**Appendix A**

**Technical Background of the FACET and AFFDEX Module**

Software-based facial expression analysis is based on algorithms that apply a set of rules based on psychological theories and statistical procedures. In general, these algorithms work in three steps: (1) detecting faces, (2) detecting facial landmarks and (3) classifying emotions.

In the first step, a face in a video frame is detected, e.g., with a Viola–Jones cascaded classifier algorithm (iMotions, 2016). This algorithm is based on a machine learning approach that uses different features to identify faces in an image and discard background regions (see Viola & Jones, 2001). This step of facial detection is applied to each video frame and hence only frame with a positive detection rate are used in subsequent analysis.

In the second step, significant facial landmarks such as eye corners, eye centers, nose tip, mouth corner and mouth center are detected. Technically speaking, a number of *regions of interest* (ROIs) that represent significant facial landmarks are identified and a likelihood value for its presence at a given (x,y) location is estimated. There are diverse methods to detect facial landmarks (e.g., by an active appearance model or a constrained local model; see Beumer, Tao, Bazen, & Veldhuis, 2006; Cootes, Edwards, & Taylor, 2001; McDuff et al., 2010). Based on these landmarks, an artificial reticular face model that represents a simplified version of the actual face is developed. This is typically done by a *support vector machine* (SVM). The output of the SVM is equivalent to estimated AU intensities. To estimate AUs, SVM uses diverse facial databases. These databases are often publicly available datasets that provide human FACS coding for essential AUs. With the SVM, the position, size, and scale of the actual face is adjusted instantly when the face moves (for more details see, e.g., Littlewort et al., 2011; McDuff et al., 2010).

In the third step, a classification algorithm is used to translate estimates for AUs into probability-like values for emotional states. While FACET uses estimates for six facial landmarks, AFFDEX uses estimates for 34 facial landmarks. In general, facial expression analysis algorithms implement a set of basic emotion detectors by feeding their AU estimates into a multivariate logistic regression classifier. This classifier is trained on facial databases and generates posterior probabilities for each emotion. Thus, the classification expresses how likely it is that the detected face expresses a certain basic emotion. Given that facial expression analysis algorithms are trained on different facial databases, it is possible that they provide different results for a certain facial expression image (iMotions, 2016).

**Appendix B**

**Data Generation and Analytics**

The translation from facial landmarks into values is based on a statistical procedure that compares the expression on the actual face with the expressions of the faces stored in the database. The values that iMotions assigns to every basic emotion is a probability-like value. In FACET these values are referred to as ‘evidence values’; in AFFDEX as ‘probabilities’.

The evidence value for the FACET module is similar to a z-score. That is, positive (negative) values indicate evidence that a certain emotion is expressed (absent). While values larger than three indicate strong evidence for the presence of a certain emotion, negative values smaller than three indicate strong evidence for the absence of an emotion. An evidence value of zero indicates that there is no evidence either way.

The probabilities returned by the AFFDEX module range between zero and one. Accordingly, a value of zero indicates no evidence and a value of one the highest evidence that a certain emotion is fully expressed. For more details, see iMotions help center, 2016.

When it comes to the interpretation of evidence values (FACET) and probabilities (AFFDEX), one has to differentiate between raw values and baseline-corrected values. Raw values represent the classification for a certain emotion of a face compared to the facial expressions in the global database. With this, an individual’s facial expression can be directly compared with countless other facial expressions. This is useful when one is interested in aggregating data of several individuals or comparing data from two or more groups of individuals (iMotions help center, 2016). In contrast to raw data, baseline-corrected data anticipate that people differ in their “neutral” expression and may change their neutral expression over time. Baseline-corrected raw values reflect the relative expression of an individual independent of the database. In some cases, it is useful to consider an individual’s baseline as this allows to detect changes in emotional expression relative to the individual’s neutral expression. This helps to avoid misestimating the extent of a certain emotion when using raw values (iMotions help center, 2016).

**Appendix C**

**Additional Results for Study 1**

Table C1   
*Non-Baseline Corrected Classification Accuracy for Emotions Separately for the iMotions Modules AFFDEX and FACET*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Emotion | Database | Number | AFFDEX | FACET | | |
| Matched | MS | sDI |
| Anger | ADFES | 22 | 8 | 22 | 0.36 | 1.00 | -0.10 | -0.29 |
| WSEFEP | 29 | 13 | 27 | 0.45 | 0.93 | -0.55 | -0.41 |
| RaFD | 39a | 23 | 39 | 0.59 | 0.97 | -1.31 | -0.51 |
| Disgust | ADFES | 22 | 22 | 22 | 1.00 | 1.00 | -0.32 | -0.40 |
| WSEFEP | 29b | 24 | 28 | 0.86 | 1.00 | -0.09 | -0.54 |
| RaFD | 39 | 39 | 39 | 1.00 | 1.00 | -0.79 | 0.24 |
| Fear | ADFES | 22c | 1 | 21 | 0.05 | 0.95 | -1.35 | -0.57 |
| WSEFEP | 29 | 0 | 21 | 0.00 | 0.72 | - | -0.65 |
| RaFD | 39 | 0 | 38 | 0.00 | 0.97 | - | -0.24 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Happiness | ADFES | 22 | 22 | 22 | 1.00 | 1.00 | -0.01 | 1.10 |
| WSEFEP | 29 | 29 | 29 | 1.00 | 1.00 | 0.57 | 1.92 |
| RaFD | 39 | 37 | 39 | 0.95 | 1.00 | 0.08 | 1.70 |
| Sadness | ADFES | 22 | 20 | 22 | 0.91 | 1.00 | 0.36 | -0.87 |
| WSEFEP | 29 | 19 | 25 | 0.66 | 0.86 | 0.15 | -0.92 |
| RaFD | 39 | 34 | 39 | 0.87 | 1.00 | 0.60 | -0.54 |
| Surprise | ADFES | 21d | 19 | 21 | 0.90 | 1.00 | 0.66 | -0.33 |
| WSEFEP | 29 | 29 | 29 | 1.00 | 1.00 | 0.48 | -0.23 |
| RaFD | 39 | 38 | 39 | 0.97 | 1.00 | 0.41 | 0.20 |
| Contempt | ADFES | 22 | 19 | 21 | 0.86 | 1.00 | 0.22 | -0.17 |
| WSEFEP | - | - | - | - |
| RaFD | 39 | 38 | 39 | 0.97 | 1.00 | 0.87 | 0.29 |
| Total | ADFES | 153e | 111 | 152 | 0.73 | 0.99 | 0.06 | -0.10 |
| WSEFEP | 174f | 115 | 160 | 0.66 | 0.92 | 0.21 | -0.09 |
| RaFD | 273g | 209 | 271 | 0.77 | 0.99 | 0.06 | -0.17 |
| Average |  | 600h | 435 | 583 | 0.73 | 0.97 | 0.10 | 0.03 |

*Note.* Number = number of classified database pictures; Matched = number of pictures that iMotions classified correctly with the database’s emotion label (true positives). MS = Matching Score. sDI = standardized Distinctness Index. FACET successfully detected all faces in the pictures. For AFFDEX, facial detection was successful for a38 of 39 | b26 of 29 | c21 of 22 and d19 of 21 database pictures. Thus, the number of detected faces for AFFDEX was e150 in the ADFES, f171 in the WSEFEP, g272 in the RaFD database and h593 in total (out of 600). While the left side of the vertical bar shows the numbers for AFFDEX, the right side shows the numbers for FACET (AFFDEX | FACET).

Table C2

*Confusion Matrix for Table C1 of Study 1 (Non-Baseline Corrected Data)*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | AFFDEX | FACET | | | | | | | |
| Database Label | Anger | Contempt | Disgust | Fear | Happiness | Sadness | Surprise | Total (Target) |
| Anger | **44 | 87** | 8 | 0 | 11 | 2 | 0 | 0 | 0 | 0 | 27 | 1 | 0 | 0 | 90 | 90 |
| Contempt | 0 | 0 | **57 | 61** | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 1 | 0 | 61 | 61 |
| Disgust | 3 | 0 | 0 | 0 | **86 | 90** | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 90 | 90 |
| Fear | 1 | 0 | 6 | 0 | 1 | 0 | **1 | 80** | 0 | 0 | 1 | 0 | 80 | 10 | 90 | 90 |
| Happiness | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **88 | 90** | 0 | 0 | 2 | 0 | 90 | 90 |
| Sadness | 4 | 3 | 1 | 0 | 5 | 1 | 0 | 0 | 0 | 0 | **73 | 86** | 7 | 0 | 90 | 90 |
| Surprise | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **86 | 89** | 89 | 89 |
| Total (Class) | 54 | 90 | 73 | 61 | 103 | 93 | 2 | 80 | 91 | 90 | 101 | 87 | 54 | 99 | 600 | 600 |

*Note.* Total (Class) is the number of times iMotions classified the basic emotion per emotion target label. Total (Target) is the number of times the emotion target label is present. While the left side of the vertical bar shows the numbers for AFFDEX, the right side shows the numbers for FACET (AFFDEX | FACET).

Table C3

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Performance Indices for AFFDEX | FACET | | | | |
| Emotion | Sensitivitya | Specificityb | Precisionc | NPVd | F1e |
| Anger | 0.49 | 0.97 | 0.98 | 1.00 | 0.81 | 1.00 | 0.92 | 1.00 | 0.61 | 1.00 |
| Contempt | 0.93 | 1.00 | 0.97 | 1.00 | 0.78 | 1.00 | 0.99 | 1.00 | 0.85 | 1.00 |
| Disgust | 0.96 | 0.89 | 0.97 | 1.00 | 0.83 | 1.00 | 0.99 | 1.00 | 0.89 | 1.00 |
| Fear | 0.01 | 0.89 | 1.00 | 1.00 | 0.50 | 1.00 | 0.85 | 1.00 | 0.02 | 1.00 |
| Happiness | 0.98 | 0.96 | 0.99 | 1.00 | 0.97 | 1.00 | 1.00 | 1.00 | 0.97 | 1.00 |
| Sadness | 0.81 | 0.96 | 0.95 | 1.00 | 0.72 | 1.00 | 0.97 | 1.00 | 0.76 | 1.00 |
| Surprise | 0.97 | 1.00 | 0.82 | 1.00 | 0.49 | 1.00 | 0.99 | 1.00 | 0.65 | 1.00 |

*Performance Indices to Assess iMotions for Study 1 (Non-Baseline Corrected Data)*

*Note.* Performance indices are derived from the confusion matrix, i.e. from *true positives* (TP), *true negatives* (TN), *false positives* (FP), *false negatives* (FN), the number of real positive occurrences (P) and the number of real negative occurrences (N):

a

b

c

d

e

Overall performance indices for AFFDEX are: Accuracy (MS) = 0.73; Kappa = 0.68

Overall performance indices for FACET are: Accuracy (MS) = 0.97; Kappa = 0.97

Kappa is an index that adjusts accuracy (MS) by accounting for a correct prediction by chance:

, whereas P(expected) = probability of a correct prediction due to chance.

While the left side of the vertical bar shows the numbers for AFFDEX, the right side shows the numbers for FACET (AFFDEX | FACET).

*../../Analysis_Studie1/Ranalysis/FigureC1.pdf*

*Figure C1.* Overview of the baseline corrected classification accuracy for basic emotions separately for the iMotions modules AFFDEX and FACET across ADFES, WSEFEP and RaFD. Contempt is not depicted since WSEFEP does not provide facial expression pictures for contempt. While figures depicting non-baseline corrected data show blue bars, figures depicting baseline corrected data show red bars (cf. Figure 1).

Table C4   
*Baseline Corrected Classification Accuracy for Emotions Separately for the iMotions Modules AFFDEX and FACET*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Emotion | Database |  | AFFDEX | FACET | | |
| Number |
|  | Matched | MS | sDI |
| Anger | ADFES | 22 | 8 | 22 | 0.36 | 1.00 | -1.06 | -0.15 |
| WSEFEP | 29 | 13 | 23 | 0.45 | 0.79 | -0.54 | -0.37 |
| RaFD | 39 | 23 | 39 | 0.59 | 1.00 | -1.31 | -0.38 |
| Disgust | ADFES | 22 | 21 | 22 | 0.95 | 1.00 | -0.24 | 0.32 |
| WSEFEP | 29 | 24 | 28 | 0.83 | 0.97 | -0.02 | -0.33 |
| RaFD | 39 | 37 | 39 | 0.95 | 1.00 | -0.72 | 0.20 |
| Fear | ADFES | 22 | 1 | 20 | 0.05 | 0.91 | -1.07 | -0.74 |
| WSEFEP | 29 | 0 | 18 | 0 | 0.62 | - | -0.67 |
| RaFD | 39 | 0 | 37 | 0 | 0.95 | - | -0.39 |
| Happiness | ADFES | 22 | 22 | 22 | 1.00 | 1.00 | 0.12 | 1.42 |
| WSEFEP | 29 | 29 | 29 | 1.00 | 1.00 | 0.60 | 2.07 |
| RaFD | 39 | 39 | 39 | 1.00 | 1.00 | 0.13 | 1.95 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sadness | ADFES | 22 | 22 | 21 | 1.00 | 0.95 | 0.18 | -0.78 |
| WSEFEP | 29 | 21 | 27 | 0.72 | 0.93 | 0.02 | -0.93 |
| RaFD | 39 | 34 | 38 | 0.87 | 0.97 | 0.62 | -0.64 |
| Surprise | ADFES | 21 | 19 | 21 | 0.90 | 1.00 | 0.58 | -0.20 |
| WSEFEP | 29 | 29 | 29 | 1.00 | 1.00 | 0.50 | -0.02 |
| RaFD | 39 | 38 | 39 | 0.97 | 1.00 | 0.32 | 0.27 |
| Contempt | ADFES | 22 | 19 | 21 | 0.86 | 0.95 | 0.32 | -0.42 |
| WSEFEP | - | - | - | - |
| RaFD | 39 | 33 | 36 | 0.85 | 0.92 | 0.78 | -0.26 |
| Total | ADFES | 153 | 112 | 149 | 0.73 | 0.97 | 0.08 | -0.06 |
| WSEFEP | 174 | 116 | 154 | 0.67 | 0.89 | 0.21 | 0.03 |
| RaFD | 273 | 204 | 267 | 0.75 | 0.98 | 0.04 | 0.12 |
| Average |  | 600 | 432 | 570 | 0.72 | 0.95 | 0.10 | 0.05 |

*Note.* In order to run the analysis on baseline corrected data, we corrected the obtained measurements for all pictures by comparing them to individual baseline values. Exemplary R-code for stimuli-wise baseline correction of iMotions data can be found online (<https://github.com/michaelschulte/FacialExpressionAnalysis>). These baseline values were generated on the basis of neutral facial expression pictures (WSEFEP, ADFES and RaFD all provide a picture of a neutral facial expression for each face model). By subtracting individual baseline values from the raw probability-like values for all emotions, we accounted for the face models’ natural variations in facial expressiveness (see iMotions, 2016).

Number = number of classified database pictures; Matched = number of pictures that iMotions classified correctly with the database’s emotion label (true positives). MS = Matching Score. sDI = standardized Distinctness Index. While the left side of the vertical bar shows the numbers for AFFDEX, the right side shows the numbers for FACET (AFFDEX | FACET).

Table C5

*Confusion Matrix for Table C4 of Study 1 (Baseline Corrected Data)*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | AFFDEX | FACET | | | | | | | |
| Database Label | Anger | Contempt | Disgust | Fear | Happiness | Sadness | Surprise | Total (Target) |
| Anger | **44 | 84** | 9 | 0 | 5 | 3 | 0 | 0 | 0 | 3 | 31 | 0 | 1 | 0 | 90 | 90 |
| Contempt | 0 | 0 | **52 | 57** | 0 | 0 | 0 | 0 | 3 | 3 | 1 | 1 | 5 | 0 | 61 | 61 |
| Disgust | 3 | 0 | 3 | 0 | **82 | 89** | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 90 | 90 |
| Fear | 1 | 0 | 5 | 0 | 0 | 0 | **1 | 75** | 0 | 0 | 1 | 0 | 82 | 15 | 90 | 90 |
| Happiness | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **90 | 90** | 0 | 0 | 0 | 0 | 90 | 90 |
| Sadness | 4 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | **77 | 86** | 8 | 0 | 90 | 90 |
| Surprise | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **86 | 89** | 89 | 89 |
| Total (Class) | 54 | 85 | 70 | 57 | 88 | 93 | 2 | 76 | 93 | 98 | 111 | 87 | 182 | 04 | 600 | 600 |

*Note.* Total (Class) is the number of times iMotions classified the basic emotion per emotion target label. Total (Target) is the number of times the emotion target label is present. While the left side of the vertical bar shows the numbers for AFFDEX, the right side shows the numbers for FACET (AFFDEX | FACET).

Table C6

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | AFFDEX | FACET | | | | |
| Emotion | Sensitivitya | Specificityb | Precisionc | NPVd | F1e |
| Anger | 0.49 | 0.93 | 0.98 | 1.00 | 0.81 | 0.99 | 0.92 | 0.99 | 0.61 | 0.96 |
| Contempt | 0.85 | 0.93 | 0.97 | 1.00 | 0.74 | 1.00 | 0.98 | 0.99 | 0.79 | 0.97 |
| Disgust | 0.91 | 0.99 | 0.99 | 0.99 | 0.93 | 0.96 | 0.98 | 1.00 | 0.92 | 0.97 |
| Fear | 0.01 | 0.83 | 1.00 | 1.00 | 0.50 | 0.99 | 0.85 | 0.97 | 0.02 | 0.90 |
| Happiness | 1.00 | 1.00 | 0.99 | 0.98 | 0.97 | 0.92 | 1.00 | 1.00 | 0.98 | 0.96 |
| Sadness | 0.86 | 0.96 | 0.93 | 1.00 | 0.69 | 0.99 | 0.97 | 0.99 | 0.77 | 0.97 |
| Surprise | 0.97 | 1.00 | 0.81 | 0.97 | 0.47 | 0.86 | 0.99 | 1.00 | 0.63 | 0.92 |

*Performance Indices to Assess iMotions for Study 1 (Baseline Corrected Data)*

*Note.* Performance indices are derived from the confusion matrix. Overall performance indices for AFFDEX are: Accuracy (MS) = 0.72; Kappa = 0.67. Overall performance indices for FACET are: Accuracy (MS) = 0.95; Kappa = 0.94. While the left side of the vertical bar shows the numbers for AFFDEX, the right side shows the numbers for FACET (AFFDEX | FACET).

**Appendix D**

**Additional Results for Study 2**

Table D1

*Non-Baseline Corrected Classification Accuracy of Valence for iMotions Modules AFFDEX and FACET*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Valence | Picture | AFFDEX | FACET | | | | | | |
| Matched | | picturewise MS | | valencewise MS | Overall MS | |
| Positive | IAPS 1710 | 25 | 32 | 0.23 | 0.29 | | 0.16 | 0.22 | | | 0.55 | 0.57 |
| IAPS 1750 | 20 | 29 | 0.18 | 0.26 | |
| GAPED P067 | 7 | 12 | 0.06 | 0.11 | |
| Negative | IAPS 9940 | 105 | 102 | 0.95 | 0.93 | |  | | |
| IAPS 9570 | 104 | 98 | 0.95 | 0.89 | | 0.95 | 0.92 | | |
| GAPED A075 | 105 | 104 | 0.95 | 0.95 | |  | | |

*Note.* Matched = number of participant faces that match the picture’s valence (true positives). MS = Matching Score. While the left side of the vertical bar shows the numbers for AFFDEX, the right side shows the numbers for FACET (AFFDEX | FACET).

Table D2

*Baseline Corrected Classification Accuracy for Instructed Facial Expressions for Basic Emotions Separately for iMotions Modules AFFDEX and FACET*

|  |  |  |  |
| --- | --- | --- | --- |
| RaFD picture | AFFDEX | FACET | | |
| Matched | MS | sDI |
| Anger | 54 | 86 | 0.49 | 0.78 | -0.10 | 0.10 |
| Contempt | 75 | 47 | 0.68 | 0.43 | 0.87 | -0.13 |
| Disgust | 87 | 90 | 0.79 | 0.82 | -0.31 | 0.48 |
| Fear | 1 | 11 | 0.01 | 0.10 | -0.95 | -0.85 |
| Happiness | 100 | 108 | 0.91 | 0.98 | -0.52 | 1.52 |
| Sadness | 39 | 44 | 0.35 | 0.40 | -0.32 | -0.66 |
| Surprise | 67 | 97 | 0.61 | 0.88 | 0.62 | -0.01 |
| Total/Average | 423 | 483 | 0.55 | 0.63 | 0.02 | 0.35 |

*Note.* Matched = Number of participant faces where the detected emotion matched the picture’s emotion as reported in the dataset (female RaFD model number 01). MS = Matching Score. sDI = standardized Distinctness Index. While the left side of the vertical bar shows the numbers for AFFDEX, the right side shows the numbers for FACET (AFFDEX | FACET).

Table D3

*Confusion Matrix for Table D2 of Study 2 (Baseline Corrected Data)*

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | AFFDEX | FACET | | | | | | | | |
| Database Label | Anger | Contempt | Disgust | Fear | Happiness | Sadness | Surprise | Total (Target) |
| Anger | **54 | 86** | 15 | 1 | 27 | 13 | 0 | 0 | 3 | 7 | 9 | 1 | 2 | 2 | 110 | 110 |
| Contempt | 2 | 5 | **75 | 47** | 7 | 7 | 0 | 6 | 5 | 28 | 4 | 6 | 17 | 11 | 110 | 110 |
| Disgust | 6 | 12 | 4 | 2 | **87 | 90** | 1 | 0 | 5 | 6 | 7 | 0 | 0 | 0 | 110 | 110 |
| Fear | 10 | 8 | 27 | 1 | 4 | 0 | **1 | 11** | 5 | 11 | 8 | 2 | 55 | 77 | 110 | 110 |
| Happiness | 1 | 0 | 7 | 0 | 1 | 0 | 0 | 0 | **100 | 108** | 1 | 0 | 0 | 2 | 110 | 110 |
| Sadness | 18 | 27 | 26 | 7 | 10 | 4 | 0 | 1 | 1 | 11 | **39 | 44** | 16 | 16 | 110 | 110 |
| Surprise | 2 | 0 | 18 | 2 | 16 | 0 | 1 | 1 | 6 | 10 | 0 | 0 | **67 | 97** | 110 | 110 |
| Total (Class) | 93 | 138 | 172 | 60 | 152 | 14 | 3 | 19 | 125 | 181 | 68 | 53 | 157 | 205 | 770 | 770 |

*Note.* Total (Class) is the number of times iMotions classified the basic emotion per emotion target label. Total (Target) is the number of times the emotion target label is present. While the left side of the vertical bar shows the numbers for AFFDEX, the right side shows the numbers for FACET (AFFDEX | FACET).

Table D4

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Performance Indices for AFFDEX | FACET | | | | |
| Emotion | Sensitivitya | Specificityb | Precisionc | NPVd | F1e |
| Anger | 0.49 | 0.78 | 0.94 | 0.92 | 0.58 | 0.62 | 0.92 | 0.96 | 0.53 | 0.69 |
| Contempt | 0.68 | 0.43 | 0.85 | 0.98 | 0.44 | 0.78 | 0.94 | 0.91 | 0.53 | 0.55 |
| Disgust | 0.79 | 0.82 | 0.90 | 0.96 | 0.57 | 0.79 | 0.96 | 0.97 | 0.66 | 0.80 |
| Fear | 0.01 | 0.10 | 1.00 | 0.99 | 0.33 | 0.58 | 0.86 | 0.87 | 0.02 | 0.17 |
| Happiness | 0.91 | 0.98 | 0.96 | 0.89 | 0.80 | 0.60 | 0.98 | 1.00 | 0.85 | 0.74 |
| Sadness | 0.35 | 0.40 | 0.96 | 0.99 | 0.57 | 0.83 | 0.90 | 0.91 | 0.44 | 0.54 |
| Surprise | 0.61 | 0.88 | 0.86 | 0.84 | 0.43 | 0.47 | 0.93 | 0.98 | 0.50 | 0.62 |

*Performance Indices to Assess iMotions for Study 2 (Baseline Corrected Data)*

*Note.* Performance indices are derived from the confusion matrix. Overall performance indices for AFFDEX are: Accuracy (MS) = 0.55; Kappa = 0.47. Overall performance indices for FACET are: Accuracy (MS) = 0.63; Kappa = 0.57. While the left side of the vertical bar shows the numbers for AFFDEX, the right side shows the numbers for FACET (AFFDEX | FACET).

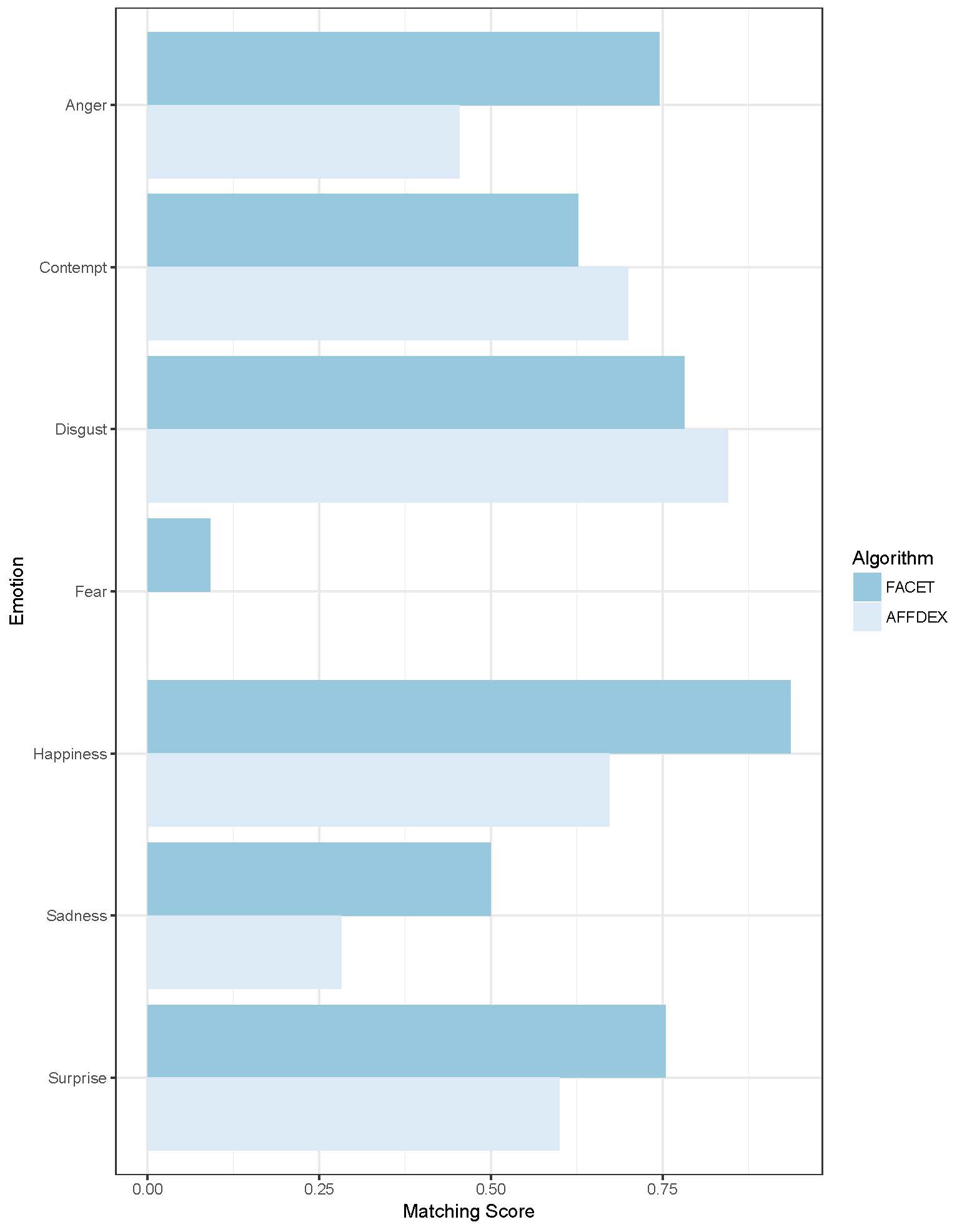
*Figure D1.* Overview of the non-baseline corrected classification accuracy for basic emotions separately for the iMotions modules AFFDEX and FACET.

Table D5

*Non-Baseline Corrected Classification Accuracy for Instructed Facial Expressions for Basic Emotions Separately for iMotions Modules AFFDEX and FACET*

|  |  |  |  |
| --- | --- | --- | --- |
| RaFD picture | AFFDEX | FACET | | |
| Matched | MS | sDI |
| Anger | 50 | 82 | 0.45 | 0.75 | -0.05 | 0.21 |
| Contempt | 77 | 69 | 0.70 | 0.63 | 1.00 | 0.22 |
| Disgust | 93 | 86 | 0.85 | 0.78 | -0.34 | 0.64 |
| Fear | 0 | 10 | 0.00 | 0.09 | - | -0.76 |
| Happiness | 74 | 103 | 0.67 | 0.94 | -0.39 | 1.06 |
| Sadness | 31 | 55 | 0.28 | 0.50 | -0.15 | -0.41 |
| Surprise | 66 | 83 | 0.60 | 0.75 | 0.51 | -0.01 |
| Total/Average | 391 | 488 | 0.51 | 0.63 | 0.11 | 0.34 |

*Note.* Matched = Number of participant faces where the detected emotion matched the picture’s emotion as reported in the dataset (female RaFD model number 01). MS = Matching Score. sDI = standardized Distinctness Index. While the left side of the vertical bar shows the numbers for AFFDEX, the right side shows the numbers for FACET (AFFDEX | FACET).

Table D6

*Confusion Matrix for Table D5 of Study 2 (Non-Baseline Corrected Data)*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | AFFDEX | FACET | | | | | | | |
| Database label | Anger | Contempt | Disgust | Fear | Happiness | Sadness | Surprise | Total (Target) |
| Anger | **50 | 82** | 16 | 10 | 34 | 14 | 0 | 0 | 1 | 1 | 9 | 3 | 0 | 0 | 110 | 110 |
| Contempt | 2 | 9 | **77 | 69** | 11 | 5 | 0 | 2 | 4 | 6 | 3 | 16 | 13 | 3 | 110 | 110 |
| Disgust | 5 | 10 | 6 | 4 | **93 | 86** | 0 | 0 | 3 | 4 | 3 | 6 | 0 | 0 | 110 | 110 |
| Fear | 9 | 13 | 25 | 17 | 4 | 0 | **0 | 10** | 12 | 13 | 8 | 11 | 52 | 46 | 110 | 110 |
| Happiness | 0 | 1 | 4 | 4 | 31 | 1 | 0 | 0 | **74 | 103** | 1 | 1 | 0 | 0 | 110 | 110 |
| Sadness | 9 | 23 | 35 | 25 | 20 | 3 | 0 | 0 | 2 | 4 | **31 | 55** | 13 | 0 | 110 | 110 |
| Surprise | 2 | 4 | 20 | 8 | 14 | 1 | 0 | 3 | 8 | 9 | 0 | 2 | **66 | 83** | 110 | 110 |
| Total (Class) | 77 | 142 | 183 | 137 | 207 | 110 | 0 | 15 | 104 | 140 | 55 | 94 | 144 | 132 | 770 | 770 |

*Note.* Total (Class) is the number of times iMotions classified the basic emotion per emotion target label. Total (Target) is the number of times the emotion target label is present. While the left side of the vertical bar shows the numbers for AFFDEX, the right side shows the numbers for FACET (AFFDEX | FACET).

Table D7

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Performance Indices for AFFDEX | FACET | | | | |
| Emotion | Sensitivitya | Specificityb | Precisionc | NPVd | F1e |
| Anger | 0.45 | 0.75 | 0.96 | 0.91 | 0.65 | 0.58 | 0.91 | 0.96 | 0.53 | 0.65 |
| Contempt | 0.70 | 0.63 | 0.84 | 0.90 | 0.42 | 0.50 | 0.94 | 0.94 | 0.53 | 0.56 |
| Disgust | 0.85 | 0.78 | 0.83 | 0.96 | 0.45 | 0.78 | 0.97 | 0.96 | 0.59 | 0.78 |
| Fear | 0.00 | 0.09 | 1.00 | 0.99 | - | 0.67 | 0.86 | 0.87 | - | 0.16 |
| Happiness | 0.67 | 0.94 | 0.95 | 0.94 | 0.71 | 0.73 | 0.95 | 0.99 | 0.69 | 0.82 |
| Sadness | 0.28 | 0.50 | 0.96 | 0.94 | 0.56 | 0.59 | 0.89 | 0.92 | 0.38 | 0.54 |
| Surprise | 0.60 | 0.75 | 0.88 | 0.93 | 0.46 | 0.63 | 0.93 | 0.96 | 0.52 | 0.69 |

*Performance Indices to Assess iMotions for Study 2 (Non-Baseline Corrected Data)*

*Note.* Performance indices are derived from the confusion matrix. Overall performance indices for AFFDEX are: Accuracy (MS) = 0.51; Kappa = 0.43. Overall performance indices for FACET are: Accuracy (MS) = 0.63; Kappa = 0.57. While the left side of the vertical bar shows the numbers for AFFDEX, the right side shows the numbers for FACET (AFFDEX | FACET).