FaceReader Methodology Note

By Dr. Leanne Loijens and Dr. Olga Krips

*Behavioral research consultants at Noldus Information Technology*

WHAT IS FACEREADER?

FaceReader is a program for facial analysis. It can detect facial expressions. FaceReader has been trained to classify expressions in one of the following categories: happy, sad, angry, surprised, scared, disgusted, and neutral. These emotional categories have been described by Ekman [1] as the basic or universal emotions. In addition to these basic emotions, contempt can be classified as expression, just like the other emotions [2]. Obviously, facial expressions vary in intensity and are often a mixture of emotions. In addition, there is quite a lot of inter-personal variation.

FaceReader has been trained to classify the expressions mentioned above. It is also possible to add custom expressions to the software yourself. In addition to facial expressions, FaceReader offers a number of extra classifications. It can, for example, detect the gaze direction and whether eyes and mouth are closed or not.

FaceReader can detect facial expressions: happy, sad, angry, surprised, scared, disgusted, and neutral.
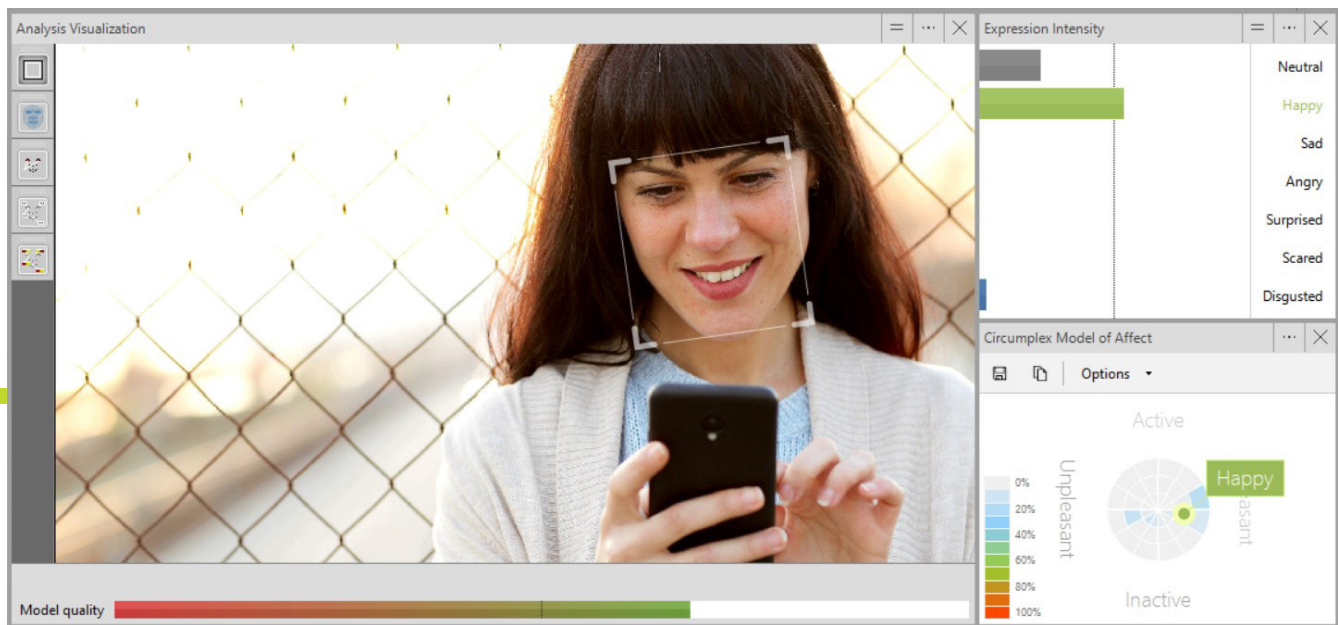
*Analyzing facial expressions
with FaceReader.*

You find a full overview of the classifications in the Technical Specifications of FaceReader that you can obtain from your Noldus sales representative. FaceReader can classify facial expressions either live using a webcam, or offline, in video files or images. Also, analysis of infrared images or video is possible.

HOW DOES FACEREADER WORK?

With the Deep Face classification method, FaceReader directly classifies the face from image pixels, using an artificial neural network to recognize patterns.
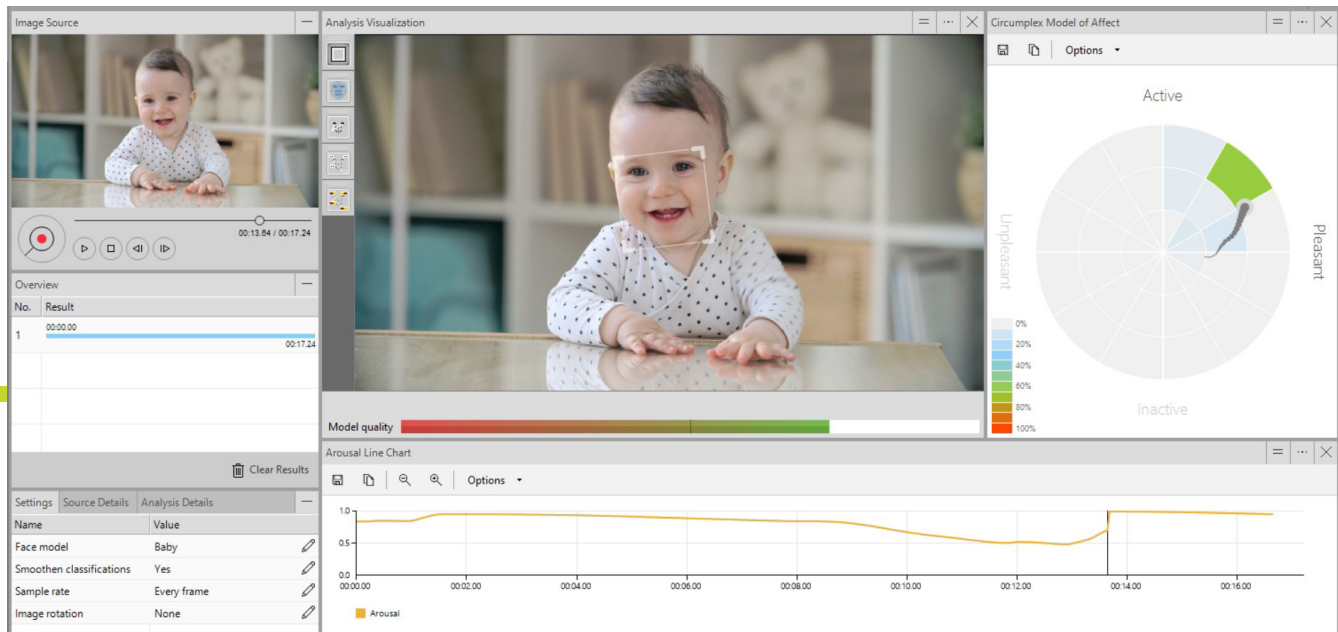
FaceReader classifies facial expressions according to the steps explained below.

1. **Face finding.** The position of the face in an image is found using a deep learning based face-finding algorithm [3], which searches for areas in the image having the appearance of a face at different scales.

2. **Face modeling.** FaceReader uses a facial modeling technique based on deep neural networks [4]. It synthesizes an artificial face model, which describes the location of 468 key points in the face. It is a single pass quick method to directly estimate the full collection of landmarks in the face.

   After the initial estimation, the key points are compressed using Principal Component Analysis. This leads to a highly compressed vector representation describing the state of the face.

3. **Face classification.** Then, classification of the facial expressions takes place by a trained deep artificial neural network to recognize patterns in the face [5]. FaceReader directly classifies the facial expressions from image pixels. Over 100,000 images that were manually annotated were used to train the artificial neural network.

The network was trained to classify the six basic or universal emotions described by Ekman [1]: happy, sad, angry, surprised, scared, and disgusted. Additionally, FaceReader can recognize a 'neutral' state and analyze 'contempt'. Moreover, the network was trained to classify the set of Facial Action Units available in FaceReader.

How FaceReader works. In the first step (left) the face is detected. A box is drawn around the face at the location where the face was found. In the next step an artificial face model is being created (middle). The model describes 468 key points in the face. Then, a trained deep artificial neural network does the classification (right).
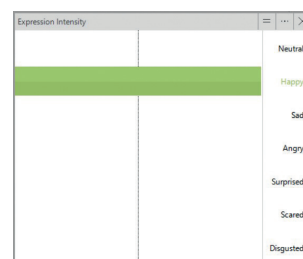
The following analyses can be carried out:

- Facial expression classification
- Valence calculation
- Arousal calculation
- Action Unit classification
- Subject characteristics analysis

There are multiple face models available in FaceReader. In addition to the general model, there are models for East-Asian people and babies (6-24 months). Based on the TFEID [6] dataset of East-Asian faces (268 images), the East-Asian model performs with an accuracy of 93%, while the General model achieves an accuracy of 91%.

Before you start analyzing facial expressions, you must select the face model which best fits the faces you are going to analyze. The model for babies is only available with Baby FaceReader.

CALIBRATION

For some people, FaceReader can have a bias towards certain expressions. You can calibrate FaceReader to correct for these person-specific biases. Calibration is a fully automatic mechanism. There are two calibration methods, participant calibration and continuous calibration. Participant calibration is the preferred method.
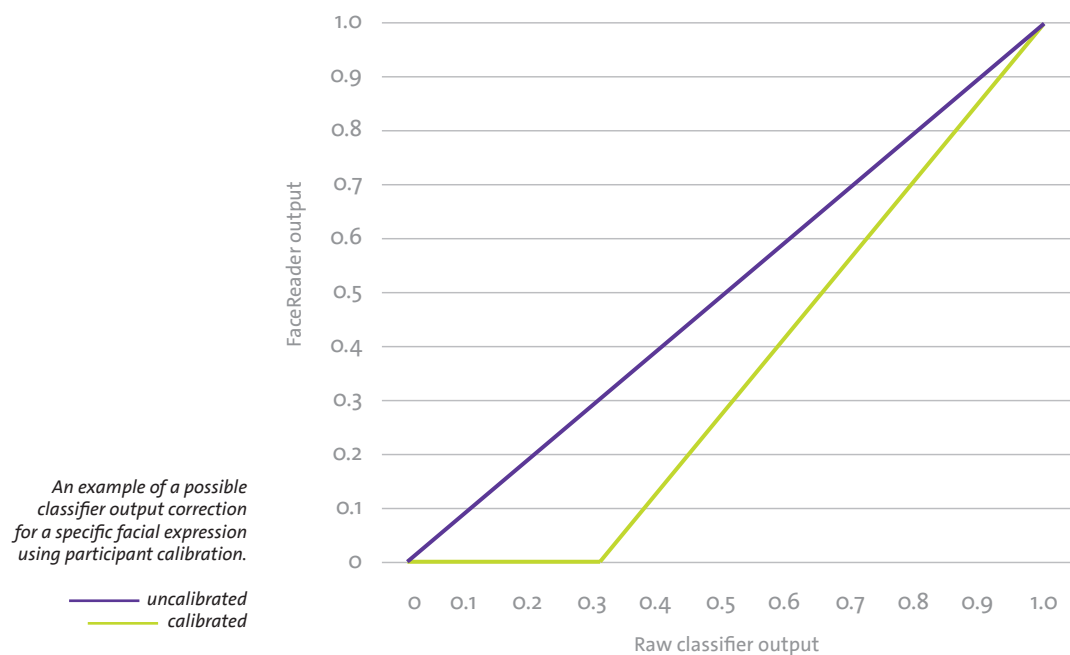
For Participant calibration, you use images or camera or video frames in which the participant looks neutral. The calibration procedure uses the image, or frame with the lowest model error and uses the expressions other than neutral found in this image for calibration.

With the Deep Face classification method FaceReader can analyze the face if part of it is hidden.

Consequently, the facial expressions are more balanced and personal biases towards a certain expression are removed. The effect can best be illustrated by an example. For instance, for a person a value of 0.3 for angry was found in the most neutral image. This means that for this test person 'angry' should be classified only when its value is higher than 0.3. The figure below shows how the classifier outputs are mapped to different values to negate the test person's bias towards 'angry'.

Continuous calibration continuously calculates the average expression of the test person. It uses that average to calibrate in the same manner as with the participant calibration.

An example of a possible classifier output correction for a specific facial expression using participant calibration.

— uncalibrated
— calibrated

FACEREADER'S OUTPUT

FaceReader's main output is a classification of the facial expressions of your test participant. These results are visualized in several different charts and can be exported to log files. Each expression has a value between 0 and 1, indicating its intensity. '0' means that the expression is absent, '1' means that it is fully present. FaceReader has been trained using intensity values annotated by human experts.

Facial expressions are often caused by a mixture of emotions and it is very well possible that two (or even more) expressions occur simultaneously with a high intensity. The sum of the intensity values for the expressions at a particular point in time is, therefore, normally not equal to 1.

VALENCE

Besides the intensities of individual facial expressions, FaceReader also calculates the valence. The valence indicates whether the emotional state of the subject is positive or negative. 'Happy' is the only positive expression, 'sad', 'angry', 'scared' and 'disgusted' are considered to be negative expressions. 'Surprised' can be either positive or negative and is therefore not used to calculate valence. The valence is calculated as the intensity of 'happy' minus the intensity of the negative expression with the highest intensity. For instance, if the intensity of 'happy' is 0.8 and the intensities of 'sad', 'angry', 'scared' and 'disgusted' are 0.2, 0.0, 0.3, and 0.2, respectively, then the valence is 0.8 − 0.3 = 0.5.

> The valence indicates whether the emotional state of the subject is positive or negative.

*Example of a Valence chart showing the valence (—) over time.*
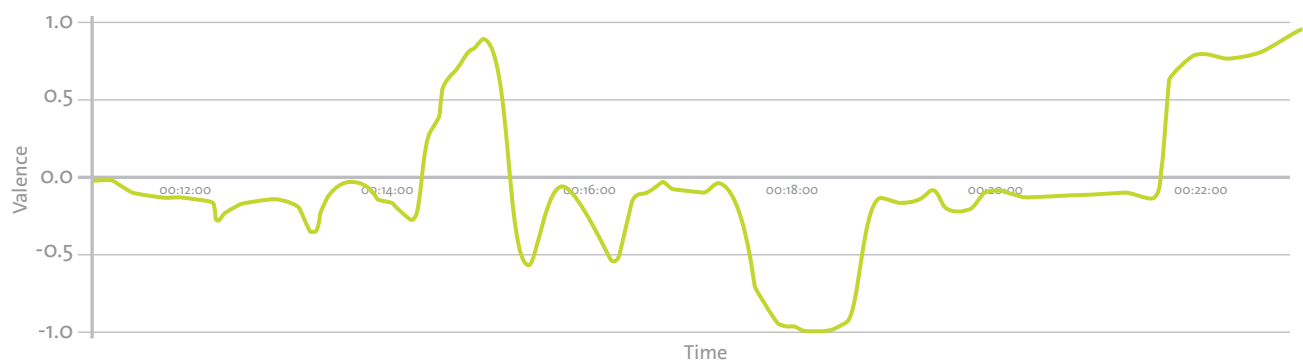
AROUSAL

Facereader also calculates facial arousal. It indicates whether the test participant is active (+1) or not active (0). Arousal is based on the activation of 20 Action Units (AUs) of the Facial Action Coding System (FACS) [7].

Arousal is calculated as follows:

1. The activation values (AV) of 20 AUs are taken as input. These are AU 1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 24, 25, 26, 27, and the inverse of 43. The value of AU43 (eyes closed) is inverted because it indicates low arousal instead of high arousal like the other AUs.

2. The average AU activation values (AAV) are calculated over the last 60 seconds. During the first 60 seconds of the analysis, the AAV is calculated over the analysis up to that moment. *AAV = Mean (AVpast 60 seconds)*

3. The average AU activation values (AAV) are subtracted from the current AU activation values (AV). This is done to correct for AUs that are continuously activated and might indicate an individual bias. This results in the Corrected Activation Values (CAV). *CAV = Max(0, AV − AAV)*

4. The arousal is calculated from these CAV values by taking the mean of the five highest values. *Arousal = Mean (5 max values of CAV)*

FaceReader's circumplex model of affect is based on the model described by Russel [8]. In the circumplex model of affect, the arousal is plotted against the valence. During the analysis, the current mix of expressions and Action Units is plotted with unpleasant/pleasant on the x-axis and active/inactive on the y-axis. A heatmap visualizes which of these expressions was present most often during the test.
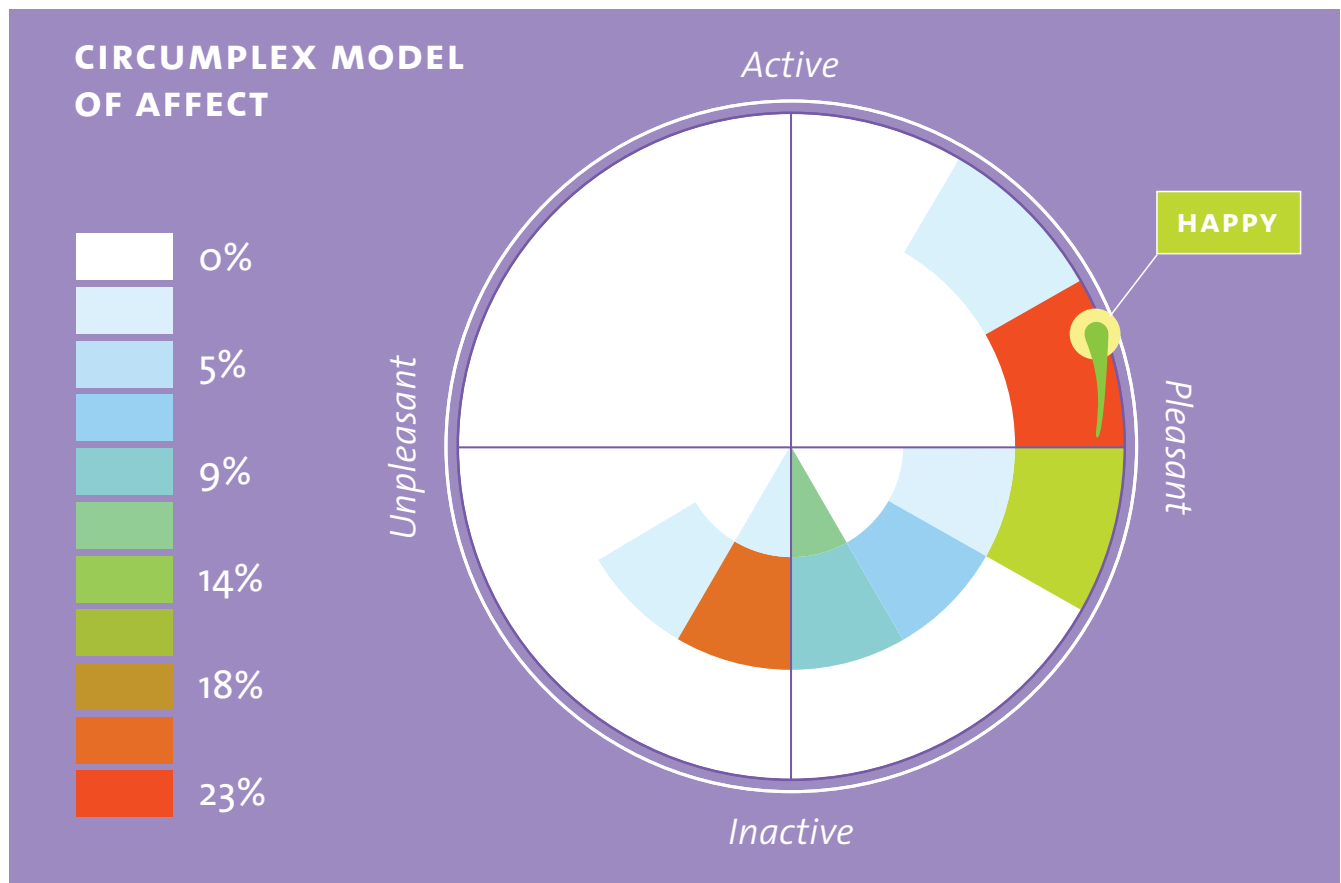
ADD-ON MODULES

Several add-on modules expand FaceReader software to meet your research needs.

THE PROJECT ANALYSIS MODULE

With the Project Analysis module, an add-on module for FaceReader, you can analyze the facial expressions of a group of participants. You can create these groups manually, but you can also create groups based on the value of independent variables. By default the independent variables Age and Gender are present, which allows you to create groups with males and females, or age groups. You can also add independent variables to create groups. Add, for example, the independent variable Previous experience to create a group with participants that worked with a program before and a group with those that did not.

You can mark episodes of interest, for example the time when the participants were looking at a certain video or image. This makes FaceReader a quick and easy tool to investigate the effect of a stimulus on a group of participants.

You can create your own filters and choose your own graphs and charts to display.
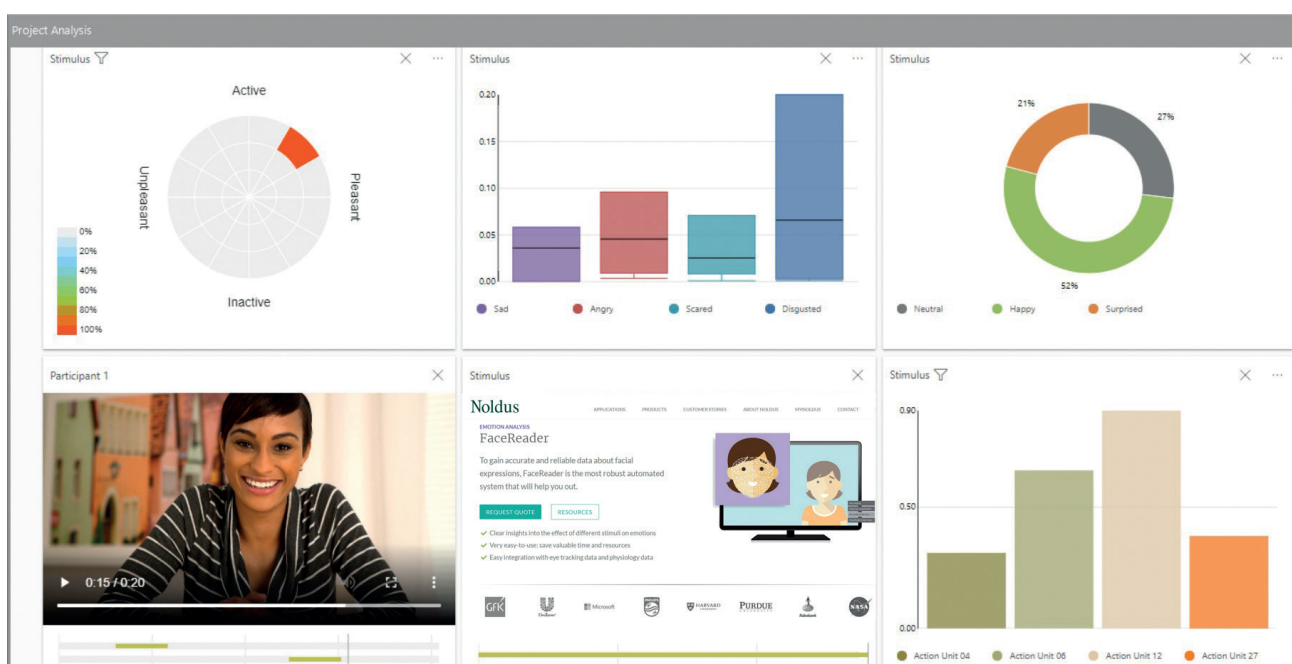
The numerical group analysis gives a numerical and graphical representation of the facial expressions, valence and arousal per participant group. With a click on a group name a T-test is carried out, to show in one view where the significant differences are.

The temporal group analysis shows the average expressions, valence and arousal of the group over time. You can watch this together with the stimulus video or image and the video of a test participant's face. This shows the effect of the stimulus on the participant's face in one view.

THE ACTION UNIT MODULE

Action Units are muscle groups in the face that are responsible for facial expressions. The Action Units are described in the Facial Action Coding System (FACS) that was published in 2002 by Ekman *et al.* [7]. With the Action Unit Module, FaceReader can analyze 20 Action Units. Intensities are annotated by appending letters, A (trace); B (slight); C (pronounced); D (severe) or E (max), also according to Ekman *et al.* [7]. Export in detailed log as numerical values is also possible.

Action Unit classification can add valuable information to the facial expressions classified by FaceReader. The emotional state Confusion is, for example, correlated with the Action Units 4 (Brow lowerer) and 7 (Eyelid tightener) [9]. Most Action Units are unilateral, that is, they can be visible in the left and/ or right part of the face. In these cases the intensity of the left and right part can be assessed independently.

Action Unit classification can add valuable information to the facial expressions classified by FaceReader.

THE RPPG (REMOTE PHOTOPLETHYSMOGRAPHY) MODULE

Photoplethysmography (PPG) is an optical technique that can be used to detect blood volume changes in the tissue under the skin. It is based on the principle that changes in the blood volume result in changes in the light reflectance of the skin. With each cardiac cycle the heart pumps blood to the periphery. Even though this pressure pulse is somewhat damped by the time it reaches the skin, it is enough to distend the arteries and arterioles in the subcutaneous tissue. In remote PPG (RPPG), FaceReader can detect the change in blood volume caused by the pressure pulse when the face is properly illuminated. The amount of light reflected is then measured. When reflectance is plotted against time, each cardiac cycle appears as a peak. This information can be converted to heart rate average and variability.

For more background information about FaceReader and Remote PPG, see Tasli *et al.* and Gudi *et al.* [10, 11, 12].

From the single heartbeats that are used to calculate the heart rate, the heart rate variability can be derived. HRV measures the variability between heartbeats, which can give important insights into the state of a person as this measure reflects changes in activation of the autonomic nervous system. One of the most commonly used measures of heart rate variability is the root mean square of successive differences (RMSSD).

The differences in time between single heart beats are called RR intervals or interbeat intervals.

$RR_n = beat_{n+1} - beat_n$

| RR$_1$ | RR$_2$ | RR$_3$ | RR$_4$ | RR$_5$ | RR$_6$ |

00.21.87   00.22.57   00.23.27   00.23.97   00.24.67   00.25.37   00.26.07   00.26.77   00.27.47   00.28.17

—— Heart Beats

Differences between the RR intervals are called successive differences (SD)
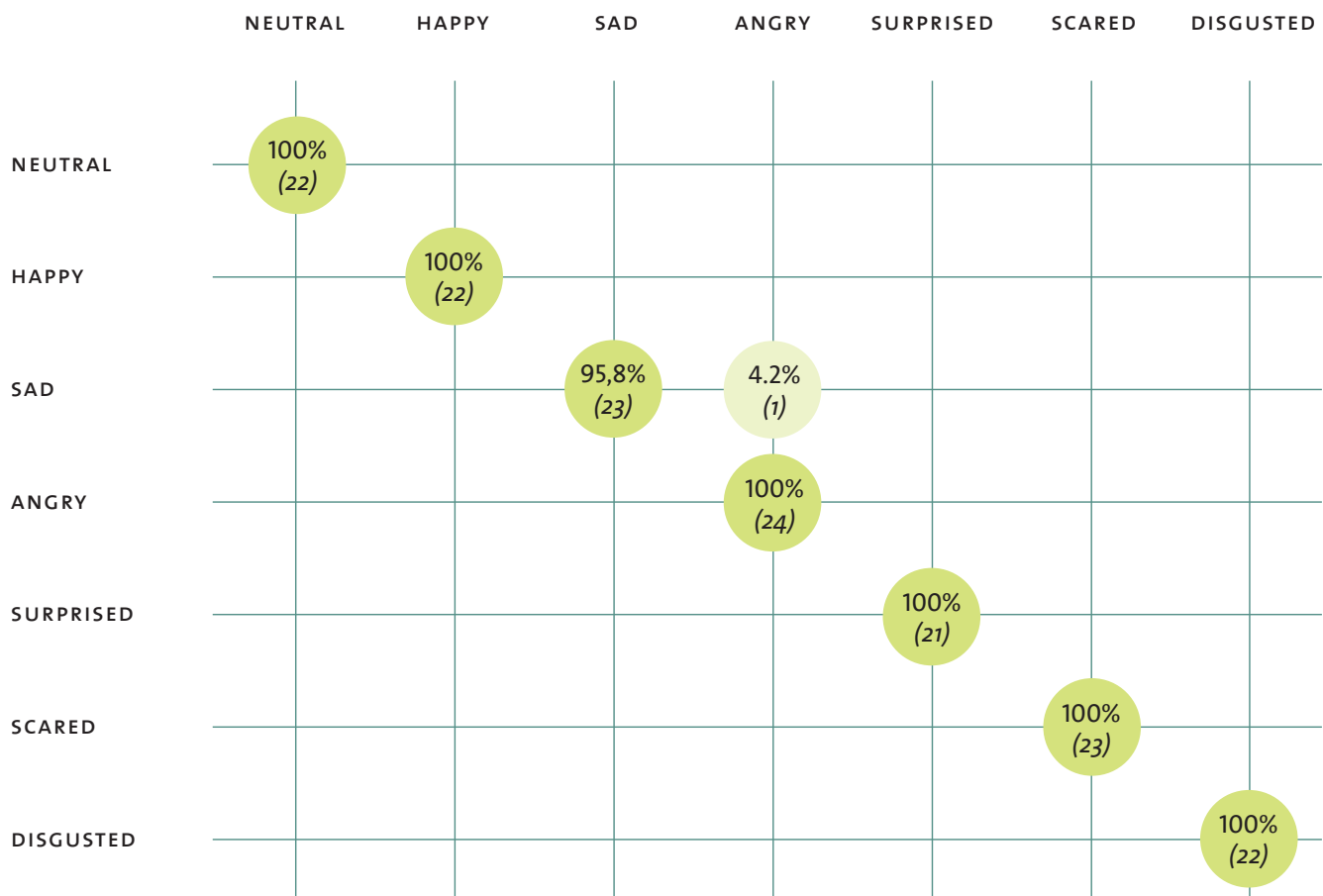
$SD_n = RR_{n+1} - RR_n$

The RMSSD is the root mean square of two successive RR intervals. The unit of RMSSD is the same as the time unit chosen in the equation above (usually it is milliseconds).

$$RMSSD = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(SD_n)^2}$$

Another measure of heart rate variability is SDNN (standard deviation of normal to normal R-R intervals). "Normal" means that abnormal beats, like ectopic beats (heartbeats that originate outside the right atrium's sinoatrial node), have been removed.

VALIDATION

To validate FaceReader, its results (version 9) have been compared with those of intended expressions. The figure below shows the results of a comparison between the analysis in FaceReader and the intended expressions in images of the Amsterdam Dynamic Facial Expression Set (ADFES) [13]. The ADFES is a highly standardized set of pictures containing images of eight emotional expressions. The test persons in the images have been trained to pose a particular expression and the images have been labeled accordingly by the researchers. Subsequently, the images have been analyzed in FaceReader. As you can see, FaceReader classifies all 'happy' images as 'happy', giving an accuracy of 100% for this expression.

	NEUTRAL	HAPPY	SAD	ANGRY	SURPRISED	SCARED	DISGUSTED
NEUTRAL	100% (22)						
HAPPY		100% (22)					
SAD			95,8% (23)	4.2% (1)			
ANGRY				100% (24)			
SURPRISED					100% (21)		
SCARED						100% (23)	
DISGUSTED							100% (22)

VALIDATION OF ACTION UNIT CLASSIFICATION

The classification of Action Units has been validated with a selection of images from the Amsterdam Dynamic Facial Expression Set (ADFES) [13] that consists of 23 models performing nine different emotional expressions (anger, disgust, fear, joy, sadness, surprise, contempt, pride, and embarrassment). FaceReader's classification was compared with manual annotation by two certified FACS coders. For a detailed overview of the validation, see the paper Validation Action Unit Module [14] that you can obtain from your Noldus sales representative.

POSED OR GENUINE

Sometimes the question is asked how relevant the results from FaceReader are if the program has been trained using a mixture of intended and genuine facial expressions. It is known that facial expressions can be different when they are intended or genuine. An intended smile is for example characterized by lifting the muscles of the mouth only, while with a genuine smile the eye muscles are also contracted [15]. On the other hand, one could ask what exactly a genuine facial expression is. Persons watching a shocking episode in a movie may show very little facial expressions when they watch it alone. However, they may show much clearer facial expressions when they watch the same movie together with others and interact with them. And children that hurt themselves often only start crying once they are picked up and comforted by a parent. Are those facial expressions that only appear in a social setting intended or genuine? Or is the question whether a facial expression is genuine or intended perhaps not so relevant?

FaceReader does not make a distinction whether a facial expression is acted or felt, authentic or posed. There is a very high agreement with facial expressions perceived by manual annotators and those measured by FaceReader [13]. One could simply say that if we humans experience a face as being happy, FaceReader detects it as being happy as well, irrespective from whether this expression was acted or not.

ARE FACIAL EXPRESSIONS ALWAYS THE SAME?

Another frequently asked question is whether the facial expressions measured by FaceReader are universal throughout ages, gender, and culture. There are arguments to say yes and to say no. The fact that many of our facial expressions are also found in monkeys supports the theory that expressions are old and therefore are independent of culture. In addition to this, we humans have until not so long ago largely been unaware of our own facial expression, because we did not commonly have access to mirrors. This means that facial expressions cannot be explained by copying behavior.

Furthermore, people that are born blind have facial expressions that resemble those of family members. This indicates that these expressions are more likely to be inherited than learned.

On the other hand, nobody will deny that there are cultural differences in facial expressions. For this purpose, FaceReader has different models, for example the East-Asian model. These models are trained with images from people of these ethnic groups. And it is true that with the East-Asian model FaceReader gives a better analysis of facial expressions of East-Asian people than with the general model and vice versa.

Feel free to contact us or one of our local representatives for more references, clients lists, or more detailed information about FaceReader and The Observer XT.

www.noldus.com

REFERENCES

1. Ekman, P. (1970). Universal facial expressions of emotion. *California Mental Health Research Digest*, **8**, 151-158.

2. Ekman, P.; Friesen, W.V. (1986). A new pan-cultural facial expression of emotion. *Motivation and emotion*, **10(2)**, 159-168.

3. Zafeiriou, S.; Zhang, C.; Zhang, Z. (2015). A survey on face detection in the wild: past, present and future. *Computer Vision and Image Understanding*. September 1, 2015, **138**, 1-24.

4. Bulat, A.; Tzimiropoulos, G. (2017). How far are we from solving the 2d & 3d face alignment problem? (and a dataset of 230,000 3d facial landmarks). In *Proceedings of the IEEE International Conference on Computer Vision 2017 (*pp. 1021-1030).

5. Gudi, A.; Tasli, H.E.; Den Uyl, T.M.; Maroulis, A. (2015). Deep learning based facs action unit occurrence and intensity estimation. In *2015 11th IEEE international conference and workshops on automatic face and gesture recognition (FG) 2015 May 4* (Vol. 6, pp. 1-5).

6. Chen, L.F.; Yen, Y.S. (2007) '*Taiwanese facial expression image database'.* Ph.D. dissertation, Brain Mapping Lab., Inst. Brain Sci., Nat. Yang-Ming Univ., Taipei, Taiwan.

7. Ekman, P.; Friesen, W.V.; Hager, J.C. (2002). *FACS manual. A Human Face.*

8. Russell, J. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, **39**, 1161–1178.

9. Grafsgaard, J.F.; Wiggins, J.B.; Boyer, K.E.; Wiebe, E.N.; Lester, J.C. (2013). Automatically recognizing facial expression: Predicting engagement and frustration. In *Proceedings of the 6th International Conference on Educational Data Mining* (pp. 43-50).

10. Tasli H. E., Gudi, A., Den Uyl, T.M. (2014a). Integrating Remote PPG in Facial Expression Analysis Framework. *Proceedings of the 16th International Conference on Multimodal Interaction*. Istanbul, November 12-16, 2014 (pp. 74-75).

11. Tasli H. E., Gudi, A., Den Uyl, T.M. (2014). Remote PPG based vital sign measurement using adaptive facial regions. I*nternational Conference on Image Processing 2014* (pp. 1410-1414).

12. Gudi, A.; Bittner, M.; Lochmans, R.; Gemert, J. van (2019). Efficient Real-Time Camera Based Estimation of Heart Rate and Its Variability. http://arxiv-export-lb.library.cornell.edu/abs/1909.01206

13. Schalk, J. van der; Hawk, S.T.; Fischer, A.H.; Doosje, B.J. (2011). Moving faces, looking places: The Amsterdam Dynamic Facial Expressions Set (ADFES). *Emotion*, **11**, 907-920. DOI: 10.1037/a0023853

14. Ivan, P.; Gudi, A. (2016). *Validation Action Unit Module.* White paper.

15. Ekman, P.; Friesen, W.V. (1982). Felt, false, and miserable smiles. *Journal of Nonverbal Behavior*, **6(4)**, 238-252.

N