CrossMark

The *blenderFace* method: video-based measurement of raw movement data during facial expressions of emotion using open-source software

Axel Zinkernagel[1] · Rainer W. Alexandrowicz[2] · Tanja Lischetzke[1] · Manfred Schmitt[1]

Abstract

This article proposes an optical measurement of movement applied to data from video recordings of facial expressions of emotion. The approach offers a way to capture motion adapted from the film industry in which markers placed on the skin of the face can be tracked with a pattern-matching algorithm. The method records and postprocesses raw facial movement data (coordinates per frame) of distinctly placed markers and is intended for use in facial expression research (e.g., microexpressions) in laboratory settings. Due to the explicit use of specifically placed, artificial markers, the procedure offers the simultaneous measurement of several emotionally relevant markers in a (psychometrically) objective and artifact-free way, even for facial regions without natural landmarks (e.g., the cheeks). In addition, the proposed procedure is fully based on open-source software and is transparent at every step of data processing. Two worked examples demonstrate the practicability of the proposed procedure: In Study 1 ($N = 39$), the participants were instructed to show the emotions happiness, sadness, disgust, and anger, and in Study 2 ($N = 113$), they were asked to present both a neutral face and the emotions happiness, disgust, and fear. Study 2 involved the simultaneous tracking of 14 markers for approximately 12 min per participant with a time resolution of 33 ms. The measured facial movements corresponded closely to the assumptions of established measurement instruments (EMFACS, FACSAID, Friesen & Ekman, 1983; Ekman & Hager, 2002). In addition, the measurement was found to be very precise with sub-second, sub-pixel, and sub-millimeter accuracy.

Keywords Measurement of facial movement · Emotion expression · Optical measurement · Video data analysis · Open-source software

Facial expression represents one crucial component of an emotion (Scherer, 2005). In the large corpus of research on emotional facial expression, three major measurement methods have been established, (a) the Facial Action Coding System (Ekman & Friesen, 1978; Ekman, Friesen, & Hager, 2002), (b) Electromyography (Fridlund & Cacioppo, 1986; Tassinary, Cacioppo, & Vanman, 2007), and (c) Automatic Facial Expression Recognition (AFER) systems (Corneanu, Oliu, Cohn, & Escalera, 2016). All three methods suffer from certain drawbacks, which will be outlined in the following. The aim of the present paper is to propose the *blenderFace* method, which offers a way to overcome the disadvantages of the established methods and provides a way to take very accurate direct measurements of facial movement at high spatial and temporal resolutions.

The necessity for a new measurement method for facial expressions

The measurement of facial movement via the Facial Action Coding System (FACS; Ekman & Friesen, 1978; Ekman, Friesen, & Hager, 2002), or the Maximally Discriminative Facial Movement Coding System (MAX; Izard, 1983), as

✉ Axel Zinkernagel
zinkernagel@uni-landau.de

[1] Personality, Psychological Assessment, and Psychological Methods, University of Koblenz-Landau, Fortstr. 7, 76829 Landau, Germany

[2] Abteilung für Angewandte Psychologie und Methodenforschung, Alpen-Adria-Universität Klagenfurt, Universitätsstraße 65-67, 9020 Klagenfurt am Wörthersee, Austria

well as the emotional interpretation of facial movements via the Emotional Facial Action Coding System (EMFACS; Friesen & Ekman, 1983), the Facial Action Coding System Affect Interpretation Dictionary (FACSAID; Ekman & Hager, 2002), or the Affect Expression Identification System (AFFEX; Izard, Dougherty, & Hembree, 1983) have emphasized the rating of mostly static facial expressions (Cohn & Schmidt, 2004; Tcherkassof, Bollon, Dubois, Pansu, & Adam, 2007). In a FACS coding, one or more trained FACS coders have to judge the extent of facial muscle movement from mostly a still picture for virtually all facial muscles on a 5-point Likert scale. The extent of movement is graded from A to E for the specific facial muscles and denoted as numbered *Action Units* (AUs). Although it is possible to rate the dynamics of facial expressions in an image sequence, this technique has turned out to be very time consuming (Cohn & Schmidt, 2004). Further drawbacks of a FACS rating with subsequent emotional interpretation are (a) that only easily discoverable but not minor, subtle facial movements can be detected and can therefore be rated, (b) the measurement accuracy of an ordinal 5-point rating scale might not be accurate enough to adequately represent the emotional dynamics of facial expression, whereas facial expression measured at a higher resolution might reveal new findings (e.g., microexpressions; Ekman & Friesen, 1969), (c) the ratings and the subsequent emotional interpretations may be subjected to a rating bias (e.g., Horstmann, 2002). In addition, in this line of research, the stimuli used to elicit a facial expression have mostly been instructional, whereas natural emotion-eliciting situations or experimental settings with induced emotions have rarely been investigated (Reisenzein, Studtmann, & Horstmann, 2013). Therefore, this line of research has raised questions about the ecological validity of the FACS and the generalizability of the results (Motley & Camden, 1988; Russell, 1994).

The second measurement method represents the measurement of emotionally relevant facial muscles via electromyogram (EMG; Fridlund & Cacioppo, 1986; Tassinary et al., 2007). Fridlund and Cacioppo proposed the deduction of 11 emotion-specific facial muscles. According to their proposals, a pair of electrodes[1] should be placed above and parallel to each emotionally relevant muscle fiber, accompanied by a ground electrode placed on the upper forehead. However, due to the physical size of the electrodes and the conductibility of the skin, crosstalk with neighboring muscle groups is possible, for example, the measurement of the *corrugator supercilii* may be confounded by activity from the *depressor supercilii* and *procerus* (Fridlund & Cacioppo, 1986). Although the EMG method can be

applied to measure dynamic activity in the specific muscles used to produce a facial expression, in many cases, the data are highly aggregated in subsequent data processing procedures: In the beginning, the muscle activity is sampled at a high frequency (usually between 10 and 2500 Hz), passed through several filters (e.g., to filter the power supply frequency), optionally integrated or smoothed (e.g., via a moving average), and often split into a pre-stimulus-onset phase and a post-stimulus-onset phase, averaged per phase and then statistically tested. In the end, this leads to a loss in dynamic information aside from an increase in the familywise error rate (Hochberg & Tamhane, 1987). Moreover, in many cases, the activity of only one or two muscles is deduced. This is problematic for the interpretation of emotional facial expressions because first, a single muscle may be involved in several emotional states. For example, according to EMFACS (Friesen & Ekman, 1983), AU1 (*frontalis, pars medialis*) may be involved in the emotions fear, sadness, and surprise, or AU4 (*corrugator supercilii*) may indicate sadness, fear, or anger. Therefore, in many cases, emotion specificity is difficult to determine, and an interpretation regarding valence is preferred. Second, a facial expression might not be shown in a prototypical manner. For example, in a natural, spontaneous facial expression, the emotion might be shown in only some parts of the face (Porter & Brinke, 2008, 2010). Therefore, depending on the position of the electrodes, an emotional expression might not be successfully detected. Besides, the measurement of facial expression via EMG is subject to several practical implications, such as the probably disturbing sensation of electrodes and cables on the facial skin and the need for an electromagnetically shielded laboratory (Fridlund & Cacioppo, 1986).

The third and most recent method for measuring facial expressions relies on the analysis of video data by applying automatic facial expression recognition (AFER) algorithms. Computer vision research in this area has reached a mature state and has provided fascinating results (for an overview, see Corneanu, Oliu, Cohn, & Escalera, 2016). Most AFER procedures include four steps (face detection in a picture or footage, detection of natural landmarks, emotion relevant feature extraction [e.g., mouth corner detection], and expression recognition) and are implemented in both open-source software (e.g., OpenFace; Baltrušaitis, Robinson, & Morency, 2016; Menpo; Alabort-i-Medina, Antonakos, Booth, Snape, & Zafeiriou, 2014; IntraFace; De la Torre, Chu, Xiong, Vicente, Ding, & Cohn, 2015; Computer Expression Recognition Toolbox; CERT; Bartlett, Littlewort, Wu, & Movellan, 2008; Littlewort, Whitehill, Wu, Fasel, Frank, Movellan, & Bartlett, 2011) and commercial software (e.g., Affectiva, 2017; Emotient Inc., 2016; Kairos, 2017; Noldus Information Technology, 2017; RealEyes, 2017). The AFER procedures have been able to achieve a

---

[1]This procedure maximizes selectivity for the deduced muscle fiber (Fridlund & Cacioppo, 1986).

reliable and valid classification of emotional expressions from footage and images, close to or even better than humans (Littlewort, Bartlett, & Lee, 2009; Terzis, Moridis, & Economides, 2010). Despite the fact that the AFER procedures achieve a very good correct classification rate, these procedures have drawbacks from a epistemological and psychometric perspective and may be problematic for psychological facial expression research for the following reasons: (a) Most AFER procedures return highly aggregated output, either through basic emotion classification or AU activation estimates. In most cases, the classification is performed by artificial neural networks that have been trained by one or more preclassified training samples. This means that only preclassified categories can be detected. For example, an immanent assumption is that the classification categories and the corresponding training samples are correct, but this assumption may be arguable, for example, regarding the number of prototypical expressions involved in basic emotions (Jack, Garrod, & Schyns, 2014) or regarding whether basic emotions have a prototypical appearance (Crivelli, Jarillo, Russell, & Fernández-Dols, 2016a; Crivelli, Russell, Jarillo, & Fernández-Dols, 2016b; Elfenbein, Beaupré, Lévesque, & Hess, 2007; Jack, Garrod, Yu, Caldara, & Schyns, 2012). Therefore, AFER procedures bear the risk of being plagued by circular reasoning because classifications will always be consistent with the predetermined classes, and thus the procedures will not be able to detect prototypical facial expressions other than the ones they were trained to detect.[2] (b) In addition, when AFER algorithms are applied to classify emotions, they suffer from the shortcoming that the intensities of emotion expressions are not measured, although there are approaches that can do this (Corneanu et al., 2016). In the case of AU activation classification, the output represents a downgrade in the scale of measurement from metrically measured 2D or 3D facial movement to a nominal scale, which is therefore accompanied by a loss in information and accuracy (e.g., AU 12 represents the lip corner puller, AU 15 represents the lip corner depressor). (c) Despite the fact that the classification rates of AFER algorithms are very good (Littlewort et al., 2009; Terzis et al., 2010), and the commonly used support vector machines (SVMs) have a strong mathematical basis, translating the classification rules of AFER algorithms into a format that can be comprehended by humans is a complex procedure. For example, when an SVM is used as a classifying algorithm, the coordinates of an n-dimensional hyperplane must be interpreted semantically. Therefore, the AFER algorithm output may be used descriptively (e.g., the occurrence or nonoccurrence of a specific emotion) rather than analytically (e.g., to test various hypotheses about measured facial movement). (d) Most AFER procedures also allow lower level output in the sense of coordinates of tracked fiducial points/natural landmarks (e.g., the corners of the mouth or the eyes). This information is gathered during the face segmentation phase and is needed to be able to classify facial expressions of emotion. These natural landmarks are determined, for example, by exploiting color and texture information along with ellipsoid fitting or via face saliency maps (Corneanu et al., 2016). Therefore, AFER algorithms try to match a predetermined pattern onto each frame of the footage or image. Because this is a relatively robust[3] but not always very accurate and reliable method, a *confidence level* is computed. This confidence level is estimated by the tracking algorithm and represents the confidence in the current landmark detection. However, the tracker and subsequently the computed confidence level are not bound to any external, "true" criteria in a psychometric sense. Therefore, the confidence level can take on high values, even when the classifications are completely wrong (Evtimov, Eykholt, Fernandes, Kohno, Li, Prakash, & Song, 2017; Sharif, Bhagavatula, Bauer, & Reiter, 2016), thus causing potentially misleading results. (e) In addition, AFER algorithms rely on tracking the natural landmarks of the face (e.g., corners of the eyes and mouth, eyebrows), but there are procedures that allow researchers to model parts of the face that lack natural facial features (e.g., Wood, Baltrušaitis, Morency, Robinson, & Bulling, 2016), for example, the analysis-by-synthesis approach. The basic idea used in this approach is that it replicates (a subaspect of) a given entity and iteratively minimizes the differences between the replica and the target entity. The process of replication and the replica aid the understanding of the entity. For example, Wood et al. (2016) estimated a 3D model (along with light conditions, skin tone, etc.) for a part of the face that lacks natural landmarks and minimized the difference between the original image and the 3D-rendered image by adapting the 3D model in an iterative process. However, this approach does not represent the measurement of facial movements in facial areas that do not have facial landmarks (e.g., movement of the skin on the checks), but it represents a synthesis which estimates the part of the facial surface iteratively to minimize the differences in visual appearance. (f) For commercial AFER software, in most cases, the emotion classifier algorithm that makes a decision about an emotion (and its intensity) is part of the secrecy of the business and is therefore unknown. This means that the interpretation and a conclusive assessment of the results remain arguable and do not conform to open science requirements.

---

[2]However, it should be mentioned that recent unsupervised learning approaches have attempted exactly this, with the goal of jointly learning feature extraction and classification (cf. Corneanu et al., 2016).

[3]If natural landmarks are hidden (e.g., if a hand moves in front of the mouth), the ellipsoid fitting algorithm finds and tracks the natural landmarks when they become visible again.

Aims of the present research

The aim of the present research was to develop a new measurement method for psychological facial expression research that avoids the disadvantages and combines the advantages of established methods while keeping the procedure simple and clear. This method should meet the following requirements: (a) The new method should allow a very accurate, metric-level procedure for measuring raw, uninterpreted (e.g., in the sense of prototypical emotions, or AUs) facial movements. The measurement of facial areas that lack natural landmarks must be accurate and reliable. The data generated by the measurement procedure must ideally be free from measurement artifacts and trustworthy in the sense of psychometric objectivity (i.e., via visual transparency and the comprehensibility of the measurement procedure). The measured raw facial movement data can later be analyzed by applying various statistical analyses for testing different theoretical approaches in describing the facial movement. (b) The measurement of the temporal dynamics of facial movements should be possible for approximately 10 min of video footage per participant and sample sizes of around 100 participants.[4] (c) The simultaneous measurement of facial movements for the whole face should be possible, without handicapping the participating subject to a greater extent (e.g., via cables) and being as unobtrusive as possible to the participant in order to facilitate the measurement of facial movements in natural settings. (d) The measurement procedure must be flexible enough to be easily adapted to various research questions/experimental settings and also be extensible by providing various interfaces at different levels of data processing, for example, via the use of open-source software. Although very sophisticated solutions for partial aspects of the proposed measurement procedure have been developed (e.g., building an individual facial surface; Jeni, Cohn, & Kanade, 2015; Suwajanakorn, Kemelmacher-Shlizerman, & Seitz, 2014, or the scaling of virtual objects; Ham, Lucey, & Singh, 2014), these approaches could not be integrated because they are used to achieve a different aim (e.g., a photorealistic 3D representation of the facial surface), have technical requirements that do not fit into our experimental setting (e.g., a moving camera with an inertial measurement unit[5]), do not support our aim of directly measuring facial movements, or are very

difficult to integrate into the rest of the measurement workflow.

The blenderFace method

The basic idea is simple because it adopts an approach developed by the film industry: Around 1990, when computers were powerful enough to allow the rendering of animated 3D faces of figures, film directors were confronted with the problem that scripted facial movements of the 3D-figures (e.g., the direction, extent, and speed of the movement of virtual facial muscles[6]) did not appear authentic and plausible to human perception. This problem was solved by employing a motion-capturing procedure: Markers were placed on the head and on emotionally relevant positions of an actor's face, and the face was recorded while the person was acting. Subsequently, the movements of the markers were digitally tracked and transferred onto the virtual face of a 3D figure. In principle, the proposed procedure follows this idea; however, it is aimed solely at measuring facial movements.[7] In contrast to the expensive equipment used in the film industry, a single standard webcam and the open-source software *Blender* (www.blender.org) are all that is needed to apply the suggested procedure. The proposed approach uses a 3D model to represent head movements and face topology and is therefore not subject to measurement bias due to head movements. In addition, the proposed method is deliberately based on markers that are applied to the face (i.e., markers painted on the skin), opposed to the naturally available landmark fitting procedures of AFER algorithms, as applied for example in OpenFace (Baltrušaitis et al., 2016). The use of applied markers means that movements of the facial skin can be measured very precisely and reliably even in parts of the face that are far from the available natural facial markers (e.g., on the cheeks). Furthermore, the proposed measurement procedure satisfies all requirements mentioned earlier: (a) very accurate measurement at (b) high temporal resolution for an arbitrary time period (e.g., 10 min) with (c) the simultaneous measurement of movements for the whole face and with (d) an open and flexible procedure.

For the postprocessing of the measured data (e.g., visualization, standardization, plausibility checks), we developed the open-source `blenderFace` package for the free and open-source R statistical programming language (R Core Team, 2016). In the following, we describe the

---

[4]This argument mainly refers to FACS coding because it is possible with EMG or AFER procedures anyway. However, it should also be a prerequisite for the proposed procedure.

[5]If the camera and the head move, the separation between head and camera movement would be possible only with the help of a static background that would have to be modeled as well. However, taking the background into account would complicate the measurement procedure.

[6]In fact, the MPEG-4 standards ISO/IEC 14496-1 & 2 contain facial animation parameters, similar to the action units used in FACS.

[7]The subsequent transfer of (aggregated) human facial movements onto a virtual 3D face is possible (e.g., for emotion recognition studies). However, it is not a central concern of this paper.

proposed procedure in principle. For a detailed description, see the supplemental material of this article, especially the documents "*Step_by_Step_Instructions.pdf*" for the tracking procedure with Blender, and the vignette from the `blenderFace` package "*blenderFace.pdf*" on how to postprocess facial movement data. The supplemental material also includes an example video and training video clips as well as example Blender data files.

Tracking setup

Illumination Because the proposed procedure is an optical measurement, a well-illuminated, shadow-less face and head is essential. Shadows may change the visual pattern, which should be tracked, and thus may cause the Blender pattern recognition algorithm to fail and to abort tracking. This does not constitute a complete fail because Blender is easily instructed to use an updated pattern; however, it decelerates the tracking procedure and requires manual intervention.[8] It is also advisable to use a flicker-free LED illumination and to switch off neon tubes because neon tubes flicker according to their power supply frequency, which may interfere with the webcam shutter frequency and lead to a small but noticeable up-and-down sliding of tracked markers, however, not impairing the blenderFace method in general.

Camera There are no special requirements for a camera. Any existing lens distortion, e.g., from a wide-angle lens, can be corrected in later steps of the blenderFace method. The camera must be firmly mounted (e.g., on top of the computer monitor) to record the face of the participant from a frontal view. We successfully used a Logitech C910, a Logitech C920 Webcam, and a Mobius Actioncam with resolutions of 640 × 480 pixels (px), 1280 × 720 px and 1920 × 1080 px at 24 and 30 frames per second (fps). The higher the resolution of the camera, the more accurate the measurement of facial movements, however, at the cost of a larger video clip file size. Moreover, the final video clip file size also depends on video length and the type of compression that is chosen for the video clip file.

Regarding the optimal frame rate, according to the Nyquist–Shannon sampling theorem (Shannon, 1949), a signal should be recorded at twice the frequency of the maximal frequency that is to be measured. For practical reasons, for example due to noise and measurement errors, the rate should be 4 to 6 times that high (Fridlund & Cacioppo, 1986). Because the fastest facial movements

are in a range of approximately 250–300 ms (Dimberg, Thunberg, & Grunedal, 2002; Porter & Brinke, 2008; Yan, Wu, Liang, Chen, & Fu, 2013), a framerate of 24 or 30 fps is still more precise by a factor of ˜10 and can therefore be considered sufficient to measure even rapid facial movements. To achieve a high and reliable video quality, automatic adaption of the frame rate, aperture, etc. should be disabled or held constant by the webcam driver configuration. In addition, the video should be optimally saved in a lossless format or at high data rates in order to prevent compression artifacts in the video clip and to increase tracking speed in a later step of this procedure.

Synchronization Options for and the expense of synchronizing the recorded footage with external events (e.g., the beginning and the ending of the stimulus presentations) depend to a large extent on experimental and laboratory settings. From the technical side, Blender is capable of processing time codes (e.g., from time-code capable cameras) or building time-code proxies, which may be used in combination with time-code-capable stimulus presentation software or data-recording devices (e.g., EEG). However, in many cases, the technical expense of using a time-code or beacon synchronization may be more than is necessary, and a burn-in of a time stamp with the option of additional information (e.g., the subject number, action markers, stimulus presentation episodes) inserted into the video data stream should be sufficient. As a fallback, it has been shown to be advantageous to have a mirror or a second monitor behind a participant's back, thus recording the stimuli as well. This provides a simple yet effective, reliable, and technically easy way to synchronize the presentation of the stimuli.

Characteristics of the participants In general, there are no restrictions on the identification of participants who are suitable for undergoing the measurement procedure. Table 1 presents an overview of the reasons that participants had to be excluded from the studies presented in this paper. Problems may occur when the markers that need to be tracked are drawn on the participants' skin: The markers should be placed in a way that prevents them from disappearing into skin folds during expression of emotion (e.g., in the fold of a smile). Heavily made up or bearded participants are also difficult to track because the markers might not be clearly visible. However, this does not mean that such participants must be completely excluded, but marker tracking needs more corrections and manual work to ensure that the markers are tracked correctly. Facial markers that are visible through participants' glasses are more problematic because they appear in a biased position. This leads to a higher tracking error because lens distortion

---

[8] With unlimited time resources, the markers could also be set manually for each frame, even if no marker is visible (more a guess than a measurement). This would also result in a "successful tracking", but probably at the expense of a decrease in model fit.

Table 1 Reasons and number of participants excluded from data analysis in studies 1 and 2

	Study 1	Study 2
N recruited	55	150
No / too weak / wrong markers	7	0
Extreme head movements (e.g., shaking with laughter, asking the investigator, "fidgety" participant)	4	12
Recording issues (e.g., footage over- or underexposed, cast shadows, technical problems)	3	19[a]
Contrary to instructions, participant's hand covers the markers during large parts of the recording	2	2
Extreme / no head movement during parallax phase; facial movement during parallax phase	0	2
The participant was heavily bearded	0	1
The participant wore large glasses	0	1
N analyzed	39	113

Note. Number of participants in Studies 1 and 2 who had to be excluded from the analysis for the reasons given in the table

[a]In Study 2, we switched off the neon tubes to avoid the flickering of the video recording. However, this led to the footage being overexposed / underexposed in the early morning and late evening hours

by glasses is not considered in Blender's optical model.[9] Therefore, markers that are visible through glasses should not be tracked. No further standardization procedures (e.g., a fixed distance between the camera and the face at the beginning of the measurements) are required for the participant.

Markers For the proposed procedure three types of markers are needed: (a) *static markers* on the head of the participant to track head movements and disentangle these movements from movements that are part of emotional facial expressions, (b) *surface markers* to estimate the individual's facial surface, and (c) *emotion markers* to finally track the emotionally relevant facial movements. To measure head movements, 12 static head markers (Blender needs a minimum of eight) that must be visible in all frames in the video clip are recommended to track. These markers must be mounted in a fixed position, for example, on a cap or on the headphones (see Fig. 1) and tracked throughout the video. If some markers are hidden by parts of the face or the head (e.g., during extreme head turns), it is possible to track additional static head markers that overlap for a few frames with existing static head markers at the beginning and end of that episode. In addition, to ensure a reliable and accurate three-dimensional tracking of head movements, the static head markers must be placed on different levels of the depth of the head (e.g., some on the forehead and some at the depth-level of the ears).

A pattern of 68 surface/emotion markers on the participant's skin (cf. Figs. 1 and 2) is needed to allow a precise measurement of facial movement. Markers, irrespective of type, should be easily recognizable as a

relatively stable pattern by the Blender pattern recognition algorithm over the sequence of the frames of the video clip. This is best achieved by a large contrast in shape, brightness, and color compared with the background. In the studies presented in this paper, we used colored glue dots on a cap and a headphone for static, distinct, and quickly drawn black fluid eyelid-liner dots on the facial skin for surface/emotion
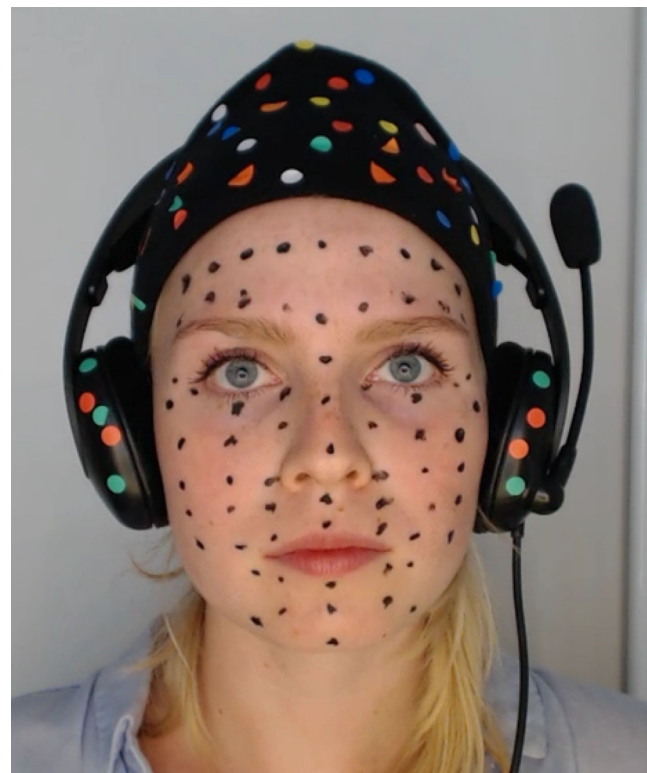

Fig. 1 Participant with glue dots as *static markers* on a cap and headphone and black fluid eyelid liner dots as *surface-* and *emotion markers* painted on the facial skin

---

[9]Although it would be possible to model the level of distortion caused by the glasses in Blender, in our opinion, the end does not justify the means as the necessary effort would be too great.
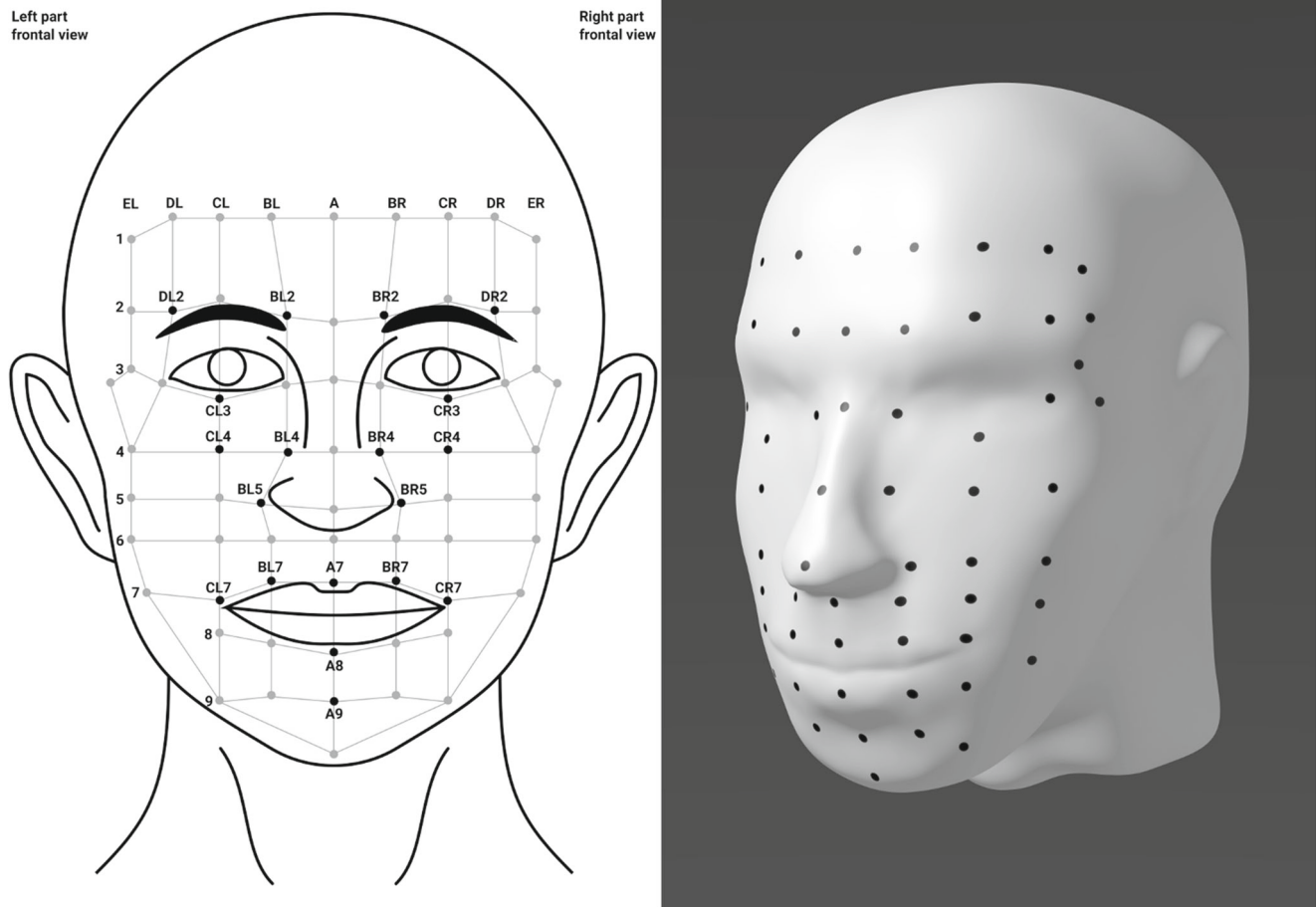
Fig. 2 Positions of proposed facial surface and facial emotion markers. The *left part* of the figure shows a 2D projection and the *right part* a 3D projection of the markers on the face. In the left part of figure, *grey markers* represent the facial surface markers needed to estimate the individual's facial surface; *black markers* additionally represent facial emotion markers that may be associated with the AUs from the FACS

markers were used, which can reliably be recognized as a stable pattern by Blender's tracking algorithm (see Fig. 1). This preparation step was completed in approximately 2 min per participant in our studies.[10]

Parallax phase To accurately estimate the three-dimensional surface of the participant's face, the 68 surface markers (see grey dots in left part of Fig. 2) are needed on the participant's face to allow for a stable and reliable assessment in three dimensions.[11] It is sufficient to track the surface markers for only a short episode (e.g., 100 – 200 frames) to estimate an individual's facial surface. It is important

that during this episode, no facial movement is shown and that the episode contains parallax of the head. Parallax is displacement in the apparent position of an object viewed along two different lines of sight. To be able to use images from the video clip to estimate the three-dimensional surface of the face, at least two images from different lines of sight are needed. In our case, this is the head with the face viewed from two slightly different perspectives.

Therefore, a short parallax phase at the beginning or the end of the video is needed to generate parallax for the face in order to provide an individual estimate of the three-dimensional facial surface: We instructed participants to direct the tips of their noses toward the camera lens and toward the middle of the right border of the computer screen for 3 s at each head position. In general, any brief episode in the footage that shows a slight head movement with no facial expression can be used. Therefore, a suitable parallax episode may also be found for participants who do not comply with the parallax phase instructions. However, participants who do not show any head movement at all

---

[10] At the beginning of data collection the research assistant needed about 5 min to prepare the test subjects; after some practice, at the end, it was less than 1 min.

[11] In general, the number of facial surface markers is arbitrary, but the suggested number of 68 markers has been confirmed to allow a good approximation of an individual's facial surface. As an alternative to an individual mask, it is also possible to use a standard facial surface or even a plane as a general projection surface for the emotion markers. However, this occurs at the cost of a loss in measurement accuracy.

(during the parallax phase or in any other phase) or always show a facial movement during the parallax phase have to be excluded from the tracking procedure. An inappropriate parallax phase (very slight head movements or facial movement during the parallax phase) will result in a high value for the *solve error* and will thus result in an imprecise 3D model and an inaccurate estimation of head movement.

The estimated individual facial surface is needed to provide a projection surface for the emotionally relevant markers. Surface markers at emotionally relevant positions on the face can be used not only for facial 3D surface estimation, but also tracked for the emotionally relevant episodes from the video clip. The emotionally relevant episodes are the sections of the video clip that are of substantial interest. This can be, for example, a social interaction or several phases of an experiment in which different stimuli are presented. The movement of these emotion markers represents the outcome of this measurement procedure and can be interpreted emotionally (see the black dots in the left part of Fig. 2). Therefore, emotion markers have to be tracked only for the episodes for which researchers want to measure emotional expression.

Tracking procedure

In this section, we describe the general blenderFace method of tracking the static markers on the head, the facial markers used to generate an individual 3D model of a participant's face, and the emotion markers. This procedure ensures that the movement of the markers is measured with high precision and independently of head movements. At the end of this section, raw motion data as well as scaling data are exported for further standardization and statistical analyses.

Starting with Blender version 2.61, the *Movie Clip Editor* provided a motion tracking module that relies on a visual pattern recognition algorithm. In this algorithm, one or more key visual patterns (markers in our case) have to be defined on a start frame. In the sequence of follow-up frames, these key patterns are searched for in a predefined search area.

Tracking head movements In the first step of the tracking procedure, the static head markers have to be tracked to obtain information about head movements. In addition, the 68 facial surface markers must be tracked for the short parallax phase of the video clip. Subsequently, from the parallax displacement of the tracked markers for the different frames of the video clip, 3D coordinates for each tracked marker, along with the Blender's virtual camera movement, are computed in the Blender 3D space. In contrast to the setting in reality, in which the head and the (neutral) face turns in front of a static camera, in Blender, the movement is reattributed to the camera. Therefore, in Blender's 3D space, head movements are
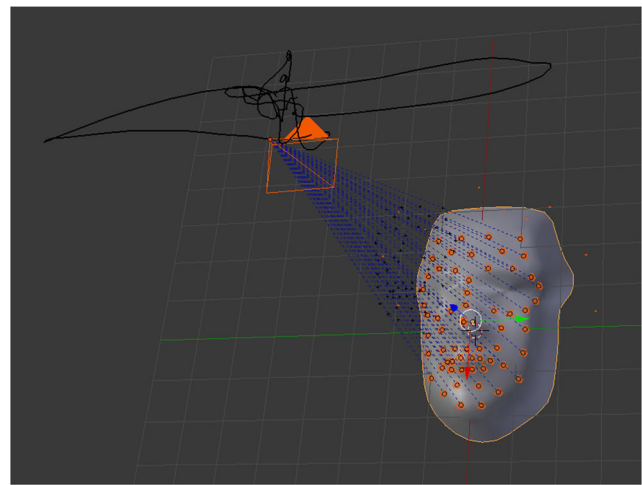

Fig. 3 Blender screenshot: the *orange pyramid* in the upper left represents the camera; the *black line* represents the movement of the camera in front of the grey static facial mask; the *orange circles* on the mask represent the tracked emotion markers projected from the camera onto the facial mask (*blue dotted lines*)

transformed into virtual camera movements and the head remains in a static position (see Fig. 3). For example, a head turning upward is represented as a camera moving downward, and a head turning to the left is represented as a camera moving to the right. This reattribution is optically,[12] logically, and mathematically equivalent to the original movement. Because of the simultaneous estimation of the 3D coordinates of the tracked markers, along with Blender's virtual camera movement, they both are represent as one "marker coordinates/camera track" unit in Blender. This has three advantages for our purpose: (a) This "marker coordinates/camera track" unit can be moved to the origin of Blender's 3D coordinate system without affecting the proportions. This is important in later steps of the blenderFace method because it facilitates the exporting of meaningful coordinates. (b) The "marker coordinates/camera track" unit can easily be rescaled, also without affecting the proportions. This is used in later steps of the blenderFace method to rescale the "marker coordinates/camera track" unit, for example, into mm. (c) Due to the reattribution of the movement to the camera the virtual head is static and does not move in Blender's virtual coordinate system. As a consequence, the emotion markers that are tracked later will no longer be affected by head movements, and this also simplifies the exporting of emotion marker movements in later steps of the blenderFace method.

The overall accuracy of this "marker coordinates/camera track" unit estimation is made available in Blender's solve error parameter. The solve error represents the mean

---

[12]Ignoring the background.

deviation of the marker position on the basis of the parallax computation from the actual tracked marker position in the video clip. The solve error should be below 0.3, which means a mean deviation of a third of a pixel between model-based tracks and the tracked markers on the video clip.[13]

Building the individual facial surface In the second step, the individual 3D surface of the participant's face is built on the basis of the 3D coordinates of the 68 surface markers. First, the facial surface is constructed by connecting four markers at a time to form rectangles (see the connection lines between the facial surface markers in the left part of Fig. 2). Subsequently, this rough approximation of the facial surface by the rectangles is interpolated and smoothed to closely fit the participant's real facial surface. Afterward, the tip of the nose in the facial surface is centered at the origin of Blenders 3D coordinate system.

Tracking emotion markers In a third step, the facial emotion markers must be tracked for the emotionally relevant episodes from the video clip. The movement of these markers is also exported into Blender's 3D space, namely as a projection from the moving camera onto the static facial surface (Fig. 3). The movement of these markers on the facial surface represents the movement of the markers painted on the participants' skin. A python script uses Blender's application programming interface (API) to access the 3D coordinates of the emotion markers per frame, exports the coordinates, and saves them in a comma-separated values file (CSV) for each participant for further data processing.

Scaling and standardization procedures Because coordinates in Blender are represented in Blender's default unit of measurement, the so-called Blender Unit (BU), two individual scaling procedures must be followed in a fourth and last step to allow the Blender data to be rescaled into meaningful measurement units. In principle, 1 BU roughly represents 1 meter in the real world.[14] However, the process of taking the data from the tracks of the video clip and transforming them into Blender's 3D space cannot be controlled and is therefore relatively arbitrary. Therefore, a BU measurement of a real-world object of known size is needed for the rescaling. In practice, the diameter of a glue dot on a participant's headphone with a known diameter of—in our

case—8 mm is measured in BUs. With this measurement of distance, it is possible to rescale the marker movement, originally measured in BUs, into mm. The second standardization procedure addresses the problem of comparing the facial expressions of participants with different face sizes, for example, a child's face with an adult's face. To prevent an effect of a potential bias in face size on the extent of facial movement, the eye–eye distance must be measured and used to rescale movement along the x-axis. Accordingly, the eye–mouth corner distance is used to rescale the movement along the y-axis. This allows marker movements to be represented in a "standardized" face so that comparisons of the movement can be made across individuals.

The complete tracking of 14 emotion markers from a 12-min video clip takes approximately 40 min.[15] In addition, it meets all the requirements described earlier when we argued for the necessity of a new method for measuring emotional facial expression.

Postprocessing of blender data

The data generated by the proposed Blender tracking procedure are saved in a CSV file for each participant and need to be postprocessed in order to be analyzed statistically. Because the amount of data and the resulting file size can get quite large,[16] and also so that we would have a standardized postprocessing procedure, we developed the `blenderFace` package[17] for the R language (R Core Team, 2016). The `blenderFace` package serves to (a) concatenate the single CSV files into one R-data file, (b) rescale the data into mm or into a standardized face, (c) center the marker movement at the onset of stimulus presentation, (d) plot raw and aggregated data for plausibility and descriptive checks, (e) compute higher order variables of movement, such as the angle and the median distance of a marker movement, and (f) makes use of several CPU cores, if available, to speed up

---

[13]Unfortunately, this value is not mentioned in the blender documentation but only in tutorials (e.g., DVD Training 9: Track, Match, Blend! (https://store.blender.org/product/track-match-blend/) or on websites (e.g., https://blender.stackexchange.com/questions/53435/solve-error-high-with-good-track).

[14]https://www.blender.org/manual/glossary/index.html

[15]This assessment is related to the video quality (including lighting, marker quality, head movements) of the sample video in the supplemental material. In the case of poor illumination, poorly visible markers, strong head movements, and so forth, more manual corrections are necessary and the processing time of a video clip is extended.

[16]For example, if 14 markers are tracked in a 12 min video of a participant, there will be 14 marker * 3 dimensions * 12 min * 60 s * 33 fps = 997920 data points per participant. Blender internally stores the data points as a "float" data type. In the CSV file, each data point is saved with 20 characters (including decimal points and value separators, but ignoring negative signs and NAs). If each character of a data point is represented by one byte (assuming an 8-Bit ASCII character set), the tracked video clip of a participant results in a 997920 * 20 = 19.96 MB CSV file.

[17]https://github.com/axzinker/blenderFace

the postprocessing. Examples of these functions will be presented in the sections describing Studies 1 and 2. For a detailed view of the procedures, see the vignette of the `blenderFace` package. In the following, the general principles of the functions are described.

Concatenating Blender's CSV data The first step in postprocessing is to concatenate the Blender data for each participant into one large data file that can be analyzed more easily. The appropriate function in the `blenderFace` package performs plausibility checks for example, it tests whether unique marker names have been used in the CSV files and also integrates the data when different numbers of markers have been tracked for the participants.

Rescaling into meaningful measurement units In a second step, the Blender data, which are scaled in BUs, need to be rescaled into more meaningful units of measurement. The `blenderFace` package contains two functions to perform a rescaling into either mm or a standardized face. In principle, the rescaling is performed via the rule of proportion.

To rescale into mm, we use the measurement in BUs of an object for which the dimensions in the real world are known. In the Blender tracking procedure, the diameter of an 8-mm glue dot is used. If the individual measurement of this glue dot diameter in BUs was 0.03, for example, the proportions would be constituted as follows:

$$\frac{\text{glue dot size in BU}}{\text{glue dot size in mm}} = \frac{\text{value in BUs to be scaled}}{\text{outcome value in mm}}$$

$$\frac{0.03}{8} = \frac{\text{value in BUs to be scaled}}{\text{outcome value in mm}}$$

After adequately solving this equation, it is possible to rescale the x-, y-, and z-coordinates of the markers into mm.

Rescaling the Blender data into a standardized face follows a similar procedure. We define the standardized face as a two-dimensional square of length 1. According to general proportional features of the face, the eye–eye, and the eye–mouth distances are each set to be 1/3 of head width and head height, respectively. The individual eye–eye distance[18] and eye–mouth corner distance were measured in the preceding Blender tracking procedure and are then used to rescale the x-axis and the y-axis, accordingly. For example, if the individual eye–eye distance in BU is 0.4, the rule of proportion is constituted as

$$\frac{0.4}{1/3} = \frac{\text{value in BUs to be scaled}}{\text{standardized outcome value}}$$

for the x-axis. The y-axis is scaled accordingly. However, we restrained from rescaling the z-axis because we were

not able to find a convincing scale factor. For example, the distance between the eyes is largely stable in proportion to the head and body size, because this distance is needed for stereoscopic vision. This is not the case for the height of the nose (which might be used to rescale the z-axis) because the height of the nose differs significantly between individuals and may be influenced by the climate zone that an individual's ancestors came from (Noback, Harvati, & Spoor, 2001). The z-axis may be considered in later versions of the `blenderFace` package; however, for the statistical analyses and two dimensional plots presented in the following, no z-axis is needed. The reason is that according to test runs only a negligible amount of variability in facial expression movements takes place along the z-axis. Therefore, ignoring the z-axis provides computational efficiency with presumably very little loss of information. Moreover, the z-values are predetermined by the facial surface on which the markers move. Therefore, using the 3D facial surface as a projection surface has the function of preventing a projection bias that, for example, a flat projection surface would produce.

Centering data at the beginning of an emotionally relevant episode Because individual faces differ in their size and topology, it is not possible to draw the markers at the exact same standardized position of the face for each participant. If uncorrected, between-persons differences in drawn marker positions would introduce unsystematic error into the measurement procedure. This means that if we aggregated uncorrected raw data across participants, the
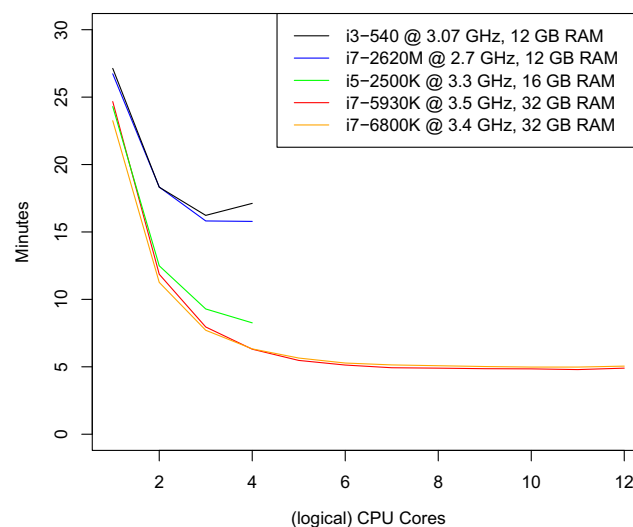

Fig. 4 Relationship between CPU cores used and the duration in minutes for the emotional episodes centering process for different processors. All calculations were done in RAM without swapping. The data set used for centering contained 113 participants, 14 emotion markers, and eight emotional episodes per participant
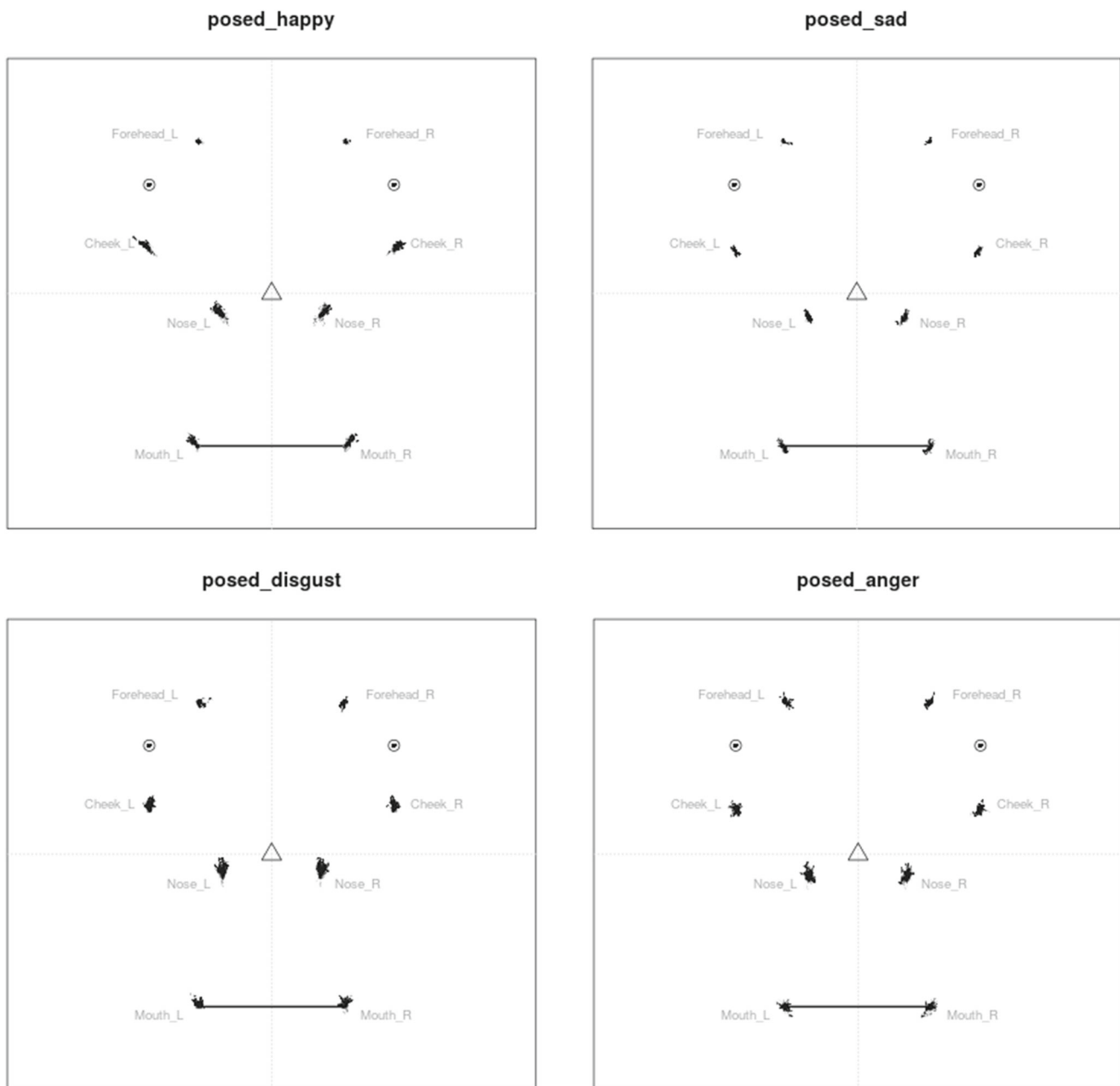
---

[18]In fact, the pupil–pupil distance is measured while the participant looks directly into the camera.

posed_happy

posed_sad

posed_disgust

posed_anger

Fig. 5 Raw data plot of centered marker movements of a standardized face per emotionally relevant episode and aggregated over participants. For each stimulus episode (posing happiness, sadness, disgust and anger), four symmetrical facial markers were plotted on the forehead, the cheeks, the nasolabial area, and on the mouth corners

variability in start positions of a marker would bias the start positions of the movement. Therefore, it is necessary to center the markers at the onset of each emotionally relevant episode. This is possible because it is not the absolute position of the marker on the face that is of interest but the marker's movement in reaction to a presented stimulus.

Depending on the experimental settings (e.g., the number of subjects, the number of experimental conditions, the number of tracked emotion markers, the length of the tracked footage), the raw data set containing the tracked

markers can become relatively large.[19] To center the emotionally relevant episodes of the raw data set, the corresponding R function can use several CPU cores to speed up the centering process. To estimate how long the centering process might take, Fig. 4 shows the relationship

---

[19]For example, the raw R data file from Study 2 with 113 participants, eight experimental conditions and 14 emotion markers was tracked for approximately 12 min per participant and had a file size of 240 MB. See also footnote 16.

between the CPU cores that were used and the duration of the centering process in minutes for different processors.

In principle, the centering is performed by selecting the values of the stimulus onset frame for the *x*-, *y*-, and *z*-axes of a marker per presented stimulus per subject. Subsequently, these values are subtracted from the corresponding values of the following frames for the duration of the episode in which the stimulus is presented. For example, if the onset frame for the stimulus episode "posing disgust" of Subject 37 for the x-values of the marker position contains "−14, 5" this value is subtracted from the *x*-axis values of all frames within the stimulus episode "posing disgust".

Visual representation of the data A visual representation of the data at different levels of aggregation offers a quick check of plausibility and also allows possible outliers and artifacts to be detected (e.g., markers disappearing in skin folds, markers hidden by a hand that is moved in front of the face, tracks jumping between two different positions because of two highly probable matching patterns, etc.). To keep things simple, all plot functions commonly ignore the *z*-axis, for the reasons given above.

The `blenderFace` package offers functions to plot (a) individual or aggregated raw data or marker movement on a standardized face to get an impression of overall marker movement and detect markers that may contain outliers (e.g., Fig. 5), (b) individual median movement per marker
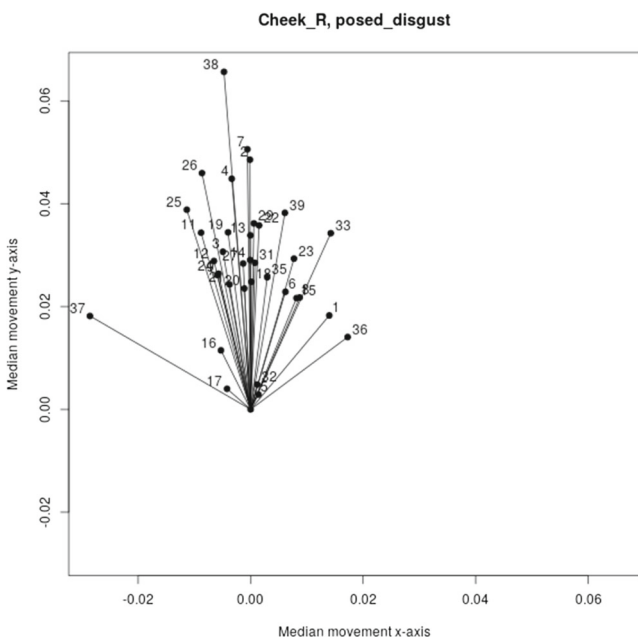
to detect individuals with unusual marker movement (e.g., Fig. 6), (c) x- and y-movement of (symmetrical painted) markers per frame to identify frames with suspicious marker movement (e.g., Fig. 7), and (d) individual or aggregate median movement per stimulus episode with quartile ellipses to get an overall impression of marker movement per presented stimulus (e.g., Fig. 8). These plots are explained in more detail in the sections in which Studies 1 and 2 are presented.

Higher order parameters of facial movement After the Blender data are postprocessed and corrected for outliers and artifacts, it is possible to analyze the data in several ways. Currently, the `blenderFace` package provides functions to compute the angle and the distance to compare marker movements with respect to direction and distance across different stimuli. However, the package will be developed continuously to extend its capabilities. One of our tasks, for example, is to add functions to compute speed, the onset, apex, and offset phases of an expressive episode as well as symmetry parameters of facial movement.

Study 1

The main purpose of Study 1 was to determine the optimal tracking settings, tracking parameters, and tracking

Fig. 6 Plot of the centered median movement of the right cheek marker (Cheek_R) for the "disgust" stimulus episode for each individual as denoted by the subject number. Movement is scaled to the standardized face. The *x*- and *y*-axes of the plot have the same range so that the plotted angle reflects the correct direction of movement

Fig. 7 Plot of Participant 37 for the disgust stimulus episode (Frames 1707–1969) for the right (*red*, *orange*) and left (*blue*, *green*) x- and y-axis movements of the cheek markers. Note that the origin of the coordinate system is at the tip of the nose; therefore, the left and right markers target the opposite directions on the x-axis (*red and blue lines*). The plot's x-axis represents the frame numbers, and the y-axis represents the movement in mm
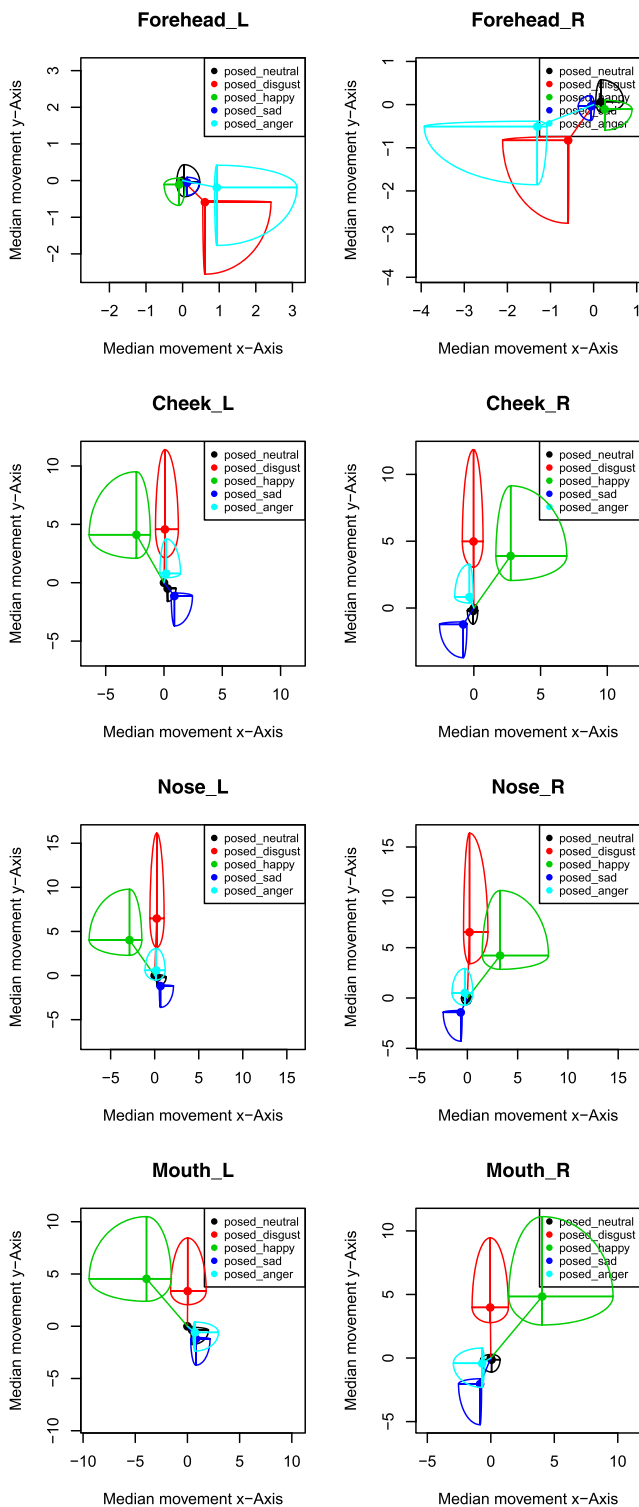
Fig. 8 Plot of the centered median movement in mm of the four symmetrical markers per stimulus episode aggregated over participants in Study 1. The ellipse represents the quartiles of the movement distribution for the *x*- and *y*-axes. The *x*- and *y*-axes of the plots are scaled to the same range so that the angle represents the true direction of movement

procedure for Blender. These properties were tested in a small sample comprising 55 participants and eight emotion markers. In addition, in this study, we tested the required hardware characteristics (e.g., camera, processing performance), illumination settings, and synchronization with the stimulus presentation procedure.

Method

Participants A total of 55 students from different disciplines at the University of Koblenz-Landau, Germany were recruited and received either course credit or were paid for their participation. However, due to improper lighting conditions, incorrectly placed markers, obliterated markers, or improper behavior by the participants (see Table 1), data from only 39 participants (age $M = 22.1$, $SD = 5.78$; 77% female) could be used in the subsequent analyses.

Design and procedure In Study 1, participants' data were collected in individual sessions. At the beginning of each session, black markers were painted on each participant's face. The participants were equipped with a black cap and a black headphone with placed colored glue dots. Four emotion markers were used: Forehead markers were placed at the positions of the left and right inner eye brow, a nasolabial marker was placed left and right beneath the nose, a mouth marker was placed at the left and right corners of the mouth, and left and right cheek markers were placed on the cheeks. These markers were placed at the cheeks to provide a measurement of movement when natural markers are not available.

Based on a script, the record of the video (FFmpeg, ver. 2.2.16) and the presentation of the stimuli was started and stopped simultaneously. The presentation of the stimuli was implemented in Milliseconds Inquisit (ver. 3.0.6.0), whereas each consecutive stimulus was presented with a predefined duration. First, the participants were instructed to direct the tip of the nose to the upper left corner and to the middle of the right border of the computer screen. This parallax phase was implemented for 10 s at each position to obtain some parallax for the facial surface estimation. Subsequently, the participants were asked to show the emotions happiness, sadness, disgust, and anger for 10 s at a time. The instructions read "Please show the emotion happiness" (in the case of happiness), accompanied by an example picture of a facial expression of the corresponding emotion taken from Olszanowski, Pochwatko, Kukliński, Ścibor-Rylski, and Ohme (2008). The instructions were presented below the picture along with a 10 s countdown. Thereafter, additional stimuli were

presented to the participant; however, these were outside the scope of the present study. Finally, the participants were debriefed and were compensated for their participation.

Measures Tracking was done for the complete video clip, including the parallax phase and facial expression episodes. However, due to missing, not visible, or obliterated facial surface markers, we decided not to estimate individual facial surfaces, but to use a standard facial surface mask that was based on averaged faces from preliminary investigations.

Thereafter, the symmetrical four facial emotion markers were tracked and projected onto the standard facial surface mask. Subsequently, the tracked emotion markers were exported into a CSV file, matched with the stimulus presentation episodes, and merged into a single raw data file in R and saved for further processing.

Results

Accuracy of the measurement procedure To estimate the accuracy of the tracking procedure, we calculated the following parameters: (a) Accuracy of measuring head movements and estimating the facial surface: The solve error represents the mean deviation of the tracked markers of the video and the marker positions estimated by the model (and projected onto a 2D surface). This solve error was below 0.3 for all participants, representing a mean deviation between the video-tracked marker position and the model-estimated marker position of maximally one third of a pixel. (b) Accuracy of measurement of the scaling parameters: The measurement of the distances between the eyes, the mouth corners, the left eye–left mouth corner, and the right eye–right mouth corner can be used to compute reliability. The reliability of these four measures was $\alpha = .996$ and also reflected interindividual differences in head proportions (i.e., participants had different mouth widths compared with their distance between their eyes). (c) Accuracy of the model building and rescaling procedure: In a pilot study a paper cuboid of $10 \times 10 \times 20$ cm, roughly representing the area and the size the blenderFace method is intended for, was constructed and equipped with glue dots. For this cuboid, the real-world positions of the glue dots were known. The cuboid was recorded on video, and the blenderFace measurement procedure with the subsequent scaling into mm was performed. The distances that were measured and scaled by Blender differed from the real-world measures by a mean of $M = 0.80$ mm ($SD = 0.54$) with a maximum of 1.82 mm for the 12 distances that were measured. Although the placement of the glue dots was performed very carefully by hand, this measurement also includes a manufacturing bias for the glue dot placement. Altogether, these parameters do not directly estimate the accuracy of the measurement of emotion marker movement but show that the measurement procedure itself is very accurate.

Outliers and artifacts The raw data file was rescaled into mm and also into the standardized face. Both rescaled data sets were centered per stimulus episode (posing happiness, posing sadness, posing disgust, and posing anger). For each stimulus episode, a raw data plot was generated to detect outliers or unusual movements (see Fig. 5).

A precise inspection of these plots[20] revealed potential outliers. A deeper analysis of these marker movements was indicated to rule out errors or artifacts. As an example, this procedure will be shown for the right cheek marker in the "posing disgust" episode; however, it was performed for all markers in all stimulus episodes. The lower left plot of Fig. 5 for the "posing disgust" episode revealed an unusual movement pattern for the right cheek marker to the left direction. This movement did not appear to be common to all participants; however, this could not be reliably determined by plotting the aggregated participants. Therefore, a second plot of individual median[21] movement for the right cheek marker of the disgust episode revealed that Participant number 37 had caused this deviant movement (Fig. 6).

Once the participant, the marker, and the episode for which the outlier or artifact occurred were identified, a very specific inspection of the x- and y-movement of the marker per frame was performed. Figure 7 shows the x- and y-movement of the right cheek marker, along with the left cheek marker for Participant 37 for the frames of the "posing disgust" episode. The deviation in the plot can be interpreted in mm, because the scaled-to-mm dataset was used.

However, the plot revealed no artifacts or outliers but showed an asymmetrical expression of disgust for the cheek markers of Participant 37. Whereas the left and the right y-axis lines (yellow and green) run parallel to a large extent, this was not true for the x-axis (red and blue lines). Note that the origin of the coordinate system is at the tip of the nose, which leads to the fact that on the x-axis in the plot, the markers on the left and the right sides of the face run in opposite directions. Nevertheless, if the x-value of the right cheek marker (red line) were to be mirrored along $y = 0$, the median deviation would be stronger (≈ 6 mm) compared with the x-value of the left cheek marker (blue line, $\approx$ 4 mm). A visual inspection of frames 1,700 to 2,000 for the video clip of the Participant 37 confirmed the assumption of asymmetry in the expression of disgust.

---

[20] Changing the alpha parameter of the `blenderFace` function, which affects the over-plotting density, is very helpful for this purpose.

[21] Because the movement is unlikely to be normally distributed, the median rather than the mean was chosen for aggregation.

For other cases in which artifacts were actually detected, for example, tracks jumping between two positions because the search pattern had a high probability of being fit to two positions in the search area, the artifacts were corrected. This was done in Blender by setting a new pattern for this track and subsequently tracking, exporting, and postprocessing the data. For cases in which a retracking was not possible (e.g., bad illumination conditions, hidden markers), the x-, y-, and z-values of this marker were set to "not available" (NA) in the raw data file in R for the corresponding frames, with the subsequent postprocessing of raw data.

Facial movement in response to the emotional stimuli

Constituting the main outcome of the procedure, we created a combined plot of marker movement per stimulus episode aggregated over participants. Figure 8 shows the markers that were arranged in accordance with their actual facial position, beginning with the left and right forehead markers, the cheek markers, the nasolabial markers, and the mouth corner markers. Data scaled to mm were used, and the x- and the y-axes were scaled to the same range. Therefore, the direction of the movement reflected the true movement of the facial skin. The ellipse around the median point represents the quartiles of the distribution of movement for the x- and the y-axis. In addition to the stimulus episodes in which participants showed happiness, sadness, disgust, and anger, a neutral episode—taken from the parallax phase

of the video clip—was added to reflect measurement noise when no emotional expression was shown. However, the neutral episode reflects not only the measurement error of the blenderFace method but also unintentional movements (e.g., mouth corner movement occurring while swallowing). The relevant parameters for this plot, the angle, and the distance for each stimulus episode per marker, can also be printed (see Table 2).

Discussion

The study was conducted to test the blenderFace method and its border conditions. As a result, the blenderFace method can be implemented with a good, shadow-free illumination on standard hardware (computer, webcam) using only on open-source software. The tracking procedure and therefore the measurement turned out to be very accurate because the mean accuracy for the tracks was below 1 pixel and below 1 mm, respectively. Because the blenderFace method uses markers that are painted on the facial skin, it becomes possible to measure movement in facial areas that lack natural landmarks (e.g., on the cheeks). In the postprocessing of the Blender data, a reliable detection of outliers and artifacts is possible. When the data are corrected, the statistical analyses of movements of the facial skin can be performed. The measured movement corresponds closely to the definition

Table 2 Angle and distance of facial movement in study 1

Stimulus	Foreh_L Angle	Foreh_L Distance	Foreh_R Angle	Foreh_R Distance
Disgust	315.97	0.85	234.49	1.02
Happy	228.11	0.14	337.24	0.28
Sad	335.81	0.13	211.61	0.07
Anger	348.65	0.95	201.26	1.40
	Cheek_L Angle	Cheek_L Distance	Cheek_R Angle	Cheek_R Distance
Disgust	89.02	4.57	90.28	4.98
Happy	119.90	4.75	54.75	4.78
Sad	308.31	1.44	237.08	1.46
Anger	76.83	0.82	111.97	0.89
	Nose_L Angle	Nose_L Distance	Nose_R Angle	Nose_R Distance
Disgust	87.92	6.48	88.13	6.54
Happy	125.37	4.94	52.35	5.33
Sad	299.22	1.36	245.19	1.57
Anger	73.10	0.62	117.18	0.57
	Mouth_L Angle	Mouth_L Distance	Mouth_R Angle	Mouth_R Distance
Disgust	89.64	3.37	90.63	3.98
Happy	130.78	6.01	50.13	6.31
Sad	304.12	1.46	247.59	2.21
Anger	320.87	0.93	211.25	0.80

Note. Angle (values in degree) and median distance (values in mm) of marker movement aggregated over participants per stimulus episode of Study 1. Note, that 0 degree describes a movement to the right, and that on the x-axis left and right markers move in the opposite direction

of the EMFACS/FACSAID (see Ekman & Hager, 2002; Friesen & Ekman, 1983). For example, for the emotion of disgust, the cheek, the nose, and the mouth markers move upwards, which represents the activation of AU 9 (nose wrinkler) and AU 10 (upper lip raiser). For the emotion happiness the cheek-, the nose-, and the mouth markers move upwards and sideways, thus representing the activation of AU 6 (cheek raiser) and AU 12 (lip corner puller).

In a second study, the blenderFace method was tested with a larger number of simultaneously recorded emotion markers, with a larger sample, and by estimating each participant's individual facial surface. In addition, technical improvements that facilitate the synchronization of the stimulus presentation with the timescale of the video clip were checked (e.g., a timestamp and the participant number burnt into the video clip).

Study 2

The aim of Study 2 was to test the full set of proposed markers (see Fig. 2), test the improved experimental settings (e.g., lighting, video compression, etc.), test the improved stimulus-video synchronization markers (e.g., video time stamp branding), and replicate the findings of Study 1.

Because there were more markers to track, it was no longer practical to use marker labels that were based on facial landmarks (e.g., "mouth corner", "inner eye brows"). A labeling scheme based on the FACS (Ekman & Friesen, 1978; Ekman et al., 2002) using Action Units (AUs) as marker labels is also inappropriate, because an AU defines the position of a marker on the face along with the direction of the movement. In contrast, the blenderFace method measures the visible movement of a marker drawn on the facial skin, which may move in virtually any direction. Therefore, we decided to use a straightforward labeling scheme defining only the position of the marker on the face (see left part of Fig. 2): Letters define the *x*-axis position in the sense of a longitude, whereas numbers define the *y*-axis position (i.e., latitude) of a marker for the surface marker mesh. The face is divided vertically by the "A"-axis—going form the center of the forehead, via the tip of the nose to the chin—in two symmetrical parts, which were labeled as "left" and "right" part of the face. Left and right refers to when looking at a face vis-à-vis (not the left and right part of the own face). Starting from the "A"-axis, the subsequent vertical axes are labeled "B", "C", "D", and "E" in the direction to the ears, combined with the label "L" for the left part, and "R" for the right part of the face. This labeling scheme facilitates the comparison of corresponding markers on the two sides of the face. The horizontal grid lines are labeled starting from top to bottom, for example, "1" at the upper forehead, via "5" intersecting the tip of the nose, to "10" at the bottom of the chin. For example, the marker "A5" denotes the tip of the nose, and "CL7" the left corner of the mouth.

Method

Participants One hundred fifty students from different disciplines at the University of Koblenz-Landau, Germany were recruited and received either course credit or were paid for their participation. However, 37 participants were excluded due to illumination problems, inappropriate adjustments of the camera drivers (e.g., autofocus, auto-brightness and contrast), hidden markers, or extreme head movements (see Table 1). A total of 113 participants (age $M = 23.1$, $SD = 2.3$; 82% female) could be used in further analyses.

Design and procedure In Study 2 participant's data were collected in individual sessions as the last part of an experimental sequence that fell outside the scope of the present article. As a cover story, the entire experimental sequence was presented as being about eye tracking. Similar to Study 1, before the experimental sequence began, the participant was equipped with a marker cap and marker headphones. In contrast to Study 1, the participants were painted with the full set of 68 facial markers (see Figs. 1 and 2). Again, the simultaneous beginning and end of the video recording and the stimulus presentation was controlled by a script. For additional video - stimulus presentation synchronization, FFmpeg was used to add the time code, the subject number, and stimulus episode to the video clip. Further on, a mirror in the back of the participants allowed the recording of the presented stimuli, which grants a synchronization on video frame level.

The participants were asked to show a neutral, a disgusted, a happy, and a fearful facial expression for 5 s each. The instructions to show an emotion were given by the sentence "Please show the emotion ..." along with a countdown of 5 s presented at the bottom of the screen. However, we did not present an example picture of the corresponding emotion to prevent participants from imitating the exact same expression that was shown in the picture. In the last part of the experimental episode relevant to the present study, the participants were asked to point their tip of the nose directly into the webcam and to the middle of the right border of the computer monitor for 5 s each. Again, this was done to obtain a parallax for estimating the individual's facial surface.

Thereafter, further stimuli were presented. However, these were outside the scope of the present investigation. Finally, the participants were debriefed and were given course credits or were paid for their participation.

Measures Analogous to Study 1, the tracking of the static head markers was performed for each participant's complete video clip, whereas the 68 surface markers were tracked only for the short parallax episode. Subsequently, the head movements were estimated and the 3D surfaces of the individual faces were built. Finally, the emotion markers were tracked for the emotionally relevant stimuli episodes, projected onto their individual facial surfaces and exported as CSV files. Subsequent postprocessing included adding stimulus presentation information, merging CSV files, standardizing the marker movements into mm and into the standardized face, and centering the marker movement at the onset of each stimulus episode.

Results

Accuracy of the measurement procedure Again, the solve error for all participants was below 0.3, indicating a mean deviation of one third of a pixel per participant between actual tracks and model-estimated tracks. The internal consistency of the measurement of the four scaling distances (eye–eye distance, mouth corner–mouth corner distance, and left/right eye–mouth corner distance) was $\alpha = .985$.

Outliers and artifacts An artifact/outlier analysis was performed for all markers in all stimulus conditions. We asserted that the data did not contain any errors and were suited for further statistical analyses.

Facial movement in response to the emotional stimuli As a main outcome, the median movements per stimulus episode aggregated over participants, along with the quartile ellipses, are presented in Figs. 9 and 10. The data set scaled to mm was used for the plots; therefore, the median movement can be interpreted in mm.

We also included an episode with no facial expression (neutral), to represent measurement noise. However, the overall measurement noise was very low (see black ellipses in Figs. 9 and 10, Table 3: mean distance in mm, irrespective of angle $M = 0.026$, $SD = 0.013$). The marker movement can also be presented with parameters such as angle and distance (Table 3).

Again, the marker movement conformed to the description of movements in EMFACS/FACSAID (Ekman & Hager, 2002; Friesen & Ekman, 1983) of activated AUs for emotional facial expression: For the emotion of happiness AU 6 (cheek raiser) and AU 12 (lip corner puller) should be activated; these are visible in Figs. 9 and 10 for the markers CL4/CR4 along with CL7/CR7 as a sideways and upwards movement (green dot and ellipse). The other markers (e.g., BL5/BR5 in Fig. 10) follow this movement to some extent because the skin is also pulled in the corresponding direction.
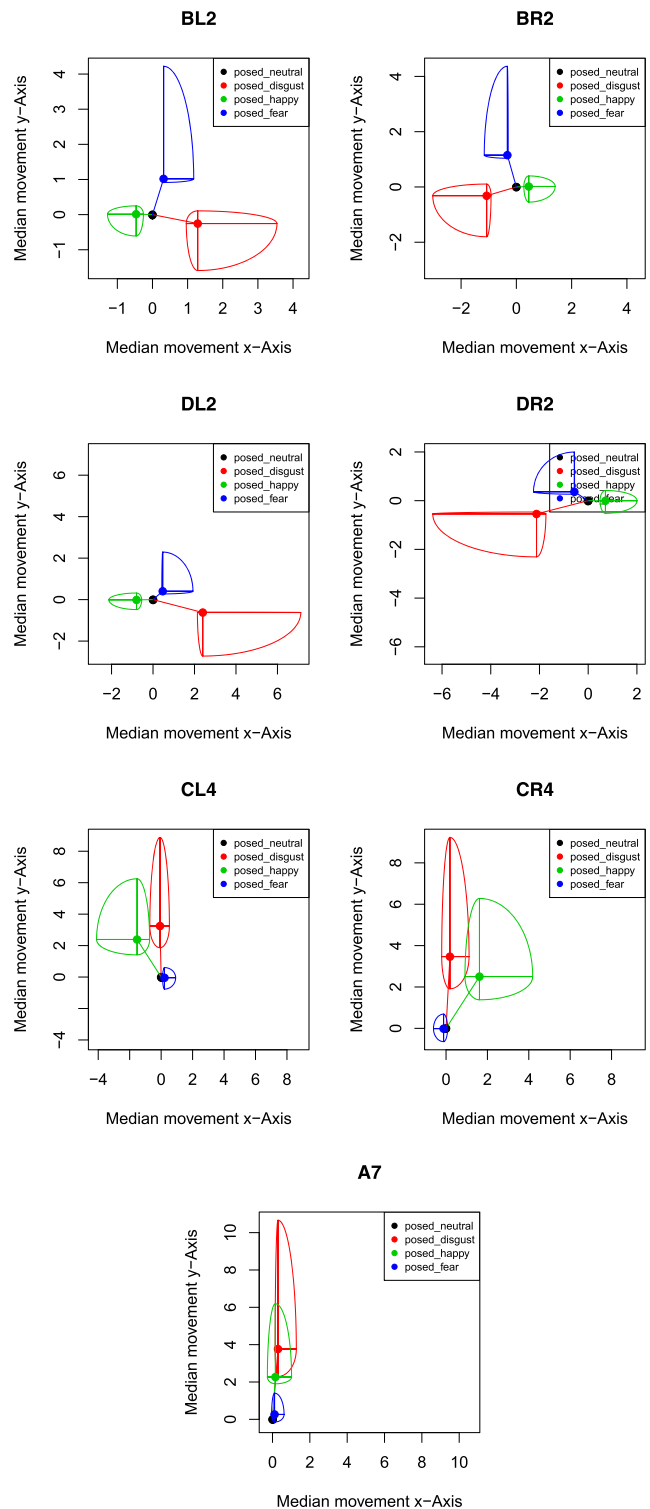

Fig. 9 Plot of the median movement in mm of the first four symmetrical markers per stimulus episode aggregated over participants in Study 2. The ellipse represents the quartiles of the movement distribution for the x- and y-axes. The x- and y-axes of the plots are scaled to the same range so that the angle represents the true direction of movement. The marker positions are shown in the left part of Fig. 2
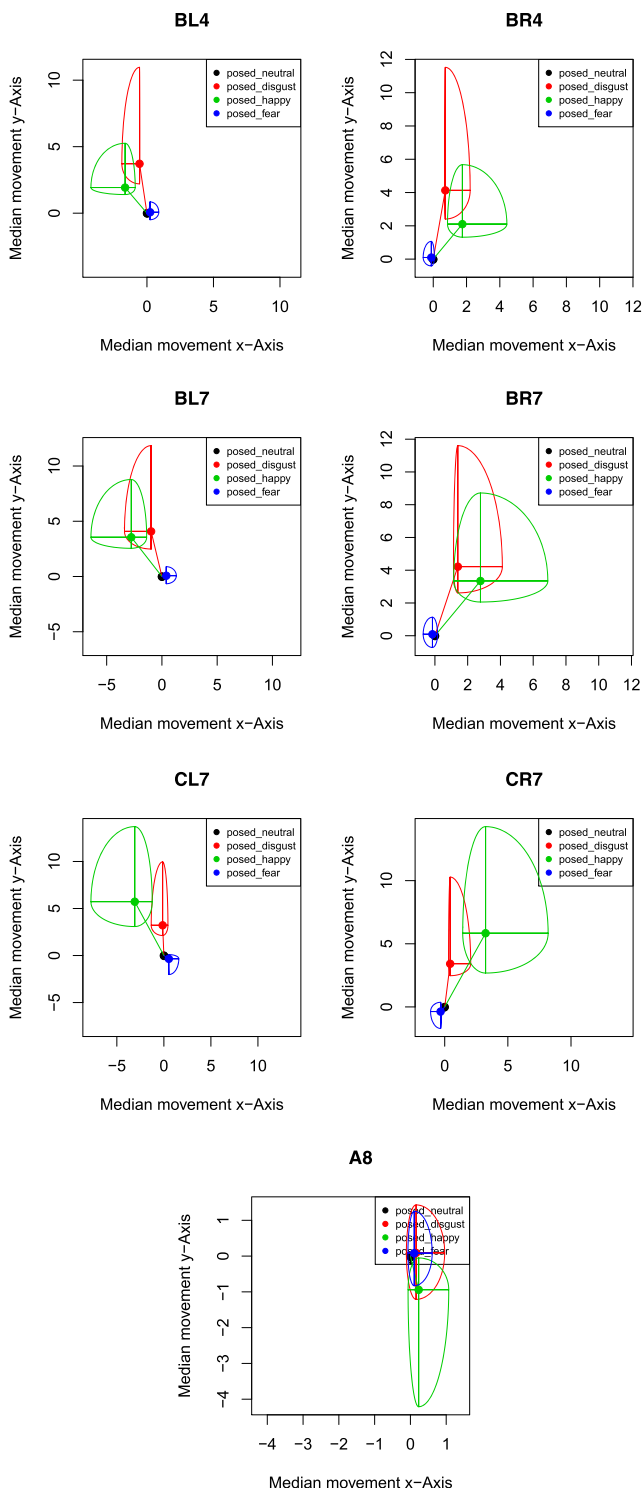
Fig. 10 Plot of the median movement in mm of the second four symmetrical markers per stimulus episode aggregated over participants in Study 2. The *ellipse* represents the quartiles of the movement distribution for the *x*- and *y*-axes. The *x*- and *y*-axes of the plots are scaled to the same range so that the angle represents the true direction of movement. The marker positions are shown in the left part of Fig. 2

For the emotion of disgust the description of activated AUs is heterogeneous. AU 9 (nose wrinkler), AU 10 (upper lip raiser) are very common, but so are pressing the lips together and the mouth corners downwards (AU 15, lip corner depressor, AU 16, lower lip depressor, AU 17, chin raiser), or a slightly opened mouth (AU 25, lips part, AU 26, jaw drop). With the red point and ellipse, Fig. 10 shows the upwards movements in BL4/BR4 and BR7/BL7 but no dragging down of the mouth corners or chin raising in the data when aggregated over participants. However, a tightening of the eyebrows toward the nose root (BL2/BR2 and DL2/DR2) was measured as this is mentioned in the literature (Ekman & Friesen, 1975; Rozin, Lowery, & Ebert, 1994).

For the emotion of fear, mainly the activation of AU 1 (inner brow raiser) and AU 2 (outer brow raiser) and sometimes AU 4 (brow lowerer, leads to a narrowing of the eyebrows) are reported. These movements are visible as blue points and ellipses in Fig. 9 for the markers BL2/BR2 and DL2/DR2 as an upwards and center directed movement. Movements in other seldom reported AUs[22] referring to pressed lips (AU 20, lip stretcher, AU 25, lips part, AU 26 jaw drop) could be measured only to a very small extent.

In summary, the findings of the blenderFace method in Study 2 replicate and extend the findings of Study 1. Again, they closely confirm the EMFACS/FACSAID (Ekman & Hager, 2002; Friesen & Ekman, 1983) assumptions of the activated AUs for emotional facial expression; however, the blenderFace method offers an objective way to measure facial movements.

Discussion

Study 2 showed that it is possible to measure up to 14 emotion markers simultaneously in a sample of 113 participants for 12 min of footage with the blenderFace method. The optical measurement of facial movements with the blenderFace method was very precise, which can be seen by the mean deviation of one third the size of a pixel between video track and model estimated track, an internal consistency of scaling measures close to 1, and a measurement of noise in a neutral condition resulting in a mean movement of 0.026 mm.

---

[22]It was not possible to measure AU 5 (upper lid raiser) or AU 7 (lid tightener) due to frequent eyelid movement, which changes the tracking pattern. These drawbacks are discussed in the Limitations section.

Table 3 Angle and distance of facial movement in study 2

Stimulus	BL2 Angle	BL2 Distance	BR2 Angle	BR2 Distance
Neutral	270.00	0.01	254.96	0.01
Disgust	348.81	1.31	196.44	1.12
Happy	178.45	0.47	2.29	0.45
Fear	72.75	1.07	105.75	1.20
	DL2 Angle	DL2 Distance	DR2 Angle	DR2 Distance
Neutral	282.29	0.02	261.92	0.02
Disgust	345.47	2.48	194.39	2.19
Happy	180.91	0.80	359.84	0.71
Fear	40.99	0.62	147.32	0.67
	CL4 Angle	CL4 Distance	CR4 Angle	CR4 Distance
Neutral	299.42	0.05	242.57	0.02
Disgust	91.25	3.25	86.89	3.47
Happy	122.63	2.84	57.17	2.98
Fear	348.47	0.22	186.43	0.12
	A7 Angle	A7 Distance		
Neutral	309.66	0.02		
Disgust	85.55	3.78		
Happy	86.10	2.27		
Fear	67.31	0.28		
	BL4 Angle	BL4 Distance	BR4 Angle	BR4 Distance
Neutral	313.66	0.03	260.50	0.03
Disgust	98.56	3.76	80.04	4.20
Happy	130.43	2.53	50.26	2.74
Fear	16.85	0.25	140.01	0.15
	BL7 Angle	BL7 Distance	BR7 Angle	BR7 Distance
Neutral	288.49	0.02	299.17	0.02
Disgust	103.48	4.20	71.66	4.44
Happy	127.95	4.51	50.27	4.35
Fear	10.52	0.41	148.31	0.18
	CL7 Angle	CL7 Distance	CR7 Angle	CR7 Distance
Neutral	306.09	0.05	235.35	0.04
Disgust	92.35	3.22	82.80	3.45
Happy	118.40	6.51	61.01	6.68
F ear	326.76	0.63	228.22	0.49
	A8 Angle	A8 Distance		
Neutral	292.30	0.03		
Disgust	29.64	0.18		
Happy	283.79	0.97		
Fear	37.82	0.13		

Note. Angle (values in degree) and median distance (values in millimeter) of marker movement aggregated over participants per stimulus episode of Study 2. Note, that 0 degree describes a movement to the right, and that on the *x*-axis left and right markers move in the opposite direction. The marker positions are shown in the left part of Fig. 2

The measured movements for the 14 emotion markers corresponded closely to the assumed AUs of the EMFACS/FACSAID (Friesen & Ekman, 1983; Ekman & Hager, 2002). The practical improvements in Study 2, such as the illumination or the time code in the video clip, strongly facilitated the synchronization of the tracking procedure and the stimulus presentation. Also the use of an individual facial surface improved the measurement accuracy as can be seen by the decrease in marker movement noise in the neutral stimulus condition in Study 2 compared with Study 1, where no individual facial masks were used.

General discussion

In this paper, we presented the blenderFace method, which circumvents the drawbacks of existing methods that are used to measure facial expressions because it offers a very accurate (sub second, sub pixel, and sub millimeter range), non intrusive simultaneous raw data measurement of several emotionally relevant markers at a high temporal resolution. Study 2 showed that it is possible to simultaneously measure the temporal dynamics of up to 14 markers at a timely resolution of 33 ms for approximately 12 min of video footage for each of 113 participants. Due to the open-source approach, the blenderFace method is very versatile and transparent at every step of data processing and in line with the desiderata of reproducible research (e.g., Fomel & Claerbout, 2009). Although we presented only descriptive findings of the blenderFace method in this paper, they correspond with findings from established measurement instruments (EMFACS, FACSAID; Ekman & Hager, 2002; Friesen & Ekman, 1983). In addition, there are many possible research applications for this method: (a) The raw data generated by the blenderFace method can be analyzed with inferential statistical methods to test various hypotheses. This can include, for example, the identification of parameters to score emotional facial expression, similar to what Olderbak, Hildebrandt, Pinkpank, Sommer, and Wilhelm (2013) did for AFER output. With the blenderFace method, this can be investigated more efficiently than with FACS, in contrast to EMG measurement it easily can be performed for facial expressions involving the whole face, and also without the assumption of basic emotions as implied by AFER algorithms, whereas the existence of basic emotions is still in scientific discourse (Gendron & Barrett, 2017). Another way to analyze raw data obtained with blenderFace could be to investigate the temporal dynamics of facial emotion expression. For example, it may be possible to identify movement parameters that allow a distinction between a posed and a spontaneous facial expression. Although very good classifiers exist for this distinction (e.g., in the case of pain; Littlewort et al., 2009), these classifiers do not allow researchers to test for specific parameters that can be applied to make this distinction. For example, distinction parameters that have been suggested in the literature are pauses, stepwise intensity changes, several onset-, apex-, and offset-phases of specific facial markers (Dubois et al., 2013; Hess & Kleck, 1990, 1994; Schmidt, Ambadar, Cohn, & Reed, 2006; Schmidt, Bhattacharya, & Denlinger, 2009; Weiss, Blum, & Gleberman, 1987). Therefore, contrary to AFER classifiers, which do a very good job at classifying, the blenderFace method may lead to gains in scientific knowledge in this area of research. A third way to use raw data from blenderFace raw could be, for example, the

analysis of microexpressions (Ekman & Friesen, 1969). Although the concept of microexpressions has existed for several years, it has not yet been extensively investigated (30 hits in PsycARTICLES and PsycINFO databases; May, 2018). (b) A novel and unique feature of the proposed blenderFace method is the possibility of superimposing the standardized measurements of several respondents as they react to an experimental stimulus condition. With this, it may be possible to identify common characteristics of facial expressions and individual deviations (e.g., asymmetries or atypical reactions) in an objective manner. This opens a large field of further research options, for example, the identification of facial reaction patterns without the restriction of a priori hypotheses such as basic emotions. (c) More generally, the blenderFace method can also be used as an optical method for measuring non-verbal behavior. For example, the head movement and head turn data that are collected in the blenderFace measurement procedure (to be disentangled from facial movements), have not yet been taken into account for data evaluation. However, head movements and head turns may be used as indicators of approach-avoidance behavior. For example, the speed with and the extent to which the face turns away from an aversive stimulus can be measured and used as an indicator. Provided the appropriate markers are used, other non verbal behavior can also be measured. For example, with markers on the hands and arms, speech-supporting illustrators could be recorded. With markers on the body (e.g., on the shoulders), the body movements during a dyadic interaction could be measured. (d) The video material also contains additional information that could be evaluated in combination with the blenderFace method. For example, it would be possible to derive biophysiological indicators such as pulse rate and increased blood flow to the skin by increasing the color intensity (Wu, Rubinstein, Shih, Guttag, Durand, & Freeman, 2012) and to combine them with participants' facial expression.

However, the proposed blenderFace method is also subject to limitations: (a) The use of markers in the blenderFace method is relatively complex, not so much in comparison with FACS rating or EMG measurement, but in comparison with AFER procedures. However, it is precisely this use of markers painted on participants' facial skin that allows the accurate measurement of facial movements. (b) At the moment, facial movement is measured in three dimensions but analyzed in a two-dimensional way. One reason to do this is data reduction. As outlined before, this can be done with reasonable costs in measurement precision. Another reason is that to transform a face into a standardized face, no suitable scaling object for the z-axis exists (e.g., there is a great deal of interindividual variability in size / height of the nose independent of head size). However, provided there is a suitable scaling object, a

three-dimensional processing is possible. (c) Markers next to the eyelids (e.g., AU 5 or AU 7) are currently very hard to track because eye-blinking constantly changes the tracking pattern and requires a lot of manual corrections. For this purpose, other techniques (e.g., a robust ellipsoid fitting), are more suitable. Due to its openness, the method can be adapted to various experimental conditions and can also be improved on the technical side. For example, in combination with the openCV library (www.opencv.org), the functionality of Blender may be enhanced to find and track a given set of markers automatically. Furthermore, a second camera could be used to get a stereoscopic view, which in turn can be used to compute a 3D surface of the face automatically. In this sense of openness, we hope the blenderFace method will be used, adapted, and improved.

References

Affectiva (2017). Affectiva. Retrieved from: www.affectiva.com

Alabort-i-Medina, J., Antonakos, E., Booth, J., Snape, P., & Zafeiriou, S. (2014). Menpo: A comprehensive platform for parametric image alignment and visual deformable models. In *Proceedings of the ACM international conference on multimedia* (pp. 679–682). New York: ACM.

Baltrušaitis, T., Robinson, P., & Morency, L.-P. (2016). OpenFace: An open-source facial behavior analysis toolkit. In *IEEE winter conference on applications of computer vision.* https://github.com/TadasBaltrusaitis/OpenFace

Bartlett, M., Littlewort, G., Wu, T., & Movellan, J. (2008). Computer expression recognition toolbox. In *2008 8th IEEE international conference on automatic face & gesture recognition* (pp. 1–2). Institute of Electrical and Electronics Engineers.

Cohn, J. F., & Schmidt, K. L. (2004). The timing of facial motion in posed and spontaneous smiles. *International Journal of Wavelets, Multiresolution and Information Processing, 2*, 1–12.

Corneanu, C. A., Oliu, M., Cohn, J. F., & Escalera, S. (2016). Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 38*(8), 1548–1568.

Crivelli, C., Jarillo, S., Russell, J. A., & Fernández-Dols, J.-M. (2016a). Reading emotions from faces in two indigenous societies. *Journal of Experimental Psychology: General*, *145*(7), 830–843.

Crivelli, C., Russell, J. A., Jarillo, S., & Fernández-Dols, J.-M. (2016b). Recognizing spontaneous facial expressions of emotion in a small-scale society of Papua New Guinea. Emotion.

De la Torre, F., Chu, W. S., Xiong, X., Vicente, F., Ding, X., & Cohn, J. F. (2015). IntraFace. In *IEEE international conference on automatic face and gesture recognition (FG).* Ljubljana, Slovenia.

Dimberg, U., Thunberg, M., & Grunedal, S. (2002). Facial reactions to emotional stimuli: Automatically controlled emotional responses. *Cognition & Emotion*, *16*(4), 449–471.

Dubois, M., Dupre, D., Adam, J.-M., Tcherkassof, A., Mandran, N., & Meillon, B. (2013). The influence of facial interface design on dynamic emotional recognition. *Journal on Multimodal User Interfaces*, *7*, 111–119.

Ekman, P., & Friesen, W. (1978). *Facial action coding system: A technique for the measurement of facial movement*. Palo Alto: Consulting Psychologists Press.

Ekman, P., Friesen, W., & Hager, J. (2002). *Facial action coding system. The manual on CD-Rom*. Salt Lake City: Network Information Research Corporation.

Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry*, *32*, 88–105.

Ekman, P., & Hager, J. (2002). Facial action coding system affect interpretation database (FACSAID). http://face-and-emotion.com/dataface/facsaid/description.jsp

Ekman, P., & Friesen, W. V. (1975). *Unmasking the face: A guide to recognizing emotions from facial cues*. New York: Englewood Cliffs.

Elfenbein, H. A., Beaupré, M., Lévesque, M., & Hess, U. (2007). Toward a dialect theory: Cultural differences in the expression and recognition of posed facial expressions. *Emotion*, *7*(1), 131–146.

Emotient Inc. (2016). Facet. Retrieved from: www.emotient.com

Evtimov, I., Eykholt, K., Fernandes, E., Kohno, T., Li, B., Prakash, A., & Song, D. (2017). Robust physical-world attacks on machine learning models. In *Cryptography and security*.

Fomel, S., & Claerbout, J. (2009). Guest Editors' Introduction: Reproducible research. *Computing in Science and Engineering*, *11*, 5–7.

Fridlund, A. J., & Cacioppo, J. T. (1986). Guidelines for human electromyographic research. *Psychophysiology*, *23*, 567–589.

Friesen, W., & Ekman, P. (1983). *EMFACS-7: Emotional facial action coding system*. California: University of California.

Gendron, M., & Barrett, L. F. (2017). Facing the past. In J.-M. Fernández-Dols, & J. A. Russell (Eds.) *The science of facial expression,* Chap. 2. New York: Oxford University Press.

Ham, C., Lucey, S., & Singh, S. (2014). Hand waving away scale. In *European conference on computer vision*. Springer International Publishing.

Hess, U., & Kleck, R. E. (1990). Differentiating emotion elicited and deliberate emotional facial expressions. *European Journal of Social Psychology*, *20*(5), 369–385.

Hess, U., & Kleck, R. E. (1994). The cues decoders use in attempting to differentiate emotion-elicited and posed facial expressions. *European Journal of Social Psychology*, *24*(3), 367–381.

Hochberg, Y., & Tamhane, A. C. (1987). *Multiple comparison procedures*. Hoboken: Wiley.

Horstmann, G. (2002). Facial expressions of emotion: Does the prototype represent central tendency, frequency of instantiation, or an ideal? *Emotion*, *2*(3), 297–305.

Izard, C. E. (1983). *Maximally discriminative facial movement coding system (MAX)*. Newark: University of Delaware. Instructional Resources Center.

Izard, C. E., Dougherty, L. M., & Hembree, E. A. (1983). *A system for affect expression identification by holistic judgements (AFFEX)*. Newark: University of Delaware. Instructional Resources Center.

Jack, R. E., Garrod, O. G. B., & Schyns, P. G. (2014). Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Current Biology*, *24*, 187–192.

Jack, R. E., Garrod, O. G. B., Yu, H., Caldara, R., & Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(19), 7241–7244.

Jeni, L. A., Cohn, J. F., & Kanade, T. (2015). Dense 3D face alignment from 2D videos in real-time. In *11th IEEE international conference and workshops on automatic face and gesture recognition (FG)* (Vol. 1). IEEE.

Kairos, A. R. (2017). Inc. Kairos. Retrieved from: www.kairos.com

Littlewort, G., Bartlett, M. S., & Lee, K. (2009). Automatic coding of facial expressions displayed during posed and genuine pain. *Image and Vision Computing*, 27(12), 1741–1844.

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*. Institute of Electrical and Electronics Engineers (IEEE).

Motley, M. T., & Camden, C. T. (1988). Facial expression of emotion: A comparison of posed expressions versus spontaneous expressions in an interpersonal communication setting. *Western Journal of Speech Communication*, 52(1), 1–22.

Noback, M. L., Harvati, K., & Spoor, F. (2001). Climate-related variation of the human nasal cavity. *American Journal of Physical Anthropology*, 599–614.

Noldus Information Technology (2017). FaceReader. Retrieved from: www.noldus.com

Olderbak, S., Hildebrandt, A., Pinkpank, T., Sommer, W., & Wilhelm, O. (2013). Psychometric challenges and proposed solutions when scoring facial emotion expression codes. *Behavior Research Methods*, 46(4), 992–1006.

Olszanowski, M., Pochwatko, G., Kukliński, K., Ścibor-Rylski, M., & Ohme, R. (2008). Warsaw set of emotional facial expression pictures - validation study. EAESP General Meeting, Opatija, Croatia. http://emotional-face.org

Porter, S., & ten Brinke, L. (2008). Reading between the lies: Identifying concealed and falsified emotions in universal facial expressions. *Psychological Science*, 19(5), 508–514.

Porter, S., & ten Brinke, L. (2010). The truth about lies: What works in detecting high-stakes deception? *Legal and Criminological Psychology*, 15(1), 57–75.

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

RealEyes (2017). RealEyes. Retrieved from: www.realeyesit.com

Reisenzein, R., Studtmann, M., & Horstmann, G. (2013). Coherence between emotion and facial expression: Evidence from laboratory experiments. *Emotion Review*, 5(1), 16–23.

Rozin, P., Lowery, L., & Ebert, R. (1994). Varieties of disgust faces and the structure of disgust. *Journal of Personality and Social Psychology*, 66(5), 870–881.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expressions? A review of the cross-cultural studies. *Psychological Bulletin*, 115(1), 102–141.

Scherer, K. (2005). What are emotions? And how can they be masured? *Social Science Information*, 44(4), 693–727. https://doi.org/10.1177/0539018405058216

Schmidt, K. L., Bhattacharya, S., & Denlinger, R. (2009). Comparison of deliberate and spontaneous facial movement in smiles and eyebrow raises. *Journal of Nonverbal Behavior*, 33(1), 35–45.

Schmidt, K. L., Ambadar, Z., Cohn, J. F., & Reed, L. I. (2006). Movement differences between deliberate and spontaneous facial expressions: Zygomaticus major action in smiling. *Journal of Nonverbal Behavior*, 30(1), 37–52.

Shannon, C. E. (1949). Communication in the presence of noise. *Proceedings of the Institute of Radio Engineers*, 37(2), 10–21.

Sharif, M., Bhagavatula, S., Bauer, L., & Reiter, M. K. (2016). Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition. In *CCS 2016 - Proceedings of the 2016 ACM SIGSAC conference on computer and communications security* (pp. 1528–1540). Association for Computing Machinery.

Suwajanakorn, S., Kemelmacher-Shlizerman, I., & Seitz, S. M. (2014). Total moving face reconstruction. In *European conference on computer vision*. Springer International Publishing.

Tassinary, L. G., Cacioppo, J. T., & Vanman, E. J. (2007). The skeletomotor system: Surface electromyography. In J. Cacioppo, L. Tassinary, & G. Berntson (Eds.), *Handbook of psychophysiology*, 3rd Edn. (pp. 267–299). Cambridge: Cambridge University Press.

Tcherkassof, A., Bollon, T., Dubois, M., Pansu, P., & Adam, J.-M. (2007). Facial expressions of emotions: A methodological contribution to the study of spontaneous and dynamic emotional faces. *European Journal of Social Psychology*, 37(6), 1325–1345.

Terzis, V., Moridis, C. N., & Economides, A. A. (2010). Measuring instant emotions during a self-assessment test: The use of FaceReader. In A. J. Spink, F. Grieco, O. E. Krips, L. W. S. Loijens, L. P. J. J. Noldus, & P. H. Zimmerman (Eds.) *Proceedings of measuring behavior* (pp. 192–195). Eindhoven: Netherlands.

Weiss, F., Blum, G. S., & Gleberman, L. (1987). Anatomically based measurement of facial expressions in simulated versus hypnotically induced affect. *Motivation and Emotion*, 11(1), 67–81.

Wood, E., Baltrušaitis, T., Morency, L.-P., Robinson, P., & Bulling, A. (2016). A 3D morphable eye region model for gaze estimation. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.) *Computer vision – ECCV 2016: 14th European conference, proceedings part I. (October 11, 2016)* (pp. 297–313). Cham: Springer.

Wu, H.-Y., Rubinstein, M., Shih, E., Guttag, J., Durand, F., & Freeman, W. T. (2012). Eulerian video magnification for revealing subtle changes in the world. *ACM Transactions on Graphics (Proc. SIGGRAPH 2012), 31*(4).

Yan, W.-J., Wu, Q., Liang, J., Chen, Y.-H., & Fu, X. (2013). How fast are the leaked facial expressions: The duration of micro-expressions. *Journal of Nonverbal Behavior*, 37(4), 217–230.