i-PERCEPTION
a Pion publication
i-Perception (2011) volume 2, pages 10 – 12
dx.doi.org/10.1068/i0415 ISSN 2041-6695 perceptionweb.com/i-perception

SHORT AND SWEET

A mismatch in the human realism of face and voice produces an uncanny valley

Wade J Mitchell
School of Informatics, Indiana University, 535 West Michigan St, Indianapolis, IN 46202, USA;
e-mail: wamitche@iupui.edu
Kevin A Szerszen, Sr
School of Informatics, Indiana University, 535 West Michigan St, Indianapolis, IN 46202, USA;
e-mail: keszersz@iupui.edu
Amy Shirong Lu
School of Informatics, Indiana University, 535 West Michigan St, Indianapolis, IN 46202, USA;
e-mail: amylu@iupui.edu
Paul W Schermerhorn
Cognitive Science Program, Indiana University, 1900 E 10th St, Bloomington, IN 47406, USA;
e-mail: pscherme@indiana.edu
Matthias Scheutz
Cognitive Science Program, Indiana University, 1900 E 10th St, Bloomington, IN 47406, USA and
Department of Computer Science, Tufts University, 161 College Ave, Medford, MA 02155, USA;
e-mail: mscheutz@cs.tufts.edu
Karl F MacDorman
School of Informatics, Indiana University, 535 West Michigan St, Indianapolis, IN 46202, USA;
e-mail: kmacdorm@indiana.edu

Received 7 November 2010, in revised form 17 February 2011; published online 1 March 2011

Abstract. The uncanny valley has become synonymous with the uneasy feeling of viewing an animated character or robot that looks imperfectly human. Although previous uncanny valley experiments have focused on relations among a character's visual elements, the current experiment examines whether a mismatch in the human realism of a character's face and voice causes it to be evaluated as eerie. The results support this hypothesis.

Keywords: anthropomorphism, facial–vocal mismatch, human realism, Masahiro Mori, social perception.

Mori (1970) proposed a nonlinear relation between a character's degree of human realism and our subjective sense of rapport: the more human the character looks the more comfortable we feel interacting with it until a point is reached at which subtle nonhuman flaws cause the character to seem eerie, like an animated corpse. Mori dubbed this dip in rapport *bukimi no tani* (the uncanny valley).

Although Mori conducted no experiments on the uncanny valley, he cited stimuli that could produce the described effect, including a prosthetic hand that looks real but feels cold and hard to the touch. In this example, there is a cross-modal mismatch: the visual appearance of the hand elicits the tactile expectation that it will feel as warm and soft as a human hand. The violation of this expectation causes more than surprise. There is a sense of the macabre, which Jentsch (1906) identified with uncertainty concerning whether the entity is animate or inanimate. This sense may be highest for an entity resembling a human being because of the viewer's self-identification (MacDorman et al 2009b; Ramey 2005). Theories ranging from the biological to the cultural have been proposed to explain the uncanny valley (MacDorman and Ishiguro 2006; Misselhorn 2009; Moosa and Minhaz Ud-Dean 2010).

Attributions of eeriness have been elicited in empirical studies by a mismatch in the human photorealism of a character's visual elements, such as eyes and face; other treatments include pairing a realistic human skin texture with atypical face height, eye separation, and eye size (MacDorman et al 2009a; Seyama and Nagayama 2007). Although Tinwell et al (2010)

have found a visual–auditory mismatch correlates with uncanniness, no experiment has yet been conducted that manipulates facial and vocal human realism as independent variables. This experiment is intended to fill that gap.

The following prediction (the hypothesis) is made by the theory that a cross-modal mismatch in human realism causes uncertainty about whether an entity is animate or inanimate, thereby eliciting feelings of eeriness: a robot with a human voice, or a human being with a synthetic voice, will be perceived as eerier than a robot with a synthetic voice or a human being with a human voice.

Forty-eight US-born participants (28 female, 20 male) were recruited in April 2010 from a sample of undergraduate students from a nine-campus Midwestern university. Their mean age was 21.2 ($SD = 3.7$). There were no significant differences in the experimental results by age or gender.

In this within-group experiment, each participant viewed, in random sequence, four 14 s videos of a character reciting neutral phrases. Each video corresponded to either matched (*robot figure–synthetic voice, human figure–human voice*) or mismatched stimulus conditions (*robot figure–human voice, human figure–synthetic voice*). Each video played in a loop until the participant completed validated indices on the character's *humanness*, *eeriness*, and interpersonal *warmth* (Ho and MacDorman 2010). Each index averaged the results of five-to-eight 7-point semantic differential scales, ranging from –3 to +3. The order of video presentation and the scales was randomized to prevent order effects. Data analysis was performed in SPSS.

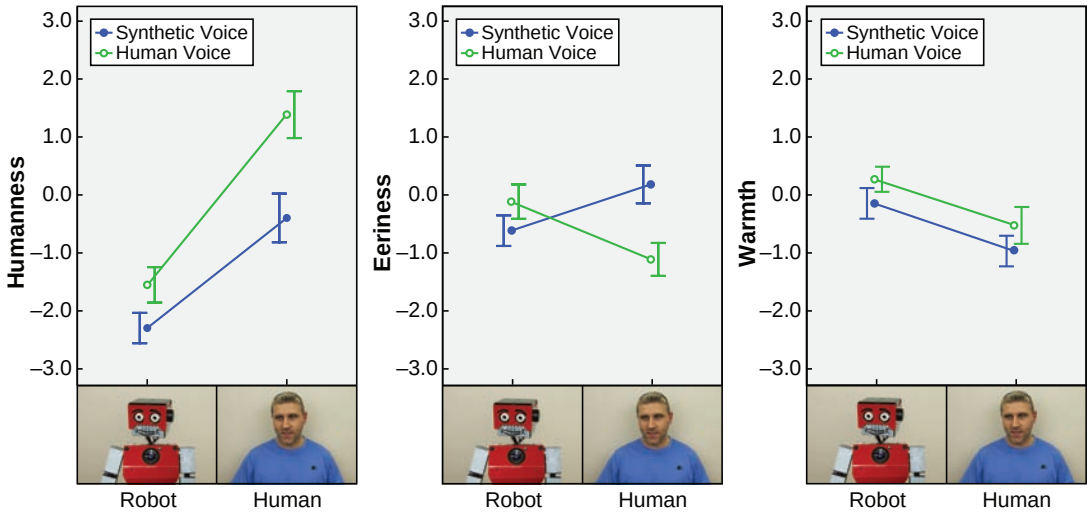

Figure 1. A human voice heightened the eeriness of the robot, while a synthetic voice heightened the eeriness of the human. The error bars indicate 95% confidence intervals.

The three indices were not significantly correlated and were normally distributed and reliable (Cronbach's αs ranged from 0.70 to 0.88). For humanness a two-way repeated measures ANOVA found a significant main effect for face realism [$F(1,47) = 110.15$, $p < 0.001$, $\eta^2 = 0.70$] and voice realism [$F(1,47) = 75.94$, $p < 0.001$, $\eta^2 = 0.62$] and a significant interaction effect [$F(1,47) = 18.65$, $p < 0.001$, $\eta^2 = 0.28$]. The human figure–human voice condition rated the highest [$M = 1.40$, $SE = 0.20$], and the robot figure–synthetic voice condition rated the lowest [$M = –2.29$, $SE = 0.13$] (figure 1). For eeriness there was a significant main effect for voice realism [$F(1,47) = 13.28$, $p = 0.001$, $\eta^2 = 0.22$] and a significant interaction effect [$F(1,47) = 36.51$, $p < 0.001$, $\eta^2 = 0.44$]. The two mismatched conditions, robot figure–human voice [$M = –0.10$, $SE = 0.15$] and human figure–synthetic voice [$M = 0.19$, $SE = 0.16$], rated significantly

higher on eeriness than the two matched conditions, robot figure–synthetic voice [$M = -0.60$, $SE = 0.13$] and human figure–human voice [$M = -1.10$, $SE = 0.14$], by a paired samples t-test [$t(47) = 6.042$, $p < 0.001$]. For warmth there was a significant main effect for face realism [$F(1,47) = 27.62$, $p < 0.001$, $\eta^2 = 0.37$] and voice realism [$F(1,47) = 11.15$, $p = 0.002$, $\eta^2 = 0.19$] but no significant interaction effect. Warmth ratings were highest for robot figure–synthetic voice [$M = 0.28$, $SE = 0.11$] and lowest for human figure–synthetic voice [$M = -0.96$, $SE = 0.13$]. The higher warmth ratings for the robot conditions may be attributed to its cuteness relative to the seriousness of the ex-Marine human actor.

These results indicate incongruence in the human realism of a character's face and voice can elicit feelings of eeriness; thus, the hypothesis is supported. This suggests a design principle for synthetic agents to avoid the uncanny valley: the human realism of a character's visual elements and voice should match.

Acknowledgements. The authors would like to express their gratitude to Leslie Ashburn-Nardo, Dennis Devine, Chin-Chang Ho, Himalaya Patel, and the reviewers for their thoughtful suggestions on improving this paper. The IUPUI/Clarian Research Compliance Administration approved this experiment (EX1007-19B). This experiment was supported by an IUPUI Signature Center grant.

References

Ho C-C, MacDorman K F, 2010 "Revisiting the uncanny valley theory: developing and validating an alternative to the Godspeed indices" *Computers in Human Behavior* **26** 1508–1518 doi:10.1016/j.chb.2010.05.015 ◂

Jentsch E, 1906 "Zur Psychologie des Unheimlichen" [On the psychology of the uncanny] *Psychiatrisch-Neurologische Wochenschrift* **8** 195–198 ◂

MacDorman K F, Ishiguro H, 2006 "The uncanny advantage of using androids in social and cognitive science research" *Interaction Studies* **7** 297–337 doi:10.1075/is.7.3.03mac ◂

MacDorman K F, Green R D, Ho C-C, Koch C, 2009a "Too real for comfort: uncanny responses to computer generated faces" *Computers in Human Behavior* **25** 695–710 doi:10.1016/j.chb.2008.12.026 ◂

MacDorman K F, Vasudevan S K, Ho C-C, 2009b "Does Japan really have robot mania? Comparing attitudes by implicit and explicit measures" *AI & Society* **23** 485–510 doi:10.1007/s00146-008-0181-2 ◂

Misselhorn C, 2009 "Empathy with inanimate objects and the uncanny valley" *Minds and Machines* **19** 345–359 doi:10.1007/s11023-009-9158-2 ◂

Moosa M M, Minhaz Ud-Dean S M, 2010 "Danger avoidance: an evolutionary explanation of the uncanny valley" *Biological Theory* **5** 12–14 doi:10.1162/BIOT_a_00016 ◂

Mori M, 1970 "Bukimi no tani" [The uncanny valley] *Energy* **7** 33–35 ◂

Ramey C H, 2005 "The uncanny valley of similarities concerning abortion, baldness, heaps of sand, and humanlike robots", in Proceedings of Views of the Uncanny Valley Workshop, Fifth IEEE-RAS International Conference on Humanoid Robots, Tsukuba, Japan, pp 8–13 ◂

Seyama J, Nagayama R S, 2007 "The uncanny valley: the effect of realism on the impression of artificial human faces" *Presence: Teleoperators & Virtual Environments* **16** 337–351 doi:10.1162/pres.16.4.337 ◂

Tinwell A, Grimshaw M, Williams A, 2010 "Uncanny behaviour in survival horror games" *Games Computing and Creative Technologies* **2** 3–25 doi:10.1386/jgvw.2.1.3_1 ◂