Interdisciplinary Practice in Industrial Design, Vol. 144, 2024, 68–78
https://doi.org/10.54941/ahfe1005119

VR-Based Evaluation of Design Elements for Restaurant Service Robots

Domagoj Bui and Yong-Gyun Ghim

University of Cincinnati, Cincinnati, OH 45219, USA

ABSTRACT

With robots being increasingly used in the service industry, the importance of human perception of robots' appearance grows bigger. Previous studies show that a robot's design affects people's perception of its characteristics and capabilities. Applying some of the design elements of lifelike robots to functional-looking robots enhances positive perception. However, questions remain about a more specific relationship between individual design elements and human perception. This study aims to explain more accurately which individual design elements positively affect the perception of functional-looking robots. To achieve this, a range of robot design variations was created with individual design elements applied from lifelike designs. Using an immersive virtual reality experience, which allows for more comprehensive and accurate evaluations, design variations were presented to 16 participants, and their perceptions were measured and recorded on-site. The results indicate that people have a more positive perception of robots that look functional but have softer shapes and visual highlights.

Keywords: Service robots, Robot morphology, VR evaluation, Industrial design

INTRODUCTION

Organizations increasingly employ service robots to assist and interact with humans (Lu et al., 2019). This means that robots are becoming encountered and used by diverse people in varying contexts. As people form conceptions about robots before the interaction begins, and robots' appearance influences users' expectations and the evaluation of their behavior and capabilities, creating the correct perception by appearance is highly important (Li et al., 2010; Bartneck et al., 2009). Lifelike robot appearances are considered to contribute to a positive perception of robots (Fink, 2012). However, they can also easily create an expectation gap (Komatsu et al., 2012) or the uncanny valley effect (Mori, 1970). Their unmatching performance to the initial positive perception has led to only a few successful implementations of service robots (Rosete et al., 2020). An alternative direction is functional morphology, which is considered more suitable for task-oriented operations (Roesler et al., 2022). To apply this approach to service robots, prior research sought to find design elements applicable to functional-looking robots that improve user perception (Bui & Ghim, 2023). Our research explores the relationship between human perception and specific design elements. We hypothesize that functional-looking robots equipped with certain lifelike design elements are perceived more favorably than functional-looking robots without adding the negative impacts of lifelike robots. We also assume that showing specific robot designs for evaluation using virtual reality (VR) environments results in more accurate assessments. Accordingly, we created computer-aided design (CAD) models of design variations for a restaurant server robot and evaluated their perception in a VR restaurant environment. Purposefully chosen design elements were individually applied to a base model of a functional-looking robot, and each variation was shown to 16 participants together with the base design. To maximize the accuracy of the assessment, the participants were asked questions about the perceived characteristics and capabilities of the presented robots during the VR virtual experience. Based on the evaluation results, we propose a design guideline to help industrial designers create service robot designs that better fit their purpose and enhance human-robot interaction (HRI).

ROBOT DESIGN ELEMENTS AND EVALUATION METHODS

Perceived Attributes of Robot Appearance and Design Elements

A robot's appearance affects how people expect and perceive its behavior and capabilities (Li et al., 2010), and people form their initial impressions of the robot based on its appearance (Phillips et al., 2018). While this first impression is crucial for people to desire to start and maintain the interaction, when robots, especially those resembling humans, do not perform up to the initial expectation, people get frustrated, and businesses stop using them (de Graaf et al., 2017; Rosete et al., 2020).

Researchers classify robots mostly into two groups based on their morphology: lifelike (anthropomorphic, zoomorphic) and functional-looking (machine-like) (Phillips et al., 2017; Bui & Ghim, 2023). Lifelike robots visually resemble the human body or animals (Bartneck et al., 2020), and functional-looking robots are designed to reflect the tasks they perform (Fong et al., 2003). The issues arise in the limitations and challenges of how humans perceive each morphology in a service context. Research suggests that lifelike robots are perceived as friendlier and are generally preferred by users. They are also rated more capable and competent (Kunold et al., 2023). However, the perception of high capability is often not met with actual performance, creating an expectation gap (Lohse, 2011; Komatsu et al., 2012). Additionally, the human likeness that makes them generally perceived as friendlier in many cases also creates discomfort because of the uncanny valley phenomenon (Mori, 1970). The disappointment from expectation gaps and unsettling feelings due to the uncanny valley have raised questions about the direction toward lifelike robot designs. Although functional-looking service robots are perceived as less friendly and capable (Kunold et al., 2023), their actual capabilities can better match the expected function in the user's perception. If designing robots in functional morphology can lead to both favorable initial perception and satisfactory service experience, it can potentially create a better design direction for service robots.

Bui and Ghim (2023) laid out the relationship between a service robot's design elements and its perceived attributes. They argued that the perception of design elements influences the attribution of characteristics and capabilities that subsequently affect robot acceptance. Their study categorized design elements under form, color, and interface. They identified elements from lifelike robot designs that contribute to positive perception, such as curved contours, symmetry, global-local relations, achromatic body, color and light highlights, and facial expressions. They suggested that these elements can improve the perception of functional-looking robots in a service environment. However, they were based largely on theories outside HRI and have not been validated with robots.

Evaluation of Robot Design

Robot design evaluations can be challenging due to the high cost, complex logistics, and safety issues. Despite these challenges, they are a key component of the design process (Berg & Vance, 2017). To overcome the challenges of using actual robots, studies often employ proxies such as images, videos, 3D CAD models, and immersive VR for participant evaluation of robot designs (Mara et al., 2021; Randall & Sabanovic, 2023). Among these studies, lower-fidelity media like photos and videos were prevalent because they are best suited for large online evaluations. However, compared to high-fidelity media, they generally hinder participants from identifying functional attributes and are less successful and consistent in evaluating robot interactions and preferences. Videos normally provide better results and help robots be perceived as more human-like than still images (Randall & Sabanovic, 2023). VR enables even more accurate evaluations (Mara et al., 2021).

There are many benefits of using VR for robot design evaluation, and it is now increasingly employed as an evaluation tool in HRI (Li et al., 2019). The ability to create immersive 3D environments and simulations reliably, flexibly, and affordably has recently led to its integration into the design process of larger organizations (Berg & Vance, 2017). One of the biggest advantages is the participant's spatial perception, which allows for a more accurate perception of the model size (Horvat et al., 2019) for better product understanding through interaction with the model (Berg & Vance, 2017) and adequate evaluation of topics in human psychology (Negi et al., 2008). At the same time, it also helps designers and organizations conduct evaluations more efficiently in terms of time and cost. VR is inexpensive for collecting large amounts of data repeatedly - for usability and aesthetic improvements and iterations, different scenario or environment simulations, or large model counts testing different design parameters (Negi et al., 2008; Roberts et al., 2020) - which is beneficial for new proposals, during the design process, and for final decision-making (Negi et al., 2008; Berg and Vance, 2016; Stadler et al., 2020). However, VR is not without its own problems. Technical limitations can exist that compromise immersion and skew the assessment, such as a limited field of view, display resolution, lack of audio input, lack of haptic feedback, and 3D CAD fidelity (Berg and Vance, 2016; Li et al., 2019).

STUDY DESIGN AND PROCEDURE

In this study, we aim to evaluate the design elements suggested in the previous study (Bui & Ghim, 2023) that contribute to better acceptance of robots:

curved contour, human-like form, color highlights, and facial features. They were chosen to offer focused and reliable insights into the application of lifelike design elements into functional-looking designs and a deeper understanding of the perception of each element (Figure 1). Five robot designs were developed as 3D CAD models for evaluation. The first one is a base model representing a functional-looking robot, with each design element purely made for the functional purpose of a restaurant server robot. Four additional robots were designed by individually introducing each of the four lifelike-derived design elements to the robot. Then, we evaluated the five robot designs to understand the influence of introduced lifelike design elements on initial human perception. We aimed to find answers to the following questions:

1. When an individual lifelike design element is introduced, does it make the functional-looking service robot more positively perceived for its characteristics and capabilities?
2. Which design elements make the biggest impact on the perception of a functional-looking service robot for its characteristics and capabilities?

This study has three main components that influence the realism, relevance, and scope of the evaluation: the design of the environment, the design of the control and test robot models, and the VR testing with simultaneous interviews.

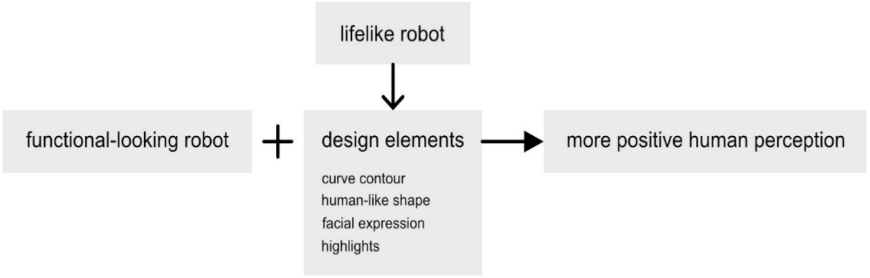

Figure 1: Introduction of design elements from lifelike robots to functional-looking robots.

Environment Design

A restaurant environment was chosen for the evaluation of the design elements. One of the restaurants in downtown Cincinnati was observed, and a virtual environment was constructed to mirror the restaurant's spatial and functional characteristics (Figure 2). It was modified to allow sufficient light for the robots to be visible and to provide ample space for them to move realistically. The environment was created in a specific shape and size to allow for longer observation time and with a detailed and realistic setting consisting of tables, bars, food and drinks, and moving people. While the environment had to be realistic enough to enable immersion, it needed to be inconspicuous to avoid detracting from the original intention of functioning as a context or overburdening the limited hardware power. The restaurant's interior, including walls, ceilings, and fixtures, and its outer environment was created for the study. Food and drinks, furniture, lights, plants and foliage, and surrounding buildings were sourced online to best fit the environment, and moving people and vehicles were imported from the software's library. Twinmotion software was utilized for creating this environment due to its capability to design and execute real-time animated models in VR and the abundance of available resources.

Figure 2: A restaurant environment for VR.

Robot Model Design

According to the context, restaurant server robots, whose tasks involve taking orders, communicating with customers, and carrying and serving food to the customers while moving around in the restaurant, were chosen for the evaluation. Each element had to be evaluated individually and equally to understand their specific design elements and perceived attributes. Firstly, a robot with a functional-looking appearance was designed in CAD by referencing designs found in conventional functional-looking robots currently on the market. To be used as a control model for the evaluation, it had 1) sharp edged contours, 2) no visual reference to a human body shape, 3) no chromatic color or color/light highlights, and 4) a text-based interface. Then, four test model robots were designed, each corresponding to one of the four chosen design elements and having one element applied to the control model individually (Figure 3). Test Robot 1 is functional-looking with curved contours. Test Robot 2 has its body changed to a human-like shape. Test Robot 3 has colored lights highlighting its functional features. Test Robot 4 has its interface changed to facial features. Each of the four test robots was modeled by applying the specific design element to the control model and changing its design accordingly while keeping the other parts the same.

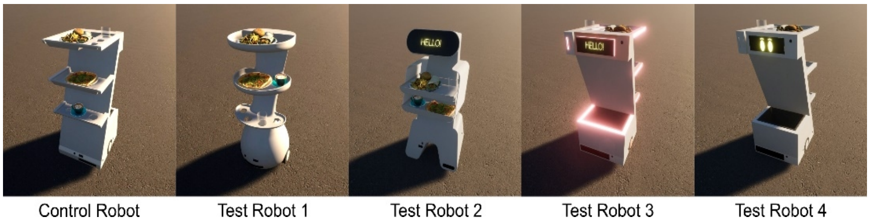

Figure 3: Control and test robot designs.

VR Testing With Simultaneous Interviews

The study was conducted in a closed room with 16 participants, aged between 21 and 31 (average age: 24 years old), comprising 11 males and 5 females. Each participant was individually called into the study and asked to provide voluntary consent to participate. After watching an introductory video displayed on a computer screen, participants were instructed to put on the VR headset and were asked questions during the virtual experience. The study lasted approximately 5 minutes per participant, and the entire session was audio-recorded for future transcription and data analysis.

The participants were situated to observe the robot approach them for 15 seconds, turn around and serve food in front of them for another 15 seconds, and then leave for 15 seconds. Additionally, the robot was placed at a distance from the participants to avoid encroaching on their personal space, which was reported to be much narrower in VR settings. Robot videos were shown to each participant in one of three randomized orders to minimize bias. The VR headset had a 4K resolution, ensuring sharp visuals without compromising immersion while meeting the hardware requirements. Also, spatial audio cues were provided through the headset speakers, and the participants were able to move and look around freely.

Interviews with the participants were conducted simultaneously with the VR experience. This was done to evaluate their perception solely from the initial impression and to eliminate any potential impression overlap or hallucination. During each robot video, participants were asked five questions and instructed to respond as quickly as possible. The first four questions were based on a 5-point Likert scale, and the final one was a short open-ended question:

Question 1. How intelligent is the robot?
Question 2. How friendly is the robot?
Question 3. How capable do you think it is to serve you in a restaurant?
Question 4. How capable do you think it is to communicate with you?
Question 5. What aspect most influenced your replies?

RESULTS

The results from the first four questions were recorded in a datasheet and analyzed based on the question, the robot model, and the total for all questions combined per robot (Figure 4).

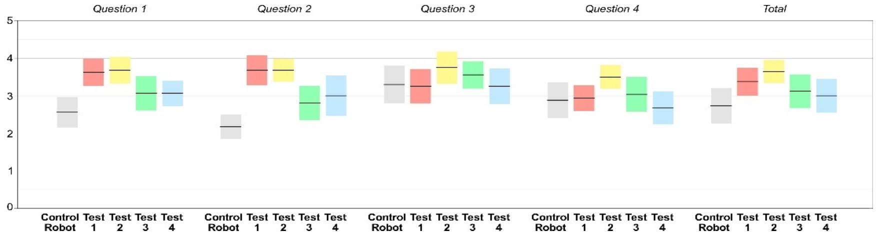

Figure 4: Evaluation results for the characteristics and capabilities of control and test robots.

The difference between the control robot and all the test robots is most noticeable in questions 1 and 2, which concern perceived characteristics of intelligence and friendliness. The curve-contoured Test Robot 1 and the human-like shaped Test Robot 2 were perceived as more intelligent and friendly than Test Robot 3 with highlights, Test Robot 4 with facial features, and the Control Robot. Notably, the Control Robot received low ratings for friendliness. The results in friendliness could be explained as an effect of lifelikeness in service robots. Lifelike robots were generally perceived as significantly friendlier and more likable in this context. Interestingly, Test Robots 1 and 2, with changes in form, were evaluated highest and very similarly in both instances. In contrast, Test Robots 3 and 4, with changes in color and interface respectively, received moderate ratings for intelligence and friendliness. Test Robot 3 particularly received the lowest ratings for friendliness among the four test robots. This suggests that changes in shape have a greater impact on positive perception of functional-looking robots compared to other alterations. Although Adding color/light highlights on functional parts slightly enhanced positive perception, it also elicited divisive opinions, as evidenced by participant comments, such as "aggressive," "rigid," and "firm." In questions 3 and 4, which focus on perceived capabilities of service and communication respectively, the curve-contoured robot and the robot with facial features were evaluated almost the same or even lower than the Control Robot. The robot with a human-like shape was the only one evaluated higher than the the others.

When the evaluation scores were averaged across the four questions, the Control Robot received the lowest average rating at 2.734. Participants explained their ratings similarly: simple, angular, utilitarian, firm, and formal. Test Robot 2, with the humanlike design, received the highest average rating at 3.656. Participants offered divided explanations: advanced, sentient, welcoming, odd, having moving arms and a face (although it had a text-based interface and no moving arms). Test Robot 1, with the curved contour, followed with an average of 3.375. Participants described it using positive terms: organic, higher quality, sophisticated, modern, smart, cute, and smaller (even though sizes were the same). Test Robot 3, highlighted with colored lights, was next with an average of 3.125. Participants expressed varied opinions including fancy, aggressive, informative, rigid, cheap, and attractive. Test Robot 4 with facial features was evaluated second to lowest, with an average of 3. Participants explained their ratings as sharp, characteristic, limited, approachable, and utilitarian.

A single-factor ANOVA analysis was conducted for each question to determine statistically significant differences in perception between each test robot and the Control Robot (Table 1). Not all differences discussed above are statistically significant. It was determined that regarding intelligence, only the curve-contoured Test Robot 1 ($F[15.318] = 4.171$, $p = 0.0005$) and the human-like shaped Test Robot 2 ($F[17.482] = 4.171$, $p = 0.0002$) were perceived more positively than the Control Robot. With regard to friendliness, all test robots were perceived significantly more positively than the Control Robot. Still, in the capability of restaurant service, none of the test robots were found to differ significantly in perception from the functional-looking design. Finally, with the communication capability, only human-like shaped Test Robot 2 was perceived significantly more positively than the Control Robot (F[4.747] = 4.171, p = 0.037). When all the results were combined, three test robots were perceived more positively than the functional-looking robot, with only Test Robot 4 failing to provide a discernible difference in perception.

Table 1. ANOVA analysis results for all control-test robot pairs.

p-values results

Compared to Control Robot	*significance*	Question 1	Question 2	Question 3	Question 4	Total
Test Robot 1	0.05	0.0004836	2.23E-06	0.8318586	0.8328618	3.4E-05
Test Robot 2		0.0002316	1.783E-07	0.1974213	0.0373406	3.9E-09
Test Robot 3		0.1157947	0.0334784	0.429391	0.5781026	0.02014
Test Robot 4		0.0690632	0.0162615	0.8571428	0.5670573	0.08658

F-statistic results

Compared to Control Robot	*F crit*	Question 1	Question 2	Question 3	Question 4	*F crit*	Total
Test Robot 1	4.17088	15.318021	34.015748	0.0458716	0.0453172	3.91755	18.4922
Test Robot 2		17.482014	45.473684	1.7375887	4.7468354		40.2203
Test Robot 3		2.6229508	4.9668874	0.6417112	0.3161593		5.5419
Test Robot 4		3.5555556	6.483376	0.032967	0.3349876		2.98396

When lifelike design elements were applied to functional-looking robots, the design element that significantly enhanced positive perception was the robot's overall human-like outer shape (F[40.220] = 3.918). The second most influential design element was the curved contour (F[18.492] = 3.918). Adding color or light highlights slightly enhanced positive perception, though to a lesser extent (F[5.542] = 3.918).

CONCLUSION

Not all design elements from lifelike robot designs resulted in an increase in positive perception about robot characteristics and capabilities in initial impressions. The most significant influence on positive perception occurred with the design elements in the 'form' category. The robot with the human-like shape received the highest evaluation. However, it already shows a potential danger of causing an expectation gap because the design was perceived more positively than all the others for questions about capabilities. On the other hand, the curved contour can have the biggest positive impact on perception by being perceived as more intelligent and friendly while retaining a similar level of perceived capability to the functional-looking design. The addition of highlights needs to be carefully considered to avoid overly emphasizing functional-looking parts. The addition of lifelike features, such as facial expressions on interfaces, has to be implemented carefully not to trigger any negative perceptions related to lifelike robots. The evaluation results for the design elements of curved contour and human-like shape are significant enough to provide insights that can influence the design of service robots in the hospitality context to deal with the issues of initial engagement.

This study is not without limitations. The robot with the facial expression was rated lowest for the capability of communication, which is contradictory to other studies. The visibility, design, and behavior of the facial expressions may have caused either the failure to capture this element or negative responses, requiring further investigation. Additionally, some participants negatively mentioned the “Z-like” shape of the original robot design. This highlights the most critical limitation of the study. The results can be skewed due to specific design decisions on the models themselves, the environment, or the immersive experience. Similar to the designs of actual products, particular designs may be perceived as more or less favorable. Further research is needed to confirm the results for each specific design element by comparing the variables of their properties, such as variations in shape, color, or material.

REFERENCES

Bartneck, C. et al. (2008) ‘Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots’, International Journal of Social Robotics, 1(1), pp. 71–81. Available at: https://doi.org/10.1007/s12369-008-0001-3.

Bartneck, C. et al. (2020) Human-Robot Interaction: An Introduction. Cambridge University Press.

Berg, L. P. and Vance, J. M. (2017) ‘An Industry Case Study: Investigating Early Design Decision Making in Virtual Reality’, Journal of Computing and Information Science in Engineering, 17(1), p. 011001. Available at: https://doi.org/10.1115/1.4034267.

Fink, J. (2012) ‘Anthropomorphism and Human Likeness in the Design of Robots and Human-Robot Interaction’, in S. S. Ge et al. (eds) Social Robotics. Berlin, Heidelberg: Springer (Lecture Notes in Computer Science), pp. 199–208. Available at: https://doi.org/10.1007/978-3-642-34103-8_20.

Fong, T., Nourbakhsh, I. and Dautenhahn, K. (2003) ‘A survey of socially interactive robots’, Robotics and Autonomous Systems, 42(3), pp. 143–166. Available at: https://doi.org/10.1016/S0921-8890(02)00372-X.

de Graaf, M., Ben Allouch, S. and van Dijk, J. (2017) ‘Why Do They Refuse to Use My Robot? Reasons for Non-Use Derived from a Long-Term Home Study’, in Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction. New York, NY, USA: Association for Computing Machinery (HRI ’17), pp. 224–233. Available at: https://doi.org/10.1145/2909824.3020236.

Horvat, N. et al. (2019) ‘Comparing Virtual Reality and Desktop Interface for Reviewing 3D CAD Models’, Proceedings of the Design Society: International Conference on Engineering Design, 1(1), pp. 1923–1932. Available at: https://doi.org/10.1017/dsi.2019.198.

IFR (2022) Sales of Robots for the Service Sector Grew by 37% Worldwide, IFR International Federation of Robotics. Available at: https://ifr.org/ifr-press-releases/news/sales-of-robots-for-the-service-sector-grew-by-37-worldwide (Accessed: 15 February 2024).

Komatsu, T., Kurosawa, R. and Yamada, S. (2012) ‘How Does the Difference Between Users’ Expectations and Perceptions About a Robotic Agent Affect Their Behavior?’, International Journal of Social Robotics, 4(2), pp. 109–116. Available at: https://doi.org/10.1007/s12369-011-0122-y.

Kunold, L., Bock, N. and Rosenthal-von der Pütten, A. (2023) 'Not All Robots are Evaluated Equally: The Impact of Morphological Features on Robots’ Assessment through Capability Attributions', ACM Transactions on Human-Robot Interaction, 12(1), pp. 3:1–3:31. Available at: https://doi.org/10.1145/3549532.

Li, D., Rau, P. L. P. and Li, Y. (2010) 'A Cross-cultural Study: Effect of Robot Appearance and Task', International Journal of Social Robotics, 2(2), pp. 175–186. Available at: https://doi.org/10.1007/s12369-010-0056-9.

Li, R. et al. (2019) 'Comparing Human-Robot Proxemics Between Virtual Reality and the Real World', in 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI). 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 431–439. Available at: https://doi.org/10.1109/HRI.2019.8673116.

Lohse, M. (2011) 'Bridging the gap between users' expectations and system evaluations', in 2011 RO-MAN. 2011 RO-MAN, pp. 485–490. Available at: https://doi.org/10.1109/ROMAN.2011.6005252.

Lu, L., Cai, R. and Gursoy, D. (2019) 'Developing and validating a service robot integration willingness scale', International Journal of Hospitality Management, 80, pp. 36–51. Available at: https://doi.org/10.1016/j.ijhm.2019.01.005.

Mara, M. et al. (2021) 'User Responses to a Humanoid Robot Observed in Real Life, Virtual Reality, 3D and 2D', Frontiers in Psychology, 12. Available at: https://doi.org/10.3389/fpsyg.2021.633178.

Mori, M. (1970) 'Bukimi no tani (the uncanny valley)', Energy, 7(4), pp. 33–35.

Negi, S. et al. (2008) 'Psychological assessment of humanoid robot appearance and performance using virtual reality', in RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication. RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication, pp. 719–724. Available at: https://doi.org/10.1109/ROMAN.2008.4600752.

Phillips, E. et al. (2017) 'What Does A Robot Look Like?: A Multi-Site Examination of User Expectations About Robot Appearance', Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 61, pp. 1215–1219. Available at: https://doi.org/10.1177/1541931213601786.

Phillips, E. et al. (2018) 'What is Human-like?: Decomposing Robots' Human-like Appearance Using the Anthropomorphic roBOT (ABOT) Database', in 2018 13th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 105–113. Available at: https://doi.org/10.1145/3171221.3171268.

Randall, N. and Sabanovic, S. (2023) 'A Picture Might Be Worth a Thousand Words, But It's Not Always Enough to Evaluate Robots', in Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction. New York, NY, USA: Association for Computing Machinery (HRI '23), pp. 437–445. Available at: https://doi.org/10.1145/3568162.3576970.

Roberts, S., Page, R. and Richardson, M. (2020) 'Designing in virtual environments: The integration of virtual reality tools into industrial design research and education', DRS Biennial Conference Series [Preprint]. Available at: https://doi.org/10.21606/drs.2020.284.

Roesler, E. et al. (2022) 'Why Context Matters: The Influence of Application Domain on Preferred Degree of Anthropomorphism and Gender Attribution in Human–Robot Interaction', International Journal of Social Robotics, 14(5), pp. 1155–1166. Available at: https://doi.org/10.1007/s12369-021-00860-z.

Rosete, A. et al. (2020) 'Service Robots in the Hospitality Industry: An Exploratory Literature Review', in H. Nóvoa, M. Drăgoicea, and N. Kühl (eds) Exploring Service Science. Cham: Springer International Publishing (Lecture Notes in Business Information Processing), pp. 174–186. Available at: https://doi.org/10.1007/978-3-030-38724-2_13.

Stadler, S. et al. (2020) 'IMPRO: Immersive Prototyping In Virtual Environments For Industrial Designers', Proceedings of the Design Society: DESIGN Conference, 1, pp. 1375–1384. Available at: https://doi.org/10.1017/dsd.2020.81.