Research Article: New Research | *Sensory and Motor Systems*

The Neural Dynamics of Facial Identity Processing: insights from EEG-Based Pattern Analysis and Image Reconstruction

Dan Nemrodov[1], Matthias Niemeier[1], Ashutosh Patel[1] and Adrian Nestor[1]

[1]*Department of Psychology at Scarborough, University of Toronto, 1265 Military Trail, Toronto, Ontario M1C1A4, Canada*

DOI: 10.1523/ENEURO.0358-17.2018

Received: 12 October 2017

Revised: 11 January 2018

Accepted: 12 January 2018

Published: 29 January 2018

Author contributions: D.N., M.N., and A.N. designed research; D.N., A.P., and A.N. performed research; D.N. and A.N. analyzed data; D.N., M.N., A.P., and A.N. wrote the paper.

Funding: http://doi.org/10.13039/501100000038Gouvernement du Canada | Natural Sciences and Engineering Research Council of Canada (NSERC)

Funding: http://doi.org/10.13039/501100007224Connaught Fund

Conflict of Interest: Authors declare no conflict of interests.

This research was supported by the Natural Sciences and Engineering Research Council of Canada (to A.N. and M.N.) and by a Connaught New Researcher Award (to A.N.).

Correspondence should be addressed to Dan Nemrodov. Tel: +1 (647) 772-0047; E-Mail: dan.nemrodov@utoronto.ca

Cite as: eNeuro 2018; 10.1523/ENEURO.0358-17.2018

The neural dynamics of facial identity processing: insights from EEG-based pattern analysis and image reconstruction

Abbreviated title: The neural dynamics of facial identity processing

Dan Nemrodov[1*], Matthias Niemeier[1], Ashutosh Patel[1], Adrian Nestor[1]

[1]Department of Psychology at Scarborough, University of Toronto, 1265 Military Trail, Toronto, Ontario, M1C1A4, Canada

* To whom correspondence should be addressed.

E-Mail: dan.nemrodov@utoronto.ca Tel : +1 (647) 772-0047

Number of figures: 8; multimedia: 1

Number of words for Abstract: 149; Introduction: 688; Discussion: 1554

Conflict of interest: The authors declare no competing financial interests.

Acknowledgments: This research was supported by the Natural Sciences and Engineering Research Council of Canada (to A.N. and M.N.) and by a Connaught New Researcher Award (to A.N.).

Abstract

Uncovering the neural dynamics of facial identity processing along with its representational basis outlines a major endeavor in the study of visual processing. To this end, here we record human electroencephalography (EEG) data associated with viewing face stimuli; then, we exploit spatiotemporal EEG information to determine the neural correlates of facial identity representations and to reconstruct the appearance of the corresponding stimuli. Our findings indicate that multiple temporal intervals support: facial identity classification, face space estimation, visual feature extraction and image reconstruction. In particular, we note that both classification and reconstruction accuracy peak in the proximity of the N170 component. Further, aggregate data from a larger interval (50-650 ms after stimulus onset) support robust reconstruction results, consistent with the availability of distinct visual information over time. Thus, theoretically, our findings shed light on the time course of face processing while, methodologically, they demonstrate the feasibility of EEG-based image reconstruction.

Keywords ERP, face space, pattern analysis, N170, spatiotemporal dynamics, image reconstruction

Significance statement

Identifying a face is achieved through fast and efficient processing of visual information. Here, we investigate the nature of this information, its specific content and its availability at a fine-grained temporal scale. Notably, we provide a way to extract, to assess and to visualize such information from neural data associated with individual face processing. Thus, the present work accounts for the time course of face individuation through appeal to its underlying visual representations while, also, it provides a first demonstration regarding the ability to reconstruct the appearance of stimulus images from electroencephalography data.

Introduction

Elucidating the dynamics of visual face processing is essential to understanding its underlying mechanisms. To this end, considerable efforts have been devoted to characterizing the time course of face processing especially through the use of electroencephalography (EEG) and magnetoencephalography (MEG) given the temporal resolution of these methods. Accordingly, much is known about the temporal profile of face processing as reflected by either traditional event-related potentials (ERP) (Bentin & Deouell, 2000; Itier & Taylor, 2002; Huddy, Schweinberger, Jentzsch, & Burton, 2003; Tanaka, Curran, Porterfield, & Collins, 2006; Rossion & Caharel, 2011; Zheng, Mondloch, & Segalowitz, 2012), or by spatiotemporal patterns (Liu, Agam, Madsen, & Kreiman, 2009; Cichy, Pantazis, & Oliva, 2014; Vida, Nestor, Plaut, & Behrmann, 2017). Comparatively less is known about the visual representations underlying the dynamics of face processing, especially as related to facial identity. To shed light on this issue, the present work employs an image-reconstruction paradigm (Miyawaki et al., 2008; Naselaris, Prenger et al., 2009; Nishimoto et al., 2011; Cowen et al., 2014; Chang & Tsao, 2017) seeking to approximate the visual appearance of individual faces from spatiotemporal EEG patterns. Concretely, this work aims to answer whether the visual information involved in face identification can be recovered from EEG signals and, further, whether such information can support the characterization of neural-based face space along with the reconstruction of individual face images.

Recent applications of pattern analysis have focused on the temporal profile of face discrimination at the category (Carlson, Tovar, Alink, & Kriegeskorte, 2013; Van de Nieuwenhuijzen et al., 2013; Cauchoix, Barragan-Jason, Serre, & Barbeau, 2014; Cichy, Pantazis, & Oliva, 2014; Kaneshiro et al., 2015) and the exemplar level (Davidesco et al., 2014; Ghuman, et al., 2014). For instance, expression-invariant identity discrimination has been carried out using MEG (Vida, Nestor, Plaut, & Behrmann, 2017), electrocorticography (Ghuman et al., 2014) and EEG (Nemrodov, Niemeier, Mok, & Nestor, 2016). These studies found multiple, distinct temporal windows sensitive to facial information and, consistent with results from monkey neurophysiology (Hung, Kreiman, Poggio, & DiCarlo, 2005; Freiwald, Tsao, & Livingstone, 2009) and human psychophysics (Lehky, 2000; Tanaka & Curran, 2001; Crouzet, Kirchner, & Thorpe, 2010), have estimated an early onset for such sensitivity (Ghuman et al., 2014; Nemrodov et al., 2016). Further, the cortical source of this information was attributed primarily to the fusiform gyrus (FG) in line with homologous investigations of face identification using functional magnetic resonance imaging (fMRI) (Nestor, Plaut, & Behrmann, 2011; Goesaert & Op de Beeck, 2013; Anzellotti, Fairhall, & Caramazza, 2014).

Yet, the representational basis of facial identity that allows successful discrimination from neural data remains to be elucidated. Arguably, neural patterns elicited by face perception speak to the properties of a representational face space (Valentine, 1991), or, more generally, of a representational similarity space (Kriegeskorte, Mur, & Bandettini, 2008). In an effort to clarify the nature of such representations, recent fMRI work (Nestor, Plaut, & Behrmann, 2016) has combined the study of face space and neural-based image reconstruction. Specifically, this work has derived visual features from the structure of FG-based face space and, then, used such features for facial image reconstruction. However, this work did not consider the temporal aspects of face processing, neither did it assess the invariant structure of a facial identity space – surprisingly, we note that face space invariance over common image transformations (e.g., viewpoint, expression) has rarely been explicitly investigated (but see (Newell, Chiroro, & Valentine, 1999; Blank & Yovel, 2011)).

To address the issues above, the current work aims to derive face space constructs from the EEG signal associated with consecutive time windows separately for different facial expressions (i.e., neutral and happy). Then, it reconstructs the appearance of one set of faces (e.g., happy) based on the structure of the face space derived for the other faces (e.g., neutral). This demonstration provides evidence that the spatiotemporal information of EEG patterns is rich enough to support: (i) identity-level face discrimination; (ii) neural-based face space estimation; (iii) visual feature synthesis, and (iv) facial image reconstruction. Further, this work characterizes the neural dynamics of expression-invariant face processing while, more generally, it provides proof of concept for the possibility of EEG-based image reconstruction.

Materials and Methods

Participants

Thirteen healthy adults (6 males, 7 females; age range: 18–27 years) with normal or corrected-to-normal vision were recruited from the (Author University) community to participate in the study in exchange for monetary compensation. All participants provided informed consent and all experimental procedures were approved by the Research Ethics Board at (Author University). The data of all participants were included in the analyses reported below.

Stimuli

A total of 140 face images of 70 individuals, each displaying a neutral and a happy expression were used as experimental stimuli. Out of these, 108 images of 54 unfamiliar males were selected from three databases: AR (Martinez & Benavente, 1998), Radboud (Langner, Dotsch, Bijlstra, Wigboldus, Hawk, & van Knippenberg, 2010) and FEI (Thomaz & Giraldi, 2010). The remaining 32 images displayed faces of 6 famous male and 10 female individuals selected from open access sources. To be clear, unfamiliar male face stimuli are the focus of the present investigation while female faces were used as go trials in a go/no-go gender recognition task (see Experimental design) and additional famous male faces were included to promote alertness (however, no results are reported for them below due to the smaller stimulus set that precluded a separate examination of famous face recognition).

All images featured young adult Caucasian individuals with frontal view, gaze and illumination. The stimuli were selected so that no facial accessories, hair or makeup obscured the internal features of the face and so that all happy expressions displayed an open-mouth smile. These images were: (a) scaled uniformly and aligned with roughly the same position of the eyes and the nose; (b) cropped to eliminate background; (c) normalized with the same mean and root mean square (RMS) contrast values separately for each color channel in CIEL*a*b* color space, and (d) reduced to the same size (95 X 64 pixels).

Experimental design

Prior to EEG testing participants were administered the Cambridge Face Memory Test (CFMT), (Duchaine & Nakayama, 2006) to confirm that their face processing abilities fall within the range of normal performance for young adults (Bowles et al., 2009). Participants also completed the Vividness of Visual Imagery Questionnaire 2 (VVIQ-2) (Marks, 1995) along with a custom familiarity-rating famous face questionnaire.

During EEG sessions participants were seated in a dimly lit room at a viewing distance of 80 cm from an LCD monitor (resolution: 1920 X 1080, refresh rate: 60Hz). The participants were instructed to perform a go/no-go gender recognition task by pressing a designated key every time they saw a female face, irrespective of expression. The experiment consisted of 32 blocks of stimulus presentation divided into 2 sessions carried out on different days. In each session, experimental blocks were preceded by one training block, subsequently discarded from all analyses. The blocks were separated by self-paced breaks.

Over the course of any given block, each image of a male face was presented twice and each image of a female face was presented once, for a total of 260 trials. Images were presented in a pseudorandom order under the constraint that no facial identity would appear consecutively. Each stimulus was presented in the center of the screen against a black background and subtended a visual angle of 3.2 x 4.9. A stimulus display lasted for 300 ms and it was followed by a fixation cross for a duration ranging randomly between 925–1075 ms. Each session, including participant and equipment setup, lasted around 2.5 hours. Stimulus presentation and response recording relied on Matlab 9.0 (Mathworks, Natick, MA) and Psychtoolbox 3.0.8 (Brainard, 1997; Pelli, 1997).

EEG acquisition and preprocessing

EEG data were recorded using a 64-electrode Biosemi ActiveTwo EEG recording system (Biosemi B.V., Amsterdam, Netherlands). The electrodes were arranged according to the International 10/20 System. The electrode offset was kept below 40 mV. The EEG and EOG were low-pass filtered using a fifth order sinc filter with a half-power cutoff at 204.8 Hz and then digitized at 512 Hz with 24 bits of resolution. All data were digitally filtered offline (zero-phase 24 dB/octave Butterworth filter) with a bandpass of 0.1–40 Hz.

Next, data were separated into epochs, from 100 ms prior to stimulus presentation until 900 ms later, and baseline-corrected. Epochs corresponding to go trials (i.e., female face stimuli) and epochs containing false alarms were discarded from further analysis. Further, noisy electrodes were interpolated if necessary (no more than 2 electrodes per subject) and epochs were re-referenced to the average reference. In addition, prior to univariate ERP analysis, data were cleaned of ocular artifacts using Infomax ICA (Delorme, Sejnowski, & Makeig, 2007).

After removing trials containing artifacts and/or false alarms, an average of 96% of trials (range: 75%-100% across participants) were selected for further analysis. In particular, we note that relatively few trials contained false alarms as participants performed the go/no-go recognition task at ceiling (accuracy range: 98.1% - 100% across participants).

All analyses were carried out using Letswave 6 (http://nocions.webnode.com/letswave) (Mouraux & Iannetti, 2008), MATLAB 9.0 and the G*Power toolbox (Faul, et al., 2007).

Univariate ERP analyses

Twelve electrodes situated over homologue occipitotemporal areas (P5, P7, P9, PO3, PO7, O1 on the left and P6, P8, P10, PO4, PO8, O2 on the right) were selected for ERP analysis. Their selection was motivated by the relevance of these electrodes for face processing as revealed by ERP analysis (e.g., robust N170 amplitudes). Data from electrodes over the left and right sites were then averaged separately creating two sets of signals, one for each hemisphere.

For the purpose of univariate tests, the data was averaged for each unfamiliar face identity across expressions. Then, P1, N170 and N250 components were visually identified on a grand-average plot and a two-way repeated measures ANOVA over facial identities and hemispheres was conducted on maximum amplitudes in the 70-180 ms range for P1, and on minimum amplitudes in the 160-250 and 260-350 ms ranges for N170 and N250, respectively. Greenhouse-Geisser correction was applied in case of violation of the sphericity assumption.

Pattern classification analysis

Epochs were linearly detrended, z-scored across time and electrodes, and corrected for outliers (i.e., values exceeding 3SD from the mean were thresholded at ±3 to minimize the deleterious effect of extreme values on SVM-based pattern classification). Then, all epochs were normalized to the same range (0–1) and mean (i.e., 0.5). In order to boost the signal-to-noise ratio (SNR) of spatiotemporal patterns (Grootswagers, Wardle, & Carlson, 2016) multiple epochs pertaining to the same condition were averaged into ERP traces. Specifically, all epochs corresponding to the same image stimulus across two consecutive blocks, for a maximum of 4 epochs, were averaged together resulting in 16 separate traces per stimulus. Further, this procedure was also instrumental in handling missing data following trial removal and in balancing the number of observations for pattern classification. Specifically, since it is possible that both trials associated with a given stimulus in a given block be removed (e.g., due to artifacts), averaging data across single blocks can lead to different numbers of observations for different stimuli. However, averaging data across pairs of consecutive blocks (i.e., 1-4 trials per stimulus following data removal) ensured that equal numbers of observations can be constructed for each pair of stimuli undergoing pattern classification.

Next, to increase the robustness of pattern analyses epochs were divided into temporal windows containing 5 consecutive bins (5 bins*1.95 ms≈10 ms)(Blankertz et al., 2011). For each window, data were concatenated into observations – for instance, data across selected occipitotemporal electrodes (see Univariate ERP analyses) were concatenated into 60-dimension observations (5 time bins X 12 electrodes). These observations were constructed for the purpose of pattern analyses in time, window by window. In addition, we constructed more inclusive observations that contain all time bins between 50 ms and 650 ms after stimulus presentation (3684-dimension vectors: 307 bins x 12 electrodes), and, thus, both early and late information relevant for face processing (Ghuman et al., 2014; Vida et al., 2017). These higher-dimensional observations were constructed for the purpose of temporally cumulative analyses able to exploit more extensive information over time.

Pairwise discrimination of facial identity was conducted with the aid of linear Support Vector Machine (SVM) classification (c=1) (SVMLIB 3.22, Chang & Lin, 2011) and leave-one-out cross-validation (i.e., one out of 16 pairs of observations was systematically left out for testing while the remaining 15 were used for training). Concretely, 1431 pairs of stimuli were classified across 16 leave-one-out cross-validation iterations and then classification accuracy was obtained for each pair by averaging across these iterations. Classification was conducted for all pairs of facial identities in two ways: (a) within expression - the classifier was trained separately on one expression, happy or neutral, and tested on the same expression; and (b) across expression - the classifier was trained on one expression and tested on the other. Significance of classification accuracy was assessed via a permutation test, by randomly shuffling identity labels 10^3 times, and correcting for multiple comparisons across temporal windows using the false discovery rate (FDR).

The analyses above were conducted for data averaged across all participants (i.e., observations were constructed from averaged ERP traces). In addition, similar discrimination analyses were performed for data from individual participants. The significance of the overall classification accuracy for this analysis was assessed using one-sample two-tailed t-tests against chance across participants separately for within- and across-expression discrimination.

Further, given that the three face databases used here for the purpose of stimulus selection and design sample different populations (i.e., individuals of different nationalities) we anticipated a possible effect of database. However, it is important to establish that any effects found here are not solely due to such differences. To assess this possibility we also examined pairwise classification results for faces extracted from the same databases and, separately, for faces extracted from different databases – that is, each face is classified only relative to faces from the same database, for a total of 472 pairs, or only relative to faces from different databases, for a total of 1918 pairs.

Relationship with behavioral performance

Average pairwise face discrimination, as indexed by temporally cumulative analysis, was correlated with behavioral markers of performance (from CFMT and VVIQ2 tests) for each participant.

For a finer-grained analysis in the temporal domain, pairwise face discrimination was correlated, window by window, with behavioral data from a previous study (Nestor, Plaut, & Behrmann, 2013, Experiment 1) in which participants rated the visual similarity of pairs of faces. Specifically, in this study participants viewed a larger set of faces, including the 108 unfamiliar face images used here, and rated the similarity of each face with every other face on a 5-point scale across a change in expression (i.e., one face in a pair displayed a neutral expression and the other a happy expression). Ratings from 22 healthy adult participants were then averaged to deliver an estimate of behavioral face similarity. Here, this estimate was compared to its neural counterpart from across-expression classification carried out on group-averaged ERP traces. Concretely, behavioral and neural-based estimates were related via Pearson correlation and tested for significance via a permutation test (10^3 permutations for each temporal window; FDR correction across windows).

Estimation of EEG-based face space

Face space constructs were derived by applying multidimensional scaling (MDS) to EEG-based estimates of face discrimination. Specifically, classification accuracies for pairs of facial identities were organized into a dissimilarity matrix in which each cell estimates the discriminability of a pair of faces; then, all values were linearly scaled between 0-1 and metric MDS was applied to approximate a corresponding space. The dimensionality of such spaces was restricted to 20 since that was found sufficient to account for most variance in the data (e.g., over 90% for temporally cumulative analyses). This procedure was conducted on within-expression estimates of discrimination, separately for each expression, resulting in separate spaces for faces with neutral and happy expressions.

Next, we examined face space invariance to image changes introduced by emotional expression. To this end, the fit between neutral and happy face spaces was estimated by aligning one space to the other, via Procrustes transformation, and measuring the badness of fit as the sum of squared errors (SSE) between the two spaces. Significance testing was then conducted through a permutation test for multidimensional spaces (Jackson, 1995): the labels of each point in one of the two spaces was randomly shuffled and the resulting space was fit to the intact one as above. This procedure was carried out for a total of 10^3 permutations, by leaving intact each of the two spaces half of the time while permuting the other space, and permutation-based SSE estimates were computed each time.

Reconstruction approach

The current procedure broadly follows a recently developed approach to facial image reconstruction designed to exploit spatial information in fMRI patterns (Nestor, Plaut, & Behrmann, 2016). This procedure capitalizes on the structure of neural-based face space for the purpose of feature derivation and image reconstruction. Here, we deployed this procedure to capture spatiotemporal information in EEG patterns and, further, to examine the ability of expression-invariant visual information to support image reconstructions of facial identity. This procedure was conducted in a sequence of steps as follows.

First, visual features accounting for face space topography were separately derived for each dimension of EEG-based face space. These features were computed as weighted sums of image stimuli following a strategy similar to reverse correlation / image classification – see (Murray, 2011) for a review and (Smith, Gosselin & Schyns, 2012) for applications to EEG data. Hence, here they are referred to as ‘classification images’ (CIM). Briefly, face stimuli, following their color conversion to CIEL*a*b*, were summed proportionally to their z-scored coordinates on a given dimension of face space. The resulting CIM (i.e., a triplet of images corresponding to L*, a* and b* channels) amounts to a linear approximation of the visual information responsible for organizing faces along that specific dimension. Thus, for each expression this procedure delivered a total of 20 different CIMs, one for each corresponding dimension of face space.

Second, since not all dimensions may encode systematic visual information, feature/dimension selection was used to identify subspaces relevant for reconstruction purposes. To this end, CIMs corresponding to each dimension were assessed regarding the inclusion of significant information. Specifically, for each dimension, permutation-based CIMs were generated after randomly shuffling the coefficients associated with stimulus images. Then, pixel intensities in the true CIM were compared to the corresponding intensities of pixels in permutation-based CIMs 10^3 permutations; FDR correction across pixels and color channels) and only CIMs that contained pixel values significantly different from chance were considered for reconstruction purposes.

Third, the coordinates of a target face were estimated in an expression-specific face space. To be clear, the estimation of face space and its CIMs were carried out using all facial identities but one. Then, the left-out face was projected in this space based on its similarity with the other faces. Thus, the procedure ensured that features were not generated from the reconstruction target guarding against circularity.

Last, the target face was constructed through a linear combination of significant CIMs. That is, relevant CIMs, as identified through feature selection above, were summed proportionally with the target's coordinates in face space and, then, added to an average face obtained from all remaining non-target faces and playing the role of a face prior. However, in contrast to previous work (Nestor, Plaut, & Behrmann, 2016), the current procedure capitalized on face space invariance for reconstruction purposes. Specifically, a happy version of face space was aligned to its neutral counterpart via Procrustes transformation using all but one face; then, the left-out face from the happy version of face space was projected into the neutral space using the parameters of the Procrustes mapping function found above. The resulting coordinates in neutral face space were next used to reconstruct the appearance of the target face, with a neutral expression, from neutral CIMs. Conversely, a happy version of the target face relied upon aligning neutral face space to its happy counterpart.

Thus, reconstruction relied here on the presence of a robust visual-based structure shared by different face space estimates across expressions. More clearly, in the absence of a common space topography the target face would be projected to a non-informative location of face space and its subsequent reconstruction would resemble the target stimulus no better than expected at chance level (per the evaluation described below).

The procedure above, including face space estimation, was conducted for separate time windows of the ERP trace as well as for temporally cumulative data.

Evaluation of reconstruction results

Image reconstructions were compared to their target stimuli via two different methods. First, image-based accuracy was estimated as the percentage of instances for which a reconstructed image in CIEL*a*b* was more similar to its target, by a pixel-wise L2 metric, than to any other stimulus with the same expression. Average reconstruction accuracy was then compared against permutation-based chance estimates by shuffling reconstruction labels and by recomputing average accuracy across reconstructions each time (for a total of 10^3 permutations). This procedure was applied to all types of reconstruction (e.g., both window-based and temporally cumulative) separately for neutral and happy faces.

Second, a single set of reconstructions, based on temporally cumulative group-based data, was subjected to experimental evaluation in a separate behavioral test. To this end, 14 new participants (six males and eight females, age range: 20-28) were requested to match image reconstructions to their targets in a two-alternative forced choice (2AFC) task. Specifically, each of 108 unfamiliar face reconstructions, including both expressions, was presented in the company of two stimuli, one of which was the actual target and the other was a foil (another face image). Thus, on each trial, a display was shown containing a reconstructed image, at the top, and two stimuli side by side, at the bottom, all of which had the same expression and the same size (as specified in Experimental Procedures). Each display was presented until participants made a response to decide which stimulus was more similar to the top image by pressing a designated left/right key. For each participant, any reconstructed image was presented 4 times in the company of different foils; thus, across participants, all 53 possible foils for a given reconstruction were exhausted. Stimulus order was pseudorandomized so that different reconstructed images appeared on consecutive trials while target stimuli appeared equally often on the left/right side. Each experimental session was completed over the course of 30 minutes.

Experimental-based estimates of reconstruction accuracy results were measured as the proportion of correct matches across participants and tested for significance tested against chance (50%) using a one-sample two-tailed t-test. Last, experimental and homologous image-based estimates of reconstruction accuracy were compared to each other via Pearson correlation across images, separately for neutral and happy faces.

Results

Univariate ERP results

Three ERP components relevant for face processing, P1, N170 and N250, were each identified across occipitotemporal electrodes (Fig 1) and examined by a two-way repeated measures analysis of variance (54 unfamiliar face identities x 2 hemispheres). The analysis of the P1 component [a- c] (see letter-indexed rows of Table 1 for details of corresponding analyses) and N170[d-f] analyses found no significant effects. Last, N250 analyses[g-i] found a main effect of identity ($F(53,636)=3.69$, $p=0.001$, $\eta_p^2= 0.235$) and a marginally significant interaction ($F(53,636)=1.32$, $p=0.07$, $\eta_p^2= 0.099$) – both hemispheres[j,k] showed a significant effect of identity ($F(53,636)=2.67$, $p=0.009$, $\eta_p^2= 0.182$; $F(53,636)=3.214$, $p=0.004$, $\eta_p^2= 0.21$ for the left and right hemispheres, respectively), but the effect was larger on the right side.

Insert Figure 1 about here

Pattern classification of facial identity

A total of 108 unfamiliar male faces (54 individuals x 2 emotional expressions) were classified based on ERP traces associated with their viewing. Specifically, spatiotemporal ERP patterns across bilateral occipitotemporal electrodes were averaged across participants and, then, evaluated for their ability to support facial identity discrimination. To assess the time course of individual face processing, classification was conducted across consecutive 10 ms temporal windows both within and across expression by training and testing the classifier on faces with the same or different expression.

This analysis found significant levels of discrimination across extensive intervals[1] (permutation test; q<0.01). Specifically, across-expression classification evinced above-chance accuracy from 152 ms after stimulus presentation until the end of the epoch, with two peaks at 170 ms and 295 ms (see Fig 2a for the group-based results). Within-expression classification yielded a similar time course but consistently higher levels of accuracy and an earlier onset, at 140 ms, in agreement with the reliance on additional, lower-level image cues for this type of discrimination (In addition, an earlier interval of significance was found for across-expression classification between 0-5 ms; however, given its very early occurrence, its reduced amplitude and the absence of its replication by within-expression classification we treat this data point as a false positive). Of note, for both types of classification we found that discrimination accuracy was maximized in the vicinity of the N170 component as identified by univariate analyses of ERP data (see Fig 1).

Insert Figure 2 about here

Following the approach described above group-based analyses were complemented next by single-participant analyses[m]. These analyses confirmed the feasibility of facial identity discrimination from the data of single participants (see Fig. 2b for the results of a representative participant). However, discrimination levels were lower than in the group-based analyses, likely due to the lower signal-to-noise ratio (SNR) of single-participant ERPs and its impact on classification (Grootswagers et al., 2017). Further, multiple intervals of discrimination emerged, as opposed to a single, uninterrupted one.

Next, pattern classification was applied to temporally cumulative data by concatenating data points from all time bins between 50 ms – 650 ms after stimulus onset[n-q]. The aim of this analysis was to maximize discrimination performance by concomitantly exploiting relevant information from all potentially relevant time points. Specifically, while the ~61-fold increase in pattern dimensionality (i.e., 12 electrodes x 307 time bins) would, by itself, reduce the effectiveness of classification, we considered the possibility that any ensuing classification decrement may be offset by the use of complementary sources of information from different time points (Blankertz et al., 2011).

Consistent with the hypothesis above, we found that this analysis yielded robust levels of discrimination for group-based data (Fig 3a): 64% and 71% for across and within-expression discrimination, respectively (permutation test; p=0.001 for both types of discrimination and both expressions). Of note, these results outperform peak performance obtained with window-based analyses in the proximity of the N170 component. Further, single-participant estimates of discrimination with temporally cumulative data were also computed[r-u] and, then, averaged across participants (Fig 3b). Again, performance was better than chance for both within-expression discrimination (two-tailed t-test across participants against 50% chance-level discrimination; $t(12)$= 9.89 and 7.27, Cohen's d=2.1 and 2.86 for neutral and happy faces, respectively; p's=0.001) and for across-expression discrimination ($t(12)$=6.84 and 7; d=2.02 and 1.97 for neutral and happy faces, respectively; p's<0.001). Further, a two-way repeated measures analysis of variance[v-x] (2 discrimination types x 2 expressions) revealed a main effect of discrimination types ($F(1,12) = 50.05$, $p<0.001$, $\eta_p^2 = 0.81$), with higher accuracy for within than across-expression discrimination, but no effect of expression and no interaction.

Insert Figure 3 about here

To examine possible effects of face database the analysis above was repeated while restricting pattern classification either to pairs of faces from the same database or from different databases. A two-way repeated measures analysis of variance[y-ab] (2 discrimination types: within/across expression x 2 pair types: within/across database) was carried out to this end – classification estimates were collapsed across neutral and happy faces given the absence of any expression effects above. This analysis revealed a main effect of discrimination types ($F(1,12) = 45.92$, $p<0.001$, $\eta_p^2 = 0.79$), with higher accuracy for within than across-expression discrimination, as well as a main effect of pair types ($F(1,12) = 38.73$, $p<0.001$, $\eta_p^2 = 0.76$), with higher accuracy for within than across-database classification. Critically though, all classification estimates were significantly above chance (two-tailed t-tests against 50% chance-level discrimination)[ac-af]; mean accuracy=56.3%, $t(12)=8.36$,$p<0.001$, *Cohen's d*=2.32 for within-expression, within-database discrimination; mean accuracy=58.7%, $t(12)=8.38$, $p<0.001$, *Cohen's d*=2.32 for within-expression, across-database discrimination; mean accuracy=53.9%, $t(12)=7.64$,$p<0.001$, *Cohen's d*=2.12 for across-expression, within-database discrimination**;** mean accuracy=55.9%, $t(12)=6.88$,$p<0.001$, *Cohen's d*=1.91 for across-expression, across-database discrimination).

Last, for completeness, classification analyses including all face pairs within and across databases were repeated with all 64 electrodes, instead of the subset of 12 occipitotemporal electrodes noted above.

However, no consistent boost in discrimination was found for any analysis in this case. Hence, for simplicity, the remainder of our analyses and results, as reported below, are based on occipitotemporal electrodes only.

Neural-based discrimination and behavioral performance

In order to assess and to quantify the relationship between behavioral and neural-based face similarity group-level estimates of EEG-based discrimination were related to behavioral estimates of pairwise face similarity. Specifically, across-expression discrimination accuracy was compared with its behavioral counterpart, across identity pairs, through Pearson correlation[ag]. This comparison revealed, first, a significant correlation for discrimination estimates from the temporally cumulative analysis (r = 0.245; p< 0.001). Then, behavioral estimates were correlated with their neural counterpart separately for each 10 ms temporal window (Fig 4)[ah]. This analysis found multiple intervals of significant correlation, between 137 -197 ms, 236 -335 ms and 340-385 ms, with peaks at 157 ms and 266 ms (q<0.01).

Insert Figure 4 about here

Next, we sought to assess whether the different levels of discrimination achieved with different participants were related to their overall face processing skills. To this end, individual estimates of discrimination from the temporally cumulative analysis, averaged across identity pairs, were correlated with CFMT[ai] scores of the same individuals. No significant correlations were found with these scores or with any other behavioral measures considered such as average familiarity ratings with famous faces or VVIQ-2[aj] scores.

Neural-based face space and expression-invariance

Face space estimates were derived through the application of MDS to within-expression face classification of temporally cumulative data. Specifically, MDS was applied to pairwise face discrimination values derived through pattern analysis of group-based data, separately for neutral and happy faces. Then, the resulting spaces were reduced to the first 20 dimensions and aligned with each other via Procrustes transformation.

An examination of the resulting spaces for neutral and happy faces, following their alignment, seemed consistent with the presence of a common topography across expressions (Fig 5a). To assess their fit more rigorously a comparison with permutation-based alignment estimates (Fig 5b)[ak] was computed next. This comparison indicated that the fit between the two spaces was considerably better than chance (p<0.001). Beyond its theoretical implications, this finding is relevant here in that it may allow exploiting the structure of visual information invariant across expression for reconstruction purposes, as detailed next.

Insert Figure 5 about here

Reconstruction results

Visual features were derived from the structure of face space, dimension by dimension, through a procedure akin to reverse correlation/image classification (Sekuler, Gaspar, Gold, & Bennett, 2000; Gosselin & Schyns, 2003; Martin-Malivel, Mangini, Fagot, & Biederman, 2006). Such features, or classification images (CIMs), were then assessed through a permutation test for the presence of significant information pixel by pixel separately for each CIEL*a*b* color channel.

An examination of CIMs containing significant[al-aw] information revealed global contrast patterns across multiple color channels – examples of such features derived from group-based temporally cumulative data are shown in Fig 6. The emergence of these features confirmed that neural-based face space is, at least partly, organized by visual information (as opposed, for instance, to higher-level semantic information). More relevantly here, it points to the potential value of CIMs as reconstruction features.

Insert Figure 6 about here

Accordingly, significant CIMs were linearly combined to deliver an approximation of face appearance broadly following an image reconstruction approach recently used with fMRI data (Nestor et al., 2016). Specifically, image reconstruction was separately applied to neutral and happy expressions. As noted above, the common face space topography for the two expressions could, in theory, allow using the relative position of a given identity in one space to deliver a reconstruction of the same facial identity with the opposite expression.

Consistent with the hypothesis above, this procedure, as performed for separate time windows with group-based data, found evidence for multiple intervals capable of supporting above-chance reconstruction (image-based permutation test; $q<0.05$) – reconstruction accuracy was assessed via a pixelwise image-matching test across reconstructed images and stimulus images. The earliest interval had an onset at 160 ms and 170 ms for neutral[ax] and happy [ay] faces, respectively, while accuracy peaked at 187 ms and 180 ms for the two expressions. Examples of reconstructions, converted from CIEL*a*b* back to RGB, are shown in Fig 7a for two different time points while average reconstruction accuracies are plotted in Fig 7b. Further, a representative example of image reconstruction for a single face, interval by interval, is shown in Movie 1 along with the temporal profile of its reconstruction accuracy.

Insert Figure 7 about here

The application of the same procedure to temporally cumulative data led to more robust results: 69.46% image-based accuracy for neutral faces[az] and 63.91% for happy faces[ba] (permutation test; p's=0.001). Fig 8a shows examples of reconstructions and Fig 8b displays average accuracy estimates and their permutation-based assessment. Experimental-based estimates of accuracy, obtained with a new group of participants, led to more modest levels of accuracy (Fig 8c); however, both neutral and happy face reconstructions were still accurate above chance (two-tailed t-test across participants against 50% chance-level discrimination: $t(13)=6.70$, 4.38; *Cohen's d*=1.86, 1.22, for neutral[bb] and happy[bc], respectively; p's<0.001 for all), with no difference between neutral and happy faces. Further, reconstruction accuracies as estimated by the two tests, image-based and experimental-based, were compared with each other across facial identities and were found to significantly correlate with each other (r=0.43 and 0.42; p's=0.001 and 0.002 for neutral[bd] and happy[be] faces, respectively), thus, mutually reinforcing their validity.

Insert Figure 8 about here

Next, across-expression classification estimates obtained with temporally cumulative data were compared with their corresponding reconstruction accuracies averaged across expressions. Specifically, Pearson correlation across facial identities found a positive relationship between across-expression discrimination and image-based accuracy[bf] (r=0.87, p<0.001). Thus, the more discriminable a facial identity is, the more accurately it can be reconstructed.

Last, reconstruction was performed for single-participant data and evaluated with the aid of the image-based test. Accuracy levels were still above chance (two-tailed t-test across participants against 50% chance-level discrimination; mean accuracy = 53.1%, $t(12)$=2.52, p=0.027, *Cohen's d*=0.73 and mean accuracy = 53.3%, $t(12)$=2.24, p=0.045, *Cohen's d*=0.65 for neutral[bg] and happy[bh] face reconstructions, respectively) while Pearson correlations between classification accuracy and reconstruction accuracy across participants were also found significant (r=0.83 and r=0.84, for neutral[bi] and happy[bj] faces, respectively; p's<0.001). Thus, participants who provided data supporting higher levels of face classification also provided more accurate reconstruction results.

Discussion

The current work investigates the neural basis of individual face processing and its temporal dynamics through the application of pattern analysis and image reconstruction to EEG data. This investigation yields several notable outcomes as follows.

First, we find that EEG data support facial identity discrimination. By and large, this finding confirms the possibility of EEG-based pattern classification of facial identity across changes in expression from EEG data (Nemrodov et al., 2016). Discrimination peaks were identified in the proximity of the N170 and N250 ERP components, consistent with univariate analyses pointing to the relevance of the former (Itier & Taylor, 2002; Heisz, Watter, & Shedden, 2006; Jacques, D'Arripe, & Rossion, 2007; Caharel et al., 2009) and the latter (Schweinberger et al., 2002; Huddy et al., 2003; Tanaka et al., 2006) for face processing. The onset of discrimination, around 150 ms, was intermediary to early estimates in the vicinity of P1 (Nemrodov et al., 2016) and later estimates around 200 ms as reported with MEG (Vida et al., 2017). One possibility is that early estimates, along with higher levels of discrimination, can be triggered by the use of low-level image properties (Cauchoix et al., 2014; Ghuman et al., 2014). In line with this consideration, we found that within versus across-expression discrimination produced earlier and consistently higher levels of discrimination accuracy. Importantly though, across-expression classification, which aims to minimize reliance upon low-level cues, exhibited robust levels of discrimination across an extensive interval (i.e., from ~150 ms onwards) while its time course was also mirrored by that of neural-behavioral correlations in the context of pairwise face similarity.

Second, temporally cumulative analyses targeted identity discrimination across a broad interval between 50-650 ms after stimulus onset. Despite the increase in dimensionality for the classification patterns, these data supported even more robust levels of accuracy for both within and across-expression discrimination, consistent with the presence of relevant information at multiple time points. Moreover, the superior levels of discrimination obtained with temporally cumulative data, as opposed to 10 ms windows, agrees with the presence of distinct sources of information at different time points. That is, we relate the boost in classification accuracy with the ability to exploit complementary information about facial identity at different times. Interestingly, this conclusion echoes that based on the lack of temporal generalization found with cross-temporal object decoding of MEG data (Carlson et al., 2013; Isik et al., 2014) – specifically, the lack of classification success with using training and testing data from distinct time intervals has been taken as evidence for the presence of different types of information over time (Grootswagers, Wardle, & Carlson, 2016). Further, the boost in classification noted above is important for practical purposes: it suggests that investigations that place less emphasis on clarifying the time course of discrimination can be better served by exploiting patterns across larger temporal intervals. Accordingly, our subsequent investigations into face space structure and image reconstruction were carried out with both window-based and cumulative data.

Third, a neural-based estimate of face space was constructed from EEG data and its organization was explained by the presence of visual information captured by CIMs. This result is significant in that it confirms that pattern discrimination relies, at least partly, on relevant visual information (e.g., as opposed to higher-level semantic cues). More importantly, we note that neural-based face space has been examined in the context of fMRI (Loffler et al., 2005; Rotshtein et al., 2005; Gao & Wilson, 2013) and monkey neurophysiology (Leopold, Bondar, & Giese, 2006; Freiwald, Tsao, & Livingstone, 2009). Yet, many of its properties, as related to facial identity representation, remain to be clarified. For instance, its invariant structure across different types of image transformation remains to be assessed and quantified. Behavioral research suggests that face space topography is largely invariant across viewpoint and lighting (Blank & Yovel, 2011). Here, we reach a similar conclusion regarding the expression invariance of neural-based face space as derived from EEG data.

Fourth, image reconstruction was carried out with the aid of CIM features derived directly from the structure of EEG data (i.e., as opposed to predefined visual features selected due to their general biological plausibility). This endeavor builds upon pattern classification while, critically, it validates its results by showcasing its reliance on relevant visual information encoded in the EEG signal. We found that multiple temporal intervals supported better-than-chance reconstruction for both neutral and happy faces with a peak in the proximity of the N170 component. Also, reconstruction accuracy was further boosted by considering temporally cumulative information, as used for pattern classification. More importantly, these results are notable in that, unlike previous work with fMRI-based facial image reconstruction (Cowen, Chun, & Kuhl, 2014; Nestor, Plaut, & Behrmann, 2016), they exploit invariant face space information for reconstruction purposes. Thus, arguably the current findings speak to the visual nature of facial identity representations rather than just to lower-level pictorial aspects of face perception.

Further, the current work provides proof of principle for EEG-based image reconstruction. Importantly, not only does this demonstrate the applicability of image reconstruction to neuroimaging modalities other than fMRI but, critically, it shows that EEG-based reconstruction can compete in terms of overall accuracy with its fMRI counterpart (Nestor, Plaut, & Behrmann, 2016).

Thus, here we build upon previous EEG investigations of face processing and on pattern analyses of neuroimaging data to address several theoretical and methodological issues. In particular, the current work capitalizes on previous attempts at clarifying the temporal profile of individual face processing via linear classification of spatiotemporal EEG patterns across facial expression (Nemrodov et al., 2016). In agreement with this previous work we find that individual faces can be discriminated from their corresponding EEG patterns, that their time course exhibits an extended interval of significant discrimination and that multiple discrimination peaks occur, including an early one in the vicinity of the N170 component. Unlike this previous work though, which only relied on a restricted set of eight male and female faces, we find that such discrimination can be performed even with a large, homogenous set of face images controlled for low and high-level face properties (e.g., through geometrical alignment and intensity normalization of 108 Caucasian male face images). Hence, differences in discrimination onset across studies (i.e., 70 ms in this previous work versus 152 ms here) are likely related to reliance on idiosyncratic image differences within a small stimulus set in this previous investigation. More importantly though, not only does the current work examine the time course of individual face classification in a more reliable and thorough manner but, critically, it utilizes its outcomes for the purpose of facial feature derivation and image reconstruction.

Naturally, boosting even further classification and reconstruction accuracy is an important future endeavor. Regarding classification, this could be achieved, for instance, through efficient techniques for feature selection, such as recursive feature elimination (Hanson & Halchenko, 2008; Nestor, Plaut, & Behrmann, 2011), aimed at reducing pattern dimensionality and optimizing discrimination performance. Since electrodes are likely to carry irrelevant or redundant information at multiple time points, eliminating this information from higher-dimensional spatiotemporal patterns (e.g., across all electrodes and time points) could benefit classification. Regarding reconstruction, more complex, biologically-plausible approaches can be developed, for instance, by considering shape and surface information separately within the reconstruction process. Since shape and surface provide complementary cues to face processing (Jiang, Blanz, & O'Toole, 2006; Andrews et al., 2016), it would be informative to derive separate types of CIMs corresponding to this distinction and to consider their separate contribution to facial image reconstruction.

Notably though, beyond overall performance, EEG-based reconstruction stands out by its ability to clarify the dynamics of visual representations as they develop in response to a given stimulus. For instance, it can speak to how a percept evolves over time in response to a static stimulus, as attempted here, by inspecting image reconstruction across consecutive time windows. Alternatively, this method could be extended to recover fine-grained dynamic information as present in moving stimuli. While reconstruction of natural movies has been previously carried out with fMRI (Nishimoto et al., 2011), the superior temporal resolution of EEG could make this modality a more efficient choice for the recovery of dynamic visual information. Further, we note that the comparatively wide availability of EEG systems could also render this modality the preferred choice for the development of new types of image-reconstruction brain-computer interfaces (BCI).

Last, while the current investigation focuses on faces as a visual category of interest, we argue that the present methodological approach can inform individual object recognition more generally. This is theoretically suggested by the presence of common neurocomputational principles underlying face and object identification (Cowell & Cottrell, 2013; Wang, Gauthier, & Cottrell, 2016) as well as, methodologically, by the ability to evaluate the dynamics of invariant object recognition (Isik, et al., 2014). Particularly encouraging in this sense is the success of efforts to construct and characterize object similarity spaces from MEG (Carlson et al., 2013) and EEG data (Kaneshiro et al., 2015).

To conclude, our investigation targets the neural dynamics of face processing as reflected by EEG patterns. Our findings shed new light on the time course of facial identity processing while providing a way to extract and to assess the underlying visual information. Last, from a methodological standpoint, our results establish the feasibility of EEG-based image reconstruction and, more generally, they confirm the rich informational content of spatiotemporal EEG patterns.

References

Andrews, T. J., Baseler, H., Jenkins, R., Burton, A. M., & Young, A. W. (2016) Contributions of feature shapes and surface cues to the recognition and neural representation of facial identity. Cortex: 83, 280–291.

Anzellotti, S., Fairhall, S. L., & Caramazza, A. (2014) Decoding representations of face identity that are tolerant to rotation. Cerebral Cortex, 24: 1988–1995.

Bentin, S., & Deouell, L. Y. (2000) Structural encoding and identification in face processing: ERP evidence for separate mechanisms. Cognitive Neuropsychology, 17: 35–54.

Blank, I., & Yovel, G. (2011) The structure of face--space is tolerant to lighting and viewpoint transformations. Journal of Vision, 11: 1–13.

Blankertz, B., Lemm, S., Treder, M., Haufe, S., & Muller, K. R. (2011) Single-trial analysis and classification of ERP components - A tutorial. NeuroImage, 56: 814–825.

Bowles, D. C., McKone, E., Dawel, A., Duchaine, B., Palermo, R., Schmalzl, L., Rivolta D., Wilson, C.E., & Yovel, G. (2009) Diagnosing prosopagnosia: effects of ageing, sex, and participant-stimulus ethnic match on the Cambridge Face Memory Test and Cambridge Face Perception Test. Cognitive neuropsychology 26: 423-455.

Brainard, D. H. (1997) The Psychophysics Toolbox. Spatial Vision, 10: 433–436.

Caharel, S., D'Arripe, O., Ramon, M., Jacques, C., & Rossion, B. (2009) Early adaptation to repeated unfamiliar faces across viewpoint changes in the right hemisphere: evidence from the N170 ERP component. Neuropsychologia, 47: 639–43.

Carlson, T. A., Tovar, D., Alink, A., & Kriegeskorte, N. (2013) Representational dynamics of object vision: The first 1000 ms. Journal of Vision, 13: 1–19.

Cauchoix, M., Barragan-Jason, G., Serre, T., & Barbeau, E. J. (2014) The neural dynamics of face detection in the wild revealed by MVPA. The Journal of Neuroscience, 34: 846–54.

Chang, C.-C., & Lin, C.-J. (2011) LIBSVM. *ACM Transactions on Intelligent Systems and Technology*, *2*: 1–27.

Chang, L & Tsao, D. Y. (2017) The Code for Facial Identity in the Primate Brain. Cell, 169: 1013:1028.

Cichy, R. M., Pantazis, D., & Oliva, A. (2014) Resolving human object recognition in space and time. Nature Neuroscience, 17: 455–62.

Cowell, R. A., & Cottrell, G. W. (2013) What Evidence Supports Special Processing for Faces? A Cautionary Tale for fMRI Interpretation. Journal of Cognitive Neuroscience, 25: 1777–1793.

Cowen, A. S., Chun, M. M., & Kuhl, B. A. (2014) Neural portraits of perception: reconstructing face images from evoked brain activity. NeuroImage: 94, 12–22.

Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010) Fast saccades toward faces: face detection in just 100 ms. Journal of Vision, 10: 1–17.

Davidesco, I., Zion-Golumbic, E., Bickel, S., Harel, M., Groppe, D. M., Keller, C. J., Schevon, C. A., McKhann, G. M., Goodman R. R., Goelman G., Schroeder, C.E., Mehta, A. D., & Malach, R. (2014) Exemplar selectivity reflects perceptual similarities in the human fusiform cortex. Cerebral Cortex, 24: 1879–1893.

Delorme, A., Sejnowski, T., & Makeig, S. (2007) Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. NeuroImage, 34: 1443–9.

Duchaine, B., & Nakayama, K. (2006) The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. Neuropsychologia, 44: 576–585.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007) G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39, 175-191.

Freiwald, W. A., Tsao, D. Y., & Livingstone, M. S. (2009) A face feature space in the macaque temporal lobe. Nature Neuroscience, 12: 1187–96.

Gao, X & Wilson H. R. (2013) The neural representation of face space dimensions. Neuropsychologia, 51: 1787-1793.

Ghuman, A. S., Brunet, N. M., Li, Y., Konecky, R. O., Pyles, J. A, Walls, S. A., Destefinom V, Wang, W. & Richardson, R. M. (2014) Dynamic encoding of face information in the human fusiform gyrus. Nature Communications, 5: 5672.

Goesaert, E., & Op de Beeck, H. P. (2013) Representations of facial identity information in the ventral visual stream investigated with multivoxel pattern analyses. The Journal of Neuroscience, 33: 8549–58.

Gosselin, F., & Schyns, P. G. (2003) Superstitious perceptions reveal properties of internal representations. Psychological Science, 14: 505–509.

Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2017) Decoding dynamic brain patterns from evoked responses: A tutorial on multivariate pattern analysis applied to time-series neuroimaging data. Journal of Cognitive Neuroscience, 29: 677-697.

Hanson, S. J., & Halchenko, Y. O. (2008) Brain reading using full brain support vector machines for object recognition: there is no “face” identification area. Neural Computation, 20: 486–503.

Heisz, J. J., Watter, S., & Shedden, J. M. (2006) Progressive N170 habituation to unattended repeated faces. Vision Research, 46: 47–56.

Huddy, V., Schweinberger, S. R., Jentzsch, I., & Burton, A. M. (2003) Matching faces for semantic information and names: an event-related brain potentials study. Cognitive Brain Research, 17: 314–326.

Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005) Fast readout of object identity from macaque inferior temporal cortex. Science, 310: 863–6.

Isik, L., Meyers, E. M., Leibo, J. Z., Poggio, T. A., Baldauf, D., & Desimone, R. (2014) The dynamics of invariant object recognition in the human visual system. Journal of Neurophysiology: 111, 91–102.

Itier, R. J., & Taylor, M. J. (2002) Inversion and Contrast Polarity Reversal Affect both Encoding and Recognition Processes of Unfamiliar Faces: A Repetition Study Using ERPs. NeuroImage, 15: 353–372.

Jackson, D. A. (1995) Protest - A Procrustean Randomization Test of Community Environment Concordance. Ecoscience.

Jacques, C., D'Arripe, O., & Rossion, B. (2007) The time course of the inversion effect during individual face discrimination. Journal of Vision, 7: 1–9.

Jiang, F., Blanz, V., & O'Toole, A. J. (2006) Probing the visual representation of faces with adaptation: A view from the other side of the mean. Psychological Science, 17: 493–500.

Kaneshiro, B., Perreau Guimaraes, M., Kim, H., Norcia, A. M., Suppes, P., Guimaraes, M. P., & Anthony, M. (2015) A Representational Similarity Analysis of the Dynamics of Object Processing Using Single-Trial EEG Classification. PloS One, 10: e0135697.

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

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010) Presentation and validation of the Radboud Faces Database. Cognition & Emotion, 24: 1377–1388.

Lehky, S. R. (2000) Fine discrimination of faces can be performed rapidly. Journal of Cognitive Neuroscience, 12: 848–55.

Leopold, D. A., Bondar, I. V, & Giese, M. A. (2006) Norm-based face encoding by single neurons in the monkey inferotemporal cortex. Nature, 442: 572–575.

Liu, H., Agam, Y., Madsen, J. R., & Kreiman, G. (2009) Timing, Timing, Timing: Fast Decoding of Object Information from Intracranial Field Potentials in Human Visual Cortex. Neuron, 62: 281–290.

Loffler, G., Yourganov, G., Wilkinson, F., & Wilson, H. R. (2005) fMRI evidence for the neural representation of faces. Nature Neuroscience, 8: 1386–1390.

Marks, D. F. (1995) New directions for mental imagery research. Journal of Mental Imagery, 19: 153–167.

Martin-Malivel, J., Mangini, M. C., Fagot, J., & Biederman, I. (2006) Do Humans and Baboons Use the Same Information When Categorizing Human and Baboon Faces? *Psychological Science*, *17*: 599-607

Martinez, A., & Benavente, R. (1998) The AR face database. No 24.

Miyawaki, Y., Uchida, H., Yamashita, O., Sato, M. A, Morito, Y., Tanabe, H. C., Sadato, N., & Kamitani, Y. (2008) Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. Neuron, 60: 915–929.

Mouraux, A., & Iannetti, G. D. (2008) Across-trial averaging of event-related EEG responses and beyond. Magnetic Resonance Imaging, 26: 1041–1054.

Murray, R. F. (2011) Classification images: A review. Journal of Vision, 11, 1–25.

Naselaris, T., Prenger, R.J., Kay, K.N., Oliver, M.D., Gallant, J.L. (2009). Bayesian reconstruction of natural images from human brain activity. Neuron, 63, 902-915.

Nemrodov, D., Niemeier, M., Mok, J. N. Y., & Nestor, A. (2016) The time course of individual face recognition: a pattern analysis of ERP signals. NeuroImage, 132: 469–476.

Nestor, A., Plaut, D. C., & Behrmann, M. (2011) Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. Proceedings of the National Academy of Sciences of the United States of America, 108: 9998–10003.

Nestor, A., Plaut, D. C., & Behrmann, M. (2013) Face-space architectures: evidence for the use of independent color-based features. Psychological Science, 24: 1294–300.

Nestor, A., Plaut, D. C., & Behrmann, M. (2016) Feature-based face representations and image reconstruction from behavioral and neural data. Proceedings of the National Academy of Sciences, 113: 416–421.

Newell, F. N., Chiroro, P., & Valentine, T. (1999) Recognizing unfamiliar faces: the effects of distinctiveness and view. The Quarterly Journal of Experimental Psychology, 52: 509–34.

Nishimoto, S., Vu, A. T., Naselaris, T., Benjamini, Y., Yu, B., & Gallant, J. L. (2011) Reconstructing visual experiences from brain activity evoked by natural movies. Current Biology, 21: 1641–6.

Pelli, D. (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spatial Vision, 10: 437–42.

Rossion, B., & Caharel, S. (2011) ERP evidence for the speed of face categorization in the human brain: Disentangling the contribution of low-level visual cues from face perception. Vision Research, 51: 1297–311.

Rotshtein, P., Henson, R. N., Treves, A., Driver, J., & Dolan, R. J. (2005) Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. Nat Neurosci, 8: 107–113.

Schweinberger, S. R., Pickering, E. C., Burton, A. M., & Kaufmann, J. M. (2002) Human brain potential correlates of repetition priming in face and name recognition. Neuropsychologia, 40: 2057–73.

Sekuler, A. B., Gaspar, C. M., Gold, J. M., & Bennett, P. J. (2004) Inversion leads to quantitative, not qualitative, changes in face processing. *Current Biology*, *14*: 391–396.

Tanaka, J., & Curran, T. (2001) A neural basis for expert object recognition. Psychological Science, 12: 43–47.

Tanaka, J., Curran, T., Porterfield, A. L., & Collins, D. (2006) Activation of preexisting and acquired face representations: The N250 event-related potential as an index of face familiarity. Journal of Cognitive Neuroscience, 18: 1488–1497.

Thomaz, C. E., & Giraldi, G. A. (2010) A new ranking method for principal components analysis and its application to face image analysis. Image and Vision Computing, 28: 902–913.

Valentine, T. (1991) A unified account of the effects of distinctiveness, inversion, and race in face recognition. The Quarterly Journal of Experimental Psychology Section A, 43: 161–204.

Van de Nieuwenhuijzen, M. E., Backus, A. R., Bahramisharif, A., Doeller, C. F., Jensen, O., & van Gerven, M. A. J. (2013) MEG-based decoding of the spatiotemporal dynamics of visual category perception. NeuroImage, 83: 1063–1073.

Vida, M. D., Nestor, A., Plaut, D. C., & Behrmann, M. (2017) Spatiotemporal dynamics of similarity-based neural representations of facial identity. Proceedings of the National Academy of Sciences of the United States of America, 114: 388–393.

Wang, P., Gauthier, I., & Cottrell, G. W. (2016) Are Face and Object Recognition Independent? A Neurocomputational Modeling Exploration. Journal of Cognitive Neuroscience, 28: 558–574.

Zheng, X., Mondloch, C. J., & Segalowitz, S. J. (2012) The timing of individual face recognition in the brain. Neuropsychologia, 50: 1451–1461.

Legends

Figure 1.

Grand-averaged ERPs across: (a) left hemisphere electrodes (P5, P7, P9, PO3, PO7, O1) and (b) right hemisphere electrodes (P6, P8, P10, PO4, PO8, O2) for 54 facial identities (averaged across expressions). Head maps show voltage distributions at (a) N170 (b) P1, N250.

Figure 2

The time course of EEG-based classification accuracy for across and within-expression discrimination of facial identity. (a) Classification was conducted across consecutive 10ms window patterns over 12 occipitotemporal electrodes for group-based ERP data. Both types of analysis exhibited above-chance discrimination across extensive temporal intervals (permutation test; FDR-correction across time, $q<0.01$); shaded areas mark intervals of better-than-chance discrimination for across-expression classification. (b) The time course of EEG-based classification accuracy for across and within-expression discrimination of facial identity for a single representative participant. Classification was conducted across consecutive 10 ms window patterns over 12 occipitotemporal electrodes. Both types of analysis exhibited above-chance discrimination across extensive temporal intervals (permutation test; FDR-correction across time, $q<0.01$); shaded areas mark intervals of better-than-chance discrimination for across-expression classification.

Figure 3

EEG-based classification accuracy for across and within-expression discrimination of facial identity with temporally cumulative data (50-650 ms after stimulus onset). Accuracy corresponding to neutral and happy faces are separately shown for: (a) group-based ERP data and (b) single-participant data (i.e., pattern classification was conducted individually for each participant and, then, its results averaged across participants). The plots display: (a) the results of permutation tests (red solid and dash lines indicate average accuracy and 99% confidence intervals estimated with 10^3 permutations) and (b) the distribution of single-participant data (green and purple solid lines indicate medians, boxes represent 1st and 3rd quartiles, whiskers represent minimum and maximum accuracy values, points represent individual participants' values and red solid lines indicate chance-level discrimination).

Figure 4

Correlation of EEG and behavioral-based estimates of pairwise face similarity. EEG-based estimates are derived from across-expression discrimination of facial identity for consecutive 10ms windows of group-based data. Multiple intervals, marked by shaded areas, exhibit significant levels of correlation (permutation test; FDR correction across time, $q<0.01$).

Figure 5

Neutral and happy face space estimates along with their fit (after Procrustes alignment). Estimates were derived through MDS analysis of similarity matrices based on within-expression face discrimination of group-based temporally cumulative data. The two face space estimates exhibit a similar topography as found with: (a) their visualization across multiple dimensions (red and green circles indicate neutral and happy faces, respectively; solid lines connect face images with the same identity with the thickness of the line proportionally reflecting shorter distances; the first four dimensions shown here account for 40% and 41% variance for neutral and happy face space); (b) badness of fit (sum of squared errors) for the two spaces compared to their permutation-based counterpart (average fits and 95% confidence intervals estimated with 10^3 permutations).

Figure 6

Examples of classification images (CIMs) extracted from EEG-based face space constructs for (a) neutral and (b) happy faces. Pairs of images show raw CIMs (odd columns) and their analysis (even columns) with a pixelwise permutation-based test (FDR-corrected across pixels; $q<0.05$). Bright/dark, red/green, and yellow/blue regions in analyzed CIs mark areas of the face brighter (L*), redder (a*), or more yellow (b*) than chance in CIEL*a*b*. Results are shown separately for the first and fourth dimensions of face spaces derived from group-based temporally cumulative data.

Figure 7

Reconstruction results for neutral and happy face images across consecutive 10 ms windows of group-based data. (a) Examples of face stimuli along with their corresponding reconstructions at two different times (numbers in the upper left corner indicate image-based estimates of reconstruction accuracy). (b) The time course of reconstruction accuracy: both neutral and happy face images exhibit above-chance discrimination across multiple temporal intervals (permutation test; FDR correction across time, $q<0.05$; shaded areas mark intervals of better-than-chance discrimination for neutral faces). Reconstruction accuracy is maximized in the vicinity of the N170 component (Fig 1) and of the discrimination peak found with pattern classification (Fig 2).

Figure 8

Reconstruction results for neutral and happy faces relying on temporally cumulative group-based data: (a) examples of face stimuli along with their corresponding reconstructions (numbers in the upper left corner indicate image-based estimates of reconstruction accuracy; numbers in the upper right indicate experimental-based accuracy); (b) average image-based reconstruction accuracy (red solid and dash lines indicate average accuracy and 95% confidence intervals estimated with 10^3 permutations) and (c) average experimental-based reconstruction accuracy (green and purple solid lines indicate medians, boxes represent 1st and 3rd quartiles, whiskers represent minimum and maximum accuracy values, points represent individual participants' values and red solid lines indicate chance-level reconstruction).

Table 1

Statistical summary

Movie 1

Illustration of (top) neutral face stimulus and its reconstruction across 10ms windows of group-based data along with (bottom) the time course of its reconstruction accuracy assessed with an image-based test. The last frame of the movie shows (top) facial image reconstruction achieved with temporally cumulative data and (b) its corresponding level of accuracy (dashed red line).

Table 1. Statistical Table.

Analysis number	Fig	Description	Data Structure	Type of test	Effect	p-values	Power/CI
a	1	P1 component	Assumed normal	repeated measures ANOVA	hemisphere	0.117	0.343
b	1	P1 component	Assumed normal	repeated measures ANOVA	identity	0.39	0.447
c	1	P1 component	Assumed normal	repeated measures ANOVA	identity X hemisphere	0.551	0.311
d	1	N170 component	Assumed normal	repeated measures ANOVA	hemisphere	0.146	0.299
e	1	N170 component	Assumed normal	repeated measures ANOVA	identity	0.513	0.373
f	1	N170 component	Assumed normal	repeated measures ANOVA	identity X hemisphere	0.307	0.532
g	1	N250 component	Assumed normal	repeated measures ANOVA	hemisphere	0.171	0.269
h	1	N250 component	Assumed normal	repeated measures ANOVA	identity	0.001	0.980
i	1	N250 component	Assumed normal	repeated measures ANOVA	identity X hemisphere	0.07	0.560
j	1	N250 component	Assumed normal	repeated measures ANOVA	identity in LH	0.009	0.926
k	1	N250 component	Assumed normal	repeated measures ANOVA	identity in RH	0.004	0.943
l	2a	Group-based discrimination	Normality not assumed	permutation test	across-expression	FDR-corrected p=0.006	
m	2b	Representative participant discrimination	Normality not assumed	permutation test	across-expression	FDR-corrected p=0.004	

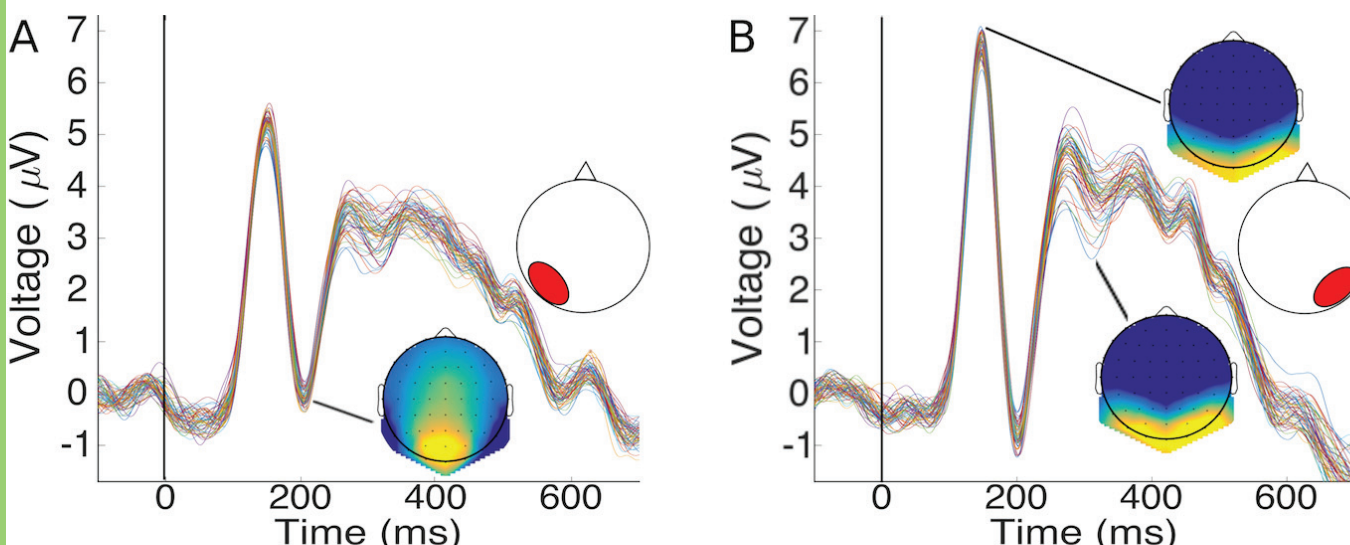
n	3A	Group-based discrimination	Normality not assumed	permutation test	across expression/neutral	0.001	95% CI: 47.1-51.3
o	3A	Group-based discrimination	Normality not assumed	permutation test	across expression/happy	0.001	95% CI: 47.2-50.9
p	3A	Group-based discrimination	Normality not assumed	permutation test	within expression/neutral	0.001	95% CI: 45.9-51.8
q	3A	Group-based discrimination	Normality not assumed	permutation test	within expression/happy	0.001	95% CI: 45.9-52.1
r	3B	Single-participant-based discrimination	Assumed normal	two-tailed t-test	across expression/neutral	0.001	95% CI: 53.1-57
s	3B	Single-participant-based discrimination	Assumed normal	two-tailed t-test	across expression/happy	0.001	95% CI: 53.2-57.2
t	3B	Single-participant-based discrimination	Assumed normal	two-tailed t-test	within expression/neutral	0.001	95% CI: 0.559-0.602
u	3B	Single-participant-based discrimination	Assumed normal	two-tailed t-test	within expression/`happy	0.001	95% CI: 54.8-60.4
v	n/a	Single-participant-based discrimination	Assumed normal	repeated measures ANOVA	discrimination type	<0.001	>0.999
w	n/a	Single-participant-based discrimination	Assumed normal	repeated measures ANOVA	expression	0.466	0.107
x	n/a	Single-participant-based discrimination	Assumed normal	repeated measures ANOVA	discrimination type X expression	0.211	0.230
y	n/a	Single-participant based discrimination within and across databases	Assumed normal	repeated measures ANOVA	discrimination type	<0.001	>0.999
z	n/a	Single-participant based discrimination within and across databases	Assumed normal	repeated measures ANOVA	pairs type	<0.001	>0.999

aa	n/a	Single-participant based discrimination within and across databases	Assumed normal	repeated measures ANOVA	discrimination type X pairs type	0.033	0.565
ab	n/a	Single-participant based discrimination within and across databases	Assumed normal	post hoc (matched two-tailed t-test)	across-within database for within-expression vs. across-within database for across-expression	0.412	95% CI: --0.1-1.5
ac	n/a	Single-participant based discrimination within and across databases	Assumed normal	two-tailed t-test against 0.5	within-expression, within-database discrimination	<0.001	95% CI: 54.6-57.9
ad	n/a	Single-participant based discrimination within and across databases	Assumed normal	two-tailed t-test against 0.5	within-expression, across-database discrimination	<0.001	95% CI: 56.4-61
ae	n/a	Single-participant based discrimination within and across databases	Assumed normal	two-tailed t-test against 0.5	across-expression, within-database discrimination	<0.001	95% CI: 52.8-55
af	n/a	Single-participant based discrimination within and across databases	Assumed normal	two-tailed t-test against 0.5	across-expression, across-database discrimination	<0.001	95% CI: 54-57.7
ag	n/a	temporally cumulative analysis	Assumed normal	Pearson correlation	EEG-based discrimination/ behavioral discrimination	<0.001	0.436
ah	4	temporal correlation	Assumed normal	Pearson correlation	across-expression discrimination/ behavioral discrimination	FDR-corrected p=0.002	
ai	n/a	temporally cumulative analysis	Assumed normal	Pearson correlation	across-expression discrimination/ CFMT	0.219	0.237
aj	n/a	temporally	Assumed	Pearson	across-expression	0.676	0.05

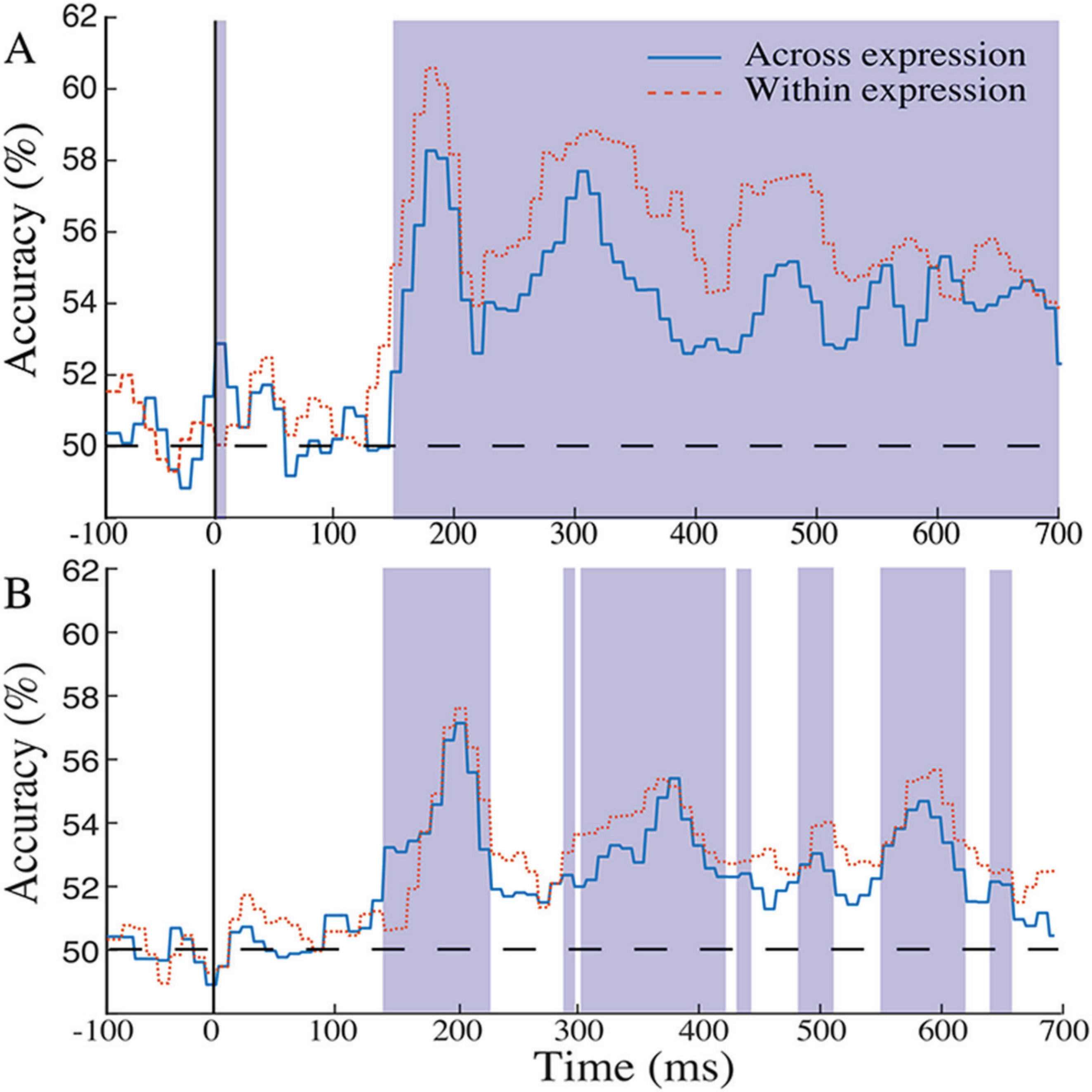
		cumulative analysis	normal	correlation	discrimination/ VVIQ-2		
ak	5B	Face space fit	Normality not assumed	Permutation test	happy-neutral similarity	<0.001	95% CI: 0.738-0.798
al	6A	Classification image	Normality not assumed	Permutation test	neutral dimension 1; luminance	FDR-corrected p=0.027	
am	6A	Classification image	Normality not assumed	Permutation test	neutral dimension 1; red-green	FDR-corrected p=0.018	
an	6A	Classification image	Normality not assumed	Permutation test	neutral dimension 1; yellow-blue	FDR-corrected p=0.024	
ao	6A	Classification image	Normality not assumed	Permutation test	neutral dimension 2; luminance	FDR-corrected p=0.012	
ap	6A	Classification image	Normality not assumed	Permutation test	neutral dimension 2; red-green	N/A	
aq	6A	Classification image	Normality not assumed	Permutation test	neutral dimension 2; yellow-blue	FDR-corrected p=0.009	
ar	6B	Classification image	Normality not assumed	Permutation test	happy dimension 1; luminance	FDR-corrected p=0.040	
as	6B	Classification image	Normality not assumed	Permutation test	happy dimension 1; red-green	FDR-corrected p=0.018	
at	6B	Classification image	Normality not assumed	Permutation test	happy dimension 1; yellow-blue	FDR-corrected p=0.024	
au	6B	Classification image	Normality not assumed	Permutation test	happy dimension 2; luminance	N/A	
av	6B	Classification image	Normality not assumed	Permutation test	happy dimension 2; red-green	FDR-corrected p=0.021	
aw	6B	Classification image	Normality not assumed	Permutation test	happy dimension 2; yellow-blue	FDR-corrected p=0.008	

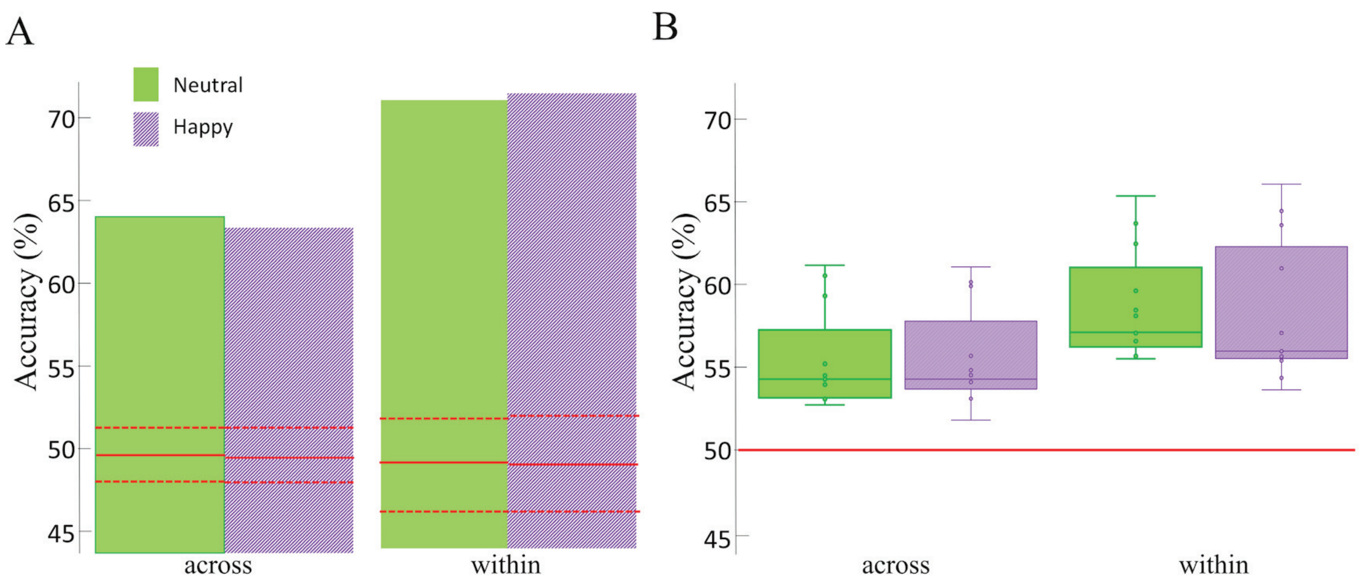
ax	7B	Temporal reconstruction accuracy	Normality not assumed	Permutation test	neutral	FDR-corrected p=0.002	
ay	7B	Temporal reconstruction accuracy	Normality not assumed	Permutation test	happy	FDR-corrected p=0.006	
bz	8B	Reconstruction accuracy (image-based)	Normality not assumed	Permutation test	neutral	0.001	95% CI: 42.3-57.8
ba	8B	Reconstruction accuracy (image-based)	Normality not assumed	Permutation test	happy	0.001	95% CI: 42.1-58.1
bb	8C	Reconstruction accuracy (experimental-based)	Assumed normal	two-tailed t-test	neutral	0.001	95% CI: 55.6-62.6
bc	8C	Reconstruction accuracy (experimental-based)	Assumed normal	two-tailed t-test	happy	0.001	95% CI: 52.4-59.2
bd	n/a	Correlation between experimental and image-based accuracies	Assumed normal	Pearson correlation	neutral	0.001	0.912
be	n/a	Correlation between experimental and image-based accuracies	Assumed normal	Pearson correlation	happy	0.002	0.898
bf	n/a	Correlation between reconstruction and discrimination	Assumed normal	Pearson correlation	averaged across expressions	<0.001	>0.999
bg	n/a	Single-participant –based reconstruction accuracy (image-based)	Assumed normal	two-tailed t-test	neutral	0.027	0.676
bh	n/a	Single-participant –based reconstruction	Assumed normal	two-tailed t-test	happy	0.045	0.580

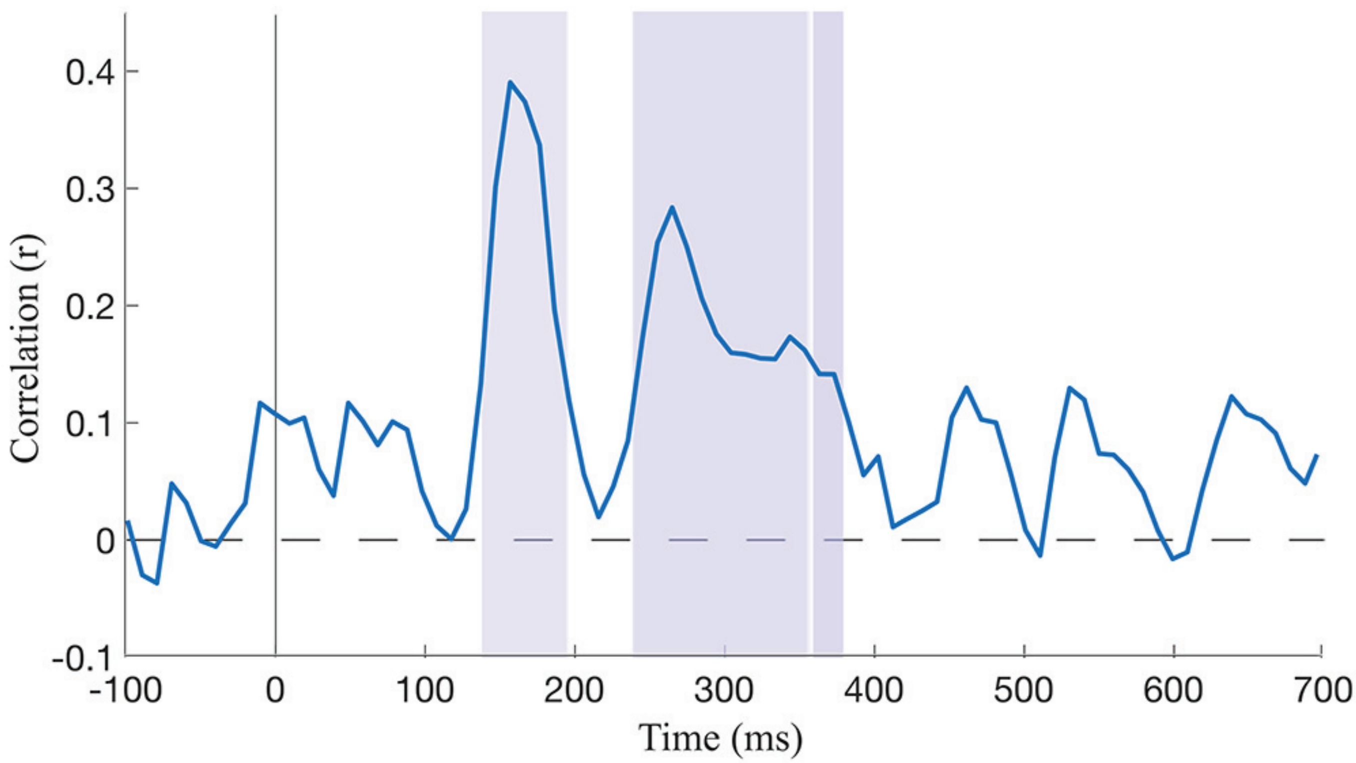
		accuracy (image-based)					
bi	n/a	Correlation between single-participant-based reconstruction and discrimination	Assumed normal	Pearson correlation	neutral	<0.001	0.974
bj	n/a	Correlation between single-participant-based reconstruction and discrimination	Assumed normal	Pearson correlation	happy	<0.001	0.980

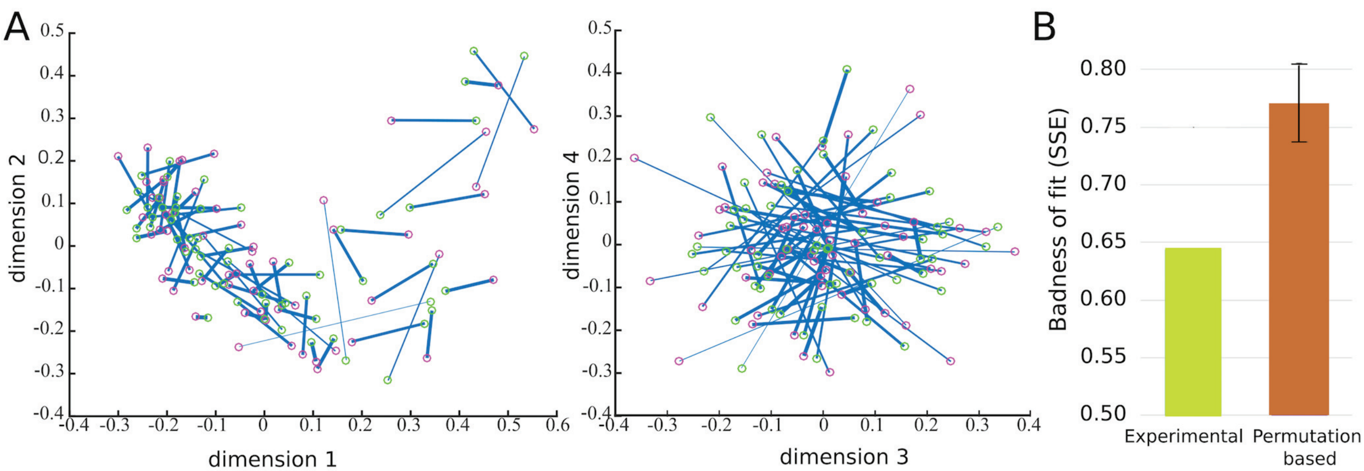


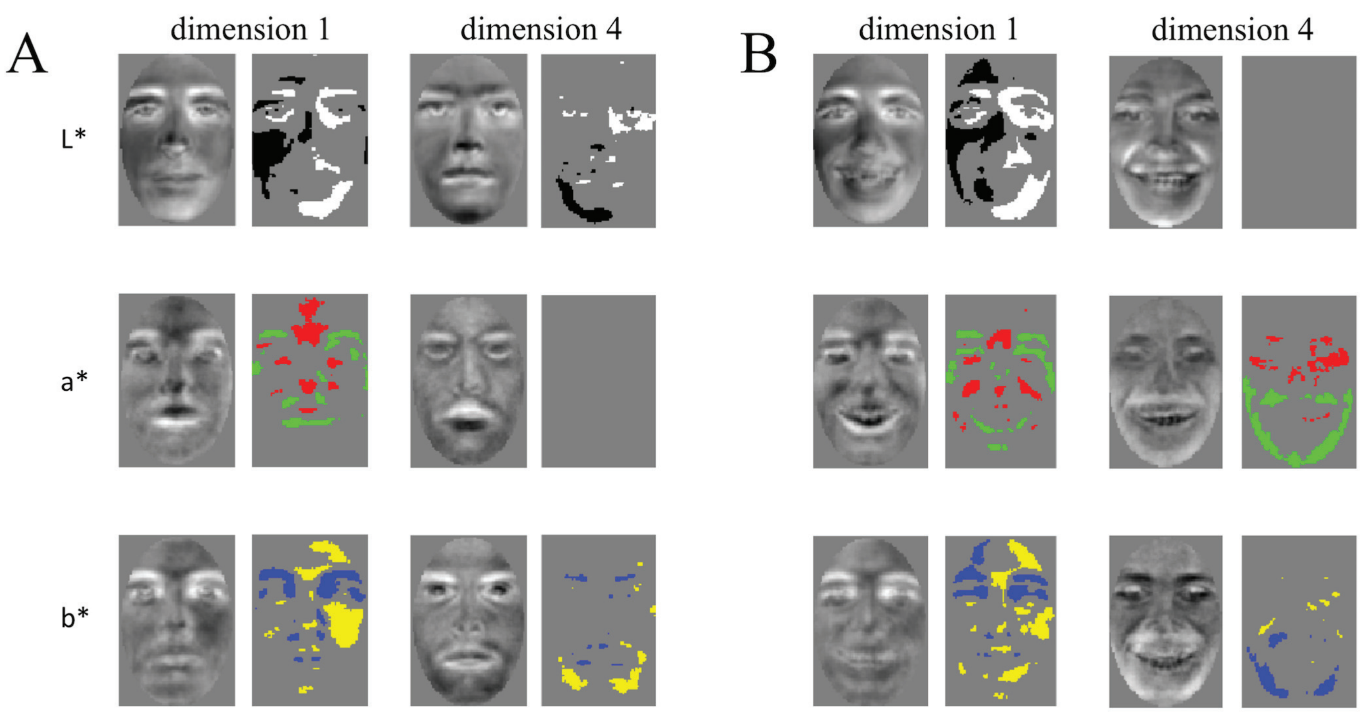
A
B
7
6
5
4
3
2
1
0
-1
Voltage (μV)
0
200
400
600
Time (ms)

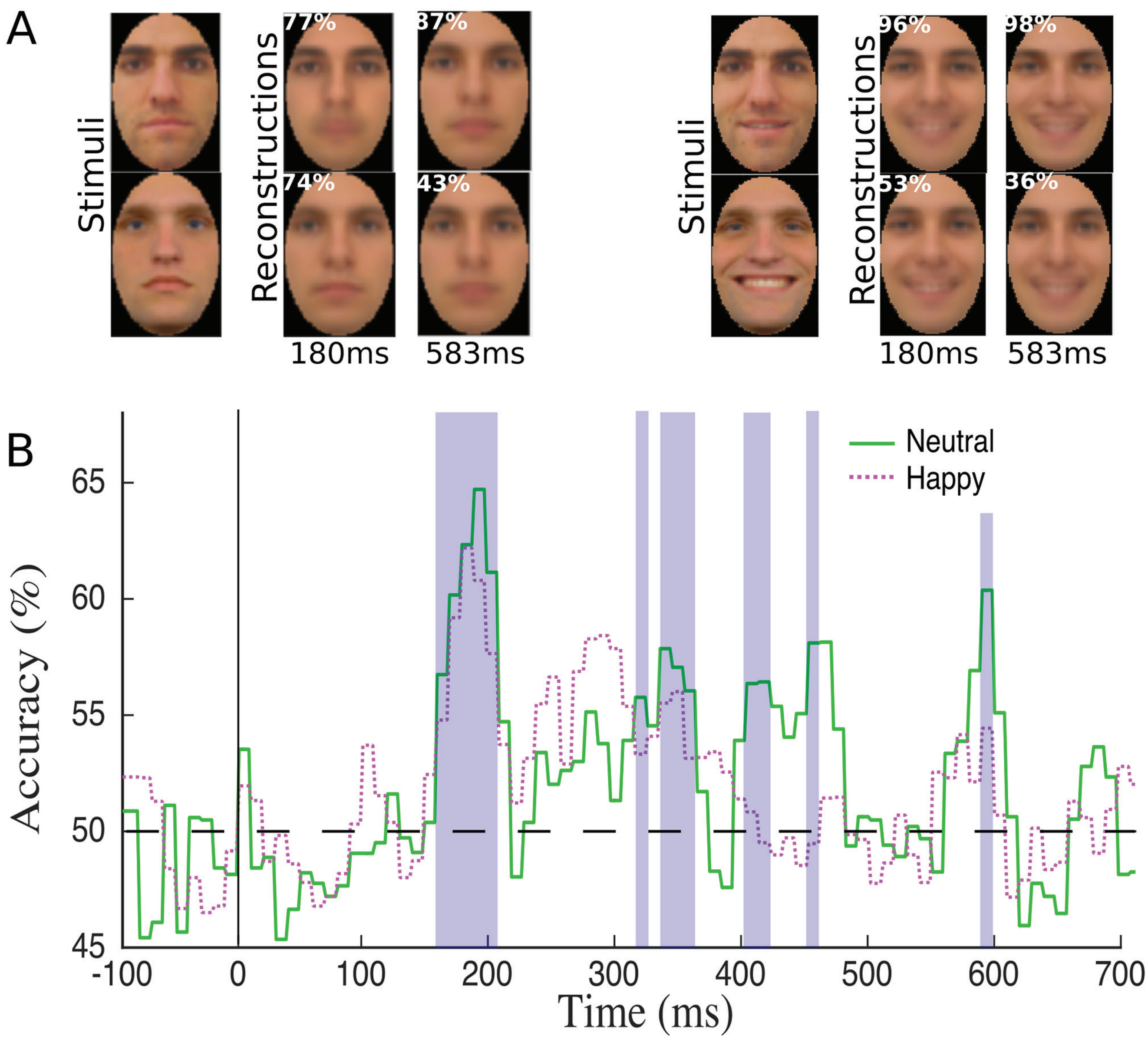

A
B
Across expression
Within expression
Accuracy (%)
Accuracy (%)
62
60
58
56
54
52
50
-100
0
100
200
300
400
500
600
700
Time (ms)

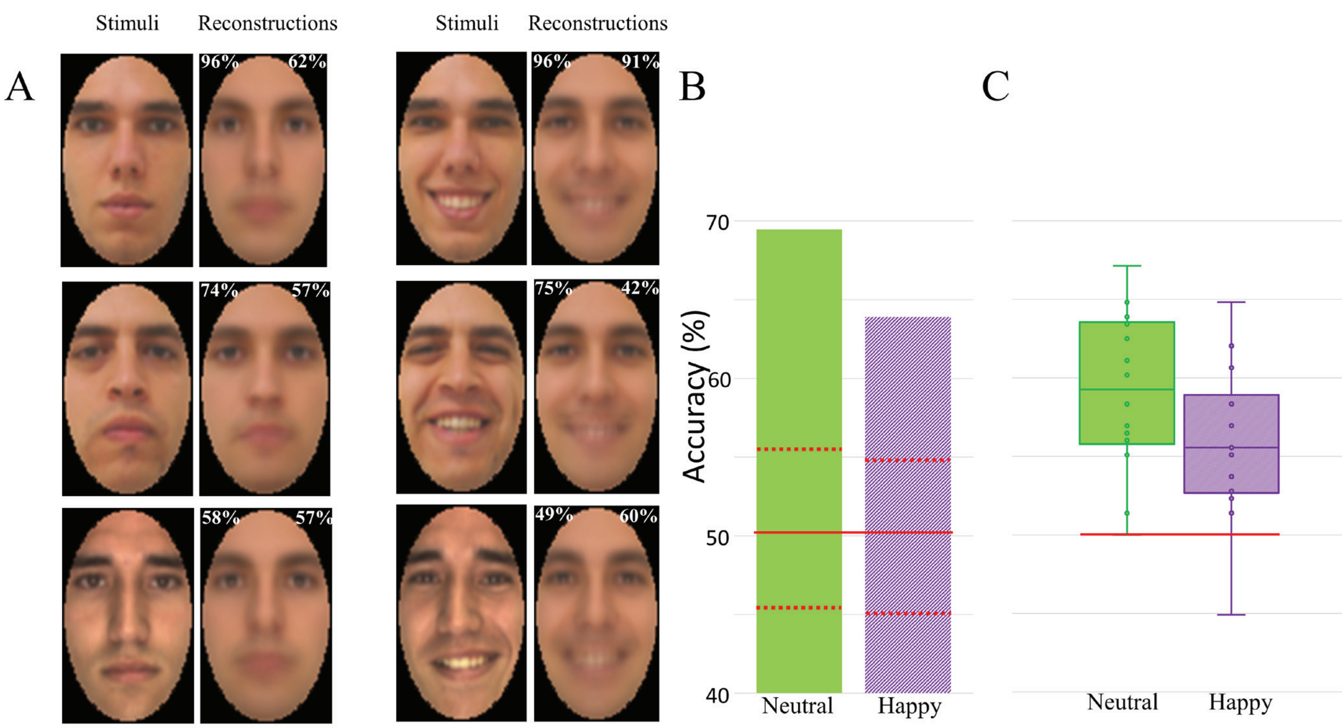

A
Neutral
Happy
70
65
60
55
50
45
Accuracy (%)
across
within
B
70
65
60
55
50
45
Accuracy (%)
across
within


0.4
0.3
0.2
0.1
0
-0.1
Correlation (r)
-100
0
100
200
300
400
500
600
700
Time (ms)


A
dimension 1
dimension 2
dimension 3
dimension 4
B
Badness of fit (SSE)
0.80
0.75
0.70
0.65
0.60
0.55
0.50
Experimental
Permutation based


A
B
dimension 1
dimension 4
dimension 1
dimension 4
L*
a*
b*


A
Stimuli
Reconstructions
77%
87%
74%
43%
180ms
583ms
Stimuli
Reconstructions
96%
98%
53%
36%
180ms
583ms
B
Neutral
Happy
Accuracy (%)
Time (ms)
65
60
55
50
45
-100
0
100
200
300
400
500
600
700



Stimuli
Reconstructions
Stimuli
Reconstructions
A
B
C
96%
62%
74%
57%
58%
57%
96%
91%
75%
42%
49%
60%
Accuracy (%)
70
60
50
40
Neutral
Happy
Neutral
Happy