

Avatar culture: cross-cultural evaluations of avatar facial expressions

Tomoko Koda · Toru Ishida · Matthias Rehm · Elisabeth André

Received: 14 October 2008 / Accepted: 27 May 2009 / Published online: 1 July 2009
© Springer-Verlag London Limited 2009

Abstract Avatars are increasingly used to express our emotions in our online communications. Such avatars are used based on the assumption that avatar expressions are interpreted universally among all cultures. This paper investigated cross-cultural evaluations of avatar expressions designed by Japanese and Western designers. The goals of the study were: (1) to investigate cultural differences in avatar expression evaluation and apply findings from psychological studies of human facial expression recognition, (2) to identify expressions and design features that cause cultural differences in avatar facial expression interpretation. The results of our study confirmed that (1) there are cultural differences in interpreting avatars' facial expressions, and the psychological theory that suggests physical proximity affects facial expression recognition accuracy is also applicable to avatar facial expressions, (2) positive expressions have wider cultural variance in interpretation than negative ones, (3) use of gestures and gesture

marks may sometimes cause counter-effects in recognizing avatar facial expressions.

1 Introduction

Since instant messenger and chat services are frequently used in our daily communication beyond nationality and languages, emoticons and expressive avatars are widely used to provide nonverbal cues to text-only messages (Kurlander and Salesin 1996). Recent growth of virtual worlds such as Second Life has attracted worldwide attention to avatar-mediated communication both from entertainment and businesses. Studies on emoticons and avatars report positive effects on computer-mediated communication. These studies indicate that emoticons and avatars improve user experiences and interactions among participants (Damer 1997; Smith et al. 2000; Pesson 2003) and build enthusiasm toward participation and friendliness in intercultural communication (Koda 2004; Isbister et al. 2000).

However, these avatars are used based on an implicit assumption that avatar expressions are interpreted universally across cultures. Since avatars work as graphical representations of our underlying emotions in online communication, those expressions should be carefully designed so that they are recognized universally. We need to closely examine cultural differences in the interpretation of expressive avatars to avoid misunderstandings in using them.

Research on examining cultural aspects of virtual embodied agents has recently started. Ruttkay addresses the importance of designing facial expressions for virtual agents with a specific culture (Ruttkay 2008), and Rehm

T. Koda (✉)
Faculty of Information Science and Technology,
Osaka Institute of Technology, Osaka, Japan
e-mail: koda@is.oit.ac.jp

T. Ishida
Department of Social Informatics,
Kyoto University, Kyoto, Japan
e-mail: ishida@i.kyoto-u.ac.jp

M. Rehm · E. André
Multimedia Concepts and their Applications,
University of Augsburg, Augsburg, Germany
e-mail: rehm@informatik.uni-augsburg.de

E. André
e-mail: andre@informatik.uni-augsburg.de

et al. integrate culture as a computational parameter for modeling multimodal interactions with virtual agents (Rehm et al. 2008).

However, few studies have compared the cultural differences in interpreting avatars' facial expressions. In one of these studies, Bartneck et al. compared Dutch and Japanese interpretations of avatars' animated gestures designed by a Japanese artist (Bartneck et al. 2004). Their results showed that there were cultural differences between Japanese and Dutch in the valences they perceived in animated characters. Japanese women perceived stronger emotions in some animated gestures of an avatar, i.e., bowing, than the Dutch subjects, although there were no overall differences in the interpretation of the presented gestures. In a previous study, we conducted a cross-cultural experiment in the form of a series of discussions between Japanese and Chinese users on a multilingual BBS that had expressive avatars designed by Japanese artists (Koda 2004). The results showed some facial expressions used in the experiment were interpreted completely differently and used for different purposes between Chinese and Japanese. The "misinterpreted" expressions were "sweat-on-the-face," "wide-eyed," and "closed-eyes." For example, the "wide-eyed" expression was interpreted as "surprised" by the Japanese subjects, while the Chinese subjects interpreted it as "intelligent" and used it when presenting a novel idea or asking questions. We observed that the Japanese subjects tried to confirm the meaning of the Chinese subject's message with the "wide-eyed" expression. This is one example of communication gaps caused by different interpretations of avatar expressions between the two countries.

The above two studies were each conducted between only two countries and used avatars designed by Japanese artists. To investigate cultural differences in avatar expression interpretation and understand what kinds of expressions are interpreted universally and what kinds are not, we need to conduct evaluations among several countries using avatars designed by artists from various countries. We believe the results would serve as a design guideline for universal avatar expression that would prevent miscommunication.

To examine cultural differences in avatar facial expression recognition, we apply findings from psychological studies on human facial expressions, since there have been a much wider variety of studies in psychology on human expressions than on avatar expressions. The most widely accepted findings come from the work of Ekman. He states that seven emotions, namely, anger, fear, disgust, surprise, sadness, happiness and contempt, are universally expressed by all cultures (Ekman 2003). However, he also argues that the implications and conno-

tations of these facial expressions are culturally dependent, and the degree to which showing or perceiving these expressions is tolerated differs socially across cultures (Ekman 1979). Recent psychological research by Elfenbein and Ambady found evidence for an "in-group advantage" in emotion recognition. That is, recognition accuracy is higher for emotions both expressed and recognized by members of the same (ethnic or regional) cultural group (Elfenbein and Ambady 2003a). Elfenbein et al. state, "This in-group advantage, defined as extent to which emotions are recognized less accurately across cultural boundaries, was smaller for cultural groups with greater exposure to one another, for example, with greater physical proximity to each other" (Elfenbein and Ambady 2003b). Their further in-depth research using Chinese and American facial expressions perceived by Chinese living in China and those living in US showed Chinese living in US for longer than 2.4 years were better at recognizing facial expressions of Americans than those of Chinese (Elfenbein and Ambady 2003c). In addition to in-group advantage, we also apply "decoding rule" in human facial expression recognition to the case of avatar expression recognition. Decoding rule implies we concentrate on recognition of negative expressions, since misinterpretation of negative expressions leads to more serious social problems than misinterpretation of positive expressions (Elfenbein and Ambady 2002).

This paper investigates cross-cultural evaluations of avatar expressions designed by Japanese and Western designers. The goals of the study were: (1) to investigate cultural differences in avatar expression evaluation and apply findings from psychological studies to human facial expression recognition, (2) to identify expressions and design features that cause cultural differences in avatar facial expression interpretation.

Three series of experiments were conducted to examine the above issues. The first experiment used avatars designed by Japanese artists. The avatars were evaluated by persons from three Asian and five Western countries. The second experiment used the same avatars, but the avatars were evaluated by persons from five Asian countries. The third experiment used Western-designed avatars that were evaluated by persons from four Western countries and Japan. All experiments were conducted on the web to gather as many participants as possible from various countries. In this paper, Sect. 2 describes the designs and results of Experiment 1, Sect. 3 describes Experiment 2, Sect. 4 describes Experiment 3, Sect. 5 discusses the cultural differences in interpreting avatar facial expressions from the results obtained by the three experiments and directs attention to future issues, and Sect. 6 concludes the study.

2 Experiment 1: cross-cultural evaluation of Japanese avatars by Western and Asian countries

Experiment 1 was conducted in 2003. Persons from eight countries, namely, Japan, South Korea, China, the United States, the United Kingdom, France, Germany, and Mexico, compared interpretations of avatars' facial expressions drawn by three Japanese designers. The experiment's website was open to the public. People from all over the world could access the website and freely participate in the experiment. Participation was voluntary. Experiment 1 investigated the following two issues:

1. Verifying cultural differences in interpreting avatars' facial expression. This was done by applying the above psychological findings on cultural differences in human facial expression recognition to the case of avatar expressions.
2. Identifying avatar facial expressions that are recognized differently across cultures.

2.1 Design of Experiment 1

2.1.1 Procedure of Experiment 1

The experiment itself was developed using Adobe Flash Player. Subjects first answered a brief questionnaire on their background profile. The main experiment started after the questionnaire, which was presented as a matching puzzle game (Fig. 1). Subjects were requested to match 12 facial expressions to 12 adjectives. The 12 facial expressions were displayed in a 4×3 matrix and the 12 adjectives as buttons below the matrix. As shown in Fig. 2, a subject could drag/drop the adjective buttons to/on the 12 expressions and continue changing the location of each button until he was satisfied with his answer. One avatar representation was chosen randomly from 40 avatars, and facial expression images were randomly placed in the 4×3 matrix. The adjective buttons were always displayed in the same order, and the 12 adjectives were always the same (see the next section for the adjectives used in the experiment).

Subjects' answers to the puzzle game and their background profile, including gender, age, county of origin, and native language, were logged in the server for later analysis. Subjects could continue the experiment with another set of avatars until they finished evaluating all 40 avatar designs, or they could stop at any time. Each avatar design was displayed only once to the same subject. The adjectives could be shown in English, French, German, Italian, Spanish, Chinese, Korean, and Japanese (all validated by native speakers). Subjects from countries



Fig. 1 Screen shot of Experiment 1: matching puzzle game between facial expressions and adjectives. Subjects could drag/drop the adjective buttons to the matching facial expressions

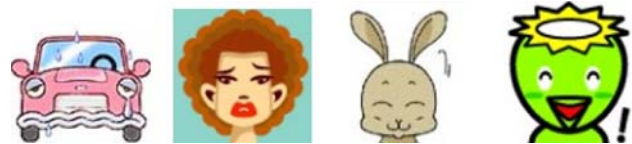


Fig. 2 Examples of avatar representations used in Experiment 1 designed by Japanese artists

where the above languages are primarily spoken could see the adjective selections in their native language according to the background profile. The default language was set to English.

2.1.2 Avatar designs of Experiment 1

Commercially used avatars were represented not by photo-realistic images but as caricatures or comic figures. We prepared 40 avatar representations drawn by three Japanese designers using Japanese comic/anime drawing style. By using avatars drawn with techniques from one culture, we were able to use these avatars as “expressers” and the subjects as “recognizers”, as in Elfenbein and Ambady (2003a, b). Accordingly, comparing the answers between Japanese users and those of other countries made it easier to validate the in-group advantage.

Avatars were represented in various forms, i.e., human figures, animals, plants, and objects. Figure 2 shows examples from the 40 avatar representations. Some designs use facial expression only (second from left), and others use gesture marks (second from right and right).



Fig. 3 Twelve facial expressions using one of the avatars. *Top row left to right* happy, sad, approving, disapproving, proud, ashamed. *Bottom row left to right* grateful, angry, impressed, confused, remorseful, and surprised. All expressions were drawn in Japanese comic style

2.1.3 Facial expression designs of Experiment 1

The 12 expressions used in the experiment were “happy,” “sad,” “approving,” “disapproving,” “proud,” “ashamed,” “grateful,” “angry,” “impressed,” “confused,” “remorseful,” and “surprised” as shown in Fig. 3. These expressions were selected from Ortony, Clore and Collins’ global structure of emotion types, known as the OCC model (Ortony et al. 1998). These are commonly used expressions in commercial chat and instant messenger services (i.e., MSN Messenger, Yahoo! Messenger) and they reflect those emotions desired by the subjects for intercultural communication (Koda 2004).

These 12 expressions are paired as valence expressions as defined in the OCC model, that is, as negative/positive emotions that arise in reacting to an event or person. “Happy,” “approving,” “proud,” “grateful,” and “impressed” are positive expressions, while “sad,” “disapproving,” “ashamed,” “angry,” “confused,” and “remorseful” are negative expressions, leaving “surprised” as a neutral expression.

2.2 Results of Experiment 1

2.2.1 Subjects and participating countries

We had 1,240 participants from 31 countries. The subjects’ gender ratio was roughly male:female = 1:1 (676 male subjects and 561 female). The subjects’ age ranges were: 6% were under 20, 43% were in their 20s, 35% were in their 30s, 12% were in their 40s, and 4% were in their 50s.

We analyzed answers from eight countries having more than 40 participants, namely, Japan ($n = 310$), South Korea ($n = 322$), China ($n = 50$), France ($n = 111$), Germany ($n = 62$), United Kingdom ($n = 49$), United States ($n = 75$), and Mexico ($n = 149$). The subjects from these eight countries saw the adjectives in their mother tongue. We used answers only in the cases where the

subject’s native language and the official language of his/her country matched.

2.2.2 Differences in interpretation of Avatar facial expression by country

Subjects’ answers to the puzzle game were analyzed by calculating matching rates between expressions and adjectives. There was no correct answer to the matching puzzle, but the avatar designers’ original intention was used as an expresser’s “standard” answer. Each expression and adjective was assigned a number (1–12) within the system. The designer’s intended pairs were described as (1,1), (2,2), (3,3), (4,4) reflecting (expression number, adjective number). We calculated each country’s number of “expression–adjective” pairs that were the same as the designers’ pairs. Consequently, here, “matching rate” means the percentage of pairs of expressions and adjectives that matched the avatar designer’s intentional pairs. For example, the matching rate of answer pairs (1,5), (2,1), (3,3), (4,9) is 25%.

The matching rate for each facial expression by country is shown in Fig. 4. The matching rate of Japanese participants was significantly higher for all expressions except “sad” and “disapproving” (by chi-squared test and Scheffe’s method of multiple comparison, $p < 0.01$), followed by the matching rate of Korean participants. Nevertheless, Japanese participants maintained high matching rates for the “sad” and “disapproving” expressions. There were no significant cultural differences in the matching rates among participants from countries other than Japan and Korea.

As stated in Sect. 2.1, avatars were designed by Japanese designers using Japanese comic/anime drawing techniques. Thus, we can regard the designers as expressers and the subjects as recognizers. Japanese subjects’ recognition accuracy of the avatar expressions was significantly higher than that of subjects from other countries, while Korean subjects’ accuracy was the second highest. This showed that there was an in-group advantage within the same country (within Japan) and one between neighboring countries (Japan and Korea).

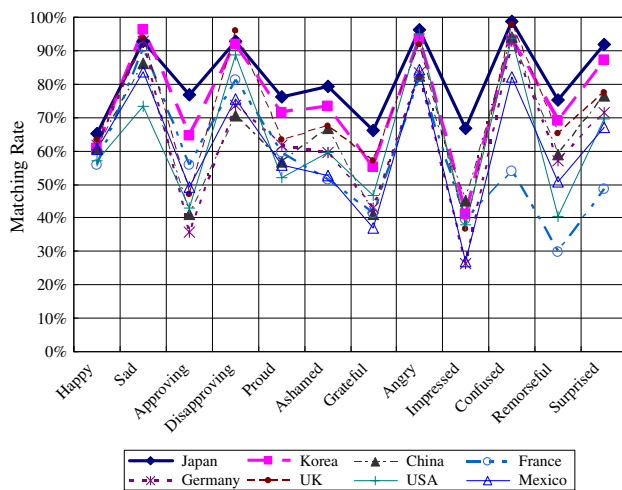


Fig. 4 Matching rate of each expression by country (Experiment 1). Matching rate means the percentage of pairs of expressions and adjectives that match the avatar designer’s intentional pairs. Numbers of answers by each country are: Japan, $n = 310$; Korea, $n = 322$; China, $n = 50$; UK, $n = 49$; France, $n = 111$; Germany, $n = 62$; USA, $n = 75$; Mexico, $n = 149$

2.2.3 Differences between positive/negative expressions

When we focused on the matching rate of each expression, the result showed that positive expressions in valence expression pairs (happy–sad, approving–disapproving, and grateful–angry) had lower matching rates than the negative expressions in the same pair. Negative expressions (sad, disapproving, angry, and confused) had significantly higher matching rates regardless of country (by analysis of variance and Scheffe’s method of multiple comparison, $p < 0.01$), while positive expressions (happy, approving, proud, grateful, and impressed) had significantly lower matching rates regardless of country. The matching rate of the “impressed” expression was significantly lower than that of any other expression (by analysis of variance and Scheffe’s method of multiple comparison, $p < 0.01$). This indicated that the subjects’ interpretations of negative expressions (sad, disapproving, angry, and confused) were similar to the designers’ intentions, regardless of country, and that the subjects’ answers to those expressions were similar across countries. On the contrary, the subjects’ interpretation of positive expressions (happy, approving, proud, grateful, and impressed) varied across countries.

We further analyzed the answers for the 12 expressions by principal component analysis. The results showed that positive expressions (happy, approving, proud, grateful, and impressed) got mixed up ($p < 0.01$). In other words, the reason for the positive expressions’ low matching rate was that each of these four expressions was not distinguished from the others.

Subjects’ comments supported this result. Both Japanese and non-Japanese commented that they had difficulty in selecting the expressions matching “approving,” “grateful,” and “impressed.”

Lastly, there were no cultural differences in the matching rates of avatar forms, i.e., human figures, animals, plants, and objects.

3 Experiment 2: cross-cultural evaluation of Japanese avatars by five Asian countries

Experiment 2 was conducted in 2005 within 5 Asian countries (Japan, South Korea, China, Malaysia, and Thailand). The reason for conducting the experiment within Asia was to validate the cultural differences found across Asia, Europe, and North America in Experiment 1 and show that these differences were again applicable in Asian countries, which are closer to each other.

In the second experiment we did the following:

1. we investigated cultural differences in avatar expression evaluation and applied findings from psychological studies on human facial expression recognition, namely, the “in-group advantage”, to Asian countries.
2. we identified design features that might cause cultural differences in avatar facial expression interpretation.

3.1 Design of Experiment 2

The procedure of Experiment 2 and the matching puzzle game was the same as the one conducted in Experiment 1 except for the following changes, which were made to control the experimental conditions more strictly.

1. Only invited, pre-registered participants from the five Asian countries could access the experiment site, while the participants in Experiment 1 had free access to the experiment site from all over the world. This change was made to gather enough number of participants for analysis only from the five Asian countries.
2. We used the same avatar designs drawn by three Japanese designers as used in Experiment 1, however, the number of avatar designs used in Experiment 2 was limited to 10 human figures instead of 40 as in Experiment 1. Avatar selection was made according to the design features used in the avatar designs in order to identify design features that might cause cultural differences in interpretation. The design features were categorized into three groups, namely, expressions only, expressions with a gesture mark, and expressions with a gesture.

3. Participants evaluated all 10 avatar designs in this experiment, while the participants could stop evaluating the avatar designs at any time during Experiment 1. Thus, the avatar designs and the number of avatars each participant evaluates were the same for all participants in Experiment 2.

3.2 Results of Experiment 2

3.2.1 Subjects and participating countries

There were 190 answers from Japan, 120 from South Korea, 300 from China, 160 from Malaysia, and 150 from Thailand. The participants were university students and researchers in their 20s and 30s, and the ratio of male to female participants was 1:1.

Japanese, Chinese, Korean subjects were shown the adjectives in their native language, and Thai and Malay subjects in English. The Thai and Malay subjects were fluent in English.

3.2.2 Differences in interpretation of avatar facial expressions

Figure 5 shows the matching rates shown by expression and country of the participants. When we focused on the matching rates by country of the participants, Japanese participants' matching rates were the highest for all expressions among participants from the five countries. This means the degree of matching the expresser (avatar designer)'s intention and the answers of the recognizers (participants) was high. Hence, an in-group advantage within the same country was identified in this experiment. This result further confirmed the results of the first experiment, in which Japanese participants' answers had significantly the highest matching rates among participants from eight countries, namely, Japan, South Korea, China, the United States, the United Kingdom, Germany, France, and Mexico. Thus, this result suggests that there are cultural differences among the five Asian countries, even though the geographical distance between them is smaller than that between the countries in Experiment 1.

When we focused on the matching rates by facial expression in Fig. 5, we again observed that negative expressions had higher matching rates than positive ones. Thus, as found in Experiment 1, the result of Experiment 2 suggested that the decoding rule was applicable to the answers from the participants from the five Asian countries.

3.2.3 Recognition accuracy by facial expression design

In this part of Sect. 3 we have analyzed the design features that would cause cultural differences in interpretation of

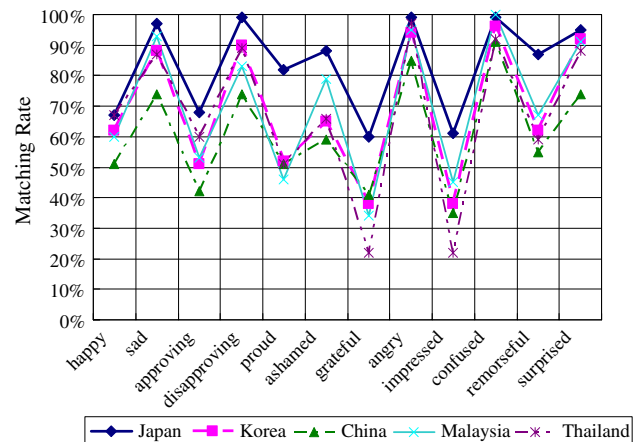


Fig. 5 Matching rates of each expression by subjects from five Asian countries in Experiment 2

avatar facial expressions. Among the facial expressions that had lower matching rates than others, we analyzed the answers to the “grateful”, and “impressed” expression by country of the participants. The design features that were used in the two expressions were categorized into three groups, namely, “facial expression only”, “facial expression with a gesture mark”, and “facial expression with a gesture.”

Figure 6 shows the design examples for the “grateful” expression. There were two designs for the “grateful” expression. The first used only a facial expression for “grateful”, and the latter used a facial expression with a gesture mark (heart mark) as shown in Fig. 6.

Figure 7 shows the details of answers to the “grateful” expression presented by only a facial expression and the one presented by a facial expression and a gesture mark. The answers of Japanese to the design with a heart mark had a higher matching rate (the percentage of answers that answered “grateful”) than the answers of Japanese to the design that used facial expression only. In contrast, in other countries using a heart mark did not necessarily result in a higher matching rate. Especially for South Korean and Thai participants, answers to the design using a “facial expression only” had a higher matching rate (the percentage of answers that answered “grateful”) than the ones to the design that used “a heart mark”. Adding a heart mark to the “grateful” expression design increased the number of



Fig. 6 Design samples of “Grateful” expression with a heart mark

participants from South Korea and Thailand who answered “impressed” rather than “grateful” for this design.

Next, we analyzed the answers to the avatar design using “a facial expression with a gesture mark”, and one using “a facial expression with gesture”, in regard to the “impressed” expression.

Figure 8 shows the design examples used for the “impressed” expression. There were two designs for the “impressed” expression. The first used “a facial expression with a gesture” (a “clapping hands” gesture) for “impressed”, and the latter used “a facial expression with a gesture mark” (the exclamation mark “!”).

Figure 9 shows the details of answers to the design that used a facial expression with a clapping hand gesture, and Fig. 10 shows the details of answers to the design that used a facial expression with “!”. The answers of Japanese participants showed the highest matching rate, about 80% (percentage of answers that answered “impressed”) among the answers of the participants from the five countries. In contrast, answers from participants from other countries varied according to the design used to express



Fig. 8 Example of the designs for the “impressed” facial expression. Right with “clapping hands” gesture, left with “!” mark

“impressed”. Chinese participants especially interpreted the “impressed” expression with a clapping hands gesture as “approving” more often than “impressed”. Thai participants interpreted the “impressed” expression with a “!” mark as “grateful” rather than “impressed”.

4 Experiment 3: cross-cultural evaluation of Western avatars by Western countries and Japan

Experiment 3 was conducted in 2008 to compare interpretations of avatars’ facial expressions drawn by Western designers. The objective of Experiment 3 was to examine whether the in-group advantage and decoding rule found in Japanese-designed avatars were also applicable to avatars’ facial expressions drawn by Western designers.

4.1 Design of Experiment 3

The procedure of Experiment 3 was the same as that of Experiment 2 except that the avatars used in Experiment 3 were drawn by Western designers, while the avatar designs used in Experiments 1 and 2 were by Japanese artists. The characteristics of the design for Experiment 3 were as follows;

1. We prepared seven avatar representations drawn by Western designers. A French designer drew two avatars, a British designer drew two, and an American drew one. They were all professional designers who used their countries’ comic/anime drawing styles. Each designer was a native of the country whose avatars he/she drew and spoke its language as a first language. A German researcher also created two avatars with a 3D facial expression modeling tool, which are referred to as German designs. The German designs have realistic faces without exaggeration or additional graphical information, which were made to compare interpretations with comic style avatars.
2. Invited participants evaluated all the seven avatar designs.

Figure 11 shows examples of the “displeased” expression drawn by French, British, and American designers and one created by the German researcher using a 3D modeling

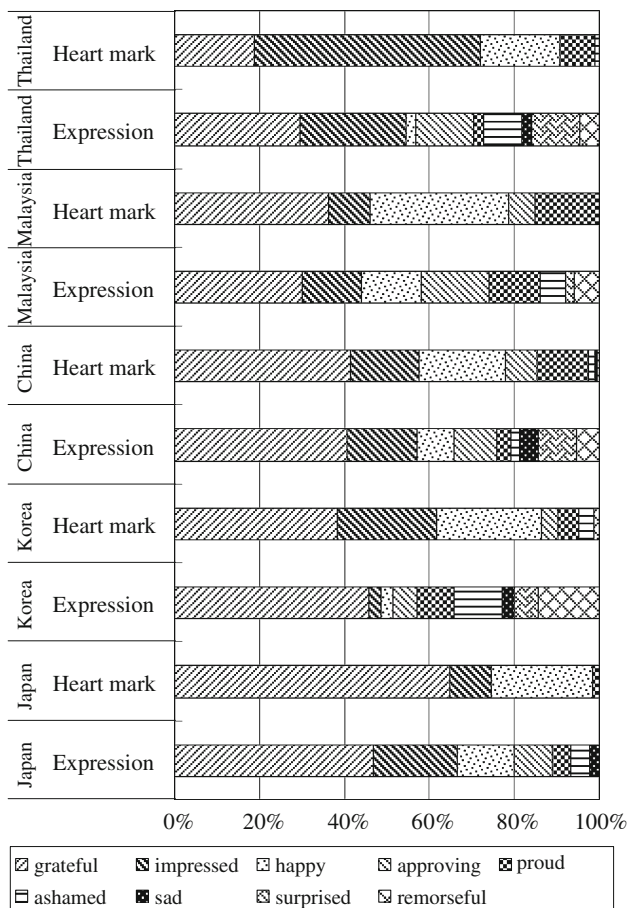


Fig. 7 Details of answers to the “grateful” expression (comparison of answers to the design that used only a facial expression and one that added a heart mark)

Fig. 9 Details of answers to the “Impressed” expression with a clapping gesture shown by country

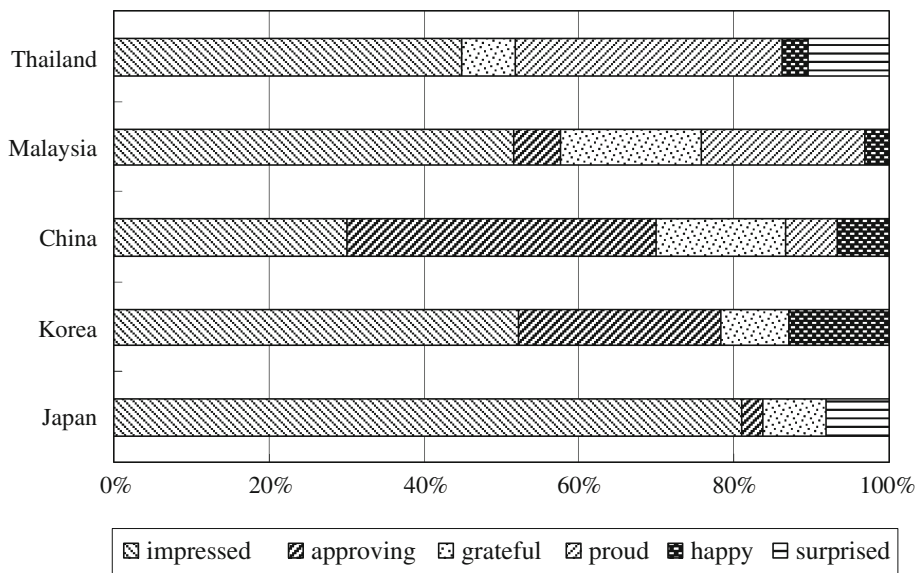
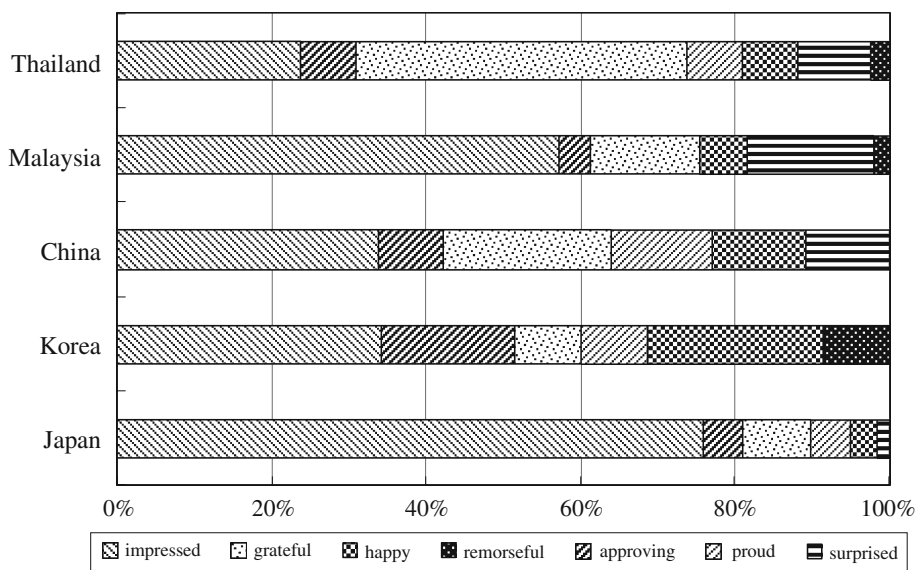


Fig. 10 Details of answers to “Impressed” expression with “!” shown by country



tool. Figure 12 shows 12 facial expressions of the avatar designed by the American designer.

4.2 Results of Experiment 3

4.2.1 Subjects and participating countries

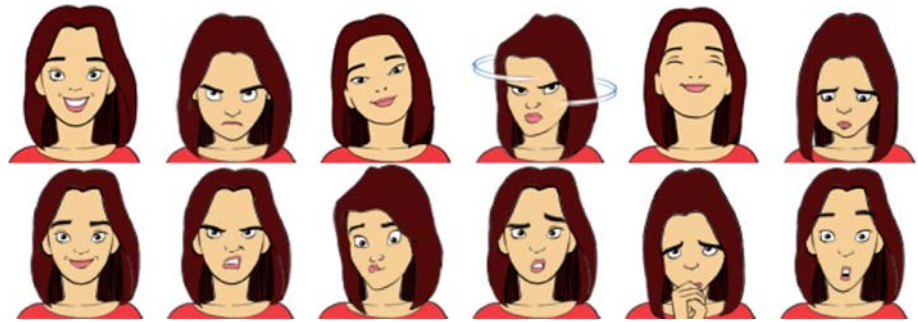
Participation in the experiment was by invitation only. The invitations were made to persons in Japan and in the countries that were the same as those of the designers. We collected a total of 293 answers. The number of answers from each country was: the United States ($n = 98$), France ($n = 23$), Germany ($n = 75$), and Japan ($n = 97$). We used answers only in the cases where the subject’s native language and the official language of his/her country



Fig. 11 Examples of “Displeased” expressions by Western designers used in Experiment 3. From left to right French, British, American and German designs

matched. The participants’ gender ratio was 65% male and 35% female. With regard to age, 80% of the participants were in their 20s, and 10% were in their 30s. With regard to computer ability, 53% of the participants rated themselves as expert computer users, and 43% as intermediate.

Fig. 12 Twelve facial expressions designed by an American designer. From top row, left to right pleased, displeased, approving, disapproving, proud, ashamed. Bottom row, left to right Grateful, angry, impressed, confused, remorseful, and surprised



4.2.2 Analysis of Average matching rates by designs

Table 1 shows the matching rates of the 12 expressions of German, UK, French, and US designs shown by country. When we compared the matching rates by country, the average matching rate of the French design by French participants, and the matching rate of the American design by American participants were the highest, compared to analogous matching rates by participants from other countries. In addition, the matching rates of Japanese participants were always the lowest for any avatar designs made by Western designers.

The average matching rates of negative expressions were always higher than those of positive expressions for any designs (the average matching rates of positive expressions to negative expressions in percentages were: German design 34.5–39.6%; UK design 46.6–73.9%; French design 65.6–79.6%; US design 40.9–67.0%).

The German designs, which were not drawn by a human artist but a 3D modeling tool, have the lowest average matching rate in any country, and 10 out of 12 expressions. The Japanese matching rate is the lowest among the four countries.

4.2.3 Analysis of matching rates by expression

In this section, we categorize the expressions into three groups according to their recognition accuracy. These categories are as follows:

1. Highly recognized expressions (expressions that had higher than a 70% matching rate by participants from all four countries): The expressions belonging to this category were: German ‘angry’; British ‘ashamed’; British ‘angry’; British ‘surprised’; French ‘disapproving’; French ‘proud’; French ‘ashamed’; French ‘angry’; and American ‘surprised’ (shown in Fig. 13).
2. Poorly recognized expressions (expressions that had lower than a 30% matching rate by participants from all four countries): The expressions belonging to this category were: German ‘disapproving’; German

‘impressed’; German ‘remorseful’; British ‘impressed’; and American ‘grateful’ (shown in Fig. 14).

3. Culture-dependent expressions (expressions that had more than a difference of 50% points between the highest and lowest matching rates, according to country) The expressions belonging to this category were: British ‘pleased’; British ‘proud’; French ‘grateful’; French ‘confused’; American ‘pleased’; American ‘disapproving’; American ‘proud’; and American ‘impressed’ (shown in Fig. 15).

Highly recognized expressions were consistently highly recognized by members from the four participating countries, and between Western and Japanese participants. Thus, we can assume these expression designs would cause fewer misinterpretations than other designs when used across countries. Poorly recognized expressions were consistently recognized as expressions different from those intended by the designer. The poorly recognized expressions were subtler expressions than the highly recognized ones, (i.e., ‘impressed’ vs. ‘angry’.) Culture-dependent expressions had wider variances in their matching rates, and Japanese matching rates were the lowest among them.

5 Discussion

Three series of experiments were conducted to examine cultural differences in interpreting avatar expressions. This section discusses the results of the three experiments.

5.1 Experiment 1: cross-cultural evaluation of Japanese avatars by Western and Asian countries

The results of overall recognition accuracy in Fig. 4 showed that avatar expressions designed by Japanese drawing techniques were recognized with significantly higher accuracy by subjects in Japan than by subjects in other countries. The recognition accuracy of participants from a neighboring country (Korea in Experiment 1) was the second highest. The in-group advantage mainly

Table 1 Matching rates shown by designers' and participants' countries

	US	France	Germany	Japan	Average		US	France	Germany	Japan	Average
<i>German 3D design</i>						<i>UK design</i>					
Pleased	42	50	36	48	35.2	Pleased	55	85	42	28	52
Displeased	14	50	20	15	19.8	Displeased	41	57	53	62	52.6
Approving	32	50	40	33	31	Approving	35	42	50	31	41.6
Disapproving	25	0	12	30	13.4	Disapproving	52	57	61	31	50.2
Proud	53	66	64	51	56.8	Proud	61	85	76	25	69.4
Ashamed	50	66	44	36	39.2	Ashamed	97	100	92	100	97.8
Grateful	35	16	44	15	32	Grateful	52	28	65	40	47
Angry	96	83	92	87	81.6	Angry	100	100	92	100	98.4
Impressed	32	16	28	12	17.6	Impressed	26	14	11	15	23.2
Confused	64	16	24	15	43.8	Confused	70	71	46	65	70.4
Remorseful	21	16	24	12	14.6	Remorseful	67	42	53	28	58
Surprised	39	33	44	39	31	Surprised	79	71	80	93	84.6
Average	41.9 ^a	38.5	39.3	32.8 ^b	34.7	Average	61.3	62.7 ^a	60.1	51.5 ^b	60.1
Positive expressions					34.5	Positive expressions					46.6
Negative expressions					39.6	Negative expressions					73.9
<i>French design</i>						<i>American design</i>					
Pleased	46	75	60	30	52.2	Pleased	37	75	25	12	49.8
Displeased	75	62	68	72	75.4	Displeased	43	75	41	50	61.8
Approving	90	87	84	72	86.6	Approving	37	0	16	0	10.6
Disapproving	96	87	92	72	89.4	Disapproving	62	75	33	100	74
Proud	78	100	76	84	87.6	Proud	75	75	100	12	72.4
Ashamed	78	87	76	78	73.8	Ashamed	50	50	50	31	56.2
Grateful	62	75	28	39	60.8	Grateful	25	25	16	18	16.8
Angry	96	100	96	96	97.6	Angry	37	50	41	62	58
Impressed	50	50	24	30	40.8	Impressed	81	50	25	18	54.8
Confused	62	87	32	27	61.6	Confused	100	75	58	93	85.2
Remorseful	75	75	52	57	61.8	Remorseful	50	25	50	31	51.2
Surprised	78	100	64	63	81	Surprised	100	100	91	100	98.2
Average	73.8	82.1^a	62.7	60.0 ^b	72.4	Average	58.1^a	56.3	45.5	43.9 ^b	57.4
Positive expressions					65.6	Positive expressions					40.9
Negative expressions					79.6	Negative expressions					67.0

The matching rates of the designer's country given in bold

^a Highest rates

^b Lowest rates

Fig. 13 Examples of expressions that were highly recognized by participants from all four countries. From top left to right German angry, British ashamed, British angry, French angry, American surprised



occurred within the same country where the expresser and recognizer belonged to the same culture, and the degree of recognition accuracy was next highest between participants from neighboring countries (Elfenbein and Ambady 2003a,

b). This result suggests that the in-group advantage that occurs in human expression recognition is applicable to avatar expression recognition within a country and between neighboring countries.



Fig. 14 Examples of expressions that were poorly recognized by participants from all four countries. *From top left to right* German remorseful, British impressed, American grateful

The results of the negative expressions having significantly higher recognition accuracy than the positive expressions may indicate that the “decoding rule” in psychological studies is applicable to avatar expressions. Mixing up expressions occurs within positive/negative expression groups other than “confused” and “surprised.” Accordingly, we can be less concerned about misunderstanding positive emotions as negative ones or vice versa. However, the connotations and implications of each expression, for example, whether one was approving or grateful within the positive expression group, were not recognized accurately across cultures. For example, the communication gap between China and Japan caused by different interpretations of the “big-eyed” expression in Koda (2004) was one example of a confusing experience for the subjects, although it did not lead to a serious misunderstanding.

The results of Experiment 1 suggest the following:

1. Cultural differences do exist in interpretation of avatar facial expressions, which confirms the psychological findings that physical proximity affects recognition accuracy. The in-group advantage was found within Japan and between Korea and Japan.
2. There are wide differences among cultures in interpreting positive expressions, while negative expressions have higher recognition accuracy. This result indicates that the decoding rule is found in avatar expression interpretation.

5.2 Experiment 2: cross-cultural evaluation of Japanese avatars by five Asian countries

The results in Fig. 5 showed that the avatar facial expressions designed using the Japanese comic drawing techniques had higher recognition accuracy for Japanese

participants than for the participants from other countries within Asia. Thus, this is another indication that the in-group advantage within the same country, which is found in recognizing human facial expressions, is also applicable to avatar expressions. There was still an in-group advantage in Experiment 2, with participants from five Asian countries, which were closer to each other than the countries of the participants in Experiment 1. We again found an indication that the decoding rule that negative expressions are more accurately recognized than positive ones in human facial expression applied to avatar expressions.

Next, analyses were made using different avatar expressions designs in the two facial expressions that had the lowest recognition accuracy. There were three different features used in the designs, namely, (1) facial expression only (i.e., “grateful” expression), (2) facial expression with a gesture mark (i.e., a heart mark in the “grateful” expression, an exclamation mark in the “impressed” expression), and (3) facial expression with a gesture (i.e., a clapping hand gesture in the “impressed” expression).

The results showed that the recognition accuracy of Japanese participants was the highest for all the three designs (facial expression only, facial expression with a gesture mark, and facial expression with a gesture) among the participants from the five Asian countries, and Japanese participants had greater recognition accuracy of designs using gesture marks than of designs using only facial expressions. Participants from countries other than Japan tended to have lower recognition accuracy of designs using gesture marks than of designs using only facial expressions. Thus, using a gesture mark does not necessarily improve the recognition accuracy for persons from countries other than the expresser’s country.

Similar cultural differences in interpreting gestures in pictograms were reported in (Cho et al. 2007). The survey was conducted using pictograms developed and used in NPO Pangaea’s communication software, which allowed children all over the world to communicate online using pictograms regardless of their mother tongues (Mori 2007; Takasaki and Mori 2007). The survey was conducted between the United States and Japan to ask meanings of 120 pictograms used in Pangaea’s communication software. The results suggest interpretations of gestures in pictograms vary according to culture (Cho et al. 2007). Cho et al. state the reason for these cultural differences can be



Fig. 15 Examples of culture-dependent expressions. *From top left to right* British pleased, British proud, French grateful, French confused, American pleased, American disapproving, American proud

explained by psychological studies by Efron (1941) and Ekman et al. (Ekman and Friesen 1969). Efron finds evidence that human gestures have different meanings according to culture. Ekman and Friesen categorize human gestures, and among these categories “emblem” gestures (symbolic gestures) are culturally dependent. Both the study of Cho et al. (2007) and this study on avatar interpretation found that “emblem” gestures (crossing arms to indicate “NO” in the former, clapping hands in the latter) had cultural differences in their interpretation. Thus, cultural differences in interpreting human gestures may be applicable to gestures in graphical representations such as avatars and pictograms.

The results of Experiment 2 suggest the following:

1. Cultural differences in interpreting avatars’ facial expressions existed among participants from Asian countries, even though these countries are closer geographically than the countries of the participants in Experiment 1. The psychological theory that suggests physical proximity affects facial expression recognition accuracy is also applicable to avatar facial expressions.
2. Use of gestures and gesture marks may sometimes cause counter-effects in recognizing avatar facial expression. Using gestures and gesture marks increased the recognition accuracy of participants from the expresser’s country. On the other hand, participants from other countries had a lower recognition accuracy of designs using gestures and gesture marks than of designs using facial expressions only.

5.3 Experiment 3: cross-cultural evaluation of Western avatars by Western countries and Japan

There was no clear indication of in-group advantage within the same country and between the neighboring countries in the results of matching rates of the Western-designed avatars (Table 1) as found with Japanese designs in Experiments 1 and 2. However, Japanese participants’ average matching rates of Western designs were always lower than Western participants’ matching rates of these designs. This suggests there was a tendency toward in-group advantage for participants from Western countries, since the five participating countries (France, Germany, UK, US, and Japan) can be divided into two groups, namely, Western countries (European countries and the United States) and Japan.

The German designs had the lowest matching rate in any country and most expressions. The German designs were created by a computer modeling tool, do not have exaggerated expressions or additional graphical information used in the human drawn American, French and British

designs. The reasons for this poor recognition rate might be the quality of the facial expression modeling tool itself, differences in design features used (or not used), and the participants’ higher expectations regarding the naturalness of the facial expressions of realistic avatars described as the Uncanny Valley model (Mori 1970).

When we categorized avatar designs by their recognition accuracy, the result showed that Japanese participants’ matching rates of the culture-dependent designs were the lowest among the four countries. Thus, again, there was a tendency for Japanese participants’ recognition accuracy of culture-dependent designs to be lower than that of participants from Western countries, which suggests an in-group advantage for participants from Western countries.

In addition, the analysis of the average matching rates by designs showed that negative expressions had higher recognition accuracy than positive expressions in Western-designed avatars. This may imply a decoding rule in Western designs.

Further issues to be examined are as follows:

1. *Variations in avatar designs* There should be more variations in avatar designs to avoid reliance on only one designer’s judgment and drawing style. We used a total of seven avatar designs drawn by four Western designers (one designer per country) in Experiment 3. We used 40 avatar designs in Experiment 1, avatar designs in Experiment 2, all of which were made by three Japanese designers. Although the four Western designers were carefully selected to have equal professional skills, we could ensure the quality of avatar designs by having more variety in designers and avatar designs.
2. *An increase in the number of participants* There were 293 answers in Experiment 3, while there were 1,240 answers in Experiment 1 and 920 in Experiment 2. The objective of Experiment 1 was to find evidences of cultural differences in interpreting avatar expressions in general; thus, we conducted an open web experiment. However, one cannot expect to have enough numbers of participants from specific counties in an open web survey. Hence, participation in Experiments 2 and 3 was by invitation only, which led to much less number of participants.
3. *Methodological issues in web surveys* As addressed in Reips 2000 and Schmidt 1997, there are disadvantages in web surveys, such as non/unfinished response errors, and measurement errors. We used a puzzle game style survey to increase the number of participants, used complete answers only for analyses to avoid unfinished errors, and used only the first answers from the same participant if he participated more than twice. Yet there are possibilities for measurement errors, such as

incorrect selection of countries and languages. We believe that effects of measurement errors can be minimized by gathering large number of participants.

4. *Translation of adjectives* Although we carefully selected each adjective and had it examined by several native speakers, translations of adjectives were not completely accurate, since some languages do not have words with the exact meaning of adjectives in other languages. Further experiment should be done with a scenario based interpretation for each avatar expression or embedding avatar expressions into an actual online activity.
5. *Definition of culture and assessment of cultural properties* We simply defined culture by the participants' country and their first language. However, the definition of culture is more complicated (Brislin 1983), since other elements of culture such as religion and values must also be taken into account. Also, cultural properties of individual participants need to be assessed via reliable and valid empirical methods as proposed in Ross 2004 and Vatrapu and Suthers 2007. Further study should consider cultural models such as Hofstede's rankings and uncertainty avoidance models (Hofstede 1984).

6 Conclusion

The goal of the study was to investigate cultural differences in avatar expression interpretations and apply findings from psychological studies in human facial expression recognition. The experiment using Japanese-designed avatars showed there were cultural differences in interpreting avatar facial expressions, and the psychological theory that suggests physical proximity affects facial expression recognition accuracy (in-group advantage) and the decoding rule were also applicable to avatar facial expressions. We again observed tendencies toward in-group advantage and application of the decoding rule among Western countries in the experiment using Western-designed avatars.

Acknowledgments Experiments 1 and 2 were conducted by a Grant-in-Aid for Scientific Research (A) (15200012, 2003–2005) from the Japan Society for the Promotion of Science (JSPS). Experiment 3 was conducted by a Grant-in-Aid for Young Scientists (Start-up: 2006–2007) and a Grant-in-Aid for Scientific Research (C) 20500196, 2008–2010) from the Japan Society for the Promotion of Science (JSPS). The work at Augsburg University was funded by the German Research Foundation (DFG) under research grant RE 2619/2-1 (CUBE-G).

References

Bartneck C, Takahashi T, Katagiri Y (2004) Cross cultural study of expressive avatars. In: Proceedings of the social intelligence design 2004

- Berger CR, Calabrese R (1975) Some explorations in initial interactions and beyond.
- Brislin RW (1983) Cross-cultural research in psychology. *Annu Rev Psychol* 34:363–400
- Cho H, Ishida T, Yamashita N, Inaba R, Mori Y, Koda T (2007) Culturally-situated pictogram retrieval. In: Ishida T, Fussell SR, Vossen PTJM (eds) *Intercultural collaboration I*, Lecture Notes in Computer Science 4568, Springer, pp 221–235
- Damer B (1997) *Avatars: exploring and building virtual worlds on the internet*. Peachpit Press, Berkeley
- Efron D (1941) *Gesture and environment*. King's Crown Press, New York
- Ekman P (1979) About brows: emotional and conversational signals. In: Cranach MV, Foppa K, Lepenies W, Plog D (eds) *Human ethology: claims and limits of a new discipline: contributions to the colloquium*. Cambridge University Press, Cambridge, pp 163–202
- Ekman P (2003) *Emotions revealed: recognizing faces and feelings to improve communication and emotional life*. Henry Holt, New York
- Ekman P, Friesen W (1969) The repertoire of nonverbal behavior: categories, origins, usage, and coding. *Semiotica* 1:48–98
- Elfenbein HA, Ambady N (2002) On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychol Bull* 128(2):203–235
- Elfenbein HA, Ambady NA (2003a) Universals and cultural differences in understanding emotions. *Curr Dir Psychol Sci* 12(5):159–164
- Elfenbein HA, Ambady NA (2003b) Cultural similarity's consequences: a distance perspective on cross-cultural differences in emotion recognition. *J Cross-Cultural Psychol* 34:92–110
- Elfenbein HA, Ambady NA (2003c) When familiarity breeds accuracy: cultural exposure and facial emotion recognition. *J Pers Soc Psychol* 85(2):276–290
- Hofstede G (1984) *Cultures consequences: international differences in work-related values*. Sage, London
- Isbister K, Nakanishi H, Ishida T (2000) Helper agent: designing and assistant for human–human interaction in a virtual meeting space. In: *Proceedings of human factors in computing systems (CHI2000)*. ACM Press, New York, pp 57–64
- Koda T (2004) Interpretation of expressive characters in an intercultural communication. In: Negoita MG, Howlett R, Jain LC (eds) *8th international conference of knowledge-based intelligent information and engineering systems (KES2004)*. Lecture Notes in Artificial Intelligence, 3214, Part II. Springer, Berlin, pp 862–868
- Kurlander D, Skelly T, Salesin D (1996) *Comic chat*. In: *Proceedings of computer graphics and interactive techniques*. ACM Press, New York, pp 225–236
- Mori M (1970) The uncanny valley. *Energy* 7(4):33–35
- Mori Y (2007) Atoms of bonding: Communication components bridging children worldwide. In: Ishida T, Fussell SR, Vossen PTJM (eds) *Intercultural collaboration I*. Lecture Notes in Computer Science 4568, Springer
- Ortony A, Clore GL, Collins A (1998) *The cognitive structure of emotions*. Cambridge University Press, Cambridge, MA
- Pesson P (2003) ExMS: an animated and avatar-based messaging system for expressive peer communication. In: *Proceedings of GROUP*, ACM Press, New York, pp 31–39
- Rehm M, Nakano Y, André E, Nishida T (2008) Culture-specific first meeting encounters between virtual agents. In: *Intelligent virtual agents 2008*. Springer, Berlin, pp 223–236
- Reips UD (2000) The Web experiment method: advantages, disadvantages, and solutions. *Psychological experiments on the Internet*. Academic Press, San Diego, pp 89–117
- Ross N (2004) *Culture and cognition: implications for theory and method*. Sage Publications, Thousand Oaks

- Ruttkay Z (2008) Cultural dialects of real and synthetic facial expressions. In: Proceedings international conference on intelligent user interfaces (IUI 2008), workshop on enculturating conversational interfaces by socio-cultural aspects of communication
- Schmidt W (1997) World-Wide Web survey research: benefits, potential problems, and solutions. *Behav Res Methods Instrum Comput* 29:274–279
- Smith MA, Farnham SD, Drucker SM (2000) The social life of small graphical chat spaces. In: Proceedings of CHI, ACM Press, New York, pp 462–469
- Takasaki T, Mori Y (2007) Design and development of a pictogram communication system for children around the world. In: Ishida T, Fussell SR, Vossen PTJM (eds) *Intercultural collaboration I*, Lecture Notes in Computer Science 4568. Springer
- Vatrapu R, Suthers D (2007) Culture and computers: a review of the concept of culture and implications for intercultural collaborative online learning. In: Ishida T, Fussell SR, Vossen PTJM (eds) *Intercultural collaboration I*, Lecture Notes in Computer Science 4568. Springer pp 260–275