JAISCR, 2011, Vol.1, No.3, pp. 215–231

INTERCULTURAL DIFFERENCES IN DECODING FACIAL EXPRESSIONS OF THE ANDROID ROBOT GEMINOID F

Christian Becker-Asano[1] and Hiroshi Ishiguro[2]

[1]*Research Group on the Foundations of AI,*
Department of Computer Science
University of Freiburg, Germany
e-mail:christian@becker-asano.de

[2]*Department of Adaptive Machine Systems*
Osaka University, Japan
e-mail:ishiguro@ams.eng.osaka-u.ac.jp

Abstract

As android robots become increasingly sophisticated in their technical as well as artistic design, their non-verbal expressiveness is getting closer to that of real humans. Accordingly, this paper presents results of two online surveys designed to evaluate a female android's facial display of five basic emotions. Being interested in intercultural differences we prepared both surveys in English, German, as well as Japanese language, and we not only found that in general our design of the emotional expressions "fearful" and "surprised" were often confused, but also that Japanese participants more often confused "angry" with "sad" than the German and English participants. Although facial displays of the same emotions portrayed by the model person of Geminoid F achieved higher recognition rates overall, portraying *fearful* has been similarly difficult for her. Finally, from the analysis of free responses that the participants were invited to give, a number of interesting further conclusions are drawn that help to clarify the question of how intercultural differences impact on the interpretation of facial displays of an android's emotions. [1]

1 Introduction and Motivation

Researchers in the field of social robotics (e.g. [2, 3]) mostly opt for rather abstract designs of their robots, which nevertheless are assumed to express "human" qualities such as emotions. Virtual reality researchers, in contrast, more often design for very human-like virtual agents, which can only be presented in two or three dimensions on the screen (e.g. [4]). In the field of "Android Science" [5] these two approaches are combined, because the robotic research platforms are explicitly designed as anthropomorphic as possible.

Our motivation behind building highly anthropomorphic robots is twofold: First, they are supposed to serve us as sophisticated tools for investigating fundamental questions about human nature, e.g., how appearance and behavior combine to fuel the impression of conversing with another human rather than a machine. Second, we aim at letting robots blend into a future society, in which humans accept such robotic counterparts as social actors at least to some extent. Therefore, we believe it crucial for an android robot to also master the non-verbal means of communication to convey its emotional state in a way that is most convenient for humans to be read and reliably interpreted.

[1]This article is an extended version of [1]. This work was partially supported by Grant-in Aid for Scientific Research (S), KAKENHI (20220002), and by a post-doctoral fellowship of JSPS.

The high degree of human-likeness of the android robot "Geminoid F", which was modeled to resemble her human counterpart's outer appearance to the finest detail, together with its sophisticated mechanical design, permits to create diverse facial expressions. Thus, it is reasonable to investigate this android's emotional expressiveness based on our belief that human-machine interaction benefits from a machine's ability to recognize, express, model, communicate, and respond to emotion [6].

The remainder of this article is structured as follows: Related work is being discussed in the following section giving rise to two research questions. In Section III two online surveys are described and the results of their first parts are summarized in Section IV. Motivated by the conclusion drawn from the results of the first part of both surveys in Section V, the free response part of the first survey is analyzed in Section VI. In the final Section VII general conclusions are drawn.

2 Related Work

"Geminoids" [7] are a special type of tele-operated robots and the term itself is derived from the Latin word "geminus" meaning twin and the ending "˜oides" meaning similarity. In contrast to the class of humanoid robots [8, 9], which are similarly designed to let people associate them with humans, the outer appearances of android robots such as "Geminoid HI-1" [7] or "Geminoid F" even feature artificial skin and hair, and they are modeled to the finest detail in the aim to make them indistinguishable from their real human counterparts at first sight. With these "androids" it is possible to pursue research in the field of "Android Science" [5], because they provide "a key testing ground for social, cognitive, and neuroscientific theories." [10]

"Geminoid HI-1" has been the first android of the Geminoid family and it was designed to resemble the outer appearance of the second author. Although it is easily mistaken for a human its facial expressivity is rather limited. In fact, visitors of an arts museum who unexpectedly encountered Geminoid HI-1, which was tele-operated such that remote conversations could take place, quite often mentioned in post-hoc interviews that improving its facial movements might further the impression of talking to a real person [11].

"Geminoid F" has been developed concentrating on its ability to perform sophisticated facial expressions. As it was also modeled after a real person, we can now compare its facial expressivity with that of its model person (cp. Figure 1).

Figure 1. Geminoid F (left) and its model person (right)

In general, emotional expressiveness has been evaluated for a number of social robots and virtual agents. Five of the basic emotions [12], namely anger, fear, happiness, sadness, and surprise, have been realized with the iCat robot [13]. In case of 100% geometrical intensity, the authors report average recognition rates between approx. 42% for fear and approx. 81% for surprise. With respect to virtual characters, esp. the emotion fear has been found to be difficult to realize as a facial expression [14]. This particular emotion, however, is known to be difficult in the human case as well [15] and the authors of [14] conclude that "affective expressions of machines are as convincing as expressions of humans."

Moreover, intercultural differences have been found in the perception of facial cues and their interpretation with regard to emotions [16]. In essence, Japanese observers tend to weight cues in the eyes more than cues displayed in the mouth, whereas American people seem to show the opposite tendency when being asked to judge facial displays of emotions. In effect, Japanese participants rate a big

smile with neutral eyes as less happy than American participants. This effect has even been confirmed for stylized facial icons (also called emoticons) which are often used in internet text mails.

On this background, with evaluating Geminoid F's facial displays of emotions we aim to gain insights into two questions: First, are we able to tune Geminoid F's facial actuators in such a way that the readability of her emotional facial expressions is comparable to that of the real person's static facial displays of the same emotions? Second, can we replicate the intercultural differences in interpretations of such facial displays of emotions?

3 The Two Online Surveys

The android robot Geminoid F was built to closely resemble her human model person's outer appearance (cp. Figure 1). Its artificial body has the same proportions, same facial features, same hair color and hairstyle as its original such that at first sight and from a distance it is difficult to tell them apart.

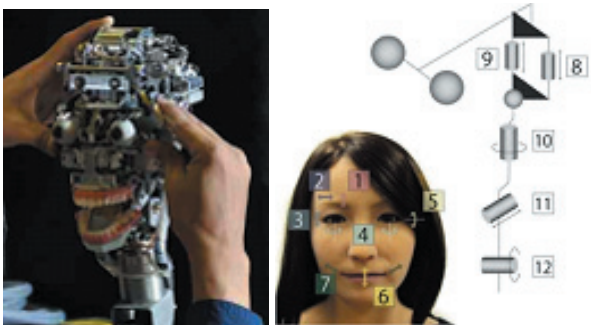

Figure 2. Geminoid F's interal configuration; left, its head without skin and hair, right, the distributions of a total of 12 degrees of freedom are (1) both eyebrows up and down, (2) both eyebrows left and right, (3) both eyelids open and close, (4) both eyes left and right, (5) both eyes up and down, (6) mouth open and close, (7) both lip corners back and forth, (8) head tilt back left, (9) head tilt back right, (10) head left and right, (11) breathing animation, and (12) upper body front and back

Geminoid F's smooth silicon skin and sophisticated internal design (cp. Figure 2) allows for a variety of facial expressions. A combination of pneumatic and electric actuators allow for a total of 12 degrees of freedom of which seven are located in its face, three in its head and neck, and two in its upper body (cp. Figure 2). In contrast to the previously developed Geminoid HI-1, the limbs of Geminoid F are immobile. This reduced complexity has the advantage that the controllers for the pneumatic actuators could be integrated into its body such that only one air pressure and one controller cable needs to be connected to Geminoid F. Of course, Geminoid F cannot stand up, perform gestures, or walk although its arms and legs look similarly humanlike as its upper torso and its face.

3.1 Purpose of the First Online Survey

We decided to limit our first empirical study to the investigation of static facial displays of emotions realized with Geminoid F. We are well aware that dynamic information plays an important role in successfully decoding the emotional content behind facial expressions [17]. Nevertheless we believe to already generate valuable insights by at first letting people from different cultures evaluate still images of its face.

Six digital pictures of Geminoid F's face were taken featuring the basic emotions angry, fearful, happy, sad, and surprised [12] plus a neutral expression. They were realized by manually adjusting the actuators through a software interface. In a similar study involving facial expressions of primary (i.e. basic) and secondary emotions of a virtual human [18] it was found that primary emotions could be identified much better than secondary ones such as hope or relief. Therefore, we decided to focus this study on solely evaluating the display of (a subset of) basic emotions. Care was taken to keep the lighting constant and comparable between all pictures. They were then scaled to 200 pixels width and 205 pixels height, before they were used in the first online survey.

3.1.1 Experimental Procedure

The first online survey was designed to test the readability of Geminoid F's facial display of emotions. As we were also interested in intercultural differences, we prepared the survey in German, Japanese, and English language. The six emotion labels with their translations into German and Japanese are presented in Table 1. Accordingly, on

the first page the participant has to choose one of these languages as his or her language for the rest of the survey.

Table 1. The seven labels with their corresponding translations

English	German	Japanese
(none of these labels)	keines dieser Labels)	「どれも該当しない」
angry	wtend	怒り（いかり）
fearful	ngstlich	恐れ（おそれ）
happy	erfreut	喜び（よろこび）
neutral	neutral	無表情 (むひょうじょう)
sad	0.768370	悲しみ（かなしみ）
surprised	berrascht	驚き（おどろき）

Subsequently, an introduction is given in which we explain that we aim "to find out, if our android robot Geminoid F can express her emotions with her face." The participant is also assured that completing the survey will not take more than five minutes and that it consists of two parts.

On page two of the survey we ask the participants for their gender, age, and nationality, of which only gender is a mandatory field. Furthermore, they can state, if the respective language chosen in the beginning is their native language. They have to confirm their entries by pressing a continue button.

Part one starts on page three with an introduction on how the participants are supposed to choose from six labels below each picture. Instead of assigning any of the labels "angry", "fearful", "happy", "neutral", "sad", or "surprised" to a picture, they can also assign "(none of these labels)" (cf. Table 1), which is set to be the default value for each of the drop down boxes. The concrete explanations are given as follows:

- You are requested to use the drop down box below the picture.
- If you are not sure which label to select, feel free to choose the option "(none of these labels)".
- You may also choose the same label for more than one picture! For example, if you think that three pictures show a happy face, you might choose "happy"' for each of these pictures.

In order to clarify the procedure, an example picture is shown together with an example of the seven choices in a drop down box. Finally, the participants are requested to press a button labeled "Start part one" to proceed to the next page of the survey.

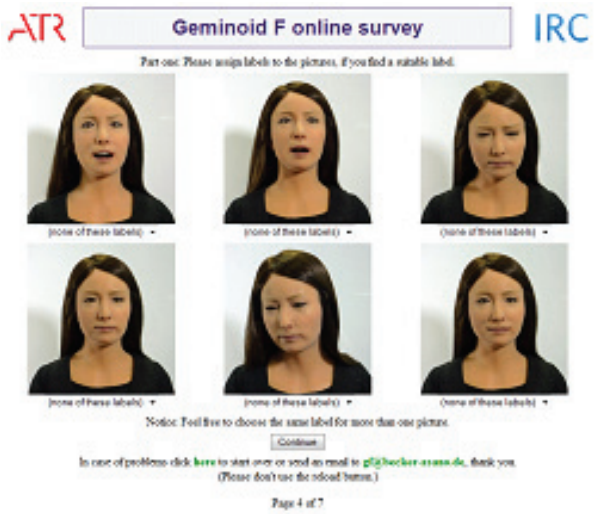

Figure 3. Screenshot of the first online survey presenting Geminoid F's portrayal of five basic emotions plus a supposedly neutral expression; from left to right, top to bottom: fearful, surprised, angry, neutral, sad, happy

Part one of the survey starts on page four with the instruction to "[p]lease assign labels to the pictures, if you find a suitable one", cp. Figure 3. The participants are reassured that they might also choose to assign any label to more than one picture, if they liked. Accordingly, all six pictures showing each facial expression of Geminoid F are presented with a drop down box below each one. The arrangement of the pictures was randomized between participants to avoid any order effects, but we chose to present all pictures at once to give the participants the opportunity to compare them with each other. Alternatively, we could have presented the facial expressions one after the other in a randomized sequence. With our setup, however, we hoped to avoid a learning effect, i.e., that participants—due to their lack of experience with an android robot's general ability to perform facial displays—get more and more experienced to the end of the sequence. This might lead to a steady change of judgment over the course of the survey, which we avoided

by presenting all pictures at once. A very similar method was used in previous studies, which aimed at evaluating cross-cultural differences of expressive avatars, e.g. [19].

After pressing the "Continue" button on page four the participants get instructions on the second part of the survey, in which they are asked to label each facial expression with one word of their own choice by typing it into a text field below each picture. On that page the instructions are summarized as follows:

- You are requested to type one word into the text field below the picture.
- If for some picture you have no idea, you might leave the corresponding text field blank.
- Anything that comes to your mind when looking at the pictures is fine. There are no 'correct' choices.

By pressing a button labeled "Start part two" they proceed to the next page, on which the same six facial expressions are presented in the same order as in part one (e.g. in Figure 3). The drop down boxes, however, are exchanged for blank input fields and the instruction above the pictures is changed to read: "Please label each facial expression with one word." Furthermore, below the pictures they are reminded that "any word is fine, because there are no 'correct' choices." After pressing "Continue" one last time the participants are thanked for their participation on page seven.

Participants have been invited over the internet, through advertisements on mailing lists, and through direct communication. This first survey was online from 5^{th} of May 2010 and after 20 days 499 internet users opened the first page. Four hundred ninety data sets were assumed valid and 235 of them were male (mean age 31.7 years; standard deviation SD = 13.7 years) and the remaining 255 female (mean age 27 years; SD = 7.7 years). With respect to the languages, 99% of those who chose Japanese, 91% of those who chose German, and 76% of those who chose English completed the survey in their native language (i.e. a total of 430 participants). As we are interested in intercultural differences and with 11 participants not stating their respective nationalities, we grouped the remaining total of 479 participants by their respective nationalities into those from the Asian region (N=168), those from the European region (N=192), and those from the American region (N=119); cp. Figure 4. We do not distinguish native speakers from non-native speakers in the analysis of the multiple choice part of the survey, which will be presented next.

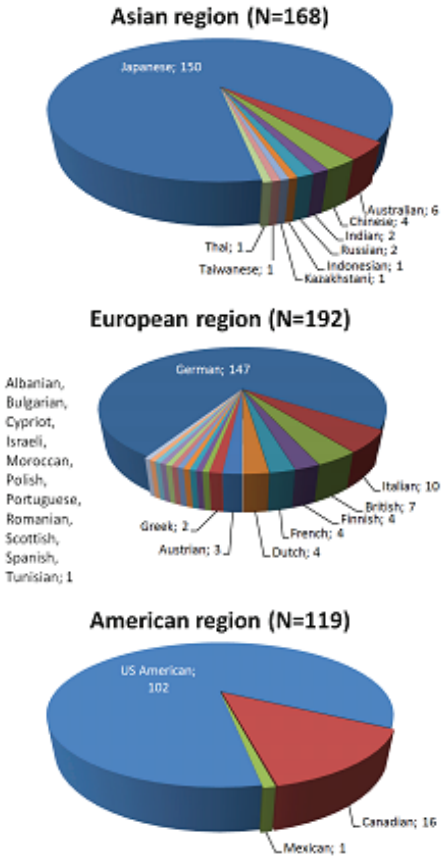

Figure 4. All participants grouped by region of origin according to the stated nationality of each participant in the first survey

3.1.2 Results of the Multiple Choice Part of the First Survey

A global confusion matrix (N=479) is presented in Table 2. Without distinguishing the participants' cultural backgrounds only the emotions *happy* (78.5%) and *sad* (74.9%) are recognized rather reliably. The best recognition rate is achieved for the *neutral* expression (83.9%). Moreover, the expressions intended to convey *fearful* and *surprised* respectively, are both labeled as "surprised" most often (*fearful*: 65.6%; *surprised*: 63.3%). In

Table 2. Confusion matrix of the global recognition rates (in percentages) of six facial displays of emotions designed for Geminoid F with presented picture (columns) against selected labels (rows); highest values are set bold face

Global results (N=479)						
Label	Picture					
	angry	*fearful*	*happy*	*neutral*	*sad*	*surprised*
angry	**54,5***	3,5	0,4	5,2	6,5	1,5
fearful	3,1	10,4	0,4	0,6	4,2	32,6*
happy	0,2	15,4*	**78,5***	0,0	0,4	0,6
neutral	1,0	0,0	15,2*	**83,9***	5,8	0,2
sad	34,0*	0,6	0,2	5,0	**74,9***	0,0
surprised	0,2	**65,6***	0,4	0,6	0,2	**63,3***
none	6,9	4,4	4,8	4,5	7,9	1,9

* above chance level of 14.8%

Figure 5. The android robot Geminoid F (top row) and its model person (bottom row) portraying five basic emotions and a neutral expression, from left to right: angry, fearful, happy, neutral, sad, surprised

addition, of all facial expressions only the *surprised* expression is labeled as "fearful" above chance level (32.6% against 14.8% chance level). Similarly, *angry* is being confused with "sad" (34%), although it is still most often labeled as "angry" (54.5%). With a Cohen's kappa for Geminoid F (GF) of $\kappa_{GF,global}$=0.536 the global agreement is satisfactory.

Thus, it seems as if our design of a *fearful* expression is most problematic. The *happy, sad*, and *neutral* expressions, however, seem to work well enough.

After splitting the data according to the three regions introduced above, the following intercultural differences can be observed:

1. The 168 Asian participants (cf. Table 3) assign "sad" (50.6%) more often than "angry" (38.1%) to the picture *angry*. In addition, they show the least confusion in assigning the label "surprised" to the picture *surprised* (73.2%) and most in assigning "surprised" to the *fearful* picture (59.5%). In general, this group features the smallest agreement ($\kappa_{GF,Asian}$=0.505).

2. The 192 participants of the European region (cf. Table IV) most often assign the label "angry" to the picture showing *angry* (66.7%) in comparison to the other two groups. In addition, they label *happy* remarkably often with "neutral" (22.4%). In case of all other pictures their judgments show less confusion than those of the Asian participants, but more confusion than those of the Americans. This is also reflected in this group's $\kappa_{GF,European}$=0.543 lying between $\kappa_{GF,Asian}$=0.505 and $\kappa_{GF,American}$=0.568.

3. The 119 participants of the American region (cf. Table 5) show the highest agreement ($\kappa_{GF,American}$=0.568), although they seem to be most confused concerning their judgment of the *surprised* picture (only 54.6% choose "surprised" and 42.9% label it with "fearful") as compared to the other two groups. Furthermore, this group's participants had the strongest tendency to assign the label "surprised" to the *fearful* picture (71.4% as compared to 67.4% and 59.5%).

In summary, the Asian group shows the worst agreement, the American group the best, and the European group lies in between. Notably, only the Asian participants label the *angry* expression with "sad" more often than "angry." In a similar fashion the American participants seem to have most difficulties in deciding between "surprised" and "fearful" for labeling the *surprised* expression—an expression, for which most of the Asian participants agree on choosing "surprised."

There are at least two possible factors, which could explain these differences: First, the general intercultural differences in the evaluation of facial expressions [16] and/or, second, the artificial nature of Geminoid F's outer appearance (cp. Figure 5, top row), which might let human observers apply different judgment standards as compared to judging a real human's facial expressions. For example, Japanese people are assumed more open to the idea of accepting robots as helpers in daily life than European and American people [19].

3.2 Purpose of the Second Online Survey

In order to clarify the reasons for the above intercultural differences and also to estimate the quality of the recognition results themselves, we conducted a second online survey featuring the model person's facial expressions (cp. Figure 5, bottom row). In particular we aimed to find out, Equation (1) if the intercultural differences would reoccur, and Equation (2) if the real person's portrayals would result in similar recognition rates.

The model person (MP) was instructed to portray the same five basic emotions angry, fearful, happy, sad, and surprised plus a neutral expression. She did not know Geminoid F's portrayals but we showed and explained to her a printout of Figure 16.1 of [15, p. 304], in which the five target emotions are portrayed by actors. During the photo session the camera was set to self-timer such that in the moment when the picture was taken no one was looking at the model person. The lighting conditions were matched to those of the previous pictures and the resulting six pictures were also resized to 200 pixels width and 205 pixels height (cp. Figure 5, bottom row).

3.2.1 Experimental Procedure

The second survey was very similar to the first one the only difference being the presentation of

Table 3. Confusion matrix of the recognition rates (in percentages) for the participants of the asian region; displays of Geminoid F (columns) against selected labels (rows); highest values are set bold face

Asian region only (N=168)						
Label	Picture					
	angry	*fearful*	*happy*	*neutral*	*sad*	*surprised*
angry	38,1*	6,0	0,6	4,8	9,5	2,4
fearful	3,6	11,9	0,6	1,2	4,2	22,0*
happy	0,0	19,0*	**75,6***	0,0	0,6	0,6
neutral	1,8	0,0	10,7*	**80,4***	6,5	0,0
sad	**50,6***	1,2	0,6	6,0	**70,2***	0,0
surprised	0,0	**59,5***	1,2	0,6	0,0	**73,2***
none	6,0	2,4	10,7	7,1	8,9	1,8

* above chance level of 14.8%

Table 4. Confusion matrix of the recognition rates (in percentages) for the participants of the European region; displays of Geminoid F (columns) against selected labels (rows); highest values are set bold face

European region only (N=192)						
Label	Picture					
	angry	*fearful*	*happy*	*neutral*	*sad*	*surprised*
angry	**66,7***	1,6	0,5	7,8	3,6	0,5
fearful	3,6	9,9	0,5	0,5	4,7	35,4*
happy	0,5	16,1*	**75,0***	0,0	0,5	1,0
neutral	0,5	0,0	22,4*	**81,8***	7,3	0,5
sad	20,8*	0,5	0,0	5,7	**76,0***	0,0
surprised	0,0	**67,2***	0,0	1,0	0,0	**59,9***
none	7,8	4,7	1,6	3,1	7,8	2,6

* above chance level of 14.8%

Table 5. Confusion matrix of the recognition rates (in percentages) for the participants of the American region; displays of Geminoid F (columns) against selected labels (rows); highest values are set bold face

European region only (N=192)						
Label	Picture					
	angry	*fearful*	*happy*	*neutral*	*sad*	*surprised*
angry	**58,0***	3,4	0,0	1,7	6,7	1,7
fearful	1,7	9,2	0,0	0,0	3,4	42,9*
happy	0,0	9,2	**88,2***	0,0	0,0	0,0
neutral	0,8	0,0	10,1	**92,4***	2,5	0,0
sad	31,9*	0,0	0,0	2,5	**79,8***	0,0
surprised	0,8	**71,4***	0,0	0,0	0,8	**54,6***
none	6,7	6,7	1,7	3,4	6,7	0,8

* above chance level of 14.8%

the model person's pictures instead of Geminoid F's pictures. We even did not change the introduction, i.e. participants were told they would have to judge facial expressions of an android robot Geminoid F. Participants were requested, however, to state if they had participated in the previous online survey already.

The invitations to this survey were distributed similarly to the previous procedure and 256 valid datasets were retrieved during the first ten days of June 2010. Of these 256 participants 110 are male (mean age 34.2 years; SD=14.1 years) and 146 are female (mean age 32.8 years; SD=11 years). In case of the male participants 25% had completed the previous survey as compared to 34% of all female participants.

Five participants did not state their respective nationalities and are, thus, excluded from the analysis such that 251 datasets remain. As presented in Figure 6, 58 participants originate from the Asian region the majority being Japanese nationals again. Seventy-nine percent of the Asian group's participants completed the survey in their native language. The 80 German nationals are the majority of all 122 participants of the European region and 70% used their native language to complete the survey. With 46 participants from the USA the American group contains a total of 71 participants, of whom 75% chose their native language. Thus, the fraction of native speakers in this second survey's data is comparably high as the one achieved in the first survey. Again, we do not distinguish native and non-native speakers in the further analysis.

3.2.2 Results of the Multiple Choice Part of the Second Survey

A global confusion matrix (N=251) is presented in Table 6. Without distinguishing the participants' cultural backgrounds all emotional displays of the model person are recognized rather reliably with *happy* (96.8%) achieving the highest and *fearful* (61.4%) the lowest recognition rate. The recognition rate of the *neutral* expression (81.7%) is similar to the one of the same expression portrayed by Geminoid F, in which case it was the highest recognition rate. For the model person, however, a recognition rate of 81.7% must be judged as average in comparison to the other facial display results.

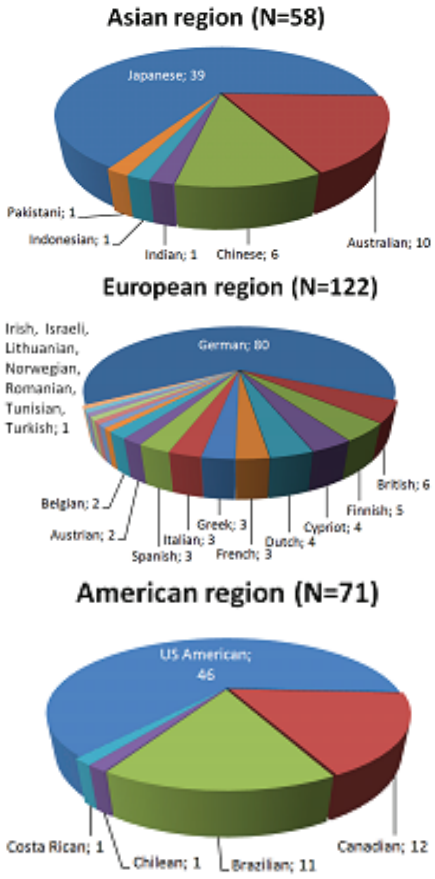

Figure 6. All participants grouped by region of origin according to the stated nationality of each participant in the second survey

Thus, it seems as if it was most difficult for our model person to portray *fearful* as a facial expression and easiest to convey *happiness*. In contrast to Geminoid F's results, much less confusion occurred when labeling the *surprised* expression (94.0% as "surprised"). The global agreement is considerably better for the model person ($\kappa_{MP,global}$=0.808) than for Geminoid F ($\kappa_{GF,global}$=0.568). Next, we split the data again according to the participants' respective regions of origin (cp. Figure 6) to check for inter-cultural differences in the perception of the model person's facial displays of emotion. The confusion matrices for the data of different regions are presented in Table 7, Table 8, and Table 9, respectively.

The following intercultural differences reappear:

1. The Asian participants (N=58, cf. Table 7) show the lowest agreement on the *fearful* expression

Table 6. Confusion matrix of the global recognition rates (in percentages) of six facial displays of emotions performed by the model person with presented picture (columns) against selected labels (rows); highest values are set bold face

Global results (N=251)						
Label	Picture					
	angry	*fearful*	*happy*	*neutral*	*sad*	*surprised*
angry	**89,6***	2,0	0,8	3,2	6,4	0,8
fearful	2,4	**61,4***	0,4	0,8	1,6	0,8
happy	0,0	0,4	**96,8***	0,4	0,0	2,0
neutral	0,4	3,6	0,4	**81,7***	1,2	0,0
sad	2,8	12,7	0,4	11,2	**79,3***	0,0
surprised	0,4	8,8	0,0	0,4	0,0	**94,0***
none	4,4	11,2	1,2	2,4	11,6	2,4

* above chance level of 14.8%

(58.6%) as well as the *surprised* (87.9%) expression as compared to both the European and the American group. They also most often chose to not assign any label to a facial display (*none* in Table 7). Their level of agreement is the lowest of all three groups ($\kappa_{MP,Asian}$=0.767).

2. The 122 participants of the European region (cf. Table 8) show a higher agreement ($\kappa_{MP,European}$=0.814) than the Asian participants again. It is comparable to the American group's level of agreement ($\kappa_{MP,American}$=0.832). The worst recognition rate is achieved for *fearful* (62.3%) and the best for *happy* (97.5%) similarly to the other two groups. Notably, none of the non-intended emotion labels are above level of chance for this group.

3. Participants of the American region (N=71, cf. Table 9) judged similarly to the European group. Notably, they achieve the highest recognition rates for four of the six facial displays with *sad* being an exceptional case (87.3% against 72.4% for the Asian and 77.9% for the European participants). They reach the highest overall level of agreement ($\kappa_{MP,American}$=0.832).

4 Summary of the Multiple Choice Results

In summary, once again for the Asian participants static facial displays of emotions seemed to be most ambiguous as reflected in their low level of agreement and their most pronounced tendency to refrain from assigning any of the labels. Interestingly, however, this tendency has not been present in case of judging Geminoid F's facial expressions (cf. Table 5).

The performances of the European and American participants are very similar again. Especially, the European participants' tendency to label *happy* as a "neutral" expression for Geminoid F (cf. Table IV) disappeared. On the contrary, they once again agree best on labeling *angry* as conveying anger ("angry", 91%) for the model person. The American group's participants are not confused any more with respect to labeling the *surprised* expression, when it is portrayed by the model person (cf. Table 9, 97.2%).

5 Conclusions for the First Part of Both Surveys

We set out to investigate, if we could (1) achieve recognition rates of facial displays of emotions with Geminoid F that are similar to the ones achieved by the model person herself, and (2) replicate the previous findings on intercultural differences in the perception of facial displays with Geminoid F. Of course, we are also interested in the engineering aspect of how to possibly improve the android's mechanical design, for which the model person's results are indicative.

Concerning our first goal, we have found that the facial expressions portrayed by Geminoid F were more ambiguous ($\kappa_{GF,global}$=0.536) than those

Table 7. Confusion matrix of the recognition rates (in percentages) for the participants of the asian region; displays of the model person (columns) against selected labels (rows); highest values are set bold face

Asian region only (N=58)						
Label	Picture					
	angry	*fearful*	*happy*	*neutral*	*sad*	*surprised*
angry	**89,7***	1,7	1,7	8,6	1,7	3,4
fearful	3,4	**58,6***	0,0	0,0	3,4	0,0
happy	0,0	0,0	**93,1***	0,0	0,0	1,7
neutral	0,0	3,4	1,7	**79,3***	1,7	0,0
sad	0,0	15,5*	0,0	5,2	**72,4***	0,0
surprised	0,0	5,2	0,0	0,0	0,0	**87,9***
none	6,9	15,5*	3,4	6,9	20,7*	6,9

* above chance level of 14.8%

Table 8. Confusion matrix of the recognition rates (in percentages) for the participants of the European region; displays of the model person (columns) against selected labels (rows); highest values are set bold face

European region only (N=122)						
Label	Picture					
	angry	*fearful*	*happy*	*neutral*	*sad*	*surprised*
angry	**91,0***	2,5	0,8	1,6	8,2	0,0
fearful	0,8	**62,3***	0,0	1,0	0,0	1,6
happy	0,0	0,0	**97,5***	0,0	0,0	1,6
neutral	0,0	3,3	0,0	**82,0***	0,8	0,0
sad	3,3	9,8	0,8	13,1	**77,9***	0,0
surprised	0,8	9,8	0,0	0,8	0,0	**95,1***
none	4,1	12,3	0,8	0,8	13,1	1,6

* above chance level of 14.8%

Table 9. Confusion matrix of the recognition rates (in percentages) for the participants of the American region; displays of the model person (columns) against selected labels (rows); highest values are set bold face

American region only (N=71)						
Label	Picture					
	angry	*fearful*	*happy*	*neutral*	*sad*	*surprised*
angry	**87,3***	1,4	0,0	1,4	7,0	0,0
fearful	4,2	**62,0***	1,4	0,0	2,8	0,0
happy	0,0	1,4	**98,6***	1,4	0,0	2,8
neutral	1,4	4,2	0,0	**83,1***	1,4	0,0
sad	4,2	15,5*	0,0	12,7	**87,3***	0,0
surprised	0,0	9,9	0,0	0,0	0,0	**97,2***
none	2,8	5,6	0,0	1,4	1,4	0,0

* above chance level of 14.8%

performed by the model person ($\kappa_{MP,global}$ =0.808). Furthermore, the *fearful* expression of Geminoid F is more often labeled with "surprised" (65.6%) than Geminoid F's *surprised* expression (63.3%). As the latter expression being most often mistaken to convey the emotion "fearful" (32.6%), it seems reasonable to at least switch the expressions with each other. They are, however, visually rather similar anyway (cp. Figure 5) and it might be best to design a new facial display of *fearful* for Geminoid F. Interestingly in this respect, the facial displays of *happy* and *neutral* are also very similarly designed for Geminoid F, but were distinguished rather reliably by the participants of the first survey. This leads us to conclude that visual similarity of emotional facial displays alone is not necessarily a predictor of categorical confusion. Finally, the participants' tendency to avoid choosing the label "fearful" in the first survey is present in the second survey as well, which can be explained by a general difficulty to portray fear in static facial displays [14, 15].

With respect to our second goal, Geminoid F's rather limited ability to change its face around the eyes (cp. Figure 5, top row) should result in more ambiguous ratings of the Asian (esp. Japanese) participants, who tend to focus more on that facial region and less on the mouth [16]. In fact, not only their global level of agreement is lower ($\kappa_{GF,Asian}$=0.505) than that of both other groups, but they also show the least agreement (in comparison to the other two groups) in labeling Geminoid F's *happy* expression (cf. Table 5, 75.6%). Even their impression of Geminoid F's *angry* face as conveying the emotion "sad" possibly results from this difference in facial expression decoding. This interpretation is supported by the results of the first part of the second survey. The model person's portrayals show much more variations around the eyes (cp. Figure 5, bottom row) and in line with our interpretation the Asian group's judgments become much less ambiguous. They are, however, most critical as they (compared to both other groups) most often decide to not assign any of the labels.

All of these findings, however, are also in line with (and can be attributed to) general cross-cultural differences in recognizing emotions from facial expressions [20]. In order to find out, which of these two factors influence our results stronger, we analyze the free response part of the Geminoid F survey next.

6 Analysis of Part Two of the Geminoid F Survey

If general cross-cultural differences in facial expression classification with regard to emotions are the major reason for the first survey's results in part one, then it would be reasonable to assume that the Japanese participants' free responses in part two of the same survey show less agreement than those responses provided by the non-Japanese participants. Of course, such diversity could also be inherent in the specific language being used to describe emotions. Thus, in the following a careful analysis of the Japanese and English speaking free responses will be presented, taking into account such language specific differences as good as possible.

6.1 Data Preprocessing

Translating emotion terms from one language to another is a notoriously difficult undertaking. With Japanese being the source language this difficulty is even getting worse, because most Japanese terms can be written either in Hiragana, Katakana, or Kanji and as in many other language a huge number of synonyms are in everyday use when it comes to emotions. Thus, the following three steps were taken to preprocess the Japanese free response data:

1. Merging exact matches on the textual level
2. Translating from Japanese to English as accurately as possible
3. Merging again based on English translations ignoring differences in the lexical category (e.g. "surprise" versus "surprised")
4. Clustering of resulting tokens according to their type of expression (e.g. "exclamations" versus "emotion terms")

In case of the English free responses only steps one and three were necessary leading to clusters that were already meaningful and clear enough to be compared. Including the German responses as well seems not to be helpful, because they would

need to be translated as well, which would cause another source of linguistic ambiguity.

All of the 156 Japanese speakers, who completed part two of the Geminoid F survey, were using their native language, but only 142 (75.1%) of all 189 English responses were given by native English speakers. Only the data of native speakers are included in the following analysis.

Furthermore, the analysis of the first part of the survey revealed that the facial expression designed to portray the emotion "fearful" was least recognized and was often assigned the label "surprised" instead. The analysis of the free responses concerning these two facial expressions is complemented by an analysis of the free responses given for the "angry" facial expression, because its recognition rate for Geminoid F in the first part also remained relatively poor in both cultural contexts; cf. Table 3 and Table 5 on page 2.

6.2 Results

In the following the free responses for each single facial expression will be compared separately.

6.2.1 Geminoid F's Facial Disply of "Angry"

In the multiple choice part of the first survey this facial expression was often misclassified as to convey *sadness* instead of *anger*. In fact, the 168 participants of the Asian region even assigned the label "sad" more often than the intended label "angry" (50.6% against 38.1%; cf. Table 3). Thus, in our aim to further clarify, which emotions are conveyed to the observer, we analyzed the free response part of the facial display *angry* for Geminoid F.

The data preprocessing explained above applied to the data of the Japanese native speakers resulted in a total of 98 responses, which were clustered into the groups "emotion terms" (58), "short exclamations" (8), and "unemotional terms and descriptions" (32). Accordingly, only 59.2% of all responses could be identified as clearly referring to emotions. The remaining 40.8% of responses are not considered further, because the English native speakers' responses did not include any such descriptions or exclamations.

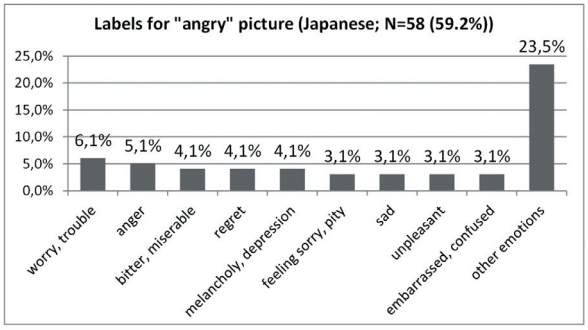

Figure 7. Those 58 responses of a total of 98 responses to the "angry" picture of Geminoid F that are clearly emotional and were given by the Japanese native speakers ("other emotions" contains all those responses that had a frequency count of less than three, i.e. less than 3.1%)

Figure 7 presents an overview of the 58 emotion terms given by the Japanese native speakers for the "angry" facial expression portrayed by Geminoid F. Not surprisingly, only 5.1% of the free responses directly refer to *anger*. Interestingly, however, even the term "sad" is only chosen by 3.1% of all participants, but its semantically close neighbors "worry", "miserable", "regret", "melancholy", and "pity" account for another 21.5%.

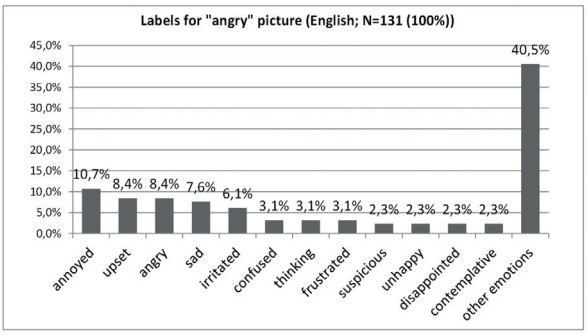

Figure 8. The 131 responses of all native English speakers when they had to judge the "angry" facial expression of Geminoid F ("other emotions" again contains all emotion responses with a frequency count of less than 3, i.e. less than 2.3%)

An overview of all 131 emotion terms given by the native English speakers in response to the "angry" face of Geminoid F is shown in Figure 8. Although, with 8.4% a higher percentage of the native English speakers label the expression with "anger" than in the Japanese participants' case, this is still

dominated by the frequency by which the label "annoyed" (10.7%) was assigned. The term "sad" was given relatively often as well (7.6%), but in contrast to the Japanese responses only the semantically close neighbor "disappointed" (2.3%) reached the minimum frequency count of three.

Thus, it seems reasonable to follow that the "angry" facial expression of Geminoid F was perceived as a more dominant emotional expression by the English group as compared to the Japanese group. Furthermore, as the Japanese participants were rather reluctant to choose clear emotion terms, to them the facial expressions seem to be more ambiguous than to the English native speakers.

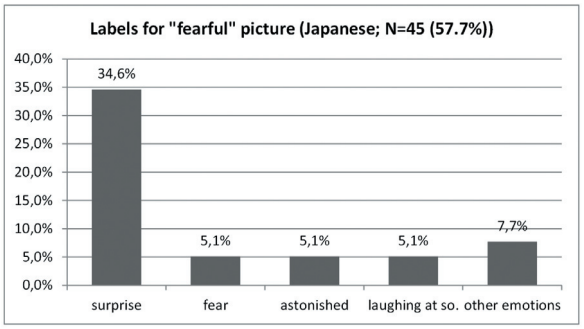

Figure 9. Those 45 responses of a total of 78 responses to the "fearful" picture of Geminoid F that are clearly emotional and were given by the Japanese native speakers ("other emotions" contains all those responses that had a frequency count of less than three, i.e. less than 5.1%)

6.2.2 Geminoid F's Facial Display of "Fearful"

The *fearful* expression was most often misclassified as "surprised" by both the Japanese (cf. Table 3, 59.5%) and the American (cf. Table 5, 71.4%) participants. The Japanese, however, also assigned the label "happy" quite frequently (19.0%), such that we were interested to find out, if this is reflected in their free responses as well.

In Figure 9 the 45 emotion terms are summarized, which represent 57.7% of all free responses given by a total of 78 Japanese native speakers for the "fearful" expression. Once again, the emotion label, which this facial expression was intended to convey, namely "fear", was given in only 5.1% of all responses. In fact, "surprised" was most often entered (34.6%) and "happy" appeared only once. Thus, it is part of the "other emotions" cluster in Figure 9. Interestingly, a small number of Japanese participants agreed in that they all entered "laughing at so." (5.1%), which is semantically close to happiness. Once more, however, more than 40% of all responses classified as either an exclamation or an unemotional description, rather than an emotion.

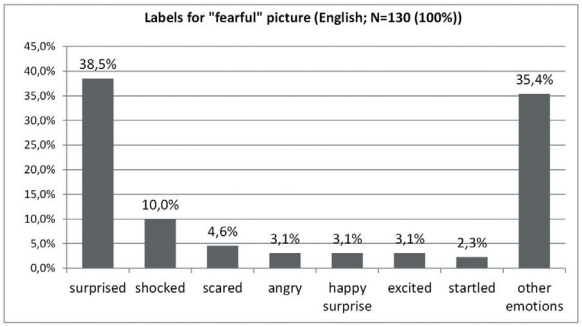

Figure 10. The 130 responses to the "fearful" picture of Geminoid F given by the English native speakers ("other emotions" contains all those responses that had a frequency count of less than three, i.e. less than 2.3%)

Figure 10 shows an overview of the 130 responses entered by the English native speakers when judging the "fearful" picture. Similarly to the Japanese participants, "surprise" was also the dominant response (38.5%). Contrary to them, however, 10% of the English native speakers entered "shocked" and another 4.6% "scared" as labels for this picture. These two emotions are clearly negative, which contrasts the rather positive response "laughing at so." given by the Japanese. Furthermore, 35.4% of the responses had to be summarized in the "other emotions" group, because each of them did not reach a frequency count of more than two.

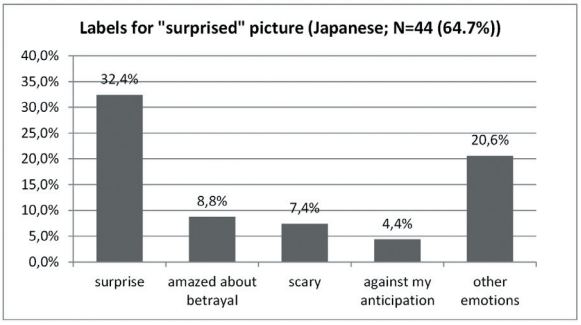

Figure 11. The 44 responses of a total of 68 responses to the ”surprised” picture of Geminoid F given by the Japanese native speakers that are clearly emotional (”other emotions” contains all those responses that had a frequency count of less than three, i.e. less than 4.4%)

In summary, the analysis of the free responses for the “fearful” facial expression are in line with the results of the first part of the Geminoid F survey (cf. Table 3 and Table 5). First, the Japanese native speakers’ responses are more positively valenced than those of the English native speakers. Second, the Japanese show a higher agreement in the choice of emotion terms (only 7.7% “other emotions”) than the English speakers (35.4% “other emotions”), which is similar to the difference in choosing “none of these labels” in the first part of the survey (Japanese: 2.4%, American: 6.7%). Once again, however, 42.3% of the Japanese entered exclamations, unemotional statements, or descriptions instead of clearly identifiable emotion terms thereby limiting the meaningfulness of this comparison.

6.2.3 Geminoid F’s Facial Display of “Surprised”

In case of Geminoid F’s “surprise” expression the first part of the survey revealed that the participants of the American region least frequently chose the label “surprised” (54.6%, cf. Table 5) in comparison to the Asian (73.2%, cf. Table 3) and also to the European region (59.9%, cf. Table 4). Moreover, the Americans also most often of all three groups chose the “fearful” label (42.9%, cf. Table 5). Thus, in analyzing the free entries in response to the “surprised” picture we were most interested to see, if the tendency can be confirmed that the American participants judge the expression more negatively than the Asian participants.

Figure 11 gives an overview of all those 44 responses by the Japanese native speakers, who gave a clearly emotional response to the “surprised” picture. Notably a rather high number of Japanese participants (35.3%) once again responded either with an exclamation, an unemotional term, or a situational description instead of an emotion term and these responses are excluded from Figure 11. With 32.4% most emotional responses qualified for “surprise.” The sole clearly negative response was “scary” with 7.4% and 20.6% of all responses have not been stated by more than two participants and are summarized under “other emotions” in Figure 11.

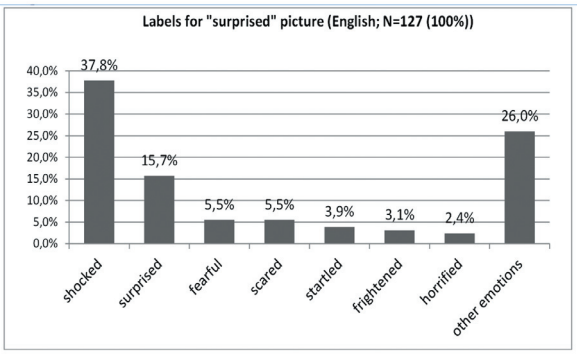

Figure 12. All 127 responses to the ”surprised” picture of Geminoid F given by the English native speakers. (”other emotions” contains all those responses that had a frequency count of less than three, i.e. less than 2.4%)

As shown in Figure 12, the English native speakers most frequently enter “shocked” in response to the “surprised” expression (37.8%), which is followed by 15.7% “surprised”. Even all of the other emotion terms, on which at least three participants agreed, are negative and semantically close to “fear.” With 26% of all responses gaining a frequency count of less than three, the “other emotions” group for “surprised” is comparable to that of the Japanese native speakers (20.6%, cp. Figure 11).

In conclusion, the free response data for “surprise” seems to confirm the tendency derived from the first part of the survey that again Japanese people seem to judge this expression as less negative than Americans. In fact, comparing Figure 12 with Figure 10, for American people exchanging the expressions “fearful” and “surprised” might be a good idea, because the “fearful” expression was labeled most often with “surprised” and the “surprised” expression most often with “shocked.” Although for the Japanese it would also be better to use the “fearful” expression to convey surprise, the “surprised” expression would not suffice to convey “fearful” to them.

6.3 Summary of Part Two of the Geminoid F Survey

In general, the Japanese native speakers judged all three facial expressions to convey less negative and more submissive/passive emotions, e.g. "worry" (cp. Figure 8), "astonished" (cp. Figure 9), or "amazed about betrayal" (cp. Figure 11). In addition, in all three cases roughly 40% of their responses were non-emotion terms, which could be taken to indicate a general reluctance to assign any clear emotional meaning to the expressions designed to convey basic emotions. In contrast, neither the English nor the Germen native speakers used any exclamations and only few Germans used descriptions when labeling the emotional facial expressions of Geminoid F.

7 Conclusions

In conclusion, this study confirms many of the previous findings surrounding the identifiability of facial displays of emotion. Thus, we successfully created facial expressions with Geminoid F to let it convey *happy, neutral,* as well as *sad*, but we have only been moderately successful with *surprised* and *angry* expressions, and our design of a *fearful* expression failed the test.

The analysis of the free response part of the first survey revealed interesting intercultural differences. First, roughly 40% of all responses given by the Japanese participants did not qualify as clearly describing an emotion term, but were either unemotional descriptions of situations or onomatopoetic descriptions of short exclamations. Second, Japanese participants show the tendency to interpret Geminoid F's facial expressions not only more positively than American participants, but also do they ascribe a less dominant stance to the android. One could speculate that the latter intercultural difference is at least in part caused by Geminoid F being designed to resemble a (Japanese) woman and more than half of the responses in the second part of the Geminoid F survey have been given by male Japanese participants (57%). Popular books and movies on Japan written by westerners often frame Japanese women as being submissive and obedient to Japanese males, but from an anthropologist's perspective this simplistic view is at least problematic, if not wrong [21]. Perhaps, the findings at hand are more likely caused by general intercultural differences in decoding facial expressions [16, 20], but we have to admit that clarifying this question remains on open topic for further investigation.

Finally, we have to acknowledge that the intuitive design of facial expressions was only our first attempt to let Geminoid F convey emotions. How these impression might change, when Geminoid F is used in an interactive situational context, is only but one of many interesting research questions that should be tackled in future research.

Acknowledgment

We would like to thank Shuichi Nishio, Kohei Ogawa, and Hiromi Ogawa for their help in preparing this study. We are also very thankful for the helpful comments provided by the anonymous reviewers.

References

[1] C. Becker-Asano,&H. Ishiguro, "Evaluating facial displays of emotion for the android robot Geminoid F." *Workshop on Affective Computational Intelligence.*. Paris: IEEE, 2011. 22–29.

[2] K. Dautenhahn, I. Nourbakhsh and T. Fong, "A Survey of Socially Interactive Robots." *Robotics and Autonomous Systems* 42 (2003): 143-166.

[3] C. Breazeal, "Emotion and sociable humanoid robots." *International Journal of Human-Computer Studies* 59 (2003): 119–155.

[4] C. Becker-Asano and I. Wachsmuth, "Affective computing with primary and secondary emotions in a virtual human." *Autonomous Agents and Multi-Agent Systems* 20, no. 1 (2010): 32–49.

[5] H. Ishiguro, "Android Science: Toward a new cross-interdisciplinary framework." *Proc. of the CogSci 2005 Workshop "Toward Social Mechanisms of Android Science"*. Stresa, Italy, 2005. 1–6.

[6] R. W. Picard, *Affective Computing.* The MIT Press, 1997.

[7] S. Nishio, H. Ishiguro, and N. Hagita, "Geminoid: Teleoperated Android of an Existing Person." In *Humanoid Robots, New Developments*, 343–352. I-Tech, 2007.

[8] T. Kanda, T. Hirano, D. Eaton, and H. Ishiguro, Interactive Robots as Social Partners and Peer Tutors

for Children: A Field Trial.“ *Human Computer Interaction (Special issues on human-robot interaction)* 19 (2004): 61–84.

[9] T. Kanda, H. Ishiguro, T. Ono, M. Imai, K. and Mase, Development and Evaluation of an Interactive Robot ‘Robovie’.“ *IEEE International Conference on Robotics and Automation.* 2002. 1848–1855.

[10] K. F. MacDorman, and H. Ishiguro, The uncanny advantage of using androids in cognitive and social science research.“ *Interaction Studies*, 2006: 297–337.

[11] C. Becker-Asano, K. Ogawa, S. Nishio, H. and Ishiguro, ”Exploring the uncanny valley with Geminoid HI-1 in a real world application.” *IADIS Intl. Conf. on Interfaces and Human Computer Interaction.* Freiburg, Germany: IADIS, 2010. 121–128.

[12] P. Ekman, ”Basic Emotions.” Chap. 3 in *Handbook of Cognition and Emotion*, 45–60. John Wiley & Sons, 1999.

[13] C. Bartneck, J. Reichenbach, and A. Breemen, ”In your face, robot! The influence of a character’s embodiment on how users perceive its emotional expressions.” *Proceedings of the Design and Emotion.* Ankara, 2004.

[14] C. Bartneck, ”How convincing is Mr. Data’s smile?: Affective Expressions of Machines.” User Modeling and User-Adapted Interaction, 2001: 279-295.

[15] P. Ekman, ”Facial Expressions.” Chap. 16 in Handbook of Cognition and Emotion, 301–320. John Wiley & Sons, 1999.

[16] M. Yuki, W. Maddux, and T. Masuda, ”Are the windows to the soul the same in the East and West? Cultural differences in using the eyes and mouth as cues to recognize emotions in Japan and the United States.” Journal of Experimental Social Psychology, 2007: 303–311.

[17] Y. Zhang, and Q. Ji, ”Active and Dynamic Information Fusion for Facial Expression Understanding from Image Sequences.” IEEE Transactions on pattern analysis and machine intelligence, 2005: 699-714.

[18] J. Tolksdorf, C. Becker-Asano, and S. Kopp, ”Do You Know How I Feel? Evaluating Emotional Display of Primary and Secondary Emotions.” Edited by H. Prendinger, J. Lester and M. Ishizuka. Intelligent Virtual Agents. Springer, 2008. 548–549.

[19] T. Koda, ”Cross-Cultural Study of Avatars’ Facial Expressions and Design Considerations Within Asian Countries.” Proc. of Intercultural Collaboration. Springer, 2007. 207–220.

[20] J.A. Russell, ”Is There Universal Recognition of Emotion From Facial Expression? A Review of the Cross-Cultural Studies.” *Psychological Bulletin* 115 (1994): 102-141.

[21] M. A. Tamanoi, ”Women’s Voices: Their Critique of the Anthropology of Japan.” *Annual Review of Anthropology* (1990):17–37.