Effects of Multimodal Cues on Children's Perception of Uncanniness in a Social Robot

Maike Paetzel
Dept. of Information Technology,
Uppsala University, Sweden
maike.paetzel@it.uu.se

Christopher Peters
KTH Royal Institute of Technology,
Stockholm, Sweden
chpeters@kth.se

Ingela Nyström,
Ginevra Castellano
Dept. of Information Technology,
Uppsala University, Sweden
{ingela.nystrom, ginevra.castellano}@it.uu.se

ABSTRACT

This paper investigates the influence of multimodal incongruent gender cues on the perception of a robot's uncanniness and gender in children. The back-projected robot head Furhat was equipped with a female and male face texture and voice synthesizer and the voice and facial cues were tested in congruent and incongruent combinations. 106 children between the age of 8 and 13 participated in the study. Results show that multimodal incongruent cues do not trigger the feeling of uncanniness in children. These results are significant as they support other recent research showing that the perception of uncanniness cannot be triggered by a categorical ambiguity in the robot. In addition, we found that children rely on auditory cues much stronger than on the facial cues when assigning a gender to the robot if presented with incongruent cues. These findings have implications for the robot design, as it seems possible to change the gender of a robot by only changing its voice without creating a feeling of uncanniness in a child.

CCS Concepts

•**Human-centered computing** → **Empirical studies in HCI;** •**Computing methodologies** → *Cognitive robotics; Intelligent agents;*

Keywords

Uncanny valley, child-robot interaction, multimodal voice and facial expressions

1. INTRODUCTION

The *categorization ambiguity* theory of uncanniness claims that a robot's uncanniness is not a function of its human-likeness, but it arises at any boundary between two opposing categories [8][19]. While the relation between human-likeness and uncanniness has already been researched in different dimensions [5][13][15], the relation between the gender of the robot and its uncanniness has rarely been investigated. Further, studies on children's perception of uncanniness are uncommon, with some noteable exceptions [16][27]. However, this demographic represents an important set of end-users for applications, e.g., robots as tutors or companions.

This paper investigates whether multimodal incongruent gender cues can trigger a feeling of uncanniness when interacting with the back-projected robot head Furhat [3] by creating an ambiguity in the category of gender. While previous work was mainly focused on the appearance of the robot, we chose to add mismatched vocal cues to the facial texture, such as a female voice to the male face texture, because people rely strongly on auditory cues when judging a character's personality [11][18]. A large user study including 106 children between age 8 and 13 was conducted. Contrary to the expectations, results show that the multimodal incongruence does not trigger the feeling of uncanniness in children. This finding supports other recent research showing that the categorical ambiguity theory cannot fully explain the uncanny valley effect [10][12].

A second focus of this paper is the children's perception of a robot's gender. We found children to rely on auditory cues much stronger than on facial texture when assigning a gender to the robot when presented with incongruent cues. We therefore conclude that it seems possible to vary the robot's gender by altering its voice without raising the feeling of uncanniness when interacting with the robot.

The paper is organized as follows: After discussing relevant related work and our research hypothesis (Section 2), we describe the experiment design (Section 3), summarize obtained results (Section 4) and end with an outline of how our current work extends the work on perception of uncanniness and gender in robots (Section 5).

2. BACKGROUND

2.1 Related Work

When Masahiro Mori [20] first described the uncanny valley, he suggested that increasing the human-likeness in a robot might at some point cause a decrease in its likeability. Much research has been carried out to determine what causes such decrease and the robot to be perceived as uncanny [5][13][15]. In the following, we consider the conceptual definition of uncanniness by Bartneck et al. [4][5] and

ICMI'16, November 12–16, 2016, Tokyo, Japan

http://dx.doi.org/10.1145/2993148.2993157

Ho and MacDorman [9] as summarized by Kätsyri et al. [10]. Still, the underlying cause of the uncanny valley is controversial. Some research suggest that uncanniness rises at any category boundary where the ambiguity about the category is high [8][19]. Kätsyri et al. [10] recently reviewed empirical studies on the effect of uncanniness and argued that they support the *perceptual mismatch* and not the *categorization ambiguity* theory as an explanation of the uncanny valley, where perceptual mismatch refers to any mismatch in the human-likeness perception of a character.

While previous work has mostly focused on categorization ambiguities and perceptual mismatch in the appearance of the robot, adding vocal cues seems promising as people rely strongly on auditory cues when perceiving personality. The voice changes the perception of human-likeness in virtual agents [18] and reveals information about the speaker's gender, current mood, perceived attractiveness or dominance [11]. Mower et al. [21] even showed that people pay more attention to auditory cues when presented with conflicting emotional cues between voice and facial expressions.

The perception of the uncanny valley in children has rarely been investigated so far. Woods et al. [27] found that children (age 9 – 11) assign a more negative *Behavioral Intention* to images of human-like robots compared to those featuring cartoon-like robots. Tinwell and Sloan [25] found children (age 9 – 11) to perceive uncanniness in virtual agents similar to adults, while Minato et al. [16] describes the perception of uncanniness to be dependent on the children's age, with infants up to the age of 13 months showing more attraction than fear compared to children between three and five years. In this work, we focus on children of age 8 and older, as Borger et al. [7] found interviewing children under the age of 8 to be difficult due to their cognitive development and other measurement methods apart from interviews are currently not well explored in uncanny valley related research.

2.2 Hypothesis

In this paper, we investigate if the categorization ambiguity between the category of male and female gender can rise the feeling of uncanniness in children. As it has already been shown that multimodal incongruent cues between voice and face can cause an uncanny feeling in the dimension of realism of a robot [17], we hypothesize that we can use multimodal incongruent gender cues to push the robot to the border in the dimension of gender and thereby create a feeling of uncanniness:

Hypothesis: A robot with incongruent multimodal gender cues is perceived as more uncanny than a robot with congruent cues.

This hypothesis is based on the theory of categorization ambiguity and with this paper we want to further investigate its influence on the perceived uncanniness for the category of gender, which has not been studied in this context so far.

3. METHOD

We conducted a between-subject experiment in which children watched a two minute introduction of a robot with either congruent or incongruent gender cues and answered a questionnaire about their perception of the robot.

3.1 Participants

106 children between 8 and 13 years (with 90.57% of the children $\geq$ 11 years) met the robot and handed in a valid questionnaire afterwards. As the allocation of the experiment conditions was randomized and not all children decided to take part in the survey, the amount of results per condition varies (cf. Table 1). The questionnaires from 23 children were incomplete and excluded from the study.

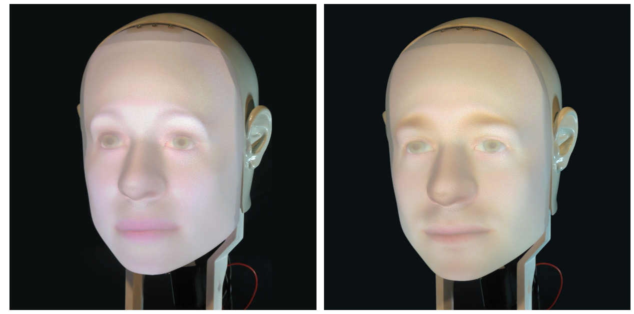

Figure 1: Furhat with the female texture to the left and the male texture to the right.

cond.	N	female	male	other	age
ff-fv	38	18	20	0	11.97 (0.21)
mf-mv	24	12	12	0	11.91 (0.22)
mf-fv	29	13	15	1	11.55 (0.18)
ff-mv	15	7	8	0	11.33 (0.35)

Table 1: Summary of participant's demographics per condition. Age is reported as average (standard deviation).

3.2 Stimulus

To test our hypothesis we used the back-projected robot head Furhat [3]. The head consists of a mask with the shape of an adult face. In a previous study [22], we found the mask to be perceived as very masculine and dominant if no face texture was projected onto it. Even though the face texture can be varied between male and female, the masculine shape of the mask limits the ability to create a feminine robot.

In our experiment, *one modality of gender was controlled by varying the face texture.* In an earlier online study [22], we selected one male and one female FaceGen [2] model with strong perceived femininity and masculinity. Even though the perception of the projection differed from the perception of the 2D image, participants mostly assigned the gender according to the design intention to the robot and the projections differed significantly in terms of perceived femininity and masculinity between the male (cf. Fig. 1 right) and female texture (cf. Fig. 1 left).

The *second modality of gender was controlled using the robot's voice.* The commercial software *CereProc* [1] was used as a human-like synthesizer with *William* as the male and *Sarah* as the female version [22]. We used the Java-based dialog framework IrisTK [24] to control the robot. Both speech and animation behavior was fully pre-script.

3.3 Procedure

The experiment was conducted during SciFest, a three-day science festival for children in Uppsala, Sweden. The experiment space was separated from the other parts of the

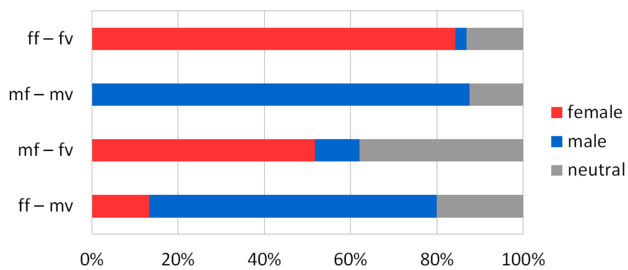

Figure 2: Distribution of assigned gender p. cond.

exhibition by curtains. Four kids could enter the experiment space at the same time. Chairs were placed in two rows to control for the horizontal and vertical space between the children and the robot. Lightning in the venue was artificial and remained the same throughout the whole experiment time. The noise in the exhibition hall could not be controlled and differed dependent on the crowdedness of the festival.

When the children entered the experiment space, the robot head was still covered by a blanket. The experimenter informed the children that they were going to meet a robot which would give a short introduction about itself and that they cannot interact with the robot directly. The experimenter then uncovered the robot and started the two-minute multimodal introduction, in which the robot talked about itself using voice, facial expressions and head movement. After the introduction, the experimenter invited the children to answer some questions about their perception of the robot. Participants received candy as a reward for answering the questionnaire. However, the participation in the survey was voluntary and not all kids who watched the robot interaction decided to take part. For answering the survey the children were brought to a separate table outside the experiment space with no visual contact to the robot.

3.4 Independent Variables

As independent variables we selected *gender of the face texture* and *gender of the voice* with the two levels *male* and *female* each. They were combined into two congruent and two incongruent conditions.

Congruent conditions: In the two congruent conditions the female face texture was paired with the female voice (condition **ff-fv**) and the male face was paired with the male voice (**mf-mv**).

Incongruent gender: The female face was paired with the male voice (**ff-mv**) and the male face was paired with the female voice (**mf-fv**).

3.5 Dependent Variables

The post-questionnaire contains three demographic questions and 17 perception questions. All perception questions except for the assignment of a gender are rated on a 5-point Likert scale. In the first part of the questionnaire, the children are presented with statements and indicate how much they agree with the statement. The second part consists of opposing traits in different dimensions, e.g., *unintelligent* vs. *intelligent*, and the children indicate their impression on the presented scales. As the children were non-native English speakers, Swedish translations by a native Swedish speaker are provided on the questionnaire.

The gender of the robot is covered by rating the perceived masculinity and femininity of the robot on separate scales and the perceived difficulty to assign a gender to the robot. In addition, the kids are asked to explicitly assign a female, male or neutral gender to the robot.

The measure of the robot's perception is mostly based on a shortened version of the Godspeed questionnaire by Bartneck et al. [6], covering the *anthropomorphism*, *likability* and *intelligence* of the robot. The questionnaire is completed by adding the dimensions of perceived *familiarity* and *strangeness* [23], *trustworthiness* [14] and *threat* [26].

4. RESULTS

The answers obtained in the four different conditions were tested for significance using One-way ANOVA with alpha level of .05. To accommodate the unequal sample size, type III sum of squares was used for calculations. The main results are summarized in Fig. 2 and Fig. 3.

In the congruent condition the robot with the male voice and face texture is perceived as significantly *more masculine* ($M = 4.21$, $SD = 0.19$) than the robot with the female voice and face texture ($M = 2.15$, $SD = 0.21$), $F(1, 60) = 45.8$, $p < .001$, and 87.5% children assigned a male gender to the robot (cf. Fig. 2). The robot with the female voice and female face texture was perceived as significantly *more feminine* ($M = 3.97$, $SD = 0.16$) than the male version ($M = 1.46$, $SD = 0.2$), $F(1, 60) = 98.08$, $p < .001$, but the children reported it to be significantly *more difficult to assign a gender* to the robot at the same time ($M = 2.05$, $SD = 0.17$), $F(1, 60) = 5.58$, $p = .021$. Still, the vast majority of the children assigned the female gender to the robot (84.2%). The difficulty in assigning a gender is in line with our previous work [22] and likely caused by the subtle incongruence between Furhat's masculine mask and the female face texture. However, this does not affect the overall rating of the congruent female robot and we even see a small preference for this condition, as it is perceived as *more trustworthy* ($M = 4.05$, $SD = 0.16$), $F(1, 60) = 5.88$, $p = .018$. There are no significant differences for the other dimensions, like *perceived strangeness*, $F(1, 60) = 1.0$, $p = .321$, or *pleasance*, $F(1, 60) = 0.96$, $p = .332$, (cf. Fig. 3).

In the incongruent gender condition, the children found it in general significantly *more difficult to determine the gender* of the robot ($M = 2.86$, $SD = 0.16$), $F(1, 104) = 25.85$, $p < .001$. Among the incongruent gender conditions, the robot with the female voice and male face texture is perceived significantly *more feminine* ($M = 3.62$, $SD = 0.18$) than the robot with the male voice and the female face texture ($M = 1.87$, $SD = 0.27$), $F(1, 42) = 30.02$, $p < .001$, but *not less masculine* ($M = 2.93$, $SD = 0.19$) than ff-mv ($M = 3.4$, $SD = 0.43$), $F(1, 42) = 1.35$, $p = .251$. This trend is also reflected in the assignment of the gender. For the male voice and female texture, 66.67% assigned a male gender, 13.33% a female gender and 20% a neutral gender. In the condition with female voice and male texture, 51.72% assigned a female gender, 10.34% a male gender and 37.93% a neutral gender (cf. Fig. 2). In the incongruent gender condition, we do not see any significant difference in terms of perception between the versions. Except for the rating of masculinity and femininity we did not find significant differences in the perception between the congruent condition and the incongruent gender condition.

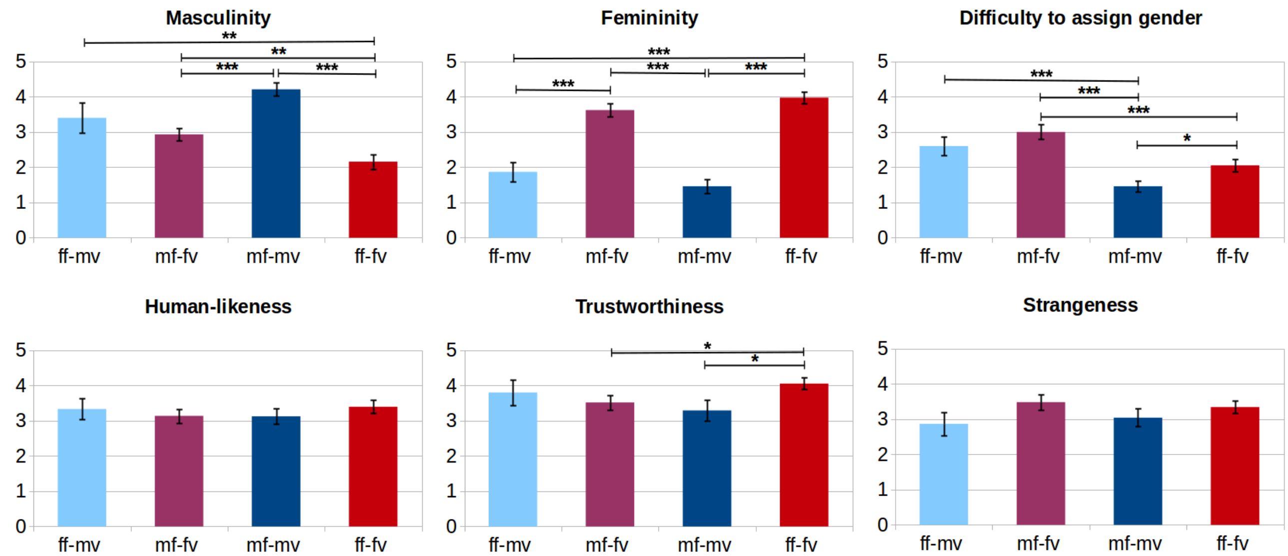

Figure 3: Differences in the 5-point Likert scale questionnaire response per condition (means and standard errors). Significant differences in One-way ANOVA are indicated by * ($p < .05$), ** ($p < .01$), * ($p < .001$)**

5. DISCUSSION

Although we see that the assigned gender and perceived masculinity and femininity differ significantly between the congruent and incongruent robot designs, we did not find significant difference in perception of uncanniness between the conditions. Contrary to what hypothesized, our findings show that *changing the gender by only using the voice does not trigger a feeling of uncanniness in children, even if the appearance of the robot shows opposing gender cues.* To investigate if the findings are related to the specific age group in this study, future research comparing the results to the responses of teenagers and adults is necessary.

Our findings *support other recent research which states that the categorical ambiguity theory cannot fully explain the uncanny valley effect* [10][12]. While the related work mostly focused on the dimension of human-likeness, our results can expand previous work questioning the influence of categorical ambiguity by showing that the findings are transferable to the dimension of gender.

If confronted with incongruent gender cues in the Furhat robot, *children tend to rely more on the auditory cues than on the facial cues when judging the overall gender of the robot*, which is in line with similar findings by Mower et al. [21] using emotional cues. This is most clearly visible for the female texture and male voice, where more than 60% of the children assigned a male gender to the robot. This high number might be influenced by the subtle incongruence in the facial cues as the shape of the face is masculine [22]. However, we see the same trend for the male texture and the female voice, where still about 50% of children assign a female and only about 10% assign a male gender.

We consider these results to be very useful for the upcoming research on back-projected robot heads as it gives confidence that we can assign a gender to the robot by altering the voice even though the mask might communicate different gender cues. It seems possible that this result can even be transferred to other robot embodiments, as the finding is in line with previous research on virtual agents [21], where the facial cues are even better visible compared to the robot's back-projection. However, we acknowledge that the facial dynamics in Furhat is different to a virtual character, so research with different embodiments is necessary to gain more confidence about the transferability. Besides the specific appearance, this study was also limited to the choice of a specific voice synthesizer. To clarify the influence of this synthesizer on the results, another study comparing different synthesizers is needed.

6. CONCLUSION

This paper investigated the effect of multimodal gender incongruence on the perception of a robot's uncanniness and gender in children. The back-projected robot platform Furhat was used to alter vocal and facial cues in two congruent and incongruent conditions. We found that multimodal incongruence does not trigger a feeling of uncanniness, which supports other research claiming the categorical ambiguity theory cannot fully explain the uncanny valley effect. Children were found to rely on the vocal cues much stronger than on the facial cues when assigning a gender to a robot. As the perception of uncanniness is not influenced by a gender mismatch between auditory and facial cues, we argue that it seems possible to alter the gender of a robot by changing the voice without being perceived as uncanny in children.

7. ACKNOWLEDGMENTS

Thanks to Y. Abileva, J. Lin, A. Schlegel, R. Nyström, S. Mohammedsalih, P. Olssen, O. Rodríguez, P. Rodríguez, D. Spiers and E. White for help with collecting data and L. Oestreicher for Fig.1. C. Peters' work is partly supported by the European Commission (EC) Horizon 2020 ICT 644204 project ProsocialLearn. The authors are solely responsible for the content of this publication. It does not represent the opinion of the EC, and the EC is not responsible for any use that might be made of data appearing therein.

8. REFERENCES

[1] CereProc Synthesizer. https://www.cereproc.com/. Accessed: 2016-05-08.

[2] FaceGen Modeller. http://facegen.com/. Accessed: 2016-05-08.

[3] S. Al Moubayed, J. Beskow, G. Skantze, and B. Granström. Furhat: a back-projected human-like robot head for multiparty human-machine interaction. In *Cognitive Behavioural Systems*, pages 114–130. Springer, 2012.

[4] C. Bartneck, T. Kanda, H. Ishiguro, and N. Hagita. Is the uncanny valley an uncanny cliff? In *The 16th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 368–373. IEEE, 2007.

[5] C. Bartneck, T. Kanda, H. Ishiguro, and N. Hagita. My robotic doppelgänger - A critical look at the uncanny valley. In *The 18th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 269–276. IEEE, 2009.

[6] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, 1(1):71–81, 2009.

[7] N. Borgers, E. De Leeuw, and J. Hox. Children as respondents in survey research: Cognitive development and response quality 1. *Bulletin de Methodologie Sociologique*, 66(1):60–75, 2000.

[8] M. Cheetham, P. Suter, and L. Jäncke. The human likeness dimension of the "uncanny valley hypothesis": behavioral and functional MRI findings. *Frontiers in Human Neuroscience*, 5(126):10–3389, 2011.

[9] C.-C. Ho and K. F. MacDorman. Revisiting the uncanny valley theory: Developing and validating an alternative to the godspeed indices. *Computers in Human Behavior*, 26(6):1508–1518, 2010.

[10] J. Kätsyri, K. Förger, M. Mäkäräinen, and T. Takala. A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in Psychology*, 6, 2015.

[11] M. Latinus and P. Belin. Human voice perception. *Current Biology*, 21(4):R143–R145, 2011.

[12] K. F. MacDorman and D. Chattopadhyay. Reducing consistency in human realism increases the uncanny valley effect; increasing category uncertainty does not. *Cognition*, 146:190–205, 2016.

[13] K. F. MacDorman, R. D. Green, C.-C. Ho, and C. T. Koch. Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior*, 25(3):695–710, 2009.

[14] R. McDonnell, M. Breidt, and H. Bülthoff. Render me real?: investigating the effect of render style on the perception of animated virtual humans. *ACM Transactions on Graphics*, 31(4):91, 2012.

[15] L. F. Meah and R. K. Moore. The uncanny valley: A focus on misaligned cues. In *6th International Conference on Social Robotics (ICSR)*, pages 256–265. Springer, 2014.

[16] T. Minato, M. Shimada, H. Ishiguro, and S. Itakura. Development of an android robot for studying human-robot interaction. In *Innovations in Applied Artificial Intelligence*, pages 424–434. Springer, 2004.

[17] W. J. Mitchell, K. A. Szerszen, A. S. Lu, P. W. Schermerhorn, M. Scheutz, and K. F. MacDorman. A mismatch in the human realism of face and voice produces an uncanny valley. *i-Perception*, 2(1):10–12, 2011.

[18] Y. Moon and C. Nass. How "real" are computer personalities? Psychological responses to personality types in human-computer interaction. *Communication Research*, 23(6):651–674, 1996.

[19] R. K. Moore. A Bayesian explanation of the 'Uncanny Valley' effect and related psychological phenomena. *Scientific Reports*, 2, 2012.

[20] M. Mori, K. F. MacDorman, and N. Kageki. The uncanny valley [from the field]. *Robotics & Automation Magazine, IEEE*, 19(2):98–100, 2012.

[21] E. Mower, M. J. Mataric, and S. Narayanan. Human perception of audio-visual synthetic character emotion expression in the presence of ambiguous and conflicting information. *Multimedia, IEEE Transactions on*, 11(5):843–855, 2009.

[22] M. Paetzel, C. Peters, I. Nyström, and G. Castellano. Congruency matters - How ambiguous gender cues increase a robot's uncanniness. In *8th International Conference on Social Robotics (ICSR)*, 2016.

[23] A. M. Rosenthal-von der Pütten and N. C. Krämer. How design characteristics of robots determine evaluation and uncanny valley related responses. *Computers in Human Behavior*, 36:422–439, 2014.

[24] G. Skantze and S. Al Moubayed. IrisTK: a statechart-based toolkit for multi-party face-to-face interaction. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction*, pages 69–76. ACM, 2012.

[25] A. Tinwell and R. J. Sloan. Children's perception of uncanny human-like virtual characters. *Computers in Human Behavior*, 36:286–296, 2014.

[26] M. L. Walters, D. S. Syrdal, K. Dautenhahn, R. Te Boekhorst, and K. L. Koay. Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion. *Autonomous Robots*, 24(2):159–178, 2008.

[27] S. Woods, K. Dautenhahn, and J. Schulz. The design space of robots: Investigating children's views. In *Robot and Human Interactive Communication, 2004. ROMAN 2004. 13th IEEE International Workshop on*, pages 47–52. IEEE, 2004.