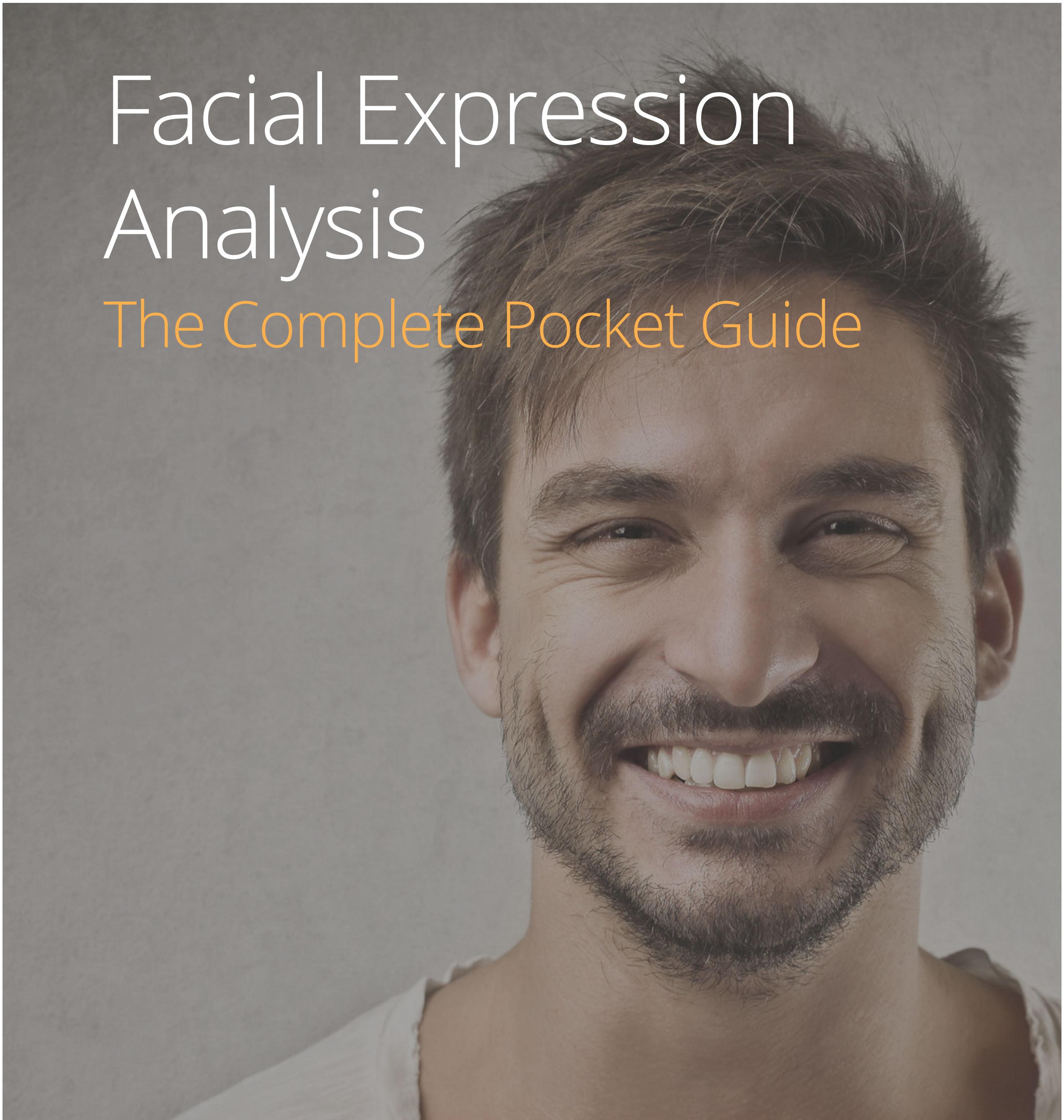
Facial Expression Analysis

The Complete Pocket Guide



iMotions – Biometric Research, Simplified

The definitive guide

Emotions are the essence of what makes us human. They impact our daily routines, our social interactions, our attention, perception, and memory.

One of the strongest indicators for emotions is our face. As we laugh or cry we're putting our emotions on display, allowing others to glimpse into our minds as they „read“ our face based on changes in key face features such as eyes, brows, lids, nostrils, and lips.

Computer-based facial expression analysis mimics our human coding skills quite impressively as it captures raw, unfiltered emotional responses towards any type of emotionally engaging content. But how exactly does it work?

We have the answers. This definitive guide is all you need to get the knack of facial coding and research into the quality of emotional behavior. Now is the right time to get started.

CONTENT

IMOTIONS®
BIOMETRIC RESEARCH PLATFORM

The basics ... and beyond

What are facial expressions?

Our face is an intricate, highly differentiated part of our body - in fact, it is one of the most complex signal systems available to us. It includes over 40 structurally and functionally autonomous muscles, each of which can be triggered independently of each other.

The facial muscular system is the only place in our body where muscles are either attached to a bone *and* facial tissue (other muscles in the human body connect to two bones), or to facial tissue only such as the muscle surrounding the eyes or lips.

Obviously, facial muscle activity is highly **specialized for expression** - it allows us to share social information with others and communicate both verbally and nonverbally.

...did you know?

Almost all facial muscles are triggered by one single nerve - the **facial nerve**. There is one exception, though:
The upper eyelid is innervated by the **oculomotor nerve**, which is responsible for a great part of eye movements, pupil contractions, and raising the eyelid.

The facial nerve controls the majority of facial muscles

All muscles in our body are innervated by nerves, which route all the way into the spinal cord and brain. The nerve connection is bidirectional, which means that the nerve is triggering muscle contractions based on brain signals (brain-to-muscle), while it at the same time communicates information on the current muscle state back to the brain (muscle-to-brain).

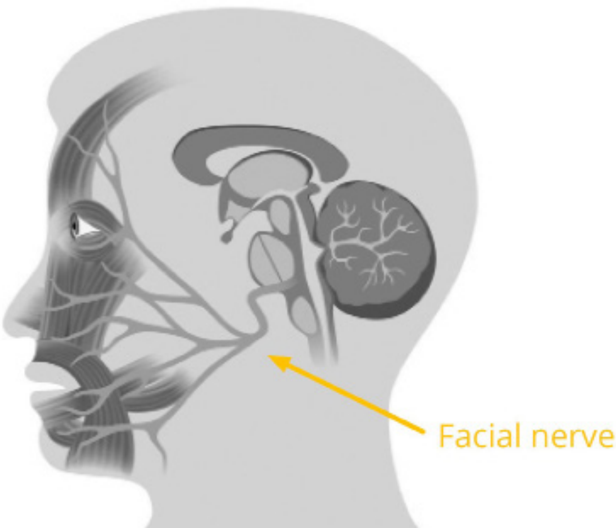
What are facial expressions?

Facial expressions are movements of the numerous muscles supplied by the facial nerve that are attached to and move the facial skin.

Almost all facial muscles are innervated by a single nerve, therefore simply referred to as **facial nerve**.

In slightly more medical terms, the facial nerve is also known under "VII. cranial nerve".

The facial nerve emerges from deep within the **brainstem**, leaves the skull slightly below the ear, and branches off to all muscles like a tree. Interestingly, the facial nerve is also wired up with much younger motor regions in our neo-cortex (neo as these areas are present only in mammalian brains), which are primarily responsible for facial muscle movements required for talking.



...did you know?

As the name indicates, the **brainstem** is an evolutionary very ancient brain area which humans share with almost all living animals.

Voluntary and involuntary facial expressions

Brainstem and motor cortex are specifically active dependent on whether a facial expression is **involuntary** or **voluntary**.

While the brainstem controls involuntary and unconscious expressions that occur spontaneously, the motor cortex is involved in consciously controlled and intentional facial expressions.

Exactly this is the reason why a fake smile just doesn't appear natural, whereas a genuine smile most likely does. You might even realize another effect: Faking a smile doesn't even *feel* right! Obviously, it doesn't trigger the same emotional reactions in your body compared to an authentic smile.

To understand this effect, take a closer look at the interaction between facial expressions and emotions in the next chapter.

Facial expressions & emotions

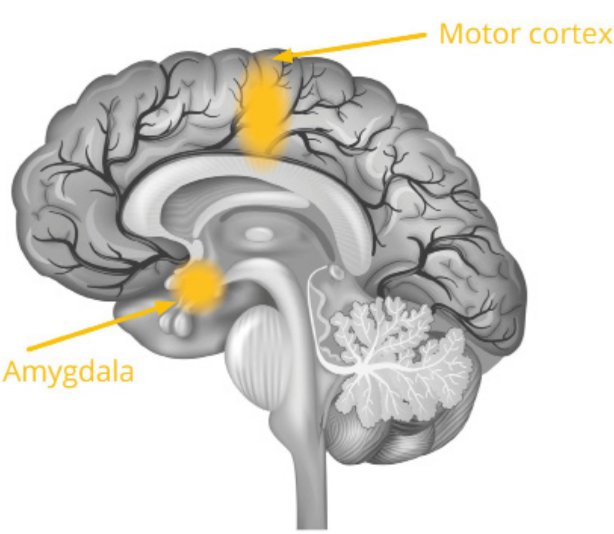
So far we know that the facial nerve connects the majority of muscles in the face with the brain. Now how does this relate to emotions and emotional behavior?

Actually, the very same regions in the brainstem that trigger our facial expressions control emotional processing and regulation.

Functional imaging studies (fMRI) have identified one specific region in the brainstem to be highly active when confronted with potential visual or auditory threats (a looming shadow or a high-pitched scream, for example):
The right and left **amygdala**.

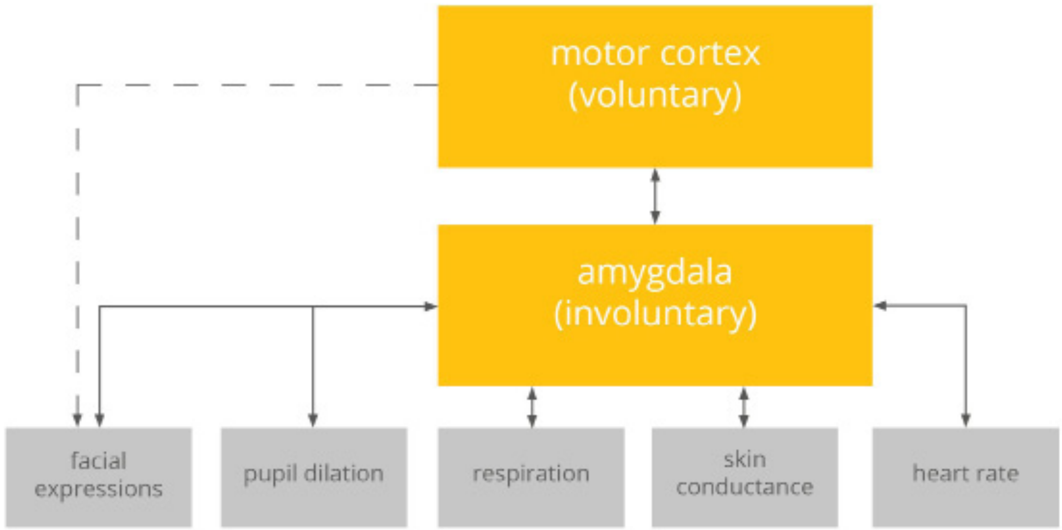
The amygdala controls emotional arousal

Often, the **amygdala** is associated with processing of live-threatening, fearful events or stimuli of high sexual appeal and bodily pleasure. Besides fear and pleasure processing, the amygdala has been found to be generally responsible for autonomic functions associated with emotional arousal.



...did you know?

The **amygdala** modulates the emission of cortisol and other stress hormones into the bloodstream, controls heart rate, skin conductance and respiration as well as observable behaviors such as changes in posture and – via the facial nerve – facial expressions.



The facial feedback hypothesis

As ingeniously discovered by *Fritz Strack and colleagues* in 1988, facial expressions and emotions are closely intertwined. In their study, respondents were asked to hold a pen in their mouths while rating cartoons for their humor content. While one group held the pen between their teeth with lips open (mimicking a smile), the other group held the pen with their lips only (preventing a proper smile).

Here‘s what Fritz Strack found out: The first group rated the cartoon as more humorous. Strack and team took this as evidence for the **facial feedback hypothesis** postulating that selective activation or inhibition of facial muscles has a strong impact on the emotional response to stimuli.

Emotions, feeling, moods

What exactly are emotions?

In everyday language, emotions are any relatively brief conscious experiences characterized by intense mental activity and a high degree of pleasure or displeasure. In scientific research, a consistent definition has not been found yet. There’s certainly conceptual overlaps between the psychological and neuroscientific underpinnings of emotions, moods, and feelings.

Emotions are closely linked to physiological and psychological arousal with various levels of arousal relating to specific emotions. In neurobiological terms, emotions could be defined as **complex action programs** triggered by the presence of certain external or internal stimuli.

These action programs contain the following elements:

1. **Bodily symptoms** such as increased heart rate or skin conductance. Mostly, these symptoms are unconscious and involuntary.
2. **Action tendencies**, for example "fight-or-flight" actions to either immediately evade from a dangerous situation or to prepare a physical attack of the opponent.
3. **Facial expressions**, for example baring one's teeth and frowning.
4. **Cognitive evaluations** of events, stimuli or objects.

Emotions vs. feelings vs. moods

Feelings are subjective perceptions of the emotional action programs. Feelings are driven by conscious thoughts and reflections - we surely can have emotions without having feelings, however we simply cannot have feelings without having emotions.

Moods are diffuse internal, subjective states, generally being less intense than emotions and lasting significantly longer. Moods are highly affected by personality traits and characteristics. Voluntary facial expressions (smiling, for example) can produce bodily effects similar to those triggered by an actual emotion (happiness,

Can you classify emotions?

Facial expressions are only one out of many correlates of emotion, but they might be the most apparent ones. Humans are obviously able to produce thousands of slightly varying sets of facial expressions – however, there is only a small set of distinctive facial configurations that almost every one associates with certain emotions, irrespective of gender, age, cultural background and socialization history.

These categorical emotions are:

Joy

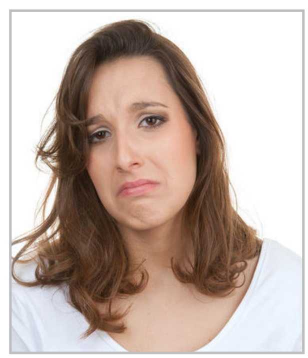
Anger

Surprise

Fear

Contempt

Sadness



Disgust

The finding that almost everyone can produce and recognize the associated facial expressions of these emotions has led researchers to the assumption that they are universal.

There is an ongoing discussion in emotion research on how the different emotions could be distinguished from each other:

1

Discrete emotion theory

assumes that the seven basic emotions are mutually exclusive, each with different action programs, facial expressions, physiological processes, and accompanying cognitions.

2

Dimensional models

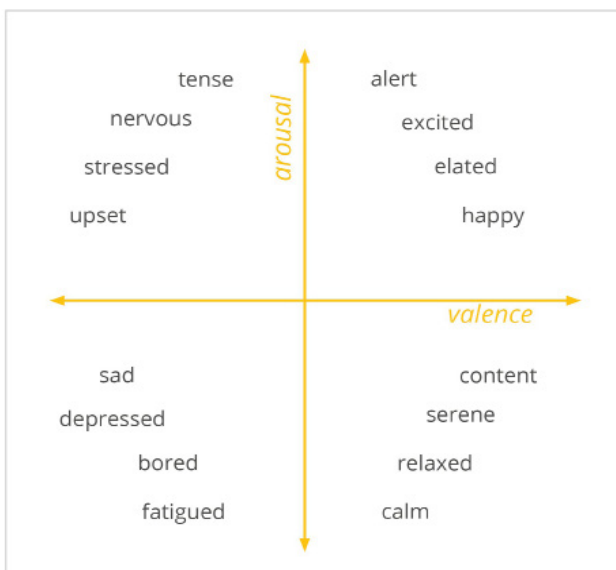
assume that emotions can be grouped and arranged along two or more dimensions. Most dimensional models use **valence** (positive vs. negative emotions) as horizontal axis and **arousal** (activating vs. calming emotions) as vertical axis.

With valence and arousal, more subtle emotional classifications are possible - breaking “happiness” into a less aroused “happy” state and a more aroused “elated” state, for example. Again, facial expressions are core indicators of underlying emotional states.

...did you know?

Testing *Darwin’s* “universality hypothesis” from 1872, *Ekman & Friesen* conducted a cross-cultural study in 1987 where respondents watched pictures of human faces and were asked to classify which emotion was present and rate the intensity of the emotion.

All respondents were able to classify which emotion was active (while minor cultural variations existed in ratings of emotional intensity).



Application fields

With facial expression analysis you can test the impact of **any content, product or service** that is supposed to elicit emotional arousal and facial responses - physical objects such as food probes or packages, videos and images, sounds, odors, tactile stimuli, etc. Particularly involuntary expressions as well as a subtle widening of the eyelids are of key interest as they are considered to reflect changes in the emotional state triggered by actual external stimuli or mental images.

Now which fields of commercial and academic research have been adopting facial expression analysis techniques lately? Here is a peek at the most prominent research areas:

...did you know?

Humans are processing visual information to a much larger extent than other senses - we might get just as thrilled by watching a video of a rollercoaster ride as we get by actually riding it.

Needless to say that immersive digital media content, video games, and software interfaces constitute a new generation of visual stimuli that elicit facial responses.

1

Consumer neuroscience and neuromarketing

There is no doubt about it: Evaluating consumer preferences and delivering persuasive communication are critical elements in marketing. While self-reports and questionnaires might be ideal tools to get insights into respondents' attitudes and awareness, they might be limited in capturing emotional responses unbiased by self-awareness and social desirability. That's where the value of facial analysis comes in: Tracking facial expressions can be leveraged to substantially enrich self-reports with quantified measures of more unconscious emotional responses towards a product or service. Based on facial expression analysis, products can be optimized, market segments can be assessed, and target audiences and personas can be identified. There's quite a lot that facial expression analysis can do for you to enhance your marketing strategy - just think about it!

2

Media testing & advertisement

In media research, individual respondents or focus groups can be exposed to TV advertisements, trailers, and full-length pilots while monitoring their facial expressions. Identifying scenes where emotional responses (particularly smiles) were expected but the audience just didn't "get it" is as crucial as to find the key frames that result in the most extreme facial expressions.

In this context, you might want to isolate and improve scenes that trigger unwanted negative expressions indicating elevated levels of disgust, frustration or confusion (those kind of emotions wouldn‘t exactly help a comedy show to become a hit series, would they?) or utilize your audience‘s response towards a screening in order to increase the overall level of positive expressions in the final release.

3 Psychological research

Psychologists analyze facial expressions to identify how humans respond emotionally towards external and internal stimuli. In systematic studies, researchers can specifically vary stimulus properties (color, shape, duration of presentation) and social expectancies in order to evaluate how personality characteristics and individual learning histories impact facial expressions.

...did you know?

The more similar your interaction partners appear to you (and the more you can relate to them), the more likely you are to take over their facial expressions.

4 Clinical psychology and psychotherapy

Clinical populations such as patients suffering from Autism Spectrum Disorder (ASD), depression or borderline personality disorder (BPD) are characterized by strong impairments in modulating, processing, and interpreting their own and others’ facial expressions. Monitoring facial expressions while patients are exposed to emotionally arousing stimuli or social cues (faces of others, for example) can significantly boost the success of the underlying cognitive-behavioral therapy, both during the diagnostic as well as the intervention phase. An excellent example is the “Smile Maze” as developed by the Temporal Dynamics of Learning Center (TDLC) at UC San Diego. Here, autistic children train their facial expressions by playing a Pacman-like game where smiling steers the game character.

5 Medical applications & plastic surgery

The effects of facial nerve paralysis can be devastating. Causes include Bell’s Palsy, tumors, trauma, diseases, and infections. Patients generally struggle with significant changes in their physical appearance, the ability to communicate, and to express emotions. Facial expression analysis can be used to quantify the deterioration and evaluate the success of surgical interventions, occupational and physical therapy targeted towards reactivating the paralyzed muscle groups.

6 Software UI & website design

Ideally, handling software and navigating websites should be a pleasant experience - frustration and confusion levels should certainly be kept as low as possible. Monitoring facial expressions while testers browse websites or software dialogs can provide insights into the emotional satisfaction of the desired target group. Whenever users encounter road blocks or get lost in complex sub-menus, you might certainly see increased “negative” facial expressions such as brow furrowing or frowning.

7 Engineering of artificial social agents (avatars)

Until recently, robots and avatars were programmed to respond to user commands based on keyboard and mouse input. Latest breakthroughs in hardware technology, computer vision, and machine learning have laid the foundation for artificial social agents, who are able to reliably detect and flexibly respond to emotional states of the human communication partner. Apple’s Siri might be the first generation of emotionally truly intelligent machines, however computer scientists, physicians, and neuroscientists all over the world are working hard on even smarter sensors and algorithms to understand the current emotional state of the human user, and to respond appropriately.

...did you know?

As a matter of fact, facial expressions are also triggered by mentalizations, memories, and thoughts.Telling your best friend of your recent vacation in Hawaii might put a bright smile on your face, whereas thinking of next Monday morning (and the alarm going off at 6am) most likely will make you frown.

The variety of application fields for facial expression analysis is incredibly wide, which might be due to the fact that it is surprisingly simple to collect, process, and analyze high-quality facial muscle activity from optimally placed sensors in order to decode involuntary emotional reactions and uncover the subconscious processes driving emotional behavior.

But how exactly can facial expressions be collected?

We’ll get to it now.

Facial expression analysis techniques

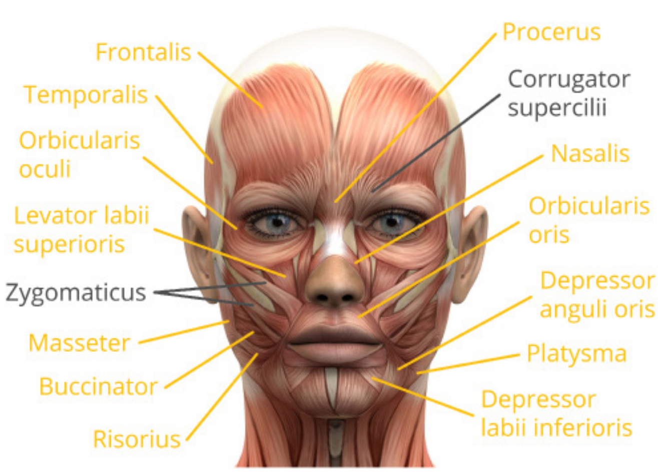
Facial expressions can be collected and analyzed in three different ways:

1. By tracking of facial electromyographic activity (fEMG)
2. By live observation and manual coding of facial activity
3. By automatic facial expression analysis using computer-vision algorithms

Let's explain them in more detail.

Facial electromyography (fEMG)

With facial EMG you can track the activity of facial muscles with electrodes attached to the skin surface. fEMG detects and amplifies the tiny electrical impulses generated by the respective muscle fibers during contraction. The most common fEMG sites are in proximity to the following two major muscle groups:



1 **Right/left corrugator supercilii ("eyebrow wrinkler")**

This is a small, narrow, pyramidal muscle near the eye brow, generally associated with frowning. The corrugator draws the eyebrow downward and towards the face center, producing a vertical wrinkling of the forehead. This muscle group is active to prevent high sun glare or when expressing negative emotions such as suffering.

2 **Right/left zygomaticus (major)**

This muscle extends from each cheekbone to the corners of the mouth and draws the angle of the mouth up and out, typically associated with smiling.

Benefits of fEMG:

1 fEMG is a non-invasive, precise, sensitive method to continuously measure facial muscle activity.

2 fEMG does not depend upon language and does not require cognitive effort or memory.

3 fEMG is able to measure even very subtle facial muscle activity even in scenarios where respondents are instructed to inhibit their emotional expression.

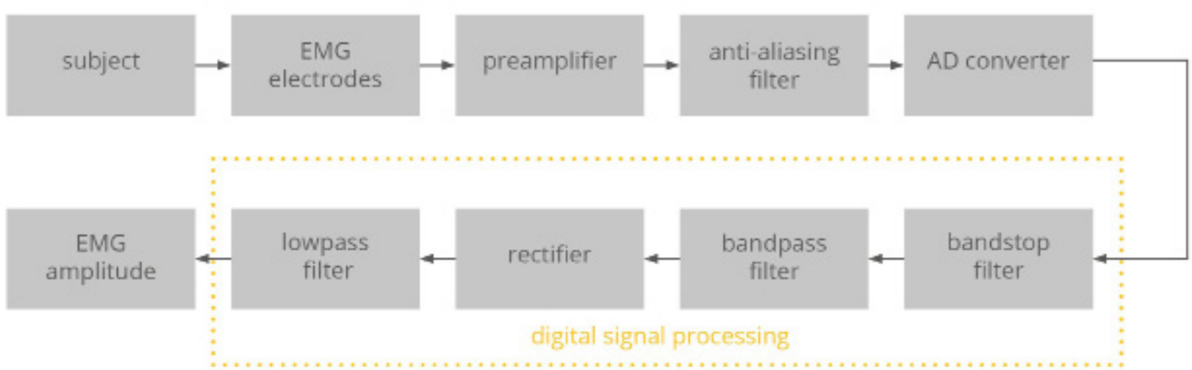
Limitations of fEMG:

1 fEMG requires electrodes, cables, and amplifiers. It therefore is intrusive, heightening the respondents' awareness of what is being measured.

2 fEMG is sensitive to motion artifacts and electrical interference.

3 fEMG analysis requires expert biosensor processing skills.

Here's how a biosensor processing pipeline for fEMG looks like:



The Facial Action Coding System (FACS)

Based on the work of Swedish anatomist *Carl-Herman Hjortsjö*, *Paul Ekman* and *Wallace V. Friesen* developed one of the most influential methods to objectively code facial behavior in 1978, which was further fine-tuned in 2002.

Their approach named the **Facial Action Coding System (FACS)** represents a fully standardized classification system of facial expressions for expert human coders based on anatomic features. Experts carefully examine face videos and describe any occurrence of facial expressions as combinations of elementary components called **Action Units** (AUs).

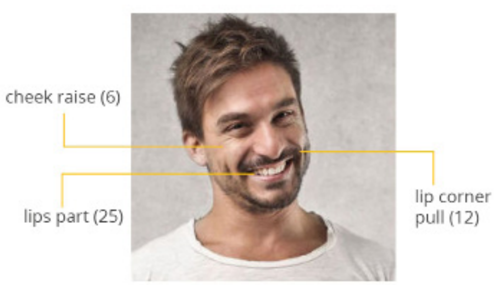
A common misbelief is that FACS is related to reading or detecting emotions. In fact, FACS is just a measurement system and does not interpret the meaning of the expressions. It's like saying the purpose of riding a bike is to go to work. Sure, you can ride a bike to work, but you can use your bike for a lot of other things as well (for leisure, sports training, etc.). You get the idea.

However, during the analytical phase, the FACS system allows for a modular construction of emotions based on the combination of AUs.

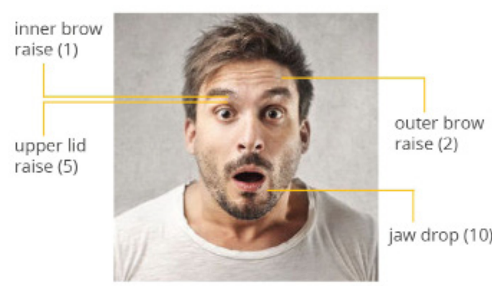
...did you know?

Each AU corresponds to an individual face muscle or muscle group. It is identified by a number (AU1, AU2, etc.). All facial expressions can be broken down into their constituent AUs. Assumed that facial expressions are "words", AUs are the "letters" that make up those words.

Happy



Surprised



With facial action unit coding you can get all the knowledge to dissociate between the following three facial expression categories:

1. **Macroexpressions** typically last between 0.5 - 4 seconds, occur in daily interactions, and generally are obvious to the naked eye.

2 **Microexpressions** last less than half a second, occur when trying to consciously or unconsciously conceal or repress the current emotional state. Microexpressions are much harder to be detected; however, with a reasonable amount of training you can become an expert coder for microexpressions.

3 **Subtle expressions** are associated with the intensity and depth of the underlying emotion. We're fur sure not perpetually smiling or raising eyebrows – the intensity of these facial actions constantly varies. Subtle expressions denote any onset of a facial expression where the intensity of the associated emotion is still considered low.

Benefits of FACS:

1 Facial action coding is a non-intrusive, objective, reliable method to describe facial expressions. Emotional interpretations emerge only during the data processing stage.

2 FACS scores have a high face validity as they are based on visible changes in facial tissue.

3 FACS scores also contain a 5-step intensity rating (A, "trace", to E, "maximum").

Limitations of FACS:

1 FACS coding requires high quality video equipment.

2 FACS scoring relies on the trained discrimination of experts, rendering the coding very laborious and expensive. A well-trained FACS coder can easily take about 100 minutes to code 1 minute of video data depending on the density and complexity of facial actions.

3 FACS training itself is resource-intense, requiring about 100 hours of studying the FACS manual, and about 12 hours to take the FACS certification.

Out of these limitations, a new generation of facial expression technologies has emerged being fully automated and computer-based.

Sounds exciting? It sure is. Let's dig deeper into the specific requirements, data outputs, and analysis techniques of these as well as their pros and cons in the following chapters.

IMOTIONS®

BIOMETRIC RESEARCH PLATFORM

Getting started with facial expression analysis

Technology

Built on groundbreaking research of core academic institutions in the United States and Europe, automatic facial coding procedures have been developed and made available to the broader public, instantaneously detecting faces, coding facial expressions, and recognizing emotional states.

This breakthrough has been mainly enabled by the adoption of state-of-the-art computer vision and machine learning algorithms along with the gathering of high-quality databases of facial expressions all across the globe.

These technologies use cameras embedded in laptops, tablets, and mobile phones or standalone webcams mounted to computer screens to capture videos of respondents as they are exposed to content of various categories.
The use of inexpensive webcams eliminates the requirement for specialized high-class devices, making automatic expression coding ideally suited to capture face videos in a wide variety of naturalistic environmental settings such as respondents' homes, workplace, car, public transportation, and many more.

...did you know?

Automated facial coding has revolutionized the field of affective neuroscience and biosensor engineering, making emotion analytics and insights commercially available to the scientific community, commercial applications, and the public domain.

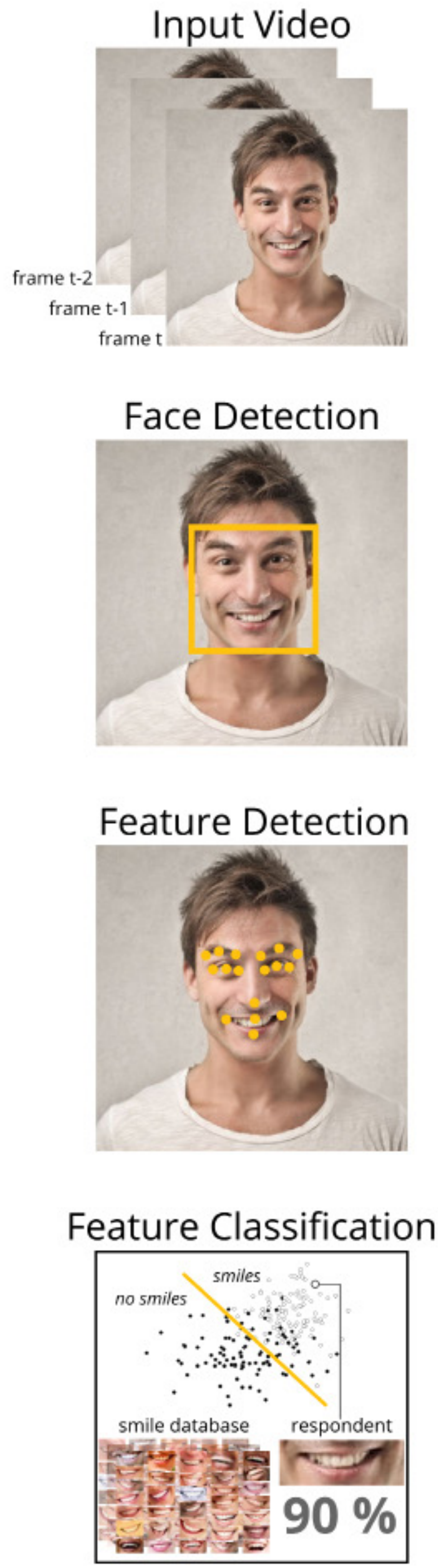
What exactly are the scientific and technological procedures under the hood of this magic black box?

Let's find out now.

The technology behind automatic facial coding

More or less, all facial expression analysis engines comprise the same steps.
Emotient FACET, for example, applies the following:

1. Face detection
2. Facial landmark detection and registration
3. Facial expression and emotion classification



1 Face detection

The position of a face is found in a video frame or image, which can be achieved by applying the **Viola Jones Cascaded Classifier** algorithm, for example. This might sound sophisticated, but you can actually find this technology in the camera of your iPhone or Android smartphone as well. The result is a face box framing the detected face.

2 Feature detection

Within the detected face, facial landmarks such as eyes and eye corners, brows, mouth corners, the nose tip etc. are detected. After this, an internal face model is adjusted in position, size, and scale in order to match the respondent's actual face. You can imagine this like an invisible virtual mesh that is put onto the face of the respondent: Whenever the respondent's face moves or changes expressions, the face model adapts and follows instantaneously. As the name indicates, the face model is a simplified version of the respondent's actual face. It has much less details, (so-called **features**) compared to the actual face, however it contains exactly those face features to get the job done. Exemplary features are single landmark points (eyebrow corners, mouth corners, nose tip) as well as feature groups (the entire mouth, the entire arch of the eyebrows etc.), reflecting the entire "Gestalt" of an emotionally indicative face area.

3 Feature classification

Once the simplified face model is available, position and orientation information of all the key features is fed as input into classification algorithms which translate the features into Action Unit codes, emotional states, and other affective metrics.

Facial expression engines differ by the available metrics that are automatically coded:

Emotient FACET

- Head orientation (yaw, pitch, roll)
- 6 facial landmarks (ADD eye and nose position)
- 7 basic emotions, valence; complex states: frustration, confusion
- 20 Action Units
- respondent's sex and whether or not respondent wears glasses

Affectiva AFFDEX

- Head orientation (yaw, pitch, roll)
- Interocular distance and 34 facial landmarks
- 7 basic emotions; valence, engagement, attention
- 14 facial expression metrics (similar to Action Units)

Noldus FaceReader

- Head orientation (yaw, pitch, roll)
- 6 basic emotions; contempt (experimental), arousal, valence
- 20 Action Units

Here's a little theory (and analysis) behind it.

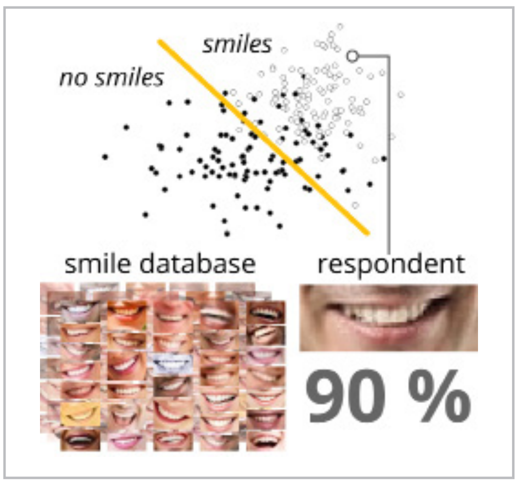
The translation from face features into metrics is accomplished **statistically**, comparing the actual appearance of the face and the configuration of the features numerically with the normative databases provided by the facial expression engines.

What does that mean? The current facial expression of a respondent is not compared one-by-one with all hundreds of thousands of pictures in the database – this would be quite tedious and take forever. Instead, the databases contain statistics and normative distributions of all feature characteristics across respondents from multiple geographic regions and demographic profiles as well as stimulus categories and recording conditions. Every facial expression analysis engine uses a different database. This is one of the reasons why you might get slightly varying results when feeding the very same source material into different engines.

The classification is done on a purely **statistical level**. Here's an example: If the respondent's mouth corners are pulled upward, a human coder would code this as activity of AU12 ("lip corner puller") – stating that the respondent is smiling. The facial expression engine instead has to compute the vertical difference between mouth corners and mouth center, returning a value of 10 mm. This value is compared to all possible values in the database (values between 0 mm and 20 mm, for example).

In our example, the respondent's expression apparently is right in the middle of the "smile" distribution (0 < 10 < 20 mm). For the engine it's definitely safe to say that the current expression is stronger than a slight smile, however not as strong as a big smile.

In an ideal world, the classifier returns either "Yes" or "No" for each emotion, AU or metric. Unfortunately, it is not that easy. Take a look again at the "smile" classifier:



The filled dots (rather to the left) are all mouth configurations in the database that are not classified as smiles, while the white dots (rather towards the right) represent mouth configurations that are regarded as smiles.

The orange line denotes the classification criterion.

Everything on the left side of the line is classified as "not a smile", whereas everything on the right side is classified as "smile".

There obviously is a diffuse transition zone between non-smiles and smiles where some mouth configurations are misclassified – they might be subtle smiles that stay under the radar or other mouth opening configurations that might be rather yawns than smiles. As a consequence, the classifier can only return a **probabilistic result**, reflecting the likelihood or chance that the expression is an authentic "smile".

In our example, the respondent's smile is rather obvious, ending up all to the right side of the "feature space". Here, the classifier is very certain about the result of a smile and returns a solid 90%. In other cases, the classifier might be less confident about its performance.

Feature classification is done for each emotion, Action Unit, and key feature independently - the classifier for smiles doesn't know anything about the classifier for frowns, they are simply coded independent of each other. The good thing about this is that the automatic coding is accomplished much more objectively than manual coding where humans – particularly novice coders – tend to interpret the activation of an Action Unit in concert with other Action Units, which significantly alters the results.

Before we dig deeper into the visualization and analysis of the facial classification results, we will have a peak into some guidelines and recommendations on how to collect best-in-class data.

Here we go.

Equipment

Let's get practical.

What should you consider when acquiring high-quality data, what are optimal stimuli, what to keep in mind when recruiting and instructing respondents? We're here to help.

1 For starters, automatic facial expression analysis can be accomplished **online**, that is, while respondents are exposed to emotional stimuli and their facial expressions are monitored (and/or recorded).

2 Apart from online data collection, facial expression analysis can be applied **offline**, for example to process facial captures that have been recorded before. You can even record a larger audience and, with minimal editing, analyze each respondents' facial expressions.

For online automatic facial coding with webcams, keep the following camera specifications in mind:

1 **Lens:** Use a stadard lens; avoid wide-angle or fisheye lenses as they may distort the image and cause errors in processing.

2 **Resolution:** The minimum resolution varies across facial expression analysis engines. You certainly don't need HD or 4K camera resolutions - automated facial expression analysis already works for low resolutions as long as the respondent's face is clearly visible.

3 **Framerate:** The camera should have a stable framerate of 10 fps or higher. However, you can certainly process video feeds with 60 fps or more.

4 **Autofocus:** You certainly want to use a camera that is able to track respondents' faces within a certain range of distance. Some camera models offer additional, software-based tracking procedures to make sure your respondents don't get out of focus.

3 **Aperture, brightness and white balance:** Generally, most cameras are able to automatically adjust to lighting conditions. However, in case of backlighting the gain control in many cameras reduces image contrast in the foreground, which may deteriorate performance.

A word on webcams: We suggest to use the *Logitech HD Pro webcam C920*. This camera is most suitable as it offers stable framerates, auto-focus, and allows access to aperture, brightness, and white balance settings.

You can of course accomplish automatic facial expression analysis with any other camera including smartphone cameras, IP cameras or internal laptop webcams (dependent on your application). However, be aware that these cameras generally come with limited auto-focus, brightness, and white balance compensation capabilities.

...did you know?

Some cameras might dynamically adjust the frame number dependent on lighting conditions (they reduce the frame rate to compensate for poor lighting, for example), leaving you with a video file with a variable frame rate. This certainly is not ideal as you will end up with variable frames per second, making it harder to compare and aggregate the results across respondents.

Offline automatic facial coding can be applied to pretty much any video material. Here's what you should keep in mind:

1 **Resolution:** The resolution of the video should be at least 640 x 480 pixels. Again, the determining factor is the size of the face in the recording.

2 **Frame rate:** A frame rate of 10 fps or higher is recommended.

3 **Video codec:** The rule of thumb? If the file can be played with Windows Media Player, chances are you can process it with automatic facial expression analysis. MP4 and WMV codecs are certainly supported. If your video is rendered in a non-supported codec, you can use freely available converter tools such as *Handbrake* or *AnyVideo Converter Free* to change the video codec. If you are not sure about the video codec, you can use *GSpot* (www.headbands.com). However, these video libraries can differ in how they read the codecs, resulting in slightly varying emotion data. If you process the same video decoded with two different libraries, this may result in two slightly different sets of emotion values.

Respondent instruction

Setting up the environment for automatic facial tracking and expression analysis under controlled laboratory settings really is no rocket science. In actual fact, there are only a few things to consider before you can get going with your study.

1 Camera placement and respondent positioning

- Place the camera at the respondent's eye level (approximately) such that it faces the respondent directly. Face angles of ±20 degrees are acceptable, however automatic facial expression analysis delivers best results if the face is roughly center and frontal in the frame.
- For screen-based stimuli (videos, pictures, websites, games, software interfaces), attach the camera either above or below the screen. If respondents are supposed to use a keyboard, chances are they look down frequently. In this case, the face detector will report missing data. To prevent this, place the camera below the monitor.
- Make sure that faces fulfil the required minimum size. Facial analysis engines recommend minimal distances from ear-to-ear for optimal encoding of emotions, attributes, demographics, pose or Action Unit coding.
- Respondents should be seated comfortably to guarantee they will stay within the frame and in focus for the entire recording. Take into account that respondents change position once in a while.

2 Illumination

- Face tracking and facial expression analysis operate best with indoor, uniform, diffuse illumination. Accomplish a setup with standard contrast.
- Use indirect light (ambient lighting) as it illuminates the face evenly. Professional setups often use a softbox diffuser to accomodate for optimal lighting conditions.
- Avoid dim lighting. Facial expression analysis doesn't work in the dark.
- Avoid strong lights in the background (from a window or an artificial light source behind the respondent's head for example). Also, strong directional lighting shining on one side of the head (direct sunlight with strong shadows, for example) causes high contrast, potentially leading to issues in the facial expression analysis.

3

Visibility of the face

- Facial expression analysis requires the visibility of emotionally sensitive facial landmarks such as eyebrows, eyes, nose, and mouth. If any of these are occluded, the face tracking and expression analysis may lead to only partial results.
- Tell respondents to not wear overly large, horn-rimmed glasses covering the eyebrows. Also, sunglasses will interfere with proper face capture.
- Long, unruly beards that occlude the mouth are not optimal (well-groomed three-day stubble should be fine, though).
- Facial jewelry, multiple mouth and eyebrow piercings can have a negative impact.
- Avoid hats, caps, beanies, etc. as they are likely to cast shadows and occlude facial landmarks. Equivalently, hair styling that partially covers the face (side-swept bangs, for example) should be excluded.
- Instruct respondents to not occlude their faces with or rest their heads on their hands.

What about talking, eating, drinking?

Talking, eating, and drinking all involve movement of the lower facial muscles, particularly in the area around the mouth and cheeks. While these activities will most likely not have any impact on face tracking, they certainly cause changes in facial expressions. As respondents talk, automatic expression coding procedures might incorrectly classify the muscular activations as indicator for the presence of certain emotions.

At the moment, this misclassification cannot be avoided, and unfortunately there is no procedure which could correct or exclude these time intervals in a video. However, you can mark those data portions where respondents talk, eat or drink in order to exclude them from the data analysis.

The gist of it? Advise your respondents to drink and eat sufficiently before the experiment. Chewing gum is an absolute no-go for facial expression analysis.

Stimulus setup

In order to evaluate the individual facial expression characteristics of your respondents, some facial expression analysis engines require the collection of a "neutral" face expression during a baseline period, which ideally is around 5 to 10 seconds long and placed at the beginning of the actual data collection.

Running a baseline

In case baselining is required, you can run the following procedures:

1

Neutral baseline

In this condition, no stimuli are presented. Respondents are just sitting in a comfortable, relaxed position and look "neutrally" towards the camera. The recorded data reflects the respondent's individual baseline expression in presence of neutral stimuli.

2

Baseline with variable stimuli

This condition contains stimuli with varying emotional content (a video with scenes that elicit neutral, positive, and negative emotions, for example). The variable baseline is considered to max out a respondent's facial expressions, comprising the full spectrum of neutral expressions and expressions with high positive and negative valence. Our suggestion: Show a static, gray slide interrupted by scenes that trigger strong facial expressions.

Setting up stimuli

Once the baseline has been collected, you are certainly ready to start a recording in the desired test environment with the right respondents containing the stimulus set of interest.

...did you know?

Keep recording conditions clean and structured: Present functional stimuli long enough and use intermediate stimuli so that the facial expressions for all respondents can cool off and start from baseline levels for the next stimulus.

How should stimuli be presented? Facial expressions are very responsive and occur within tens to hundreds of milliseconds after stimulus onset. As a consequence, present any material long enough for respondents to process its content. This holds true for all sensory modalities: vision, hearing, taste, smell, and touch. Further, it might be useful to place cool-off stimuli of appropriate duration between the stimuli of interest in order to allow facial expressions to return to the neutral baseline state.

Pictures, pictures, pictures

The International Affective Picture System (IAPS) is a database that has been specifically designed for emotion and attention research, comprising about 1,000 standardized color photographs that have been rated based on their emotional content. The IAPS is being widely used in several fields of academic and commercial research. Access to the database is restricted and requires signing an official request:

http://csea.phhp.ufl.edu/media/requestform.html

Quick recap

What do we know so far? We talked about the technology of facial coding, recommended equipment, respondent instruction, and stimulus setup. That‘s already quite some information to work with.

You snooze, you lose. Let‘s move one step further and sneak a peek beneath the surface of theory to get an even bigger grasp of automatic facial expression analysis: What can facial epression data *tell*?

In case you were wondering all along, be excited - we‘re about to find out.

Data output & visualization

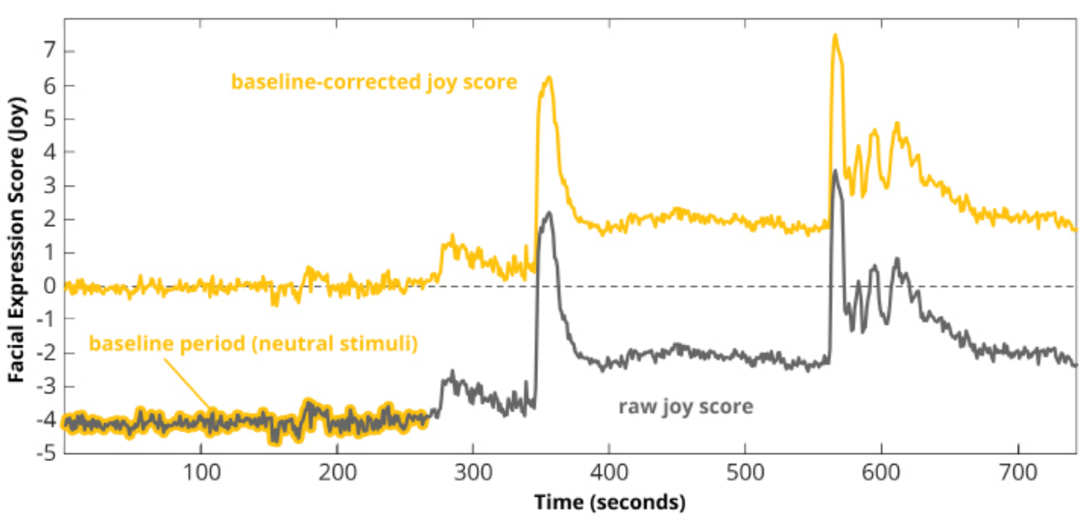
Automatic facial expression analysis returns **numeric scores** for facial expressions, Action Units, and emotions along with the degree of confidence. Most generally, you can think of these metrics as detectors: As the facial expression or emotion occurs and/or intensifies, the confidence score rises from 0 (no expression) to 100 (expression fully present).

The numerical values representing the facial expression can be plotted and visualized for both static images and videos (frame by frame time-course of emotion scores):

Take a look at the image. You might wonder why the facial expression is not classified as a single emotion - instead, it returns the three emotions joy (99%), surprise (10%), and disgust (2%). The other emotions returned values of 0%. How come?

Obviously, the current facial expression is a mixture of the three basic emotions. According to the classification, the expression most likely represents a highly joyful face with a slight indication for surprise.

Now that we covered static images, let's have a closer look into the time course of emotional data. For this purpose, we analyzed the facial expressions of respondents watching a 4 minute trailer with scenes of varying emotional arousal.



Plotting raw scores vs. baseline-corrected scores

1 Raw score

First, you can analyze the raw expression scores, which represents the classification results of the facial expression engine for a certain respondent compared to the facial expressions stored in the global database. The benefit of this procedure is that the respondent's data can be compared directly with the expressions of thousands of other participants, giving you straightforward insights into emotional reactions. This procedure is recommended whenever you would like to aggregate data of several respondents, compare the data of two or more groups of respondents, or collate two or more stimuli (different versions of an advertisement, for example).

2 Baseline-corrected scores (if applicable)

Respondents differ in their "neutral" expression (baseline), and in some scenarios it makes sense to first baseline-correct the raw values in order to generate facial coding scores that reflect the relative expressions of a respondent independent of the database. Here's an example: Let's assume a respondent has a rather joyful baseline expression. If this individual baseline is not considered and raw scores are used, the amount of smiling and other joyful expressions within a certain time period will be misestimated.

To avoid this, do the following within each emotion:

1. Compute the median across all samples during the baseline period.
2. Subtract the median from all other samples.

...did you know?

Baseline-correcting values is recommended whenever you are interested in facial expressions of individual respondents, for example when evaluating training progress in a facial muscle training, which is supposed to be adaptive to the patient's individual base expression levels.

By baseline-correcting the respondent's data, the signal is shifted up- or downwards, allowing you to see those periods in time where the respondent shows changes in emotional expressions relative to their own baseline/neutral expression.

A word on thresholding

Another important aspect for the analysis of facial expression data is whether to focus on actual scores or on thresholded data.

You can apply thresholds both in time and amplitude:

1

Thresholding in time
This thresholding technique defines a minimum duration in which a certain facial expression has to be active in order to be considered for further analysis. Temporal thresholding masks off short occurrences of facial expressions, no matter how strong they are.

If you're interested in longer-lasting emotional expressions, however not so much in quick facial expression "bursts", thresholding in time (using a time interval of 0.5 seconds, for example) might be a good solution for you. This means that any occurrence that is shorter than the defined window is disregarded.

...did you know?

Thresholding is a technique where you define a cut-off criterion – a threshold value – to highlight only the strongest or longest facial expressions while at the same time masking off the smaller, more subtle (and shorter) expressions.

2

Amplitude-based thresholding
Thresholding based on amplitutes specifies a minimum amplitude in which a certain facial expression has to be active in order to be considered for further analysis. Amplitude-based thresholding masks off low-amplitude occurrences such as a slight smile or a subtle furrow, no matter how long they are present.

If you're interested in stronger, rather obvious emotional expressions, however not too much into low-amplitude facial activity, we recommend amplitude-based thresholding.

Particularly when thresholding based on amplitude, you can set either an absolute or relative cut-off. What exactly is the difference between those two cut-offs?

1 **Absolute thresholds**
Let's assume your facial expression data ranges on a scale from 0 (completely absent) to 100 (strongly present). You could set an absolute (hard-coded) threshold at a value of 50. In this case, all values above the threshold "survive", whereas all other values are set to zero. Absolute thresholding certainly is a good choice if respondents show high-amplitude expressions above the threshold, and if the respondent group is very consistent in their neutral baseline expression.

2 **Relative thresholds**
To identify the strongest facial expressions of a respondent irrespective of their absolute score, you have to define a relative threshold. Relative thresholds are set dependent on the data distribution of a respondent. Here's how it could work: You sort all scores according to their value and define a relative cut-off criterion at the 80th percentile, for example. This is the value which separates the lower 80% from the upper 20% of the data. With relative thresholding you can identify those moments in time where respondents showed their strongest facial expression (the top 20%, for example). Be aware that a relative threshold typically has different numerical values across respondents since the raw expression scores of each respondent differ (while the 80th percentile for respondent A might be at a value of 55, the 80th percentile for respondent B might be at a value of 20).

Agony of choice - which threshold should you use?

Which thresholds to use or to combine varies quite a lot dependent on the application. In any case, avoid setting thresholds arbitrarily. Instead, check existing papers and studies using the same stimulus material, a comparable participant pool, and/or the same set of facial expressions or Action Units.

Searching for papers on Google Scholar (scholar.google.com) is a first start: Just enter the facial expression analysis technique or engine in combination with the desired keywords – you will be surprised how many studies on specific facial expressions already have been conducted, serving as source of inspiration (and as necessary backup for the methodical description of the parameters in your report or paper). Definitely give it a try!

Aggregation of facial data – raw or thresholded?

The aggregation of facial expression data across respondents can be accomplished based on the actual values, but also on the thresholded and binarized signal of each and every respondent. Here's how - step by step:

1. Break the continuous data of each respondent during the selected stimulus period into intervals with a certain window length (500 milliseconds, for example).

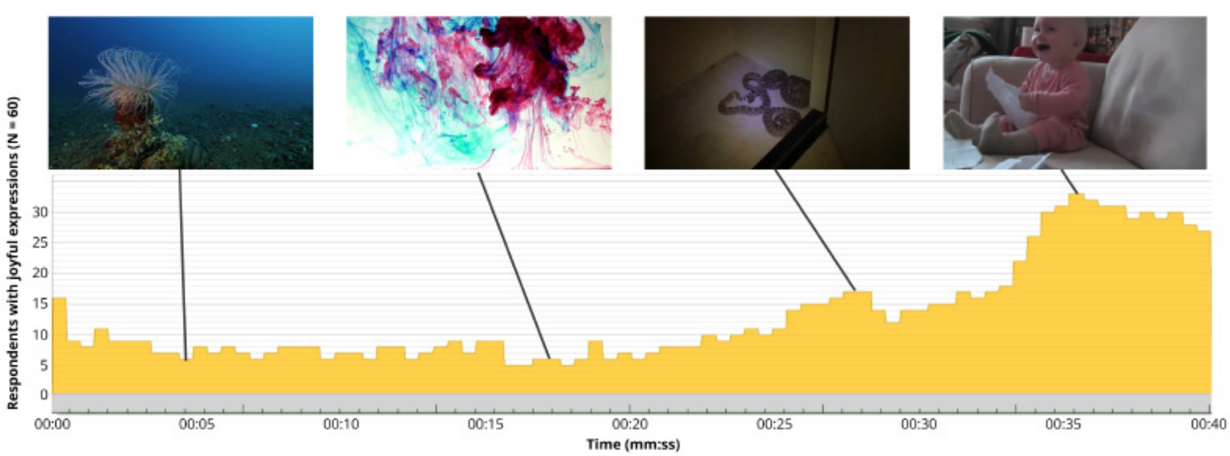
2. Compute the median of the facial expression score within each window.

3. If the median exceeds the threshold, assign a value of 1 („true"). Assign a value of 0 („false") to intervals where the data doesn't exceed the threshold. Instead of the actual facial expression scores, the binary values 0 or 1 are used.

4. In each time interval, sum up the binary scores across all respondents. Here's an example: In case you collected facial expression data from 10 respondents and all of them showed a facial expression above the threshold in a certain interval, the aggregated value for this interval is 10. If none of your respondents had a facial expression that exceeded the threshold, the aggregated value for this interval is 0.

The resulting curve is a visualization of the consistency of your respondents' facial expressions triggered by the provided stimulus material. The aggregated result relates directly to the physiological and psychological arousal of the audience.

You can now easily extract the number of occurrences of a specific facial expression, both within a single respondent as well as across all respondents. How? Just count!



Facial expression analysis ... reloaded

Adding biometric sensors

While facial recognition delivers valuable information on the quality of an emotional response, generally referred to as its **valence**, one core limitation of computer-based facial coding lies in its inability to assess someone's emotional **arousal**, that is, the intensity of an emotion.

You might ask why this is the case – doesn't a bright smile say it all? Doesn't it represent an evident indicator for a highly arousing feeling of happiness?

The truth is both yes and no. While it might very well apply for the emotional assessment of a single person, it definitely becomes more complex when comparing emotional expressions across different stimuli, respondents or groups.

...did you know?

The dynamics of facial muscle activity change across the day, the life cycle, and certainly across cultures. In addition there surely are strong individual differences in emotional expressions. Imagine watching a horror movie with your dad - while you are frightened to the core, he probably just keeps a straight face (although he might be as terrified by the film as you are).

If you aim to portray the emotional engagement of a larger audience or consumer group in response to emotionally loaded stimuli (ads, shows, pictures, videos) in its full complexity, it is essential to assess both the valence of the emotional expression and the associated arousal.

That's exactly where the tremendous value of multimodal research comes in.

Here's a list of the most widely used biomarkers for emotional arousal.

Eye tracking

Like no other experimental method, eye tracking offers incredible insights into visual attention. While eye tracking is commonly used to monitor where we direct our eye movements at a certain point in time, it also tracks the dilation of the pupil.

Changes in pupil dilation happen whenever the brightness of the stimulus material changes – when we step out of bright sunlight into a dark room, the pupils dilate to let in more light. The identical dilation effect occurs when we encounter stressful situations or when we're exposed to arousing stimuli or thoughts – the amount of dilation is proportional to the strength of arousal.

As pupil dilation is an autonomic process, it cannot be controlled consciously. Exactly this circumstance renders the tracking of pupil dilation an excellent method to assess immediate emotional arousal.

GSR/EDA

Electrodermal activity (EDA), also referred to as galvanic skin response (GSR), reflects the amount of sweat secretion from sweat glands in our skin. Increased sweating results in higher skin conductivity. When exposed to emotional stimulation, we "sweat emotionally" – particularly on forehead, hands, and feet.

Just as pupil dilation, skin conductance is controlled subconsciously, therefore offering tremendous insights into the unfiltered, unbiased emotional arousal of a person. GSR measurements can be done with lightweight and mobile sensors, which makes data acquisition very easy.

In addition, automatic data analysis procedures extract key metrics on the fly, giving you immediate access to the underlying changes in emotional arousal.

EEG

Electroencephalography is a neuroimaging technique measuring electrical activity generated by the brain from the scalp surface using portable sensors and amplifier systems. It undoubtedly is your means of choice when it comes to assess brain activity associated with perception, cognitive behavior, and emotional processes.

Among all biosensors, EEG has the highest time resolution, thereby revealing substantial insights into sub-second brain dynamics of engagement, motivation, frustration, cognitive workload, and further metrics associated with stimulus processing, action preparation, and execution.

Simply put: EEG impressively tells which parts of the brain are active while we perform a task or are exposed to certain stimulus material.

Unlike facial recognition, EEG is able to monitor the global emotional state of a person, which cannot be controlled consciously - you can fake your smile, but you can't trick your brain. Combining the two modalities allows you to get insights into both the moment-by-moment changes in emotional expression as well as variations in emotional states across a longer time span.

EMG

Electromyographic sensors monitor the electric energy generated by bodily movements of the face, hands or fingers, etc. Use EMG to monitor muscular responses to any type of stimulus material to extract even subtle activation patterns associated with consciously controlled hand/finger movements (startle reflex). Also, facial EMG can be used to track smiles and frowns in order to infer one's emotional valence.

ECG/PPG

Track heart rate, or pulse, from ECG electrodes or optical sensors (PPG) to get insights into respondents' physical state, anxiety and stress levels (arousal), and how changes in physiological state relate to their actions and decisions.

Complementing facial expression analysis with eye tracking, GSR or EEG allows you to get insights into both the valence (quality) of an emotional response as well as the amount of arousal (intensity) it triggers in your respondents.

Collecting synchronized data from multiple modalities adds more to the picture as every sensor contributes a valuable aspect of emotional responses that you cannot get with any other. You might even uncover a previously unknown, entirely new effect in cognitive-emotional processing – just think about how your face would brighten up!

Biometric sensors in concert...

Each sensor reveals a specific aspect of human cognition, emotion, and behavior. Depending on your individual research question, consider to combine facial expression analysis with two or more additional biosensors in order to gain meaningful insights into the intricate relationships between the autonomic regulation of emotional arousal and valence, cognition, attention, and motivation.

Facial expression analysis done right with iMotions

Let's recap

While automatic facial expression detection itself is quite easy to do, the analysis and interpretation of the collected data is a bit trickier. Also, as facial expressions only monitor the valence of a response, we recommend to combine facial expressions with other biosensors in order to get the "bigger picture" on how arousal and valence, motivation and emotion interact in response to physical or psychological stimuli.

Before you can kick off your facial expression research, you definitely need to think about which recording and data analysis software to use. Usually, separate software is required for data acquisition and data processing. Although some manufacturers offer integrated solutions, you will most likely have to export the raw data to a dedicated analysis software for data inspection and further processing.

Which multimodal software solution is the one you need to integrate any type of stimulus with eye tracking and other biometric sensors without having to piece everything together?

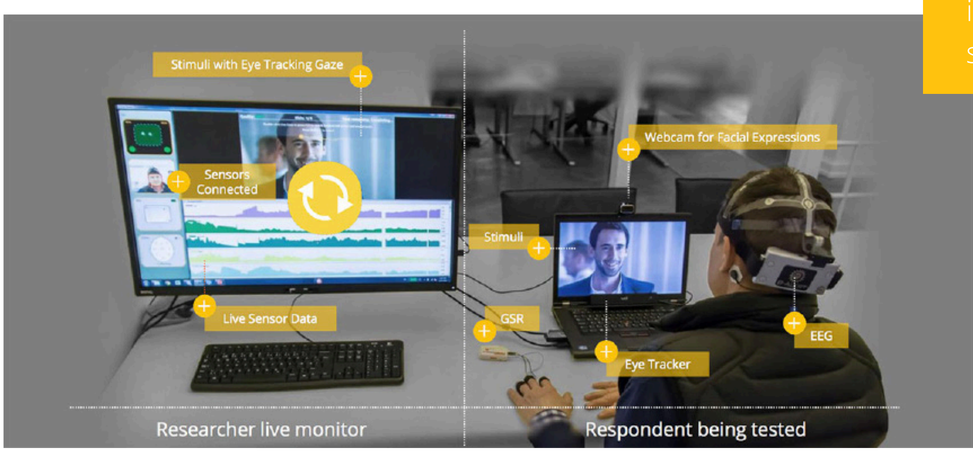
iMotions Biometric Research Platform

IMOTIONS
BIOMETRIC RESEARCH PLATFORM

iMotions Biometric Research Platform is one easy-to-use software solution for study design, multisensor calibration, data collection, and analysis.

Out of the box, iMotions supports over 50 leading biometric sensors including facial expression analysis, GSR, eye tracking, EEG, and ECG/EMG along with survey technologies for multimodal human behavior research.

iMotions multi-modal sensor setup



What's in it for you?

Get the most from facial expression analysis. From start to finish, iMotions got you covered:

1. Run your multimodal study on **one single computer.**

2. Forget about complex setups: iMotions requires **minimal technical skills** and effort for **easy experimental setup** and data acquisition.

3. Get **real-time feedback** on calibration quality for highest measurement accuracy.

4. Draw on unlimited resources: **Plug & play any biometric sensor** and **synchronize it with any type of stimulus**.

5. Receive **immediate feedback on data quality** throughout the recording across respondents.

6. Worried about data synchronization? Don't be. While you can work on what really matters, **iMotions takes care of the synchronization of data streams across all sensors**.

iMotions introduces innovative aggregation of individually coded facial expression data

1. You can use this innovative facial expression aggregation procedure to get **insights into the consistency of facial expressions** elicited by the provided stimulus material as well as the underlying valence of the response.

2. Working with time windows, iMotions can **pool together significant occurrences of facial expressions** that are happening approximately at the same time.

3. iMotions can **count the number of respondents that showed a response** (happy, sad, angry etc.) in a certain time window. The aggregated result relates directly to the audience.

Further reading

Here are our recommended reads to dig deeper into facial expression analysis:

1 Adolphs, R. (2002). Neural systems for recognizing emotion“. Current Opinion in Neurobiology 12 (2): 169–177.

2 Fox, E. (2008). Emotion Science: An Integration of Cognitive and Neuroscientific Approaches. Palgrave MacMillan.

3 Kodra, E., Senechal, T., McDuff, D., el Kaliouby, R. (2014). From Dials to Facial Coding: Automated Detection of Spontaneous Facial Expressions for Media Research. http://www.affectiva.com/wp-content/uploads/2014/09/From_Dials_to_Facial_Coding_Automated_Detection_of_Spontaneous_Facial_Expressions_fo.pdf

4 Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J., & Bartlett, M. (2011). The computer expression recognition toolbox (CERT). In Face and Gesture 2011 (pp. 298–305).

5 Matsumoto, D. (2001). Culture and Emotion. In D. Matsumoto (Ed.), The Handbook of Culture and Psychology (pp. 171-194). New York: Oxford University Press.

6 Müri, R. M. (2015). Cortical control of facial expression. Journal of Comparative Neurology. Accepted article, ePub ahead of print.

7 Rinn, W. E. (1991). Neuropsychology of facial expression. In R. Feldman & B. Rime (Eds.), Fundamentals of nonverbal behavior (pp. 370). New York: Cambridge University Press.

8 Viola, P., Jones, M. (2004). Robust Real-Time Face Detection. International Journal of Computer Vision, 57, 137–154.

North America
285 Summer Street, Suite 201
Boston, MA 02210
United States
Tel +1 617-520-4958

Europe
Frederiksberg Allé 1-3 Fl. 7
Copenhagen V, 1621
Denmark
Tel +45 71 998 098

Synchronize, Visualize and Analyze your research in Eye Tracking, Facial Expression Analysis, Galvanic Skin Response, Surveys, EEG and much more in one software platform.

www.imotions.com