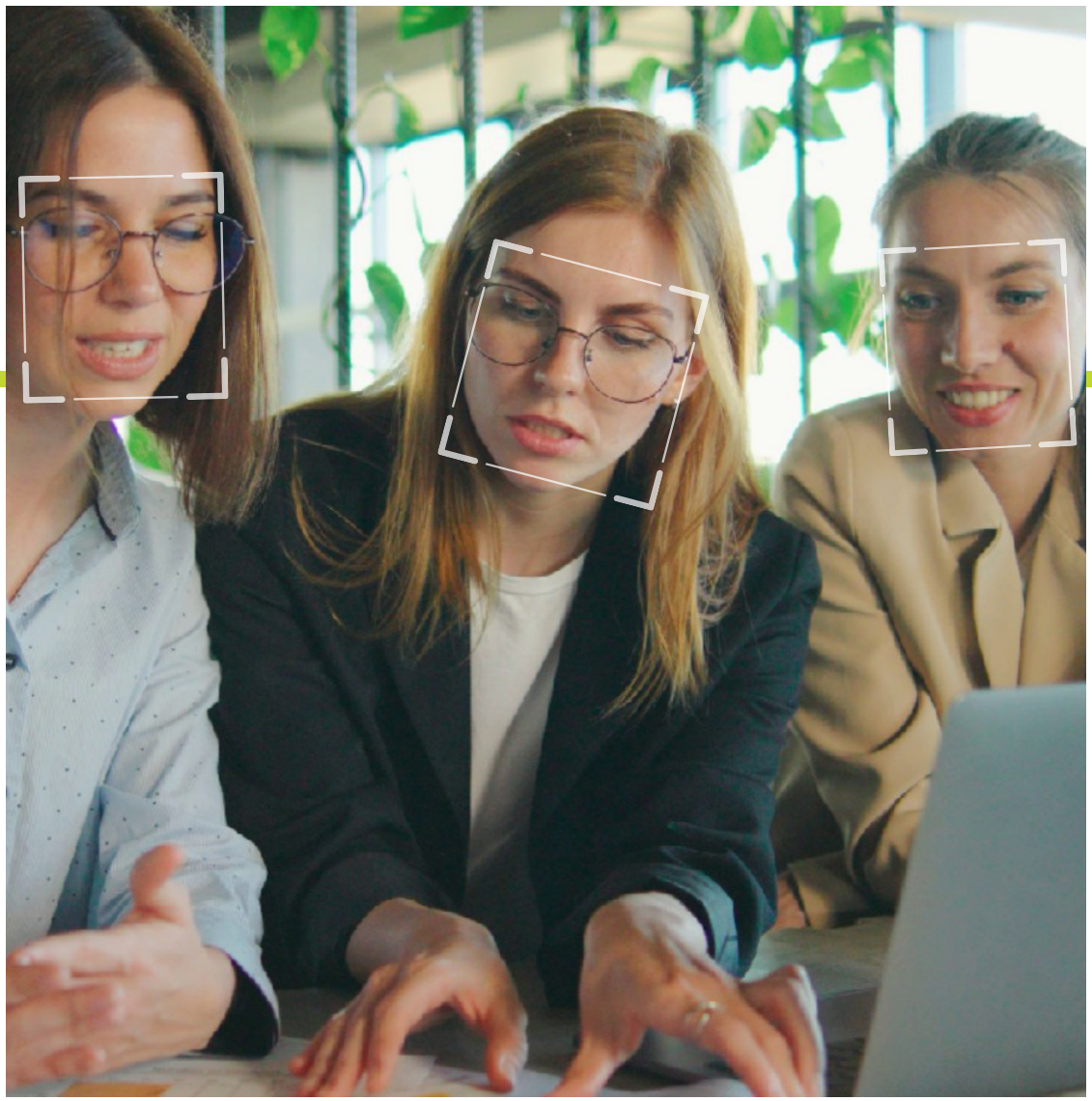


FaceReader Methodology Note

A white paper by Noldus Information Technology

CONTENTS

WHAT IS FACEREADER?

FaceReader made history in 2007 as the world's first-ever software system capable of automatically detecting and analyzing human facial expressions. Since its debut, FaceReader has grown into a powerful platform for behavioral research. It now enables detailed analysis of emotions, attention, physiological signals, and group dynamics.

These key features make FaceReader a complete suite for a wide range of research applications:

CLASSIFY EXPRESSIONS

Recognize key emotions

Automatically detect basic emotions: happy, sad, angry, surprised, scared, disgusted, and neutral. Based on the model developed by psychologist Paul Ekman [1].

FACIAL ACTION UNITS (FACS)

Reveal hidden emotional nuances

Understand the subtle mechanisms underneath facial expressions by measuring the activity of individual facial muscles.

MULTIPLE SUBJECTS

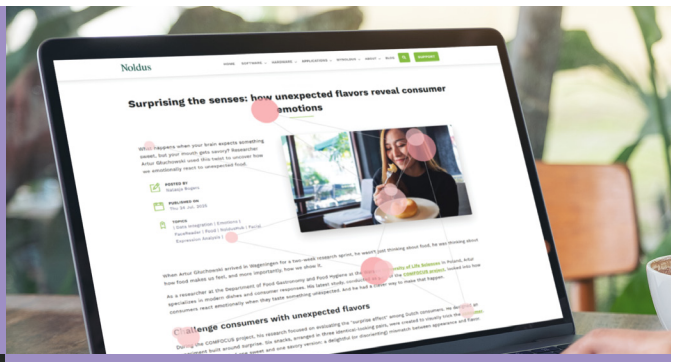
Study multiple people at once

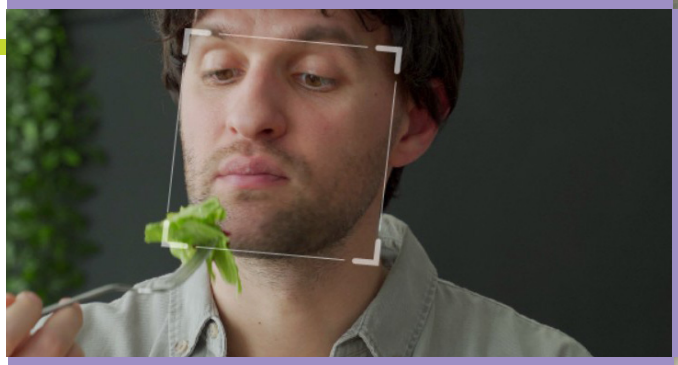
Speed up research while capturing richer, more natural group responses and shared emotional dynamics.

EYE TRACKING

Follow every glance and focus point

Use eye tracking with visual outputs like heatmaps and gaze plots. No bulky hardware required.





CONSUMPTION BEHAVIOR

Catch every bite, sip, and subtle reaction

Spot micro-actions (sipping, chewing, etc.) to capture honest, unconscious responses.

BIOMETRIC SIGNALS

Measure without sensors

Use webcam-based signals like heart rate and breathing to explore stress, attention, or emotional arousal in context.

VOICE ANALYSIS

Hear what emotion sounds like

FaceReader now also measures voice intensity and speech patterns to uncover how people really feel, beyond their words.

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 in several steps:

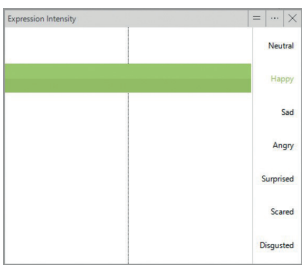
1. **Face finding.** The position of the face in an image or video is found using a deep learning-based face-finding algorithm [2], which searches for areas in the image having the appearance of a face.
2. **Face modeling.** FaceReader uses a facial modeling technique based on deep neural networks [3]. It synthesizes an artificial face model, which describes the location of almost 500 key points in the face. The predicted key points of such a model are learned from a database of annotated images. It is a single pass quick method to directly estimate the full collection of landmarks in the face.

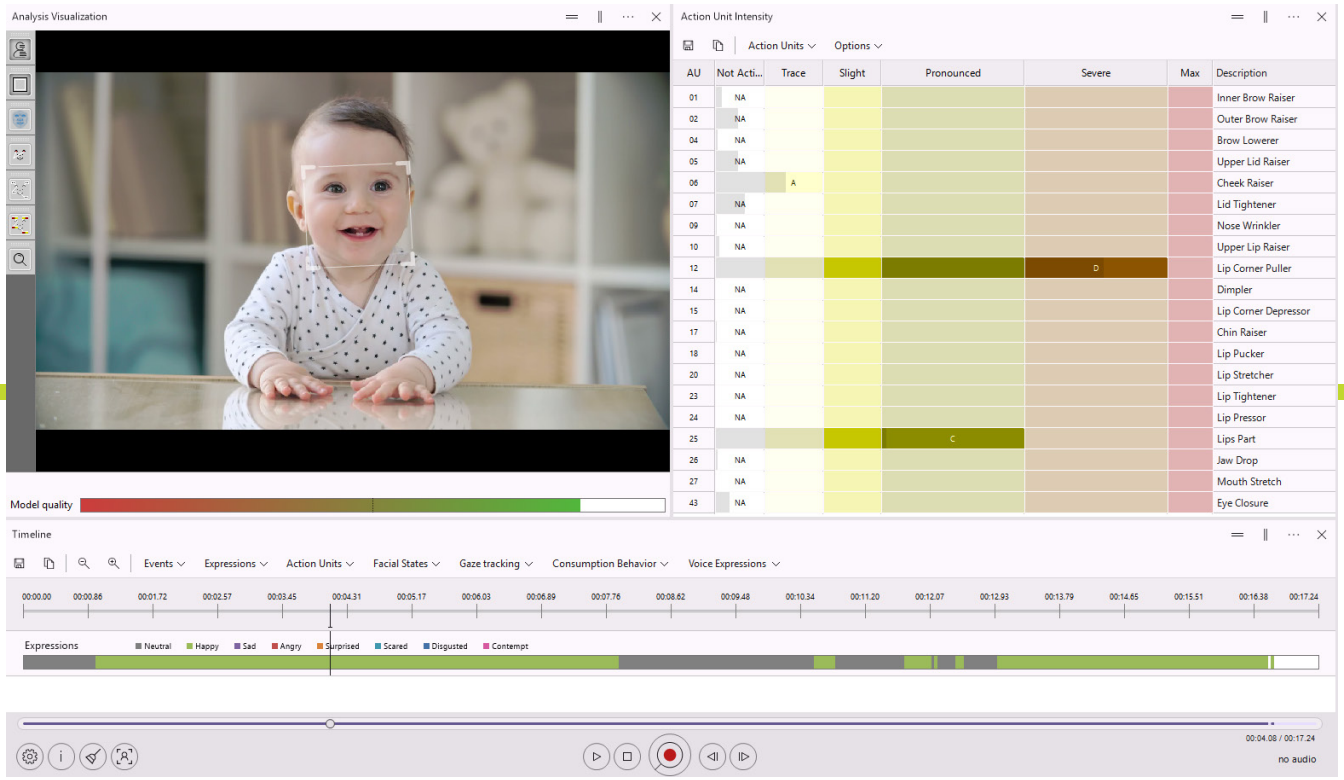
   After the initial estimation, the key points are compressed using Principal Component Analysis. This leads to a highly compressed vector representation describing the state of the face.
3. **Facial expression classification.** Classification of the facial expressions is done by a trained deep artificial neural network to recognize patterns in the face [4]. FaceReader directly classifies the facial expressions from image pixels.

   More than 100,000 manually annotated images of people's faces from around the world were used to train the artificial neural network. This large and diverse dataset helps the system perform well across a wide range of individuals and experimental conditions.

   The network was trained to classify the six basic or universal emotions described by Ekman [1]: happy, sad, angry, surprised, scared, disgusted and a neutral state. Additionally, 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

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

There are various face models available in FaceReader. In addition to the general model, there are models for East-Asian people and babies (6-24 months). Evaluated on the TFEID [5] 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

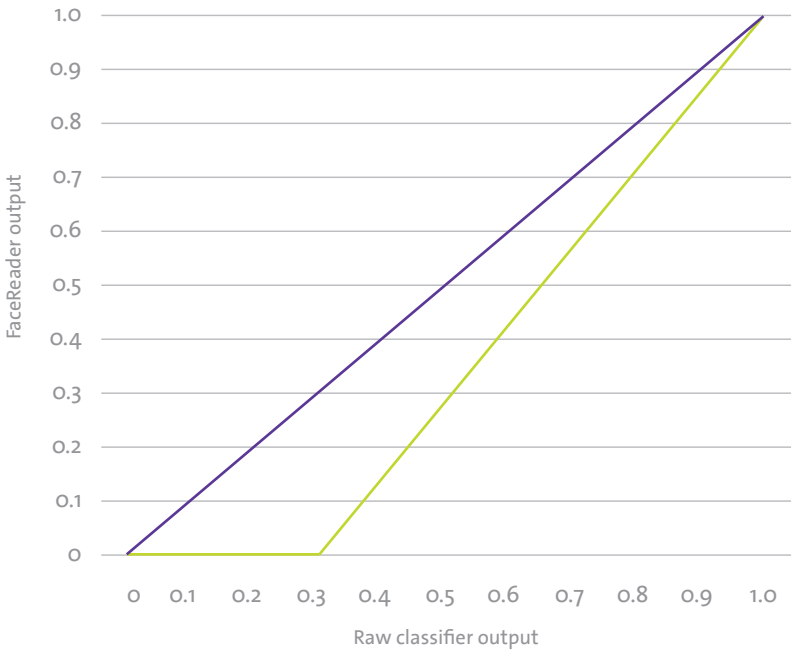
For some people, FaceReader can have a bias towards certain expressions. You can calibrate FaceReader to correct these person-specific biases. Calibration is a fully automatic mechanism. There are two calibration methods, participant calibration and continuous calibration.

1. **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.

Calibration is a fully automatic mechanism. It makes your data more accurate by adjusting for each person's natural facial expression style.

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, a person with 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'.

2. **Continuous calibration.** This method automatically tracks and updates the average facial expression of a test participant during the session. FaceReader uses this ongoing average as a baseline to adjust the results - just like in participant calibration. This helps correct for personal expression habits and makes the analysis more accurate over time.



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

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.

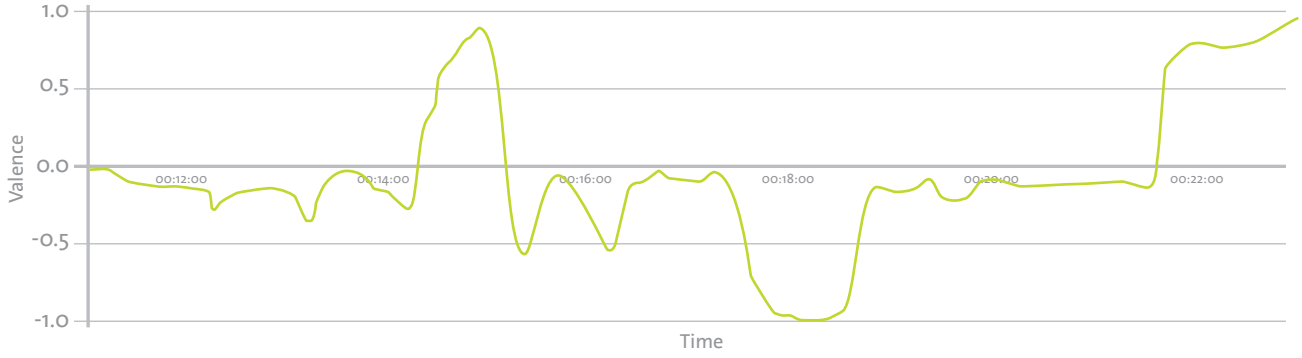
Facial expressions are often the result of a mixture of emotions, and it is possible for two or more expressions to occur simultaneously, although typically not at a high intensity. The sum of the intensity values for the expressions at a particular point in time is normally not equal to 1.

VALENCE

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

Besides the intensities of individual facial expressions, FaceReader also calculates valence. Valence indicates whether the emotional state of the subject is positive or negative. 'Happy' is classified as the only positive expression, whereas 'sad', 'angry', 'scared', and 'disgusted' are categorized as 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.

Example of a Valence chart showing valence over time.



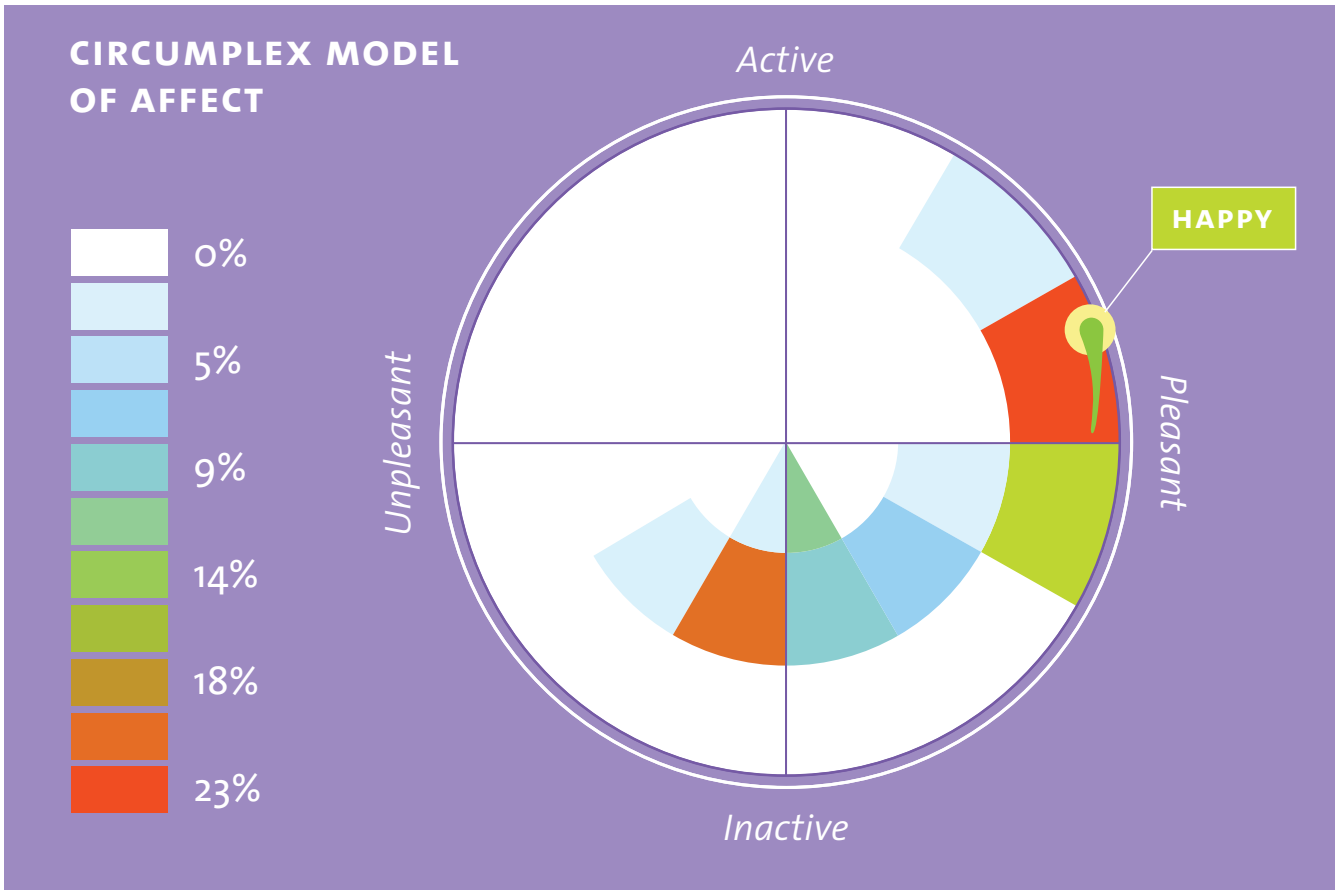
AROUSAL

FaceReader also measures facial arousal, indicating whether a participant's facial activity is active (+1) or inactive (0). Arousal is based on the activation of 20 Action Units (AUs) of the Facial Action Coding System (FACS) [6].

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, 18, 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 arousal is calculated by taking the mean of the three highest values. *Note: Contempt is not included in the calculation of arousal.*

FaceReader's circumplex model of affect is based on the model described by Russel [7]. 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.



EXPAND YOUR RESEARCH OPTIONS

Add extra layers of insight with features built for professional research applications.

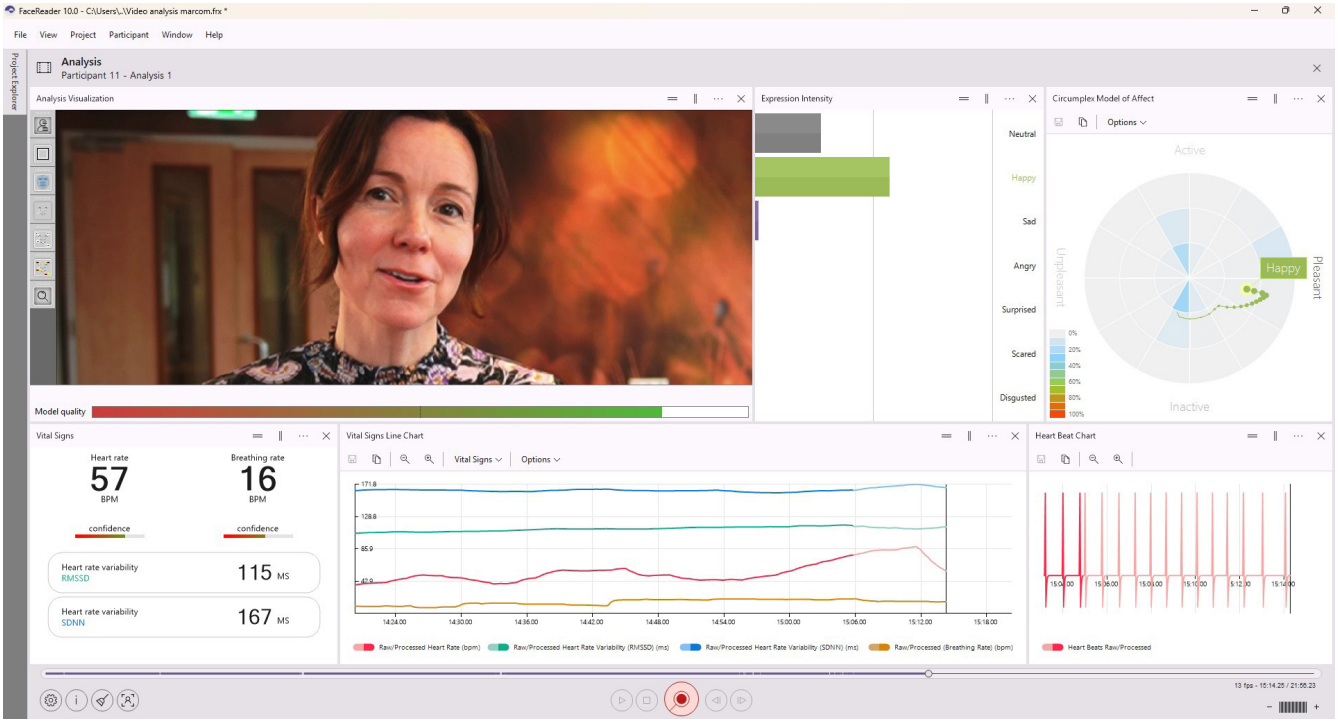
IN-DEPTH PROJECT ANALYSIS

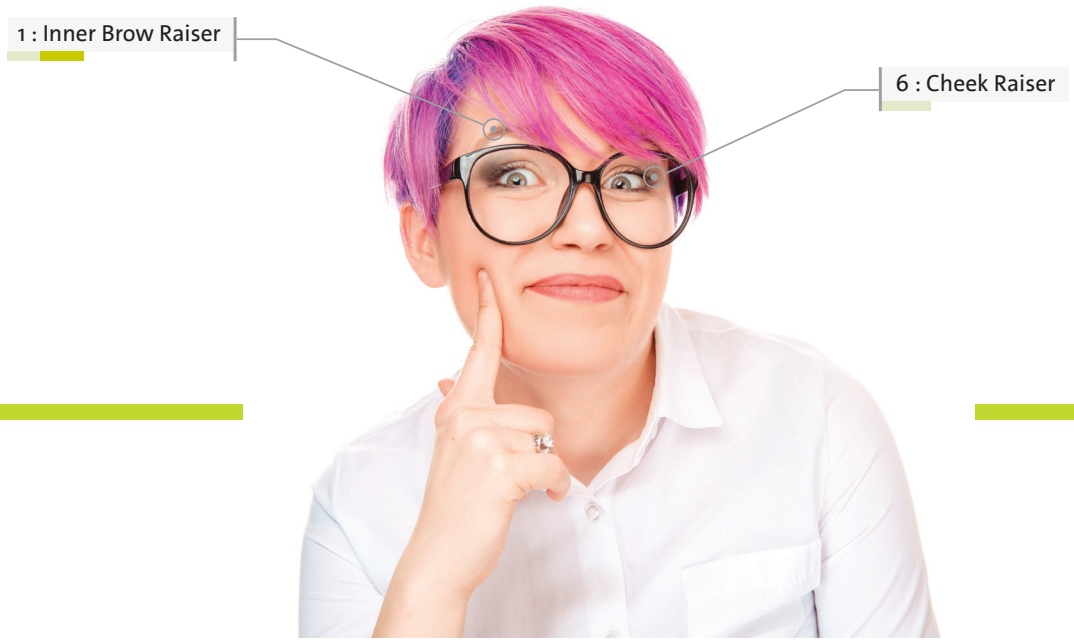
With your smart study assistant in FaceReader, you can analyze the facial expressions of a group of participants. You can create these groups manually, but you can also create groups based on the value of independent variables.

You can add these independent variables yourself. 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.

With FaceReader, you can instantly tag key moments, such as when participants engage with a video or image, making it simple and fast to understand how your audience responds to different stimuli.

You can create your own filters and choose your own graphs and charts to display.





The numerical group analysis shows both numbers and graphs for the facial expressions, valence, and arousal to help you compare groups of participants. With one click, you can run a T-test, to quickly spot where the biggest differences between groups are.

The temporal group analysis shows how the group's average emotional responses shift over time. You can view these changes alongside the stimulus video or image and the participants' facial recordings, making it easy to see exactly how each stimulus influenced their expressions.

UNDERSTANDING EXPRESSIONS THROUGH ACTION UNIT CLASSIFICATION (FACS)

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

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.* [6].

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.* [6]. Export in detailed log as numerical values is also possible.

1 Inner Brow Raiser*
2 Outer Brow Raiser*
4 Brow Lowerer**
5 Upper Lid Raiser*
6 Cheek Raiser*
7 Lid Tightener*
9 Nose Wrinkler
10 Upper Lip Raiser
12 Lip Corner Puller*
14 Dimpler
15 Lip Corner Depressor*
17 Chin Raiser
18 Lip Pucker
20 Lip Stretcher*
23 Lip Tightener
24 Lip Pressor
25 Lips Part
26 Jaw Drop
27 Mouth Stretch
43 Eyes Closed*

** Unilateral analysis is possible*
***In Baby FaceReader AU4 is replaced by AU3+4 (Brow knitting and knotting)*

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) [8].

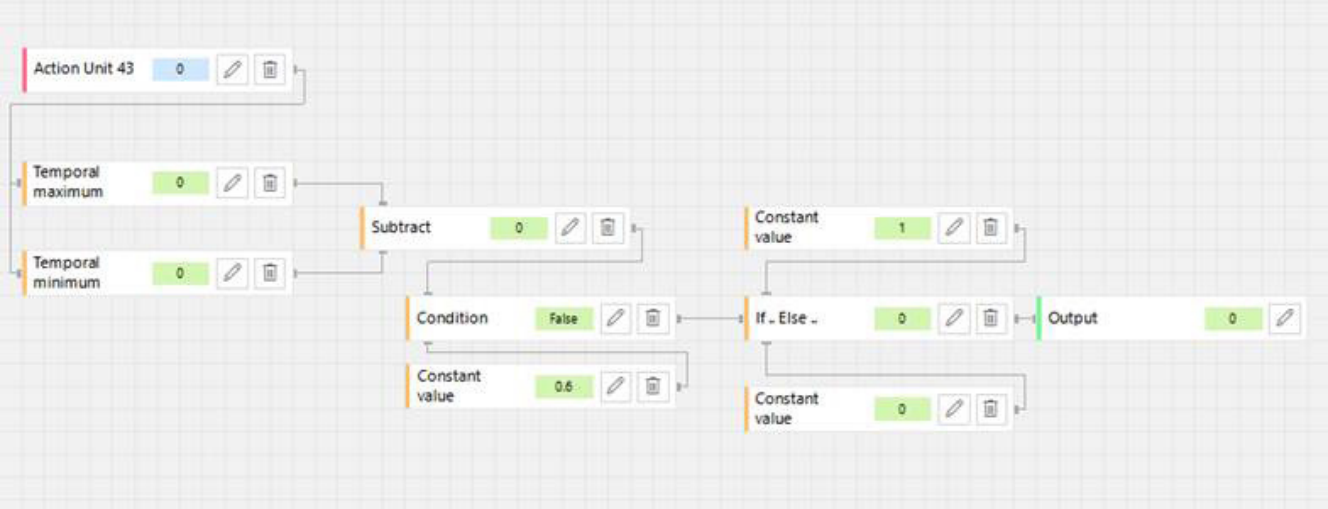
Most facial muscle movements can appear on just one side of the face. FaceReader measures the intensity separately for the left and right side, making it possible to detect asymmetrical expressions. This can be useful in studies on authenticity, individual expression patterns, or neurological effects.

CREATE, CUSTOMIZE, AND REFINE

In some research contexts, basic emotion categories are not enough. That's why you can create user-defined expressions — tailored emotion labels based on your own data and specific research needs. A separate white paper, Best practices when creating custom expressions, explains how to set this up.

For a full overview of FaceReader's classifications, input formats, and output variables, the Technical Specifications document is available for download.

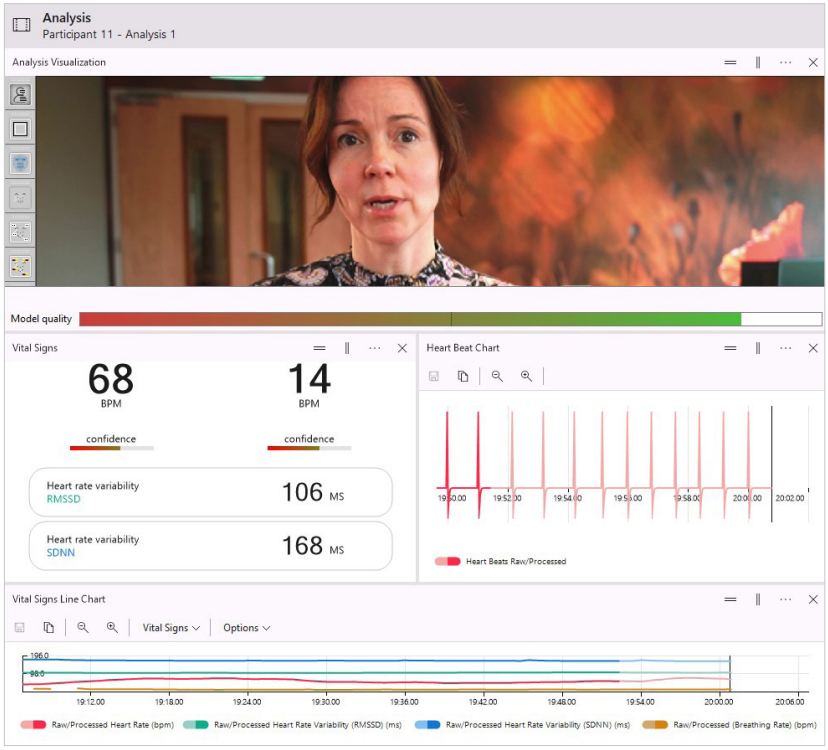
Formula for the custom expression blinking: a relevant psychophysiological metric that is scored in FACS as AU45. Blinking is quickly (within less than 0.5 second) closing and opening the eyes.



NON-INVASIVE TRACKING OF VITAL SIGNS

FaceReader offers non-contact physiological measurements using standard video. Heart rate is measured through remote photoplethysmography (RPPG), which detects subtle changes in skin color caused by blood flow. These changes, invisible to the eye, are captured frame by frame and used to calculate heart rate and heart rate variability (HRV). This is the natural variation in time between successive heartbeats. HRV is a widely used indicator of autonomic nervous system activity. A higher HRV is often linked to calmness and resilience, while a lower HRV can reflect stress or reduced flexibility.

In addition, breathing rate can be estimated by analyzing small movements in the upper body. To do this, the participant's chest and shoulders must be clearly visible in the video. Because breathing is slower and more subtle than heart activity, the software needs about 15 seconds for calibration, and another 15 seconds to produce stable readings.



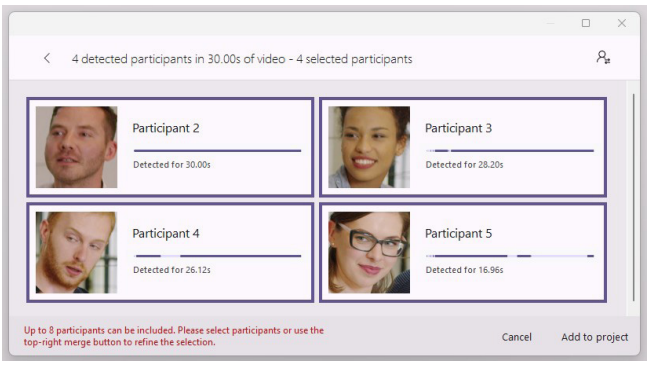
Together, these features offer valuable insight into someone's internal state, without needing physical sensors, wires, or wearables.

For more background information about FaceReader and the non-invasive way of tracking vital signs, download the white paper.

TRACK MULTIPLE FACES IN REAL-WORLD SETTINGS

Multi-subject analysis lets you quantify facial expressions from a group of subjects in a single recording, supporting research studies on audience engagement, emotional contagion, or interaction dynamics in natural settings. It allows you to analyze facial expression data from up to eight participants captured in the same pre-recorded video.

FaceReader processes each video frame using an appearance-based deep learning model to generate face descriptors for all detected faces. These



Screenshots of a group discussion analyzed in FaceReader and the selection of participants.

descriptors are used to create face tracks – sequences of detections – based on spatial consistency and visual similarity. When a person moves, a new track starts.

After the entire video has been analyzed, a clustering algorithm merges tracks with high appearance similarity to group segments that belong to the same person, even when participants have left or (re-)entered the scene.

The resulting participant analyses can be reviewed, merged if needed, and added to the project, similar to regular single-face analyses. Multi-subject analysis supports an offline workflow to conduct a facial analysis on multiple faces. To be clear: detailed analysis with the advanced FaceReader research options requires analysis per single face.

Gain group-level insights

Multi-subject analysis enables deeper insights into group-level dynamics when you combine this with the functionalities of project analysis:

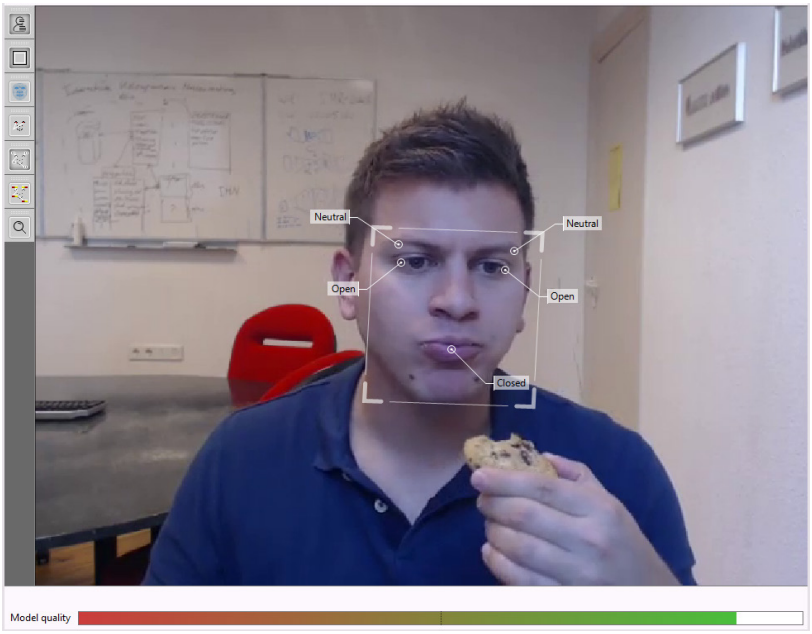
- **Side-by-side view.** Plots the intensity of individual parameters (e.g. valence, happy, surprise) for all selected participants over time, facilitating moment-by-moment comparison.
- **Synchrony chart.** Displays the rolling 10-second Pearson correlation of a chosen parameter across participants, highlighting segments when emotional responses align (e.g. shared laughter).
- **Correlation table.** Presents the Pearson correlation coefficients between each participant pair over the full session, showing overall response similarity.

Speed up research while capturing richer, more natural group responses and shared emotional dynamics.

For researchers interested in the emotional interplay among individuals, multi-subject analysis offers a valuable extension to traditional single-subject facial expression recognition.

Multi-subject analysis cannot be combined with Baby FaceReader.

Synchrony chart showing the rolling 10-second Pearson correlation of valence between four participants over time.



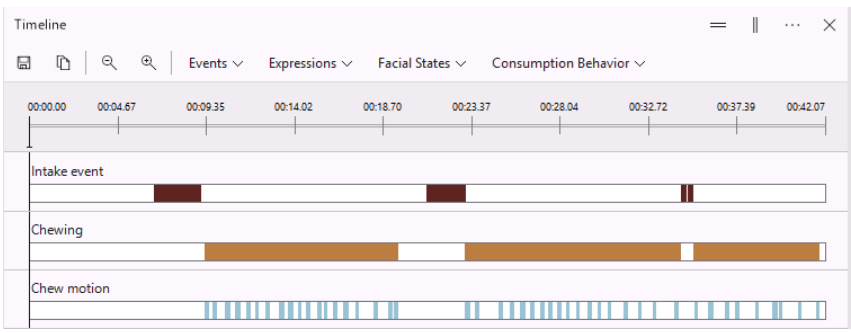
Catch every bite, sip, and subtle reaction. Spot micro-actions (sipping, chewing, etc.) to capture honest, unconscious responses.

DETECT EATING AND DRINKING BEHAVIOR AUTOMATICALLY

FaceReader includes functionality to automatically detect and classify moments of eating and drinking during live or pre-recorded videos. It identifies three types of behavior: intake events (bites or sips), periods of chewing, and individual chew movements. This makes it possible to link consumption behavior to other facial analysis measures, enhancing research on product evaluation and natural eating behavior in real-life contexts.

By integrating facial analysis with automated detection of eating and drinking behavior, this functionality provides researchers with an efficient tool to study consumer responses to food and beverages, without the need for manual annotation.

This functionality combines several detection methods. First, it identifies when a participant's hand moves toward the mouth. Simultaneously, the system monitors facial landmarks to track mouth opening and jaw movement. A temporal classifier then analyzes these motion patterns to distinguish chewing from similar actions such as talking, and to count discrete chew events. If the participant briefly changes position, FaceReader continues tracking as long as the face and hands remain sufficiently visible.



During analysis, consumption behavior events are visualized on FaceReader's timeline and summarized in the Consumption Behavior Statistics window. This interface provides direct counts of intake events and total chew movements. The data can be exported together with other facial analysis metrics for further processing.

Measuring consumption behavior is designed for single-subject recordings using video or webcam input. It is not compatible with the analysis of multiple subjects or Baby FaceReader. It is good to know that the use of straws or cutlery may reduce detection accuracy, as the current system is optimized for clear hand-to-mouth gestures.

UNDERSTAND EMOTIONS THROUGH VOICE

In addition to analyzing facial expressions and other functionalities, FaceReader also supports voice-based emotion recognition, adding another valuable layer to your research.

This technique - known as Speech Emotion Recognition (SER) - uses machine learning to detect emotions based on how something is said rather than what is said. By analyzing vocal cues such as pitch, volume, tone, and speaking rate, FaceReader reveals a speaker's emotional state without depending on the actual words.

> See how people feel, sound, and react all at once. This gives you a more complete picture of behavior, without switching tools or losing context.

Want to learn more? A separate white paper, Voice Analysis in FaceReader, explains this functionality in greater detail.

EYE TRACKING WITH A STANDARD WEBCAM

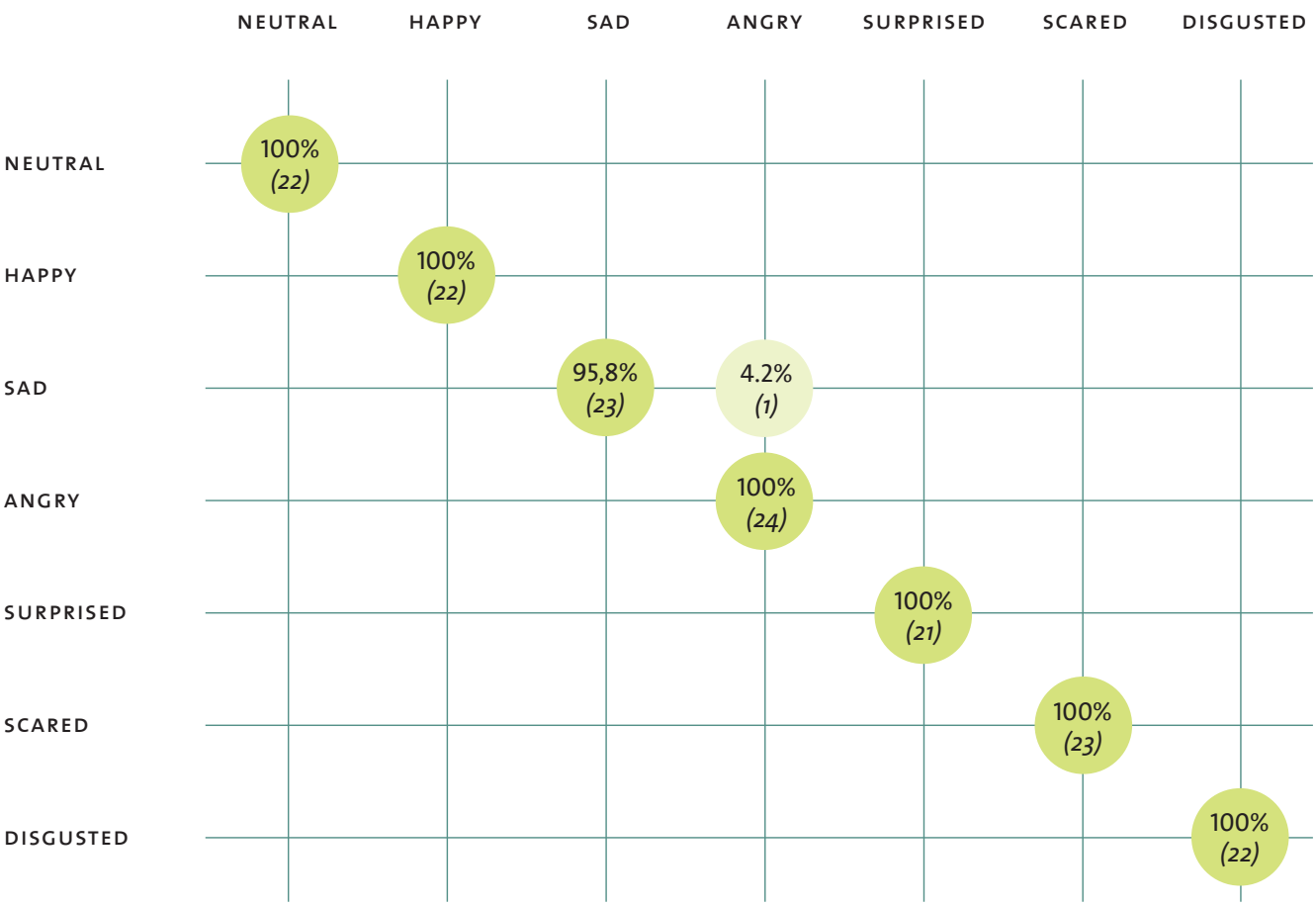
What people look at often tells you what matters to them — whether it's instinctive attention or a conscious choice. FaceReader makes it easy to capture that information with built-in eye tracking that works on any regular webcam. No special hardware required.

This makes eye tracking more flexible and accessible, whether you're studying attention in psychology, user behavior in UX research, or visual engagement in marketing.

Curious how it works and how accurate it is? The white paper Webcam-based eye tracking explains the method in detail, shares validation data, and shows how you can get started in your own study.

VALIDATION

To validate FaceReader's performance, its output was compared with the intended emotional expressions in a standardized dataset. The figure below presents the results of this comparison using images from the Amsterdam Dynamic Facial Expression Set (ADFES) [9]. The ADFES contains a highly standardized collection of images depicting eight distinct emotional expressions. Individuals in these images were trained to display specific expressions, which were subsequently labeled by researchers. These images were then analyzed using FaceReader. As shown, FaceReader correctly classified all images labeled as 'happy' as 'happy', achieving an accuracy rate of 100% for this expression.



VALIDATION OF ACTION UNIT CLASSIFICATION

The classification of Action Units was validated using a subset of images from the Amsterdam Dynamic Facial Expression Set (ADFES) [9], which features 23 models displaying nine emotional expressions: anger, disgust, fear, joy, sadness, surprise, contempt, pride, and embarrassment. FaceReader's automated classifications were compared to manual annotations performed by two certified FACS coders. For a comprehensive description of the validation process, please refer to the paper Validation of the Action Unit Module [10], available from your Noldus sales representative.

POSED OR GENUINE

You may wonder how relevant FaceReader's results are when it's trained on both posed and genuine facial expressions. After all, we know facial expressions can look quite different depending on whether someone is deliberately posing or showing genuine emotion.

Want to know more on different scientific theories on emotion? Check out this blog post.

Take smiles, for example. An intended smile usually involves only the muscles around the mouth, while a genuine smile - often called a Duchenne smile - also engages the muscles around the eyes [11]. But what really counts as a 'genuine' facial expression?

Think about this: someone watching a shocking scene in a movie alone might barely react, but watching the same scene with friends could bring out much more obvious expressions. Or consider children who hurt themselves — they often don't start crying until a parent picks them up and comforts them.

So are these expressions that appear mainly in social settings truly genuine, or are they intended? Or maybe the distinction between genuine and intended expressions isn't as clear-cut — or even as relevant — as we think.

Want to learn more? Check out our blog post Smile like you mean it.

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 [9]. 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?

A common question about FaceReader is whether facial expressions are universal across age, gender, and culture. The answer is both yes and no.

On one hand, there's strong evidence that many facial expressions are biologically hardwired rather than learned. For example, expressions similar to ours are observed in monkeys, suggesting that they are ancient and not culturally dependent. Historically, humans also had limited access to mirrors, which means facial expressions could not have been developed through mimicry alone. Furthermore, individuals born blind display facial expressions that closely resemble those of their family members, pointing to a genetic rather than a learned origin.

On the other hand, cultural differences in expression and interpretation certainly exist. This is why FaceReader offers different models, such as its East Asian model, which is specifically trained on images of people from that demographic. Using these tailored models produces more accurate results for specific populations compared to the general model, and vice versa

> Want to know more on different scientific theories on emotion? Check out this blog post.

REFERENCES

1. Ekman, P. (1970). Universal facial expressions of emotion. *California Mental Health Research Digest*, **8**, 151-158.
2. 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.
3. 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).
4. 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).
5. 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.
6. Ekman, P.; Friesen, W.V.; Hager, J.C. (2002). *FACS manual. A Human Face.*
7. Russell, J. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, **39**, 1161–1178.
8. 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).
9. 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
10. Ivan, P.; Gudi, A. (2016). *Validation Action Unit Module.* White paper.
11.. Ekman, P.; Friesen, W.V. (1982). Felt, false, and miserable smiles. *Journal of Nonverbal Behavior*, **6(4)**, 238-252.

INTERNATIONAL HEADQUARTERS
Noldus Information Technology bv
Wageningen, The Netherlands
Phone: +31-317-473300
Fax: +31-317-424496
E-mail: contact@noldus.com

NORTH AMERICAN HEADQUARTERS
Noldus Information Technology Inc.
Leesburg, VA, USA
Phone: +1-703-771-0440
Toll-free: 1-800-355-9541
Fax: +1-703-771-0441
E-mail: info@noldus.com

REPRESENTATION
We are also represented by a worldwide network of distributors and regional offices. Visit our website for contact information.

NOLDUS.COM