Evaluating facial displays of emotion for the android robot Geminoid F

Christian Becker-Asano
Intelligent Robotics and Communication laboratories
Advanced Telecommunication Research Institute Int.
Kyoto, Japan
christian@becker-asano.de

Hiroshi Ishiguro
Department of Adaptive Machine Systems
Osaka University
Japan
ishiguro@ams.eng.osaka-u.ac.jp

Abstract*—With android robots becoming 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. We prepared both surveys in English, German, and Japanese language allowing us to analyze for inter-cultural differences. Accordingly, we not only found that our design of the emotional expressions "fearful" and "surprised" were often confused, but also that many Japanese participants seemed to confuse "angry" with "sad" in contrast to the German and English participants. Although similar facial displays portrayed by the model person of Geminoid F achieved higher recognition rates overall, portraying *fearful* has been similarly difficult for the model person. We conclude that improving the android's expressiveness especially around the eyes would be a useful next step in android design. In general, these results could be complemented by an evaluation of dynamic facial expressions of Geminoid F in future research.

Android science; Affective computing; Facial expressions; Online survey

I. Introduction and Motivation

Researchers in the field of social robotics (e.g. [1], [2]) 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. [3]). In the field of "Android Science" [4] 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.

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[5].

The remainder of this paper is structured as follows: The following section discusses related work giving rise to two research questions. In Section III two online surveys are described and their results are presented. These are then summarized in Section IV, before in Section V general conclusions are drawn.

II. Related Work

"Geminoids" [6] 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 [7; 8], which are similarly designed to let people associate them with humans, the outer appearances of android robots such as "Geminoid HI-1"[6] 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" [4], because they provide "a key testing ground for social, cognitive, and neuroscientific theories." [9]

"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 [10].

"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 (cf. Figure 1).

This work was partially supported by Grant-in Aid for Scientific Research (S), KAKENHI (20220002) and by a post-doctoral fellowship of the Japan Society for the Promotion of Science.

978-1-61284-082-6/11/$26.00 ©2011 IEEE

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 [11], namely anger, fear, happiness, sadness, and surprise, have been realized with the iCat robot [12]. 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 [13]. This particular emotion, however, is known to be difficult in the human case as well [14] and the authors of [13] 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 [15]. In essence, Japanese observers tend to weight cues in the eyes more than cues displayed in the mouth, whereas American people seemed to show the opposite tendency when being asked to judge facial displays of emotions. In effect, a big smile without neutral eyes was rated as less happy by Japanese participants than compared to American participants. This effect has even been confirmed for stylized facial icons (also called emoticons) which are often used in internet text mails. In addition, the uncanny valley is also being investigated based on the animation of virtual characters; e.g., [16].

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?

III. THE TWO ONLINE SURVEYS

The android robot Geminoid F was built to closely resemble her human model person’s outer appearance (cf. 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. Even when both of them are moving slightly, this difficulty remains.

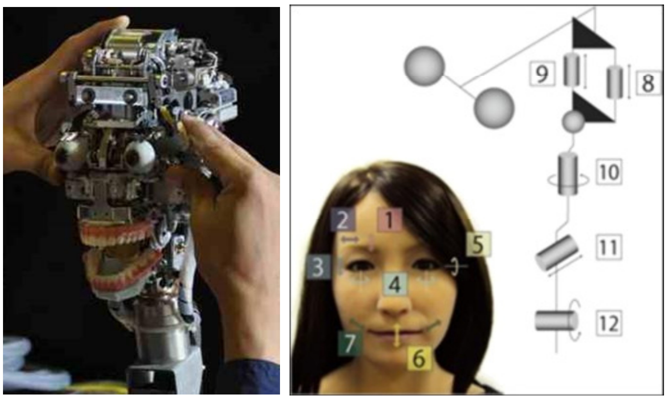

Figure 2. Geminoid F’s interal configuration; left, its head without skin and hair, right, the distributions of a total of 12 degrees of freedom are (1) both eye brows up and down, (2) both eye brows left and right, (3) both eye lids open and close, (4) both eyes left and right, (5) both eyes up and 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) head back and forth, and (12) upper body front and back

Geminoid F’s smooth silicon skin and sophisticated internal design (cf. 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 (cf. 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 (as well as its predecessor) cannot stand up, perform gestures, or walk although its arms and legs look similarly human-like as its upper torso and its face.

The repeated—but so far unjustified—claim that it were able to “laugh, smile, and exhibit other facial expressions more naturally than Ishiguro’s previous androids” [16] motivated us to start investigating the emotional expressiveness of Geminoid F empirically.

A. *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] and indeed we plan to extend our evaluation of Geminoid F’s facial display by presenting videos in the future. Nevertheless we believe to already generate valuable insights by letting people from different cultures evaluate still images of its face at first.

Six digital pictures of Geminoid F’s face were taken featuring the basic emotions angry, fearful, happy, sad, and surprised [11] 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.

TABLE I. 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	wütend	怒り（いかり）
fearful	ängstlich	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	traurig	悲しみ（かなしみ）
surprised	überrascht	驚き（おどろき）

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.

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, Japanse, and English language. The six emotion labels with their translations into German and Japanese are presented in Table I. 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.

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 I), 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.

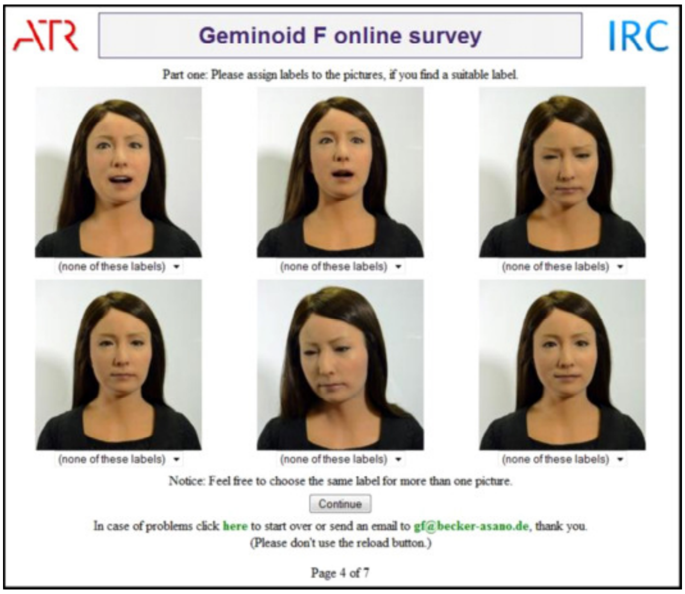

Figure 3. Screenshot of the first online survey presenting Geminoid F's portrayal of five basic emotions plus a supposedly neutral expression; from left to right, top to bottom: fearful, surprised, angry, neutral, sad, happy

Part one of the survey starts on page four with the instruction to please "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.

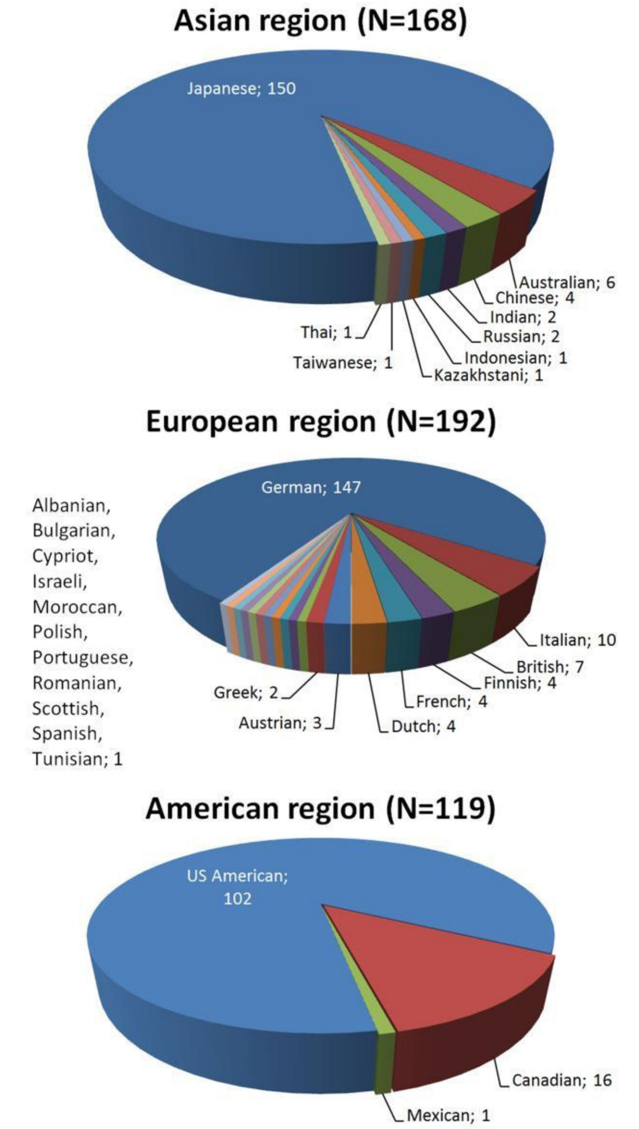

Figure 4. All participants group by region of origin according to the stated nationality of each participant in the first survey

- 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."[1] 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 5th 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 nation-nalities into those from the Asian region (N=168), those from the European region (N=192), and those from the American region (N=119); cf. Figure 4. We do not distinguish native speakers from non-native speakers in the further analysis.

[1] The results of the second part of the two surveys are still being analyzed and, thus, not further discussed in this paper.

TABLE II. 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	4,6	7,9	1,9

* above chance level of 14.8%

2) *Results*

A global confusion matrix (N=479) is presented in Table II. 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 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 [20] 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 III) assign "sad" (50.6%) more often than "angry" (38.1%) to the picture *angry*. In addition, they show the least confu-sion 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).

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

TABLE III. 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 IV. 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 V. 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

American region only (N=119)						
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

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 V) 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 [15] 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 [20].

B. 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, (1) if the intercultural differences would reoccur, and (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 [14 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 camera is triggered 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 (cf. Figure 5, bottom row).

1) Experimental procedure

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

2) Results

A global confusion matrix (N=251) is presented in Table VI. 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

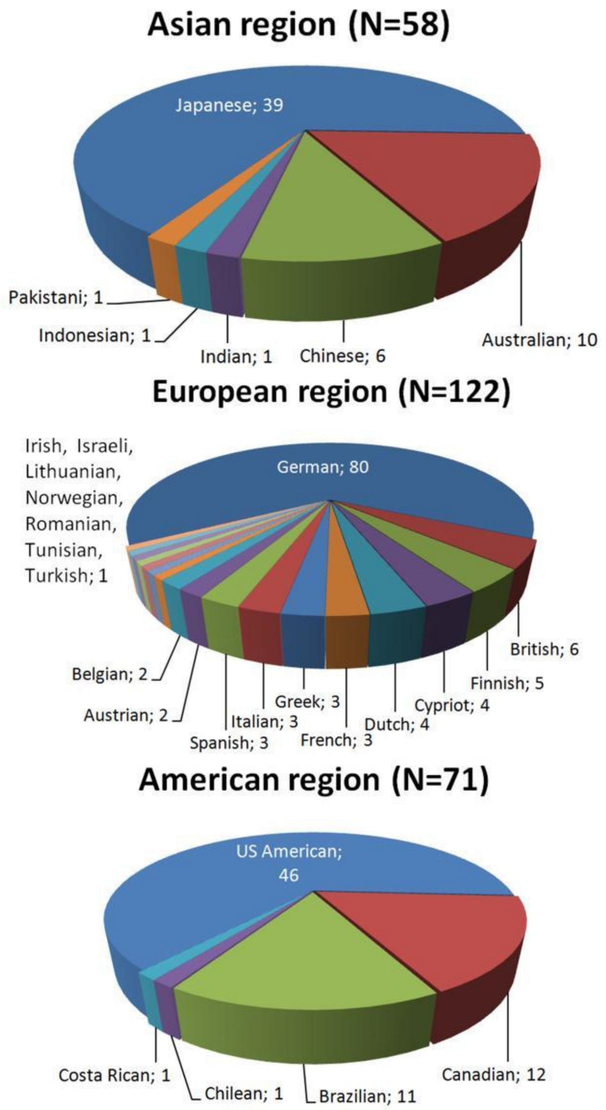

Figure 6. All participants group by region of origin according to the stated nationality of each participant in the second survey

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.

Thus, it seems as if it was most difficult for our model person to portray *fearful* as a facial expression and easiest to

TABLE VI. 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

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 (cf. Figure 6) to check for intercultural 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 VII, Table VIII, and Table IX, respectively.

The following intercultural differences reappear:

1. The Asian participants (N=58, cf. Table VII) show the lowest agreement on the *fearful* expression (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 VII). Their level of agreement is the lowest of all three groups ($\kappa_{MP,Asian}$=0.767).
2. The 122 participants of the European region (cp. Table VIII) 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 IX) 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).

IV. SUMMARY

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 V).

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 IX, 97.2%).

TABLE VII. 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 VIII. 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	1,6	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 IX. 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

V. CONCLUSIONS

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 androids 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 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 (cf. 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 the general difficulty portraying fear in static facial displays [13; 14].

With respect to our second goal, Geminoid F's rather limited ability to change its face around the eyes (cf. 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 [15]. 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 V, 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 second survey. The model person's portrayals show much more variations around the eyes (cf. 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 as reported in [23].

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. We have to admit that the intuitive design of facial expressions was only our first attempt to let Geminoid F convey emotions. A more systematic approach in designing facial expressions as well as complementing these results by an evaluation of non-static facial displays of emotions remain interesting opportunities for 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] Dautenhahn, K.., Nourbakhsh, I., and Fong,T. "A Survey of Socially Interactive Robots." *Robotics and Autonomous Systems* 42 (2003): 143-166.

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

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

[4] Ishiguro, H. "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.

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

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

[7] Kanda, T., Hirano, T., Eaton, D., and Ishiguro, H. „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.

[8] Kanda, T., Ishiguro, H., Ono, T., Imai, M., and Mase, K. „Development and Evaluation of an Interactive Robot 'Robovie'." *IEEE International Conference on Robotics and Automation.* 2002. 1848-1855.

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

[10] Becker-Asano, C., Ogawa, K., Nishio, S., and Ishiguro, H. "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.

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

[12] Bartneck, C., Reichenbach, J., and Breemen, A. "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.

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

[14] Ekman, P. "Facial Expressions." Chap. 16 in *Handbook of Cognition and Emotion*, 301-320. John Wiley & Sons, 1999.

[15] Yuki, M., Maddux, W., and Masuda, T. "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.

[16] Tinwell, A., Grimshaw, M., Nabi, D.A., and Williams, A. "Facial expression of emotion and perception of the Uncanny Valley in virtual characters." *Computers in Human Behavior* (2010), in press.

[17] Guizzo, E. *Geminoid F: Hiroshi Ishiguro Unveils New Smiling Female Android.* IEEE Spectrum. April 03, 2010. http://spectrum.ieee.org/automaton/robotics/humanoids/040310-geminoid-f-hiroshi-ishiguro-unveils-new-smiling-female-android (accessed October 26, 2010).

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

[19] Tolksdorf, J., Becker-Asano, C., and Kopp, S.. "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.

[20] Koda, T. "Cross-Cultural Study of Avatars' Facial Expressions and Design Considerations Within Asian Countries." *Proc. of Intercultural Collaboration.* Springer, 2007. 207-220.

[21] Cohen, J. "A Coefficient of Agreement for Nominal Scales." *Educational and Psychological Measurement* 20 (1960): 37-46.

[22] MacDorman, K. F., Vasudevan, S. K., and Ho, C. "Does Japan really have robot mania? Comparing attitudes by implicit and explicit measures." *AI & Society*, 2009: 485-510.

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