Available online at www.sciencedirect.com
SCIENCE DIRECT®
Int. J. Human-Computer Studies 62 (2005) 179–192
International Journal of Human-Computer Studies
www.elsevier.com/locate/ijhcs

Subtle emotional expressions of synthetic characters

Christoph Bartneck[a,*], Juliane Reichenbach[b]

[a]*Department of Industrial Design, Eindhoven University of Technology, Industrial Design, Den Dolech 2, 5600 MB Eindhoven, The Netherlands*
[b]*Technische Universität Berlin, Institute of Aeronautics and Astronautics, Marchstr. 12, 10587 Berlin, Germany*

Available online 2 January 2005

Abstract

This study examines the influence of the geometrical intensity of an emotional facial expression on the perceived intensity and the recognition accuracy. The stimuli consisted of synthetic faces at ten geometrical intensity levels in each of the five emotional categories. A curve–linear relationship was found between geometrical and perceived intensity. Steps of 20% geometrical intensity appear to be appropriate to enable the participants to distinguish the intensity levels. At about 30% geometrical intensity the recognition accuracy reached a level that was not significantly different from each emotions maximum recognition accuracy. This point indicates a categorical perception of the facial expressions. The results of this study are of particular importance for the developers of synthetic characters and might help them to create more subtle characters.

1. Introduction

Many synthetic characters are used for entertainment, communication, and work. They range from movie stars (Thomas and Johnson, 1981) and pets (Sony, 1999)

*Corresponding author. Tel.: +31402475175; fax: +31402475376.
E-mail addresses: christoph@bartneck.de (C. Bartneck), Juliane.Reichenbach@ilr.tu-berlin.de (J. Reichenbach).

1071-5819/$ - see front matter
doi:10.1016/j.ijhcs.2004.11.006

Fig. 1. Aibo, eMuu and microsoft agent.

(see Fig. 1) to helper agents (Bell et al., 1997) (see Fig. 1) and avatars for virtual cooperative environments (Isbister et al., 2000). Characters can also have a physical body, e.g. robots. The range of robots is very wide and therefore this paper focuses on robots that interact with humans and not on industrial or military robots. The interesting robots for this study help the elderly (Hirsch et al., 2000), support humans in the house (NEC, 2001), improve communication between distant partners (Gemperle et al., 2003) and are research vehicles for the study on human–robot communication (Okada, 2001; Breazeal, 2003). A survey of relevant characters is available (Bartneck, 2002).

The ability to communicate emotions is essential for a natural interaction between characters and humans because it is not possible not to communicate. The absence of a character's emotional expressions could already be interpreted as indifference towards the human. Therefore, it is important that characters express their emotional state. Some of these characters can express emotions to improve the interaction between the character and the user (Bartneck, 2003; Breazeal, 2003) (see Fig. 1) or to visually support synthetic speech (CSLU, 1999). The CWI institute in Amsterdam developed a talking screen character that is able to express emotions based on an emotion disc (Ruttkay et al., 2000).

Three parameters and their interaction are important for the design of emotional expressions for characters: geometrical intensity, perceived intensity and recognition accuracy. We will now take a closer look at the three parameters.

1.1. Geometrical intensity

The synthetic face has certain components, such as eyebrows and a mouth, which can be manipulated. Usually, a maximum for each emotional expression is defined by reproducing already validated faces, such as the well-known Ekman faces (Ekman and Frieser, 1976). The spatial difference of each component between the neutral and the maximum expression is then divided into equal parts. To express 30% happiness, for example, the components are moved 30% of distance between neutral and maximum.

1.2. Perceived intensity

Humans are able to judge the intensity of a human's or character's expression. Several studies have been carried out in which participants evaluated expressions (Etcoff and Magee, 1992; Hess et al., 1997).

1.3. Recognition accuracy

Each emotional expression has a certain distinctness, which can be measured by the recognition accuracy of humans observing the expression. In this study, when we refer to recognition accuracy, we do not mean the differentiability between intensity levels within one emotion. We mean the differentiability between emotion categories measured as recognition rate. In such recognition tests, the participants have to identify which emotion was expressed. Low-intensity expressions are usually less distinct (Etcoff and Magee, 1992; Bartneck, 2001) but can play an important role in human communication (Suzuki and Bartneck, 2003).

1.4. Focus of this study

We now take a look at the relationships of these three parameters. Clearly, the geometrical intensity has a direct influence on the perceived intensity and the recognition accuracy of the expression. The closer the emotional expression is to its maximum the higher is the perceived intensity of the expression. However, it cannot be assumed that this relationship is as simple as the function *perceived intensity* = *geometric intensity*. A 30% geometrical intense expression of happiness may not be perceived to be 30% intense or correctly recognized in 30% of the cases. This study attempts to shed some light on this particular relationship.

1.5. Research questions

Based on the background given above we would like to define the three research questions of this study:

1. What is the relationship between the geometrical and perceived intensity?
2. What is the influence of the geometrical intensity on the recognition accuracy of the expression?
3. What is the relationship between perceived intensity and the recognition accuracy of the expression?

1.6. Relevance of this study

With this study we hope to provide a better insight into the perception of the emotional expressions of synthetic characters. Synthetic characters are used to an increasing degree in computer games, virtual environments, or robots. The results could be of great interest to the developers of these characters and might help them to gain more control of their designs.

1.7. Related work

Hess et al. (1997) studied the relationship between the physical intensity of an emotional expression and the perceived intensity and the recognition of that expression using pictures of natural faces as stimuli. They changed the physical intensity by combining a neutral face with an intense expression of an emotion using graphic morphing software in 20% steps. This is problematic since it is impossible to control how the morphing software merges the pictures and therefore generates steps of 20% intensity.

Hess et al. found a significant main effect of physical intensity for both perceived intensity and recognition accuracy. With increasing physical intensity, perceived intensity increased in a linear way. For recognition accuracy a significant linear and quadratic trend was found. Furthermore, task difficulty was rated lower for higher intensities. Besides, happiness was the easiest to recognize and it was recognized the best: almost 100% correct identifications even for low physical intensities. This happy face advantage has been reported before (Ekman and Friesen, 1971). Hess et al. argue that their results support the theory of categorical perception only for happiness, not for the other emotions.

In our study, we hope to replicate their results regarding the perceived intensity with different stimuli, namely schematic faces. Regarding the recognition accuracy, we want to find out if we can support a categorical or a dimensional perception of emotional expressions. In the present study, however, we do not use the critical morphing procedure to create different intensity levels. Instead, we use an animation tool as described in the Methodology section below.

Differences in identification of emotions between natural and synthetic faces was researched by Kätsyri et al. (2003). They found that emotional expressions shown by a synthetic talking head that they developed (Frydrych et al., 2003) was recognized worse than emotional expressions displayed by natural faces. This suggests that synthetic faces are not an adequate alternative for emotions research. On the other hand, there is research that shows that emotional expressions by synthetic faces are recognized as well or even better than emotions on natural faces (Katsikitis, 1997; Bartneck, 2001).

Another aspect of emotional expressions is of interest to this study. The space of human emotions is frequently modeled either with dimensions, such as arousal and valence (Schlossberg, 1954; Osgood et al., 1957; Russel, 1979; Hendrix et al., 2000) or in categories such as happiness and sadness (Ekman et al., 1972; Izard, 1977; Plutchik, 1980). It has already been shown that a two-dimensional space is insufficient to accurately model the perception of emotional facial expressions (Schiano et al., 2000). Etcoff and Magee (1992) showed that emotional facial expressions are perceived categorically.

They used line drawings of emotional faces to study the relationship between physical intensity of an emotional facial expression and the recognition. They had their subject identify an emotion on 11 evenly spaces facial expression continua. The continua were based on merging either a neutral face with an emotional expressive face or on merging two faces with different emotional expressions. It was found that emotions were perceived categorically, except for surprise. That means that small physical differences in emotional facial expressions are easier to distinguish when at boundaries between emotions and harder when within one emotion category. In our study, we only use neutral—emotion continua for 5 emotions. We expect to find a boundary for each emotion where it is possible to recognize an expression as a particular emotion.

2. Methodology

We reproduced the method used by Hess et al. (1997) to allow a comparison of the results with two exceptions. First, we used an 11-point scale instead of a continuous slider, which should not have any effect on the validity of the comparison. Second, we used 10% steps of geometrical intensity instead of Hess' et al. 20% steps. This offers more fine-grained analysis, while still enabling us to compare the results by only considering every second intensity step.

Unlike Hess et al. who did their study with morphed natural faces, we used schematic faces. These faces differed in the percentage of the angles of the mouth, the eyebrows and the eyes from the neutral position (0%) to the extreme position (100%).

2.1. Subjects

In total, 24 men and 7 women ranging from 18 to 31 years of age ($M = 21.32$, $SD = 3.41$) participated in the experiment. They were paid for their participation.

2.2. Design

We used a 5 (emotion) $\times$ 10 (intensity) within subject design. The dependent variables were perceived intensity, recognition accuracy and task difficulty.

Perceived intensity: participants were asked to rate the intensity of the emotions anger, contempt, disgust, fear, happiness, sadness, and surprise on 11-point scales for each presented schematic face. Each scale was labeled with an emotion and anchored with "not intense at all" and "very intense".

Recognition accuracy: the intended emotion was considered correctly identified if it received a highest rating on the correct scale. The recognition rate defines the distinctness of an emotion.

Task difficulty: the task difficulty had to be rated on a 5-point scale anchored by the labels "very easy" and "very difficult".

2.3. Material

We used pictures of schematic faces displaying five basic emotions anger, fear, happiness, sadness, and surprise (see Fig. 2). The mouth, eyebrows and eyes were manipulated to create the emotions. However, the nose was not manipulated. This could be a reason for the low recognition rates for the disgust expression in a pretest. A similar problem was found in the pilot test of a previous study (Bartneck, 2001). Because of the low recognition accuracy in a pretest the disgust expressions were excluded from the experiment. The faces were developed at Centrum voor Wiskunde en Informatica (CWI), Amsterdam (Ruttkay et al., 2000).

The intensity of the emotion was varied by manipulating the angle of the eyebrow, the mouth and the eyes. The intensity of each expression started with 10% of the maximum angle, and was increased by 10% steps, ending with the full-blown emotion at 100% geometric intensity, thus resulting in 10 faces per emotion. To create those faces with the intensity steps, the CharToon software was used (Noot and Ruttkay, 2000).

2.4. Procedure

The experiment took place at TU Eindhoven and lasted about 30 min. After the participants read instructions on a computer screen they were shown the most intense faces and the neutral face before they entered a practice session. In the practices session, they had to evaluate three different faces. The participants were shown a face randomly for five seconds on a computer screen. Afterwards the face was replaced by a questionnaire (see Fig. 3). They had to fill in seven intensity rating

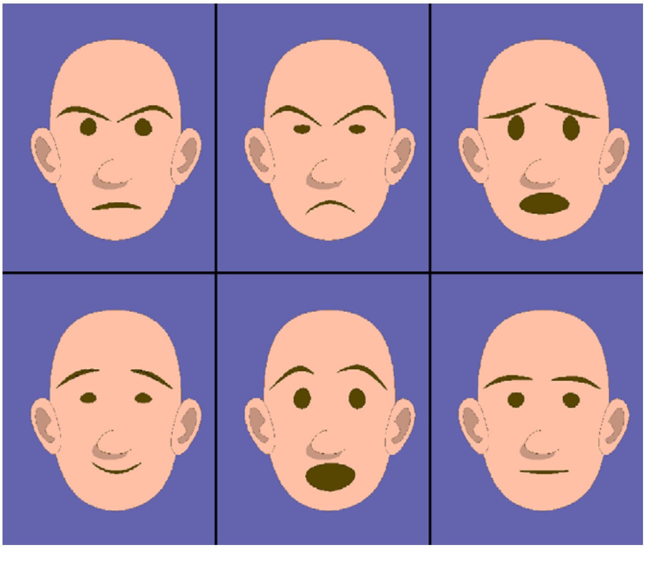

Fig. 2. The five most intense faces and the neutral face.

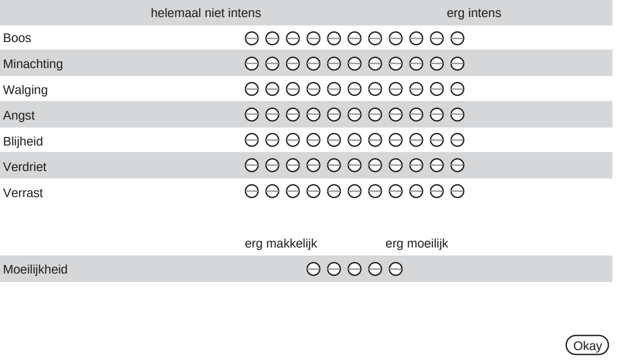

Fig. 3. The questionnaire. On top: *anger*, *contempt*, *disgust*, *fear*, *happiness*, *sadness*, and *surprise*. Below: *difficulty*. The scale markings are on top: *not intense at all—very intense* and at the bottom: *very easy—very difficult*.

scales and one difficulty scale. They could not continue before all scales were filled in. After the practice session the participants could ask questions about the process and the software of the experiment. Afterwards, the experiment started. The structure of the experiment is identical to the practice session. However, now the participants had to evaluate all 50 faces. After the experiment the participants were debriefed.

3. Results

3.1. Relationship between geometrical intensity and perceived intensity

A 5 (emotion) $\times$ 10 (geometric intensity) repeated measures ANOVA was conducted. The geometric intensity had a significant effect on the perceived intensity ($F(9, 270) = 129.069$; $p < .001$). Faces with high geometric intensity received higher intensity ratings. The intensity ratings differed significantly between emotions ($F(4, 120) = 37.135$; $p < .001$). There was a significant interaction effect between emotion and intensity ($F(36, 1080) = 4.274$; $p < .001$). See Fig. 4 for the mean intensity ratings.

To see which consecutive intensities of a single emotion differed, we calculated the repeated contrasts for each emotion. F and p values for the significant differences between consecutive levels can be seen in Table 1.

3.2. Relationship between geometrical intensity and recognition accuracy

A two-way ANOVA with emotion and geometric intensity as within-subjects factors was performed. There was a significant main effect for geometric intensity ($F(9, 270) = 43{,}314$; $p < .001$). Faces with high geometric intensity were recognized better. The intensity ratings differed significantly between emotions ($F(4, 120) = 72{,}552$;

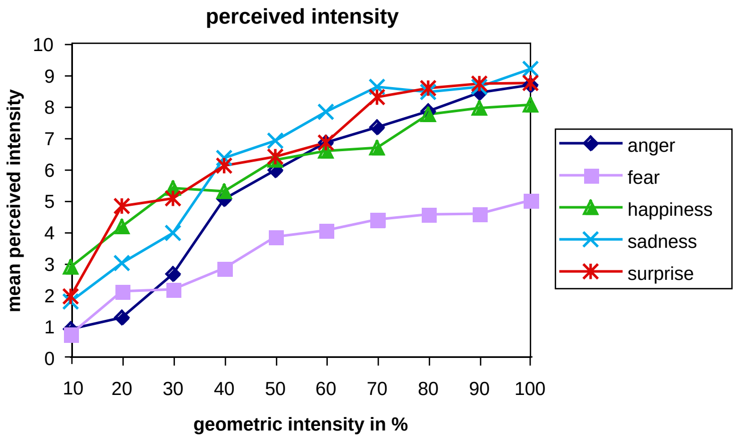

Fig. 4. Mean intensity ratings of the intended emotion as a function of the geometric intensity of the face.

Table 1
Significant differences in perceived intensity between consecutive geometric intensities for each emotion

Emotion	Intensity in %	*F*	*p*
Anger	20 vs. 30	7.603	.010
	30 vs. 40	13.073	.001
Fear	10 vs. 20	8.489	.007
Happiness	10 vs. 20	7.576	.010
	20 vs. 30	6.803	.014
	70 vs. 80	4.218	.049
Sadness	10 vs. 20	8.766	.006
	30 vs. 40	19.107	.000
	50 vs. 60	9.400	.005
	60 vs. 70	7.872	.009
Surprise	10 vs. 20	29.315	.000
	60 vs. 70	14.030	.001

$p < .001$). There was a significant interaction effect of emotion × intensity ($F(36, 1080) = 4.302$; $p < .001$). See Fig. 5 for mean recognition rates.

To find out for what intensities the recognition rate was significantly lower compared to the maximum intensity of 100% for each emotion, we tested simple contrasts with the highest intensity level. See the *F* and *p* values for all significant differences in Table 2.

We recoded the recognition rate by dividing the intensity rating for the intended emotion by the sum of the intensity ratings on all 7 emotion scales. Thus, we got a number between 0 and 1 for each facial expression that told us more about the

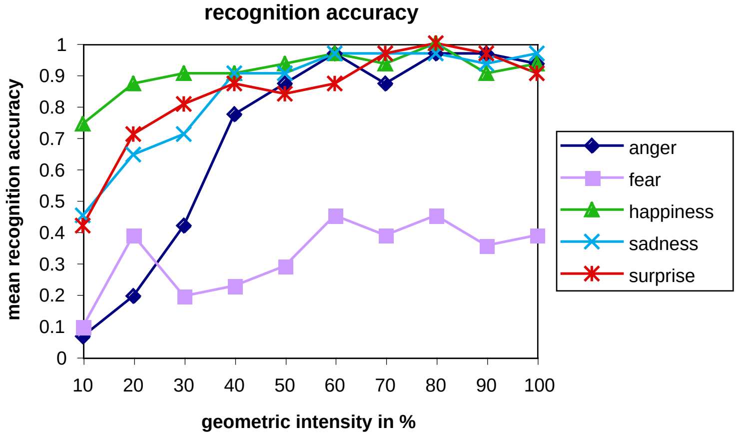

Fig. 5. Mean recognition of the intended emotion as a function of the geometric intensity of the face.

Table 2
Significant differences in recognition rate between the highest geometric intensity of 100% and lower intensities for each emotion

Emotion	Intensity in %	*F*	*p*
Anger	10	202.500	.000
	20	86.250	.000
	30	25.430	.000
Fear	10	9.346	.005
Sadness	10	32.000	.000
	20	11.029	.002
	30	10.435	.003
Surprise	10	18.544	.000

recognition accuracy of an expression than a mere 0 vs. 1 coding of either being correctly identified or not. Confusion with other emotions are taken into account. Fig. 6 shows the mean recoded recognition accuracy.

We conducted a two-way ANOVA with emotion and geometric intensity as within-subjects factors. There was a significant main effect for geometric intensity ($F(9, 270) = 40.525$, $p < .001$). Faces with high geometric intensity were recognized better. The intensity ratings differed significantly between emotions ($F(4, 120) = 71.598$, $p < .001$). There was a significant interaction effect emotion $\times$ intensity ($F(36, 1080) = 3.863$, $p < .001$). See Fig. 6 for mean recoded recognition rates. It can be seen that all emotions were recognized above chance level. When

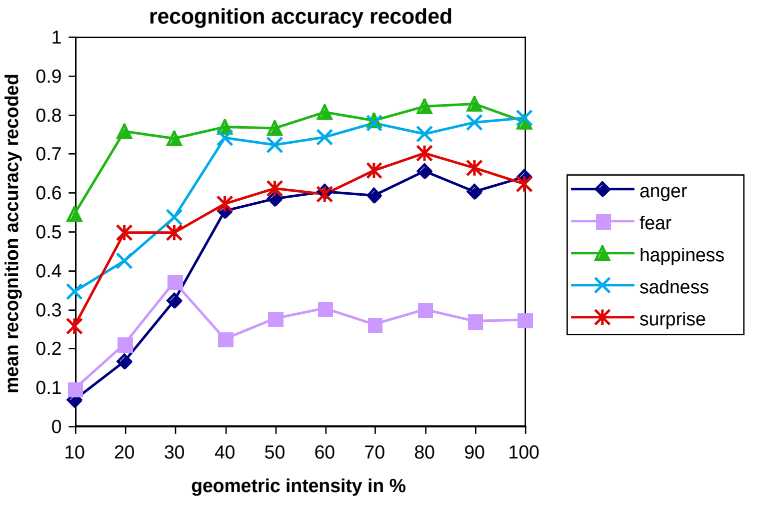

Fig. 6. Mean recoded recognition of the intended emotion as a function of the geometric intensity of the face.

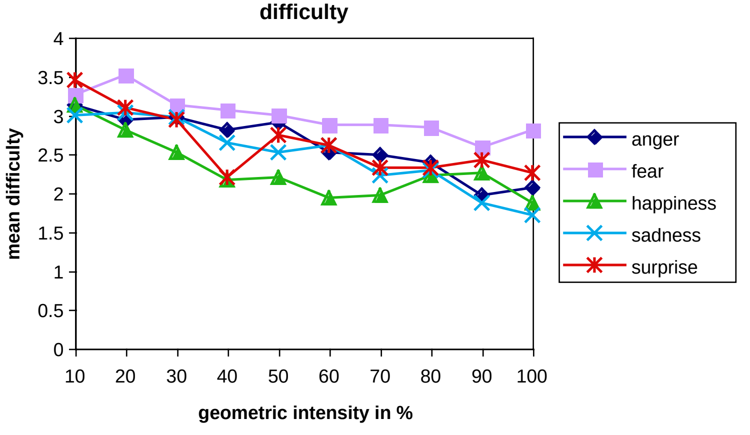

Fig. 7. Mean difficulty ratings of the intended emotion as a function of the geometric intensity of the face.

comparing Figs. 5 and 6 it can be seen that with this recoded recognition a differentiated picture of the recognition accuracy of the emotions emerges.

3.3. Relationship between geometrical intensity and difficulty

A 5 (emotion) $\times$ 10 (geometric intensity) repeated measures ANOVA was conducted. There was a significant main effect for geometric intensity ($F(9, 270) = 13{,}982$; $p<.001$) and emotion ($F(4, 120) = 13{,}597$; $p<.001$. See Fig. 7 for mean difficulty ratings.

Table 3
Significant differences in difficulty rating between the highest geometric intensity of 100% and lower intensities for each emotion

Emotion	Intensity in %	*F*	*p*
Anger	10	15.528	0.000
	20	7.237	0.012
	30	12.099	0.002
	40	9.857	0.004
	50	14.162.	0.001
Fear	20	8.307	0.007
Happiness	10	20.462	0.000
	20	9.478	0.004
	30	7.290	0.011
Sadness	10	20.815	0.000
	20	27.678	0.000
	30	14.729	0.001
	40	11.991	0.002
	50	14.119	0.001
	60	16.243	0.000
	70	8.930	0.006
	80	6.083	0.020
Surprise	10	23.309	0.000
	20	12.071	0.002
	30	4.321	0.046

To see if it was any more difficult to judge a low-intensity emotion we tested simple contrast with the highest intensity. See Table 3 for the significant results.

4. Discussion

4.1. Relationship between geometrical intensity and perceived intensity

The perceived intensity increased with higher geometric intensity. Given geometrical intensity level steps of 10% the consecutive perceived intensity levels differed mainly at low geometrical intensity levels but not at the higher levels. Given geometrical intensity level steps of 20%, as used by Hess et al. (1997), different results emerge. All consecutives levels differed significantly for anger. For happiness, the first 4 consecutive levels differed, but the 80% level was not significantly different from the 100% level. For sadness, all consecutive levels but the 60% vs. 80% differed significantly. For surprise, the 20% vs. 40% level and the 60% vs. 80% level were significantly different. For fear, none of the consecutive levels differed significantly. It seems that the 10% geometrical intensity level steps are too small to be discriminated. For a practical application it appears useful to use 20% steps to ensure that the user can distinguish the different levels.

Fig. 4 showed the relationship between geometrical and perceived intensity. The graph shows that this relationship cannot be modelled by a simple function, such as *perceived intensity* = *geometric intensity* but that a curve-linear trend is visible consisting of a linear trend ($F(1, 30) = 476.89$, $p < .001$) and a quadratic trend ($F(1, 30) = 64.532$, $p < .001$).

4.2. Relationship between geometrical intensity and recognition accuracy

The recognition accuracies for each emotion increased with the geometric intensity up to a certain point where the recognition accuracy did not significantly differ anymore from the recognition accuracy at the maximum geometrical intensity of each emotion. This point was reached at 40% geometrical intensity for anger and sadness and at 20% geometrical intensity for fear and surprise. The recognition accuracy for happiness already reached this point at 10%. This happy-face bonus was previously observed (Ekman and Friesen, 1971).

4.3. Relationship between geometrical intensity and difficulty

Although participants were able to recognize the emotions even at low intensities, it was still more difficult for them compared to high intensity expressions. This result is in line with our expectations. Fear remains a problematic emotional expression because it was difficult to identify at low and high intensity. In addition it was the most difficult emotion to identify.

5. Conclusions

We conducted a study of synthetic facial expressions that explored the relationship of geometrical intensity, perceived intensity and recognition accuracy. Our results show that it is possibly to communicate emotions also at low intensity levels and thereby enable characters and robots to act more subtle.

Fear and happiness remain two special emotional categories for facial expressions. The happy-face advantage shows how sensitive humans are in perceiving positive expressions. Since the repertoire of positive expressions is limited to smiling it is good to know that it is also correctly recognized at low intensities. Fear is a problematic expression since it is difficult to recognize and to judge its intensity.

The results of our study indicate that emotional expressions might be perceived categorically. The strong increase of recognition accuracy at about 30% geometrical intensity could be interpreted as categorical perception as described by Etcoff and Magee (1992). However, we only explored facial expression between neutral face and most intense face for each emotion and not between two different emotions. Therefore, our results can only be an indication.

Acknowledgements

We would like to thank Han Noot and Zsofie Ruttkay for providing the faces for this study. Kees Overbeeke was of great help for the methodology and analysis.

References

Bartneck, C., 2001. How convincing is Mr. Data's smile: affective expressions of machines. User Modeling and User-Adapted Interaction 11, 279–295.

Bartneck, C., 2002. eMuu—an embodied emotional character for the ambient intelligent home. Ph.D. thesis, Eindhoven University of Technology, Eindhoven, unpublished.

Bartneck, C., 2003. Negotiating with an embodied emotional character. Paper presented at the Design for Pleasurable Products Conference (DPPI2004), Pittsburgh.

Bell, G., Ling, D., Kurlander, D., Miller, J., Pugh, D., Skelly, T., Stankosky, A., Thiel, D., Dantzich, M., Wax, T., 1997. Lifelike computer characters: The persona project at microsoft research. In: Bradshaw, J.M. (Ed.), Software Agents. AAAI Press, London, pp. 191–222.

Breazeal, C., 2003. Designing Sociable Robots. MIT Press, Cambridge.

CSLU, 1999. CSLU Toolkit, from http://cslu.cse.ogi.edu/toolkit/

Ekman, P., Friesen, W.V., 1971. Constants across cultures in the face and emotion. Personality and Social Psychology 17 (2), 124–129.

Ekman, P., Friesen, W.V., Ellsworth, P., 1972. Emotion in the Human Face: Guidelines for Research and an Integration of Findings. Pergamon Press, New York.

Ekman, P., Frieser, W., 1976. Pictures of Facial Affects. Consulting Psychologist Press, Palo Alto.

Etcoff, N.L., Magee, J.J., 1992. Categorical perception of facial expressions. Cognition 44, 227–240.

Frydrych, M., Kätsyri, J., Dobsik, M., Sams, M., 2003. Toolkit for animation of finnish talking head. Paper presented at the International Conference on Audio-Visual Speech Processing (AVSP 2003), St. Jorioz.

Gemperle, F., DiSalvo, C., Forlizzi, J., Yonkers, W., 2003. The hug: a new form for communication. Paper presented at the Designing the User Experience (DUX2003), New York.

Hendrix, J., Ruttkay, Z., Hegen, P.T., Noot, H., Lelievre, A., Ruiteer, B.D., 2000. A facial repertoire for avatars. Paper presented at the Workshop on Interacting Agents, Enschede.

Hess, U., Blairy, S., Kleck, R.E., 1997. The intensity of emotional facial expressions and decoding accuracy. Journal of Nonverbal Behavior 21 (4), 241–257.

Hirsch, T., Forlizzi, J., Hyder, E., Goetz, J., Stroback, J., Kurtz, C., 2000. The ELDeR Project: social and emotional factors in the design of eldercare technologies. Paper presented at the Conference on Universal Usability, Arlington.

Isbister, K., Nakanishi, H., Ishida, T., Nass, C., 2000. Helper agent: designing an assistant for human–human interaction in a virtual meeting space. Paper presented at the Conference on Human Factors in Computing Systems (CHI2000), Den Hague.

Izard, C.E., 1977. Human Emotions. Plenum Press, New York.

Katsikitis, M., 1997. Classification of facial expressions of emotions: a multidimensional scaling approach. Perception 26, 613–626.

Kätsyri, J., Klucharev, V., Frydrych, M., Sams, M., 2003. Identification of synthetic and natural emotional facial expressions. Paper presented at the International Conference on Audio-Visual Speech Processing (AVSP 2003), St. Jorioz.

NEC, 2001. *PaPeRo*, from http://www.incx.nec.co.jp/robot/PaPeRo/english/p_index.html

Noot, H., Ruttka, Z., 2000. CharToon 2.0 Manual (No. INS-R0004). Amsterdam: CWI.

Okada, M., 2001. Muu: artificial creatures as an embodied interface. Paper presented at the ACM Siggraph 2001, New Orleans.

Osgood, C.E., Suci, G.J., Tannenbaum, P.H., 1957. The Measurements of Meaning. University of Illinois Press, Champaign.

Plutchik, R., 1980. Emotion: A Psycho Evolutionary Synthesis. Harper & Row, New York.

Russel, J.A., 1979. Affective space is bipolar. Journal of Personality and Social Psychology 37, 345–356.
Ruttkay, Z., Noot, H., Ruiteer, B.D., Hagen, P.T., 2000. CharToon Faces for the Web. Paper presented at the Ninth International WWW conference, Amsterdam.
Schiano, D.J., Ehrlich, S.M., Rahardja, K., Sheridan, K., 2000. Face to InterFace: facial affect in (Hu)Man and Machine. Paper presented at the CHI 2000, Den Hague.
Schlossberg, H., 1954. Three dimensions of emotion. Psychological review 61, 81–88.
Sony, 1999. Aibo, from http://www.aibo.com
Suzuki, N., Bartneck, C., 2003. Subtle Expressivity of Characters and Robots. Paper presented at the CHI2003, Fort Lauderdale (Extended Abstracts).
Thomas, F., Johnson, O., 1981. Disney animation: the illusion of life: Walt Disney Productions.