Cross-Cultural Study on Facial Regions as Cues to Recognize Emotions of Virtual Agents

Tomoko Koda[1], Zsofia Ruttkay[2], Yuka Nakagawa[1], and Kyota Tabuchi[1]

[1] Faculty of Information Science and Technology, Osaka Institute of Technology
1-79-1 Kitayama, Hirakata, 573-0196 Osaka, Japan
koda@is.oit.ac.jp
[2] Moholy-Nagy University of Art and Design
Zugligeti ut 9-25, Budapest, Hungary
ruttkay@mome.hu

Abstract. This paper reports the preliminary results of a cross-cultural study on facial regions as cues to recognize the facial expressions of virtual agents. The experiment was conducted between Japan and Hungary using 18 facial expressions of cartoonish faces designed by Japanese. The results suggest the following: 1) cultural differences exist when using facial regions as cues to recognize cartoonish facial expressions between Hungary and Japan. Japanese weighed facial cues more heavily in the eye regions than Hungarians, who weighed facial cues more heavily in the mouth region than Japanese. 2) The mouth region is more effective for conveying the emotions of facial expressions than the eye region, regardless of country. Our findings can be used not only to derive design guidelines for virtual agent facial expressions when aiming at users of a single culture, but as adaptation strategies in applications with multicultural users.

Keywords: facial expression, virtual agents, character, cross-culture.

1 Introduction

Virtual agents are frequently used in virtual worlds, online applications and (serious) games. In the current state of development of the virtual agent technology, virtual agents can express emotions in their bodily behavior, first of all, by displaying facial expressions. The culture of the virtual agent is relevant in relation to the culture of the real human interlocutor. Moreover, there are training and simulation applications emerging where the cultural identity and particularly, the bodily behavior is the major learning component of the training application [1]. Such agents have been designed under the assumption that their expressions are interpreted universally among all cultures. The basis for this assumption is the early finding about the universality of the 6 basic expressions – joy, surprise, fear, sadness and disgust - by Ekman [2]. Later works by Ekman and colleagues indicated cultural differences in perceived intensity of emotions [3, 4]. However, recent research indicates cultural differences in recognizing human facial expressions. Elfenbein et. al. [5] have coined the term cultural dialects of facial expressions: the cultural dialect, unlike a personal

T. Ishida (Ed.): Culture and Computing, LNCS 6259, pp. 16–27, 2010.

idiosyncratic variant, is a well identifiable specific usage of some facial signal. Such cultural dialect can be seen as an in-group advantage in human facial expression recognition, whereby human facial expression recognition is generally more accurate for perceivers from the same cultural group as expressers. These works all have been using photographs, where the cultural identity of the face was clear. Ruttkay [6] discusses further empirical findings and provides a detailed scheme of the possible factors of cultural differences in interpreting facial emotions of virtual agents.

When designing virtual agents, it is a challenging and not very much exploited possibility to use non-realistic faces. There are two motivations for going for cartoon-like faces. On the one hand, it has been shown that the more realistic the design is, the more critical the human perceivers are. As the realism increases, the "uncanny valley" effect occurs [7, 8]. However, in most application contexts it is the "suspension of disbelief" which is to be achieved, not the full realism. Further on, when using non-realistic faces there are additional means of expressivity (exaggeration, usage of non-realistic features or additional signals). Hence it is interesting to study possible cultural variations in perception of cartoon-like faces. Do findings from psychology on interpreting realistic facial expressions carry over to cartoon-like faces? How do the drawing style and familiarity with non-realistic facial expressions (e.g. in the tradition of comics) influence the interpretations of cartoon-like facial expressions? Koda's cross-cultural study [9] on the recognition of cartoon-like agent facial expressions drawn by Asian and Western designers suggests that the recognition accuracy of facial expressions is higher for virtual agents designed by the same cultural group as the subjects. E.g. Japanese cartoon-like facial expressions are recognized most accurately by Japanese, and Western cartoon-like facial expressions are recognized most accurately by western countries. The results suggest an in-group advantage is also applicable to cartoon-like facial expressions of virtual agents.

Recent psychological study also investigates the cultural differences of facial expressions by focusing on the facial regions. Research on human eye movements to interpret photo realistic human facial expressions showed East Asian participants mostly focused on the eyes, and Western participants scanned the whole face [10]. Yuki et al. used pictograms and photorealistic human facial images and suggest that Americans tend to interpret emotions based on the mouth, while Japanese tend to focus on the eyes [11]. Yuki et al. state this cultural difference arises from cultural norms: that people in cultures where emotional subduction is the norm (such as Japan) would focus on the eyes, and those in cultures where overt emotional expression is the norm (such as U.S.) would focus on the mouth shape.

This study applies the findings of [10, 11] to animated cartoon-like virtual agent faces to improve the culturally effective facial expression design of virtual agents. Such findings can be used not only to derive design guidelines when aiming at users of a single culture, but as adaptation strategies in applications with multicultural users. E.g. in an ATM machine or on-line shop, if the user's cultural identity is established; the virtual agent's facial expressions may be fine-tuned for optimal recognition for the given culture. Animating virtual agents' facial expression is rather easily done, but much more difficult in case of physical robots. Recent social robots have cartoonish faces with limited facial expressions, e.g., Kismet [12], Nexi MDS

Robot [13], iCat [14] when their research focus is not on increasing realism of a humanoid robot such as geminoids [8]. We believe providing research results on the perception of cartoonish virtual agents' facial expressions is also meaningful in order to minimize the effort to develop social robot's facial expressions.

We investigated cultural differences in using the eye and mouth regions as cues to recognize the facial expressions of cartoonish agent faces between Hungary and Japan. We conducted a web-based survey to confirm the following hypothesis. In cartoonish faces, Japanese weigh facial cues in the eye regions more heavily than Hungarians, who weigh facial cues in the mouth region more heavily than Japanese. Section 2 describes the facial expression design and experiment design, section 3 describes the results, section 4 discusses the results, and section 5 concludes the research.

2 Experiment

2.1 Design of Facial Expressions

Two Japanese designers designed the three agent faces shown in Fig. 1. Each face design has neutral, happy, and sad expressions. The examples of the three original facial expressions are shown in Fig. 2. The face and facial expressions were created using CharToon [15], a design and animation tool for 2D cartoon faces. The facial expressions were designed by taking the emotional expressions displayed in Ekman's FACS (Facial Action Coding System) training material. The facial features were discussed in case of happy and sad expression [16].

Pre-evaluation of the original expressions was conducted by ten Japanese and eight Hungarians to validate that each expression conveyed the intended emotions of the designer. The static images of happy and sad expressions of the three face designs were shown randomly to the evaluators after showing the neutral expression in each session. They were asked to select the perceived emotion of the each expression (happy or sad expression) from the following four adjectives: happy, sad, surprised, fear. They wrote an adjective if they don't find appropriate adjectives from the four provided. As a result of the pre-evaluation, all the original face designs had higher than 90% recognition accuracy among Japanese, higher than 85% among Hungarians, although we used cartoon faces and facial expressions designed by Japanese. Thus, we can assume the happy and sad expressions correctly convey the intended emotions in both countries. We also asked the perceived age of each agent face (discussed in section 4).

We then created six static expressions per agent design by combining the eyes and mouths. The six combinations are: happy eyes and neutral mouth (HN), happy eyes and sad mouth (HS), neutral eyes and happy mouth (NH), neutral eyes and sad mouth (NS), sad eyes and happy mouth (SH), and sad eyes and neutral mouth (SN). The total number of combined facial expressions is 18 (six expressions x three agents). Fig. 3 shows the six combined expressions created from the original expressions in Fig. 2.

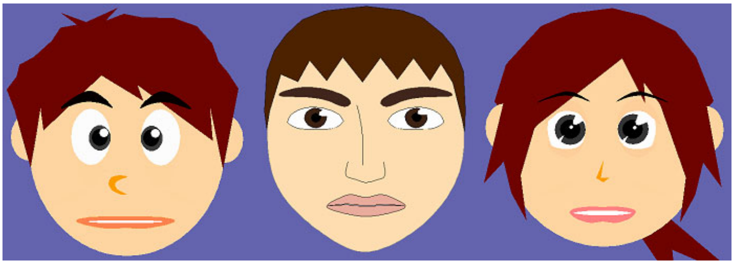


Fig. 1. Three agent designs created by two Japanese designers. Each design shows neutral expression. From left: Boy 1, Boy 2, Girl designs.

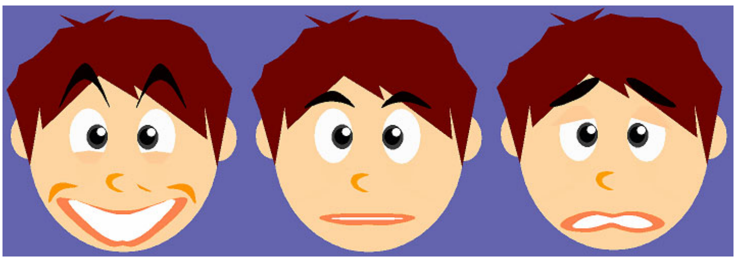


Fig. 2. Three original facial expressions of the Boy 1 design. From left: happy, neutral, and sad.

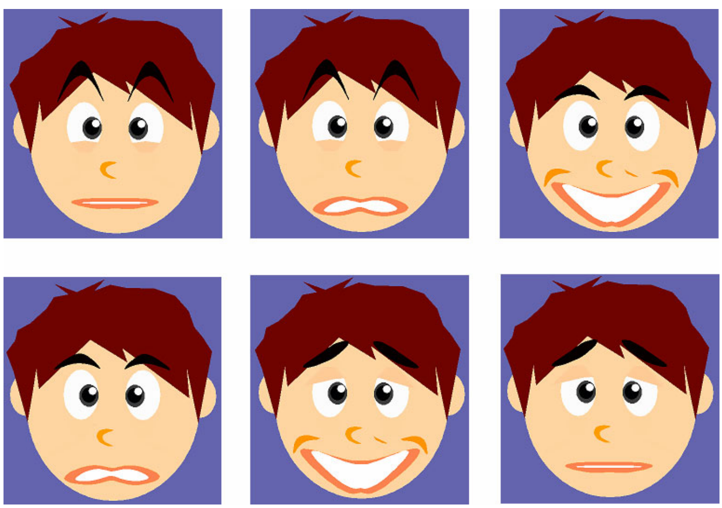


Fig. 3. Examples of six combined facial expressions created from images in Fig. 2. From top left: happy eyes and neutral mouth (HN), happy eyes and sad mouth (HS), neutral eyes and happy mouth (NH); from bottom left: neutral eyes and sad mouth (NS), sad eyes and happy mouth (SH), sad eyes and neutral mouth (SN).

Next, we created 18 animations that start from the neutral expressions of each agent design and end with one of the combined expressions by using CharToon's animation editor. Each animation lasted for four seconds; one second for a neutral expression, two seconds for the transition to a combined expression, and one second for the combined expression, and the animation ends with the combined expression. The animations were converted to the Adobe Flash movie format.

2.2 Experiment Procedure

The experiment was conducted on the web from December 2009 to January 2010. Subjects were invited to participate in the experiment by email. Invitations were made to gather participants with minimum exposure to and experiences with other cultures.

The experiment procedure was as follows:

1) Participants accessed the experiment web site. Instructions were made in Japanese for the Japanese participants and in English for the Hungarians.
2) Participants first read the instructions on the procedure. No explanation on the objective of the experiment was made.
3) They accept the terms and condition by clicking the accept button.
4) They answered a questionnaire about their demographics: name or nickname, age, gender, nationality, first language, and experiences of living abroad.
5) The 18 animations are randomly shown one by one. Subjects watched each animation by clicking the play button. They can replay the animation multiple times.
6) They evaluated the perceived emotion of the final expression of each animation using a 6-point Likert scale (6: very happy - 1: very sad). The instruction was made to follow their first impression upon rating.
7) The experiment continued until they finished all 18 animations. The estimated time to complete the experiment is approximately 5 minutes.
8) Participant answers (demographic information and ratings) were stored on a secured disk on the web server. Each answer was labeled with the name/nickname (which participants entered) and the timestamp when the participants clicked the accept button in order to differentiate participants with the same name or nickname.

We gathered 53 valid answers from Japan and 31 from Hungary. Valid participants resided in Hungary or Japan, spoke either Japanese or Hungarian as their first language, and had no experience or had lived abroad less than one year. The Hungarian participants are fluent in English. Incomplete answers that lack participants' demographic information or those did not finish evaluating the all 18 animations were not used for later analysis. The participant demographics of valid answers in Japan were 29 male and 24 female, whose average age is 17.9 years old (standard deviation (SD) 4.5 years). The demographics in Hungary were 15 male and 16 female, whose average age is 24.3 years old (SD=11.0 years).

Fig. 4 shows a screenshot of one of the animation and evaluation pages of the experiment.

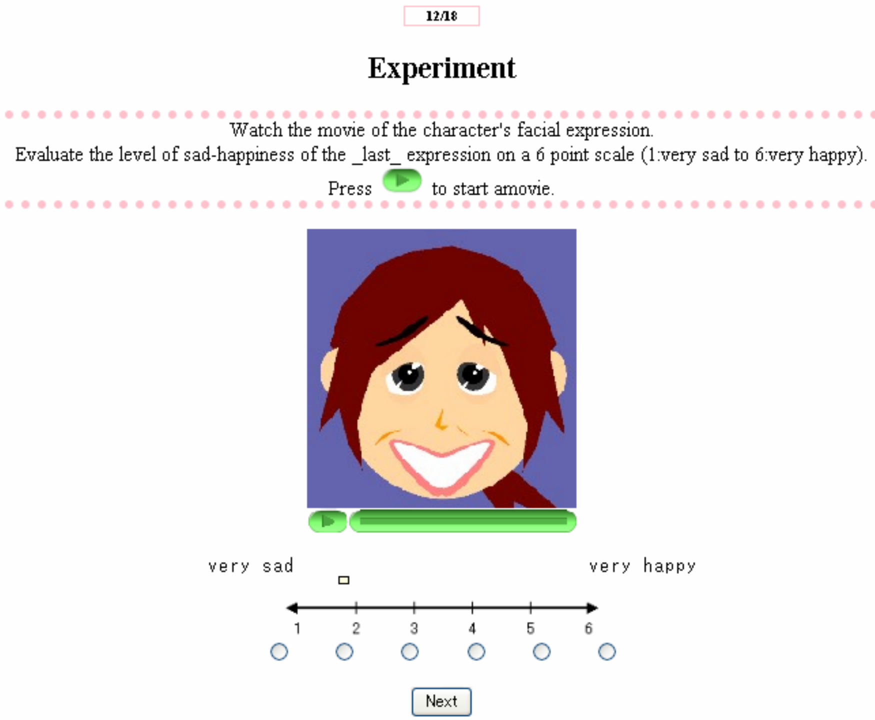


Fig. 4. Screenshot of web experiment. Flash movie shows animation from a neutral expression to one of 18 combined expressions shown randomly. Subjects rated perceived emotion of each combined expression on 6-point Likert scale.

3 Results

3.1 Analysis of Perceived Emotions by Facial Regions

The participant answers about the perceived emotions are shown in Fig. 5. By focusing on the shape of either the eye or mouth regions, the combined expressions were placed into four categories: happy eyes, sad eyes, happy mouth, and sad mouth. For example, the happy eyes category includes happy eyes and neutral mouth and happy eyes and sad mouth expressions.

First, we focused on the answers by facial categories regardless of the country. As shown in Fig. 5, the happy mouth category was rated the happiest and the sad mouth category was rated the saddest in all categories by both countries. The eye region (happy eyes/sad eyes) did not convey the intended emotions as effectively as the mouth region (happy mouth/sad mouth).

Second, we focused on the answers between Hungary and Japan. Japanese rated the happy eyes category as significantly happier than Hungarians regardless of the

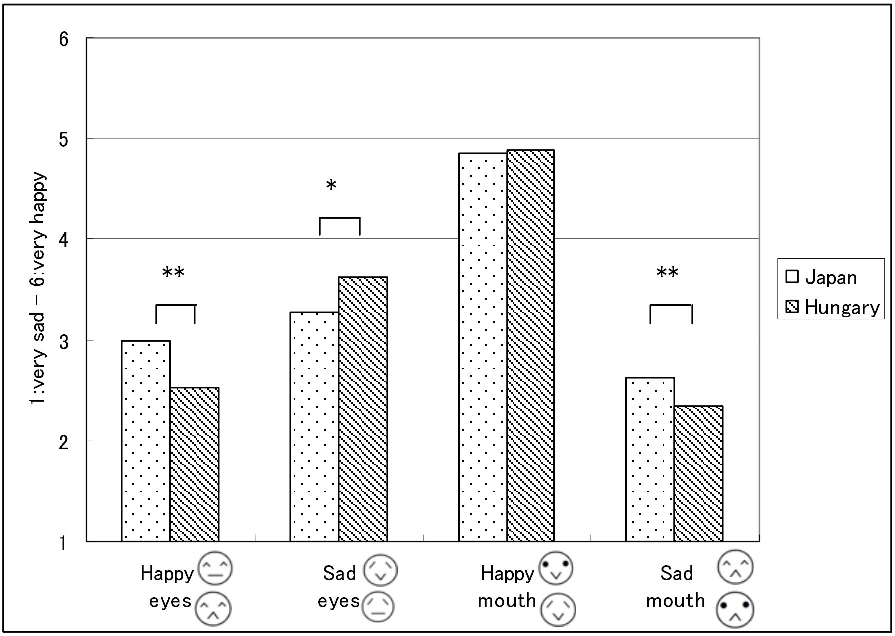


Fig. 5. Perceived emotions of Japan and Hungary categorized by facial regions. Scale indicates 1: very sad to 6: very happy. Number of subjects: Japan (n=53), Hungay (n=31). ** indicates p<0.01, * indicates p<0.05).

mouth's shape (Japan=3.00 (n=53), Hungary=2.52 (n=31), F=24.77, p<0.01) and the sad eyes category as sadder (Japan=3.27 (n=53), Hungary=3.63 (n=31), F=5.98, p<0.05). Hungarians rated the sad mouth category significantly sadder than Japanese regardless of the shape of the eyes (Japan=2.63 (n=53), Hungary=2.34 (n=31), F=14.59, p<0.01). There were no significant differences in the ratings of the happy mouth category (Japan=4.87 (n=53), Hungary=4.76 (n=31), F=0.13, not significant).

3.2 Analysis of Perceived Emotions by Combined Expressions

In this section, we examine the results by the six combined expressions for more detailed analysis. Fig. 6 shows the perceived emotions of the combined expressions. Analysis by Two-way ANOVA indicates significant cultural differences in the perceived expressions between Japan and Hungary in the HN (Japan=3.57 (n=53), Hungary=2.91 (n=31), F=27.31, p<0.01), SN (Japan=2.09 (n=53), Hungary=2.49 (n=31), F=10.74, p<0.01) and NS (Japan=2.82 (n=53), Hungary=2.56 (n=31), F=8.43, p<0.05) expressions. There were no other significances in the perceived emotions between Japan and Hungary.

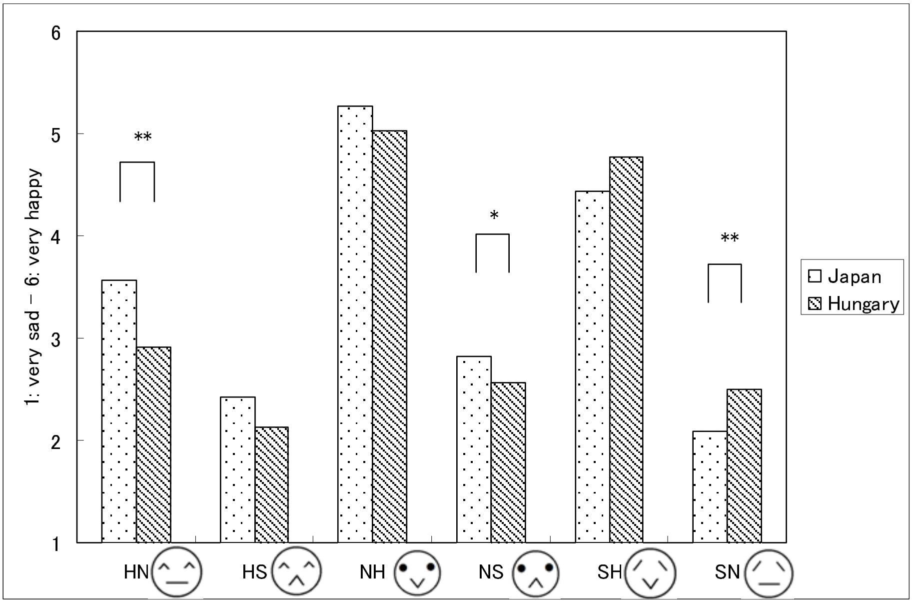


Fig. 6. Perceived emotions of six combined expressions in Japan and Hungary. Scale indicates 1: very sad to 6: very happy. Number of subjects: Japan (n=53) , Hungay (n=31). ** indicates p<0.01, * indicates p<0.05). Acronyms: HN: happy eyes and neutral mouth, HS: happy eyes and sad mouth, NH: neutral eyes and happy mouth, NS: neutral eyes and sad mouth, SH: sad eyes and happy mouth, SN: sad eyes and neutral mouth (SN).

3.3 Analysis of Perceived Emotions by the Shape of the Facial Regions

This section analyzes the changes of the perceived emotions within country by the shape of the facial regions. Table 1 shows the perceived emotions shown by the shape of the mouth region when the shape of the eye region is fixed. Table 2 shows the perceived emotions shown by the shape of the eye region when the shape of the mouth region is fixed.

The perceived emotions between HN and HS (mouth shape change when the eye region is fixed on happy eyes), and SN and SH (mouth shape change when the eye region is fixed on sad eyes) are significantly different in both countries (Table 1). This result again indicates that the shape of the mouth effectively displays the emotions of cartoon faces both in Japan and Hungary.

The emotions perceived by the shape of the eyes are not as consistent as those by the shape of the mouth (Table 2). The emotions perceived by Japanese differ significantly between NH and SH (eye shape change when the mouth region is fixed on a happy mouth), but the differences of emotions perceived by Hungarians are not significant. Contrary to designer intentions, Hungarians rated the perceived emotions of NS and HS (eye shape change when the mouth region is fixed on sad mouth) significantly differently, but the differences in emotions perceived by Japanese are not significant.

Table 1. Perceived emotions by shape of mouth region

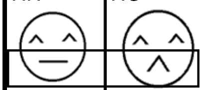
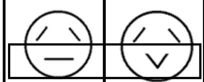
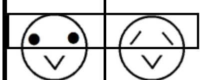
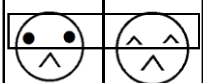
Country	Design	Happy Eyes			Sad Eyes		
		HN 	HS	p<0.01	SN 	SH	p<0.01
Japan	boy 1	3.87	2.53	**	1.79	4.43	**
	boy 2	3.17	2.4	**	3.02	4.94	**
	girl	3.66	2.36	**	1.47	3.94	**
Hungary	boy 1	3.06	2.25	**	2.35	4.57	**
	boy 2	2.55	2.16	**	2.87	4.77	**
	girl	3.13	1.87	**	2.76	4.74	**

Table 2. Perceived emotions by shape of eye region

Country	Design	Happy mouth			Sad mouth		
		NH 	SH	p<0.01	NS 	HS	p<0.01
Japan	boy 1	5.47	4.43	**	3	2.53	
	boy 2	4.75	4.94		2.75	2.4	
	girl	5.58	3.94	**	2.72	2.36	
Hungary	boy 1	5.32	4.77		2.48	2.35	
	boy 2	4.32	4.77		2.58	2.16	**(opposite)
	girl	5.42	4.74		2.61	1.87	**(opposite)

4 Discussion

First, we examined the results by categorizing the facial regions. The results from 3.1 suggest cultural differences in the perceived emotions in the happy eyes, sad eyes, and sad mouth categories in the expected direction. The highest happiness ratings in the happy mouth category in both countries suggest that the mouth region more effectively conveys the emotions of the facial expressions than the eye region. The mouth's effectiveness is understandable since the mouth is the most expressive part of the face, since it is evolved as a primary means of verbal communication [17, 18] while the eyes are more difficult to control than the mouth when people express emotions.

Second, we examined the results by the combined expressions. The results from 3.2 again indicated cultural differences in the expected direction in the hypothesis.

Japanese rated the perceived emotions of HN (happy eyes and neutral mouth) significantly happier than Hungarians and SN (sad eyes and neutral mouth) as significantly sadder than Hungarians. Hungarians rated the perceived emotions of NS (neutral eyes and sad mouth) significantly sadder than Japanese. This means the expressions where either eyes or mouth is/are neutral confirmed the hypothesis, except NH (neutral eyes and happy mouth). However, the contradictory expressions (HS, SH), where the eye expressions and mouth expressions show opposite expressions, fail to show significant cultural differences.

The effectiveness of the mouth region is shown again in NH and SH (both with happy mouths). Both countries' perceived emotions of NH have the happiest scores. SH's score are higher than HN.

Third, we focused on the differences in ratings by the shape of the eye/mouth when the other facial region was fixed. Both countries responded to the differences in the mouth shape as significantly and dynamically as expected. However, Japanese responded to the differences of the eye shape more significantly and dynamically than Hungarians, implying that Japanese focused more strongly on the eyes even when the mouth region has the most convincing happy expression; Hungarians continued to focus on the mouth region.

However, the differences in the eye shape with a sad mouth did not result in the expected direction. Perhaps the reason reflects the nature of the effectiveness of the mouth region in facial expression recognition and eye designs. Some subjects commented that the happy eye designs resembled surprised expressions when isolated from the mouth region. Although the perceived emotions of the happy eyes using static images of the eye region in the pre-evaluation test had 90% in Japan and 85% accuracy in Hungary, the result implies we must design the eyes more carefully when they are used solely apart from other facial regions.

Former studies [10, 11] used emoticons and photorealistic human facial expressions. Emoticons can convey emotions simply and effectively because we use them in our daily lives. Since the photorealistic human facial expressions used in these studies were made by professional actors, they naturally and effectively conveyed the intended emotions. However, cartoonish facial designs have wide variations caused by the designers' drawing styles and techniques. Although we used CharToon as a design tool to minimize possible variations, the agent faces in the experiment had more room for improvement.

Another consideration should be put on cultural diversity in designing cartoon faces. Koda reports in [9] that the recognition accuracy of facial expressions is higher for virtual agents designed by the same cultural group as the subjects. The facial expressions used in the experiment were designed by Japanese, thus the drawings followed Japanese stylistic conventions for expressing emotions, and Japanese have more exposure and experience in such drawings. Although the pre-evaluation of facial expressions did not show significant cultural differences, we cannot exclude the possibility that cultural differences in recognizing expressions in cartoon faces might be a result of cultural differences in drawing cartoon faces. We had an interesting result from the pre-evaluation of the agent faces. Hungarians' perceived age of the agent faces used in the experiment are much younger than the Japanese ones. Hungarian perceived the age of the agent faces as in their teens or younger, and Japanese as in their twenties.

In order to investigate the cultural differences in recognizing cartoon facial expressions in both ways, we plan to conduct a subsequent experiment with Hungarian designed cartoon faces. We expect to have different drawing styles of virtual agent faces, since Hungary has not been exposed to Japanese comic/anime culture compared to other European/American countries. Evaluation from other countries than Hungary and Japan is also needed to eliminate in-group advantage within a country or a culture.

5 Conclusion

This paper applies the psychological findings on cultural differences on facial regions used as cues to recognize human facial expressions to the cases of animated virtual agent faces. Our preliminary results support the hypothesis. There are cultural differences when using facial regions as cues to recognize cartoonish facial expressions between Hungary and Japan. Japanese weighed facial cues in the eye regions more heavily than Hungarians, who weighed facial cues in the mouth region more heavily than Japanese. We also confirmed that, regardless of the country, the mouth region more effectively conveys the emotions of facial expressions than the eye region.

We believe the results can be used not only to derive facial expression design guidelines of virtual agents when aiming at users of a single culture, but as adaptation strategies in applications with multicultural users. The virtual agent's facial expressions may be fine-tuned for optimal recognition for the given culture. The results also can be applied to designing a physical robot facial expressions to minimize mechanical movements to create its facial expressions according to the culture the robot serves.

Acknowledgements. This research is supported by a Grant-in-Aid for Scientific Research (C) 20500196 (2008-2010) from the Japan Society for the Promotion of Science. CharToon was developed by Zsofia Ruttkay and Han Noot.

References

1. Rickel, J., Marsella, S., Gratch, J., Hill, R., Traum, D., Swartout, B.: Towards a New Generation of Virtual Humans for Interactive Experiences. IEEE Intelligent Systems, 3–38 (July/August 2002)
2. Ekman, P., Friesen, W.V.: Constants Across Cultures in The Face And Emotion. Journal of Personality and Social Psychology 17, 124–129 (1971)
3. Ekman, P., Friesen, W.V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W.A., Pitcairn, T., Ricci-Bitti, P.E., Scherer, K.R., Tomita, M., Tzavaras, A.: Universals and Cultural Differences in The Judgment of Facial Expressions of Emotion. Journal of Personality and Social Psychology 53, 712–717 (1987)
4. Matsumoto, D., Ekman, P.: American-Japanese Cultural Differences in Intensity Ratings. Motivation and Emotion 13, 143–157 (1989)

5. Elfenbein, H., Beaupré, M., Leveque, M., Hess, U.: Toward a Dialect Theory: Cultural Differences in Expressing and Recognizing Facial Expressions. Emotion 7, 131–146 (2007)
6. Ruttkay, Z.: Cultural Dialects of Real and Synthetic Emotional Facial Expressions. Journal of AI & Society 24(3), 307–315 (2009)
7. Mori, M.: Bukimi no Tani: The Uncanny Valley (Translated by MacDorman, K.F., Minato, T.). Energy 7(4), 33–35 (1970)
8. Ishiguro, H.: Android Science: Toward a New Cross-disciplinary Framework. In: CogSci 2005 Workshop: Toward Social Mechanisms of Android Science, pp. 1–6 (2005)
9. Koda, T., Ishida, T., Rehm, M., Andre, E.: Avatar Culture: Cross-cultural Evaluations of Avatar Facial Expressions. Journal of AI & Society 24(3), 237–250 (2009)
10. Jack, R.E., Blais, C., Scheepers, C., Schyns, P.G., Caldara, R.: Cultural Confusions Show Facial Expressions are Not Universal. Current Biology 19(18), 1543–1548 (2009)
11. Yuki, M., Maddux, W.W., Masuda, T.: Are the Windows to the Soul the Same in the East and West? Journal of Experimental Social Psychology 43, 30–311 (2007)
12. Breazeal, C.: Emotion and Sociable Humanoid Robots. International Journal of Human-Computer Studies 59(1-2), 11–155 (2003)
13. Breazeal, C.: Sociable Robots. Journal of the Robotics Society of Japan 24(5), 591–593 (2007)
14. Van Breemen, A.J.N.: ICat: Experimenting with Animabotics. In: AISB Creative Robotics Symposium (2005)
15. Ruttkay, Z., Noot, H.: Animated CharToon Faces. In: International Symposium on Non-photorealistic Animation and Rendering, pp. 91–100 (2000)
16. Ekman, P., Friesen, W.V.: Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto (1978)
17. Ekman, P., Friesen, W.V.: Unmasking the Face: A Guide to Recognizing Emotions from Facial Cues. Prentice Hall, Englewood Cliffs (1975)
18. Fridlund, A.J.: Human Facial Expression: An Evolutionary Review. Academic Press, San Diego (1994)