity toward robots, which is supported by the studies in human–human interaction, for example, too long eye contact and pointing finger at others [55, 56]. Additionally, age or lifespan is also considered as one of the influence factors as well. The previous studies suggested that young people have more positive attitudes toward robots since they have more exposure to robots via media compared to elder people [54, 57]. Such studies will accomplish the generalization of the results with respect to the Uncanny Valley hypothesis.

5 Conclusion

In conclusion, we found a biphasic relationship between human-likeness and affinity ratings on the robot's nonverbal behaviours and the point where the affinity rating significantly dropped down is found. This study provides evidence suggesting the existence of the Uncanny Valley from the viewpoint of the robot's nonverbal behaviour. The findings suggest that the Uncanny Valley from the viewpoint of the robot's nonverbal behaviour is context-dependent. In giving a talk context, the robot's head nodding behaviour is prone to cause and fall into the Uncanny Valley while the robot's gestures should be considered as a fundamental element of the robot's nonverbal behaviour in order to enhance more likability from humans. This study explores the Uncanny Valley from the viewpoint of the robot's nonverbal behaviour in order to gain a better understanding of the Uncanny Valley as a step closer to achieve the highest likability from humans in human–robot interaction.

Authors' Contribution CT designed the experiment, collected and analyzed data, and wrote the paper. KO provided statistical analysis approach and conceptual advice regarding the experiments and results. YM supervised the study and experimental design. All authors contributed to writing the paper.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical Approval This study was carried out in accordance with the recommendations of Tokyo Institute of Technology's Ethical Review Board for Epidemiological Studies. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Tokyo Institute of Technology's Ethical Review Board for Epidemiological Studies.

References

1. Mori M (1970) The uncanny valley. Energy 7(4):33–35
2. MacDorman KF (2006) Subjective ratings of robot video clips for human like- ness, familiarity, and eeriness: an exploration of the uncanny valley. In ICCS/CogSci 2006 long symposium: toward social mechanisms of android science, pp 26–29
3. MacDorman KF, Ishiguro H (2006) The uncanny advantage of using androids in cognitive and social science research. Interact Stud 7(3):297–337
4. MacDorman KF, Entezari SO (2015) Individual differences predict sensitivity to the uncanny valley. Interact Stud 16(2):141–172
5. Mathur MB, Reichling DB (2016) Navigating a social world with robot partners: a quantitative cartography of the uncanny valley. Cognition 146:22–32
6. Seyama J, Nagayama RS (2007) The uncanny valley: effect of realism on the impression of artificial human faces. Presence Teleoperators Virtual Environ 16(4):337–351
7. Bartneck C, Kanda T, Ishiguro H, Hagita N (2007) Is the uncanny valley an uncanny cliff? In: The 16th IEEE international symposium on robot and human interactive communication, 2007. RO-MAN 2007. IEEE, pp 368–373
8. Bartneck C, Kanda T, Ishiguro H, Hagita N (2009) My robotic doppelgänger-a critical look at the uncanny valley. In: The 18th IEEE international symposium on robot and human interactive communication, 2009. RO-MAN 2009. IEEE, pp 269–276
9. Bartneck C, Croft E, Kulic D (2008) Measuring the anthropomorphism, animacy, likeability, perceived intelligence and perceived safety of robots. Metrics HRI Workshop Tech Rep 471:37–44
10. Burleigh TJ, Schoenherr JR, Lacroix GL (2013) Does the uncanny valley exist? An empirical test of the relationship between eeriness and the human likeness of digitally created faces. Comput Hum Behav 29(3):759–771
11. Ho C-C, MacDorman KF (2017) Measuring the uncanny valley effect. Int J Soc Robot 9(1):129–139
12. Mitchell WJ, Szerszen KA Sr, Lu AS, Schermerhorn PW, Scheutz M, MacDorman KF (2011) A mismatch in the human realism of face and voice produces an uncanny valley. i-Perception 2(1):10–12
13. Hanson D, Olney A, Prilliman S, Mathews E, Zielke M, Hammons D, Fernandez R, Stephanou H (2005) Upending the uncanny valley. In: Proceedings of the national conference on artificial intelligence, vol 20. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press, p 1728
14. Ho C-C, MacDorman KF (2010) Revisiting the uncanny valley theory: developing and validating an alternative to the godspeed indices. Comput Hum Behav 26(6):1508–1518
15. Ambady N, Rosenthal R (1998) Nonverbal communication. In: Encyclopedia of mental health, vol 2, pp 775–782
16. Duncan S Jr (1969) Nonverbal communication. Psychol Bull 72(2):118
17. Knapp ML, Hall JA, Horgan TG (2013) Nonverbal communication in human interaction. Cengage Learning, Boston
18. Argyle M, Cook M (1976) Gaze and Mutual Gaze. Cambridge University Press, Cambridge.
19. Thepsoonthorn C, Ogawa KI, Miyake Y (2017) Does robot's human-based gaze and head nodding behavior really win over non-human-based behavior in human–robot interaction?. In: Proceedings of the companion of the 2017 ACM/IEEE international conference on human–robot interaction. ACM, pp 301–302
20. Tickle-Degnen L, Rosenthal R (1990) The nature of rapport and its non-verbal correlates. Psychol Inq 1(4):285–293
21. McNeill D (1992) Hand and mind: what gestures reveal about thought. University of Chicago press, Chicago

22. Imada AS, Hakel MD (1977) Influence of nonverbal communication and rater proximity on impressions and decisions in simulated employment interviews. J Appl Psychol 62(3):295

23. Edinger JA, Patterson ML (1983) Nonverbal involvement and social control. Psychol Bull 93(1):30

24. Thepsoonthorn C, Yokozuka T, Miura S, Ogawa K, Miyake Y (2016) Prior knowledge facilitates mutual gaze convergence and head nodding synchrony in face-to-face communication. Sci Rep 6:38261

25. Rosenthal-von der Pütten AM, Krämer NC (2014) How design characteristics of robots determine evaluation and uncanny valley related responses. Comput Hum Behav 36:422–439

26. Saygin AP, Chaminade T, Ishiguro H, Driver J, Frith C (2011) The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. Soc Cogn Affect Neurosci 7(4):413–422

27. Kätsyri J, Förger K, Mäkäräinen M, Takala T (2015) A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. Front Psychol 6:390

28. Moore RK (2012) A Bayesian explanation of the 'Uncanny Valley' effect and related psychological phenomena. Sci Rep 2:864

29. Wiemann JM, Knapp ML (1975) Turn-taking in conversations. J Commun 25(2):75–92

30. Kleinke CL (1986) Gaze and eye contact: a research review. Psychol Bull 100(1):78

31. Kellerman J, Lewis J, Laird JD (1989) Looking and loving: the effects of mutual gaze on feelings of romantic love. J Res Personal 23(2):145–161

32. Andrist S, Mutlu B, Tapus A (2015) Look like me: matching robot per- sonality via gaze to increase motivation. In: Proceedings of the 33rd annual ACM conference on human factors in computing systems. ACM, pp 3603–3612

33. Darwin C, Prodger P (1998) The expression of the emotions in man and animals. Oxford University Press, Oxford

34. Kwon J, Ogawa K-I, Ono E, Miyake Y (2015) Detection of nonverbal synchronization through phase dierence in human communication. PLoS ONE 10(7):e0133881

35. Goldin-Meadow S, McNeill D, Singleton J et al (1996) Silence is liberating: removing the handcuffs on grammatical expression in the manual modality. Psychological Rev 103(1):34–54

36. Kendon A (1980) Gesticulation and speech: two aspects of the process of utterance. In: The relationship of verbal and nonverbal communication, vol 25, pp 207–227

37. McNeill D (1985) So you think gestures are nonverbal? Psychol Rev 92(3):350

38. Franz S, Nolte-Holube R, Wallho F (2013) Nafome: Nao follows me- tracking, reproduction and simulation of human motion. Jade University of Applied Sciences, Wilhelmshaven

39. Rouneau A (2015) pykinectnao, github repository

40. Cheetham M, Pavlovic I, Jordan N, Suter P, Jancke L (2013) Category processing and the human likeness dimension of the uncanny valley hypothesis: eye-tracking data. Front Psychol 4:108

41. Schwind V, Jäger S (2016) The uncanny valley and the importance of eye contact. i-com 15(1):93–104

42. Shimada M, Minato T, Itakura S, Ishiguro H (2007) Uncanny valley of androids and its lateral inhibition hypothesis. In: The 16th IEEE international symposium on robot and human interactive communication, 2007. RO-MAN 2007. IEEE, pp 374–379

43. Silvera-Tawil D, Garbutt M (2015) 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). IEEE, pp 740–745

44. Findlay JM, Gilchrist ID (2003) Active vision: the psychology of looking and seeing. Number 37. Oxford University Press, Oxford

45. Marschner L, Pannasch S, Schulz J, Graupner S-T (2015) Social communication with virtual agents: the effects of body and gaze direction on attention and emotional responding in human observers. Int J Psychophysiol 97(2):85–92

46. Rayner K (1995) Eye movements and cognitive processes in reading, visual search, and scene perception. Stud Vis Inf Process 6:3–22

47. Goldin-Meadow S, McNeill D (1999) The role of gesture and mimetic representation in making language the province of speech

48. Prakash A, Rogers WA (2015) Why some humanoid faces are perceived more positively than others: effects of human-likeness and task. Int J Soc Robot 7(2):309–331

49. Paepcke S, Takayama L (2010) Judging a bot by its cover: an experiment on expectation setting for personal robots. In: 2010 5th ACM/IEEE international conference on human–robot interaction (HRI). IEEE, pp 45–52

50. Bateson G, Jackson DD, Haley J, Weakland J (1956) Toward a theory of schizophrenia. Syst Res Behav Sci 1(4):251–264

51. Shibata T, Wada K, Tanie K (2004) Tabulation and analysis of questionnaire results of subjective evaluation of seal robot in Japan, UK, Sweden and Italy. In: Proceedings. ICRA'04. 2004 IEEE international conference on robotics and automation, 2004, vol 2. IEEE, pp 1387–1392

52. Bartneck C, Nomura T, Kanda T, Suzuki T, Kennsuke K (2005) A cross-cultural study on attitudes towards robots. In: HCI international conference

53. Bartneck C, Suzuki T, Kanda T, Nomura T (2007) The influence of people's culture and prior experiences with Aibo on their attitude towards robots. AI Soc 21(1–2):217–230

54. MacDorman KF, Vasudevan SK, Ho C-C (2009) Does Japan really have robot mania? Comparing attitudes by implicit and explicit measures. AI Soc 23(4):485–510

55. DeVito JA (2011) Human communication: the basic course. Pearson Higher Ed, New York

56. Axtell RE, Axtell RE (1993) Do's and taboos around the world. Wiley, New York

57. Turkle S, Taggart W, Kidd CD, Dasté O (2006) Relational artifacts with children and elders: the complexities of cybercompanionship. Connect Sci 18(4):347–361

58. Jokinen K (2010) Non-verbal signals for turn-taking and feedback. In: LREC

59. Novick DG, Hansen B, Ward K (1996) Coordinating turn-taking with gaze. In: Proceedings, fourth international conference on spoken language, 1996. ICSLP 96, vol 3. IEEE, pp 1888–1891

Chidchanok Thepsoonthorn received the BS.c. Degree in Computer Science (Informatics) with 1st class honour from Sirindhorn International Institute of Technology (SIIT), Thammasat University, Thailand in 2011, the MS.c. Degree in Artificial Intelligence (Intelligent Robotics) from University of Edinburgh, UK in 2012, and the Ph.D. Degree in Computational Intelligence and Systems Science (Human-robot Interaction) from Tokyo Institute of Technology, Japan in 2018. She received a scholarship for good academic performance and good conduct student from Sirindhorn International Institute of Technology (SIIT) and Keidanren Scholarship Fund for her bachelor degree and full scholarship from Thai government for her master and doctoral degree. Her research interests include social robotics, human-robot interaction, robot behaviours, Uncanny Valley, and artificial intelligence.

Ken-ichiro Ogawa is an associate professor in the department of Logistics and Information at the Ryutsu Keizai University through an assistant professor in the department of computer Science. He received Ph.D. degree in Theoretical Physics from Tokyo Institute of Technology in 2011. His currently research interests include the modelling of biological communication based on the active inference such as the free energy principle, some effects of a preferential attachment rule covering a long-range area on the temporal development of social networks.

Yoshihiro Miyake received the Ph.D degree from the University of Tokyo, Japan, in 1989. He currently works as a professor at the Department of Computer Science, Tokyo Institute of Technology. He has been studied the mechanism of human communication from the viewpoint of co-creation system. His current research interests include the visualization of implicit communication based on nonlinear dynamics, the communication support based on human-computer interaction and gait assist robot based on rhythmic interaction.