A Visit on the Uncanny Hill

Petr Švarný

Charles University, Prague, Czech Republic
svarnyp@logici.cz

Abstract. The article introduces shortly the Uncanny valley hypothesis and sums up some of the research done in the field connected to it. It explores the possible new options in research or robot design which could help to subdue this uncanny obstacle on the way to a pleasant human-robot interaction. It also mentions the possible risk of an uncanny valley phenomenon the other way around, from the view of artificial intelligence (AI).

Keywords: human-robot interaction, uncanny valley

1 Introduction

This paper explores the so called Uncanny valley hypothesis in the light of the use of humanities and art in human-robot interaction. As all sorts of AI systems take a bigger part in our day to day lives, we more often face the question how to make human-robot interactions pleasant and natural-seeming. This problem was studied already in 1970 by M. Mori [1, 2], who introduced the hypothesis how people react to human-like entities. We will describe this hypothesis briefly and show some results concerning its verification. Thereafter we focus on possibilities how the hypothesis of an uncanny valley could be treated with inspiration coming from art. We suggest that the valley should be approached from the side of the AI also.

2 The Valley Ahead

The Uncanny valley hypothesis claims that the familiarity, affinity, or comfort of our contact with an entity that is similar in some respects to humans is not a simple linear function. Although it is true that the more human-like an entity is, the more we are comfortable while interacting with it, Mori supposed that there is a sudden drop in comfort as we reach a certain point of realism

and it does not cease unless we face a real human entity. According to this hypothesis, a human test subject should feel little affinity towards robots that are not similar to humans (see industrial robots). The subject should have some level of affinity to humanoid robots but should have an eerie sensation when confronted with an actroid[1]. It was already in the original article that the difference between motionless and moving entities was explored. Mori mentioned the different feeling we have when facing a simple prosthetic arm that is still and when we observe a myoelectric hand[2].

The topic caught more attention today than at the time when the article was published. Generally the hypothesis finds support in today's research. For example, we can see the attempts to broaden the studied aspects in [3]. However, there is present also an opposite view. We can take [4] as an example of an article that tries to eliminate the valley. Medical investigations also could be taken into account as prosopagnosia or the way how we react to first time exposure to unfamiliar faces might play an important role in the subject.[3]

3 Valley Hiking in the Modern World

One of the main questions to answer before we try to venture into the valley is, if it is necessary to climb up the hill to realism and affinity. A good artistic example of this could be Johnny 5 – he has rudimental options how to express emotions, he is not human-like but has some basic human characteristics, and he reacts similarly as a human being would. He represents a robot that is comfortable to interact with, although he does not have human-like features.

However, Hanson et al. present the following reason why it is worth trying to achieve realistic human robots:

> ... realistically depicted humanlike robotics will serve as an unparalleled tool for investigating human social perception and cognition. In our experiments, our robots have demonstrated clearly that realistic robots can be appealing. We conclude that rendering the social human in all possible detail can help us to better understand social intelligence, both scientifically and artistically. [4] (p. 31)

[1] An android that is visually very human-like.

[2] Basically a moving prosthetic arm. The mentioned example is directed by electric signals received from human skin surface.

[3] See for example [5] showing that basic observation of facial behaviour is deep-rooted and it is present already at a very young age. The great speed with which people react to facial stimuli is shown in the study [6].

This quote mentions social perception and cognition. Therefore, we can point out one of the possible problems connected to the studies of the uncanny valley – they do not use commitment and longer term cooperation. These are present in many human interactions and often play an important role in the formation of our social life. Any feelings of eeriness and discomfort connected to human-like robots could possibly vanish after a few days of interaction and be replaced with genuine affection.

However, we might not need realistically humanoid robots in order to have a comfortable human-robot interaction. As the first idea coming from art, we mention McCloud's observation from the art of drawing comics. He claims [7] (p. 31) that simple shapes allow the reader for more immersion as they allow for more universality. Any character that is depicted in a realistic manner is understood by the reader automatically as something different, something exterior to which he cannot relate that easily. This takes into account also the human tendency to recognize faces in many simple shapes (for example due to pareidolia) and allows us to construct robots with simple forms of facial expressions. Nevertheless, we need to pay attention to the fact that the immersion present in comics is due to some other factors also: we are often the witnesses of the character's thoughts, the character is expressing emotions, and she is reacting to the situations she faces in an unsurprising way. This would suggest that a successful comics based interaction is given by a robot that has a simple facial interface and reacts in a way we would expect it to react.

We can drop the option to share inner thought processes for two reasons. First, it is a common and quite accepted response in a conversation between people to answer: "I don't know", when one is asked about a difficult thought process. Second, if the robot achieves the other two mentioned points, it will be attributed a mind by his human colleagues.

We cannot leave the other two demands aside. Being confronted with humanoid robots that do not react in an expected way can be similar to facing a human that reacts abnormally. It leads to a reaction of fear and panic because the theory of the mind of the encountered person fails to predict or explain his actions. The fact that unexpected behaviour is alien to us already from early age is shown for example in [5]. Infants react strongly if their communication counterpart does not follow the usual pattern of behaviour and suddenly stops reacting to stimuli.

For the second demand, if we would not request a simple facial interface, we would return to the original idea of trying to make human-like robots instead of making only robots that are pleasant to interact with or we would remain with machine-like robots. At this point it is our main concern to ameliorate

the interaction between humans and robots at the lowest cost possible. If we focus on facial realism, we might end up with a machine that is great at expressing emotions but is too complex for a daily use in our lives. On the other hand, if we omit facial features altogether we fail to facilitate the human-machine interaction. For this reason we want to stay with a design as simple as possible.

In many respects the fact that human communication is nowadays often also dependent on a computer interface facilitates our attempts to befriend humans with robots. Many people grow up expressing their emotions in emoticons and text messages and receiving emotional responses in a similar way. A recent movie named Moon has shown a robot called Gerty that communicated with an emotionally neutral voice but his statements were accompanied with an emoticon on his main screen showing his mood. It was thanks to this small screen that communication with Gerty seemed much more pleasant than communication with HAL9000 from the movie 2001: A Space Odyssey.

Many other interactions do not even need any visual interface to work properly. Already the old psychoanalysis program called Eliza has proven somewhat effective in fooling people into believing she had some mind or intelligence, although she had none [8]. A modern counterpart of Eliza is Apple's Siri, an intelligent personal assistant that responds to voice commands and reacts only in voice or by giving the demanded output behaviour (for example, sending an email). Obviously such applications do not fall into the uncanny valley, but they show how minute the trouble with the valley can be. Emotional modulation of the AI's voice could be enough to give people (already used to talking over phones) enough feedback to make the interaction close to a human-human exchange. The crucial point is the difference in importance people ascribe to visual and auditory stimuli. In order for the conversation to meet our two demands, the robot could even have a static chassis and demonstrate all its reactions by his audio systems. This view also leads to the important question of application. What would be the use of a human-like realistic robot?

As the subtitle of the conference is "artificial dreams", the reference to P. K. Dick's "Do androids dream of electronic sheep?" comes into mind. The human-like androids in that world are used for mining and similar labour. Such use seems simply unrealistic as it would probably be more cost effective to have specialized machines for these purposes. The scenario of personal assistants is a more realistic and probable one. Following in the footsteps of Siri they could take the form of an audio responding humanoid with supressed or simplistic and non-changeable facial features. We return here again to the question if the valley needs to be crossed. Employing a realistic humanoid assistant would

only lead to affinity towards this assistant and possible impairment on the effectiveness of its use (for example one would want his assistant to take some rest or go for vacation). On the other hand, a well-designed assistant – let us say still on the hill before the steep drop into the valley – could already make its human user comfortable enough but prevent him from ascribing too many human characteristics to the assistant. This could be achieved by maintaining an illusion of correct emotional response and simplistic representation.

4 Foreign Visitors to the Valley

We focused the whole time on the human-robot interaction. If we imagine, however, a robot already capable of genuine emotional response, we can ask also about the robot-human interaction. If there is a human-robot uncanny valley, would there be also one for the artificial participants in the conversation? How would their emotions react to other robots, perceived by humans as uncanny? Obviously it is a question closely tied to the mechanisms that would be incorporated into these robots and thus for now unanswerable.

However, it might already be the time to start evaluating whether we shouldn't prepare artificial/AI/robot equivalents of some humanities. Especially psychology could be transformed into a tool to work with AI from a top-down perspective. This might need to be as specialized as its human counterpart and couldn't be simply presented as some interdisciplinary effort between psychology and AI. A more "biological" approach to robots and AI could also help to classify any eeriness or bizarre behaviour as AI counterparts of human abnormal states without getting lost in too complex bottom-up descriptions and at the same time it would allow the treatment of AI in a similar manner as humans or animals are treated. A good example of a topic from psychology that could be useful for our cause is the Asperger syndrome. A person suffering from this disorder might often make other people uncomfortable and thus slip into the uncanny valley.

The ultimate use of many of the here mentioned ideas – even the use of non-human like assistants or psychological classifications – is closely tied to the ethics of AI. Do we want to ascribe the same status to beings evolved from human research and effort as to those that evolved from the chaos of the universe?

5 Conclusion

We have introduced the idea of the uncanny valley from M. Mori that robots that are human-like might make people feel eerie because of their imperfect

similarity to humans. We suggested that the valley does not have to be taken as an obstacle with regards to the design and goals of many AIs and robots even if they would be interacting with people on a daily basis. Some questions still need to be answered before the valley could be left for good. What stimuli are more relevant in human-human interaction? Aren't contemporary humans already used to computerized interactions? If so, is it enough to overcome the valley and make interactions with robots comfortable? Shouldn't a holistic approach, as AI-psychology, be introduced into AI to deal with similar problems?

References

1. Mori, M.: Bukimi no tani. Energy 7(4), 33–35 (1970)
2. Mori, M., MacDorman, K.F. da Kageki, N.: The uncanny valley [from the field]. Robotics & Automation Magazine 19(2), 98–100 (2012)
3. Ho, C., MacDorman, K.: Revisiting the uncanny valley theory: Developing and validating an alternative to the godspeed indices. Computers in Human Behavior 26(6), 1508–1518 (2010)
4. Hanson, D., Olney, A., Prilliman, S., Mathews, E., Zielke, M., Hammons, D., Fernandez, R., Stephanou, H.: Upending the uncanny valley. In: Proceedings of the National Conference on Artificial Intelligence. Volume 40(4). AAAI Press, MIT Press (2005)
5. Cleveland, A., Kobiella, A., Striano, T.: Intention or expression? four-month-olds reactions to a sudden still-face. Infant Behavior and Development 29(3), 299–307 (2006)
6. Hadjikhani, N., Kveraga, K., Naik, P., Ahlfors, S.: Early (n170) activation of face-specific cortex by face-like objects. Neuroreport 20(4) (2009)
7. McCloud, S.: Understanding comics: The invisible art. Harper Paperbacks (1993)
8. Weizenbaum, J.: Eliza—a computer program for the study of natural language communication between man and machine. Communications of the ACM 9(1), 36–45 (1966)