Task-dependent enhancement of facial expression and identity representations in human cortex

Katharina Dobs [a,b,c,*], Johannes Schultz [a,d], Isabelle Bülthoff [a], Justin L. Gardner [b,e]

[a] Department Human Perception, Cognition and Action, Max Planck Institute for Biological Cybernetics, Max-Planck-Ring 8, 72076 Tübingen, Germany
[b] Laboratory for Human Systems Neuroscience, RIKEN Brain Science Institute, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan
[c] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 43 Vassar Street, Cambridge, MA 02139, USA
[d] Division of Medical Psychology and Department of Psychiatry, University of Bonn, Sigmund Freud Str. 25, 53105 Bonn, Germany
[e] Department of Psychology, Stanford University, 450 Serra Mall, Stanford, CA 94305, USA

ABSTRACT

What cortical mechanisms allow humans to easily discern the expression or identity of a face? Subjects detected changes in expression or identity of a stream of dynamic faces while we measured BOLD responses from topographically and functionally defined areas throughout the visual hierarchy. Responses in dorsal areas increased during the expression task, whereas responses in ventral areas increased during the identity task, consistent with previous studies. Similar to ventral areas, early visual areas showed increased activity during the identity task. If visual responses are weighted by perceptual mechanisms according to their magnitude, these increased responses would lead to improved attentional selection of the task-appropriate facial aspect. Alternatively, increased responses could be a signature of a sensitivity enhancement mechanism that improves representations of the attended facial aspect. Consistent with the latter sensitivity enhancement mechanism, attending to expression led to enhanced decoding of exemplars of expression both in early visual and dorsal areas relative to attending identity. Similarly, decoding identity exemplars when attending to identity was improved in dorsal and ventral areas. We conclude that attending to expression or identity of dynamic faces is associated with increased selectivity in representations consistent with sensitivity enhancement.

Introduction

Humans effortlessly discern two distinct and, indisputably, ecologically important visual aspects of faces: identity and expression. Identity does not change with expression, and, while idiosyncrasies of expression can give clues to identity (O'Toole et al., 2002; Xiao et al., 2014; Lander and Butcher, 2015; Dobs et al., 2016, 2017), is largely associated with static features such as the shape of the eyes and mouth. Conversely, expression is eminently changeable and associated with movement of facial features. What neural mechanisms allow selective extraction of facial form features, such as identity, and facial motion features, such as dynamic expression, from this complex interplay of features? Much work has focused on how attention operates on simpler visual features (Maunsell and Treue, 2006; Bisley, 2011; Carrasco, 2011; Maunsell, 2015) or distinct high-level object categories (e.g., Peelen et al., 2009; Çukur et al., 2013; Peelen and Kastner, 2014). Considerably less is known about mechanisms for selecting different complex aspects of the same stimuli like facial identity and expression.

Attention to identity or expression is known to modulate activity in the diverse set of cortical areas associated with face processing. In humans, face-selective areas have been found along the fusiform (e.g., FFA), the inferior occipital cortex (OFA), and the posterior superior temporal sulcus (STS) (Kanwisher et al., 1997; Haxby et al., 2000), with similar organization in macaques (Tsao et al., 2008; Fisher and Freiwald, 2015; Weiner and Grill-Spector, 2015). While anatomical segregation of processing for expression and identity has been reported (Sergent et al., 1994; Haxby et al., 2000; O'Toole et al., 2002; Andrews and Ewbank, 2004), other reports suggest overlap (Calder and Young, 2005; Bernstein and Yovel, 2015; Lander and Butcher, 2015; Fisher et al., 2016). When subjects selectively attend to expression or eye gaze, greater average BOLD response in STS has been observed, while ventral areas such as FFA showed larger average responses when attending to identity (Hoffman and Haxby, 2000; Narumoto et al., 2001; but see also Ganel et al., 2005).

Attention thus modulates the average BOLD response in face

processing areas; however, this existing evidence is equivocal as to how attentional selection for identity and expression functions. Average modulations of BOLD responses in cortical areas could be interpreted in at least two different ways. First, increased activity in late stage areas could be a signature of processes that improve sensory representations. This sensitivity enhancement would result in enhanced representations of the attended facial aspect, and might be similar to effects reported in early visual areas (Motter, 1994; McAdams and Maunsell, 2000; Reynolds et al., 2000; Martinez-Trujillo and Treue, 2004; Cohen and Maunsell, 2011; for review see Maunsell, 2015) or to effects reported in later face-selective areas when attention is directed to visual features which distinguish individual faces (Gratton et al., 2013). Second, aside from the fact that stronger responses may have higher signal-to-noise ratio, the magnitude of response and the fidelity of response patterns could be unrelated. This would be evident in an increase of overall activity but no increased distinctiveness of the neural representations of the attended facial aspect. Indeed, for representations of the image contrast of visual stimuli, BOLD responses in early visual cortex are increased with spatial attention (Buracas and Boynton, 2007; Li et al., 2008; Murray, 2008; Pestilli et al., 2011), but do not change the slope of their relationship with contrast as would be expected by a higher fidelity representation to differences in contrast. These increased responses can, nonetheless, account for behavioral performance enhancement by an efficient selection model (Pestilli et al., 2011; Hara and Gardner, 2014) in which responses with larger magnitude carry a larger effect on perceptual decisions (Pelli, 1985; Lee et al., 1999; Verghese, 2001; Eckstein et al., 2009; Mante et al., 2013). Thus, at least these two, non mutually-exclusive possibilities of sensitivity enhancement and selection exist for attentional selection for expression and identity of faces.

To test whether a selection mechanism or sensitivity enhancement provides a better account of the neural mechanisms underlying attentional selection of identity or expression, we measured cortical responses while subjects were instructed to attend to the identity or the expression of dynamic face stimuli. Specifically, subjects detected changes in either expression or identity of naturalistically animated avatar faces (Fig. 1). Importantly, these stimuli allow fine-tuning of changes along these dimensions (Dobs et al., 2014), which in turn allows matching task difficulty online during the experiment. We assessed fidelity of representation by our ability to decode (Kamitani and Tong, 2005) individual exemplars of identity and expression in functionally (Kanwisher et al., 1997; Kanwisher and Yovel, 2006; Kanwisher, 2010) and topographically (Wandell, 1999; Wandell et al., 2007) localized cortical areas across the visual hierarchy. We hypothesized that a pure selection account would result in no improvement in classification accuracy, but simply an increase in overall activity. In contrast, sensitivity enhancement would be evident in improved classification accuracy when each facial aspect is attended relative to when that aspect is not attended.

Material and methods

Subjects

Six observers (two female; mean age: 32 years) from the RIKEN Brain Science Institute volunteered as subjects. All observers were right-handed and had normal or corrected-to-normal vision and provided informed written consent prior to the experiment. All procedures for psychophysical and neuroimaging experiments were approved in advance by the RIKEN Functional MRI Safety and Ethics Committee.

Stimuli and display

Each of the four basic stimuli used in this experiment consisted of a short video displaying a female avatar face animated by a facial expression (from neutral to the peak facial expression). We used two female avatar faces and two dynamic facial expressions (Fig. 1A) to create these four clips. Each clip differed from the others either by the face or by the

expression displayed. Two additional avatar faces and two additional facial expressions served as change stimuli for the change detection task. Briefly, the procedures used to make these basic and change animations were as follows (for details see Dobs et al., 2014). First, four facial expressions (angry, happy, disgust, surprise) were motion-recorded from one non-professional female actor following a previously validated and published procedure (Dobs et al., 2014). All facial expressions were 2 s long and started from a neutral expression that proceeded to the peak expression. Second, four avatar faces (ID1−ID4) were designed in Poser 8 (SmithMicro, Inc., Watsonville, CA, USA). We selected two identities (ID1 and ID2) and two facial expressions (angry and happy) to create the basic stimuli. The remaining two identities and two facial expressions were used to create change stimuli that served as targets (see Design and procedure). Change stimuli were created by linearly morphing a basic face into a target face (ID1 to ID3, ID2 to ID4; 10% morph steps) or by linearly morphing one basic facial expression into another target facial expression (angry to disgust, happy to surprise; 5% morph steps). Third, the two basic faces and their facial morphs were animated by the motion-captured facial expressions and their motion morphs (for details about the motion-retargeting procedure see Curio et al., 2006). Importantly, each basic clip could be modified parametrically either in terms of its identity information or its dynamic expression information, allowing matching and continuous control of task difficulty in both tasks (see Design and procedure below). Finally, the animations were rendered as Quicktime movies of 2 s duration (450×600 pixels, 30 frames at 60 Hz) in 3ds Max 2012 (Autodesk, Inc., San Rafael, CA, USA).

Stimuli were presented using MGL (http://justingardner.net/mgl) and Matlab (version R2010a; The MathWorks, Inc., Natick, MA, USA). Images were back-projected on a projection screen (Stewart Filmscreen, Torrance, CA, USA; 28.3×20.0 cm size, 800×600 pixel resolution, 60 Hz refresh rate) located inside the scanner at 50 cm viewing distance from the subject. Animation stimuli were scaled to a size of approximately $9° \times 12°$ and positioned such that the tip of the avatar's nose was located at the center of the screen.

Design and procedure

Subjects performed tasks in which they detected changes in either expression (expression task) or identity (identity task) on a stream of one of the four basic moving face stimuli (Fig. 1A). The design of the task was adapted from a previously published paradigm used to investigate feature-based attention in low-level stimuli (Liu et al., 2011). Each 20 s trial (Fig. 1B) began with a 0.4 s period in which a centrally presented letter instructed subjects to either detect changes in the expression (letter 'E'), identity (letter 'I'), or to perform no task (letter '0'). Throughout the rest of the trial, subjects were required to fixate a cross presented at the center of the screen. Following a 0.2 s gray screen period, eight animated 2 s long face stimuli (Fig. 1A) were shown, each followed by a short gray screen of 0.3 s between stimuli (inter-stimulus interval, ISI). Trials were either no-change or change trials. In no-change trials, the stream of eight face stimuli consisted of eight repetitions of one of the four basic stimuli, whereas change trials consisted of repetitions of one basic stimulus interspersed with change stimuli containing small changes in expression or identity. Subjects had to press a button within the presentation of a stimulus (2 s) or up to 0.5 s afterwards to report a change in the identity (identity task) or expression (expression task) of the stimulus. Reaction time was measured from the beginning of stimulus presentation until the response occurred, or from the beginning of the previous stimulus for responses faster than 0.3 s. The last stimulus was followed by an extra 0.3 s gray screen (total of 0.6 s) to allow for a 2.6 s response window for this final stimulus. Each trial was followed by a 0.5 s inter-trial interval (ITI).

Changes in expression or identity of stimuli were controlled by staircase procedures to maintain performance across tasks, stimuli and subjects at a similar threshold level. In each trial, one to four of the eight presentations of the basic stimuli were randomly replaced by a change
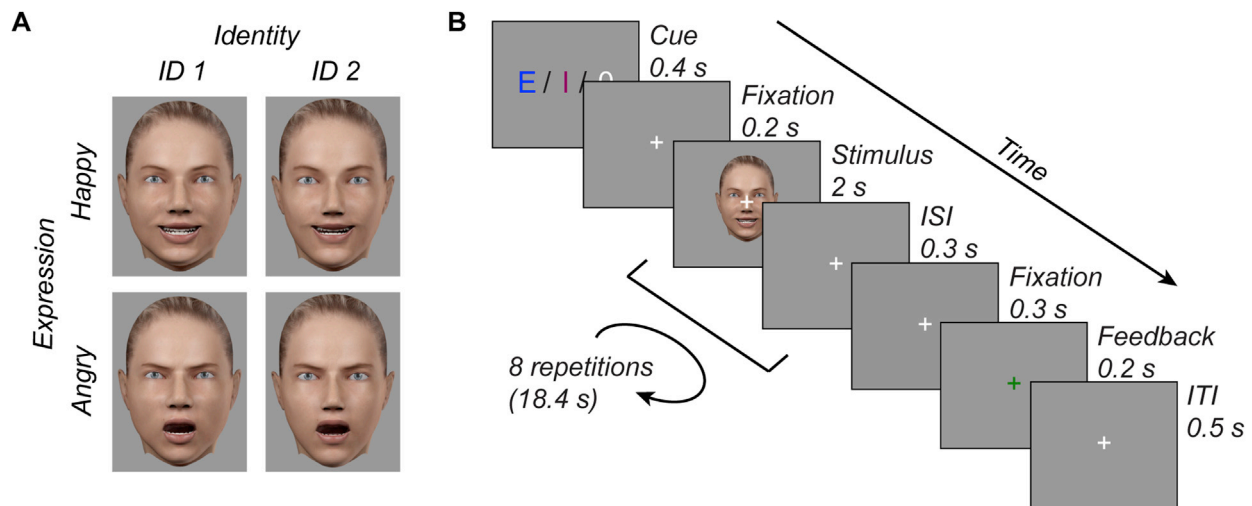
Fig. 1. Experimental stimuli and task. (A) Stimuli. Representative frames of the four animation videos (duration: 2 s each) used as basic stimuli in the experiment. (B) Trial sequence. Subjects were cued to detect changes in expression (letter 'E'), in identity (letter 'I'), or to perform no task (letter '0'). In no-change trials, one of the four stimuli (panel A) was repeatedly shown eight times, followed by a fixation period. Change trials were identical except that between one and four stimuli contained small changes in expression or identity, with the amount of change determined by a staircase procedure running during the experiment. ISI, Inter-stimulus interval; ITI, Inter-trial interval.

stimulus which differed either in expression or identity. In the following, we refer to a change stimulus as a target stimulus if the change was one the subject was instructed to report (e.g., a change in identity in the identity task). In both expression and identity tasks, irrelevant changes were also presented, at the same threshold level as when they were relevant changes, and subjects were required to ignore these changes. For instance, in the identity task, subjects had to respond to a change in identity (i.e., a target stimulus) but ignore changes in expression. For each of the four basic stimuli, the amount of stimulus change (i.e., the morph level) was controlled via a one-up, two-down staircase procedure to maintain performance at a threshold level of 70.7% correct (Levitt, 1971). Note that this adaptive procedure allowed controlling for individual differences in face processing (e.g., due to ethnicity) by equating difficulty across stimuli, tasks and subjects. Detection of target stimuli was counted as correct and both misses and false alarms were counted as errors. At the end of each trial, subjects were given feedback on their performance via a 0.2 s change in the color of the fixation cross. Green indicated that all target changes were detected and all other changes correctly rejected, yellow indicated partial detection or rejection, and red indicated that responses contained only false alarms or misses.

Before the first scanning session, subjects practiced the same change detection task for approximately 1 h in a psychophysics laboratory until they reached a stable threshold level across all four stimuli and both tasks. For each subject, the final threshold level was used as the starting threshold in the following scanning experiments (see Results). To ensure that subjects clearly discriminated the two basic identities (ID1 from ID2) and expressions (angry from happy), we further asked them to label each of the four stimuli before the start of the scanning experiment.

Scanning experiments were identical to the pre-scan psychophysics experiments, except that we included trials without any changes (i.e., no-change trials) and only analyzed the BOLD response obtained in those trials to avoid potential stimulus and motor response difference related confounds. We note that the subjects were not aware in advance about whether a trial contained changes or not and so were expected to maintain the same behavioral state for trials with and without changes.

Each scanning run consisted of eight identity trials, eight expression trials and four no-task trials for a total of 20 trials (400 s/run). One-fourth of task trials (two identity and two expression trials, randomized order) were change trials while the remaining were no-change trials. The order of trials was pseudorandomized such that the first trial in a run was always a no-task trial and each trial type was combined with each of the

four basic stimuli equally often (each stimulus was shown twice in an identity trial, twice in an expression trial and once in a no-task trial). Moreover, presentation of each facial aspect (e.g., expression) was pseudo-randomized such that one exemplar (e.g., "happy") followed another one (e.g., "angry") equally often. No-task trials were discarded from the analysis, thereby eliminating effects of transient T1 signal change commonly seen at the start of an fMRI run. Furthermore, imaging data acquired during change trials were discarded from the analysis to avoid confounds with motor response or effects due to perceptual differences across trials. Thus, only 12 of 20 trials in each run entered the analysis. Together with the presentation time of 18.4 s per facial stimulus in each trial, these manipulations should minimize physiological carry-over effects between trials that could otherwise compromise the validity of the stimuli. Subjects completed a minimum of 16 and up to 20 runs in the scanner divided into two sessions on separate days.

Behavioral data analysis

For each subject, we calculated mean sensitivity (d' value) and reaction time for the two tasks (i.e., expression or identity) and their two corresponding stimuli (e.g., angry or happy) from the responses to all change trials. To test for differences in task difficulty between the experimental conditions, we submitted the sensitivity and reaction time data to a 2 tasks x 2 stimuli repeated-measures ANOVA.

Eye tracking

To ensure that eye movements did not confound our experimental manipulations (i.e., subjects maintained accurate fixation), we monitored eye position using an infrared eye tracker (Eyelink 1000, SR research, 500 Hz) during psychophysical training and scanning. Some eye tracking data was discarded due to insufficient calibration results in the scanner. However, eye tracking data were available for at least one scanning session per subject. Eye position data were analyzed offline and evaluated for stability of fixation using custom Matlab code. We baseline-corrected eye tracking data in each run and calculated the median horizontal and vertical eye position across runs for the experimental conditions. We assessed whether eye position differed across experimental conditions using t-tests (separately for vertical and horizontal eye position) and Hotelling's T^2 test (for both vertical and horizontal eye position).

Magnetic resonance imaging protocol

Scanning was performed at the RIKEN Brain Science Institute (Wako, Japan) on a 4 T Varian Unity Inova whole-body MR system (now Agilent Technologies, Santa Clara, CA, USA) equipped with a head gradient system. In a separate scanning session, a T1-weighted high-resolution 3D anatomical image (MPRAGE; TR, 13 ms; TI, 500 ms; TE, 7 ms; flip angle, 11; voxel size, $1 \times 1x1$ mm; matrix, $256 \times 256 \times 180$) and T2-weighted (TR, 13 ms; TE, 7 ms; flip angle, 11; voxel size, $1 \times 1x1$ mm; matrix, $256 \times 256 \times 180$) fast, low-angle shot sequence were acquired from each subject with a birdcage radio frequency coil. The T1-weighted volume was then divided by the T2-weighted volume to form the reference high-resolution 3D anatomical volume (Van de Moortele et al., 2009). The functional images were acquired using an echo-planar imaging (EPI) pulse sequence with two shots per image (TR, 1.5 s; TE, 30 ms; flip angle, 55; voxel size, $3 \times 3x3$ mm; matrix, 64×64; interleaved) and a 16-channel LifeService coil. Scans were collected in 34 axial slices at an angle approximately perpendicular to the calcarine sulcus. To align the functional data to the high-resolution anatomical images, we also acquired a T1-weighted anatomical at the beginning of each scanning session.

Preprocessing

Preprocessing contained various steps to reduce potential artifacts as follows. Respiration and heartbeat were recorded during scanning and used to attenuate physiological signals in the imaging time series (Cheng et al., 2001) by estimating and correcting physiological fluctuation in k space (Hu et al., 1995). Using customized software, we then corrected the functional data for potential head movements (Nestares and Heeger, 2000) and linearly detrended and high-pass filtered the data with a cutoff at 0.01 Hz.

Task localizer scan

To avoid any circularity issues in choosing voxels within each cortical area for classification analyses, we ran a separate task localizer. At the beginning of each scanning session, we ran two independent scans to identify voxels responding to the experimental task and visual stimuli. We designed the task localizer as a block design experiment. Subjects were instructed to perform the same task as the main experiment. The trial procedure was identical to the main experiment except that the trial duration was reduced to 10 s by repeating stimulus presentation four instead of eight times and task trials were alternated with a 10 s blank (no task) fixation period. Similar to the main experiment, localizer trials consisted of identity and expression trials of which one-fourth were change trials containing one or two stimulus changes.

To identify voxels whose response was modulated by the experimental paradigm (i.e., expression and identity trials), we performed a Fourier-based analysis (Engel et al., 1994; Vintch and Gardner, 2014; for a review see Boynton et al., 2012). For each session, we concatenated the functional data from the two task localizer runs. For each voxel, the resulting time series were Fourier transformed. We then computed coherence as the ratio between the amplitude at the stimulus frequency and the square root of the sum of squares of the amplitudes at all frequencies. The coherence value is a measure of how well the activity of each voxel is modulated by the experimental task. We sorted all voxels in each area in descending order according to their coherence value and selected voxels with a coherence value of 0.4 or higher. However, to keep the number of selected voxels balanced, we restricted the number of voxels in each area to a lower bound of 20 (i.e., selecting also voxels with coherence < 0.4 if necessary) and an upper bound of 150 voxels. For each subject and each area, this selection of voxels was used in all the following analyses.

Retinotopic mapping

In a separate scanning session, we mapped early visual cortex for each subject using standard topographic mapping methodology (Wandell et al., 2007; Gardner et al., 2008). Subjects were shown 10–12 runs of high-contrast sliding radial checkerboard patterns as expanding and contracting rings (two runs each) or clockwise and counterclockwise rotating wedges (three to four runs each) for 10.5 cycles of 24 s. The first half cycle of response was removed and corresponding runs were appropriately time-reversed, shifted in time by a few volumes to compensate for hemodynamic lag and averaged. These averaged voxel time series were Fourier transformed and the coherence and phase of response at the stimulus frequency was displayed on flattened maps of the cortical surface. We defined borders between visual areas as phase reversals in a polar angle map of the visual field. For each subject, we defined the following early visual areas across hemispheres: V1, V2, V3, hV4, and hMT+. The definition of hV4 was a hemifield representation anterior to V3v (Brewer et al., 2005).

Face localizer scan

For each subject, we defined face-sensitive areas in the occipital and temporal lobe based on a separate category localizer scan. The fMRI protocol for this localizer slightly differed from the other scans. Functional images were acquired using an EPI pulse sequence (TR, 1.07 s; TE, 25 ms; flip angle, 63.5; voxel size, $3 \times 3x3$ mm; matrix, 64×64; interleaved) and a 16-channel LifeServices coil. Scans were collected in 27 axial slices aligned with the inferior surface of the occipital and temporal lobes. In a block design, subjects viewed six runs of gray-scale images belonging to nine different categories (human face, building, human body, car, flower, fruit or vegetable, musical instrument, scrambled and gray images) while performing a 1-back repetition detection task. All images were equalized in their magnitude spectrum and DC offset (i.e., mean image luminance) to control for low-level image feature differences, while they were still easily visible as each category. Each run contained 50 randomized blocks (block duration 12.9 s) of different image categories (image display time, 0.75 s; size, $14° \times 14°$) for a total run duration of about 10 min. These imaging data were analyzed using a general linear model with faces and buildings as predictors and modeling blocks of stimulation as box-car functions filtered through a double gamma HRF model. We defined human face-sensitive areas as clusters of voxels which responded more to human faces than to buildings (FDR-corrected $p < 0.05$; Genovese et al., 2002; Kret et al., 2011; Avidan et al., 2014; Goffaux et al., 2016; Watson et al., 2016). Note that although subjects viewed only static images of faces, we found voxels in bilateral superior temporal sulcus responding more to faces than buildings for all subjects. Finally, we compared face-sensitive areas obtained from the face localizer with voxels active in the task localizer to ensure that voxels in these areas were also activated by our experimental task. We then used the combined activation (i.e., the union of activation across localizers) to define face-sensitive areas. We localized OFA in the lateral inferior occipital gyrus (Gauthier et al., 2000), FFA in the mid-fusiform gyrus (Kanwisher et al., 1997; McCarthy et al., 2003), and STS in the posterior part of the superior temporal sulcus. Since studies reported differences in face processing mechanisms across hemispheres (Perrett et al., 1988; Kanwisher et al., 1997; Ishai et al., 1999; Haxby et al., 2000; Rossion et al., 2003; Kanwisher and Yovel, 2006; Meng et al., 2012), we defined face-sensitive areas separately for each hemisphere to assess differences in activation between the two hemispheres.

fMRI data analysis

We asked whether the identity and expression task differently modulated responses in visual cortical areas. More specifically, we asked whether task modulated the fidelity of neural representation of specific exemplars for each facial aspect. We performed different analyses to

address these research questions, as described in the following sections.

BOLD response instances for classifier and amplitude analyses

We computed a BOLD "response instance" for each cortical area representing the pattern of BOLD activity in response to each trial (only task trials without stimulus changes were analyzed). These response instances consisted of scalar values for each voxel indicating its activity as percent signal change for each trial. We obtained response instances as follows. First, we converted the data of each run to percent signal change by dividing the time course of each voxel by its mean signal over this run. Second, we calculated the single-trial fMRI response for all voxels in a cortical area by averaging their response in a 3–21 s time window after trial onset. The time window started 3 s after trial start to compensate for the hemodynamic delay. We performed a time window analysis to rule out the possibility that the selected time window used to average a voxel's response affected decoding performance. A shifted and shortened time window (i.e., from 3-21 s to 7.5–19.5 s) did not significantly change the decoding results (see Supplementary Material). Third, from these scalar responses, we selected responses of voxels that fulfilled the coherence criterion (see above) to form a response instance. This sub selection of voxels was performed to extract voxels within each cortical area that responded specifically to the task using an independent data set to avoid circularity. Finally, for each subject, we concatenated response instances across both sessions. Collections of these instances were then used to run classifier analyses and to calculate the mean response amplitude for experimental conditions. For the data that was used to decode specific exemplars of identity and expression, we further removed the mean across voxels of each response instance to additionally assess whether potential effects were carried in the mean.

Multivoxel pattern classification

To assess whether there were different patterns of response to different conditions, we used binary linear classifier analyses. Specifically, we trained and tested classifiers on the two experimental tasks, and on the specific exemplars of each facial aspect (e.g., angry vs. happy for expression) separately for the identity and expression task. For each subject and each cortical area, we used cross-validation to build and test linear classifiers using Fisher's linear discriminant analysis (Liu et al., 2011). Briefly, to predict the class of a novel instance in the test data set, Fisher's linear discriminant projects the instance onto a weight vector and compares the resulting scalar to a bias point. This weight vector is constructed from the build data based on the difference between the means of the two classes of instances multiplied by the inverse of the covariance matrix. The bias point is calculated as the mean point between the projected instances of the two classes in the build set. To avoid overfitting the covariance matrix, we used a regularization parameter that weights the diagonal (i.e., the variance of each voxel) of the covariance matrix according to the standard deviation of all instances and voxels (Misaki et al., 2010; Haufe et al., 2014). The weight assigned to each voxel using these procedures can be indicative of the contribution each voxel makes to each classification. If there is little or no covariance between voxels (c.f. Haufe et al., 2014) then positive weights indicate higher mean activity in the first class (for example, in our case, expression task), while negative weights indicate higher mean activity in the second class (in our case, identity task).

Classifier accuracy analysis

For classification analyses, we used leave-one-run-out cross-validation and permutation analyses to evaluate classifier accuracy. For each subject and each cortical area, a test set was constructed from one run, and the remaining runs were used to build a classifier. This procedure was repeated for each run of the data. The classifier accuracy was calculated as the number of correct classifications divided by the total number of classifications (across instances and left-out runs). We obtained the mean classifier accuracy by averaging accuracy values across subjects and performed permutation analyses to assess the statistical

significance of this mean classifier accuracy. That is, for each subject, we again performed leave-one-run-out cross-validation but randomly reassigned the labels of the training data. We repeated this procedure 1000 times and averaged the accuracy values across subjects to compute a distribution of expected mean classifier accuracies for a null effect. To obtain a two-tailed p-value, we calculated twice the proportion of the permutation distribution that was greater than or equal to the observed mean classifier accuracy (Nichols and Holmes, 2002). To assess whether these effects were carried by differences in mean activity, we additionally performed the same classification analysis but using the mean-subtracted response instances. To analyze task differences within each facial aspect, we used the same distributions of expected mean classifier accuracies to compute two-tailed permutation tests (Efron and Tibshirani, 1993). Specifically, we calculated the difference in empirical mean classifier accuracies between the two tasks for expression and identity decoding, respectively. This procedure was repeated with the distributions of expected mean classifier accuracies. Finally, we subtracted the empirical difference value from the distribution of expected mean differences and computed the proportion of the permutation distribution which was smaller or greater than or equal to the observed mean difference. The statistical significance was then defined by twice the minimum of these proportions.

Cross-classification analysis

To determine whether the ability to decode one facial aspect (e.g., expression) was invariant over the other facial aspect (e.g., identity), we used a cross-classification analysis. For expression, we trained classifiers on the two expressions performed by the first identity (i.e., ID1) and tested them on the expressions from the second identity (i.e., ID2), and vice versa. Accordingly, for identity, we trained classifiers on both identities performing one expression, and tested on the other expression, and vice versa. For each subject and each facial aspect (e.g., expression), we concatenated all instances across runs, and a test set was constructed by selecting all instances from one condition (happy versus angry expression for ID1), while the remaining instances from the other condition were used to build the classifier (happy versus angry expression for ID2). This procedure was repeated for the other direction (e.g., trained on ID2 and tested on ID1). Classification accuracy for each condition was evaluated on instances from the test sets (i.e., across directions), and we again used permutation analyses to evaluate classifier accuracy (see above).

Single-subject jackknife analysis

To determine whether the attention effects found in the group-averaged data were consistent within subjects or driven by a few outlier data points, we analyzed the classifier performance for each task on a single-subject level. We used a jackknife procedure to estimate the confidence interval of each subject's decoding accuracy. In contrast to a bootstrap procedure which samples with replacement from the available data, jackknife estimates the confidence interval of a data set by iteratively removing one data point (Efron, 1979, 1981; Efron and Tibshirani, 1993). We chose the jackknife procedure instead of the bootstrap to avoid repetitive use of the same data point resulting from sampling with replacement, as this could have lead to a decrease of noise in the data and thus to artificially enhanced classifier accuracy. In contrast, the jackknife is sensitive to outliers and should reveal if classification accuracy is due to some influential data points. Therefore, for each cortical area, we ran a leave-one-trial-out ("jackknife") procedure to estimate the expected accuracy for decoding of identity and expression in each task. To this end, we removed one trial for each condition (e.g., the response instance for one exemplar of each identity, respectively) from the data, and calculated the classifier accuracy as described above based on the reduced dataset. This procedure was repeated for each trial. This resulted in four distributions of classification accuracies (2 facial aspects x 2 tasks) for each cortical area in each subject. For each subject, we used t-tests to compare the accuracy data obtained from the jackknife analysis against chance

level and to compare between conditions.

Classifier weight analysis

To determine if the ability to decode task was due to overall differences in activity across voxels (e.g., an increase or decrease in activity across all voxels for identity or expression task) as found in previous studies (Hoffman and Haxby, 2000; Narumoto et al., 2001), we examined classifier weights along with determining classifier accuracy. We therefore analyzed the weights assigned by the task classifier. For each subject, we constructed a task classifier based on all response instances across runs and read out the corresponding weights assigned by the classifier. We then calculated the mean weights across subjects for each area to assess whether voxels showed higher response in the expression (i.e., positive mean weights) or the identity (i.e., negative mean weights) task. To determine whether these mean weights were statistically different from chance, we performed permutation analyses to obtain the chance level of each area. To that end, for each subject, we constructed a classifier based on randomized labels and calculated the mean weights for 1000 repetitions. The mean of the distribution across permutations and subjects was taken as the expected chance level mean weight. We then compared the observed mean weights for each area with its expected chance level via t-tests. We note that classifier weights may not be equivalent to individual voxel selectivity, particular when there is covariance between voxel responses (Haufe et al., 2014). However, our classifiers were built with a regularization parameter (see Multivoxel pattern classification above) which acts to weaken the effect of covariance on classifier weight determination. Moreover, the classifier weight analysis matched the response magnitude analysis (see below) suggesting that weight maps were little influenced by voxel covariance.

Response amplitude analysis

To assess whether our experimental conditions differed in response amplitude, in addition to response pattern, we examined the mean response amplitude in cortical areas. For each subject and each area, we calculated the mean response amplitude for each condition by concatenating data across response instances and runs and averaging across all voxels. To examine differences in mean activity between the conditions, we used the same response instances as for the classifier analyses. We first assessed whether the experimental tasks (i.e., expression and identity task) differ in mean activity. Compared to the classifier weight analysis reported above, this analysis is expected to be less sensitive as all voxels contribute similarly to the average response. In contrast, only voxels that are informative for the classification are regarded in the classifier weight analysis. To assess the direction of difference in response amplitude between the expression and the identity task, we subtracted the response amplitude observed in the identity task from the response observed in the expression task and compared this difference to zero using t-tests. Further, we determined if there were overall differences in response amplitude between specific exemplars of identity and expression for the experimental tasks. We used t-tests to compare the mean response amplitude between specific exemplars of expression and exemplars of identity for the expression and the identity task, respectively.

Results

Behavioral results

Subjects were able to report selectively about changes in identity and expression (Fig. 2A), as assessed by analysis of their behavioral performance. The average amounts of change at threshold detection were a partial morph between the basic and change stimuli and thus did not reach ceiling for either expression (mean ± STD % morph level: angry: 36.0 ± 7.8, happy: 25.1 ± 9.1) or identity (ID1: 52.1 ± 16.0, ID2: 69.9 ± 16.15). Sensitivity to changes in the face stimuli was high in both tasks (expression task: *d'* mean and SEM = 2.21 ± 0.24, identity task: *d'* mean and SEM = 2.10 ± 0.16). No difference in sensitivity to changes

were found between the identity and expression tasks ($F_{(1,5)} = 0.12$, $p = 0.734$, two-way repeated-measures ANOVA), thus suggesting that our procedures to equate difficulty between the two tasks were successful. Furthermore, no difference in sensitivity between the two expressions (i.e., angry versus happy) or the two identities was found ($F_{(1,5)} = 0.13$, $p = 0.726$).

Despite matching difficulty between tasks, we did find some differences in reaction times (Fig. 2B). Subjects took longer ($F_{(1,5)} = 100.43$, $p < 0.001$, two-way repeated measures ANOVA) to detect a change in expression (mean reaction time and SEM: 1.71 ± 0.14 s, measured from the beginning of each animation) than a change in identity (mean reaction time and SEM: 1.20 ± 0.18 s). This difference likely arises because expression was encoded in facial motion that unfolds in time, compared to static identity information that is available immediately at the beginning of the stimulus. This difference in task demands is a potential confound for the interpretation of higher magnitude BOLD signals (as attention might be deployed longer for the expression task and may lead to different strategy biases). We discuss alternative explanations for the differential activation of the two tasks in the Discussion.

Analysis of eye tracking data showed that subjects maintained stable fixation during all conditions. The median horizontal and vertical eye position during a trial did not depend on the task (horizontal eye position: $t_{(5)} = 1.57$, $p = 0.177$; vertical eye position: $t_{(5)} = 0.44$, $p = 0.681$) and there was no difference in eye position across tasks ($F_{(2,4)} = 3.10$, $p = 0.154$, Hotelling's T^2 test). Moreover, no pairwise comparison of median vertical and horizontal eye position revealed significant differences between facial expressions (angry versus happy: $t_{(5)} <= 1.86$, $p > 0.121$) or identities (ID1 versus ID2: $t_{(5)} <= 1.24$, $p \geq 0.269$) across tasks. Accordingly, in both tasks, subjects' eye position did not depend on the type of expression ($F_{(2,4)} <= 3.64$, $p \geq 0.126$) or identity ($F_{(2,4)} <= 2.08$, $p \geq 0.241$). Thus, we can exclude the contribution of eye movements to any changes in BOLD response across tasks.

Modulations of cortical responses by attention

We asked whether any visual cortical areas showed different patterns of responses between the identity and expression tasks. Any such difference could be attributed to differences in the task and not visual
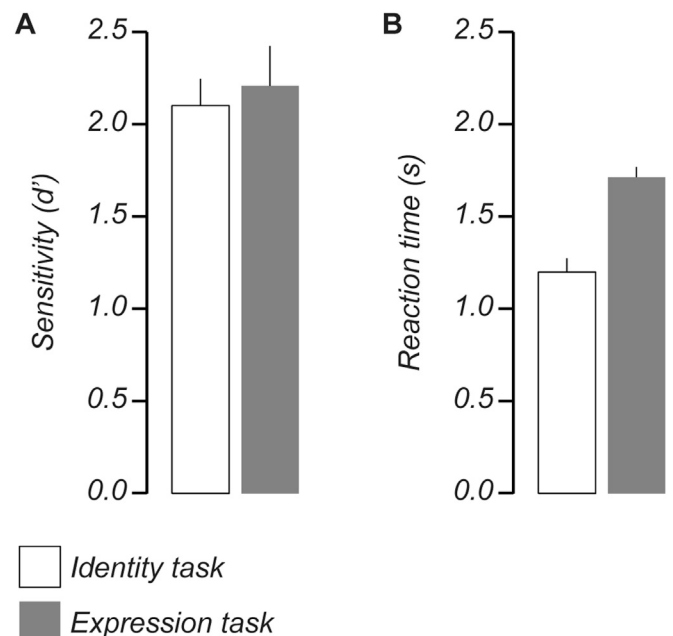
Fig. 2. Behavioral results obtained from change trials. (A) Mean sensitivity (d') and (B) mean reaction times in test trials for each task. Error bars indicate +1 SEM across subjects ($n = 6$).

stimuli, as the actual physical stimuli were identical between these conditions. Classifier performance was above chance in early visual and all face-sensitive areas (Fig. 3, $p \leq 0.008$, permutation test) suggesting that all areas showed different patterns of response for the expression and identity tasks.

We examined whether the ability to decode task was due to overall increases or decreases in response across whole regions of cortex. The weights assigned to each voxel will be positive if a voxel shows higher response in the expression task and negative if the voxel has a higher response in the identity task (when there is little or no covariance between voxels which the regularization of the Fisher discriminant analysis we used promotes; see Methods). We found that all areas were modulated by task but areas were dissociated by which task was associated with greater response as assessed by the classifier weight analysis (Fig. 4; white bars). The expression task recruited dorsal face-sensitive areas (hMT+: t (5) = 2.67; $p = 0.022$; lSTS: t (5) = 6.57, $p < 0.001$), while ventral areas showed enhanced activity in the identity task (lOFA: t (5) = −2.51, $p = 0.027$; lFFA: t (5) = −2.04, $p = 0.049$). Similar to ventral areas, early visual cortex areas responded more to the identity than expression task (V1: t (5) = −2.53, $p = 0.026$; V2: t (5) = −2.52, $p = 0.026$; V3: t (5) = −2.50, $p = 0.027$; trend for hV4: t (5) = −1.97, $p = 0.053$). Areas involved in later stages of face processing showed a distinction between the left and the right hemisphere such that areas in the left hemisphere showed significant differences in overall activity between tasks, while right hemisphere areas did not (rOFA: t (5) = −1.20, $p = 0.142$; rFFA: t (5) = −1.36, $p = 0.115$; rSTS: t (5) = 1.39, $p = 0.111$).

We corroborated the classifier weight analysis showing segregated responses in dorsal and ventral pathways using the simple average amplitude response for each area (Fig. 4; filled bars). We expected the average response analysis to be less sensitive than the classifier weight analysis as all voxels contribute similarly to the average response. In contrast, voxel responses are reflected according to how informative they are in the classifier weight analysis (i.e., non-informative voxels have weights close to zeros and thus do not contribute). Consistent with the classifier weight analysis, we found higher response amplitudes in the expression than identity task in dorsal areas (hMT+: t (5) = 2.88; $p = 0.017$; lSTS: t (5) = 5.33, $p = 0.002$). While ventral areas showed the opposite trend suggesting that these areas yielded larger overall responses for the identity than the expression task (trend for lOFA: t (5) = −1.92, $p = 0.056$; lFFA: t (5) = −1.55, $p = 0.091$). The response amplitude analysis for early visual cortex was also consistent with the classifier weight analysis showing larger responses for the identity than the expression task (V1: t (5) = −2.59, $p = 0.025$; V2: t (5) = −2.98, $p = 0.015$; V3: t (5) = −2.85, $p = 0.018$; trend for hV4: t (5) = −1.99, $p = 0.091$). We also found similar differences between face-sensitive areas in the left and right hemisphere such that mean response amplitudes in right hemisphere areas did not differ between tasks (rSTS: t (5) = 1.28, $p = 0.128$; rOFA: t (5) = −1.32, $p = 0.121$; rFFA: t (5) = −0.42, $p = 0.345$).

Task-dependent modulation of representations of expression and identity exemplars

We asked whether expression information could be selectively enhanced, by examining whether decoding of individual exemplars of expression from cortical responses would change as a function of task. Indeed, we found selective enhancement for decoding of expression
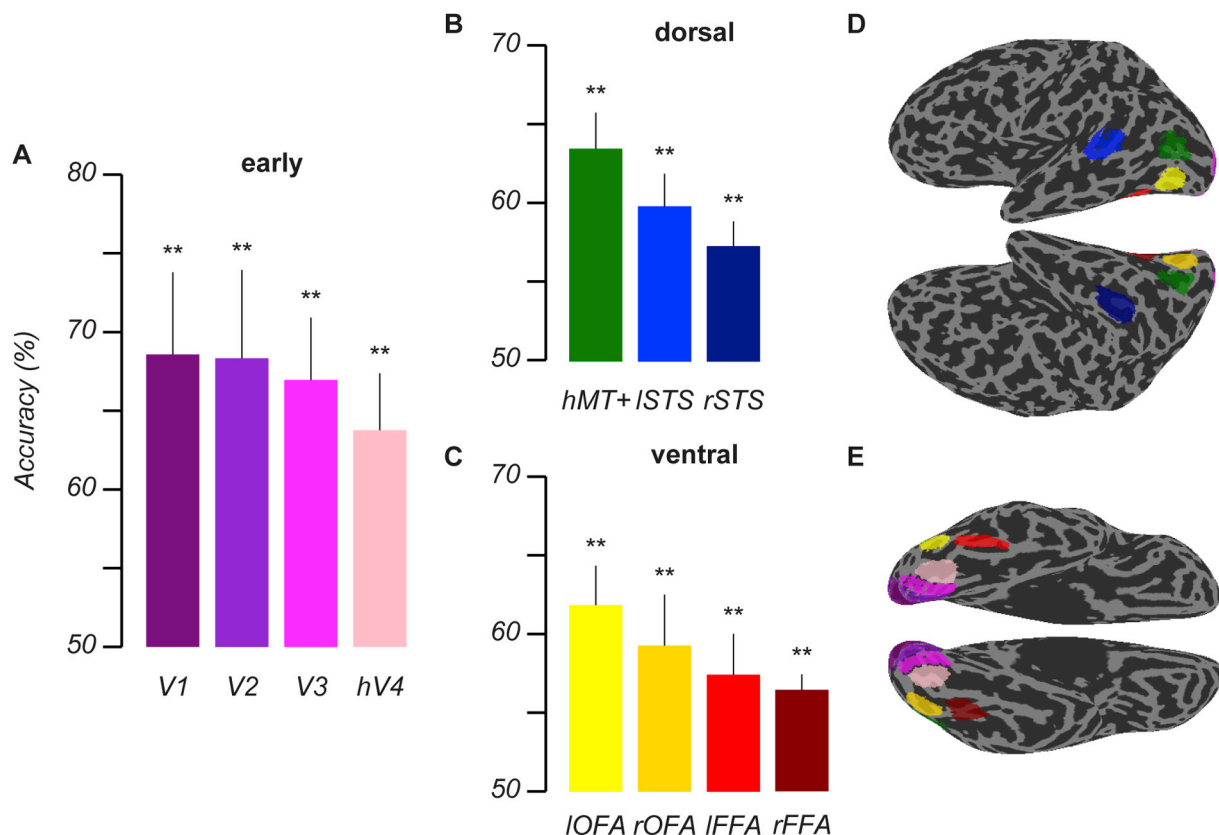
Fig. 3. Both early and late visual areas carried information about task. Mean classifier accuracy for predicting the attended cue across subjects within topographically and functionally defined areas for voxels selected from an independent localizer ($r^2 > 0.4$) separated for early visual (A), dorsal (B) and ventral (C) face-sensitive cortical areas. Chance performance corresponds to 50% correct. Error bars indicate +1 SEM across subjects ($n = 6$). Asterisks indicate significance level obtained from permutation analyses (*: p < 0.05, **: p < 0.01, permutation test). Localized early visual and face-sensitive cortical areas projected on the surface of a sample subject's inflated cortical surface in lateral (D) and inferior (E) view.
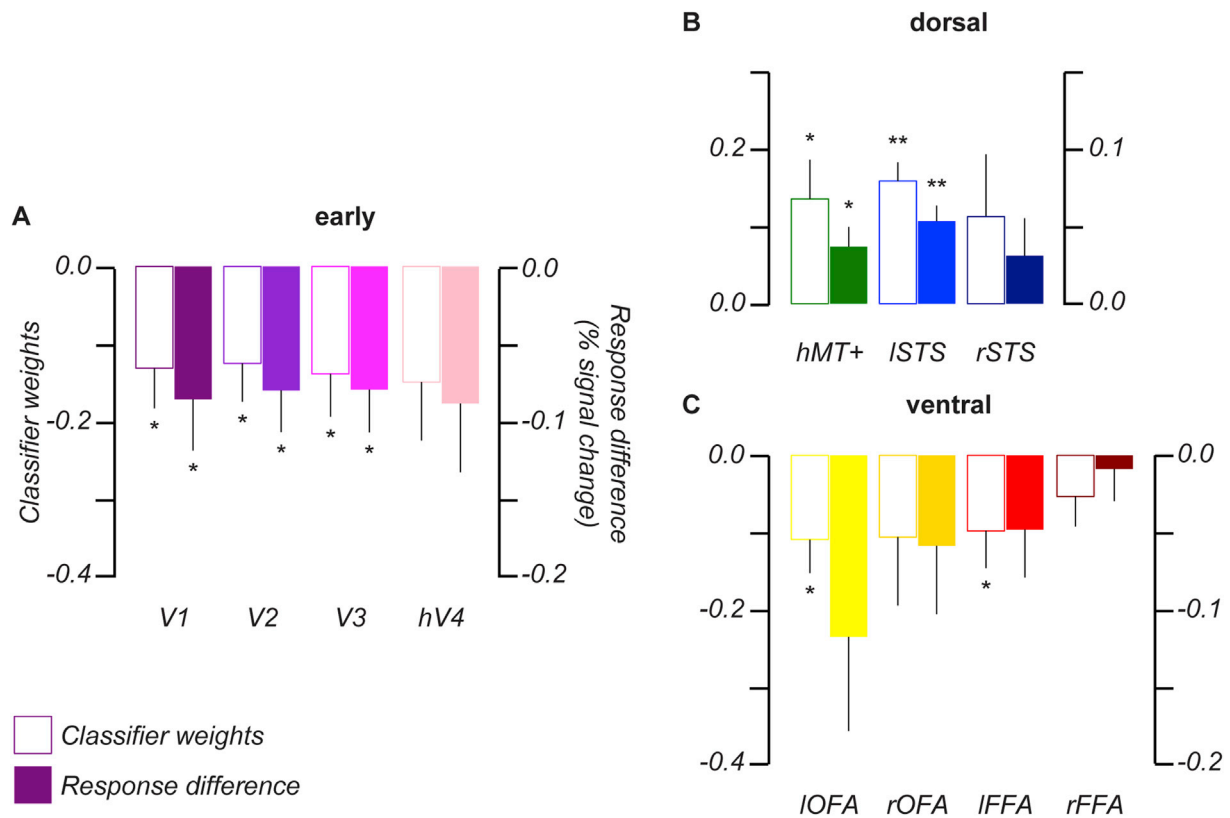
Fig. 4. Pathways segregated by differences in response magnitude between identity and expression tasks. White bars plot mean classifier weights, whereas filled bars plot mean response differences at voxels selected by a localizer ($r^2 > 0.4$) in early visual (A), dorsal (B) and ventral (C) face-sensitive cortical areas. Positive values indicate higher activity in the expression task and negative values indicate higher activity in the identity task. Error bars indicate $+1$ SEM across subjects ($n = 6$). Asterisks indicate significance level obtained from comparing to expected chance level (*: $p < 0.05$, **: $p < 0.01$, t-test).

exemplars during the expression task (Fig. 5, solid bars) compared to the identity task (open bars). In the expression task, early visual cortex areas V1, V2, V3 and hV4 ($p < 0.01$, permutation test), dorsal areas hMT+ ($p < 0.001$) and right STS ($p = 0.004$) all showed above chance classification performance, thus demonstrating that they had distinctive patterns of response for the two different expressions. The ability to decode exemplars of expression was dependent on subjects performing the expression task; when performing the identity task, we could not decode the expressions in any area (accuracy <0.58, $p \geq 0.07$), thus suggesting that the pattern of responses for expressions did not differ significantly. The enhanced decoding of expression exemplars during the expression task was verified as a significant difference between task conditions in all areas being able to decode expression (all $p < 0.05$, permutation test). Overall, expression exemplars could be decoded in early and dorsal visual areas, but only when subjects performed the expression task.

We corroborated these results on averaged data by a single-subject analysis using a jackknife procedure to estimate confidence intervals for each classification accuracy. Confirming the group analysis, we could decode expression exemplars during the expression task from early visual cortex areas V1, V2, V3 and hV4 ($t (26) > 4.43$, $p < 0.001$ for at least five of six subjects; jackknife analysis), hMT+ and right STS ($t (26) > 4.28$, $p < 0.001$ for at least five subjects). During the identity task, similar to the group analysis, area V2, hV4 and rSTS no longer showed above chance classification accuracy ($t (26) > 2.61$, $p < 0.01$ for only three subjects). However, in contrast to the group analysis, we could still decode expression from early visual cortex areas V1, V3 and hMT+ ($t (26) > 3.41$, $p < 0.01$; for at least four subjects). Like the group analysis, enhanced decoding accuracy of expression exemplars with attention was supported by significant task differences in all areas being able to decode expression ($t (52) > 2.13$, $p < 0.05$ for at least four subjects).

We further asked whether the ability to decode expression was

invariant over identity using a cross-classification analysis. Similar to the previous analyses, we could decode expression exemplars only during the expression task in early visual cortex areas V1, V2, V3 and hV4 (mean accuracy: V1: 0.59; V2: 0.57; V3: 0.56; hV4: 0.56; all $p \leq 0.038$, permutation test) and in right STS (rSTS: 0.56; $p = 0.020$). In contrast, we could decode expression in left STS irrespective of the task (expression task: 0.55, $p = 0.034$; identity task: 0.58; $p = 0.004$). Only area hMT+ no longer showed above chance classification performance during the expression task (0.53; $p = 0.306$). We note that this cross-validation scheme resulted in lower classification accuracies in general, presumably due to the reduction of the training data set size by a factor of two. These results suggest that decoding of exemplars of expression was not identity-specific but could generalize across identities.

We performed a similar analysis to search for selective enhancement of identity exemplar representations. Similar to decoding of expression, decoding of identity exemplars improved during the identity task compared to the expression task. Specifically, identity could be decoded from multiple areas when subjects performed the identity task (Fig. 6, open bars higher than solid bars): early visual areas V1 ($p < 0.001$, permutation test) and V3 ($p = 0.020$), face-sensitive areas in the bilateral occipital (lOFA: $p = 0.028$; rOFA: $p = 0.050$), the bilateral ventral temporal lobe (lFFA: $p = 0.010$; rFFA: $p < 0.001$) and the right STS ($p < 0.001$). However, in contrast to expression, decoding of identity remained above chance in the expression task in V1 ($p < 0.001$), right OFA ($p = 0.050$) and right STS ($p = 0.010$). Analysis of task differences in decoding performance revealed enhanced decoding of identity with task in rFFA and rSTS ($p \leq 0.01$, permutation test). This reduced task-dependent enhancement compared to expression decoding could be due to the possibility that facial features useful to discriminate the basic identities from their target identities (as required for the identity task) are different from those useful to discriminate both basic identities (as
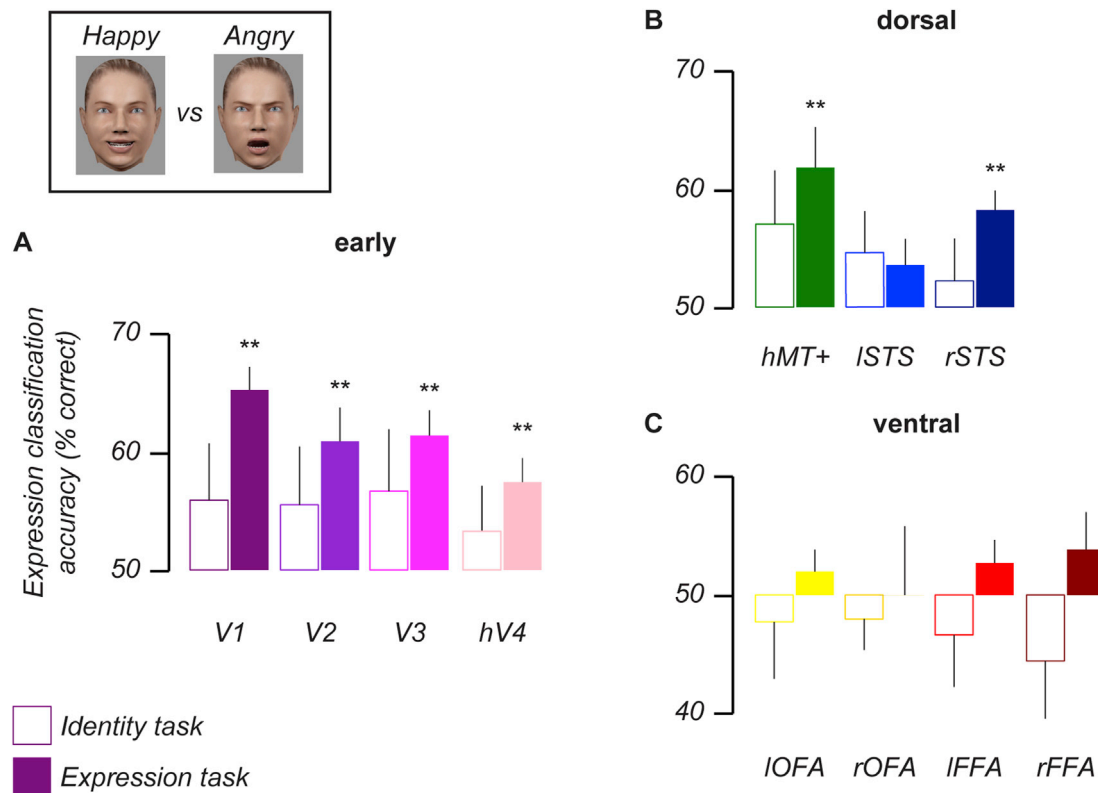
Fig. 5. Representation of expression exemplars was selectively enhanced during the expression task. Mean classifier accuracy for decoding exemplars of expression, shown separately for each task in early visual (A), dorsal (B) and ventral (C) face-sensitive cortical areas. Chance performance corresponds to 50% correct. Error bars indicate +1 SEM across subjects ($n = 6$). Asterisks indicate significance level obtained from permutation analyses (**: $p < 0.01$, permutation test).

required by the decoder). To rule out this possibility, we compared our stimuli using a Gabor similarity analysis. Results, which suggest that the facial features relevant for discriminating both basic identities and for discriminating the basic from the target identities are the same, are reported in the Supplementary Material. In sum, identity exemplars could be decoded in early visual, ventral and dorsal areas, during the identity and the expression task. Similar to expression decoding, performing the identity task enhanced decoding of identity exemplars, but these effects were only found to be statistically significant in later face-processing areas.

To corroborate these group results, we ran a jackknife procedure for each subject. Similar to the group analysis, identity exemplars could be decoded during the identity task from early visual areas V1 and V3 (t (26) > 4.42, $p < 0.001$ for at least four subjects), bilateral OFA (t (26) > 4.35, $p < 0.001$ for at least four subjects), right (t (26) > 5.61, $p < 0.001$ for four subjects) but not left FFA (t (26) > 22.49, $p < 0.001$ for only three subjects), and the right STS (t (26) > 2.72, $p < 0.01$ for all subjects). Confirming the group analysis, decoding of identity remained above chance for the expression compared to the identity task in V1 and V3 (t (26) > 5.17, $p < 0.001$ for at least four subjects), right OFA (t (26) > 2.94, $p < 0.01$ for four subjects) and right STS (t (26) > 13.46, $p < 0.001$ for four subjects). In contrast to the group analysis, enhanced decoding of identity in rFFA and rSTS with task was not supported on the single-subject level. The jackknife analysis revealed high inter-subject variability in these areas (t (26) > 6.13, $p < 0.01$ for two or three subjects, while t (26) < −4.38, $p < 0.01$ in two other subjects).

To investigate whether the decoding of identity exemplars was invariant over expression, we used cross-classification analysis. Similar to the previous analyses, performing the identity task improved identity decoding in later face-processing areas. Above chance decoding performance of identity exemplars during the identity task was observed in

early visual cortex areas V1 and hV4 (mean decoding accuracy: V1: 0.63, $p < 0.001$; hV4: 0.59, $p = 0.002$; permutation test), a trend for left OFA (0.54, $p = 0.080$), left FFA (0.58; $p < 0.001$), and right STS (0.59, $p = 0.004$). Like the classifier analysis reported above, decoding performance remained significant during the expression task in area V1 (0.58, $p = 0.004$) and was marginally significant in right STS (0.55, $p = 0.098$). In contrast, decoding performance remained also significant during the expression task in left OFA (0.56, $p = 0.022$) and left FFA (0.56, $p = 0.012$), and was only significant during the expression task for area V2 and V3 (V2: 0.58, $p < 0.001$; V3: 0.58, $p = 0.002$). In addition, we could decode identity in left STS irrespective of task (identity task: 0.55, $p = 0.030$; expression task: 0.56, $p = 0.044$), while we could no longer decode identity during the identity task from right FFA (0.53, $p = 0.358$) and right OFA (0.46, $p = 0.17$). Overall, decoding accuracy was lower, similar to the cross-classification analysis for expression, presumably due to the lower numbers of training exemplars used. However, we cannot rule out that some of this reduction may also be due to a lack of invariance across expression.

Differences in mean activity for specific exemplars of identity and expression

We asked whether specific exemplars of expression or identity were associated with overall differences in neural responses and could thus explain decoding performance, by examining whether we could still decode exemplars from mean-subtracted cortical responses. For the decoding of expression exemplars, the response patterns were not the result of overall differences in response magnitude to different exemplars of expressions. In all cortical areas, classification accuracy remained significant for data in which the mean across voxels was removed ($p < 0.05$, permutation test). For the decoding of identity exemplars, all
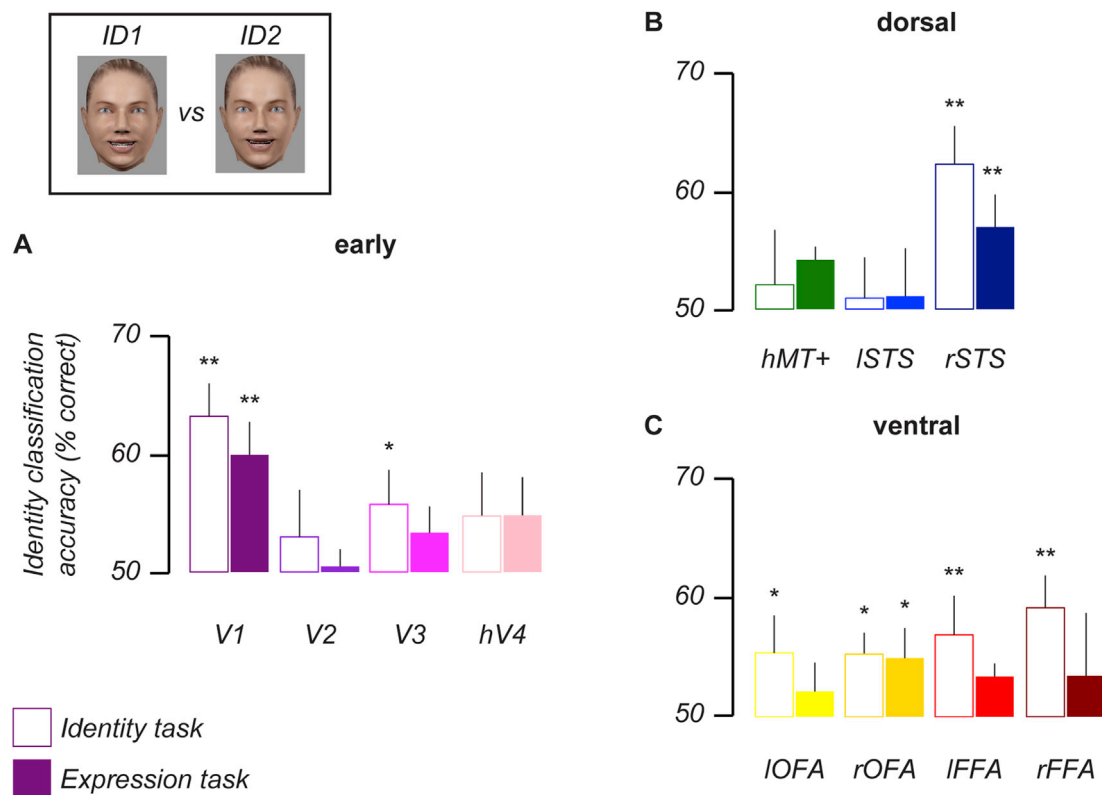
Fig. 6. Representation of identity exemplars was selectively enhanced during the identity task. Mean classifier accuracy for decoding exemplars of identity, shown separately for each task in early visual (A), dorsal (B) and ventral (C) face-sensitive cortical areas. Chance performance corresponds to 50% correct. Error bars indicate +1 SEM across subjects (*n* = 6). Asterisks indicate significance level obtained from permutation analyses (*: $p < 0.05$, **: $p < 0.01$, permutation test).

effects were carried by differences in the pattern of responses, with the exception of OFA, for which removing the mean across voxels made decoding identity exemplars non-significant (lOFA: $p = 0.158$, rOFA: $p = 0.180$ for mean-subtracted data). In line with the results from mean-subtracted responses, analysis of differences in mean activity in cortical areas did not reveal any differences between exemplars of expressions in the expression task ($t(5) <= 1.71, p \geq 0.148$ in all areas) and the identity task ($t(5) < 1.98, p > 0.105$ in all areas), nor between identities in the identity task ($t(5) <= 1.88, p > 0.118$ in all areas) and the expression task ($t(5) <= 2.55; p > 0.051$ in all other areas) except for a significant difference in area V1 ($t(5) = 3.67, p = 0.014$). We conclude that the observed effects were mainly due to differences in pattern and not to overall differences in mean activation suggesting that decoding results could not be simply explained by enhanced neural responses or differences in reaction times to specific exemplars of expression or identity, respectively.

Discussion

We conclude that attending to expression or identity significantly improves cortical representations of these face aspects in later stage cortical areas, consistent with attention acting on representations through sensitivity enhancement mechanisms (Motter, 1994; McAdams and Maunsell, 2000; Martinez-Trujillo and Treue, 2004; Liu et al., 2011; Gratton et al., 2013). Using multivoxel pattern analysis, we found that attention to expression increased decoding accuracy of exemplars of expression (angry versus happy) in early visual and dorsal areas, indicative of increased discriminability of the representations of the attended facial aspect. Decoding accuracy of identity exemplars was significantly improved during the identity task in two areas (rFFA and rSTS), but did not reach significance in early visual cortex.

Consistent with previous studies (Hoffman and Haxby, 2000;

Narumoto et al., 2001; but see also Ganel et al., 2005), attention to expression increased overall BOLD activity in dorsal areas whereas attention to identity increased cortical responses in ventral face-sensitive areas. We extended these findings by showing that attention to identity generally increased responses in early visual cortex areas similarly to ventral areas, while response in motion-processing area hMT+ was higher for the expression than the identity task (LaBar et al., 2003; Furl et al., 2013).

Methodological considerations and advancements

Our experimental protocol avoided potential confounds due to differences in stimulus or task difficulty between identity and expression tasks. We used an adaptive staircase procedure and computer-generated stimuli, such that we could independently manipulate expression and identity differences to keep both at individual subjects' discrimination thresholds. Further, we only analyzed imaging data from trials in which no changes in stimuli were presented and subjects thus made no overt responses. As a result, we can rule out effects of differences in response times and motor processes across tasks related to making faster or slower reports. Longer duration of attention during the expression compared to the identity task, suggested by longer response times observed in expression- compared to identity-change trials, cannot fully explain the pattern of results either. If longer reaction times occurred because subjects attended longer and this resulted in higher neural activity, then we would expect larger overall BOLD responses for the expression than the identity task. Indeed, previous studies reported a positive relationship between duration of physiological processes (as measured by reaction time) and BOLD amplitude (Binder et al., 2005; Yarkoni et al., 2009). In contrast, for most of the areas, mean responses were larger when attending to identity even though reaction times were shorter than when attending to expression. The only exceptions were dorsal areas hMT+

and left STS, consistent with previous studies reporting neural activity in dorsal areas during processing of facial expressions based on static (Hoffman and Haxby, 2000; Narumoto et al., 2001; Ganel et al., 2005) and dynamic faces (LaBar et al., 2003; Furl et al., 2013).

However, it is important to note that, despite all stimuli being dynamic faces, changes in expression and identity in this study differently affected dynamic and static features: while identity changes only led to changes in static features, expression changes affected dynamic and (particularly near the end of a stimulus) static features. As a result, certain differences in task demands, for example different response strategies, may have played a role in the differential activation across expression and identity tasks. These differences may correspond to the inherent inequality of processing expression and identity: When seeing a face, static identity information is available immediately, while information about expression evolves over time. We found that activation in early visual cortex areas was enhanced during the identity task, while area hMT+ showed higher BOLD signal during the expression task. This fits well with the notion that discriminating identity required attending to subtle form features, while discriminating expression required attending to visual motion. We thus cannot determine whether differences between the identity and the expression task are due to differences in the weighting of form versus motion processing or more specifically to differences between identity versus expression processing. Future work will be necessary to disentangle these intermingled factors.

Despite an equal level of difficulty, the tasks had different effects on response magnitude in different cortical areas. Larger BOLD magnitude might be expected to have higher signal-to-noise ratio and thereby enhance classification accuracy (Carlin, 2015). However, the pattern of differences in response magnitude across tasks alleviates the concern that the enhanced classification accuracy on different tasks could be a simple consequence of larger BOLD responses. Specifically, some areas, particularly in early visual cortex (V1-V3, hV4), showed better classification accuracies despite lower responses when attending to expression (c.f. Kok et al., 2012; Pratte and Tong, 2014). However, modulation of BOLD magnitude with task still reflects different task demands for the identity and the expression task. These differences in task demands might explain why we could no longer decode expression exemplars from STS during the identity task, whereas other studies using orthogonal tasks still reported significant decoding of expressions in STS (Wegrzyn et al., 2015; Zhang et al., 2016).

Our experimental task bears some similarity with Garner's inference paradigm (Garner, 1976; for a recent review see Algom and Fitousi, 2016), where observers are asked to make judgments using one attribute of a visual stimulus while ignoring another. The Garner paradigm was designed to investigate the interference between stimulus attributes. Using this paradigm, symmetric interference effects were found for expression and identity processing (Ganel and Goshen-Gottstein, 2004), but also contradictory results were reported (Schweinberger and Soukup, 1998; Schweinberger et al., 1999). While our task could be used to address similar questions, here, we adapted a task that was previously used to investigate feature-based attention of color and motion stimuli (Liu et al., 2011). This task served three major purposes. First, it controls attentional state by requiring subjects to maintain attention throughout the trials. Second, it allows online matching difficulty between the two attentional tasks. This latter property is an advantage compared to the Garner paradigm in which the difficulty of discrimination between two dimensions must be equated through pretesting and is typically not updated during the experiment. Third, we excluded confounds due to motor processes related to overt behavioral responses. Whereas in the Garner paradigm observers have to respond during the conditions under investigation, in our experimental paradigm, we only analyze conditions in which there was no change in any stimulus attribute, and thus no motor response occurred.

We used naturalistically moving, computer generated faces to investigate expression and identity information in the human cortex. In contrast to other studies, we found a strong segregation of expression

exemplar information in regions that did not include the ventral temporal cortex. The number of subjects in our study limits the interpretations of this result and we cannot exclude that the absence of expression information in the ventral stream is due to a lack of power. However, another possible explanation is that previous studies have used uncontrolled natural movies of expressions (Skerry and Saxe, 2014) or may have confounded identity with expression by using only static faces as stimuli (Vuilleumier et al., 2001; Pessoa et al., 2002; Surguladze et al., 2003; Winston et al., 2003, 2004; Ishai et al., 2004; Fox et al., 2009b; Kadosh et al., 2010; Xu and Biederman, 2010; Nestor et al., 2011; Harry et al., 2013; Wegrzyn et al., 2015; Zhang et al., 2016). Indeed, recent work suggests that ventral face-sensitive areas process expressions by extracting form variations in expressions (Said et al., 2011; Bernstein and Yovel, 2015). In line with this hypothesis, one study directly compared static with dynamic faces and found that FFA showed sensitivity for static expressions, whereas STS showed sensitivity for dynamic expressions (Furl et al., 2013). In our study, we were able to apply the same facial motion to both identities, thus eliminating any identity information carried by individual differences in the way an expression is performed (O'Toole et al., 2002; Knappmeyer et al., 2003; Lander et al., 2006). Future studies should further investigate if expression information in the ventral stream can be explained by form variations in static images of expression.

The use of dynamic face stimuli might be important for decoding identity information in the STS. Previous studies have suggested that STS may integrate identity, transmitted by form, and expression, transmitted by motion in humans (Giese and Poggio, 2003; Puce et al., 2003; Lange and Lappe, 2006; Furl et al., 2015; Dobs et al., 2017) and monkeys (Desimone and Ungerleider, 1986; Oram and Perrett, 1996). While studies based on adaptation techniques (Grill-Spector and Malach, 2001; Naccache and Dehaene, 2001) using static face stimuli reported identity as well as expression information in right STS (Winston et al., 2004; Fox et al., 2008), recent studies based on decoding techniques did not find significant identity information in STS (Axelrod and Yovel, 2015; Wegrzyn et al., 2015; Zhang et al., 2016). However, this may be because STS responds more weakly to static compared to dynamic faces (Kanwisher et al., 1997; Fox et al., 2009a; Schultz and Pilz, 2009; Schultz et al., 2013).

While our stimuli and task procedures were thus designed to control for task difficulty while allowing for enough repeated trials to perform classification analyses, these experimental conditions necessitated limited number of exemplars of identity and expression. With just a couple identities and expressions, the generalizability of our results to a wider range of identities and expressions is unknown. It is possible that discrimination is based on a very limited set of low-level features of the stimuli. However, we note that subjects were able to discriminate expression over different identities and vice-versa. Similarly, the results from the cross-classification analyses suggest that decoding of one facial aspect was, at least partly, invariant over the other facial aspect. The ability to generalize the results from one exemplar (e.g., ID1) to the other (e.g., ID2) suggests that the strategies used to discriminate one exemplar from its target are at least not tailored to an individual exemplar. It does not, however, rule out that the attentional strategies used in our identity and expression change detection tasks may to some degree differ from strategies used in everyday face processing tasks, such as identity recognition.

Another potential concern with our stimuli is that individual exemplars of identity and expression were not matched in terms of visual similarity. Thus, differences in decoding expression exemplars, for example, might show larger effects of task because their visual differences were more distinct. However, at least three mitigating factors suggest that this was not the case. First, differences in the identity exemplars and expression exemplars were well-above psychophysical threshold and all subjects could easily discriminate them. Second, identity and expression produced cortical representations that were similarly discriminable in terms of classifier accuracy (median accuracy: identity:

0.55, expression 0.54, max. accuracy: identity: 0.63, expression: 0.65), thus nearly equating representational similarity. This representational similarity is of key importance for our analysis as lack of improvement in classification accuracy could otherwise be a ceiling or floor effect. Third, with our stimuli, identity discrimination may rely on more subtle featural differences and thus be less discriminable than expression. Note that this is in contrast to other stimulus sets where identity is typically more discriminable than expression. In our stimulus set, we used the same texture for both facial identities, thereby reducing variance, and the two expressions chosen differed largely in the mouth area. Thus, if identity was less discriminable than expression, then enhancement with attention might be expected to have improved the separability of identity representations more than representations of expression, which were already more separable at the start. As this is opposite to what we found, this alternative explanation is unlikely.

We also cannot rule out the possibility that highly idiosyncratic and exemplar specific attentional strategies may have contributed to our findings. For example, subjects could have developed an attentional strategy to detect changes between the basic identities and their targets that may have been less efficient in enhancing (i.e., separating) the basic identities themselves. However, our Gabor analysis of local image similarity (see Supplementary Material) did not find any evidence for specific facial features that were more distinct than others that might promote such a strategy. Ideally, of course, many different identities and expressions should be tested to rule-out these concerns, but given the multiple sessions per subject needed for the current study, collecting sufficient data to perform classification analysis across a larger set of identities and expressions may require a prohibitive amount of data.

Mechanisms of attention for complex high-level visual representations

The enhanced accuracy in decoding expression from early visual cortex during the expression task suggests that the mechanism for attentional selection of complex, high-level aspects of faces is accomplished, at least in part, by modulating responses in the earliest parts of the cortical processing stream. That we were able to decode identity and expression from early visual cortex is not surprising, given that these facial aspects differ in low level visual features (Henriksson et al., 2015), which are known to be represented by neurons in these areas (Hubel and Wiesel, 1962). While it is possible that identity and expression information is fed back (Williams et al., 2008; Petro et al., 2013; Cohen and Tong, 2015), or at least interacts (Goffaux et al., 2016) with information from higher cortical areas, our decoding results do not constitute evidence for face-specific representations in these early visual areas. We found enhanced decoding accuracy for expression when subjects performed the expression task even though the low-level stimulus features were not different between task conditions. For identity, we found enhancement of representation in only two late-stage areas and nominal but not statistically significant effects in early visual cortex. We cannot rule out the possibility that measurement of BOLD responses and our analysis lacked the resolution and power to resolve stronger effects of task on identity in early visual cortex as were evident for expression. Also it is important to note that we directed subjects' attention to subtle changes in facial features. As this task emphasizes attending to the physical appearance of an identity, we cannot exclude that another task which involves more abstract identity processing may have different effects on identity representations. Nonetheless, our data are consistent with the idea that enhancement of sensitivity to low-level visual features needed to encode more complex facial aspects are a relevant mechanism of attentional selection for high-level visual stimuli like faces.

The human cortical face processing system affords an excellent opportunity to understand mechanisms of attentional selection for a complex high-level perceptual system. Our results provide evidence for sensitivity enhancement of expression and identity with selective attention, and thus flexibility of representation depending on task, consistent with recent behavioral evidence (Stoesz and Jakobson, 2013; Xiao et al.,

2013, 2014; Lander and Butcher, 2015). This flexibility together with the finding that we could decode identity in right STS, which was previously reported to process facial expression (Sergent et al., 1994; Haxby et al., 2000; O'Toole et al., 2002; Andrews and Ewbank, 2004), is aligned with recent reports suggesting that identity and expression are not as rigidly and independently represented in the neural face processing system as previously assumed (Calder and Young, 2005; Bernstein and Yovel, 2015; Lander and Butcher, 2015; Fisher et al., 2016). The attentional mechanisms we have found that enhance representations for expression and identity in this high-level system appear to share characteristics found in studies reporting selective response enhancement (e.g., Motter, 1994; McAdams and Maunsell, 2000) and improved representation of features (e.g., Martinez-Trujillo and Treue, 2004) in early visual areas (for reviews see Maunsell and Treue, 2006; Carrasco, 2011; Maunsell, 2015), or enhanced representation of object categories (e.g., Peelen et al., 2009; Çukur et al., 2013; for a review see Peelen and Kastner, 2014). Taken together, attentional enhancement mechanisms may be shared across multiple levels of the visual system. Enhancement for expression could have been found only at the highest representational levels of processing when representations of faces are more complete. Instead, our findings suggest that top-down modulation even for highly complex visual stimuli such as faces are implemented across all levels of the visual system with similar mechanisms operating at different levels of the processing hierarchy.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi. org/10.1016/j.neuroimage.2018.02.013.

References

Algom, D., Fitousi, D., 2016. Half a century of research on Garner interference and the separability–integrality distinction. Psychol. Bull. 142, 1352–1383.

Andrews, T.J., Ewbank, M.P., 2004. Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. Neuroimage 23, 905–913.

Avidan, G., Tanzer, M., Hadj-Bouziane, F., Liu, N., Ungerleider, L.G., Behrmann, M., 2014. Selective dissociation between core and extended regions of the face processing network in congenital prosopagnosia. Neuroimage 24, 1565–1578.

Axelrod, V., Yovel, G., 2015. Successful decoding of famous faces in the fusiform face area. PLoS One 10, e0117126.

Bernstein, M., Yovel, G., 2015. Two neural pathways of face processing: a critical evaluation of current models. Neurosci. Biobehav. Rev. 55, 536–546.

Binder, J.R., Medler, D.A., Desai, R., Conant, L.L., Liebenthal, E., 2005. Some neurophysiological constraints on models of word naming. Neuroimage 27, 677–693.

Bisley, J.W., 2011. The neural basis of visual attention. J. Physiol. 589, 49–57.

Boynton, G.M., Engel, S.A., Heeger, D.J., 2012. Linear systems analysis of the fMRI signal. Neuroimage 62, 975–984.

Brewer, A.A., Liu, J., Wade, A.R., Wandell, B.A., 2005. Visual field maps and stimulus selectivity in human ventral occipital cortex. Nat. Neurosci. 8, 1102–1109.

Buracas, G.T., Boynton, G.M., 2007. The effect of spatial attention on contrast response functions in human visual cortex. J. Neurosci. 27, 93–97.

Calder, A.J., Young, A.W., 2005. Understanding the recognition of facial identity and facial expression. Nat. Rev. Neurosci. 6, 641–651.

Carlin, J.D., 2015. Decoding face exemplars from fMRI responses: what works, what Doesn't? J. Neurosci. 35, 9252–9254.

Carrasco, M., 2011. Visual attention: the past 25 years. Vis. Res. 51, 1484–1525.

Cheng, K., Waggoner, R.A., Tanaka, K., 2001. Human ocular dominance columns as revealed by high-field functional magnetic resonance imaging. Neuron 32, 359–374.

Cohen, E.H., Tong, F., 2015. Neural mechanisms of object-based attention. Neuroimage 25, 1080–1092.

Cohen, M.R., Maunsell, J.H.R., 2011. Using neuronal populations to study the mechanisms underlying spatial and feature attention. Neuron 70, 1192–1204.

Curio, C., Breidt, M., Kleiner, M., Vuong, Q.C., Giese, M.A., Bülthoff, H.H., 2006. Semantic 3d motion retargeting for facial animation. ACM Trans. Appl. Percept. 77–84.

Çukur, T., Nishimoto, S., Huth, A.G., Gallant, J.L., 2013. Attention during natural vision warps semantic representation across the human brain. Nat. Neurosci. 16, 763–770.

Desimone, R., Ungerleider, L.G., 1986. Multiple visual areas in the caudal superior temporal sulcus of the macaque. J. Comp. Neurol. 284, 164–189.

Dobs, K., Bülthoff, I., Breidt, M., Vuong, Q.C., Curio, C., Schultz, J., 2014. Quantifying human sensitivity to spatio-temporal information in dynamic faces. Vis. Res. 100, 78–87.

Dobs, K., Bülthoff, I., Schultz, J., 2016. Identity information content depends on the type of facial movement. Sci. Rep. 6, 1–9.

Dobs, K., Ma, W.J., Reddy, L., 2017. Near-optimal integration of facial form and motion. Sci. Rep. 7, 1–9.

Eckstein, M.P., Peterson, M.F., Pham, B.T., Droll, J.A., 2009. Statistical decision theory to relate neurons to behavior in the study of covert visual attention. Vis. Res. 49, 1097–1128.

Efron, B., 1979. Bootstrap methods: another look at the jackknife. Ann. Stat. 7, 1–26.

Efron, B., 1981. Nonparametric estimates of standard error: the jackknife, the bootstrap and other methods. Biometrika 68, 589–599.

Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap. CRC Press.

Engel, S.A., Rumelhart, D.E., Wandell, B.A., Lee, A.T., Glover, G.H., Chichilnisky, E.J., Shadlen, M.N., 1994. fMRI of human visual cortex. Nature 369, 525.

Fisher, C., Freiwald, W.A., 2015. Contrasting specializations for facial motion within the macaque face-processing system. Curr. Biol. 25, 261–266.

Fisher, K., Towler, J., Eimer, M., 2016. Facial identity and facial expression are initially integrated at visual perceptual stages of face processing. Neuropsychologia 80, 115–125.

Fox, C.J., Iaria, G., Barton, J.J.S., 2009a. Defining the face processing network: optimization of the functional localizer in fMRI. Hum. Brain Mapp. 30, 1637–1651.

Fox, C.J., Moon, S.Y., Iaria, G., Barton, J.J.S., 2009b. The correlates of subjective perception of identity and expression in the face network: an fMRI adaptation study. Neuroimage 44, 569–580.

Fox, C.J., Oruc, I., Barton, J.J.S., 2008. It doesn't matter how you feel. The facial identity aftereffect is invariant to changes in facial expression. J. Vis. 8, 11–11.

Furl, N., Henson, R.N., Friston, K.J., Calder, A.J., 2013. Top-down control of visual responses to fear by the amygdala. J. Neurosci. 33, 17435–17443.

Furl, N., Henson, R.N., Friston, K.J., Calder, A.J., 2015. Network interactions explain sensitivity to dynamic faces in the superior temporal sulcus. Neuroimage 25, 2876–2882.

Ganel, T., Goshen-Gottstein, Y., 2004. Effects of familiarity on the perceptual integrality of the identity and expression of faces: the parallel-route hypothesis revisited. J. Exp. Psychol. Hum. Percept. Perform. 30, 583–597.

Ganel, T., Valyear, K.F., Goshen-Gottstein, Y., Goodale, M.A., 2005. The involvement of the "fusiform face area" in processing facial expression. Neuropsychologia 43, 1645–1654.

Gardner, J.L., Merriam, E.P., Movshon, J.A., Heeger, D.J., 2008. Maps of visual space in human occipital cortex are retinotopic, not spatiotopic. J. Neurosci. 28, 3988–3999.

Garner, W.R., 1976. Interaction of stimulus dimensions in concept and choice processes. Cognit. Psychol. 8, 98–123.

Gauthier, I., Tarr, M.J., Moylan, J., Skudlarski, P., Gore, J.C., Anderson, A.W., 2000. The fusiform "face area" is part of a network that processes faces at the individual level. J. Cognit. Neurosci. 12, 495–504.

Genovese, C.R., Lazar, N.A., Nichols, T., 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage 15, 870–878.

Giese, M.A., Poggio, T., 2003. Neural mechanisms for the recognition of biological movements. Nat. Rev. Neurosci. 4, 179–192.

Goffaux, V., Duecker, F., Hausfeld, L., Schiltz, C., Goebel, R., 2016. Horizontal tuning for faces originates in high-level Fusiform Face Area. Neuropsychologia 81, 1–11.

Gratton, C., Sreenivasan, K.K., Silver, M.A., D'Esposito, M., 2013. Attention selectively modifies the representation of individual faces in the human brain. J. Neurosci. 33, 6979–6989.

Grill-Spector, K., Malach, R., 2001. fMR-adaptation: a tool for studying the functional properties of human cortical neurons. Acta Psychol. 107, 293–321.

Hara, Y., Gardner, J.L., 2014. Encoding of graded changes in spatial specificity of prior cues in human visual cortex. J. Neurophysiol. 112, 2834–2849.

Harry, B., Williams, M.A., Davis, C., Kim, J., 2013. Emotional expressions evoke a differential response in the fusiform face area. Front. Hum. Neurosci. 7, 1–6.

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., Bießmann, F., 2014. On the interpretation of weight vectors of linear models in multivariate neuroimaging. Neuroimage 87, 96–110.

Haxby, J.V., Hoffman, E.A., Gobbini, M.I., 2000. The distributed human neural system for face perception. Trends Cognit. Sci. 4, 223–233.

Henriksson, L., Mur, M., Kriegeskorte, N., 2015. Faciotopy- A face-feature map with face-like topology in the human occipital face area. Cortex 72, 156–167.

Hoffman, E.A., Haxby, J.V., 2000. Distinct representations of eye gaze and identity in the distributed human neural system for face perception. Nat. Neurosci. 3, 80–84.

Hu, X., Le, T.H., Parrish, T., Erhard, P., 1995. Retrospective estimation and correction of physiological fluctuation in functional MRI. Magn. Reson. Med. 34, 201–212.

Hubel, D.H., Wiesel, T.N., 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. 160, 106–154.

Ishai, A., Pessoa, L., Bikle, P.C., Ungerleider, L.G., 2004. Repetition suppression of faces is modulated by emotion. PNAS 101, 9827–9832.

Ishai, A., Ungerleider, L.G., Martin, A., Schouten, J.L., Haxby, J.V., 1999. Distributed representation of objects in the human ventral visual pathway. PNAS 96, 9379–9384.

Kadosh, K.C., Henson, R., Kadosh, R.C., 2010. Task-dependent activation of face-sensitive cortex: an fMRI adaptation study. J. Cognit. Neurosci. 22, 903–917.

Kamitani, Y., Tong, F., 2005. Decoding the visual and subjective contents of the human brain. Nat. Neurosci. 8, 679–685.

Kanwisher, N., 2010. Functional specificity in the human brain: a window into the functional architecture of the mind. PNAS 107, 11163–11170.

Kanwisher, N., McDermott, J., Chun, M.M., 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci. 17, 4302–4311.

Kanwisher, N., Yovel, G., 2006. The fusiform face area: a cortical region specialized for the perception of faces. Philos. Trans. R. Soc. Lond. B Biol. Sci. 361, 2109–2128.

Knappmeyer, B., Thornton, I.M., Bülthoff, H.H., 2003. The use of facial motion and facial form during the processing of identity. Vis. Res. 43, 1921–1936.

Kok, P., Jehee, J.F.M., de Lange, F.P., 2012. Less is more: expectation sharpens representations in the primary visual cortex. Neuron 75, 265–270.

Kret, M.E., Pichon, S., Grèzes, J., de Gelder, B., 2011. Similarities and differences in perceiving threat from dynamic faces and bodies. An fMRI study. Neuroimage 54, 1755–1762.

LaBar, K.S., Crupain, M.J., Voyvodic, J.T., McCarthy, G., 2003. Dynamic perception of facial affect and identity in the human brain. Neuroimage 13, 1023–1033.

Lander, K., Butcher, N., 2015. Independence of face identity and expression processing: exploring the role of motion. Front. Psychol. 6, 255.

Lander, K., Chuang, L., Wickham, L., 2006. Recognizing face identity from natural and morphed smiles. Q. J. Exp. Psychol. B 59, 801–808.

Lange, J., Lappe, M., 2006. A model of biological motion perception from configural form cues. J. Neurosci. 26, 2894–2906.

Lee, D.K., Itti, L., Koch, C., Braun, J., 1999. Attention activates winner-take-all competition among visual filters. Nat. Neurosci. 2, 375–381.

Levitt, H., 1971. Transformed up-down methods in psychoacoustics. J. Acoust. Soc. Am. 49, 467–477.

Li, X., Lu, Z.L., Tjan, B.S., Dosher, B.A., Chu, W., 2008. Blood oxygenation level-dependent contrast response functions identify mechanisms of covert attention in early visual areas. PNAS 105, 6202–6207.

Liu, T., Hospadaruk, L., Zhu, D.C., Gardner, J.L., 2011. Feature-specific attentional priority signals in human cortex. J. Neurosci. 31, 4484–4495.

Mante, V., Sussillo, D., Shenoy, K.V., Newsome, W.T., 2013. Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature 503, 78–84.

Martinez-Trujillo, J.C., Treue, S., 2004. Feature-based attention increases the selectivity of population responses in primate visual cortex. Curr. Biol. 14, 744–751.

Maunsell, J.H.R., 2015. Neuronal mechanisms of visual attention. Annu Rev. Vis. Sci. 1, 373–391.

Maunsell, J.H.R., Treue, S., 2006. Feature-based attention in visual cortex. Trends Neurosci. 29, 317–322.

McAdams, C.J., Maunsell, J.H., 2000. Attention to both space and feature modulates neuronal responses in macaque area V4. J. Neurophysiol. 83, 1751–1755.

McCarthy, G., Puce, A., Gore, J.C., Allison, T., 2003. Face-specific processing in the human fusiform gyrus. J. Cognit. Neurosci. 9, 605–610.

Meng, M., Cherian, T., Singal, G., Sinha, P., 2012. Lateralization of face processing in the human brain. CerCor 279, 2052–2061.

Misaki, M., Kim, Y., Bandettini, P.A., Kriegeskorte, N., 2010. Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. Neuroimage 53, 103–118.

Motter, B.C., 1994. Neural correlates of feature selective memory and pop-out in extrastriate area V4. J. Neurosci. 14, 2190–2199.

Murray, S.O., 2008. The effects of spatial attention in early human visual cortex are stimulus independent. J. Vis. 8, 2–2.

Naccache, L., Dehaene, S., 2001. The priming method: imaging unconscious repetition priming reveals an abstract representation of number in the parietal lobes. Neuroimage 11, 966–974.

Narumoto, J., Okada, T., Sadato, N., Fukui, K., Yonekura, Y., 2001. Attention to emotion modulates fMRI activity in human right superior temporal sulcus. Cognit. Brain Res. 12, 225–231.

Nestares, O., Heeger, D.J., 2000. Robust multiresolution alignment of MRI brain volumes. Magn. Reson. Med. 43, 705–715.

Nestor, A., Plaut, D.C., Behrmann, M., 2011. Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. PNAS 108, 9998–10003.

Nichols, T.E., Holmes, A.P., 2002. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum. Brain Mapp. 15, 1–25.

Oram, M.W., Perrett, D.I., 1996. Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey. J. Neurophysiol. 76, 109–129.

O'Toole, A.J., Roark, D.A., Abdi, H., 2002. Recognizing moving faces: a psychological and neural synthesis. Trends Cognit. Sci. 6, 261–266.

Peelen, M.V., Fei-Fei, L., Kastner, S., 2009. Neural mechanisms of rapid natural scene categorization in human visual cortex. Nature 460, 94–97.

Peelen, M.V., Kastner, S., 2014. Attention in the real world: toward understanding its neural basis. Trends Cognit. Sci. 18, 242–250.

Pelli, D.G., 1985. Uncertainty explains many aspects of visual contrast detection and discrimination. J. Opt. Soc. Am. A 2, 1508–1532.

Perrett, D.I., Mistlin, A.J., Chitty, A.J., Smith, P.A.J., Potter, D.D., Broennimann, R., Harries, M., 1988. Specialized face processing and hemispheric asymmetry in man and monkey: evidence from single unit and reaction time studies. Behav. Brain Res. 29, 245–258.

Pessoa, L., McKenna, M., Gutierrez, E., 2002. Neural processing of emotional faces requires attention. PNAS 99, 11458–11463.

Pestilli, F., Carrasco, M., Heeger, D.J., Gardner, J.L., 2011. Attentional enhancement via selection and pooling of early sensory responses in human visual cortex. Neuron 72, 832–846.

Petro, L.S., Smith, F.W., Schyns, P.G., Muckli, L., 2013. Decoding face categories in diagnostic subregions of primary visual cortex. Eur. J. Neurosci. 37, 1130–1139.

Pratte, M.S., Tong, F., 2014. Spatial specificity of working memory representations in the early visual cortex. J. Vis. 14, 22–22.

Puce, A., Syngeniotis, A., Thompson, J.C., Abbott, D.F., Wheaton, K.J., Castiello, U., 2003. The human temporal lobe integrates facial form and motion: evidence from fMRI and ERP studies. Neuroimage 19, 861–869.

Reynolds, J.H., Pasternak, T., Desimone, R., 2000. Attention increases sensitivity of V4 neurons. Neuron 26, 703–714.

Rossion, B., Joyce, C.A., Cottrell, G.W., Tarr, M.J., 2003. Early lateralization and orientation tuning for face, word, and object processing in the visual cortex. Neuroimage 20, 1609–1624.

Said, C.P., Haxby, J.V., Todorov, A., 2011. Brain systems for assessing the affective value of faces. Philos. Trans. R. Soc. Lond. B Biol. Sci. 366, 1660–1670.

Schultz, J., Brockhaus, M., Bülthoff, H.H., Pilz, K.S., 2013. What the human brain likes about facial motion. Neuroimage 23, 1167–1178.

Schultz, J., Pilz, K.S., 2009. Natural facial motion enhances cortical responses to faces. Exp. Brain Res. 194, 465–475.

Schweinberger, S.R., Burton, A.M., Kelly, S.W., 1999. Asymmetric dependencies in perceiving identity and emotion: experiments with morphed faces. Percept. Psychophys. 61, 1102–1115.

Schweinberger, S.R., Soukup, G.R., 1998. Asymmetric relationships among perceptions of facial identity, emotion, and facial speech. J. Exp. Psychol. Hum. Percept. Perform. 24, 1748.

Sergent, J., Ohta, S., MacDonald, B., Zuck, E., 1994. Segregated processing of facial identity and emotion in the human brain: a pet study. Vis. Cognit. 1, 349–369.

Skerry, A.E., Saxe, R., 2014. A common neural code for perceived and inferred emotion. J. Neurosci. 34, 15997–16008.

Stoesz, B.M., Jakobson, L.S., 2013. A sex difference in interference between identity and expression judgments with static but not dynamic faces. J. Vis. 13, 26–26.

Surguladze, S.A., Brammer, M.J., Young, A.W., Andrew, C., Travis, M.J., Williams, S.C.R., Phillips, M.L., 2003. A preferential increase in the extrastriate response to signals of danger. Neuroimage 19, 1317–1328.

Tsao, D.Y., Moeller, S., Freiwald, W.A., 2008. Comparing face patch systems in macaques and humans. PNAS 105, 19514–19519.

Van de Moortele, P.-F., Auerbach, E.J., Olman, C., Yacoub, E., Uğurbil, K., Moeller, S., 2009. T_1 weighted brain images at 7 Tesla unbiased for Proton Density, T_2* contrast and RF coil receive B1 sensitivity with simultaneous vessel visualization. Neuroimage 46, 432–446.

Verghese, P., 2001. Visual search and attention: a signal detection theory approach. Neuron 31, 523–535.

Vintch, B., Gardner, J.L., 2014. Cortical correlates of human motion perception biases. J. Neurosci. 34, 2592–2604.

Vuilleumier, P., Armony, J.L., Driver, J., Dolan, R.J., 2001. Effects of attention and emotion on face processing in the human brain: an event-related fMRI study. Neuron 30, 829–841.

Wandell, B.A., 1999. Computational neuroimaging of human visual cortex. Annu. Rev. Neurosci. 22, 145–173.

Wandell, B.A., Dumoulin, S.O., Brewer, A.A., 2007. Visual field maps in human cortex. Neuron 56, 366–383.

Watson, R., Huis in t Veld, E.M.J., de Gelder, B., 2016. The neural basis of individual face and object perception. Front. Hum. Neurosci. 10, 189.

Wegrzyn, M., Riehle, M., Labudda, K., Woermann, F., Baumgartner, F., Pollmann, S., Bien, C.G., Kissler, J., 2015. Investigating the brain basis of facial expression perception using multi-voxel pattern analysis. Cortex 69, 131–140.

Weiner, K.S., Grill-Spector, K., 2015. The evolution of face processing networks. Trends Cognit. Sci. 19, 240–241.

Williams, M.A., Baker, C.I., Op de Beeck, H.P., Mok Shim, W., Dang, S., Triantafyllou, C., Kanwisher, N., 2008. Feedback of visual object information to foveal retinotopic cortex. Nat. Neurosci. 11, 1439–1445.

Winston, J.S., Henson, R.N.A., Fine-Goulden, M.R., Dolan, R.J., 2004. fMRI-adaptation reveals dissociable neural representations of identity and expression in face perception. J. Neurophysiol. 92, 1830–1839.

Winston, J.S., O'Doherty, J., Dolan, R.J., 2003. Common and distinct neural responses during direct and incidental processing of multiple facial emotions. Neuroimage 20, 84–97.

Xiao, N.G., Perrotta, S., Quinn, P.C., Wang, Z., Sun, Y.-H.P., Lee, K., 2014. On the facilitative effects of face motion on face recognition and its development. Front. Psychol. 5, 1–16.

Xiao, N.G., Quinn, P.C., Ge, L., Lee, K., 2013. Elastic facial movement influences part-based but not holistic processing. J. Exp. Psychol. Hum. Percept. Perform. 39, 1457–1467.

Xu, X., Biederman, I., 2010. Loci of the release from fMRI adaptation for changes in facial expression, identity, and viewpoint. J. Vis. 10, 36–36.

Yarkoni, T., Barch, D.M., Gray, J.R., Conturo, T.E., Braver, T.S., 2009. BOLD correlates of trial-by-trial reaction time variability in gray and white matter: a multi-study fMRI analysis. PLoS One 4 e4257–15.

Zhang, H., Japee, S., Nolan, R., Chu, C., Liu, N., Ungerleider, L.G., 2016. Face-selective regions differ in their ability to classify facial expressions. Neuroimage 130, 77–90.