

# The Influence of Naturalness, Attractiveness and Intensity on Facial Emotion Recognition\*

Kerstin Limbrecht, Stefanie Rukavina, Andreas Scheck, Steffen Walter,  
Holger Hoffmann, Harald C. Traue  
University of Ulm, Ulm, Germany

Understanding the determinants facial emotion recognition is still one of the main topics in emotion research. Mimic expressions are not only a representation of feelings caused by emotions, but are also an important communication channel in social interaction. Research of the last 20 years could show that emotion recognition abilities differ within and between individuals. Moreover, every basic emotion seems to be processed differently in the human brain. Underlying processes are still not clear. In this study, the common practice of presenting pictures of faces showing basic emotions to participants was used in a more methodological sense. Possible determining factors in facial emotion recognition, like naturalness and intensity of the expressed emotion and the attractiveness of the photographed person, were analyzed.

*Keywords:* facial emotion recognition, validated stimuli, standardized picture set, naturalness, attractiveness, intensity

## Introduction

The recognition of emotions is an essential element of social interaction (Buck, 1984; Ekman, 1993). A common paradigm for studying the ability of emotion recognition is the presentation of photographs with different emotional expressions, usually taken at the time of the strongest expression. Although facial expressions do not necessarily reflect an internal affective state (Kappas, 2003), they are an important communication channel and one of the major information cues in social interaction. Several studies revealed that there are differences within and between individuals in the emotion recognition ability. For example, persons with an anxiety disorder have more false alarms for anger compared to a healthy population (Kessler, Roth, von Wietersheim, Deighton, & Traue, 2007) or women recognize subtle emotions better than men (Hoffmann, Kessler, Eppel, Rukavina, & Traue, 2010). In general, people recognize happiness best and fear worst in most studies (Kessler, Hoffmann, Bayerl, Neumann, Basic, Deighton, & Traue, 2005). As a result of poor emotion recognition performance and problems in social interaction (e.g., social isolation and lack of

---

\***Acknowledgments:** This research was supported by grants from the Transregional Collaborative Research Centre SFB/TRR 62 "Companion-Technology for Cognitive Technical Systems" funded by the DFG (German Research Foundation).

Kerstin Limbrecht, Clinic for Psychosomatic Medicine and Psychotherapy, University of Ulm.

Stefanie Rukavina, Clinic for Psychosomatic Medicine and Psychotherapy, University of Ulm.

Andreas Scheck, Clinic for Psychosomatic Medicine and Psychotherapy, University of Ulm.

Steffen Walter, Ph.D., Clinic for Psychosomatic Medicine and Psychotherapy, University of Ulm.

Holger Hoffmann, Ph.D., Clinic for Psychosomatic Medicine and Psychotherapy, University of Ulm.

Harald C. Traue, Ph.D., professor, Clinic for Psychosomatic Medicine and Psychotherapy, University of Ulm.

empathy) can come up. To prevent these severe consequences, underlying processes in emotion recognition must be understood more in detail in order to create and implement, e.g., standardized diagnostic tests or training procedures in the second step.

Emotion recognition (for basic emotions) seems to be quite complex. On a neurological basis, it can be shown that there are some general processes including face detection and emotion recognition. Here, the “gyrus fusiformis” and “sulcus temporalis inferior” are involved. For detailed emotion recognition, independent neural circuits are assumed. At least for recognition of fear, anger and disgust impressive results can be reported (Murphy & Nimmo-Smith, 2003). Murphy and colleagues conducted a meta analysis with 106 PET (positron emission tomography) and fMRI (functional magnetic resonance imaging) studies. The activation associated with fear, disgust and anger differed significantly. These emotions were most consistently associated with the “amygdale”, the “insula” and “globus pallidus”, and the “lateral orbitofrontal cortex”, respectively. If there are independent structures for the different basic emotions, this should result in very specific and stable differences in emotion recognition for the different basic emotions.

Derived from the above descriptions, it can be summarized that facial expressions of emotion play an important role in communicating the needs and intentions of humans. But how is facial emotion recognition proceeded? In Figure 1, a very simplified scheme of an interaction model can be seen: Person 1 converts her emotional status into a facial expression. Person 2 interprets this facial expression by several decoding mechanisms. This process works vice versa, if person 2 intends to answer.

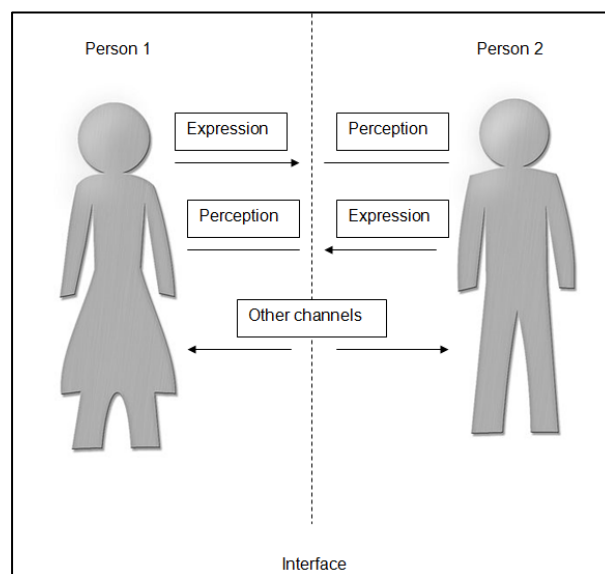


Figure 1. Interaction scheme for facial expressions.

As explained above, there are some basic results in facial emotion recognition research, e.g., that happiness is recognized best and fear worst. So, why is this? There are several determining factors (see Figure 1: other channels) that can simplify or complicate the decoding of facial expressions. In the picture set, the authors wanted to control some of the factors to prevent systematic influences on decoding. Examples for the additional social information examined by the human face are, e.g., physiognomy of the face, age, gender or ethnical background (Langner, Dotsch, Bijlstra, Wigboldus, Hawk, & Knippenberg, 2010). Besides, the recognition rates for each basic emotion itself, other possible co-varying factors can be taken into account.

The relationship between recognition accuracy and the rating of other stimulus information is a relatively unexplored area. So far, no numbers for the power of the different factors can be revealed. Thus, investigating these influencing factors systematically seems to be inevitable. But therefore, those influencing elements have to be detected to create standardized stimulus material which allows the measurement of recognition performance solely. In this study, naturalness and intensity of the expressed emotion and the attractiveness of the photographed person were analyzed. There are some references in literature why these variables may have an impact on facial emotion recognition. For intensity, other studies have shown that the emotion recognition rate depends on the strength of an emotion (Hoffmann et al., 2010; Palermo & Coltheart, 2007). Emotions shown with maximum intensity are recognized best, independent of other factors like gender. Naturalness was studied to prevent lacks in ecological validity (Carroll & Russell, 1997). Ekman and others pointed out that—apart of those prototypic expressions described in EmFACS (Emotional Facial Action Coding System)—other variations of mimic expressions for the basic emotions are common (Ekman, 2003). Elfenbein, Beaupré, Lévesque, and Hess (2007) introduced the dialect theory in which differences in natural mimic for diverse cultural groups are discussed. So, in a standardized picture set, a compromise between naturalness and standardization should be aspired. As a third factor, the attractiveness of the photographed person was considered. The physical appearance of a person influences emotion recognition in that way that people and already children are more interested in attractive persons. Often attractiveness is correlated with facial symmetry. These are rated, e.g., smarter, successful and full of moral fiber (O'Doherty, Winston, Chritchley, Perret, Burt, & Dolan, 2003; Dion, Berscheid, & Walster, 1972). It can be hypothesized that the preference for attractive partners goes along with evolutionary arisen concepts like survival of the fittest (Zaidel, Aarde, & Baig, 2005).

In summary, emotion recognition accuracy can be defined as a result of different processing steps. An involvement of naturalness, attractiveness and intensity is conceivable. Therefore, to understand underlying processes better in a more psychological sense, a systematic analysis of these several possible influencing factors seems to be an important step. Results can provide an important basis for future research. As a first step towards this direction, the analysis and results of the stimulus selection process of the emotion picture set “PFA-U” (Pictures of Facial Affect-Ulm) are presented here.

## **Experimental Setting**

### **General Methods**

A new set of FACS-based pictures for the six emotions (happiness, sadness, disgust, anger, surprise and fear) called PFA-U was created and tested in advance (Limbrecht, Hoffmann, Walter, Gruss, Hrabal, & Traue, under review). Therefore, a total of 2,810 pictures of 48 people out of five different perspectives were taken under standardized conditions (see Figure 2). Frontal pictures were evaluated first in order to select the expression with the highest recognition rate for each person.

Recognition results varied between 71% for fear and 99% for happiness, but were quite homogeneous within the distinct emotional categories. Ninety six pictures (two emotional pictures per person) with recognition rates above 55% were chosen for further analysis and final picture selection. For the final image set consisting of 48 people showing one emotional facial expressions, naturalness, attractiveness and intensity should be concerned to create a highly standardized picture set to document the ability of facial emotion recognition. Pictures can be embedded in the FEEL (Facially Expressed Emotion Labeling) test (Kessler,

Bayerl, Deighton, & Traue, 2002) to ensure a standardized testing of human emotion recognition ability. This new picture set can be used for future studies in terms of emotion management, and especially emotion recognition and stimulation. The additionally available perspectives offer the possibility for a more detailed analysis of recognition processes.



Figure 2. Setup for stimulus development out of five perspectives.

### Participants

A total of 68 medical students (average age = 21.29;  $SD = 2.58$ ) participated in this study. All participants confirmed in writing that their participations were voluntary in accordance with the ethical guidelines (ethics commission 245/08\_UBB/se).

### Hypotheses

As previous research demonstrated (see introduction), different neural circuits can be assumed for the six basic emotions. On the basis of this empirical background, it can be hypothesized that there are differences in the rating of naturalness, attractiveness and intensity as well. Additionally, we assume to find gender differences. Supplementary, we wanted to find out, if there is a systematic correlation between recognition accuracy and naturalness, attractiveness and intensity.

### Procedure

The participants looked at the 96 portrait pictures that passed the first rating process in the following manner: In a randomized order, every picture was displayed together with the correct emotional label on a computer screen. The three scales (naturalness, attractiveness and intensity) were also displayed in addition to the picture to be rated. The facial pictures were evaluated by clicking on the scale values. The intensity was rated on a 11-point Likert scale (0%-100%). This scaling was selected due to the practicability of this classification and the ability to compare our results with data from other studies (Hoffmann et al., 2010). The subjects rated how intensive the emotion was reflected in the facial expression. The attractiveness of the photographed person was rated, using a seven-point Likert scale (from “I think that this person is not attractive at all” to “I think that person is very attractive”) as used, e.g., in a study of Stirrat, Gumert, and Perret (2011). Finally, naturalness was assessed on a seven-point scale (from “I think that this person displays the emotion in a very unnatural manner” to “I think this person displays the emotion in a very natural manner”). The raters were asked to indicate how natural the shown emotional expression seemed to them.

## Results

The 96 remaining pictures were analyzed as follows with regard to the three scales: intensity (11-point scale), attractiveness (seven-point scale) and naturalness (seven-point scale). The mean values and standard deviations are shown in Table 1.

Table 1

*Means and SD (Standard Deviations) (96 Stimuli)*

Emotion	Intensity		Attractiveness		Naturalness	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Fear	6.77	1.13	4.05	0.63	4.23	0.72
Anger	6.13	1.19	3.82	0.66	3.81	0.77
Disgust	7.14	1.12	3.25	0.69	4.17	0.94
Happiness	7.35	1.19	4.74	0.66	5.77	0.53
Sadness	5.68	1.29	4.09	0.62	4.75	0.66
Surprise	7.18	1.21	3.98	0.61	3.75	0.82
Total	6.71	0.97	3.99	0.54	4.41	0.53

A selection process with the obtained data was carried out with the objective to include one emotional expression from each person into the picture set (total of 48 pictures). Parametric statistical approaches were used, due to the confirmation of normal distribution prerequisites with Shapiro-Wilks test ( $p > 0.05$  for intensity, attractiveness and naturalness). Results of a ANOVA (analysis of variance) (emotion  $\times$  rating) showed significant differences between the different emotions for intensity ( $F_{(1, 5)} = 21.147$ ;  $p < 0.000$ ), naturalness ( $F_{(1, 5)} = 86.530$ ;  $p < 0.000$ ) and attractiveness ( $F_{(1, 5)} = 37.670$ ;  $p < 0.000$ ). Post-hoc analysis resulted in several significant pair comparisons ( $p < 0.05$ ). Therefore only results that did reach significance are reported here: For intensity, this is the case for the comparison of fear and sadness ( $p < 0.000$ ), anger with disgust, happiness and surprise ( $p < 0.000$  for all comparisons), and sadness with disgust, happiness and surprise ( $p < 0.000$  for all comparisons). For attractiveness, the comparison of disgust with all other emotions and happiness with all other emotions shows meaningful differences ( $p < 0.000$  for all comparisons). For naturalness, pair comparisons of happiness and sadness with all other emotions reached significance ( $p < 0.000$  for all comparisons).

As females are able to rate facial expressions correctly when displaying with little intensity (Hoffmann et al., 2010), gender was afterwards introduced as a possible differentiating factor.

At first, pictures were analyzed in respect to differences in ratings for intensity, naturalness and attractiveness caused by gender. Therefore,  $t$ -tests were conducted. For intensity, a significant difference between male and female actors was detected for fear ( $t_{(134)} = -2.051$ ;  $p = 0.042$ ), disgust ( $t_{(134)} = -2.043$ ;  $p = 0.043$ ) and sadness ( $t_{(134)} = -0.2696$ ;  $p = 0.008$ ). In all three cases, female expressions were rated as more intense. For attractiveness all emotional expressions besides disgust were rated more attractive when expressed by women ( $p < 0.05$  for anger, fear, happiness, sadness and surprise). Female actors did also show disgust ( $t_{(134)} = -2.334$ ;  $p = 0.021$ ) and sadness ( $t_{(134)} = -2.831$ ;  $p = 0.005$ ) in a more natural way.

Neither differentiating among the diverse emotions nor significant difference between male and female actors can be reported at all for intensity ( $t_{(134)} = -2.334$ ;  $p = 0.244$ ), but for naturalness ( $t_{(134)} = 2.777$ ;  $p = 0.006$ ) and attractiveness ( $t_{(134)} = 4.392$ ;  $p < 0.000$ ).

In the second step, means of male and female raters were compared. For none of the three scales,

significance could be reached when exploring gender of the raters and its possible influence (ANOVA:  $F_{(1, 6)} = 0.105$ ;  $p = 0.746$  for intensity;  $F_{(1, 6)} = 1.842$ ;  $p = 0.175$  for attractiveness;  $F_{(1, 6)} = 0.970$ ;  $p = 0.325$  for naturalness). It can be concluded that there is no systematic rater effect due to gender. But, there seems to be a difference in the evaluation of intensity, attractiveness and naturalness between male and female actors.

The achieved results were used for final picture selection with only one emotional expression per actor (for results of recognition accuracy see Limbrecht et al., under review). It was intended to protect homogeneity for recognition accuracy on the one hand and to improve consistency for intensity, attractiveness and naturalness on the other hand. All analyses reported so far were run again with the selected 48 pictures.

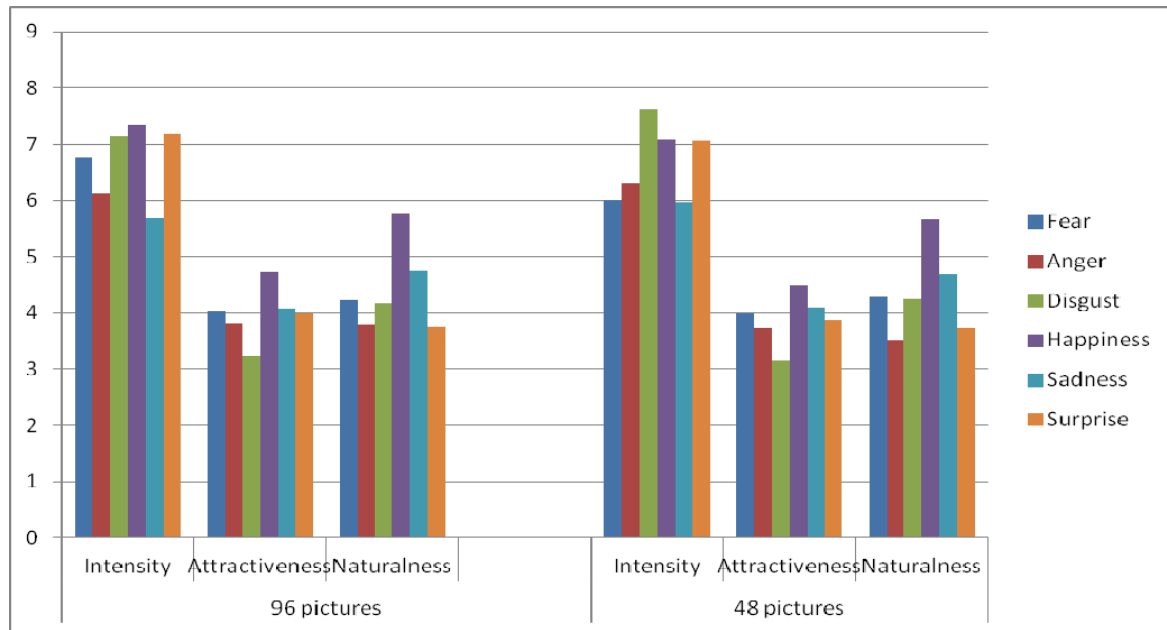


Figure 3. Ratings of the three factors before and after stimulus selection.

The final picture selection resulted in only some significant changes on emotional level. The evaluation of the intensity of fear ( $p < 0.000$ ) and disgust ( $p = 0.003$ ) resulted in some significant changes as well as the evaluation of the naturalness of anger ( $p = 0.045$ ). No changes can be reported for attractiveness (see Figure 3 for an overview of the ratings with 96 and 48 stimuli).

Results relating to means and SD are shown in Table 2.

Table 2

Means and Standard Deviations (Final Stimuli)

Emotion	Intensity		Attractiveness		Naturalness	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Fear	6.01	1.14	3.98	0.65	4.29	0.83
Anger	6.48	1.29	3.93	0.74	3.53	0.92
Disgust	7.63	1.09	3.14	0.72	4.25	1.02
Happiness	7.09	1.24	4.80	0.61	5.66	0.53
Sadness	5.97	1.32	4.12	0.66	4.70	0.70
Surprise	7.06	1.32	3.87	0.66	3.72	0.84
Total	6.82	0.99	3.97	0.54	4.36	0.57

The conducted ANOVA test showed significant differences between the different emotions for intensity ( $F_{(5)} = 16,609$ ;  $p < 0.000$ ), naturalness ( $F_{(5)} = 59.108$ ;  $p < 0.000$ ) and attractiveness ( $F_{(5)} = 42.841$ ;  $p < 0.000$ ). Some of the subsequent post-hoc analyses reached significance ( $p < 0.05$ ). Therefore, only results that did reach significance are reported here: For intensity, this is the case for the comparison of fear with happiness and surprise ( $p < 0.000$  for all comparisons), disgust with fear, anger and sadness ( $p < 0.000$  for all comparisons), and sadness with happiness and surprise ( $p < 0.000$  for all comparisons). For attractiveness, the comparison of disgust with all other emotions and happiness with all other emotions show meaningful differences ( $p < 0.000$  for all comparisons). For naturalness, pair comparisons of happiness with all other emotions reached significance ( $p < 0.000$  for all comparisons) as well as the comparison of anger with fear, disgust, happiness and sadness ( $p < 0.000$  for all comparisons).

Compared to the results reported for the 96 stimuli, less pair comparisons reached significance. This can be seen as an improvement in consistency.

Subsequent calculations were done again including gender of actors respectively raters. In the first step, the pictures were again analyzed for gender of the photographed persons to detect possible differences in rating intensity, naturalness and attractiveness of the stimuli. For intensity, male and female actors differed significantly only when looking at fear ( $t_{(134)} = -2.165$ ;  $p = 0.032$ ) and disgust ( $t_{(134)} = -2.723$ ;  $p = 0.007$ ). Intensity was rated higher for female actors in this case. Female and male actors also differed in their attractiveness. Females were rated more attractive on the emotional level of fear ( $t_{(134)} = -2.165$ ;  $p = 0.043$ ), anger ( $t_{(134)} = -2.269$ ;  $p = 0.025$ ), happiness ( $t_{(134)} = -2.273$ ;  $p = 0.025$ ) and sadness ( $t_{(134)} = -8.046$ ;  $p < 0.000$ ). Differences in gender concerning naturalness of the expression can be proved for fear ( $t_{(134)} = -2.568$ ;  $p = 0.024$ ), happiness ( $t_{(134)} = -2.334$ ;  $p < 0.000$ ), sadness ( $t_{(134)} = -2.421$ ;  $p = 0.009$ ) and surprise ( $t_{(134)} = -2.831$ ;  $p = 0.005$ ). Gender differences are nearly the same as those of the previous calculations.

Neither differentiating among the diverse emotions nor significant difference between male and female actors can be reported at all for intensity ( $t_{(134)} = -1.028$ ;  $p = 0.306$ ), but for naturalness ( $t_{(134)} = -1.747$ ;  $p = 0.083$ ) and attractiveness ( $t_{(134)} = -4.372$ ;  $p < 0.000$ ).

In a second step, means of male and female raters were compared. For none of the three scales, significance could be reached when exploring gender of the raters and its possible influence (ANOVA:  $F_{(1, 6)} = 1.442$ ;  $p = 0.231$  for intensity;  $F_{(1, 6)} = 1.923$ ;  $p = 0.166$  for attractiveness;  $F_{(1, 6)} = 3.038$ ;  $p = 0.082$  for naturalness).

Table 3

*Correlations of the Scales With Recognition Rates*

Emotion	Intensity	Attractiveness	Naturalness
Fear	0.139	0.599	0.676
Anger	-0.191	0.574	-0.293
Disgust	0.663	0.168	0.710*
Happiness	-0.307	-0.349	-0.406
Sadness	0.182	-0.390	-0.510
Surprise	0.492	-0.014	0.114

Note. \*:  $p < 0.05$ .

It can be concluded again that there is no systematic rater effect due to gender. But, there seems to be a difference in the evaluation of intensity, attractiveness and naturalness between male and female actors.

It was hypothesized that there is a systematic relationship between recognition accuracy and the three variables described in this paper. To prove these assumption correlations between intensity, attractiveness and naturalness with recognition accuracy were conducted. Therefore, means of each stimulus were adducted for comparison (see Table 3). Unfortunately, no emotion recognition rates were recorded in this study. Correlation, therefore, had to be done with recognition accuracy rates of a former study (Limbrecht et al., under review). Although population was comparable in cases of gender, educational background and age only approximations can be expected from this comparisons. Analyzes confirm the supposed limitations of this procedure. Only one significant correlation can be reported: the more natural disgust was shown the better it was recognized. All other possible correlations could not reach significance.

As a unit of measurement for consistency within the different emotions deviations around the mean were analyzed for the three scales individually. Results fluctuate from 1.1154 multiplied by the SD as the lowest score (attractiveness ratings for surprise) up to 1.0915 multiplied by the SD as the highest score (naturalness rating for fear). An overview about all results can be found in Table 4. It can be assumed that variance is even smaller when differentiating for gender.

Table 4

*Consistency for the Three Scales*

Emotion	Intensity	Attractiveness	Naturalness
Fear	$1.35495 \times SD$	$1.5806 \times SD$	$1.90155 \times SD$
Anger	$1.4962 \times SD$	$1.5173 \times SD$	$1.6165 \times SD$
Disgust	$1.24355 \times SD$	$1.4 \times SD$	$1.253 \times SD$
Happiness	$1.3131 \times SD$	$1.3702 \times SD$	$1.65895 \times SD$
Sadness	$1.42445 \times SD$	$1.4641 \times SD$	$1.11755 \times SD$
Surprise	$1.57135 \times SD$	$1.1154 \times SD$	$1.48995 \times SD$

*Note.* Numbers represent the factor that has to be multiplied with the SD as a reference for variance around mean.

## Discussion

In contrast to the two other scales (naturalness and attractiveness), only intensity was considered when creating the stimuli. All actors were asked to show each facial expression with maximum intensity. Because of evolutionary relevance, individual differences in physiognomy and competence in artificial emotion expression differences in the evaluation of intensity were still expected, especially for sadness and fear. Ekman and Friesen (2003) explained that emotions like fear and sadness are often described as less intense when looking at a picture instead of being confronted with these feelings in real life. This is because these two emotions are negative emotions on the one hand and notably connected to survival (fear) and social bonding (sadness) on the other hand. Both expressions are very dynamic, too (Ekman, 2003). For example, sadness is often associated with crying and sobbing, fears with screaming and distancing movements. This information is not accessible when using static picture material. Harwood, Hall, and Shinkfield (1999) could demonstrate that sadness is recognized better when using dynamic stimulus material. So, it may be reasonable that these two emotions would have been rated more intense on a dynamic presentation. It was hypothesized that at least some emotions were rated more intense for female than for male actors. Several studies (Kring & Gordon, 1998) show that females are more expressive than males as a product of human evolution. Women were dependent on the protection and help of males because of their physical inferiority. As a consequence, women had to learn to



express emotions more accurate than men, or rather they can show them with more nuances. One reference showed similar discussions (Bijlstra, Holland, & Wigboldus, 2010). Thus, we expected that gender role thinking could influence involuntarily the intensity ratings of emotional expressions (Hess, Senécal, Kirouac, Herrera, Philippot, & Kleck, 2000).

For naturalness and attractiveness, the emotions happiness, anger and disgust show most significant differences compared with the ratings of the other emotions which do not differ among each other. This indicates a high consistency of picture material within the different scales. But, what is different for anger, disgust and happiness when rating their naturalness and attractiveness, respectively?

Naturalness has certain impact on ecological validity (Carroll & Russell, 1997). Happiness was rated most natural in all studied conditions. This result was expected, because happiness was the only expression solely photographed when occurring non-acted as Duchenne smiles. In contrast, anger was rated very low. Maybe, this is due to the fact that anger, as well as sadness and fear (see above) is a very dynamic emotion that does not appear as threatening, as if it would have been present dynamically. Studies exploring naturalness are very rare. Langner et al. (2010) analyzed authenticity on a five-point Likert scale. Further statistical results were not presented, hence the results and their data cannot be compared. Elfenbein et al. (2007) showed that in cultural comparisons, the naturalness of emotions is quite important to the recognition accuracy rates. Prototypical FACS-based material is recognized well, completely independent from culture, but accuracy rates are higher within the same cultural area. These results inspired the authors of this study to induce the emotional expressions most naturally. A retrospective selection of those stimuli with the highest concordance to the prototypical expressions and the AUs (Action Units) leads to a higher homogeneity of the stimuli set. In general, it can be hypothesized that an improvement of the naturalness scale could be achieved with dynamic stimuli (Hoffmann et al., 2010; Hoffmann, Traue, Bachmayr, & Kessler, 2010).

Attractiveness is a relevant factor in social interaction. Within seconds the authors realize, whether they like another person and find him/her attractive or not. The results have pointed out that women and men participating in the current study were considered equally attractive. This contradicts, e.g., the findings of Tracy and Beall (2011) that women's faces are regarded to be more likeable and attractive than men's. Maybe, this is due to the fact that actors were collected from a very homogeneous sample of students. We could not find any differences in order of attractiveness between males and females, but some general differences. Happiness was rated most attractive, disgust less attractive. In general, it can be hypothesized that emotions send different messages that have an effect on attractiveness, e.g., happiness communicates approachability and disgust aversion (Becker, Kenrick, Neuberg, Blackwell, & Smith, 2007). Therefore, it is possible that positive emotions go along with a higher attractiveness rating. Additionally, socio-cultural gender norms have to be considered when thinking about further studies.

Interestingly, there are no significant differences between female and male raters regarding the scales attractiveness, naturalness and intensity. Therefore, a systematic gender effect of the raters can be excluded. For the sex of the photographed person, no differences could be observed, too.

Looking at the consistency, it can be concluded that it is necessary to differentiate on the level of basic emotions. Then, results are sufficient homogeneous. This fact is well known for hit rates in emotion recognition. Recognition accuracy differs highly for fear, anger, sadness, disgust, happiness and surprise with happiness often recognized best and fear mostly recognized worse (Langner et al., 2010; Tracy, Robins, & Schriber, 2009; Goeleven, De Raedt, Leyman, & Verschuere, 2008). Although in this study, no correlations between emotion

recognition and the estimation of intensity, attractiveness and naturalness could be found, it is likely that coherences would occur in an experimental setup that includes the request of both ratings. This would offer the possibility for further analyses like multiple regressions or interaction analyses.

In summary, our results indicate that the individual emotions (discussed earlier for estimating the intensity of fear and anger) are evaluated differently because of their relevance and importance in our society and our evolutionary history. Therefore, it is important to be aware of all co-varying factors. This study makes a first contribution, although there are some limitations. In conclusion, the authors point out that for creating a stimulus set, not only the emotion recognition accuracy should be considered. They showed differences in all three variables—intensity, attractiveness and naturalness. To reduce or rather exclude this effect systematically, a higher degree of similarity should be achieved for every single picture.

### Conclusions and Further Steps

Our picture set, PFA-U, is one of the first that claims standardization for other important influencing factors, besides the prototypical emotional expression and the resulting recognition rates. The differentiation of emotion ensures the highest homogeneity of the mentioned parameters. Further studies should focus on systematic variation of the three parameters to gain better insight of their relevance. For example, morphing tools offer the possibility to morph different intensities. One interesting research question is how recognition accuracy would change if, e.g., intensity was kept constant in advance for the different basic emotions.

Finally, PFA-U offers a lot of new hypotheses, aiming to clarify the fundamentals and mechanisms of the human emotion recognition, because of the additional perspectives and its embedding into a well-proved system (the FEEL test) (Kessler et al., 2002).

### References

- Becker, D. V., Kenrick, D. T., Neuberg, S. L., Blackwell, K. C., & Smith, D. M. (2007). The confounded nature of angry men and happy women. *Journal of Personality and Social Psychology*, 92, 179-190.
- Bijlstra, G., Holland, R.W., & Wigboldus, D. H. J. (2010). The social face of emotion recognition: Evaluations versus stereotypes. *Journal of Experimental Social Psychology*, 46, 657-663.
- Buck, R. (1984). *The communication of emotion*. New York: Guilford Press.
- Carroll, J. M., & Russell, J. A. (1997). Facial expressions in Hollywood's portrayal of emotion. *Journal of Personality and Social Psychology*, 72(1), 164-176.
- Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, 24(3), 285-290.
- Ekman, P. (1993). Facial expression and emotion. *American Psychologist*, 48, 384-392.
- Ekman, P. (2003). *Emotions revealed*. New York: Times Books.
- Ekman, P., & Friesen, W. V. (2003). *Unmasking the face: A guide to recognizing emotions from facial expressions*. Cambridge: Malor Books.
- Elfenbein, H. A., Beaupré, M., Lévesque, M., & Hess, U. (2007). Toward a dialect theory: Cultural differences in the expression and recognition of posed facial expressions. *Emotion*, 7(1), 131-146.
- Goeleven, E., De Raedt, R., Leyman, L., & Verschuere, B. (2008). The Karolinska directed emotional faces: A validation study. *Cognition and Emotion*, 22(6), 1094-1118.
- Harwood, N. K., Hall, L. J., & Shinkfield, A. J. (1999). Recognition of facial emotional expressions from moving and static displays by individuals with mental retardation. *American Journal of Mental Retardation*, 104(3), 270-278.
- Hess, U., Senécal, S., Kirouac, G., Herrera, P., Philippot, P., & Kleck, R. E. (2000). Emotional expressivity in men and women: Stereotypes and self-perception. *Cognition and Emotion*, 14, 609-642.
- Hoffmann, H., Kessler, H., Eppel, T., Rukavina, S., & Traue, H. C. (2010). Expression intensity, gender and facial emotion recognition: Women recognize only subtle emotions better than men. *Acta Psychologica*, 135(3), 278-283.

- Hoffmann, H., Traue, H. C., Bachmayr, F., & Kessler, H. (2010). Perceived realism of dynamic facial expressions of emotion: Optimal durations for the presentation of emotional onsets and offsets. *Cognition and Emotion*, 24(8), 1369-1378.
- Kappas, A. (2003). What facial activity can and cannot tell us about emotions. In M. Katsikitis (Ed.), *The human face: Measurement and meaning* (pp. 215-234). Dordrecht: Kluwer Academic Publishers.
- Kessler, H., Bayerl, P., Deighton, R. M., & Traue, H. C. (2002). Facially Expressed Emotion Labeling (FEEL): A computer-based test for emotion recognition. *Verhaltenstherapie Undverhaltensmedizin*, 23(3), 297-306.
- Kessler, H., Hoffmann, H., Bayerl, P., Neumann, H., Basic, A., Deighton, R. M., & Traue, H. C. (2005). Measuring emotion recognition with computer morphing: New methods for research and clinical practice. *Nervenheilkunde*, 24, 611-614.
- Kessler, H., Roth, J., von Wietersheim, J., Deighton, R. M., & Traue, H. C. (2007). Emotion recognition patterns in patients with panic disorder. *Depression and Anxiety*, 24, 223-226.
- Kring, A. M., & Gordon, A. H. (1998). Sex differences in emotion: Expression, experience, and physiology. *Journal of Personality and Social Psychology*, 74, 686-703.
- Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & Knippenberg, A. (2010). Presentation and validation of the radboud faces database. *Cognition and Emotion*, 24(8), 1277-1388.
- Limbrecht, K., Hoffmann, H., Walter, S., Gruss, S., Hrabal, D., & Traue, H. C. (under review). Pictures of Facial Affect-Ulm (PFA-U): A new FACS-based set of pictures for basic emotions. In *Proceedings of Eighth International Conference on Emotion and Organizational Life (EMONET VIII)*.
- Murphy, F. C., Nimmo-Smith, I., & Lawrence, A. D. (2003). Functional neuroanatomy of emotions: A meta-analysis. *Cognitive, Affective and Behavioral Neuroscience*, 3, 207-233.
- O'Doherty, J., Winston, J., Critchley, H. D., Perrett, D., Burt, D. M., & Dolan, R. J. (2003). Beauty in a smile: The role of medial orbitofrontal cortex in facial attractiveness. *Neuropsychologia*, 41(2), 147-55.
- Palermo, R., & Coltheart, M. (2004). Photographs of facial expression: Accuracy, response times, and ratings of intensity. *Behavior Research Methods, Instruments, and Computers*, 36, 634-638.
- Stirrat, M., Gumert, M., & Perret, D. (2011). The effect of attractiveness on food sharing preferences in human mating markets. *Evolutionary Psychology*, 9, 79-91.
- Tracy, J. L., & Beall, A. T. (2011). Happy guys finish last: The impact of emotion expressions on sexual attraction. *Emotion*, 11, 1379-1387.
- Tracy, J. L., Robins, R. W., & Schriber, R. A. (2009). Development of a FACS-verified set of basic and self-conscious emotion expressions. *Emotion*, 9(4), 554-559.
- Zaidel, D. W., Aarde, S. M., & Baig, K. (2005). Appearance of symmetry, beauty, and health in human faces. *Brain and Cognition*, 57, 261-263.