Experimental Study Toward Modeling of the Uncanny Valley Based on Eye Movements on Human/Non-human Faces

Yoshimasa Tawatsuji[1], Kazuaki Kojima[2], and Tatsunori Matsui[2]

[1] Graduate School of Human Sciences, Waseda University
[2] Faculty of Human Sciences, Waseda University
2-579-15, Mikajima Tokorozawa Saitama, 359-1192, Japan
wats-kkoreverfay@akane.waseda.jp, koj@aoni.waseda.jp,
matsui-t@waseda.jp

Abstract. In the research field of human-agent interaction, it is a crucial issue to clarify the effect of agent appearances on human impressions. The uncanny valley is one crucial topic. We hypothesize that people can perceive a human-like agent as human at an earlier stage in interaction even if they finally notice it as non-human and such contradictory perceptions are related to the uncanny valley. We conducted an experiment where participants were asked to judge whether faces presented on a PC monitor were human or not. The faces were a doll, a CG-modeled human image fairly similar to real human, an android robot, another image highly similar and a person. Eyes of the participants were recorded during watching the faces and changes in observing the faces were studied. The results indicate that eye data did not initially differ between the person and CG fairly similar, whereas differences emerged after several seconds. We then proposed a model of the uncanny valley based on dual pathway of emotion.

Keywords: The uncanny valley, eye movements, dual pathway of emotion, humanlike agent.

1 Introduction

In the research field of human agent interaction, it is a crucial issue to clarify the effect of appearance of a robot in the real world or a virtual character on a PC monitor on human impressions toward it [1]. Here, we refer to such a robot or character as an *agent*. *The uncanny valley*, which was coined by Mori [2], is one critical topic in considering appropriate appearances of agents in terms of impressions from people. It hypothesized that although human familiarity toward an agent increases as an appearance of the agent gets more humanlike, it drastically decreases to the bottom of the valley when the agent-appearance is almost the same to but slightly different from real human as illustrated in Fig.1. Several studies have approached to what causes the phenomenon hypothesized in the uncanny valley and how, even though its mental mechanism has not been yet clarified.

M. Kurosu (Ed.): Human-Computer Interaction, Part IV, HCII 2013, LNCS 8007, pp. 398–407, 2013.

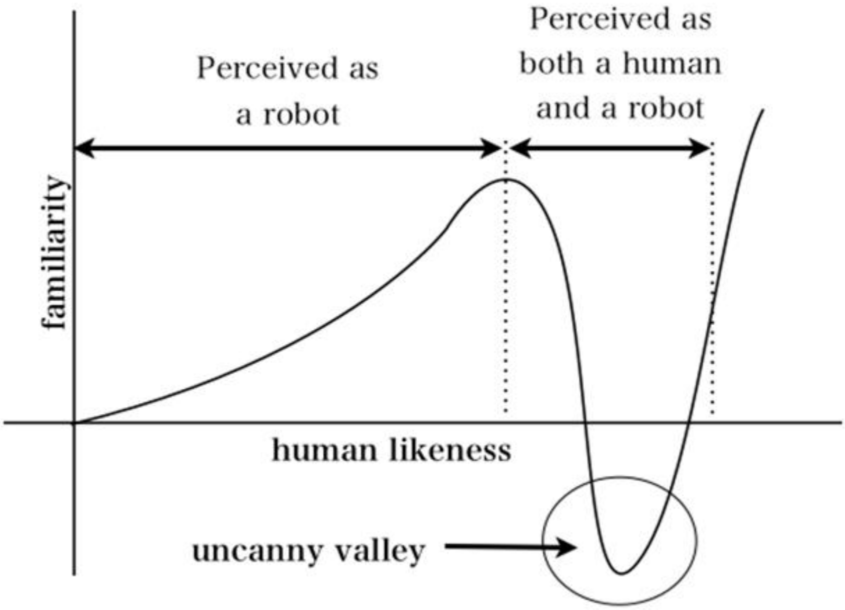

Fig. 1. Basic concept of the uncanny valley (partially altered from that in [2])

Some studies provided empirical data regarding how people observed a humanlike agent and how observation of an agent differed from that of a person. Noma et al. [3] revealed that about 75% of participants recognized a moving humanlike robot as human when observing it in short time, whereas they noticed that it was not human as time passed. These facts must imply that people can perceive a humanlike agent as human in a short interaction and that people can simultaneously recognize such an agent as both human and non-human. Minato et al. [4] proved that perceptual responses to a human and a humanlike agent were different through an experiment where participants had a conversation with a person or an android highly similar to human. Behaviors of the participants were observed by recording eye fixations. They reported that the proportion of eye fixations on the person's face was smaller than that on the android's face. This result indicates that perceptual processes of human and non-human differ from each other. From these studies, we hypothesized that people can perceive an agent as human at an earlier stage in interaction even if they finally notice it as non-human and such contradictory perception are related to the uncanny valley. According to findings by Minato et al., changes of perceptions when facing an agent can be verified by obtaining data of eyes. Our hypothesis predicted that eye movements of those who observed a person or an agent extremely similar to real human did not differ initially, whereas differences between them emerged after observing for a while.

This study experimentally investigated changes in human perceptual processes of a person and humanlike agents. We conducted an experiment where participants judged whether each of the person or agents was human or non-human. Their perceptual processes were studied by recording and analyzing their eye movements. We then discussed perceptual processes of humanlike agents based on the experimental results and modeled the processes to provide explanation regarding how the uncanny valley occurs.

2 Experiment

2.1 Method

This experiment used five pictures of faces of (a) a doll, (b) a CG-modeled human image fairly similar to real human, (c) an android robot, (d) another image highly similar, and (e) a person as shown in Fig. 2. These faces were selected from several web pages to present faces whose similarities to real human got gradually higher.

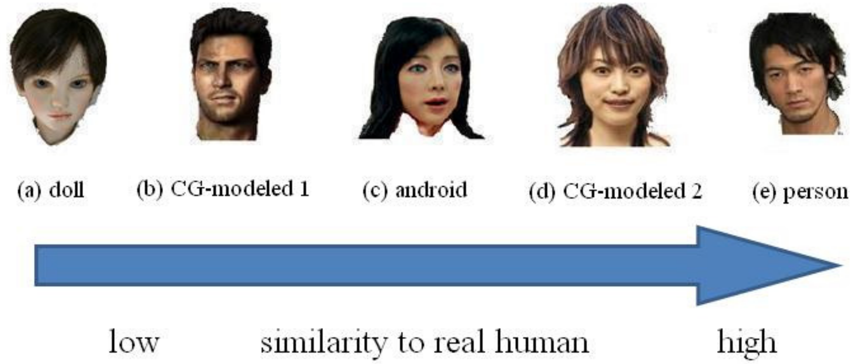

Fig. 2. Five faces used in experiment

In our experiment, participants were asked to judge whether each of the faces presented on a PC monitor was human or not. Each face was located at the center of the monitor. To control the initial location of eye of the participants, a white page where a cross was depicted at its central point was inserted before presenting each face. Eyes of the participants were recorded during watching the faces and eye fixations on the faces were estimated with EMR-AT VOXER produced by nac Image Technology. The participants were told that each face was presented for one minute and asked to write their judgments on a paper sheet. The faces were presented in the order of the doll, CG-modeled 2, android, person, and CG-modeled 1.

Some of the participants were asked to respond to two questionnaires regarding the faces after the judging task. The questionnaires included *(Q1) how difficult was the judgments of each face?* and *(Q2) which parts of faces did you pay attention to when judging?* The participants responded to Q1 on a three-point scale where 1 denoted easy and 3 denoted difficult.

2.2 Data Analysis

According to Yarbus [5], people frequently gaze at a region including the eyes, nose and mouth during watching at a person's face. These facial areas are important for human to seize some social information about others. Thus, we calculated a length of time when each area was gazed at for each face.

We used dFactory, analysis software for eye movement data, to calculate how long the participants gazed at each facial area. The calculation was conducted in three steps. First, the screen of the monitor was divided into 16 x 16 small blocks. Second, areas denoting the right eye, left eye, nose and mouth were defined. Each area comprised a block of the respective face part and its surrounding blocks. For example, the right-eye area comprised a block including the center of the pupil of the right eye and eight blocks surrounding the pupil block. Fig. 3 indicates the four areas in case of the CG-modeled 1. Finally, total time length of eye fixations on each area was calculated by adding time of eye fixations on each of comprised blocks. To confirm changes in perceptual processes, the analysis of eye-fixation time was performed in three time spans: 5 seconds, 10 seconds and 30 seconds from the start of face presentation.

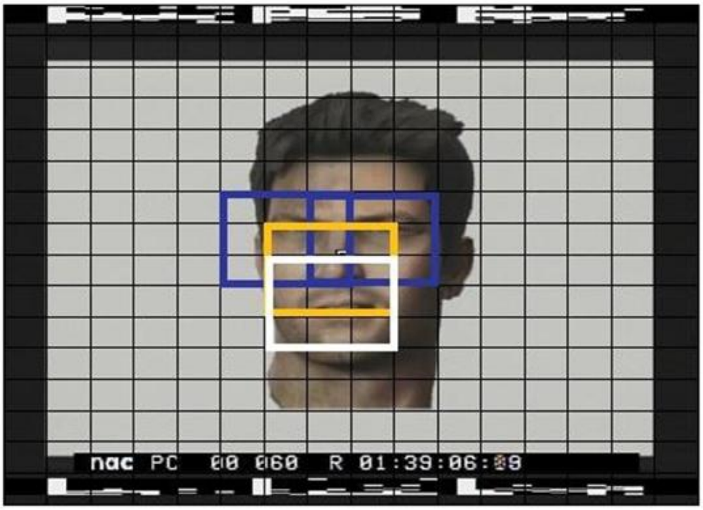

Fig. 3. Areas of right eye, left eye, nose and mouth

3 Results

Twenty one undergraduates (18 males and 3 females) participated in the experiment.

3.1 Judgment of Human/Non-human to Each Face

The proportions of participants who judged each face as human were as 28.6% for the doll, 19.1% for the CG-modeled 1, 19.1% for the android, 90.5% for the CG-modeled

2 and 100.0% for the person. Fig.4 indicates the proportions of judgments. Those results were mostly corresponding to our assumption of the similarities to real human. The android and CG-modeled 1 can be considered as the most unsimilar. Although the doll was more similar than the two, it was also evaluated as less humanlike. The CG-modeled 2 was the most humanlike, and the person was correctly judged as human. To precisely study differences in observing human and non-human faces, we analyzed eye data of the CG-modeled 1 which was mostly judged as non-human, that of the CG-modeled 2 which was judged as human although it was actually non-human and that of the person.

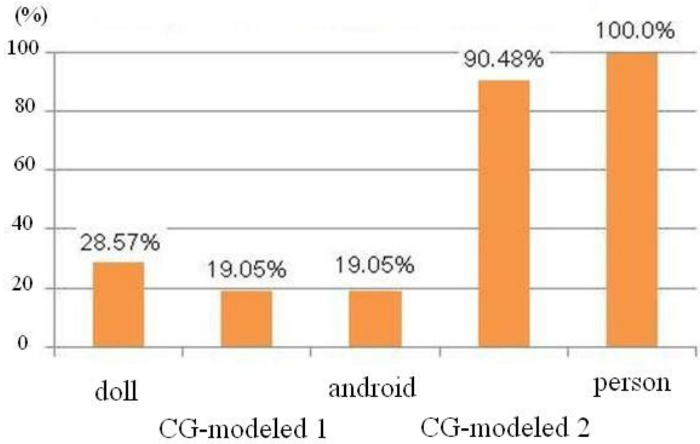

Fig. 4. Proportions of participants who judged each face was human

Fig. 5. Transactions of eye fixations of a participant in initial five seconds

3.2 Time of Eye Fixations on Areas of Each Face

Gaze data of 15 participants who judged the CG-modeled 2 as human and the CG-modeled 1 as non-human was analyzed. However, data of 7 participants was excluded due to its poor quality. Thus, data of the other 8 was actually used. Fig.5 shows examples of transactions of eye fixations during observing each face in the initial five seconds. The size of each circle denotes the length of total time of eye fixations at the respective point.

Table 1 indicates averages of time length of eye fixations on the four areas of each face in each time span. The t-test revealed significant differences of time length of eye fixations on the right eye areas among the three faces. Fig.6 shows average time of eye fixations on the right eye area of each face in each time span. In initial 5 seconds, eye fixations on the right eye area of the CG-modeled 2 was significantly longer than that of the CG-modeled 1 (p<.01) and that of the person (p<.01). In initial 10 seconds, eye fixations on the right eye area of the CG-modeled 2 was significantly longer than that of the CG-modeled 1 (p<.01) and that of the person (p<.01), and the difference between the CG-modeled 1 and person was moderately significant (p<.10). In entire 30 seconds, the participants observed the right eye area of the CG-modeled 2 more frequently than that of the person (p<.01), and the difference between the CG-modeled 1 and person was moderately significant (p<.10).

Table 1. Average of time length of eye fixations on left eye area, right eye area, nose area and mouth area of each face (sec.)

5 seconds	CG-modeled 2	CG-modeled 1	Human
On left eye	2.496	2.580	2.732
On right eye	2.995	1.873	1.729
On nose	3.735	2.847	2.698
On mouth	1.414	1.097	0.980

10 seconds	CG-modeled 2	CG-modeled 1	Human
On left eye	4.300	4.893	6.315
On right eye	5.544	3.385	2.529
On nose	6.676	5.622	5.812
On mouth	2.340	2.605	1.593

30 seconds	CG-modeled 2	CG-modeled 1	Human
On left eye	11.858	14.598	13.164
On right eye	14.682	12.356	8.680
On nose	18.506	16.223	14.333
On mouth	7.257	5.985	5.385

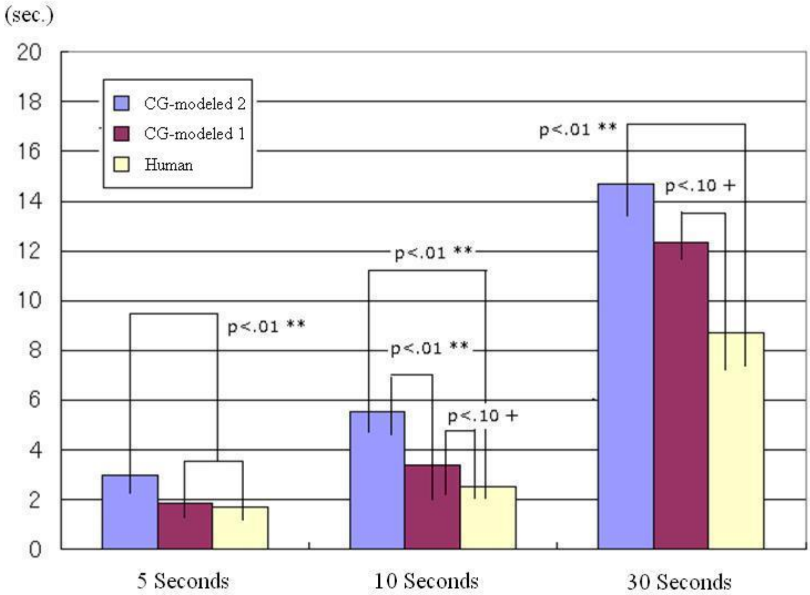

Fig. 6. Averages of time length of eye fixations on the right eyes of each face

3.3 The Questionnaires

Twelve participants (9 males and 3 females) responded to the questionnaires. Fig. 7 shows average scores of responses to Q1. The averages were significantly different between in the CG-modeled 1 and CG-modeled 2 (p<.05), and between in the

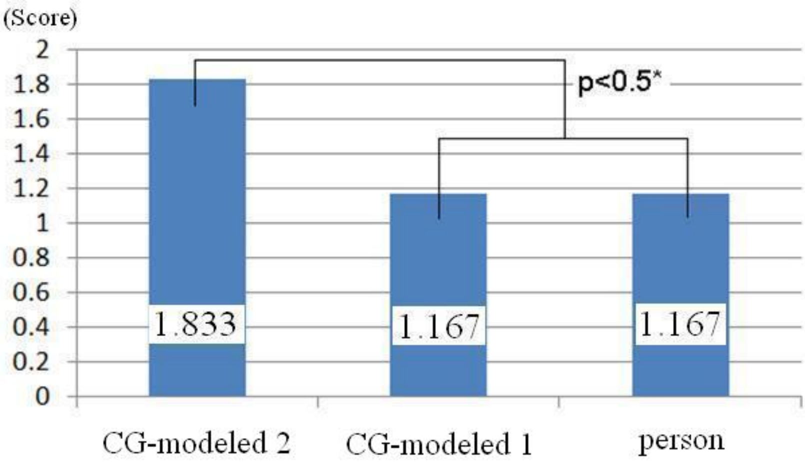

Fig. 7. Average scores of responses to Q1

CG-modeled 2 and person (p<.05). According to responses to Q2, the proportion of participants who described the skin and its texture was 66.7% and that of the eyes was 25.0% in the CG-modeled 1. In the CG-modeled 2, the proportion of the eyes was 50.0%, that of the mouth was 41.7% and that of the facial expression was 16.7% and the eyes was 25.0%. In the person, the proportion of the eyes was 16.7%, that of the beard was 16.7% and that of the hair was 16.7%. Moreover, the participants responded positively and negatively to CG-modeled 1 and CG-modeled 2. Most of them felt humanlike in the facial expression of CG-modeled 1 but strange in its skin and recognized easily that it was not human but CG-modeled. On the other hand, they felt humanlike in the shadows under the eyes of CG-modeled 2 but unnatural in its facial expression.

4 Discussion

4.1 Perceptual Processes in Two Steps during Judgment

The participants judged the CG-modeled 1 as non-human and the CG-modeled 2 and person as human. Time length of eye fixations on the right eye area of the CG-modeled 2 was significantly longer than that of the person in every time span. According to the responses to Q1, they found more difficult when judging the CG-modeled 2 than when the person and CG-modeled 1. Yarbus [5] mentioned that people pay much attention to unnatural or unfamiliar elements of pictures. Thus, the participants must have found the right eye of the CG-modeled 2 unnatural after observing for several seconds, and that caused their eyes to fix on it. These results are consistent with the report by Minato et al. [4].

In case of the CG-modeled 1, it was remarkable that there was no significant difference of time length of eye fixations on the right eye area from that of the person in initial 5 seconds, whereas the difference emerged as time passed. Thus, these results must imply that the participants had once perceived the CG-modeled 1 as human in the short-time observation. The emergence of the difference of time length of eye fixations must have been brought by shift of the participants' attention to differences of the CG-modeled 1 from a real human. Although the participants reported that the CG-modeled 1 was non-human, it can be assumed that people initially perceive a humanlike agent fairly similar to real human as human and then perceive it as non-human. Thus, perceptual processes of a humanlike agent can be considered to include the two steps.

4.2 The Dual-Pathway of Emotion

LeDoux [6] proposed that human brain functions and neural mechanisms of processes have two pathways of high and low roads and both of them play important roles in responding to dangerous stimuli from the external world. A stimulus is processed at first by the thalamus, and then the low road crudely leads information of the stimulus directly to the amygdala, which enables immediate response to the stimulus. On the other hand, the high road is another pathway to the amygdala via the celebral cortex, which simultaneously processes the stimulus carefully in detail. This dual pathway, especially response by the low road, is an essential function to avoid a danger.

The results in our experiment can be explained along with the concept of the dual pathway of emotion. When human observes a humanlike agent, the low road transmits information roughly but immediately so that he/she perceives the agent as human. Consequently, the observer initially responds to the agent as if it is human. On the other hand, the high road provides more elaborate information of the agent and that allows the observer to realize that the agent is actually non-human. These parallel functions by the dual pathway form two different perceptual processes.

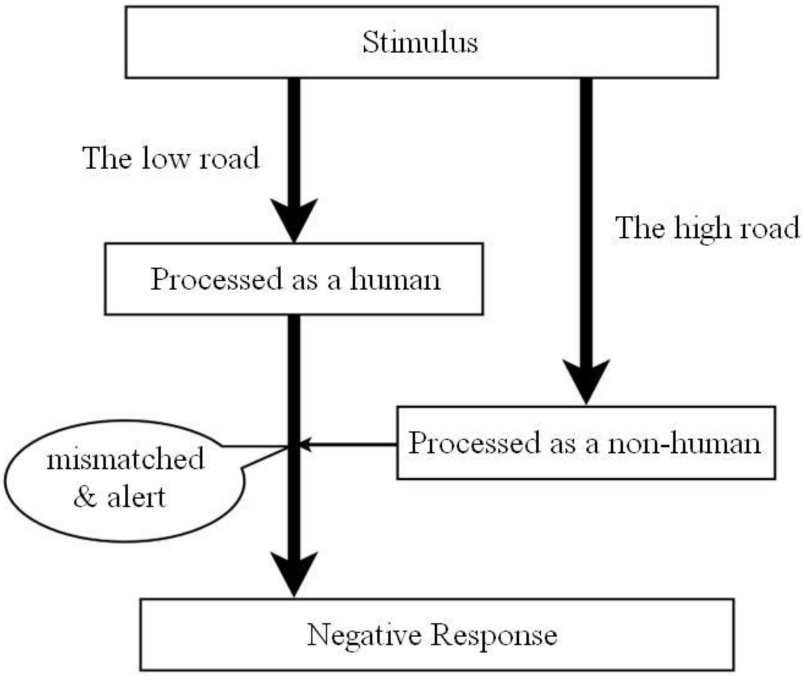

Fig. 8. Conceptual model of human perception toward a humanlike agent

This model can provide an explanation of the mental mechanism behind the uncanny valley from the viewpoint of the function to avoid dangers. For human, perceiving a humanlike agent as a human by the low road means a mistake of a different species for an individual akin to them. After the high load transmits the information, he or she receives an alert to the agent to avoid a danger which may be caused by the mistake. Accordingly, two pieces of information between the low and high road are mismatched, and the alert makes human generate negative response to the agent. Fig.8 illustrates this conceptual model.

5 Conclusion

This study experimentally investigated changes in human perceptual processes of a person and humanlike agents based on the hypothesis that differences in the perceptions emerged after observing for several seconds. The results confirm that there was no significant difference between the person and CG-modeled 2 initially, whereas the difference emerged as time passed. According to the results indicating the two steps in perceptions of a humanlike agent, we modeled the uncanny valley based on the concept of the LeDoux's dual-pathway of emotion. The model insists that mismatch between pieces of information transmitted by the low and high roads can be served as a trigger of the uncanny valley.

As described above, the model explains what causes human negative response to a humanlike agent. However, it cannot explain evocation of the feeling of eeriness and the drastic decrease of familiarity documented by the uncanny valley. One important future work is hence to clarify how the mismatch can be connected with the feeling of eeriness, and to describe the more detailed model.

References

1. Kanda, T., Miyashita, T., Osada, T., Haikawa, Y., Ishiguro, H.: Analysis of humanoid appearances in human-robot interaction. Journal of the Robotics Society of Japan 24(4), 497–505 (2006)
2. Mori, M.: Uncanny Valley. Energy 7(4), 33–35 (1970)
3. Noma, M., Saiwaki, N., Itakura, S., Ishiguro, H.: Composition and Evaluation of the Humanlike Motions of an Android. In: Proc. Int. Conf. Humanoid Robots, pp. 163–168 (2006)
4. Minato, T., Shimada, M., Ishiguro, H., Itakura, S.: Development of an Android Robot for Studying Human-Robot Interaction. In: Proc. IEA/AIE. Conf. 2004, pp. 424–434 (2004)
5. Yabus, A.L.: Eye Movements and Vision. Prenum Press, New York (1967)
6. LeDoux, J.E.: The Emotional Brain: The Mysterious Underpinnings of Emotional Life. Simon & Schuster (1998)