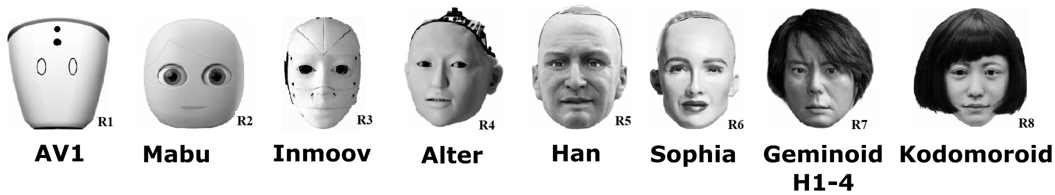
The Valley of non-Distraction: Effect of Robot's Human-likeness on Perception Load

Daisy Ingle
UNSW Sydney
Sydney, Australia
d.ingle@student.unsw.edu.au

Nadine Marcus
UNSW Sydney
Sydney, Australia
nadinem@unsw.edu.au

Wafa Johal*
UNSW Sydney
Sydney, Australia
wafa.johal@unsw.edu.au



	AV1	Mabu	Inmoov	Alter	Han	Sophia	Geminoid H1-4	Kodomoroid
Human-Likeness	14.04	37.56	59	61.3	77.04	78.88	92.6	93.44
Surface Look	0.04	0.139	0.236	0.518	0.661	0.742	1.000	0.98

Figure 1: Selection of robot subjects used as distractor ordered based on their human-likeness score in the ABOT Database [8]

ABSTRACT

Previous research in psychology has found that human faces have the capability of being more distracting under high perceptual load conditions compared to non-face objects. This project aims to assess the distracting potential of robot faces based on their human-likeliness. As a first step, this paper reports on our initial findings based on an online study. We used a letter search task where participants had to search for a target letter within a circle of 6 letters, whilst an irrelevant distractor image was also present. The results of our experiment replicated previous results with human faces and non-face objects. Additionally, in the tasks where the irrelevant distractors are images of robot faces, the human-likeness of the robot influenced the response time (RT). Interestingly, the robot Alter produced results significantly different than all other distractor robots. The outcome of this is a distraction model related to human-likeness of robots. Our results show the impact of anthropomorphism on distracting potential and thus should be taken into account when designing robots.

CCS CONCEPTS

• **Human-centered computing → User centered design**; *Human computer interaction (HCI).*

KEYWORDS

anthropomorphism; robot design; human-likeness; perception load

*Corresponding Author

HRI '21 Companion, March 8–11, 2021, Boulder, CO, USA

ACM ISBN 978-1-4503-8290-8/21/03…$15.00
https://doi.org/10.1145/3434074.3447137

ACM Reference Format:
Daisy Ingle, Nadine Marcus, and Wafa Johal. 2021. The Valley of non-Distraction: Effect of Robot's Human-likeness on Perception Load. In *Companion of the 2021 ACM/IEEE International Conference onHuman-Robot Interaction (HRI '21 Companion), March 8–11, 2021, Boulder,CO, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3434074.3447137

1 INTRODUCTION

One idea behind the design of anthropomorphic robots is that their human like appearance can enhance natural interaction and communication; making them more socially acceptable. However this assumption has been questioned in the past, as on one hand, people could respond socially with some robots and find unnatural others that "overuse anthropomorphic" features [1]. While human perception and the likeness of anthropomorphic features is an important component to take into account in designing social robots, the effect of anthropomorphism on other dimensions of the interaction has been poorly explored in comparison.

Considering a task in which the robot collaborates or competes against a human, we are particularly interested in how robots could be a distractor [4]. Recently, Urgen et al. [9] investigated if two robot appearances (mechanical or android) could distract as much as a human in a similar search task. They concluded that there was an interaction effect between human-likeness and cognitive load (task difficulty) when measuring response time. However, this study presented several limitations: 1) the robots used (Repliee Q2 with or without skin) are both quite high in terms of human-likeness; 2) images used contained the upper body (which is not the case for the human grounding study by Lavie et al. [2]) and 3) the number of participants is only 22. Lavie et al. [2] has shown that under high perceptual load, human faces as irrelevant distractors interfere with task performance more than non-face objects. This work investigates the area between objects and human faces: anthropomorphic robots that lie on a human-likeness scale. At a certain degree of human-likeness, the likeability of robots has been shown to drop dramatically, called the Uncanny Valley [6, 7]. The interference of

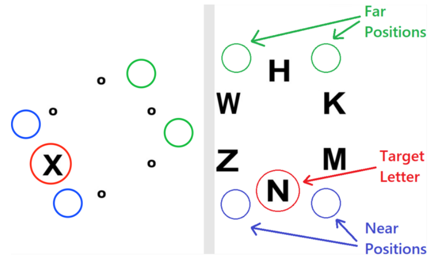


Figure 2: Examples of low (on the left) and high (on the right) perceptual load stimuli showing the different position of the distractors (in green or blue) relatively to the target (in red)

robot distractors on the boundary between non-human and human on task performance has not been studied and this work investigates whether an Uncanny Valley effect is observed when looking at robots as irrelevant distractors.

The primary aim of this work is to identify whether a relationship exists between perception of distractor robot faces, perceptual load tasks and the human-likeness of the distractor robot faces. In order to do so, we use a similar protocol (see section 2.1) to Urgen et al., but taking into account the limitation of their study we extend it by: 1) considering the whole span of the human-likeness spectrum; 2) considering only robot's faces to accurately reproduce the related study by Lavie et al. [2] in psychology; and 3) increasing the number of participants and number of trials per participants.

- **H1** The response times for high perceptual load tasks will depend on the human-likeness of the robot face distractor.
- **H2** The "Uncanny Valley" curve will be present within the response time results for high perceptual load tasks.

2 METHODOLOGY

2.1 Stimuli Design

The design of the task aims to replicate and extend the work by Lavie et al. [2, 5]. The stimuli presented to participants can be divided into two components: the primary stimuli – circle of letters and the secondary stimuli – the distractor image.

2.1.1 Circle of Letters. The primary stimulus display consisted of six letters arranged in a circle with one target letter in each arrangement. The target letter chosen for each arrangement was either an **X** or an **N**. The non-target letters were chosen so that when the participant searches for the target letter, there are both low perceptual load and high perceptual load conditions for the search. In the low perceptual load condition, the non-target letters were all the same – **o** letter. The non-target letters were designed as 1/3 the height and width of the size of the target letter for each arrangement (Figure 2). In the high perceptual load condition, **H**, **K**, **M**, **W**, and **Z** were used as the non-target letters. All six letters in the high perceptual load conditions were the same size as the target letters (Figure 2). The selection of letters, size, and spacing of letters in this experiment was designed to replicate previous experiments by Lavie et al. [2, 5]. In designing each stimulus display, we took care to ensure randomness of each configuration of letters (variation of target location over all 6 locations within the circle).

2.1.2 Distractor Images. Two categories of images were used as distractors for this experiment: faces of anthropomorphic robots, and non-robot images (control) consisting of a non-face object and human faces. For consistency, each image was black and white, the same resolution, and with no background.

Robot Images. The robot subjects chosen for this study were from the ABOT database [8] (containing a categorisation of 251 robots), where each robot listed in the database has determined human-likeness score between 0 and 100 which is calculated from four other scores between 0-1 based on robot appearance (surface look, facial features, body-manipulators, mechanical locomotion). The two most significant scores used in this study are the human-likeness score which is an overall rating of the human-likeness of the robots, and the surface look score which focuses on facial details (e.g., gender, skin, hair, eyelashes). As this study features only the faces of the robot subjects, the surface look score was considered as an alternative human-likeness score for the robot subjects. In order to evaluate the effect of different levels of anthropomorphism, the robot subjects were chosen to span the entire human-likeness scale. We sub-sampled the robots over the human-likeness scale and selected 5 robots to cover the spectrum. For a robot subject to be suitable, we decided that robots with screens, animal-like features, alien-like features, gendered features, non-neutral facial expressions and hats were to be avoided. Our aim was to remove as many different variables as possible between each subject, to ensure that results found from the study are valid.

Interestingly, for robots with high human-likeness scores (>70), attributes such as gender and ethnicity become unavoidable and had to be taken into account. To control these possibly significant variables, measures were taken in the selection of robots with high human likeness scores, in the questionnaire, and in the selection of non-robot images. When selecting robot subjects with high human-likeness scores, two robots (one designed as female, one designed as male) were chosen such that apart from their perceived gender, they were equivalent (in score). As detailed in Figure 1, robots *Han* and *Sophia* were selected as equivalent with opposite genders as their human-likeness scores are within 2% of each other and they are both developed by Hanson Robotics with similarities (lifelike facial features, no hair, Caucasian features). Similarly, robots *Geminoid H1-4* and *Kodomoroid* were selected as equivalent with human-likeness scores within 1% of each other and both developed by Hiroshi Ishiguro Laboratory with similarities (lifelike facial features, hair colour, similar hair lengths).

The initial decision of having one robot subject per category was extended to two robot subjects for the higher two described categories. However, as detailed in Figure 1, three robots were chosen within the human-likeness category (61-80). One of the initial choices for this category taking into account the perceived genders of the robots was to use images of robots *Han* and *Sophia*. Upon viewing the entire selection of robot subjects, there seemed quite a gap visually between robot *Inmoov* and robots *Han* and *Sophia* as the former robot has a hard plastic face with visible hardware whilst the latter robots have a face with very life-like skin and details. This gap was evident when looking at the corresponding surface look scores (0.236, 0.661). For the final selection of robot subjects, another robot subject *Alter* was added with a similar human-likeness score to *Inmoov* but with a face with more life-like skin and a higher surface look score. After the selection of robot images was completed, a bit of image editing was performed to ensure a similar aspect (i.e. no extra objects were in view, resolution was clear enough, edit background to white, resize).

Non-Robot Images. To be able to replicate key results from Lavie et al.[2], images of both human faces and an non-face object, a violin, were also selected as distractors for this experiment. The female human faces were sourced from stock photography and made into a male version using a gender morphing image tool. The violin image was sourced from stock images and chosen as images of musical instruments have been previously used in similar experiments [2].

2.2 Reaction Time Task

For each circle of letters search task, the location of the distractor in relation to the letter was designed to replicate Lavie et al. experiments' [2, 5]. We used two positions for each distractor and perceptual load. Figure 2 explains the two types of positions, namely near and far positions, in relation to the target letter for each search task. Categorising and limiting the distractor positions in this way allowed for the distractor position to seem randomised when completing the task. Between each search task, a fixation screen consisting of a white background and a black **X** in the centre of the screen was displayed for 1 second. The use of a fixation screen is common in search tasks [5] to realign participant's gaze to the centre of the display in preparation for the next search task. Each participant was presented a series of 104 trials. To ensure that there would be no predictable trend in the order that the search tasks were displayed, the order of each task was randomised each time the experiment was run. Similarly to Lavie et al. [2, 5], participants used keyboard keys to give their answers. Two measures were extracted: the number of errors and the response time in milliseconds.

2.3 Participants

98 participants (39 females, 57 males, 2 other, Mean Age = 30.89, SD = 14.97) were recruited online through the university email service and social media posts using Facebook, Twitter, Reddit, and Yammer. No participants were reimbursed for participating.

3 RESULTS

The analysis reported here focuses on the RT for correct responses and RTs within 3 standard deviations from the mean (i.e. removing outliers). A total of 9689 trials was then included in the analysis.

Analysing the RT results for under high load conditions, the RT with the non-face distractor image (musical instrument) (M = 1181ms, SD = 520) was significantly faster than for the human face distractor image (M = 1377ms, SD = 612) (T = -3.54, $p < 0.001$). Additionally, there was a significant interaction between distractor type and perceptual load (T = -4.80, $p < 0.001$). These results support previous findings [2] that interference of RT performance from human face distractors is not affected by perceptual load, whilst interference from non-face distractors is affected by perceptual load.

H1: Effect of the Distractor's Human-Likeness on Response Time

Performing a least squares regression analysis on the data found that there is no significant relationship between the RT results and the human-likeness score of the distractor under low perceptual load (T = -0.03, p = 0.98, N = 2969). Under low perceptual load, no significant relationship was found between the human-likeness score of the robot distractors and the RTs for the search task. These results support previous findings [3] that under low perceptual load, the interference of RT performance from irrelevant distractors is unaffected by the distractor type.

Performing a least squares regression analysis on the data found that there is a significant relationship between the RT results and the human-likeness score of the distractor under high perceptual load (T = 3.03, $p < 0.05$, N = 2987). To further analyse this relationship between the RTs and human-likeness, a statistical paired t-test was performed for each robot distractor. Table 1 displays the related t-test results comparing the response time data for each robot to the next robot on the human-likeness scale under high perceptual load conditions. The results show that there was no significant difference in response times between robots R1 (AV1) and R2 (Mabu), and between robots R5 (Han) and R6 (Sofia). However, the response time for robot R4 (Alter) was significantly faster than R3 (Inmoov) and R5 (Han). We also find significant difference between R2 (Mabu) and R3 (Inmoov), the latter giving longer RTs. R7 (Geminoid) was also found to give longer response time compare to R6 (Sophia) and R8 (Kodomoroid).

Robot 1	RT (ms)	Robot 2	RT (ms)	N	T-test
R1 (AV1)	M = 1230 SD = 494	**R2** (Mabu)	M = 1219 SD = 456	718	T = 0.36 p = 0.72
R2 (Mabu)	M = 1212 SD = 448	**R3** (Inmoov)	M = 1321 SD = 600	730	T = -2.92 $p < 0.05$
R3 (Inmoov)	M = 1303 SD = 592	**R4** (Alter)	M = 1063 SD = 421	720	T = 7.58 $p < 0.001$
R4 (Alter)	M = 1076 SD = 427	**R5** (Han)	M = 1313 SD = 647	722	T = -3.56 $p < 0.001$
R5 (Han)	M = 1299 SD = 596	**R6** (Sophia)	M = 1280 SD = 617	736	T = 0.45 p = 0.65
R6 (Sophia)	M = 1282 SD = 612	**R7** (Geminoid H1-4)	M = 1366 SD = 602	706	T = -2.16 $p < 0.05$
R7 (Geminoid H1-4)	M = 1349 SD = 584	**R8** (Kodomoroid	M = 1265 SD = 502	684	T = 2.13 $p < 0.05$

Table 1: Statistical results of response times under high perceptual load conditions.

These results support the hypothesis that the response times for high perceptual load tasks depend on the human-likeness of the robot face distractor. However, we notice that it doesn't seem that the effect is regulated by a linear relationship, which is unlike what Urgen et al.'s study seemed to have found [9].

H2: RT & Human-likeness reflects the Uncanny Valley Curve

To test whether the response time data shows similarity to the "Uncanny Valley" curve [7], a polynomial function was fit to the response times under high load conditions against the human-likeness score of the distractors. The work by Mathur et al. [6]

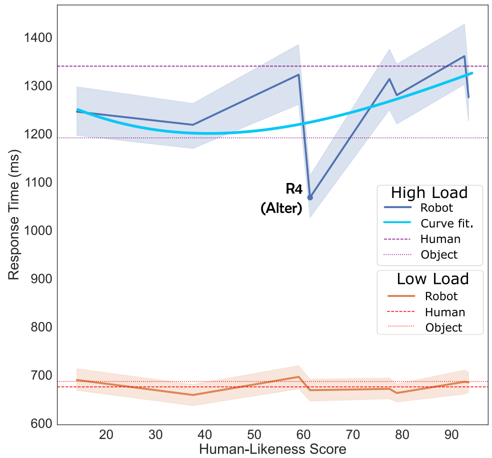


Figure 3: RT according to *human-likeness* score under high load conditions, with possible curve fit

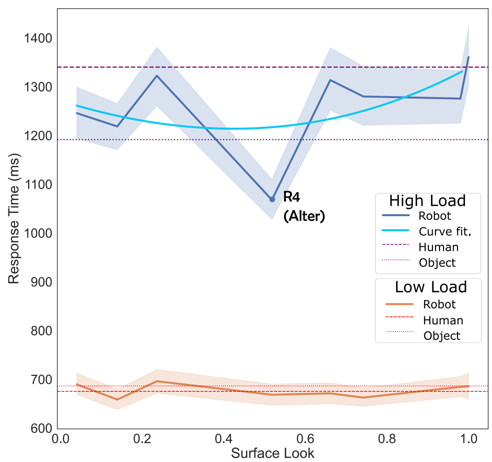


Figure 4: RT according to *surface* score under high load conditions, with possible curve fit

found that a third degree polynomial was the best fit polynomial to model the "Uncanny Valley" curve. Therefore, we fit a third degree polynomial to all response time results versus the human likeness score of each distractor under high load condition. Figure 3 shows the resulting polynomial against the mean response times under high load conditions for each distractor type together with the the mean response times for the non-face object and human face distractors for comparison. The following polynomial shows the calculated coefficients for the curve in Figure 3 to 3 significant figures:

$$f(x) = 1330 - 6.73x + 0.104x^2 - 0.000341x^3$$

For each coefficient value, a standard deviation value was also calculated. In the same order as the coefficients, the following are the standard deviation results for each calculated polynomial coefficient to 3 significant figures:

$$e_0 = 89.0,\ e_1 = 6.85,\ e_2 = 0.141,\ e_3 = 0.000844$$

As detailed in Figure 1, each robot distractor also has a surface look score and with the exception of robots Geminoid and Kodomoroid, the surface look scores rank the robots in the same order as the human-likeness score. Considering the surface look scores as an alternative measure of human-likeness of robot faces, we performed a similar polynomial fit to the reaction time results against the surface score of each robot. Figure 4 shows the resulting polynomial plotted with the mean response time results under high load conditions for each distractor together with the the mean response times for the non-face object and human face distractors for comparison. The following polynomial shows the calculated coefficients for the curve to 3 significant figures:

$$f(x) = 1270 - 257x + 264x^2 - 53.0x^3$$

As above, for each coefficient value, there is a corresponding standard deviation value. In the same order as the coefficients, the following are the standard deviation results for each calculated polynomial coefficient to 3 significant figures:

$$e_0 = 37.8,\ e_1 = 370,\ e_2 = 821,\ e_3 = 498$$

These analyses result in polynomial fitted curves with a similar shape to the trough present in the 'Uncanny Valley' curve as modelled in [6]. However, their curve model is not available for us to quantitatively compare our results.

4 DISCUSSION AND CONCLUSION

This work has successfully reproduced previous research on the psychological theory of perceptual load and built upon this research by replicating the same experiments with new stimuli focused on anthropomorphic or human-like robots. When using images of robot faces in high perceptual load search tasks, the results confirmed that the degree of human-likeness or anthropomorphism has an effect on the performance of the search task. These results imply that the degree of human-likeness of robots may impact the distracting potential of the robot when present in high perceptual load conditions (difficult task). This work has also found that the robot R4 (Alter) had a different distracting effect than a non-face object and other robot faces, which could imply that this robot is not recognised as a familiar object (aka not triggering any automatized recognition cognitive stream). Finally, this work has attempted to model the relationship by curve fitting the response time results in high perceptual load tasks, giving opportunity for future research to continue this model.

In the future, we would like to increase the set of robot subjects to obtain more granularity in our model, especially around R4 (Alter). Another potential area to extend this work would be to look at full body robots (instead of just faces) or to look at robots in motion. Finally, we are aware that this work presents the limitation of not being in person and potentially affected by the novelty effect. It could be interesting to integrate a similar search task protocol in a more realistic scenario featuring physically present robots.

The results from this work form a preliminary scale to assess the effect of anthropomorphic features on perception load and their potential distracting effect. We believe that our model can be used to inform choices by HRI designers and experimenters in the way they design and use robots for collaboration.

REFERENCES

[1] Julia Fink. [n.d.]. Anthropomorphism and Human Likeness in the Design of Robots and Human-Robot Interaction. In *Social Robotics*, Shuzhi Sam Ge, Oussama Khatib,

John-John Cabibihan, Reid Simmons, and Mary-Anne Williams (Eds.). Vol. 7621. Springer Berlin Heidelberg, 199–208.

[2] Nilli Lavie, Tony Ro, and Charlotte Russell. 2003. The role of perceptual load in processing distractor faces. *Psychological science* 14, 5 (2003), 510–515.

[3] Nilli Lavie and Yehoshua Tsal. 1994. Perceptual load as a major determinant of the locus of selection in visual attention. *Perception & psychophysics* 56, 2 (1994), 183–197.

[4] Alvin X. Li, Maria Florendo, Luke E. Miller, Hiroshi Ishiguro, and Ayse P. Saygin. 2015. Robot Form and Motion Influences Social Attention. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (Portland, Oregon, USA) *(HRI '15)*. Association for Computing Machinery, New York, NY, USA, 43–50. https://doi.org/10.1145/2696454.2696478

[5] James SP Macdonald and Nilli Lavie. 2008. Load induced blindness. *Journal of Experimental Psychology: Human Perception and Performance* 34, 5 (2008), 1078.

[6] Maya B. Mathur and David B. Reichling. 2016. Navigating a social world with robot partners: A quantitative cartography of the Uncanny Valley. (2016), 22–32.

[7] M. Mori, K. F. MacDorman, and N. Kageki. 2012. The Uncanny Valley [From the Field]. *IEEE Robotics Automation Magazine* (2012).

[8] Elizabeth Phillips, Xuan Zhao, Daniel Ullman, and Bertram F Malle. 2018. What is Human-like? Decomposing Robots' Human-like Appearance Using the Anthropomorphic roBOT (ABOT) Database. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. 105–113.

[9] Burcu A Urgen, İlayda Guneysu, Selin Yilmaz, Begüm Cerrrahoğlu, and Ece Dinçer. 2020. Do Robots Distract us as much as Humans? The Effect ofHuman-like Appearance and Perceptual Load. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. 493–495.