# Consistent behavioral and electrophysiological evidence for rapid perceptual discrimination among the six human basic facial expressions

Qiuling Luo [1,2] · Milena Dzhelyova [2]

## Abstract

The extent to which the six basic human facial expressions perceptually differ from one another remains controversial. For instance, despite the importance of rapidly decoding fearful faces, this expression often is confused with other expressions, such as Surprise in explicit behavioral categorization tasks. We quantified implicit visual discrimination among rapidly presented facial expressions with an oddball periodic visual stimulation approach combined with electroencephalography (EEG), testing for the relationship with behavioral explicit measures of facial emotion discrimination. We report robust facial expression discrimination responses bilaterally over the occipito-temporal cortex for each pairwise expression change. While fearful faces presented as repeated stimuli led to the smallest deviant responses from all other basic expressions, deviant fearful faces were well discriminated overall and to a larger extent than expressions of Sadness and Anger. Expressions of Happiness did not differ quantitatively as much in EEG as for behavioral subjective judgments, suggesting that the clear dissociation between happy and other expressions, typically observed in behavioral studies, reflects higher-order processes. However, this expression differed from all others in terms of scalp topography, pointing to a qualitative rather than quantitative difference. Despite this difference, overall, we report for the first time a tight relationship of the similarity matrices across facial expressions obtained for implicit EEG responses and behavioral explicit measures collected under the same temporal constraints, paving the way for new approaches of understanding facial expression discrimination in developmental, intercultural, and clinical populations.

**Keywords** Facial expressions · Perceptual similarity · EEG · FPVS

## Introduction

Facial expressions are important communicative cues, guiding behavior and social interactions. To be able to reliably and quickly interpret these signals, humans have evolved complex systems to successfully categorize facial expressions, i.e., visually discriminate among different expressions and generalize these differences across other facial variations, for

✉ Milena Dzhelyova
dzhelyova@gmail.com

1 Department of Psychology, Zhejiang Normal University, Jinhua, China

2 Institut de Recherche en Sciences Psychologiques (IPSY), Université catholique de Louvain (UCL), Place du Cardinal Mercier, 10, B-1348 Louvain-la-Neuve, Belgium

instance, face identity. In humans, since the pioneering work of Ekman and colleagues, six categories of facial expressions (Fear, Anger, Disgust, Happiness, Sadness, Surprise), also known as the "basic emotions" (Ekman & Friesen, 1971; Ekman, 1993), have been suggested to be effectively transmitted and decoded (Smith et al., 2005), and thus universally recognized (Matsumoto, 2001; Izard, 1994; Elfenbein & Ambady, 2002 for meta-analysis; but see also Jack, Sun, Delis, Garrod, & Schyns, 2016, revealing only four universally recognized expressions).

One of the most common methods for measuring human expression categorization is by asking participants to select a given label for a static displayed facial picture (Calvo & Lundqvist, 2008; Ekman, 1993; Palermo & Colheart, 2004; Russell, 1993, 1994; Tottenham et al., 2009). A common finding of these studies is that these six expressions are recognized well above chance level and many of them even above 70% (Calvo & Lundqvist, 2008; Ekman, 1993; Palermo & Colheart, 2004; Russell, 1993, 1994; Tottenham et al., 2009). Fear is generally the least accurate and the

slowest expression identified in explicit behavioral categorization tasks (e.g., for a review Calvo & Lundqvist, 2008), while the categorization of Happiness outperforms all other expressions. Yet, not all expressions are as easily recognized as Happiness: Fearful facial expression often is misjudged as Surprise. Similarly, disgusted faces often are perceived as angry and vice versa. This confusion among the facial expressions could be partly due to overlapping signals used to transmit them, thus making some expressions perceptually more similar to one another. For example, for both Fearful and Surprised expressions, the upper lids of the eyes are widely pulled and the mouth often is open. A common signal for Disgust and Anger is the wrinkling around the nose.

Often this simplified method of studying facial expression categorization has been criticized (Barrett, Adolphs, Marsella, Martinez, & Pollak, 2019; Russell, 1993). For instance, participants' performance varies depending on the task format. Providing the participants with a label to choose from can increase the recognition performance compared with a free labeling task (for a review Russell, 1993). In particular circumstances, it also can lead to high agreement rates despite the presented expression not being correctly labeled (Russell, 1993). For example, in a forced-choice experiment, when participants were shown an angry expression they categorize it as Anger on average 12.5%, and more often as Frustration (on average 40%) or Determination (on average 31.7%) (Russell, 1993). Furthermore, only modest correlations have been found between a perceptual matching task in which a participant has to leave an odd expression and an explicit labeling task (Palermo et al., 2013), suggesting at least partial independence of labeling and discrimination of facial expressions.

Finally, when facial expressions are presented briefly and backward masked, precluding further processing of the stimulus (Maxwell & Davidson, 2004, Milder et al., 2008, Neath & Itier, 2014), a different pattern of observations emerges. In these conditions, Happiness is still the expression that is categorized the fastest, but the differences found in the free-viewing literature between the other facial expressions are not observed (Milders et al., 2008). In fact, in some cases, the categorization of fearful faces is performed better than the categorization of angry faces (Maxwell & Davidson, 2004) and can be as good as the categorization of happy faces.

The short review of the literature above indicates that an important and unresolved issue is the extent to which these laboratory observations obtained during explicit categorization tasks truly reflect *visual* discrimination among facial expressions. Indeed, the findings of rapidly presented facial expressions suggest that performance at facial categorization in explicit behavioral tasks when images are presented for fairly long periods of time might not only be influenced by visual similarities between facial expressions but also by higher-order conceptual or decisional factors. For instance, a happy

expression may not be particularly distinctive in terms of physical information, but because it is the only clearly positive expression, it may be systematically judged as being more distinctive than others in explicit behavioral tasks, especially when observers can make their judgments over several seconds. Alternatively, most of the basic emotions are negative, their categorization might be more difficult as the perceiver needs to exclude more options as a possible answer, thus making Happiness categorization the easiest.

To clarify the extent to which the facial expressions of the basic emotions are different from one another at the perceptual level, researchers have asked human observers to rate pairs of expressions on similarity (Adolphs, 1999; Hamann & Adolphs, 1999). Participants tend to perceive as more similar expressions that are often confused with each other, such as Fear and Surprise; Disgust and Anger; Sadness and Neutral (e.g., the control group in Adolphs, 1999). Recent studies have further advanced this topic by comparing responses within face-selective areas and the perceived similarity among the expressions by using functional magnetic resonance imaging (fMRI) to estimate the similarity structure of the neural representation of the basic emotions with multivariate pattern analysis (MVPA; Haxby et al., 2011; Kriegeskorte et al., 2008). The results of these studies have suggested that pairs of expressions that are categorized as more similar to one another show similar neural patterns within face-selective temporal cortices (Said et al., 2010; Saarimäki et al., 2015; Sormaz et al., 2016). The resulting classification accuracies are frequently interpreted as evidence for a discrete signature of the particular expression (Saarimäki et al., 2015; Said et al., 2010). However, classification performance—i.e., discrimination of facial expressions—in such studies is often very low (slightly significant above chance, where chance level could be as low as 16.66% when all six basic expressions are considered) and is distributed across several brain structures, challenging the interpretation that these neural patterns are fingerprints of discrete representations of facial expressions (Clark-Polner et al., 2017). More generally, the slow accumulation of indirect neural activity in the visual cortex as measured in fMRI can be influenced by feedback from higher-order regions and is thus not well suited for isolating responses reflecting visual discrimination.

In light of the above considerations, it should be clear that the extent to which the six basic human facial expressions differ perceptually remains unknown. To shed light on this issue, we quantified humans' discrimination of the six basic facial expressions under tight temporal constraints with an implicit and highly sensitive electrophysiological measure. In the main experiment of our study, participants were first presented with a given facial identity expressing an emotion (e.g., a happy face) at a rapid rate of 12 Hz (i.e., 83.33 ms stimulus onset asynchrony).

This stimulation sequence was presented while directly recording neural activity with high-density scalp electro-encephalography (EEG) to isolate perceptual facial expression discrimination without a potentially confounding explicit facial expression discrimination task. Critically, changes in facial expression (e.g., from happy to sad) were inserted periodically—here every nine images (1.33 Hz). In these conditions of fast periodic visual stimulation (FPVS), the brain synchronizes its activity to this presentation rate, generating neural responses at 12 Hz and 1.33 Hz and their harmonics that can be observed in the EEG spectrum exactly at these frequency rates (Regan, 1966, 1989; Norcia et al., 2015; Dzhelyova et al., 2017). Critically though, the 1.33 Hz response is observed only if the two expressions are discriminated from one another. This response can then be objectively (i.e., at predefined frequencies) identified and quantified in the EEG frequency domain. Hence, we are in a position to isolate and to quantify the contrast between basic facial expressions in the context of brief changes occurring in a continuous perceptual flow, without any explicit task. Using such a paradigm previously validated by contrasting neutral to expressive faces (Dzhelyova et al., 2017; see also Leleu et al., 2018), we contrasted the changes between the 6 universal facial expressions against each other (e.g., Disgust to Fear, etc.), i.e., 30 two-by-two discriminations while recording EEG. In doing that, we were able to objectively measure visual discrimination between facial expressions and evaluate if this discrimination is context-dependent (i.e., contingent on the expression presented as a common face). We expected that facial expressions that are often misjudged as others due to perceptual similarity would lead to a weaker discrimination response than a contrast between perceptually distinct expressions. In addition, we compared the similarities of the discrimination response obtained with FPVS-EEG and the perceptual similarities as measured by explicit behavioral tasks. In comparing these two measures, we were able to assess perceptual and neural similarities among brief changes in facial expressions.

## Methods

### Participants

Fifteen right-handed volunteers (9 males, 22.87 ± 2.774 years) participated in the study. They had no prior history of neurological or psychiatric problems and had a normal or corrected-to-normal vision. All participants provided signed and informed consent and were paid an amount according to their testing time. The study was approved by the Biomedical Ethical Committee of the University of Louvain.

### Stimuli

Sixteen identities (8 males) from the Karolinska database (Lundqvist et al., 1998), displaying the six basic emotions (Anger, Disgust, Fear, Happiness, Sadness, and Surprise) were used as stimuli. The faces were in a frontal view with forward eye gaze. They were colored photographs with global luminance equalized during the presentation. The image size was set to 4.81° (width) x 7.19° (height) with external features, such as ears and hair, placed against a grey background (RGB: 153, 153, 153, examples in Fig. 1). To estimate the image similarities, a physical dissimilarity index for each expression contrast was calculated. Expressive faces were first converted to greyscale images. Then, pixel-wise correlations between every two expressive faces were computed and averaged across individual faces. By subtracting these correlations from 1 and multiplying by 100, a physical dissimilarity index was obtained for each expression contrast (Leleu et al., 2018). The results for this image similarity measure can be found in the supplementary material: Pixel Similarity Measure.

### Procedure

The experiment consisted of two separate sessions: an EEG session followed by a behavioral session, conducted a week apart for all participants. The EEG session was always recorded first, because having the participant performed expression related tasks might influence their performance during the EEG recording session. The behavioral session is reported first.

#### Behavioral session

Participants were comfortably seated at 1m distance to the screen and performed three explicit behavioral tasks: a forced-choice face-labeling task, an expression similarity rating task, and an expression change detection task. The forced-choice face labeling and expression similarity rating tasks were similar to the tasks used in previous studies, which generally demonstrated the lowest discrimination and categorization performance for fearful faces (Calvo & Lundqvist, 2008; Ekman, 1993; Palermo & Coltheart, 2004; Russell, 1993, 1994; Tottenham, et al., 2009). The order of the tasks was counterbalanced across participants. Participants performed the tasks at their own pace and were allowed to take breaks between them. The duration of the whole behavioral session lasted approximately 1.5 hours.

**Forced-choice face labeling task** All facial expression stimuli (16 identities x 6 basic expressions) were randomly presented four times, resulting in 384 trials in total. Repetitions were used to guarantee data reliability and as the performance did
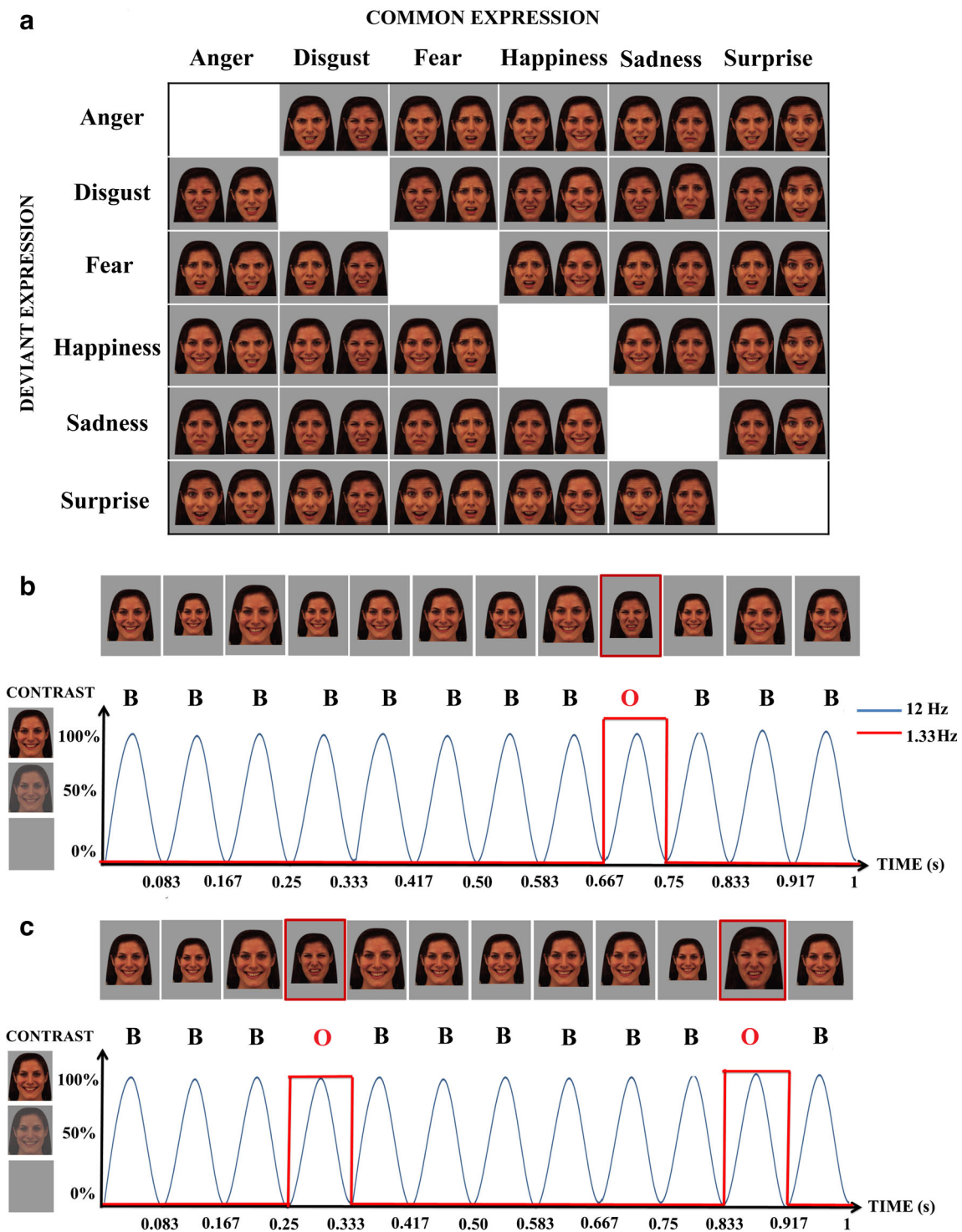
**Fig. 1** Stimuli examples and experimental design of the FPVS experiments. **A)** The combination of all facial expressions, resulting in 30 pairwise contrasts. Note that the same expressions (e.g., Happy and Happy, hence the blank grids) are not contrasted in the EEG design. **B)** Illustration of the fast periodic visual stimulation (FPVS) paradigm during EEG recording. Facial images of one identity with one of the basic expressions (e.g., Happiness - common emotional face) are sinusoidally presented at 12 Hz, reaching full contrast halfway through the cycle. Every ninth image is the same identity but with a different expression (e.g., Disgust – deviant emotional face or change of expression to disgust). The trial duration is 54 sec, including a sequence of 50 sec and 2 sec of gradual fading in and fading out of the images. **C)** Expression change detection task in FPVS. An expressive face (e.g., happy) is presented at 12 Hz (~ 83 ms). Fifteen times during the 54-sec trials pseudo-randomly separated at least 2-sec apart, the same individual appears with a different expression (e.g., disgusted). The participants' task is to detect changes in facial expressions

not differ among them, $F(3, 285) = 0.836$, $p = 0.462$, data were averaged across the repetitions. In each trial, one facial image was centrally presented on the screen along with six emotion labels (1. Happiness; 2. Fear; 3. Surprise; 4. Sadness; 5. Disgust; 6. Anger), appearing below the facial stimulus. Participants were instructed to choose the label that best matched the given facial expression by pressing the corresponding number on the computer keyboard. Each display was presented until the participants responded. Once the response was given, a blank screen was shown for a random period, varying between 50 to 150 ms.

**Expression similarity rating task** Each trial started with a black fixation cross, centrally shown for a period varying between 50 and 150 ms, followed by a pair of facial stimuli. The pair of stimuli was made of two different identities, displaying either the same or different expressions, presented on each side of the fixation cross. A 7-point scale was shown below the faces (1 = identical, and 7 = extremely different), on which the participants rated the similarity of the given facial expressions. Participants were explicitly instructed to provide their ratings not based on the identity of the faces, which differed at every trial, but solely based on the emotional expression. All possible combinations, including the combinations of the same expressions, were examined in this task, thus resulting in 36 expression contrasts. The gender of facial stimuli was counterbalanced by presenting the following four gender combinations for each emotion contrast: a male with a male; a female with a female; a male with a female; a female with a male. Moreover, each combination of an emotional contrast and gender combination was repeated twice, resulting in 288 trials in total: 36 expression contrasts x 4 gender combinations x 2 repetitions. The performance on the two repetitions was not significantly different from each other for all emotion conditions across the 15 participants, $t(143) = 1.453$, $p = 0.149$, indicating consistency of the data between the first and the second repetition.

**Expression change detection task** A last behavioral task designed based on the EEG experiment also was implemented. In this task, for each participant, the same facial stimuli as in the EEG session were presented. In a given sequence, an individual displaying one of the facial expressions (e.g., fearful face) from the 30 pairwise expression contrasts was repeatedly presented at a rapid 12 Hz rate by means of sinusoidal contrast modulation (Dzhelyova et al., 2017, Experiment 2) for 54 seconds, including 2 seconds of fade in and fade out. The image of the same individual but with another facial expression (i.e., Happiness) appeared at nonperiodic intervals 15 times in a sequence. The distance between the changes to a different expressive face varied between 2 and 6 sec. All participants were instructed to look at the fixation cross and to press the space key as soon as they detect the change in emotional expression.

### EEG session

Participants were seated comfortably in a dimly illuminated room at a distance of 1 m from the computer screen. Fifteen pairs of facial expressions were used in the experiment combined by grouping each of the facial expressions with the remaining five facial expressions (e.g., Happiness with Sadness, Happiness with Fear, etc.). Each expression from the 15 pairs was used once as a deviant face (measuring changes in an expression) and once as a common face, resulting in 30 pairwise contrasts (Fig. 1A). Therefore, there were 60 stimulation sequences in total (each contrast was presented with 1 female and 1 male individual), the order of which was fully randomized for each participant. Each participant saw only one of the female and male facial identities for all of the expression contrasts. A different combination of two facial identities was selected and displayed for every participant. Each sequence started with a fixation cross displayed for 2-5 s, followed by 2 s of gradual stimulus fade in, 50 s of stimulation sequence, and 2 s of gradual stimulus fade out (Fig. 1). Thus, the EEG session lasted around 1.5 hours including breaks. The stimulation fade in and fade out were set to avoid abrupt eye-movements or blinks at the beginning or near the end of a sequence. Matlab 7.8 (R2009a) with PsychToolbox was used for stimulus display. Stimuli were presented through a sinusoidal contrast modulation as in many previous studies (e.g., Alonso-Prieto et al., 2013; Rossion & Boremanse, 2011). A practical advantage of this mode of stimulation is that it is a smoother stimulation mode than the square wave stimulation, thus making the experiment more comfortable for the participant. In addition, the visual stimulation was presented almost all the time, creating a continuously changing percept.

In each sequence, an emotion contrast (e.g., Fear and Happiness) was randomly selected. The common emotional face of the contrast was repeated throughout the whole sequence at a periodic frequency rate of 12 Hz (F). A presentation rate above 10 Hz for face stimulation should evoke reduced activity over occipito-temporal regions, with a maximum at medial occipital sites (Alonso-Prieto et al., 2013), thus spatially separating better the general visual response (i.e. onset of the face against the background) from the expression change response. The deviant emotional face (change of emotion) was presented at fixed intervals of every nine faces (i.e., F/9, 1.33 Hz). Moreover, as in our previous study (Dzhelyova et al., 2017), facial images randomly changed in size (between 90% and 110%, 4% steps) at every cycle to avoid facial expression discrimination based on low-level image properties (Fig. 1B). In order to maintain the attention of the participants high and constant during the experiment, they were asked to press a response key when a fixation cross randomly changed its color from black to red, 10 times within every sequence. The duration of the color change was 300 ms. The fixation

cross was presented in the center of the face stimuli, just below the eyes, which is the optimal fixation point for face perception (Peterson & Eckstein, 2012). Participants also were instructed to pay attention to the faces, which appeared on the screen. The performance of the color change detection of the fixation cross task was averaged across conditions with the same change of expression independently of the facial expressions used as common faces (e.g., fearful face as deviant and all other facial expressions as a common emotional face). Thus, the accuracy (ACC) rates and correct response times (RTs) of the color change detection task were calculated for each of the six basic emotions presented as deviant facial expressions (*Expression Change*: Anger, Fear, Disgust, Happiness, Sadness, Surprise). No significant differences were found for the accuracy rates, $F(5, 70) = 1.192$, $p = 0.32$, or RTs for correct trials, $F(5, 70) = 0.269$, $p = 0.929$.

## EEG acquisition

EEG activity was recorded via a BIOSEMI Active two amplifier system (Biosemi, Amsterdam, Netherlands), with 128 Ag/AgCl electrodes at 512 Hz. The electrodes include standard 10-20 system locations as well as additional intermediate positions. Eye movements were monitored with four electrodes placed at the outer canthi of the eyes (HEOG) and above and below the right eye (VEOG).

## EEG preprocessing and statistical analysis

**Preprocessing** All EEG pre-processing steps were performed using Letswave 5 (http://nocions.webnode. com/letswave) and Matlab 7.8 (R2009a) (The Math works) and were almost identical to our previous study (Dzhelyova et al., 2017). EEG data were segmented to include 2 s before and after each sequence, resulting in 58 s segments (−2 to 56 s). Then, it was digitally band-pass filtered at 0.10–100 Hz with a Butterworth filter (4th order). Artifact-ridden or noisy channels were replaced using linear interpolation of the three neighboring channels; less than 5% of all channels were interpolated for a given sequence. After that, a common average reference computation was applied to all channels.

**Frequency analysis** Preprocessed data segments were cropped to an integer number of 1.33 Hz cycles beginning 4 s after the onset of the sequence until approximately 52 s (~48 sec, 64 expression change cycles, 24576-time bins in total). The first 4 s of each sequence were excluded to avoid any contamination by the initial transient responses. The two sequences were averaged for each of the 30 emotion contrasts. A Fast Fourier Transform (FFT) was then applied to these averaged segments, and amplitude spectra were extracted for all channels (square root of the sum of squares

of the real and imaginary parts divided by the number of data points). Thanks to the long time-window (48 s), frequency analysis yielded spectra with a high-frequency resolution of 0.021 Hz (1/48), thus increasing SNR (Regan, 1989) and allowing unambiguous identification of the response at the frequencies of interest (i.e., 1.33 Hz and its harmonics for the oddball rate; 12 Hz and its harmonics for the base rate). To identify the presence of statistically significant responses at the frequency of deviant stimuli presentation and its harmonics, Z-scores were calculated (the difference between amplitude at the frequency of interest and mean amplitude of 20 surrounding frequency bins divided by the standard deviation of the 20 surrounding bins; Rossion et al., 2012; Dzhelyova et al., 2017). Only significant responses were taken for analysis (Z-score > 1.64, $p < 0.05$ one-tailed). First amplitude spectra across subjects separately for each condition were averaged (i.e. grand averages), and then the resulting grand-averaged spectra were pooled across all 128 channels. The majority of the 30 pairwise emotion contrasts were significant for the first eleven harmonics of the expression change frequency (up to 14.66 Hz). For each participant, those harmonics (excluding the 12 Hz, corresponding to the general response) were included in the quantification of the discrimination response to changes in expression. This summed baseline-corrected amplitudes response accounted for $95.70 \pm 2.68$ percent of the response of the 20 harmonics (up to 26.67 Hz) for the 30 emotion contrasts. The 12 Hz response showed significant 8 harmonics (up to 96 Hz), which decreased in magnitude with increasing harmonic frequencies. The first four harmonics were characterized by a similar middle occipital topography and thus were included in the quantification of the 12 Hz response.[1] To quantify these responses, the baseline-corrected amplitudes were calculated by subtracting the average amplitude of the 20 surrounding bins (10 on each side, excluding the immediately adjacent bin and the bins containing the highest and lowest amplitudes).

### General visual response – 12 Hz

We pooled the data across all conditions having the same change of expression, independently of the facial expression presented as a common stimulus (e.g., *Anger* among *Disgust* was combined with *Anger* among *Fear*, *Anger* among *Happiness*, etc.). This resulted in 6 averages, for which we could analyze separately the 12 Hz rate.

The 12 Hz response focused over the medial occipital region, a distribution typical for stimulation rates above 10 Hz (e.g., Alonso-Prieto et al., 2013; Dzhelyova et al., 2017).

---

[1] The pattern of results for the base rate did not change after the inclusion of the second harmonic (24 Hz).

Based on previous studies (Dzhelyova et al., 2017) and the scalp distribution obtained when combining all facial expressions, the medial occipital region was defined to contain channels POz, POOz, Oz, Oiz, and Iz. The extracted summed baseline-corrected amplitudes over the MO region were evaluated with repeated measures ANOVA with factors *Expression Change* (Anger, Fear, Disgust, Happiness, Sadness, Surprise).

### Response to deviant expressions (change in expression) – 1.33 Hz

The change in expression response was characterized by a bilateral occipito-temporal response, a response characteristic to high-level processing of visual stimuli (Alonso-Prieto et al., 2013; Liu-Shuang et al., 2014; Lochy et al., 2016) particularly for faces. To evaluate any hemispheric differences, two regions of interest, ROIs (left and right occipital-temporal region) were defined. The summed baseline-subtracted amplitudes were averaged across five electrodes for each ROI: left occipito-temporal (LOT): PO7, PO9, PO11, P7, P9; right occipital-temporal (ROT): PO8, PO10, PO12, P8, P10. To assess the response to deviant faces, we performed three types of analyses.

First, we analyzed differences in the amplitude of the expression change response (Anger, Fear, Disgust, Happiness, Sadness, Surprise) independently of the facial expressions presented as a common stimulus. The extracted summed baseline-corrected amplitudes were subjected to repeated-measures ANOVA with factors *Expression Change* (Anger, Fear, Disgust, Happiness, Sadness, Surprise) and *ROI* (left, right).

Second, following this initial analysis, each of the six facial expressions was evaluated separately to examine any differences in the amplitude of the discrimination response due to a specific contrast. Because there were no significant hemispheric differences (see *Results* section) in the response to deviant expressions (expression change), the summed baseline-corrected amplitudes were averaged over LOT and ROT ROIs and were subjected to repeated-measures ANOVA with a factor *Common Expression* (the remaining 5 expressions).

Lastly, we compared the two contrasts with the same pair of expressions to evaluate whether the amplitude varied as a function of the facial expression presented as a common stimulus. To adjust for multiple comparisons, a sequential Holm-Bonferroni correction (Hochberg & Tamhane, 1987) was applied to this additional analysis. A Greenhouse-Geisser correction for degrees of freedom was applied if Mauchly's test of sphericity was significant for all ANOVAs.

### Individual analysis

To analyze the response at an individual level, individual's FFT spectrum for each of the basic emotion expressions presented as deviant stimuli independently of the expression presented as a common stimulus was cropped, centered at the presentation of a deviant emotion (1.33 Hz and the subsequent harmonics until 14.66 Hz, excluding the response at 12 Hz), surrounded by the 20 neighboring bins, representing the noise level. Then, we summed the spectrum, including the response and the noise level and calculated a z-score for each frequency bin. In the absence of a signal at the central bin of interest, the value at this bin has 1 chance out of 21 (i.e., $p < 0.05$) to be the highest in the spectrum. To determine the participants' hemispheric dominance for each expression, we calculated a lateralization index (RH-LH/RH+LH). The lateralization index was subsequently classified as left if it was $< -0.2$; as bilateral, if it was between $-0.2$ and $0.2$, and as right-lateralized if it was $> 0.2$ (Centeno et al., 2014; Trimmel et al., 2017).

### Brain topographical analysis

Brain topographical analysis was performed to clarify whether there were scalp distribution differences among the six facial expression changes. In previous studies, a brief change from neutral to happy expression was found to elicit a more dorsal response over posterior electrodes compared with the other expressions (Dzhelyova et al., 2017; Leleu et al., 2018). Consequently, to evaluate qualitatively whether the response among the basic human expressions differed, normalization (McCarthy & Wood, 1985) of the individual baseline-corrected amplitudes was applied. For each participant, the baseline-corrected spectrum for each condition was first scaled by dividing the value at each electrode by the scalp-wide root-mean-square value (i.e. the square root of the sum of squares for all 128 electrodes) (McCarthy & Wood, 1985). The normalized amplitudes were summed for the significant expression change frequency and its harmonics and then analyzed with repeated measures ANOVAs with *Expression Change* (Anger, Disgust, Fear, Happiness, Sadness, Surprise) and *Electrode* at anterior (66 electrodes) and at posterior (62 electrodes) sites. Importantly, a significant interaction between *Expression Change* and *Electrode* would indicate topographical differences among the expressions.

### Perceptual and neural similarity measures

To estimate the relationship between the perceptual and neural similarity among the expressions, we compared the expression change detection efficiency score (ES = ACC (proportion correct)/RT (in sec) (Townsend & Ashby, 1978, 1983) and the discrimination amplitudes. This comparison allowed us to explore the 30 contrast and thus investigate whether the asymmetries observed in the behavioral response matched the ones found in the neurophysiological one (see *Results*). For

each expression contrast, we averaged across trials and subjects ES and the baseline-corrected amplitudes and z-transformed them. These z-scores were arranged into similarity matrices (Fig. 6A). We performed the analysis at a group and at an individual level. A group r value was calculated by correlating the perceptual similarity based on the expression detection task and the neural similarity based on the discrimination response, averaged across all participants. The individual relationship was estimated by applying, a linear mixed model (function "lmer" in R; Bates, Maechler, Bolker, & Walker, 2015) predicting the behavioral performance as a function of the discrimination response over the PO region. The explanatory variables were: *Condition* (within-participant, categorical fixed factor), baseline-corrected amplitudes of the *Expression Change* (within-participant, continuous fixed covariate), and *Participant* (random factor). These two types of analyses were performed for both the 15 contrasts resulting from averaging the contrasts with the same pairs of expressions as well as for all 30 contrasts. Finally, to control for the image characteristics of the stimuli we re-estimated the correlation between the perceptual and neural similarity controlling for the dissimilarity index of the images. Results with the other behavioral tasks can be found in Supplementary material: Perceptual and neural similarity measures.

## Results

### Behavioral tasks

#### Forced-choice face labeling test and expression similarity test

Both explicit behavioral tasks, requiring to match facial expressions to labels or to relate/dissociate facial expressions confirmed the types of confusions that are typically observed in such tasks (Adolphs, 1999; Calvo & Lundqvist, 2008; Palermo & Coltheart, 2004; Tottenham, et al., 2009). Fear was the least distinguishable expression, by far, and often was misjudged as Surprise or Disgust and also the slowest expression matched to its label. In contrast, Happiness was the most distinguishable expression. In addition, disgusted and angry faces often were confused as each other and perceived to be more similar. A detailed description of the results can be found in the Supplementary material: Behavioral Results.

#### Expression change detection task

We performed three types of analyses. First, we evaluated the behavioral performance on the task, independently of the expression presented as a common face. This allowed us to compare the results of this novel task to the other two explicit tasks. Second, we evaluated the confusability among the expressions by evaluating the accuracy rates and reaction time latencies necessary to detect changes for each facial expression depending on the expression presented as a common face. Finally, contrasts with the same pair of expressions were compared to evaluate whether behavioral performance differed as a function of the expression presented as a common face. For all post hoc tests, we applied sequential Holm-Bonferroni correction for multiple comparisons.

**Analysis independent of the common expression** First, we collapsed the accuracy (ACC) and correct response times (RT) data of each expression presented as deviant stimuli across the facial expressions used as common faces. A repeated-measures ANOVA with a within-subject factor *Expression change* (Fear, Anger, Disgust, Happiness, Sadness, Surprise) revealed that the RTs were significantly different among the basic emotions used as a deviant expression, $F (5, 70) = 10.683$, $p < 0.0001$, $\eta^2_p = 0.433$. The discrimination of *Happiness* and *Surprise* was performed the fastest. After applying Holm-Bonferroni correction for multiple comparisons, results revealed that *Happiness* was detected significantly faster than all other expressions ($ps \leq 0.007$) except *Surprise*, whereas changes to *Surprise* were detected faster than changes to *Anger* and *Sadness* ($ps < 0.0001$). In addition, changes to *Disgus*t were detected faster than changes to *Anger* ($p = 0.014$).

Comparable results were found when evaluating the accuracy and efficiency scores data and can be found in the Supplementary material—Expression change detection task: accuracy and efficiency scores.

**Analysis depending on the common face** To evaluate whether the perceptual similarity of the expressions affected the discrimination among the expressions in the explicit task, we conducted additional analysis on response times and accuracy scores for each expression separately dependent on the facial expression used as a common face.

The response time data showed significant differences for changes to fearful $[F (2.138, 29.935) = 6.982, p = 0.003, \eta^2_p = 0.333]$, happy $[F (4, 56) = 5.480, p = 0.001, \eta^2_p = 0.281]$, sad $[F (4, 56) = 10.078, p < 0.001, \eta^2_p = 0.419]$, and surprised $[F (1.975, 27.644) = 9.738, p = 0.001, \eta^2_p = 0.410]$ faces.

When present among angry and disgusted faces, fearful faces were faster discriminated than when presented among faces with happy, sad, or surprised expression ($ps \leq 0.048$). Similarly, the discrimination of sad and surprised faces took longer when present among fearful faces compared with their discrimination among the other expressions ($ps \leq 0.013$).

The fastest discrimination of happy faces was among sad expression ($ps < 0.017$). Additionally, the discrimination of happy faces among surprised expression resulted in a significantly shorter time than the discrimination of happy faces among disgusted faces ($p = 0.038$; Table 1A). The lowest accuracy for angry faces was observed when the common face

**Table 1.**  Results for the expressions change detection task. Response time in ms (A) and accuracy (B), score for all expression contrasts

|  |  | Anger | Disgust | Fear | Happiness | Sadness | Surprise |
|---|---|---|---|---|---|---|---|
| **A** |  | Deviant expression |  |  |  |  |  |
| Common Expression | Anger | - | 528.33 | 476.75 | 485.88 | 519.54 | 470.54 |
|  | Disgust | 567.99 | - | 481.36 | 493.99 | 510.07 | 478.26 |
|  | Fear | 551.31 | 523.81 | - | 495.56 | 593.50 | 547.27 |
|  | Happiness | 534.84 | 491.29 | 514.01 | - | 502.01 | 465.48 |
|  | Sadness | 536.58 | 539.80 | 515.59 | 451.88 | - | 489.97 |
|  | Surprise | 508.26 | 502.71 | 561.58 | 480.47 | 528.18 | - |
|  | **MEAN** | **539.79** | **517.19** | **509.86** | **481.55** | **530.66** | 490.30 |
| **B** |  | Deviant expression |  |  |  |  |  |
| Common Expression | Anger | - | 0.77 | 0.84 | 0.89 | 0.71 | 0.92 |
|  | Disgust | 0.66 | - | 0.87 | 0.85 | 0.79 | 0.94 |
|  | Fear | 0.70 | 0.81 | - | 0.82 | 0.54 | 0.71 |
|  | Happiness | 0.77 | 0.87 | 0.83 | - | 0.83 | 0.94 |
|  | Sadness | 0.76 | 0.80 | 0.82 | 0.91 | - | 0.88 |
|  | Surprise | 0.79 | 0.86 | 0.62 | 0.91 | 0.78 | - |
|  | **Mean** | **0.73** | **0.82** | **0.80** | **0.87** | **0.73** | **0.88** |

was disgusted, reaching significance compared with the accuracy score when presented among happy ($p = 0.049$), sad ($p = 0.016$), and surprised ($p = 0.009$) faces. Additionally, the accuracy dropped when angry faces were presented among fearful faces than among surprised faces ($p = 0.033$).

The discrimination of fearful faces was least accurate when they were presented among surprised ($ps \leq 0.001$).

In addition to these expressions: fearful $[F (4, 56) = 11.851, p < 0.001, \eta^2_p = 0.458]$, happy $[F (4, 56) = 3.158, p = 0.021, \eta^2_p = 0.184]$, sad $[F (4, 56) = 14.122, p < 0.001, \eta^2_p = 0.502]$, and surprised $[F (2.192, 30.682) = 19.6, p < 0.001, \eta^2_p = 0.583]$ faces, the evaluation of the accuracy data added significant differences in the accuracy performance for changes to angry faces $[F (4, 56) = 2.742, p = 0.037, \eta^2_p = 0.164]$.

Analogously to the low accuracy for the discrimination of fearful faces among surprise expression, the lowest accuracy for surprised faces was found when they were presented among fearful faces ($ps \leq 0.028$). Additionally, the discrimination of surprised faces when presented among happy faces led to less accurate discriminations than the discrimination of surprised faces when presented among sad faces ($p = 0.017$).

The lowest accuracy scores when fearful faces acted as a common face also were observed for sad expressions ($ps \leq 0.007$). Additionally, a difference in the accurate detection of sad faces was found among happy and angry faces ($p = 0.042$). Finally, the discrimination of happiness among fearful faces produced lower accuracy scores than among sad ($p = 0.026$) or surprised ($p = 0.015$) faces (Fig. 1S).

**Analysis of expressions contrasts with the same pairs of expressions** The final analysis evaluated if the explicit discrimination performance was affected whether the expression acted as a deviant or common face (e.g., context-dependent). To this end, we compared all contrasts with the same pairs of expressions. Five contrasts (Fear vs. Anger, Fear vs. Sadness, Anger vs. Happiness, Happiness vs. Sadness, and Surprise vs. Sadness) revealed significant differences for both accuracy rate and response time latencies ($ps < 0.023$). The Anger versus Surprise and Disgust versus Fear pairs differed only in the response time latencies ($ps < 0.005$), whereas the Anger versus Disgust and Disgust versus Surprise pairs differed only in the accuracy score ($ps \leq 0.01$).

In summary, the behavioral discrimination task under tight temporal constraints confirmed *Happiness* as the most distinguishable expressions from the other five basic expressions. Nonetheless, *Surprise* also was efficiently discriminated from the other expressions, except when compared to *Fear*. Interestingly, discriminating a deviant expression from common fearful faces often was associated with the lowest performance. However, when *Fear* was used as a deviant expression, it was not the most difficult to discriminate; *Sadness* and *Anger* were much more difficult to discriminate overall. Compared with the other two behavioral tasks, the discrimination of *Fear* reached a better performance in the expression change detection task. Regarding the two-by-two comparisons, *Sadness, Disgust,* and *Fear* were easily confused with each other and resulted in low accuracy and long response time for detection. Importantly, we also observed asymmetries in the responses depending on whether the expression was presented as a common or as deviant face, highlighting the importance of context when studying expression processing.
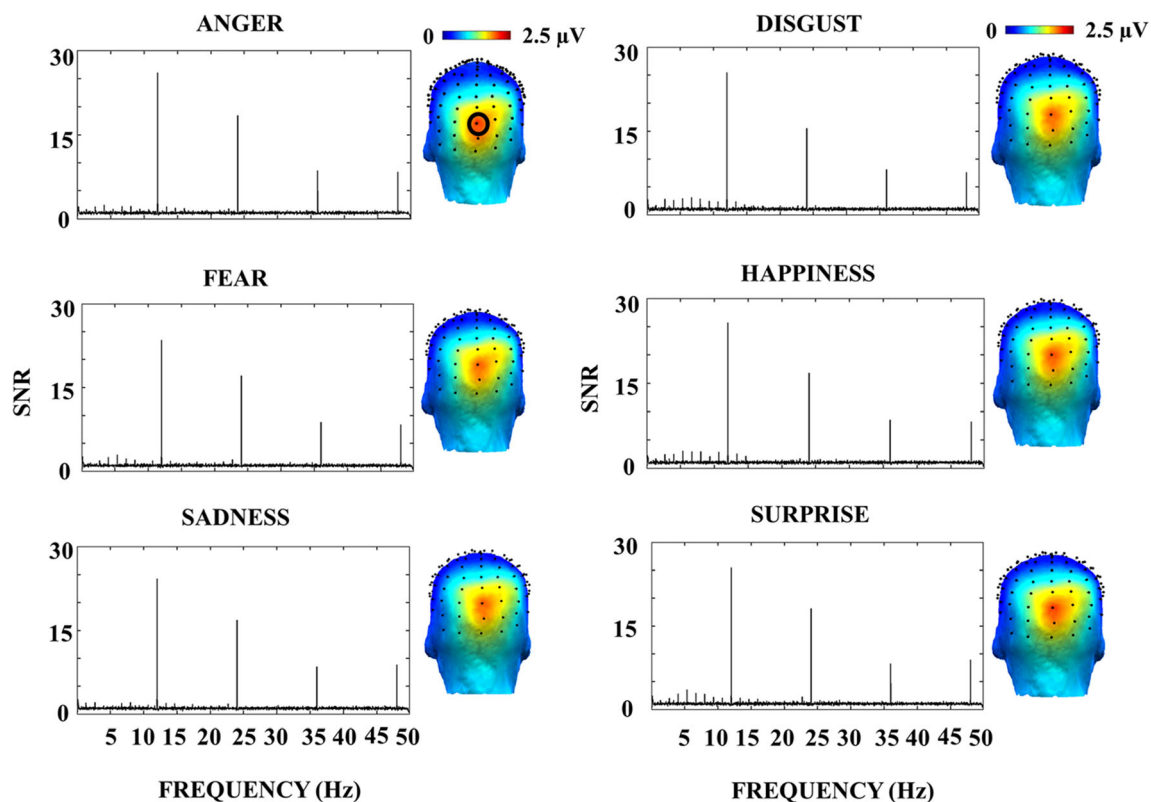
**Fig. 2** General visual response and its harmonics. Grand-averaged SNR spectra (0-50 Hz) at Oz for the six deviant emotions. Topographical maps display the summed (12 Hz, 24 Hz, 36 Hz, 48 Hz) baseline-corrected amplitudes for the base rate response

## EEG data

### Frequency domain analysis

**General response (12 Hz) and its harmonics** Grand-averaged SNR spectra showed clear responses at the 12 Hz base stimulation frequency and its harmonics (24 Hz, 36 Hz, etc.) (Fig. 2). The analysis of this response, which should not differ across the different conditions, because it merely reflects the contrast between the facial images and the background, focused over MO ROI. The repeated measures ANOVA on the summed baseline-corrected amplitudes for the 4 harmonics of the 12 Hz response (12 Hz, 24 Hz, 36 Hz, 48 Hz) over the middle occipital ROI indicated no difference in the amplitude among the basic emotions presented as deviant expressions, *Expression change*, $F(1, 14) = 1.574$, $p = 0.179$.

**Expression change response (1.33 Hz) and its harmonics** Figure 3A shows clear facial expression discrimination responses for each of the six facial expressions presented as deviant emotions. As found in previous studies (Dzhelyova et al., 2017; Leleu et al., 2018), the discrimination response to changes of *Happiness* was characterized by a dorsal-parietal response over the right hemisphere. The weakest responses were found for *Sadness* and *Anger*, followed by *Fear*,

whereas *Surprise* resulted in the largest contrast response. The baseline-corrected amplitudes, summed for the significant harmonics (1.33 Hz and the 11 consecutive harmonics excluding the 9th harmonic, corresponding to the general visual response), were analyzed over the left and right ROIs where the maximal discrimination responses were found. The same pattern of results was found when the response to deviant emotion was evaluated over the whole scalp (averaged of the 128 channels), providing an unbiased measurement of the strength of the discrimination response (see Supplementary material: EEG results over the whole scalp).

Similarly to the behavioral task, we analyzed differences in the amplitude of the expression change response (Anger, Fear, Disgust, Happiness, Sadness, Surprise) in three ways. First, we evaluated the response independently of the facial expressions presented as a common stimulus. The extracted summed baseline-corrected amplitudes were subjected to repeated-measures ANOVA with factors *Expression Change* (Anger, Fear, Disgust, Happiness, Sadness, Surprise) and *ROI* (left, right). Second, following this initial analysis, each of the six facial expressions was evaluated separately in order to examine any differences in the amplitude of the discrimination response due to a specific contrast. Because there were no significant hemispheric differences (see *Results* section) in the response to deviant expressions (expression
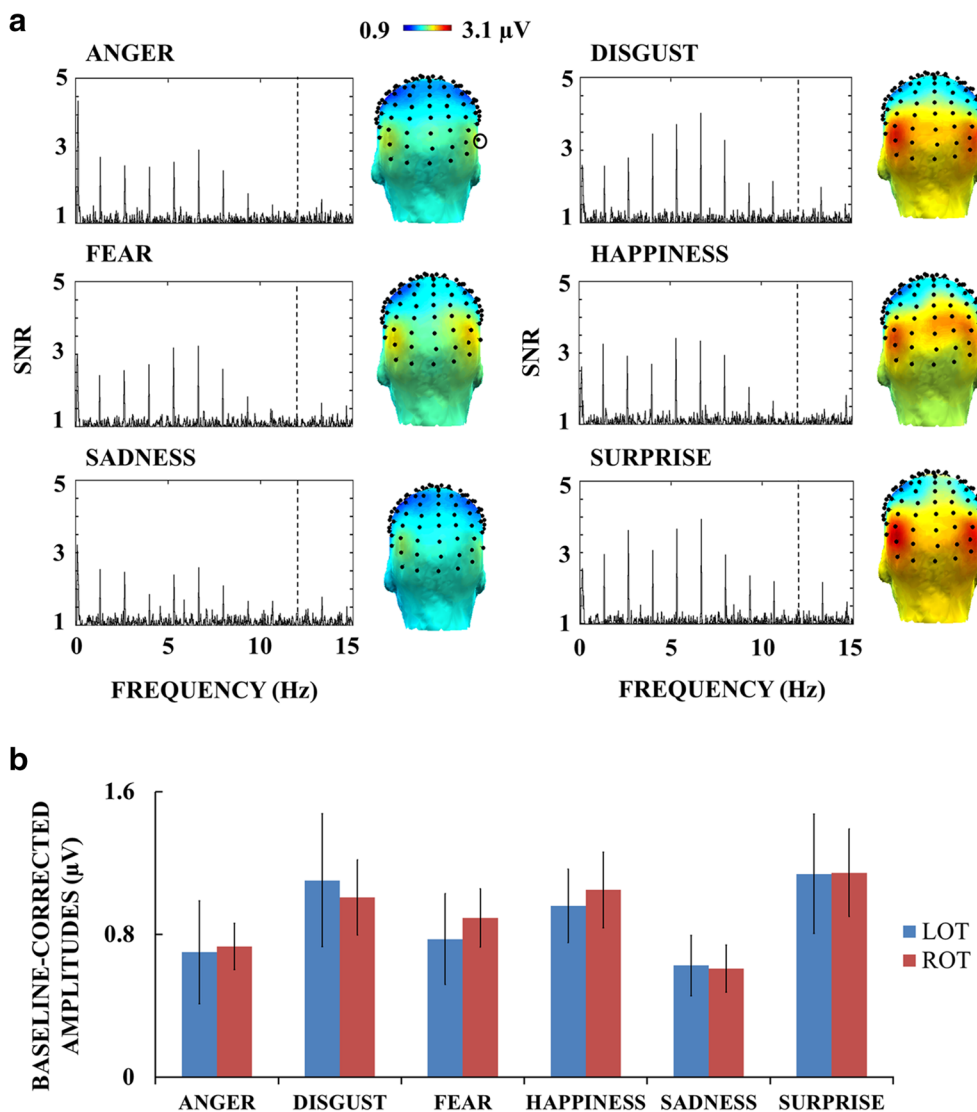
**Fig. 3** Expression change discrimination response. **A)** Grand-averaged SNR spectra at electrode P10 over the right hemisphere, for all facial expressions changes, averaged across the remaining five basic expressions presented as a common face. Topographical maps illustrate the distribution of the discrimination response summed for the significant harmonics (1.33 Hz and its harmonics to 14.66 Hz, excluding the base rate 12 Hz) for each of the basic expressions. SNR = 1 indicates noise level. **B)** Summed baseline-corrected amplitudes for each of expression changes across left and right OT (occipito-temporal) ROIs

change), the summed baseline-corrected amplitudes were averaged over LOT and ROT ROIs and were subjected to repeated-measures ANOVA with a factor *Common Expression* (the remaining 5 expressions). Last, we compared the two contrasts with the same pair of expressions in order to evaluate if the amplitude varied as a function of the facial expression presented as a common stimulus.

*Analysis independent of the common expression*

Repeated measures ANOVA with factors *ROI* (LOT and ROT) and *Expression change* (Anger, Disgust, Fear, Happiness, Sadness, Surprise) revealed no hemispheric differences in the processing of the facial expressions, *F*

(1, 14) = 0.018, *p* = 0.89 (Fig. 3B). The main effect of *Expression change* was significant, *F* (2.53, 35.39) = 4.356, *p* = 0.014, $\eta_p^2$ = 0.237. The discrimination response to sad faces led to the weakest discrimination response, which was significantly weaker than the discrimination of happy (*p* = 0.005), surprised (*p* = 0.011), disgusted (*p* = 0.041), and only demonstrated a trend to be smaller than the response to changes to fearful (*p* = 0.058) faces. The discrimination of Anger also produced low discrimination response, significantly lower than the discrimination of Disgust (*p* = 0.024) and Surprise (*p* = 0.005). In contrast, the strongest response was obtained from changes to surprised faces, followed by changes to disgusted and happy faces. In addition to being stronger
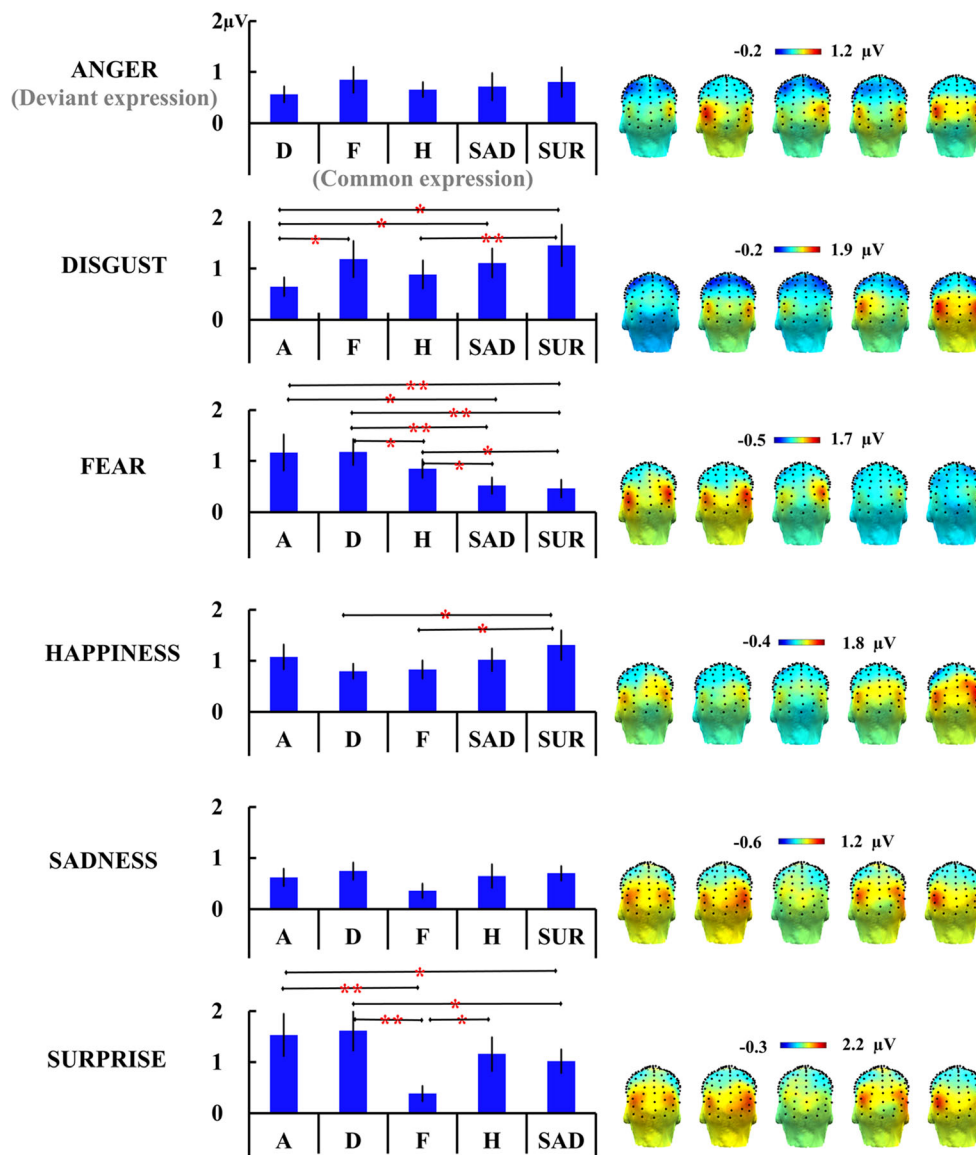
**Fig. 4** Summed baseline-corrected amplitudes (1.33 Hz and its harmonics to 14.66 Hz, excluding the base rate 12 Hz) for the discrimination of each facial expression presented as deviant stimuli (marked on the y-axis) from the other five expressions, presented as common stimuli (marked on the x-axis) across left and right OT (occipito-temporal) ROIs. Topographical maps for each of the emotion contrasts. In different colors are presented the individual responses. Common stimuli: A – anger; D – disgust; F – fear; H – happiness; SAD – sadness; SUP – surprised.*p < 0.05; **p < 0.01; ***p < 0.001

**Table 2.** Summed baseline-corrected amplitudes for the discrimination of each deviant facial expression (expression change) from the other common expressions across left and right OT ROIs. *Common stimuli: A — anger; D — disgust; F — Fear; H — Happiness; SAD — sadness; SUP — surprised*

|  |  | Common expression | | | | | |
|---|---|---|---|---|---|---|---|
|  |  | A | D | F | H | SAD | SUR |
| Oddball expression | Anger |  | 0.561 | 0.847 | 0.654 | 0.712 | 0.803 |
|  | Disgust | 0.643 |  | 1.186 | 0.884 | 1.110 | 1.453 |
|  | Fear | 1.165 | 1.175 |  | 0.847 | 0.516 | 0.461 |
|  | Happiness | 1.076 | 0.794 | 0.828 |  | 1.018 | 1.307 |
|  | Sadness | 0.622 | 0.746 | 0.363 | 0.649 |  | 0.701 |
|  | Surprise | 1.530 | 1.615 | 0.387 | 1.160 | 1.019 |  |

than the response to changes to sad and angry faces, the expression change response to Surprise was stronger than expression change response to Fear ($p = 0.033$). The interaction between the *ROI* and *Expression Change* did not reach significance, $F (5, 70) = 0.86$, $p = 0.51$.

*Analysis depending on the common face*

The additional analysis conducted to investigate the response to each facial expression separately depending on the facial expression used as a common face showed differences for changes to fearful $[F (4, 56) = 5.598, p = 0.001, \eta^2_p = 0.277]$, surprised $[F (2.325, 32.550) = 8.122, p = 0.001, \eta^2_p = 0.367]$, disgusted $[F (2.367, 33,143) = 4.276, p = 0.017, \eta^2_p = 0.234]$, and happy $[F(2.615, 36.608) = 3.057, p = 0.047, \eta^2_p = 0.179]$ faces (Fig. 4; see Table 2 for amplitude data).

When presented among sad and surprised faces, fearful faces elicited a smaller discrimination response than when presented among faces with any of the other expressions. Applying Holm-Bonferroni correction revealed a significant difference in the magnitude of the discrimination response when fearful faces were presented among surprised and sad faces compared with when fearful faces were presented among disgusted ($p = 0.004$ among surprised and $p = 0.007$ among sad) and among angry ($p = 0.019$ among surprised and $p = 0.027$ among sad) and reached only a trend to be significantly smaller than the discrimination response to fearful faces among happy faces ($p = 0.047$ fear among surprise and $p = 0.032$ fear among sadness). In addition, fearful faces among disgusted faces resulted in a larger discrimination response compared with fearful among happy faces ($p = 0.035$).

Analogously to the decreased amplitudes for the discrimination response for fearful faces among surprised faces, when presented among fearful expressions, surprised faces led to the smallest discrimination response, significantly smaller than the response when presented among faces displaying any of the other basic emotions: Disgust ($p = 0.002$), Sadness ($p = 0.024$), Anger ($p = 0.005$), or Happiness ($p = 0.013$). Moreover, the discrimination of surprised faces among sad faces revealed the second lowest response, which was significantly lower than that of discriminating surprised faces among disgusted faces ($p = 0.016$) and tended to be smaller than the discrimination among angry faces ($p = 0.047$).

The discrimination of disgusted faces among angry faces revealed the lowest response, which was significantly smaller than the discrimination of disgusted faces among sad ($p = 0.018$) and surprised ($p = 0.019$) faces and tended to be smaller than among fearful ($p = 0.031$). The discrimination of disgusted faces among happy faces also elicited a lower response than the other discrimination responses, which also was significantly smaller than the discrimination of disgusted faces among surprised faces ($p = 0.009$).

The last discrimination response that varied contingent on the expression presented as a common face was the discrimination response to happy faces, due to smaller response to changes to happiness when presented among fearful and disgust faces compared with the other expression, significantly smaller than when presented among surprise faces ($p = 0.015$ for fearful faces, $p = 0.036$ for disgust faces). A decreased response to changes in another expression among fearful expressions also could be observed for sadness and surprise. Specifically, response to changes to surprise when presented among fear was significantly smaller than when presented among anger ($p = 0.005$), disgust ($p = 0.002$), and happiness ($p = 0.013$).

*Analysis of expressions contrasts with the same pairs of expressions*

Finally comparing the contrast with the same expressions, we found that there were no consistent differences for the majority of the contrasts with the same pair of expressions when they appeared as a common or deviant face (expression change) except for two contrasts: sad-happy and angry-surprised that survived Bonferroni-Holm correction $ps \geq 0.024$ while the contrast happy-angry tended to differ ($p = 0.032$) depending on the position of the two expressions.

### Individual response

At an individual level (Fig. 5A), significant responses ($z > 1.64$, $p < 0.05$) were observed for all expressions in 14 of 15 participants (1 participant did not show a significant response for changes to disgusted faces). Restricting the response to the left and right occipito-temporal region, again most participants showed at least one significant electrode from all ten electrodes. Participant 15 did not show a significant response for changes to Disgust and Happiness (the topography of the responses was not specific to the frequencies of interest), and participant 4 did not show a significant response for changes to Fear (distributed over the MO region) and Sadness (distributed over the parietal region). The bilateral distribution at a group level could be explained by the strong variation of individual topographies across the different expressions (Fig. 5B).

### Scalp topography analysis

Scalp topography analysis was performed to examine the scalp distribution differences among the responses to changes to angry, disgusted, fearful, happy, sad, and surprised faces, separately for the anterior and posterior half of the scalp split along the midline. Normalized (McCarthy & Wood, 1985) amplitudes were evaluated with repeated measures ANOVA with *Expression change* and *Electrode* as factors. Importantly, a significant interaction between *Expression change* and
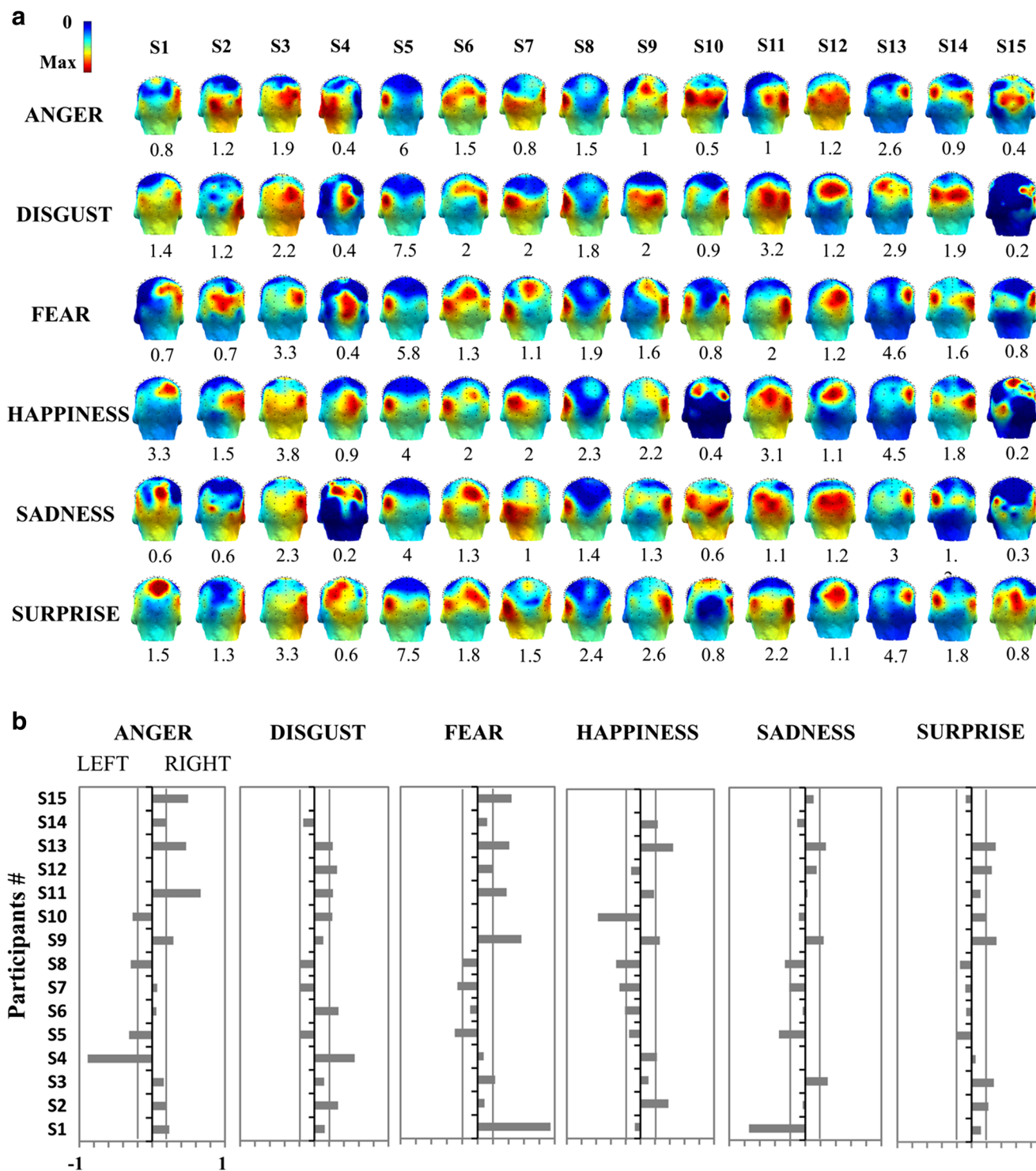
**Fig. 5** Individual response. **A)** Individual topographical maps for each of the facial expressions presented as deviant stimuli, independently of the expression shown as a common stimulus. **B)** Hemispheric lateralization for each facial expression presented as a deviant face, independently of the expression shown as a common stimulus. Differences between the response over the ROIs in the left and the right hemisphere $< \pm 0.2 \ \mu V$ are considered as bilateral and highlighted in grey. Note: S15's lateralization index for Disgust and Happiness discrimination is missing, S4 lateralization index for Sadness discrimination also is not displayed on the figure due to low/absent responses within the ROIs

*Electrode* would indicate differences in the topographies, hinting at a contribution from different sources. The interaction was significant for posterior channels, $F (305, 4270) =$ 1.403, $p < 0.001$, $\eta^2_p = 0.091$, but not for anterior channels, $F (325, 4550) = 1.110$, $p = 0.091$, indicating that the topographical differences were mostly driven by posterior sites, possibly

due mainly to the specific topographical distribution for discrimination of Happiness compared with the discrimination of the other expressions.

Based on the scalp distributions of *Happiness* discrimination and previous findings (Dzhelyova et al., 2017; Leleu et al., 2018), we tested first specifically for topographical differences for Happiness. Indeed, discrimination of Happiness produces different topographies compared with all other expressions ($ps \leq 0.002$ happy vs. angry, $F(61, 854) = 1.690$, $p = 0.001$, $\eta^2_p = 0.108$; happy vs. disgusted, $F(61, 854) = 1.654$, $p = 0.002$, $\eta^2_p = 0.106$; happy vs. fearful, $F(61, 854) = 1.727$, $p = 0.001$, $\eta^2_p = 0.110$; happy vs. sad, $F(61, 854) = 2.741$, $p < 0.001$, $\eta^2_p = 0.164$; happy vs. surprised, $F(61, 854) = 1.641$, $p = 0.002$, $\eta^2_p = 0.105$).

Additionally, topographical differences were found for Anger versus Surprise ($F(61, 854) = 1.423$, $p = 0.021$, $\eta^2_p = 0.092$), Fear versus Sadness ($F(61, 854) = 1.529$, $p = 0.007$, $\eta^2_p = 0.098$), and Sadness versus Surprise ($F(61, 854) = 1.786$, $p < 0.001$, $\eta^2_p = 0.113$).

## Summary of the EEG data

In summary, we found clear responses to brief changes (at a single fixation, SOA = 83 ms) from an expressive face to another expressive face, as indexed by a robust discrimination response in the EEG, for all basic emotion. This response was generally distributed over bilateral occipito-temporal sites, typical of face perception, although, notably, there was no right hemispheric advantage, unlike when contrasting a facial expression to neutral faces (Dzhelyova et al., 2017; Leleu et al., 2018). The amplitude of the response to changes of expressions was contingent on the expression that needed to be recognized; more subtle expressions, such as *Sadness* led to a weaker discrimination response compared with more distinguishable expressions, such as *Happiness, Disgust,* and *Surprise.* Furthermore, the magnitude of the discrimination response depended on the expression used as a common stimulus. The magnitude of the discrimination response was reduced when the common expression was perceptually similar
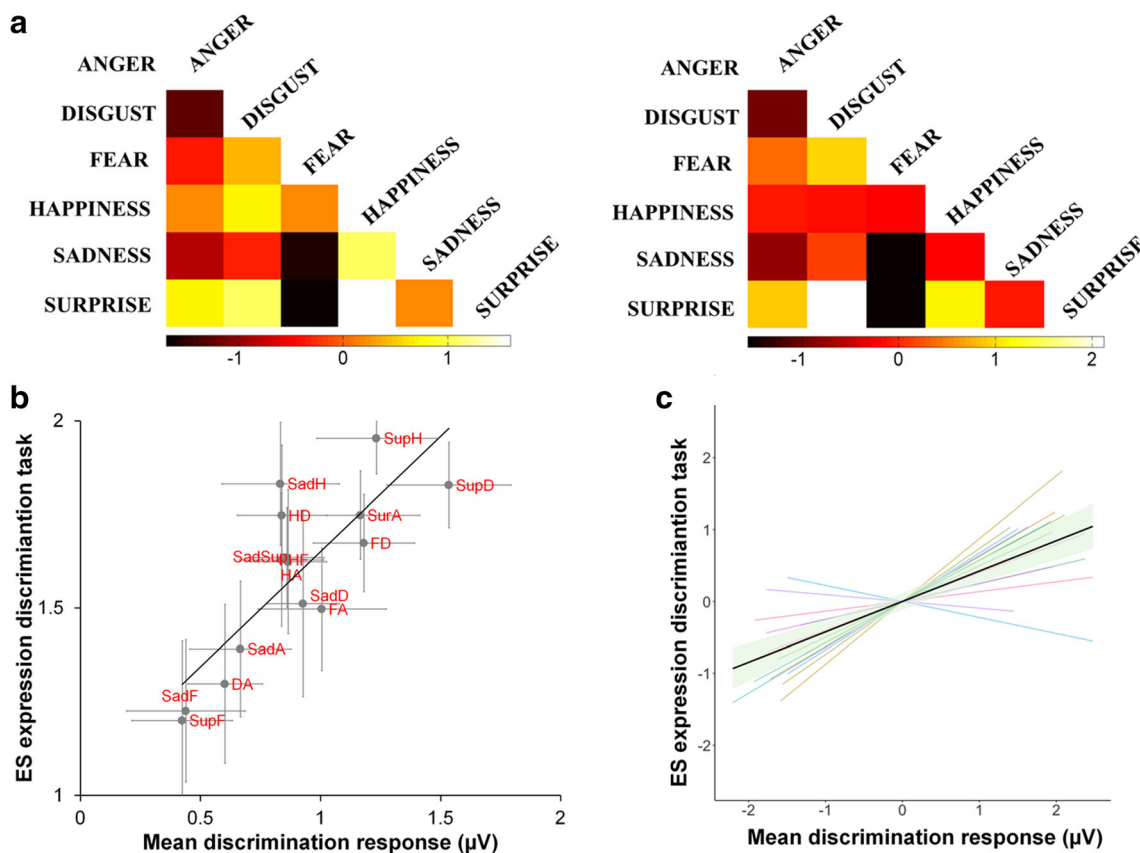


**Fig. 6** Comparison of the perceptual similarity measures and the discrimination response similarity. **A)** Similarity matrix for ES of the expression detection task and the discrimination response over the PO region. Scores and amplitudes are z-transformed. **B)** Correlations between the perceptual similarity as measured by the ES of the expression change detection task and the discrimination response over the PO region, providing a direct measure of neural similarity between the 15 pairs of expressions over the left and right occipito-temporal regions of interest. The positive correlations indicate that perceptually more similar expressions (i.e., lower ES) lead to weaker EEG discrimination responses. **C)** Individual estimates of the relationship between perceptual and neural similarity among the expressions. ES = efficiency score (ES = Accuracy (proportion)/RT(in sec)), PO = occipito-temporal region averaged over the 10 electrodes of interest within the left and right hemisphere

to the expression change observed in the sequence. For instance, changes to fearful faces produced weak discrimination responses when presented among sad and surprised faces. Similarly, changes to disgusted faces elicited a weaker response when presented among angry facial expressions than among any of the other expressions. Interestingly, when fearful faces were the common faces, they led to the weakest discrimination response, not only for the perceptually similar expressions, such as *Sadness* and *Surprise* but also for *Happiness*. Importantly, these variations of the discrimination response were not observed in the general visual response at 12 Hz and harmonics, suggesting that they highlighted higher-level processing of facial expressions.

## Relationship between the perceptual and neural similarity among the six basic emotions

The pattern of results for facial expression discrimination responses was very similar to the behavioral performance, hinting that perceptual similarity among the expressions could explain the behavioral confusability among them as well as the strength of the neural response (Fig. 6A). *Happiness* was the most distinguishable facial expression across three behavioral tasks. It was faster and more effectively identified and detected than any of the other expressions. Similarly, changes to *Happiness* elicited a large EEG discrimination response, which was characterized by specific scalp topography, distributed dorsally compared with the other basic emotions. On the contrary, *Sadness* and *Anger* were easily confused with other expressions and resulted in low accuracy and long detection time. These results agreed with the findings of the lowest discrimination response for changes of *Sadness* and *Anger*. It is noteworthy that *Fear* was the most confusable expression in the forced-choice face labeling task and the expression similarity rating task. The discrimination of *Fear*, however, reached a better performance in the expression change detection task and a relatively large discrimination response for changes to fear in EEG recording. In addition, when fearful faces masked the other expressions, these led to decreased expression discrimination responses. The z-transformed values for the ES and baseline-corrected amplitudes for each expression pair (averaged across trials and subjects) are arranged into a similarity matrix (Fig. 6A).

To directly test the relation between the perceptual and neural similarity among the different expressions, we calculated the correlations between the ES of the expression detection change task and the neural similarity response measured as the discrimination response among the 15 contrasts over the averaged occipito-temporal ROIs. There was a strong correlation between the neural and perceptual similarity of the basic expressions (group $r = 0.81$, N = 15, $p < 0.0001$, confidence interval [CI] = [0.51; 0.93], Fig. 6B, LMM estimate = 0.43, $SE = 0.08$, $t = 5.65$, $p < 0.001$, Fig. 6C). These high correlations

indicate that when the perceived similarity of the six basic expressions increases, as indicated by smaller ES, the FPVS-EEG discrimination response is weaker. To exclude the possibility that this relationship was due to image characteristics, we also evaluated the relationship between the indexes of perceptual and neural similarity controlling for the image physical dissimilarity index. The strong relationship between the visual and perceptual similarities was still observed (group $r = 0.79$, $p = 0.001$).

Similarly, to account for the asymmetries found in both responses behavioral performance and neural expression discrimination we calculated the correlations among the 30 contrasts. The relation between the perceptual and neural similarity remain strong (group $r = 0.78$, N = 30, $p < 0.001$, CI = [0.58; 0.89], LMM estimate = 0.36, $SE = 0.06$, $t = 5.98$, $p < 0.001$).

## Discussion

We examined discrimination among the six basic facial emotions both with behavioral and neurophysiological measures. We collected implicit and explicit measures of facial expression discrimination by using FPVS while recording EEG or behavioral responses. Three different behavioral tasks revealed similar patterns of expression discrimination, suggesting that expression misjudgments could be at least partially explained by the perceptual similarity among them. During the implicit discrimination response, different deviant facial expressions were interleaved into a series of identical expressive faces at regular intervals (12/9 Hz). Consequently, a robust discrimination response specific to the change in expressions in the EEG spectrum was identified beyond explicit behavioral requirements. This response was precisely localized at the frequency rate of the expression change and was distributed bilaterally over occipito-temporal regions. Obvious expressions such as *Happiness*, *Surprise,* and *Disgust* led to higher discrimination responses than subtle expressions, such as *Sadness* and *Anger,* which produced the weakest discrimination response. Furthermore, the response to specific expressions was modulated by the common expression among which they were presented with more similar expressions leading to weaker discrimination responses. Interestingly, fearful faces showed an asymmetry in the response with high discrimination response when needed to be distinguished among most of the other basic emotions (e.g., Anger, Disgust, and Happiness). Yet, when presented as a common face, fearful expression led to weak discrimination response not only for perceptually similar expressions: Surprise and Sadness, but also for perceptually dissimilar expressions (e.g., Happiness). Similar results were obtained with the behavioral data. These observations were confirmed by a correlational analysis revealing that the perceptual similarity

among the expressions was highly correlated (r > 0.80) with the discrimination response in the EEG, with a larger response for expression contrasts which were more distinct (less similar).

## A robust and sensitive index of brief expression changes

### Behavioral evidence

In both behavioral and neutral FPVS measures, clear and robust expression discrimination responses were reliably found across participants even at a short face stimulus SOA (83.33 ms). The short presentation of the stimuli forces the detection of expression change to be done at a single glance. The target expression is backward- and forward-masked by other expressive faces in FPVS, which could impair elaborate processing of the stimuli and further isolate the detection procedure from general perceptual and cognitive functions (Dzhelyova et al., 2017). Despite discrepancies found in previous behavioral studies regarding the time necessary to process facial expressions, 50 ms presentation or longer could guarantee the effective distinguishable discrimination performance for different expressions (Neath & Itier, 2014). In agreement with these findings, during our expression detection task, participants show high accuracy for the detection of all expression contrasts, with an average ACC of 80.59%. Most importantly, none of the contrasts showed accuracy lower than chance level. Yet, although we observed high correlations with explicit tasks measuring the similarity among expressions frequently used in the literature—similarity rating task and the labeling of expressions (Adolphs, 1999; Calvo & Lundqvist, 2008; Hamann & Adolphs, 1999; Palermo & Coltheart, 2004; Tottenham, et al., 2009), compared with standard behavioral findings, there were some clear differences in the discrimination of fearful and happy faces. The discrepancy in the discrimination of these two expressions could be accounted for by several factors. In the explicit behavioral tasks, all emotional stimuli were presented isolated until an explicit forced evaluation is provided. The unlimited presentation duration can provide enough time to process in-depth the perceived stimuli rather than just to rely on perceptual encoding. Therefore, the superiority effect for happiness could be explained not only due to its distinctiveness of physical features (e.g., open mouth, Leppänen & Hietanen, 2004, 2007) but also due to the influence of other cognitive processes, such as a higher frequency of happy expressions occurrences than the other negative expressions in everyday life (Öhman et al., 2001) or an increased difficulty to discriminate among the other negative expressions (Leppänen & Hietanen, 2004). On the contrary, the fast presentation of serial stimuli in FPVS could effectively rule out the possible influence of elaborate processing on emotion perception. Alternatively, it is

possible that the tasks are tapping into different levels of expression categorization and hence different neural processes. Admittedly, the discrimination of the facial expression does not necessarily require expression recognition, while the expression labeling task, for example, requires recognizing the expression and retrieving the emotional label. Yet, language can have an impact on the explicit categorization of facial expressions: A meta-analysis revealed that despite all expressions being recognized well above chance, the recognition rate varied substantially as a function of the language that the participants spoke, with higher recognition rates for Indo-European languages compared to non-Indo-European (Nelson & Russell, 2013). This evidence raises questions about what is facial expression recognition and does recognition require a verbal label? Given these findings, the results could challenge the traditional understanding of expert expression processing and open a window to evaluate human expertise in expression perception under tight temporal constraints.

### Neural evidence

In addition to the explicit discrimination and detection of changes in facial expressions among the basic emotions, we provided a robust implicit measure of expression discrimination. Despite the high interest in the topic and the discovery of common (Haxby et al., 2000) and separable (Damasio et al., 2000; Blair et al., 1999) neural mechanism for processing the facial expressions associated with the six basic emotions, currently, there is little consensus whether the six basic expressions have a clear "brain marker" (see Barrett & Wager, 2006 and for opposing view Vytal & Hamann, 2010). Finding a specific pattern of activation for each expression that distinguishes among the basic emotions is challenging using typical neuroscience approaches (e.g., fMRI; EEG in response to transient events, i.e., event-related potentials). Here with FPVS coupled with EEG, we were able to provide for the first time sensitive neural discrimination responses between all facial expressions, at both the group and individual level. This success is due to several factors: the high sensitivity of the FPVS-EEG approach, the objective quantification of the response in the frequency-domain, but also because, rather than measuring separately the response to each of these expressions (against a uniform visual field), the fact that the paradigm directly contrasts two facial expressions, isolating their difference. Due to its high sensitivity and objectivity, this approach, therefore, opens an avenue to investigate difficult to test populations, such as infants, young children, patients with socioemotional dysfunctions (see Leleu et al., 2019; Van der Donck et al., 2019, 2020), or cross-cultural samples where language barriers could account for at least some of the discrepancies observed in expression perception.

The periodic changes of the expressions led to a robust discrimination response distributed over bilateral occipito-

temporal sites. Hemispheric specialization for facial expressions has been a topic of debate for several decades (Borod, 1998; Etcoff, 1984; Davidson, 1992; Gazzaniga & Smylie, 1990; Gur, Skolnicj & Gur, 1994; Harris et al., 2012; Mandal, Tandon & Asthana, 1991; Sato et al., 2004; Tsuchiya et al., 2008). Indeed, the right hemisphere involvement in processing of facial expressions as evidenced by lesion studies and neuroimaging (Borod, et al., 1998; Etcoff, 1984; Gazzaniga & Smylie, 1990; Harris et al., 2012; Sato et al., 2004; Tsuchiya et al., 2008) has been challenged if the onset of injury is controlled (Abbott et al., 2014). In our study, we observed changes in the hemispheric specialization for the different expressions but also within participants, the pattern of hemispheric dominance varied depending on the discriminated expression, which can partially explain the absent lateralization.

One might ask, which are the structures responsible for coding of the facial expressions? Admittedly, it is difficult to estimate the neural sources leading to the occipito-temporal distribution due to the inversion problem characteristic for the EEG method. Yet, a distributed pattern of activation within face-selective regions, particularly the pSTS (Said et al., 2010; Sormaz et al., 2016) has been linked to the perceptual similarity among expressions. The distributed representation of facial expression, however, also might extend beyond this region. Using intracerebral recordings, facial expressions can be decoded in the fusiform face area (FFA), yet evidence that the perceptual similarity is linked to the activation within the FFA is lacking (Sormaz et al., 2016), possibly due to low sensitivity of fMRI decoding approaches.

The occipito-temporal distribution was common to all expressions, except for changes to Happiness, which had a specific distribution spreading dorsally. This particular spatial signature on the scalp for Happiness hints at least at partially separated neural population coding for this expression and was previously observed in the studies investigating changes of expression from neutral to expressive faces (Dzhelyova et al., 2017; Leleu et al., 2018). These findings are found using two data sets of posed expressions (KDEF and stimulus set used in Durand et al., 2007; Leleu et al., 2018 ), yet in future studies, it is important to assess their generalizability to more natural sets of stimuli such as subtle and/or dynamic facial expressions.

### The relation between perceptual and neural similarity

The amplitude of the discrimination response to changes in expressions varied depending on the perceptual similarity among the expressions which need to be discriminated. Both behavioral tasks and EEG measurements suggested that the subtle expressions produced weaker responses than more explicit expressions. Generally, the discrimination of changes to *Happiness, Disgust,* and

*Surprise* elicited the largest oddball response in EEG and implied a quicker detection compared with the other expressive faces, possibly due to larger perceptual discrepancy (e.g., open mouth, Leppänen & Hietanen, 2004, 2007). On the contrary, *Sadness* and *Anger*, which were easily confused with other expressions (Bombari et al., 2013; Calvo & Lundqvist 2008; Palermo & Coltheart, 2004), were related to decreased neural discrimination response. Fearful expressions fell somewhere in between. Interestingly, *Fear* did not produce the smallest discrimination response often observed in behavioral studies. From an evolutionary perspective, fearful expressions are crucial for survival and thus attention grabbing expression (Öhman et al., 2001; Wilson & MacLeod, 2003; Yiend & Mathews, 2001).

Most importantly, the neural measures of expression discrimination were highly correlated with their corresponding detection performance in a fast presentation nonperiodic oddball paradigm. This relation could not be explained solely by the image characteristics, because the image dissimilarity indices were not related to the EEG or behavioral responses obtained with the FPVS paradigm. Furthermore, we observed asymmetries in some pairs of expressions depending on whether they were presented as common or deviant stimuli. Compared with earlier studies investigating the relationship between the perceptual and neural measures, our results revealed stronger correlations possibly since the neural similarity was estimated directly by comparing the discrimination response between a pair of expressions without any post calculations. For a comparison purpose, the correlation between the neural similarity extracted through decoding approaches of the basic expressions and the mean similarity rating of behavioral tasks reached only moderate relationships in previous studies (e.g., Said et al., 2010, r = 0.36; Saarimäki et al., 2015, r = 0.43). In addition, the high correlation found here could be explained by the similar structure of the tasks relying predominantly on perceptual discrimination, while in previous studies neural similarities are compared to perceptual similarities based on behavioral tasks (e.g., similarity ratings), which might be influenced not only perceptual similarity but also other higher cognitive processes.

### Asymmetries in facial expression processing

The last important finding of our study is that we observed context-dependent modulation of the responses depending on the expression presented as a common or a deviant face. These asymmetries were present both at the behavioral and the neuronal levels. Yet, it seemed that participants were more affected at a behavioral level than the implicit neuronal response, suggesting that these asymmetries are probably based on

higher-order conceptual or decisional factors reflected in the explicit judgments.

Particularly strong asymmetries were observed for fearful faces. Fearful faces were well discriminated when presented as a deviant face. Yet, when presented as a common face, they led to reduced discrimination not only for perceptually similar expressions but also for perceptually different expressions. This finding is well-aligned with a recent study that used repetition suppression design and observed that fearful expression was the expression producing the strongest repetition suppression responses (Turano et al., 2017). In line with the evolutionary explanation of the importance of quick detection of fearful expression, this facial signal also leads to delayed attention disengagement (Georgiou et al., 2005) even in young infants (Peltola et al., 2008). Some evidence supporting this delayed disengagement hypothesis could be found in our discrimination measures: when fearful faces were presented as common face stimuli, a decreased discrimination response was obtained not only to the perceptually similar expressions but also to expressions that are more distinct such as happiness for example.

In addition to fear, other expressions also produced asymmetrical responses when they were presented as a deviant or a common face. Happiness, for example, also was detected faster among negative expressions in explicit behavioral tasks. This could be a result of a "pop-out effect" of this clearly positive expression, a finding often reported in visual search or discrimination tasks (e.g., Calvo & Marrero, 2009; Calvo & Nummenmaa, 2008). Taken together, this evidence highlighted the importance of context in perceiving facial expressions (e.g., Aviezer, Bentin, Dudarev, & Hassin, 2011; Barrett, Mesquita, & Gendron, 2011; Carroll & Russell, 1996) and emphasized its importance for interpreting and decoding facial expressions.

## Conclusions

With a few minutes of recordings, we provided a robust and sensitive discrimination response for each expression of the basic emotions over occipito-temporal sites. In contrast to behavioral studies, expressions of Happiness did not differ quantitatively as much in EEG as behavioral subjective judgments suggest, but it differed from all other expressions in terms of scalp topography. Fearful faces showed an asymmetry in the response with high discrimination response when needed to be recognized among the other basic emotions; however, when presented as a common face led to weaker discrimination response even for perceptually dissimilar expressions. Despite these differences with behavioral studies, we found a tight relationship of the similarity matrices across facial expressions obtained for implicit EEG responses and behavioral

explicit measures collected under the same tight temporal constraints. Given the implicitness of the task, the short recording time, and the robustness of the results, applying this new approach in developmental, intercultural, and clinical populations could pave the way for more thoroughly understanding facial expression discrimination.

## References

Turano, M. T., Lao, J., Richoz, A. R., Lissa, P. D., Degosciu, S. B., Viggiano, M. P., & Caldara, R. (2017). Fear boosts the early neural coding of faces. *Social Cognitive and Affective Neuroscience, 12*(12), 1959-1971.

Abbott, J. D., Wijeratne, T., Hughes, A., Perre, D., & Lindell, A. K. (2014). The influence of left and right hemisphere brain damage on configural and featural processing of affective faces. *Laterality: Asymmetries of Body, Brain and Cognition*, 19(4), 455-472.

Adolphs, R. (1999). The human amygdala and emotion. *The Neuroscientist, 5*(2), 125-137.

Alonso-Prieto E, Van Belle G, Liu-Shuang J, Norcia AM, & Rossion B. (2013). The 6Hz fundamental stimulation frequency rate for individual face discrimination in the right occipito-temporal cortex. *Neuropsychologia* 51, 2863-2875.

Aviezer, H., Bentin, S., Dudarev, V., & Hassin, R. R. (2011). The automaticity of emotional face-context integration. *Emotion, 11*(6), 1406-1414. https://doi.org/10.1037/a0023578

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. *Psychological Science in the Public Interest, 20*(1), 1-68. https://doi.org/10.1177/1529100619832930

Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in Emotion Perception. *Current Directions in Psychological Science, 20*(5), 286-290. https://doi.org/10.1177/0963721411422522

Barrett, L. F., & Wager, T. D. (2006). The structure of emotion: Evidence from neuroimaging studies. *Current Directions in Psychological Science*, 15(2), 79-83.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., … Bolker, M. B. (2015). Package 'lme4'. *Convergence* , 12 (1), 2.

Blair, R. J. R., Morris, J. S., Frith, C. D., Perrett, D. I., & Dolan, R. J. (1999). Dissociable neural responses to facial expressions of sadness and anger. *Brain*, 122(5), 883-893.

Bombari, D., Schmid, P. C., Schmid Mast, M., Birri, S., Mast, F. W., & Lobmaier, J. S. (2013). Emotion recognition: The role of featural and configural face information. *The Quarterly Journal of Experimental Psychology*, 66(12), 2426-2442.

Borod, J. C., Cicero, B. A., Obler, L. K., Welkowitz, J., Erhan, H. M., Santschi, C., … Whalen, J. R. (1998). Right hemisphere emotional perception: evidence across multiple channels. *Neuropsychology*, 12(3), 446.

Calvo, M. G., & Lundqvist, D. (2008). Facial expressions of emotion (KDEF): Identification under different display-duration conditions. *Behavior Research Methods*, 40(1), 109-115.

Calvo, M. G., & Marrero, H. (2009). Visual search of emotional faces: The role of affective content and featural distinctiveness. *Cognition*

*and Emotion, 23*(4), 782-806. https://doi.org/10.1080/02699930802151654

Calvo, M. G., & Nummenmaa, L. (2008). Detection of emotional faces: salient physical features guide effective visual search. *Journal of Experimental Psychology: General, 137*(3), 471.

Carroll, J. M., & Russell, J. A. (1996). Do facial expressions signal specific emotions? Judging emotion from the face in context. *Journal of Personality and Social Psychology, 70*(2), 205.

Centeno, M., Koepp, M. J., Vollmar, C., Stretton, J., Sidhu, M., Michallef, C., … Duncan, J. S. (2014). Language dominance assessment in a bilingual population: validity of fMRI in the second language. *Epilepsia, 55*(10), 1504-1511.

Clark-Polner, E., Johnson, T. D., & Barrett, L. F. (2017). Multivoxel pattern analysis does not provide evidence to support the existence of basic emotions. *Cerebral Cortex, 27*(3), 1944-1948.

Damasio, A. R., Grabowski, T. J., Bechara, A., Damasio, H., Ponto, L. L., Parvizi, J., & Hichwa, R. D. (2000). Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nature Neuroscience, 3*(10), 1049.

Davidson, R. J. (1992). Anterior cerebral asymmetry and the nature of emotion. *Brain and Cognition, 20*(1), 125-151.

Durand, K., Gallay, M., Seigneuric, A., Robichon, F., & Baudouin, J. Y. (2007). The development of facial emotion recognition: The role of configural information. *Journal of Experimental Child Psychology, 97*(1), 14-27.

Dzhelyova, M., Jacques, C., & Rossion, B. (2017). At a single glance: fast periodic visual stimulation uncovers the spatio-temporal dynamics of brief facial expression changes in the human brain. *Cerebral Cortex, 27*(8), 4106-4123.

Ekman, P. (1993). Facial expression and emotion. *American Psychologist, 48*(4), 384.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*(2), 124.

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychological Bulletin, 128*(2), 203.

Etcoff, N. L. (1984). Perceptual and conceptual organization of facial emotions: Hemispheric differences. *Brain and Cognition, 3*(4), 385-412.

Gazzaniga, M. S., & Smylie, C. S. (1990). Hemispheric mechanisms controlling voluntary and spontaneous facial expressions. *Journal of Cognitive Neuroscience, 2*(3), 239-245.

Georgiou, G., Bleakley, C., Hayward, J., Russo, R., Dutton, K., Eltiti, S., & Fox, E. (2005). Focusing on fear: Attentional disengagement from emotional faces. *Visual Cognition, 12*(1), 145-158.

Gur, R. C., Skolnick, B. E., & Gur, R. E. (1994). Effects of emotional discrimination tasks on cerebral blood flow: Regional activation and its relation to performance. *Brain and Cognition, 25*(2), 271-286.

Hamann, S. B., & Adolphs, R. (1999). Normal recognition of emotional similarity between facial expressions following bilateral amygdala damage. *Neuropsychologia, 37*(10), 1135-1141.

Harris, R. J., Young, A. W., & Andrews, T. J. (2012). Morphing between expressions dissociates continuous from categorical representations of facial expression in the human brain. *Proceedings of the National Academy of Sciences, 109*(51), 21164-21169.

Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., … Ramadge, P. J. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron, 72*(2), 404-416.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4*(6), 223-233.

Hochberg, J., & TAMHANE, A. C. (1987). *Multiple comparison procedures* (No. 519.535 H655m Ej. 1). John Wiley & Sons.

Izard, C. E. (1994). Innate and universal facial expressions: evidence from developmental and cross-cultural research. *Psychological Bulletin,* 115 (2), 288-299.

Jack, R. E., Sun, W., Delis, I., Garrod, O. G. B., & Schyns, P. G. (2016). Four not six: Revealing culturally common facial expressions of emotion. *Journal of Experimental Psychology: General, 145*(6), 708-730. https://doi.org/10.1037/xge0000162

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience, 2,* 4.

Leleu, A., Dzhelyova, M., Rossion, B., Brochard, R., Durand, K., Schaal, B., & Baudouin, J. Y. (2018). Tuning functions for automatic detection of brief changes of facial expression in the human brain. *Neuroimage, 179,* 235-251.

Leleu, A., Favre, E., Yailian, A., Fumat, H., Klamm, J., Amado, I., … Demily, C. (2019). An implicit and reliable neural measure quantifying impaired visual coding of facial expression: evidence from the 22q11. 2 deletion syndrome. *Translational Psychiatry, 9*(1), 67.

Leppänen, J. M., & Hietanen, J. K. (2004). Positive facial expressions are recognized faster than negative facial expressions, but why?. *Psychological Research, 69*(1-2), 22-29.

Leppänen, J. M., & Hietanen, J. K. (2007). Is there more in a happy face than just a big smile?. *Visual Cognition, 15*(4), 468-490.

Liu-Shuang, J., Norcia, A. M., & Rossion, B. (2014). An objective index of individual face discrimination in the right occipito-temporal cortex by means of fast periodic oddball stimulation. *Neuropsychologia, 52,* 57-72.

Lochy, A., Van Reybroeck, M., & Rossion, B. (2016). Left cortical specialization for visual letter strings predicts rudimentary knowledge of letter-sound association in preschoolers. *Proceedings of the National Academy of Sciences, 113*(30), 8544-8549.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska directed emotional faces (KDEF). *CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, 91,* 630.

Mandal, M. K., Tandon, S. C., & Asthana, H. S. (1991). Right brain damage impairs recognition of negative emotions. *Cortex, 27*(2), 247-253.

Matsumoto, D. (2001). Culture and emotion. In D. Matsumoto (Ed.), *The handbook of culture and psychology* (pp. 171–194). New York: Oxford University Press

Maxwell, J., & Davidson, R. (2004). Unequally masked: Indexing differences in the perceptual salience of" unseen" facial expressions. *Cognition & Emotion, 18*(8), 1009-1026.

McCarthy, G., & Wood, C. C. (1985). Scalp distributions of event-related potentials: an ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology/ Evoked Potentials Section, 62*(3), 203-208.

Milders, M., Sahraie, A., & Logan, S. (2008). Minimum presentation time for masked facial expression discrimination. *Cognition and Emotion, 22*(1), 63-82.

Neath, K. N., & Itier, R. J. (2014). Facial expression discrimination varies with presentation time but not with fixation on features: a backward masking study using eye-tracking. *Cognition & Emotion, 28*(1), 115-131.

Nelson, N. L., & Russell, J. A. (2013). Universality revisited. *Emotion Review, 5*(1), 8-15.

Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottereau, B. R., & Rossion, B. (2015). The steady-state visual evoked potential in vision research: a review. *Journal of Vision, 15*(6), 4-4.

Öhman, A., Lundqvist, D., & Esteves, F. (2001). The face in the crowd revisited: a threat advantage with schematic stimuli. *Journal of Personality and Social Psychology, 80*(3), 381.

Palermo, R., & Coltheart, M. (2004). Photographs of facial expression: Accuracy, response times, and ratings of intensity. *Behavior Research Methods, Instruments, & Computers, 36*(4), 634-638.

Palermo, R., O'Connor, K. B., Davis, J. M., Irons, J., & McKone, E. (2013). New tests to measure individual differences in matching and labelling facial expressions of emotion, and their association with ability to recognise vocal emotions and facial identity. *PLoS One*, *8*(6), e68126.

Peltola, M. J., Leppänen, J. M., Palokangas, T., & Hietanen, J. K. (2008). Fearful faces modulate looking duration and attention disengagement in 7-month-old infants. *Developmental Science*, *11*(1), 60-68.

Peterson, M. F., & Eckstein, M. P. (2012). Looking just below the eyes is optimal across face recognition tasks. *Proceedings of the National Academy of Sciences*, *109*(48), E3314-E3323.

Regan, D. (1966). Some characteristics of average steady-state and transient responses evoked by modulated light. *Electroencephalography and Clinical Neurophysiology*, *20*(3), 238-248.

Regan, D. (1989). Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine. Elsevier, New York

Rossion, B., & Boremanse, A. (2011). Robust sensitivity to facial identity in the right human occipito-temporal cortex as revealed by steady-state visual-evoked potentials. *Journal of Vision*, *11*(2), 16-16.

Rossion, B., Prieto, E. A., Boremanse, A., Kuefner, D., & Van Belle, G. (2012). A steady-state visual evoked potential approach to individual face perception: effect of inversion, contrast-reversal and temporal dynamics. *NeuroImage*, *63*(3), 1585-1600.

Russell, J. A. (1993). Forced-choice response format in the study of facial expression. Motivation and Emotion, 17, 41–51.

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

Saarimäki, H., Gotsopoulos, A., Jääskeläinen, I. P., Lampinen, J., Vuilleumier, P., Hari, R., … Nummenmaa, L. (2015). Discrete neural signatures of basic emotions. *Cerebral Cortex*, *26*(6), 2563-2573.

Said, C. P., Moore, C. D., Engell, A. D., Todorov, A., & Haxby, J. V. (2010). Distributed representations of dynamic facial expressions in the superior temporal sulcus. *Journal of Vision*, *10*(5), 11-11.

Sato, W., Kochiyama, T., Yoshikawa, S., Naito, E., & Matsumura, M. (2004). Enhanced neural activity in response to dynamic facial expressions of emotion: an fMRI study. *Cognitive Brain Research*, *20*(1), 81-91.

Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and decoding facial expressions. Psychological Science, 16(3), 184-189.

Sormaz, M., Watson, D. M., Smith, W. A., Young, A. W., & Andrews, T. J. (2016). Modelling the perceptual similarity of facial expressions from image statistics and neural responses. *Neuroimage*, *129*, 64-71.

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., … Nelson, C. (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Research*, *168*(3), 242-249.

Townsend, J., & Ashby, F. G. (1978). Methods of modeling capacity in simple processing systems. Castellan J, Restle F, editors. *Cognitive theory*.

Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. CUP Archive.

Trimmel, K., Sachsenweger, J., Lindinger, G., Auff, E., Zimprich, F., & Pataraia, E. (2017). Lateralization of language function in epilepsy patients: a high-density scalp-derived event-related potentials (ERP) study. *Clinical Neurophysiology, 128*(3), 472-479.

Tsuchiya, N., Kawasaki, H., Oya, H., Howard III, M. A., & Adolphs, R. (2008). Decoding face information in time, frequency and space from direct intracranial recordings of the human brain. *PLoS One*, *3*(12), e3892.

Van der Donck, S., Dzhelyova, M., Vettori, S., Mahdi, S. S., Claes, P., Steyaert, J., & Boets, B. (2020). Rapid neural categorization of angry and fearful faces is specifically impaired in boys with autism spectrum disorder. *Journal of Child Psychology and Psychiatry* https://doi.org/10.1111/jcpp.13201

Van der Donck, S., Dzhelyova, M., Vettori, S. *et al.* Fast Periodic Visual Stimulation EEG Reveals Reduced Neural Sensitivity to Fearful Faces in Children with Autism. *Journal of Autism and Developmental Disorders* 49**,** 4658–4673 (2019). https://doi.org/10.1007/s10803-019-04172-0

Vytal, K., & Hamann, S. (2010). Neuroimaging support for discrete neural correlates of basic emotions: a voxel-based meta-analysis. *Journal of Cognitive Neuroscience*, *22*(12), 2864-2885.

Wilson, E., & MacLeod, C. (2003). Contrasting two accounts of anxiety-linked attentional bias: selective attention to varying levels of stimulus threat intensity. *Journal of Abnormal Psychology*, *112*(2), 212.

Yiend, J., & Mathews, A. (2001). Anxiety and attention to threatening pictures. *The Quarterly Journal of Experimental Psychology Section A*, *54*(3), 665-681.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.