Rostalski Sophie-Marie (Orcid ID: 0000-0003-1661-195X)
Johnston Patrick (Orcid ID: 0000-0001-7703-1073)

Person identity-specific adaptation effects in the ventral occipito-temporal cortex

Sophie-Marie Rostalski[1], Jonathan Robinson[2,3], Géza Gergely Ambrus[1], Patrick Johnston[3*], Gyula Kovács[1*]

[1] Department of Biological Psychology and Cognitive Neurosciences, Institute of Psychology, Friedrich Schiller University, Jena, Germany,

[2] Department of Philosophy, Monash University, Melbourne, Australia

[3] School of Psychology & Counselling, Faculty of Health, Queensland University of Technology, Brisbane, Australia

* These two authors contributed to the publication equally as senior authors.

Corresponding Author: Sophie-Marie Rostalski

KEY WORDS

Identity recognition, image invariance, fMRI

ACKNOWLEDGEMENTS

Running title: Adaptation to identity in FFA and OFA   **Word count:** 7310

ABSTRACT

Identifying the faces of familiar persons requires the ability to assign several different images of a face to a common identity. Previous research showed that the occipito-temporal cortex, including the fusiform and the occipital face areas, is sensitive to personal identity. Still, the viewpoint, facial expression, and image-independence of this information are currently under heavy debate. Here we adapted a rapid serial visual stimulation paradigm (Johnston et al., 2016) and presented highly variable ambient-face images of famous persons to measure fMRI adaptation. FMRI adaptation is considered as the neuroimaging manifestation of repetition suppression, a neural phenomenon currently explained as a correlate of reduced predictive error responses for expected stimuli. We revisited the question of image-invariant identity-specific encoding mechanisms of the occipito-temporal cortex, using fMRI adaptation with a particular interest in predictive mechanisms. Participants were presented with trials containing eight different images of a famous person, images of eight different famous persons, or seven different images of a particular famous person followed by an identity change to violate potential expectation effects about person identity. We found an image-independent adaptation effect of identity for famous faces in the fusiform face area. However, in contrast to previous electrophysiological studies, using similar paradigms, no release of the adaptation effect was observed when identity-specific expectations were violated. Our results support recent multivariate pattern analysis studies, showing image-independent identity encoding in the core face-processing areas of the occipito-temporal cortex. These results are discussed in the frame of recent identity-processing models and predictive mechanisms.

INTRODUCTION

The efficient processing of human faces is an important aspect of social perception. In addition to detecting or recognizing faces in general, the identification of specific familiar faces across a variety of changes in low-level features, such as viewing angle, expression, illumination, or image contrast, is an important ability (Jenkins and Burton, 2011), giving us a great advantage in social situations. Although we are able to perceive a great deal of information even from unfamiliar faces (e.g., sex, age, emotional state, etc.), the identification of a familiar person requires a pre-existing internal representation of that particular person. Tasks involving face recognition or identification (e.g., face matching) that are easy for familiar faces can be very challenging for faces of persons we do not know (Bruce et al., 2001). Thus, it seems more and more likely that familiar and unfamiliar faces are processed differently (Kovács, 2020). Although forming a differential representation of an unfamiliar identity (ID) by performing perceptual tasks, such as sorting faces has been found to be stable enough to elicit a differential electrophysiological response in later event-related potential (ERP) components (Andrews et al., 2017), most studies found that this type of familiarization is not sufficient for creating a stable ID representation in the brain (Dubois et al., 1999; Natu and O'Toole, 2011). Hence, the exact process of forming robust identity representations is still unknown (Kovács, 2020). However, an image-invariant representation is essential for the identification of a person across a wide variety of possible situations and stimulus material.

In an influential model of face processing (Gobbini and Haxby, 2007; Haxby et al., 2000) areas in the fusiform gyrus represent the ID of a perceived face. Although we now know that a broader network of regions is involved in ID-specific information processing (Duchaine and Yovel, 2015), the fusiform gyrus is still considered to be part of the so-called core face network, supporting higher, more holistic, level of face representation compared to the feature-based representations supported by the inferior occipital gyrus (for a recent multivariate pattern analysis paper, supporting this conclusion see Tsantani et al., 2021). Therefore, the fusiform gyrus is a strong candidate structure for focusing experiments testing image-invariant face identification processes – that is to say, theoretically, the FFA is proposed to be responsible for identifying facial identity at a conceptual level, rather than simply categorizing familiar images as belonging to different ID categories. A well-established method for testing the extent of stimulus and stimulus-attribute sensitivity of an area is fMRI-

adaptation (fMR-a; Grill-Spector and Malach, 2001; Krekelberg et al., 2006). This technique has already been used extensively for twenty years to test for functional characteristics of cortical neurons in several studies (Grill-Spector et al., 1999). For example, signal adaptation in the lateral occipital cortex could be observed despite changes in position or size of objects (Malach et al., 1995). If a neuronal population keeps showing reduced responses to a specific object repeatedly presented but in different orientations, these neurons are considered viewpoint-invariant (Andresen et al., 2009). A release from this stimulus-repetition-related adaptation, on the other hand, would be regarded as evidence that this neuronal population encodes the viewpoint. By similar logic, across a set of highly variable ("ambient") face images, if what remains stable with respect to repetition across those set of images is person identity at a conceptual level, such modulation of the neuronal response would be considered as evidence that the neuronal population encodes person identity.

A reduced response to repeated face images in the fusiform face area (FFA; Kanwisher et al., 1997) has been shown by many previous studies, using similar techniques (Andrews and Ewbank, 2004; Loffler et al., 2005), improving our understanding of face representations. Still, it is not fully clarified under which circumstances an adaptation effect can be found in ventro-temporal face processing regions. For example, changes in viewpoint were found to result in a release from adaptation in the FFA for unfamiliar, but not for familiar faces (Ewbank and Andrews, 2008), suggesting their differential encoding within the area. This led to the conclusion that the representation of familiar faces is rather viewpoint-independent while that of unfamiliar faces is viewpoint-specific. In other words, different viewpoints of unfamiliar IDs may be perceived as different IDs, whereas familiar ID representations in the FFA are more stable.

The fact that this difference could be found in the FFA provided further evidence for a representation of facial ID in this region. Eger et al. (2005) used familiar and unfamiliar faces differing in rotation angle and expression and found a stronger image-independent reduction in activity due to ID repetition in anterior than in the middle fusiform gyrus, especially for familiar faces. Later, Xu et al., (2009) measured the BOLD signal in the FFA elicited by two artificially generated and, therefore, unfamiliar faces from either the same person or different identities and either with the same or with varying viewing angles. They demonstrated that both the change in ID and viewpoint elicited a larger BOLD signal in the FFA compared to

presenting identical images. Notably, the ID and viewpoint changing conditions led to similar BOLD signals, indicating that both led to a release of adaptation equally. This finding puts doubt on the role of FFA in the encoding of invariant aspects of face processing as shown in other fMR-a studies (Eger et al., 2004; Winston et al., 2004). For example, Davies-Thompson et al., (2009) tested whether occipito-temporal face-selective regions use an image-invariant neural code for familiar face representation in a block design. They contrasted blocks of repetitions of identical images of the same ID with blocks of different images of the same ID and blocks containing different images of different identities – for familiar and unfamiliar identities, separately. A reduced response to the same image - same ID condition, compared to different images of different identities, was found both for familiar and unfamiliar faces in the FFA. Surprisingly, showing different images from the same or different identities resulted in different responses neither for unfamiliar nor familiar faces. In a subsequent study, Davies-Thompson et al., (2013) replicated these results with a slightly different design: presenting blocks of eight different images of the same ID in contrast to blocks with eight images of different identities did not result in a significantly reduced response in the FFA nor occipital face area (OFA). Altogether, these neuroimaging results suggest the existence of a relatively low-level, image-dependent representation of ID within the core network areas. This conclusion, however, is at odds with more recent studies, which used machine learning techniques to perform multivariate pattern analyses (MVPA) on the FFA. Axelrod and Yovel (2015) were able to discriminate between the response patterns obtained for the different images of two highly familiar identities reliably. Recently, Tsantani et al. (2021) used short video clips and tested the available information in FFA and OFA. They found that both the OFA and the FFA contain ID-specific information and that the FFA reflects higher-level and more image-independent information than the OFA. Thus, so far, no consensus exists in the literature regarding the nature of ID representation of the core face-processing network areas.

Notwithstanding, there is clear evidence for the existence of an early-mid latency ID-specific adaptation effect from recent EEG and MEG studies (Simpson et al., 2015). Simpson and colleagues (2015) used an MEG adaptation design to show that faces but not objects showed clear adaptation effects localized to the FFA at around 170ms post-stimulus onset. This region also showed a release from adaptation to different identity faces at a latency of around 250-

300ms post-stimulus onset. This implies that the FFA is engaged, at different latencies, both in the holistic processing and individuation of face stimuli and, at a slightly later latency, of attaching these holistic face images to particular person identities.

Additional evidence comes from Johnston and colleagues (2016). They found a modulation of the face-sensitive N170 amplitude by introducing expectations about the appearance of a given ID, using several highly variable, ambient (Jenkins and Burton, 2011) face images of the same person in the EEG. A rarely occurring similar image of a different ID in a stream of various face images of the same person was perceived as a deviant. Accordingly, the release from adapting to one ID by encountering a different one manifested in a higher N170 amplitude. The authors related the underlying processes of their observation to predictive coding theories (Friston, 2005; Rao and Ballard, 1999). Those could explain this increase in activity since it is assumed that perception is based on expectations that rely on the prior statistical probability of events. These rely on experiences and are called "prior beliefs," which are continuously updated in the course of our everyday life. For example, a stream of images showing the same ID would therefore strengthen our expectations about the future appearance of another image of the same person. Thus, the sudden occurrence of a different ID is unexpected and manifests in a greater prediction error, which is measurable as an enhanced neural response in the prediction estimator areas. The key idea here is that when a particular identity is expected to occur, perceptual "templates" corresponding to that identity are preactivated to prioritize the rapid confirmation of that identity (Parr et al., 2020).

More recently, in a paper describing several EEG and one MEG experiment, Johnston and colleagues (Johnston et al., 2017) deployed a "contextual trajectory paradigm," wherein a series of trials consisting of 5 images depicted a specific contextual trajectory with the final stimulus transition either confirming that trajectory or violating the expectation. They tested trajectories for facial expressions (e.g., from neutral to happy), body rotation (e.g., turning from left to right), and locations of stimuli on the screen (e.g., clockwise or anticlockwise motion). For each experiment, a robust pattern was found, showing a modulatory effect of predictability of the last image in the N170. Irrespective of stimulus type, an enhanced N170 amplitude was found for unpredictable versus predictable stimulus transitions. These results show that a contextual modulation of early ERP components can be found after only four informative images (priors), offering a basis for expectations. In line with this idea, the effects

of expectation violations were found to be more pronounced both after five compared to three priors (Robinson et al., 2018) corresponding to the higher precision of the prior belief (Friston and Kiebel, 2009), and where the size of the perceptual distance between the violation event and the expected event was greater (Robinson et al., 2018).

FMR-a is considered as the neuroimaging manifestation of repetition suppression (RS), a phenomenon which is explained currently by many as the correlate of predictive error reduction of repeated or frequent, thereby expected stimuli (for a review, see Kovács and Schweinberger, 2016).

There is also evidence from a recent MEG study showing that prediction error-signals to violations of expected head orientation and facial ID could be spatially dissociated. Whereas prediction-error responses to stimulus orientation were localized to the dorsal visual processing stream, error signals to facial ID were localized to the right fusiform gyrus, among other locations (Robinson et al., 2020). Therefore, adopting the logic of the Johnston et al. (2016) ERP experiment to fMRI, we implemented an fMR-a design to measure the magnitude of release from adaptation in violated ID-specific expectation in key face processing regions. Additionally, this design allows us to revisit the issue of image-invariant ID-specific encoding mechanisms of the FFA by using fMR-a.

Specifically, this work aims at investigating ID-specific processing in the fusiform and occipital face areas, using ambient images of celebrities, very well known to our participants. Furthermore, by generating and violating expectations about person ID, we aimed at testing if prediction error responses are manifest in the neuroimaging signal.

MATERIALS AND METHODS

Participants

Thirty healthy participants took part in this experiment. They gave their informed consent for participation in accordance with the guidelines of the Declaration of Helsinki and with the approval of the ethics committee of the University of Jena. No participant had any history of neurological or psychiatric illness, and all had normal or corrected to normal vision. Three

participants were excluded from the analysis due to excessive head movements during the scanning session and another three were not analyzed further because they had too low performance on the behavioral task (2) or did not know the presented IDs (1). Thus, 24 right-handed participants (16 female; mean age 25,9 (± 3,6) years) were included in the current analysis. Please note, that certain regions of interest (ROI) could not be localized reliably in every participant. Therefore, the number of participants can slightly differ for the different areas.

Stimuli

Colorful images of 16 celebrities (8 males: Chris Hemsworth, Chris Pratt, David Beckham, Ewan McGregor, Gerard Butler, Jude Law, Matt Damon and Tom Hardy, 8 females: Cameron Diaz, Charlize Theron, Gwyneth Paltrow, Jennifer Lawrence, Kate Hudson, Kirsten Dunst, Reese Witherspoon and Scarlett Johansson) that were freely available on the internet were used for this experiment (for examples see Figure 1). The photographs vary in their physical properties (e.g., expression, head-position, eye-gaze, hairstyle, camera-angle, camera exposure, luminance). These types of images are also known from the literature as "ambient images" as they contain natural day-to-day variations under different conditions and can be compared to situations during daily life face recognition (Bortolon et al., 2018; Jenkins et al., 2011). Apart from the fact that we did not use duplicates or flipped image versions, the only other image selection criterion was that the viewing direction of the faces was at least roughly directed towards the camera. Eighteen such ambient images per ID were selected, aligned, and scaled to a resolution of 250*250 pixels (3.3° in radius). Thus, the stimulus set contained 288 different images of 16 different identities. By using ambient exemplar images, we ensured that any observed effect reflects higher-level ID processing, independently of the physical features of the images.

In order to functionally localize specific areas of interest (FFA and OFA), a sequence of blocks with images depicting faces, objects, and Fourier noise images were used. Each stimulus category consisted of 40 different stimuli. Faces were randomly occurring colored images of different famous and unknown persons. Identities in the localizer were different from those from the experimental task. Image blocks of objects encompassed a mixture of various items (e.g., food, clothing, etc.). All stimuli were scaled to a resolution of 600x600 pixels (8.5° in

radius). The Fourier noise images were created by an algorithm described in (Dakin et al., 2002).

Experimental design

The experiment was presented using MATLAB 2013a (The Mathworks) and Psychtoolbox v.3.0.14 (Brainard, 1997). A trial was composed of eight subsequently presented face images of same-sex identities (female and male face trials were presented 50%). Each image was presented for 500ms without any inter-stimulus-interval (ISI) and was slightly jittered spatially around the screen center to avoid low-level adaptation processes. Thus, a trial lasted 4 seconds, and it ended with a fixation cross. The inter-trial interval (ITI) was randomized to 2,4 or 6 seconds. Four different conditions of such stimulus sequences were created:

First, *alternation* sequences (ALT) consisted of the ambient images of eight different identities.

Second, in *adaptation* sequences (ADA), eight different images of the same ID were presented to test image-independent ID-specific adaptation effects.

Third, in the e*xpectation violation* condition (EV), a series of seven different images of the same ID was followed by the face of another ID. We reasoned that if ID-specific expectation modulates the observed adaptation effects, then the unexpected change of facial ID should lead to a release of adaptation.

Theoretically, after seeing the second image in the row, participants could expect the appearance of images of the same ID both in ADA and EV and the appearance of different identities in ALT. Therefore, we created a fourth condition which was similar to alternation sequences but ended with the repetition of one ID. In this condition, the first six images depicted six different identities, followed by two images of the same ID. These *Alternation with final repetition* (ALT-II) trials were not subject to any specific hypotheses but ensured us that the participants had equal number of trials where the first seven images were depicting the same ID (ADA, EV) or different identities (ALT, ALT-II). Examples for all four trial types are shown in Figure 1.

The trials of these four conditions appeared with the same probability randomly, with the only constraint being that a maximum of three subsequent trials could depict same sex identities and a maximum of two trials could come from the same condition.

Since attention is known to modulate response suppression and expectation violation (Larsson and Smith, 2012), participants had to perform a task, unrelated to the above-described manipulations. They had to respond to images with reduced size (1.98°), which could occur at any position within a trial sequence (detection rate of these target trials: 71.7 % plus minus 18.6%). This task was set to ensure that participants focused their attention on the stimuli without diverting their attention to the different conditions. To avoid potential effects of attention these target trials were removed from all further analyses. The main experimental procedure comprised 4 runs with one run including 80 trials (20 trials per condition), of which 10% were target-sequences.

We hypothesized that in areas encoding facial ID in an image-independent manner, ADA should lead to a reduced average BOLD signal, when compared to ALT sequences. Specifically, if an area is involved in image-independent facial ID processing, a lower BOLD response should be observable when different images of the same famous ID are repeated compared to when images of different famous identities are presented. In addition, previous studies suggested that the electrophysiological measures of face processing reflect the expectation of the occurrence of the same or different identities and the violation of these expectations (Johnston et al., 2016). We reasoned that if predictive mechanisms explain ID-specific signal reductions, then the violation of such expectations in the EV condition should manifest in a release of adaptation as well.

Procedure and imaging parameters

Participants were introduced to the MRI center, and a medical briefing was conducted. Next, they were asked to make familiarity judgments about the female and male identities used in the main experiment. For this, sample images of each ID (which were not used in the main experiment) were presented first alone and then together with their names and professions. Participants had to indicate whether they are familiar with them. Only if they reported to know the face and the name, the ID was evaluated as being familiar to the participant.

The scanning was conducted with a 3 Tesla MR Scanner (Siemens Prisma fit). All functional data was obtained using an Echo Planar Imaging (EPI) Sequence (35 slices; TR = 2000ms; TE = 30ms; flip angle = 90°; 64*64 matrices; in-plane resolution: 3x3 mm²; slice thickness: 3mm). A magnetization-prepared rapid gradient-echo sequence (MP-RAGE; TR=2300ms; TE=3,03, 1mm isotropic voxel size) was used to acquire high-resolution T1-weighted sagittal images to generate 3D structural scans. All images were acquired using a 20-channel head coil.

Behavioral data were recorded by a button box. There was only one button to signal the detection of target stimuli. First, within each scanning session, two experimental runs were administered, followed by the anatomical scan and another two experimental runs. Finally, the localizer scan completed the session of approximately one hour.

We implemented a localizer sequence to determine the relevant regions of interest. Here, blocks of images (presented for 250 ms) showing faces, objects, and Fourier noise were used. Each block was repeated five times, interleaved with blank periods of 12 seconds. Stimuli occurred randomly within one block.

Data processing and statistical analyses

Data and code required to reproduce all analyses will be available at OSF (https://osf.io/m3pwt/) at the time of publication. The conditions of our ethics approval do not permit the publishing of the raw MRI data. We will therefore provide extracted fMRI data from individual coordinates, which will be made available as well.

Neuroimaging data were preprocessed using SPM12 (Wellcome Department of Imaging Neuroscience, London, UK). In brief, the functional data were corrected for shifts in acquisition time of slices, realigned to correct for movement, co-registered to the anatomical images, normalized to the MNI-152 space, resampled to 2 mm isotropic voxel size, and finally spatially smoothed with an 8mm FWHM Gaussian kernel. A general linear model was specified, using the onsets of the trials of the four different conditions and the six movement parameters as regressors. For the experimental functional data, hemodynamic derivatives were added to the model.

To identify the location of FFA and OFA, we contrasted face blocks with blocks of objects and Fourier-randomized noise from the localizer sequence with a threshold of either $p < .05$

family-wise error (FWE) corrected (n = 18) or p < .0001 uncorrected (n = 6). The right FFA could be localized in all 24 participants (average MNI coordinates (±SE): 42 (0), -51 (1), -20 (1)) and in 23 participants in the left hemisphere (average MNI coordinates (±SE): -40 (1), -51 (1), -21 (1)). For every subject, we used the same contrast and threshold to identify OFA. The Right OFA could be localized in 23 participants (average MNI coordinates (±SE): 42 (1), -76 (1), -12 (1)). In the left hemisphere, OFA was also localized in 23 subjects (average MNI coordinates (±SE): -40 (1), -77 (1), -12 (1)). Individual coordinates can be found in the supplementary material Table 1.

The BOLD signals evoked by the stimulation, defined as the entire sequence of eight images, were extracted from each individual ROI, defined as the peak face responding voxel, using a 4 mm radius sphere (Brett, 2011). Hemodynamic response functions (HRF) were inspected to assure that the ROIs were identified reliably, and the extracted signal was evaluated. We observed that the HRFs peaked on the fourth timepoint consistently for all of our participants (supplementary material Figure 1). Therefore, peak HRF values were entered into the statistical models. We performed a two-way repeated-measures ANOVA with the within-subject factors of hemisphere (*right*, *left*) and condition (*adaptation*, *expectation violation*. *alternation*. *alternation-II*). Finally, we conducted an exploratory, second-level whole-brain analysis.

RESULTS

To assess whether participants were familiar with the presented identities, they filled out a questionnaire prior to the experiment. Mean familiarity ratings are 76,0% for male and 84,9% for female IDs and show that our participants were familiar with the stimuli.

FMRI Results

The mean BOLD signal in the FFA for all conditions is presented in Figure 2A. The repeated measures ANOVA revealed a main effect of condition ($F_{(3,54)}$ = 3.333, p = 0.026, η_p^2 = 0.156). Post-hoc tests showed a significant difference between *ALT* and *ADA* trials, t(18) = -2.915, p_{holm} = 0.031. This shows that the presentation of different, highly variable natural images of the same ID leads to response reduction in the FFA, suggesting that it plays a role in the

encoding of ID in an image-independent manner. All other comparisons remained nonsignificant. No interaction of hemisphere and condition was found ($F_{(3,54)}$ = 0.972, p = 0.413, η_p^2 = 0.051), suggesting similar effects over the left and right FFA.

The repeated measures ANOVA showed a significant main effect of condition in the bilateral OFA as well ($F_{(2.033,38.632)}$ = 4.194, p = 0.022, η_p^2 = 0.181; Greenhouse-Geisser corrected). Post-hoc tests revealed that the only significant difference is between ALT and EV t(19) = 3.313, p_{holm} = 0.010. Unlike in the FFA, the *ALT-ADA* comparison remained nonsignificant for the OFA t(19) = 2.358, p_{holm} = 0.109. We did not find an interaction of hemisphere and condition for the OFA either ($F_{(3,57)}$ = 1.887, p = 0.142, η_p^2 = 0.090).

To better evaluate the evidence for differences of BOLD responses to our conditions, we conducted a Bayesian repeated-measures ANOVA, including post hoc tests. We report Bayes factor BF_{10} for both analyses, reflecting how much more likely our data occur under alternative hypotheses than the null hypothesis.

In the FFA, the Bayes factor for a main effect of condition was 2.011, signaling that the effect of condition is more likely than the null hypothesis. More interestingly, post hoc tests revealed a Bayes factor for the *ADA* versus *ALT* comparison of 3.606, confirming our previous analysis and a Bayes factor of 1.438 for the difference between *ADA* and *ALT-II* trials. All other comparisons revealed Bayes factors < 1, suggesting that the evidence favors the null hypothesis.

The Bayesian repeated-measures ANOVA of OFA data revealed a Bayes factor of 4.603, favoring the alternative hypothesis over the null hypothesis by a factor of 4. Post hoc tests showed Bayes factors for the *EV* trial comparison with *ALT-II* of 23.586 and with *ALT* trials of 4.171. The Bayes factor for the comparison of *ADA* and *ALT* trials was 3.213, which is similar to the EV and ALT comparison. All other Bayes factors remained < 1.

Whole-brain Analyses

In order not to overlook any area that might show activation differences to the different conditions outside the pre-defined ROIs, we computed a whole-brain random design analysis contrasting adaptation > alternation on the group level. Applying a threshold of p <

$0.0001_{uncorrected}$ revealed only one cluster of activation (k = 5) in the right inferior frontal gyrus (MNI[x,y,z]: 62, 8, 24) (Figure 3), an area close to the inferior frontal face area (iFFA), which is part of the extended face-processing network and is known to play a role in the processing of eye-gaze and the semantic aspects of faces (Chan and Downing, 2011; Duchaine and Yovel, 2015; Ishai, 2008). No other contrast revealed significant clusters.

DISCUSSION

In the present study, we investigated ID-specific adaptation effects within the occipito-temporal face processing areas. We found an image-independent adaptation effect of identity for famous faces in the fusiform face area. This difference of presenting highly variable, ambient images of the same versus different identities was only significant for the FFA. In contrast, the OFA showed significantly lower activation for a condition where expectations are violated compared to alternating identities and a tendency for ID specific adaptation.

Previously, (Ewbank and Andrews, 2008) found fMR-a across different viewpoints in the FFA to familiar, but not to unfamiliar faces. However, their adaptation condition contained images, although varying in viewpoint, that still came from the same original images. Also, the implemented range of viewpoint change of this study was relatively small (12°). Therefore, these images were very similar in low-level features. Their interpretation of an ID-specific adaptation effect for familiar faces in the FFA is consequently only partly justified. Still, this conclusion is confirmed by our current study with the application of highly variable, ambient images (Jenkins and Burton, 2011).

Other studies in which blocks with different images from the same ID and blocks with different images of different IDs were contrasted, failed to find a difference in FFA responses to those conditions (Davies-Thompson et al., 2009). Thus, the current study shows that it is possible to discriminate familiar identities in FFA activity, providing evidence for a stable and image-independent ID representation (at least for familiar identities) in this area. One explanation for the discrepant results of the current and previous studies could come from small but significant differences in the applied designs. While we used colorful ambient images, (Davies-Thompson et al., 2009) used gray-scale frontal faces. Also, they presented ten images per

block and chose stimuli from a larger stimulus set. However, the latter differences should have made it easier to find an ID-specific adaptation effect in their study. In a follow-up experiment, they specifically investigated the responses to blocks of 8 different images showing either the same or different IDs (Davies-Thompson et al., 2013). Again, there was no significant fMR-a effect for ID present. In this study, responses to familiar and unfamiliar faces were compared and the same results were found for both stimulus types. Later, Weibert and colleagues (2016) replicated those results with a much larger sample. These authors found ID-specific adaptation only in the medial temporal lobe (MTL) and not in the FFA.

More recent studies, which applied multivariate pattern analyses (MVPA) to fMRI data have been able to discriminate between identities even when they were unknown (Anzellotti et al., 2014; Nestor et al., 2011). Although the decoding performance of these studies was generally moderate and sometimes barely exceeded chance level, our current results validate these multivariate findings on the univariate level and support the idea that the FFA discriminates between identities independently of images. Our results also fit those of recent MVPA studies of highly familiar faces which are more comparable to the stimulus material of the current study (Axelrod and Yovel, 2015). Both studies found identity-specific information in the bilateral FFA.

Although the adaptation versus alternation comparison did not reach significance in the OFA, we could show that the expectation violation was significantly different to the alternation condition. More interestingly the neural response to trials in which expectations were violated was almost the same as for the adaptation condition. This marginally significant effect of ID-specific adaptation hints towards an effect similar to that of the FFA but requires some more detailed analyses, such as previously mentioned multivariate ones. Note that the exclusive application of celebrity faces limits the interpretation of our findings to famous familiar faces. For the differences of famous and personally familiar faces regarding fMR-a see the review of Kovács (2020).

Overall, our study shows image-independent adaptation effects to faces in the fusiform gyrus, suggesting the image invariance of the representation within the area. However, as we did not include a same-image condition in our design (the repetition of the very same image, similar to Davies-Thompson et al., 2013), we cannot make a direct conclusion about the degree of this image invariance in the FFA. Previous studies, however, compared such same-

image and different-image (similar to our adaptation condition) blocks directly and found highly significant differences between them in the FFA as well as in the OFA for familiar faces (e.g. Davies-Thompson et al., 2013). On the basis of these results, we conclude that a fully image-invariant representation might not be present in the FFA yet.

Unlike the previous ERP studies which used similar paradigms (Johnston et al., 2016) we were unable to find effects of ID-specific expectation-suppression and expectation-violation related response enhancements in the occipito-temporal cortex. This may be due to the low temporal resolution of fMRI. Because of the limitation of the duration of the entire experiment, we opted for a paradigm in which the BOLD response to the entire trial is modeled. Since the images within a trial are not separated by sufficiently long ISIs, it was not possible to separate the response to the individual images (specially to the last one or two images) from the rest. However, predictive processes can still play a role in the presently described effects, specifically in the difference between adaptation and alternation sequences. We assume that presenting one famous ID activates a neural representation that facilitates processing of images of the same ID by top-down influences (Blom et al., 2020). This can also be considered as a trace of the accumulation of prior evidences, which, in turn, results in lower neuronal responses compared to alternation sequences in which no such predictions are possible.

The whole-brain analysis revealed a single cluster, being more active for ALT when compared to ADA in the inferior frontal gyrus, corresponding closely to the recently described area of iFFA. This area is supposed to be part of the face-processing network and is known to play a role in the processing of dynamic face properties as well as eye-gaze information. (Chan and Downing, 2011; Duchaine and Yovel, 2015; Ishai, 2008). As identity was kept constant in ADA, but changed continuously in ALT, our results raise the possibility that this area is also specifically involved in high-level predictions about identity continuity within an image sequence. The confirmation of this hypothesis, however, will require specific future studies.

CONCLUSION

Confirming results from multivariate pattern analyses, the present study shows image-independent ID-specific adaptation effects in the fusiform face area for famous familiar faces.

Especially in combination with the results of the occipital face area, our results suggest that the ID representations in occipito-temporal regions are not yet sufficiently clarified and that further research is needed. We could not replicate results from MEG studies, showing expectation violation effects related to facial identity in the fusiform gyrus.

AUTHOR CONTRIBUTIONS

PJ and GK devised and supervised the project, SMR, JR, GA, GK, and PJ designed the experiment, GA programmed the experiment. SMR carried out the experiment and performed all analyses. SMR and GK wrote the manuscript with support from JR, GA, and PJ.

FUNDING AND CONFLICTS OF INTEREST

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on https://osf.io/m3pwt/.

LIST OF ABBREVIATIONS

ADA – adaptation condition
ALT – alternation condition
ALT-II – alternation with final repetition condition
ANOVA – analysis of variance
BOLD – blood-oxygen-level dependent
EEG – electroencephalography
EPI - Echo Planar Imaging sequence
ERP – event-related potential
EV – expectation violation condition
FFA – fusiform face area
fMR-a – fMRI-adaptation
fMRI – functional magnetic resonance imaging
FWE – family-wise error
HRF – hemodynamic response function
ID – identity
iFFA – inferior frontal face area
ISI – inter-stimulus interval

ITI – inter-trial interval
MEG – magnetencephalography
MNI - Montreal Neurological Institute
MP-RAGE - magnetization-prepared rapid gradient-echo sequence
MTL – medio temporal lobe
MVPA – multivariate pattern analyses
OFA – occipital face area
ROI – regions of interest
RS – repetition suppression
SE – standard error

REFERENCES

Andresen, D.R., Vinberg, J., Grill-Spector, K., 2009. The representation of object viewpoint in human visual cortex. NeuroImage 45, 522–536. https://doi.org/10.1016/j.neuroimage.2008.11.009

Andrews, S., Burton, A.M., Schweinberger, S.R., Wiese, H., 2017. Event-related potentials reveal the development of stable face representations from natural variability. Quarterly Journal of Experimental Psychology 70, 1620–1632. https://doi.org/10.1080/17470218.2016.1195851

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. https://doi.org/10.1016/j.neuroimage.2004.07.060

Anzellotti, S., Fairhall, S.L., Caramazza, A., 2014. Decoding representations of face identity that are tolerant to rotation. Cerebral Cortex 24, 1988–1995. https://doi.org/10.1093/cercor/bht046

Axelrod, V., Yovel, G., 2015. Successful decoding of famous faces in the fusiform face area. PLoS ONE 10, 19–25. https://doi.org/10.1371/journal.pone.0117126

Blom, T., Feuerriegel, D., Johnson, P., Bode, S., Hogendoorn, H., 2020. Predictions drive neural representations of visual events ahead of incoming sensory information. Proceedings of the National Academy of Sciences of the United States of America 117. https://doi.org/10.1073/pnas.1917777117

Bortolon, C., Lorieux, S., Raffard, S., 2018. Self- or familiar-face recognition advantage? New insight using ambient images. Quarterly journal of experimental psychology (2006) 71, 1396–1404. https://doi.org/10.1080/17470218.2017.1327982

Brainard, D.H., 1997. The Psychophysics Toolbox. Spatial Vision 10, 433–436. https://doi.org/10.1163/156856897X00357

Brett, M., 2011. MarsBaR Documentation. Functional Imaging 1–45.

Bruce, V., Henderson, Z., Newman, C., Burton, A.M., 2001. Matching identities of familiar and unfamiliar faces caught on CCTV images. Journal of Experimental Psychology: Applied 7, 207–218. https://doi.org/10.1037/1076-898X.7.3.207

Chan, A.W.-Y., Downing, P.E., 2011. Faces and Eyes in Human Lateral Prefrontal Cortex. Frontiers in Human Neuroscience 5, 51. https://doi.org/10.3389/fnhum.2011.00051

Dakin, S.C., Hess, R.F., Ledgeway, T., Achtman, R.L., 2002. What causes non-monotonic tuning of fMRI response to noisy images? Current Biology 12, R476–R477. https://doi.org/10.1016/S0960-9822(02)00960-0

Davies-Thompson, J., Gouws, A., Andrews, T.J., 2009. An image-dependent representation of familiar and unfamiliar faces in the human ventral stream. Neuropsychologia 47, 1627–1635. https://doi.org/10.1016/j.neuropsychologia.2009.01.017

Davies-Thompson, J., Newling, K., Andrews, T.J., 2013. Image-invariant responses in face-selective regions do not explain the perceptual advantage for familiar face recognition. Cerebral Cortex 23, 370–377. https://doi.org/10.1093/cercor/bhs024

Dubois, S., Rossion, B., Schlitz, C., Bodart, J.M., Michel, C., Bruyer, R., Crommelinck, M., 1999. Effect of familiarity on the processing of human faces. NeuroImage 9, 278–289. https://doi.org/10.1006/nimg.1998.0409

Duchaine, B., Yovel, G., 2015. A Revised Neural Framework for Face Processing. Annual Review of Vision Science 1, 393–416. https://doi.org/10.1146/annurev-vision-082114-035518

Eger, E., Schweinberger, S.R., Dolan, R.J., Henson, R.N., 2005. Familiarity enhances invariance of face representations in human ventral visual cortex: fMRI evidence. NeuroImage 26, 1128–1139. https://doi.org/10.1016/j.neuroimage.2005.03.010

Eger, E., Schyns, P.G., Kleinschmidt, A., 2004. Scale invariant adaptation in fusiform face-responsive regions. NeuroImage 22, 232–242. https://doi.org/10.1016/j.neuroimage.2003.12.028

Ewbank, M.P., Andrews, T.J., 2008. Differential sensitivity for viewpoint between familiar and unfamiliar faces in human visual cortex. NeuroImage 40, 1857–1870. https://doi.org/10.1016/j.neuroimage.2008.01.049

Friston, K., 2005. A theory of cortical responses. Philosophical Transactions of the Royal Society B: Biological Sciences 360, 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K., Kiebel, S., 2009. Predictive coding under the free-energy principle. Philosophical Transactions of the Royal Society B: Biological Sciences 364, 1211–1221. https://doi.org/10.1098/rstb.2008.0300

Gobbini, M.I., Haxby, J. V., 2007. Neural systems for recognition of familiar faces. Neuropsychologia 45, 32–41. https://doi.org/10.1016/j.neuropsychologia.2006.04.015

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., Malach, R., 1999. Differential processing of objects under various viewing conditions in the human lateral occipital complex. Neuron. https://doi.org/10.1016/S0896-6273(00)80832-6

Grill-Spector, K., Malach, R., 2001. fMR-adaptation: A tool for studying the functional properties of human cortical neurons. Acta Psychologica 107, 293–321. https://doi.org/10.1016/S0001-6918(01)00019-1

Haxby, J. V., Hoffman, E.A., Gobbini, M.I., 2000. The distributed human neural system for face perception. Trends in Cognitive Sciences 4, 223–233. https://doi.org/10.1016/S1364-6613(00)01482-0

Ishai, A., 2008. Let's face it: It's a cortical network. NeuroImage. https://doi.org/10.1016/j.neuroimage.2007.10.040

Jenkins, R., Burton, M., 2011. Stable face representations. Philosophical Transactions of the Royal Society B: Biological Sciences 366, 1671–1683. https://doi.org/10.1098/rstb.2010.0379

Jenkins, R., White, D., Van Montfort, X., Mike Burton, A., 2011. Variability in photos of the same face. Cognition 121, 313–323. https://doi.org/10.1016/j.cognition.2011.08.001

Johnston, P., Overell, A., Kaufman, J., Robinson, J., Young, A.W., 2016. Expectations about person identity modulate the face-sensitive N170. Cortex 85, 54–64. https://doi.org/10.1016/j.cortex.2016.10.002

Johnston, P., Robinson, J., Kokkinakis, A., Ridgeway, S., Simpson, M., Johnson, S., Kaufman, J., Young, A.W., 2017. Temporal and spatial localization of prediction-error signals in the visual brain. Biological Psychology 125, 45–57. https://doi.org/10.1016/j.biopsycho.2017.02.004

Kanwisher, N., McDermott, J., Chun, M.M., 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. The Journal of neuroscience : the official journal of the Society for Neuroscience 17, 4302–11. https://doi.org/10.1098/Rstb.2006.1934

Kovács, G., 2020. Getting to know someone: Familiarity, person recognition, and identification in the human brain. Journal of Cognitive Neuroscience 32, 2205–2225. https://doi.org/10.1162/jocn_a_01627

Kovács, G., Schweinberger, S.R., 2016. Repetition suppression – An integrative view. Cortex 80, 1–4. https://doi.org/10.1016/j.cortex.2016.04.022

Krekelberg, B., Boynton, G.M., van Wezel, R.J.A.A., 2006. Adaptation: from single cells to BOLD signals. Trends in Neurosciences 29, 250–256. https://doi.org/10.1016/j.tins.2006.02.008

Larsson, J., Smith, A.T., 2012. fMRI Repetition Suppression : Neuronal Adaptation or Stimulus Expectation ? Cerebral Cortex 22, 567–576. https://doi.org/10.1093/cercor/bhr119

Loffler, G., Yourganov, G., Wilkinson, F., Wilson, H.R., 2005. fMRI evidence for the neural representation of faces. Nature Neuroscience 8, 1386–1390. https://doi.org/10.1038/nn1538

Malach, R., Reppas, J.B., Benson, R.R., Kwong, K.K., Jiang, H., Kennedy, W.A., Ledden, P.J., Brady, T.J., Rosen, B.R., Tootell, R.B., 1995. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. Proceedings of the National Academy of Sciences 92, 8135–8139. https://doi.org/10.1073/pnas.92.18.8135

Natu, V., O'Toole, A.J., 2011. The neural processing of familiar and unfamiliar faces: A review and synopsis. British Journal of Psychology 102, 726–747. https://doi.org/10.1111/j.2044-8295.2011.02053.x

Nestor, A., Plaut, D.C., Behrmann, M., 2011. Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. Proceedings of the National Academy of Sciences 108, 9998–10003. https://doi.org/10.1073/pnas.1102433108

Parr, T., Rikhye, R.V., Halassa, M.M., Friston, K.J., 2020. Prefrontal Computation as Active Inference. Cerebral Cortex 30, 682–695. https://doi.org/10.1093/cercor/bhz118

Rao, R.P.N., Ballard, D.H., 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nature neuroscience 2, 79–87. https://doi.org/10.1038/4580

Robinson, J.E., Breakspear, M., Young, A.W., Johnston, P.J., 2018. Dose-dependent modulation of the visually evoked N1/N170 by perceptual surprise: A clear demonstration of prediction-error signalling. European Journal of Neuroscience 1–11. https://doi.org/10.1111/ejn.13920

Robinson, J.E., Woods, W., Leung, S., Kaufman, J., Breakspear, M., Young, A.W., Johnston, P.J., 2020. Prediction-error signals to violated expectations about person identity and head orientation are doubly-dissociated across dorsal and ventral visual stream regions. NeuroImage 206, 116325. https://doi.org/10.1016/j.neuroimage.2019.116325

Simpson, M.I.G., Johnson, S.R., Prendergast, G., Kokkinakis, A. V., Johnson, E., Green, G.G.R., Johnston, P.J., 2015. MEG adaptation resolves the spatiotemporal characteristics of face-sensitive brain responses. Journal of Neuroscience 35, 15088–15096. https://doi.org/10.1523/JNEUROSCI.2090-15.2015

Tsantani, M., Kriegeskorte, N., Storrs, K., Williams, A.L., McGettigan, C., Garrido, L., 2021. FFA and OFA encode distinct types of face identity information. The Journal of Neuroscience 41, JN-RM-1449-20. https://doi.org/10.1523/jneurosci.1449-20.2020

Weibert, K., Harris, R.J., Mitchell, A., Byrne, H., Young, A.W., Andrews, T.J., 2016. An image-invariant neural response to familiar faces in the human medial temporal lobe. Cortex 84, 34–42. https://doi.org/10.1016/j.cortex.2016.08.014

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. Journal of Neurophysiology 92, 1830–1839. https://doi.org/10.1152/jn.00155.2004

Xu, X., Yue, X., Lescroart, M.D., Biederman, I., Kim, J.G., 2009. Adaptation in the fusiform face area (FFA): Image or person? Vision Research 49, 2800–2807. https://doi.org/10.1016/j.visres.2009.08.021
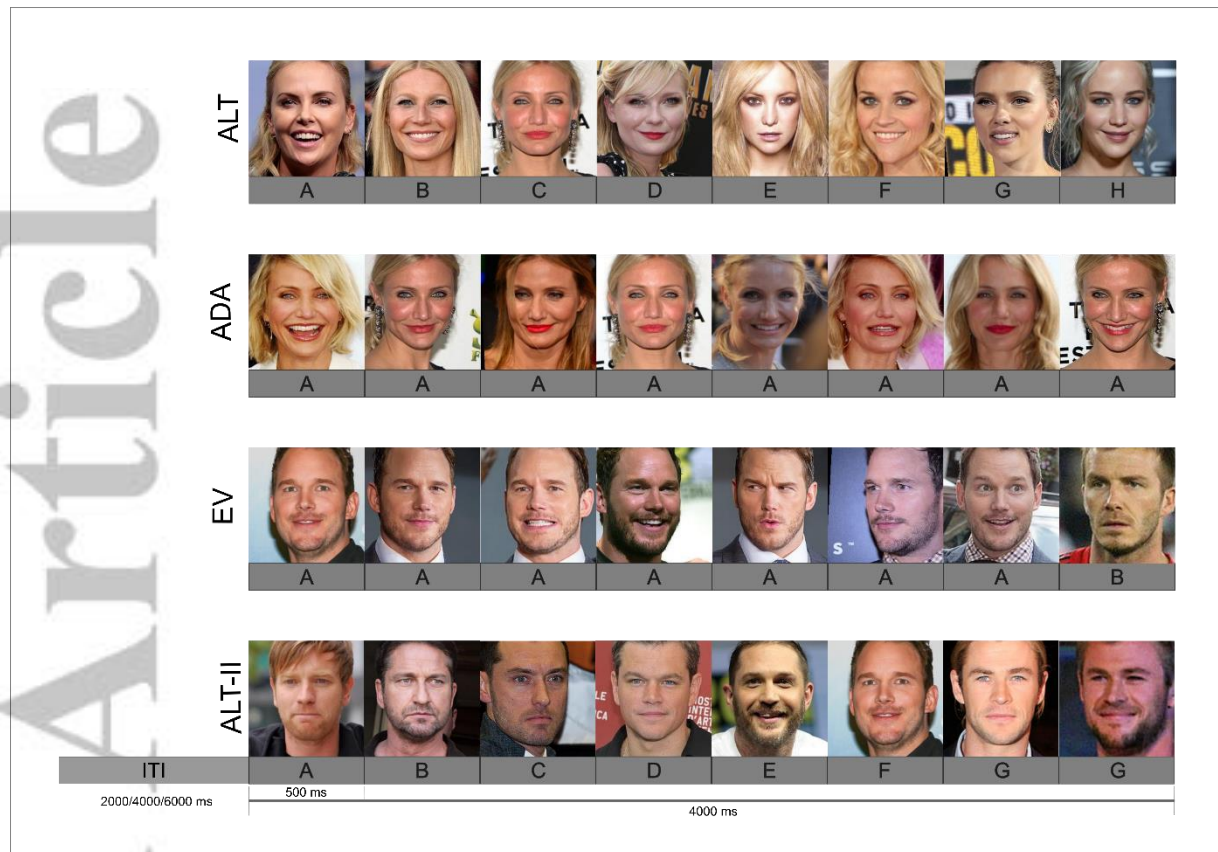
Figure 1. Conditions and trial structure in the four applied conditions. Each condition was composed of 8 faces. ALT: different same-sex identities. ADA: 8 different images of the same ID. EV: 7 different images of the same ID, followed by another same-sex ID. ALT-II: 6 images of different same-sex identities, followed by two different images of another, same-sex ID. Please note that these images are just examples and might not have been part of the actual stimulus set, as we do not have permission for publishing all exemplars we used. All images shown here are subject to either Creative Commons Attribution-Share Alike license or have no copyrights (public domain).
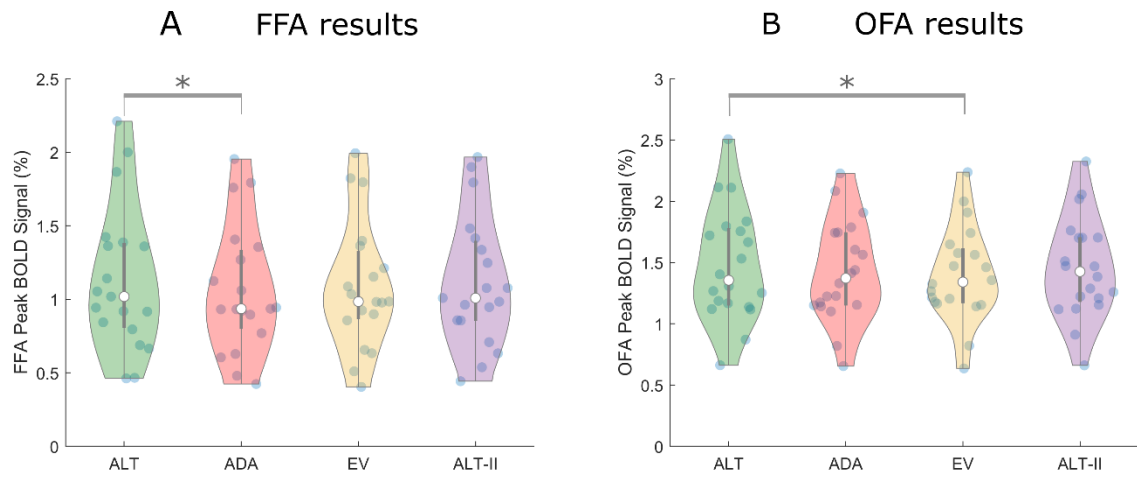
*Figure 2. Peak BOLD signals from individual participants to the different conditions for the bilateral FFA (left panel) and OFA (right panel). * represents p<0.05.*
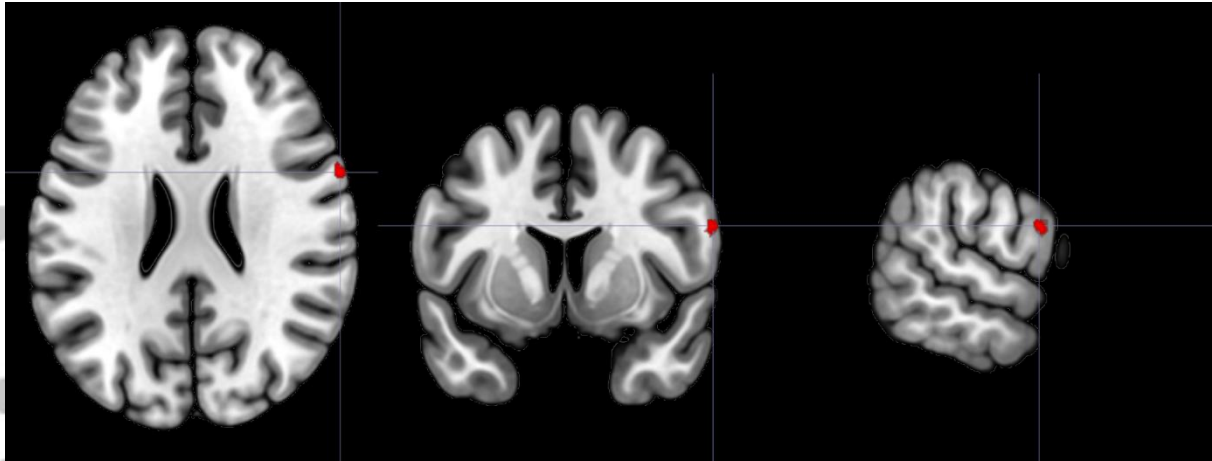
Figure 3. Results of the whole-brain analyses. Significant (p < 0.0001$_{uncorrected}$) cluster for the contrast ALT > ADA is marked by red.

Graphical Abstract Text to: Person identity-specific adaptation effects in the ventral occipito-temporal cortex

Sophie-Marie Rostalski[1], Jonathan Robinson[2,3], Géza Gergely Ambrus[1], Patrick Johnston[3*], Gyula Kovács[1*]

[1] Department of Biological Psychology and Cognitive Neurosciences, Institute of Psychology, Friedrich Schiller University, Jena, Germany,

[2] Department of Philosophy, Monash University, Melbourne, Australia

[3] School of Psychology & Counselling, Faculty of Health, Queensland University of Technology, Brisbane, Australia

* These two authors contributed to the publication equally as senior authors.

To investigate image-invariant identity-specific encoding in the human brain, participants were presented with ambient face images of famous identities in a rapid serial visual stimulation paradigm, with the potential of an identity change to induce expectation violation. We observed an image-independent adaptation effect in the fusiform face area, but no release of adaptation when expectations were violated. Our results support recent findings, demonstrating image-independent identity encoding in the core face-processing areas of the occipito-temporal cortex.