Spatio-temporal dynamics of face perception

Muukkonen, I.*[1], Ölander, K.[1], Numminen, J.[2], & Salmela, V.R.[1,3]

[1]Department of Psychology and Logopedics, University of Helsinki, Finland

[2]Helsinki Medical Imaging Centre at Töölö Hospital, Helsinki University Hospital, Finland

[3]Ami centre, Aalto Neuroimaging, Aalto University, Finland

*Corresponding author: ilkka.muukkonen@helsinki.fi

Abstract

The temporal and spatial neural processing of faces have been studied rigorously, but few studies have unified these dimensions to reveal the spatio-temporal dynamics postulated by the models of face processing. We used support vector machine decoding and representational similarity analysis to combine information from different locations (fMRI), timepoints (EEG), and theoretical models. By correlating information matrices derived from pair-wise decodings of neural responses to different facial expressions (neutral, happy, fearful, angry), we found early EEG timepoints (110-150 ms) to match fMRI data from early visual cortex (EVC), and later timepoints (170 – 250 ms) to match data from occipital and fusiform face areas (OFA/FFA) and posterior superior temporal sulcus (pSTS). The earliest correlations were driven by information from happy faces, and the later by more accurate decoding of fearful and angry faces. Model comparisons revealed systematic changes along the processing hierarchy, from emotional distance and visual feature coding in EVC to coding of intensity of expressions in right pSTS. The results highlight the importance of multimodal approach for understanding functional roles of different brain regions.

Keywords: *EEG, fMRI, RSA, face expression, decoding, face perception*

1 Introduction

Faces contain information on different socially important categories, such as identity and the emotional state of an individual. Behaviorally, we excel at dividing these different sources of information, and can easily identify a person just from their face despite changes in expressions or viewpoints. Indeed, the processing of faces has been suggested to be 'special' (Richler & Gauthier, 2014), and to be processed more holistically than other visual stimuli (Shen & Palmeri, 2015; Tanaka & Simonyi, 2016; but see also Gold, Mundy, & Tjan, 2012). Newborns recognize their mother's face (Bushnell, 2001; Bushnell, Sai, & Mullin, 1989) despite their still crude visual skills (Dobson & Teller, 1978), prenatal children focus more on face-shaped than non-face-shaped objects (Reid et al., 2017) and inversion distorts face recognition more than recognition of other objects (Taubert, Apthorp, Aagten-Murphy, & Alais, 2011; Valentine, 1988; Yin, 1969). Thus, faces provide excellent case for studying how our brain processes nuanced multidimensional information, in which the change of one feature (e.g. expression) must be separated from the changes of another feature (e.g. identity).

Several models explaining the neural processing of faces have been suggested (Duchaine & Yovel, 2015; Haxby, Hoffman, & Gobbini, 2000). One main feature in the models is the separation of the processing of changeable and invariant aspects. Changeable aspects, (or motion: Bernstein & Yovel, 2015) such as expressions, are mainly processed in a dorsal stream, especially in superior temporal sulcus (STS; Greening, Mitchell, & Smith, 2018; Said, Haxby, & Todorov, 2011; Zhang et al., 2016), while invariant aspects, such as identity, are processed in a ventral stream from part-based processing in occipital face area (OFA; Atkinson & Adolphs, 2011; Henriksson, Mur, & Kriegeskorte, 2015; Pitcher, Walsh, & Duchaine, 2011), to fusiform face area (FFA; Anzellotti, Fairhall, & Caramazza, 2014; Carlin & Kriegeskorte, 2017; Dobs, Schultz, Bülthoff, & Gardner, 2018; Kanwisher & Yovel, 2006), and finally to highest-level, viewpoint invariant processing in ventral anterior temporal lobe (vATL; Anzellotti & Caramazza, 2016; Anzellotti et al., 2014; Collins & Olson, 2015). Temporally, the processing starts around 90 ms (Dima, Perry, Messaritaki, Zhang, & Singh, 2018; Sugase, Yamane, Ueno, & Kawano, 1999), perhaps already somewhat sensitive to differences in expression (Dima et al., 2018; Müller-Bardorff et al., 2018). Most robust evidence for separating faces from other objects is found in the N170-component (at ~170 ms; for a review, see Rossion & Jacques, 2011), also sensitive to some expressions (Hinojosa, Mercado, & Carretié, 2015). Coding of identities is usually found later. Although earliest results are found from ~100 ms onwards (Ambrus, Kaiser, Cichy, & Kovács, 2019; Vida, Nestor, Plaut, & Behrmann, 2017), higher-level identity representations surpassing low-level visual information (Vida et al., 2017), as well as sensitivity to familiar faces (Schweinberger & Neumann, 2016), are found later, at around 250 ms, and differentiation of same sex identities only

after 400 ms (Ambrus et al., 2019).

Most studies have, however, investigated only one facial feature (e.g. expressions or identities) with one method (M/EEG or fMRI) at a time. The recent increase in the use of multivariate methods, such as representational similarity analysis (RSA; Kriegeskorte, Mur, & Bandettini, 2008), in cognitive neuroscience has provided new ways to study more nuanced information structures in different parts and at different timepoints in the brain. Following this, studies have compared how different visual object categories (e.g. animate vs. inanimate, faces vs. bodies) are coded (Cichy, Pantazis, & Oliva, 2014, 2016), which brain areas contain information on the identity of a face stimuli (Guntupalli, Wheeler, & Gobbini, 2017), and which models best describe the information structure in the brain (Kriegeskorte et al., 2008; Vida et al., 2017). Promising results for unraveling how the brain codes facial identities have also accumulated in single-cell studies on primates (Chang & Tsao, 2017).

Representational similarity analysis makes it also possible to compare data from different methods. As the data is analyzed on information-level rather than activity-level, the information in for example a certain timepoint in EEG can be compared to the information in different spatial areas from fMRI-data, and thus combine the good temporal accuracy of M/EEG with the good spatial accuracy of fMRI. This approach has been used in studying audiovisual attention networks (Salmela, Salo, Salmi, & Alho, 2018), visual object categorization (Cichy & Pantazis, 2017; Cichy et al., 2014, 2016), differences between neural responses to tasks and images (Hebart, Bankson, Harel, Baker, & Cichy, 2018), and processing of facial shape and surface features (Nemrodov, Behrmann, Niemeier, Drobotenko, & Nestor, 2019). These studies showed that, within the first few hundred milliseconds, processing of visual object categories spreads from early visual areas (V1) to lateral occipital complex (LOC; Cichy et al., 2016) and reconstruction accuracies of neutral and happy faces' shapes correlate between EEG and fMRI (Nemrodov et al., 2019). Furthermore, a simultaneous EEG-fMRI study found face selectivity (response magnitude between faces and images of chairs) in fMRI to correlate with earlier EEG timepoints in OFA than in FFA and STS (Sadeh, Podlipsky, Zhdanov, & Yovel, 2010). In the current study, we combined EEG and fMRI data of morphed faces that varied dimensionally both in their identity and expression. We compared the representational dissimilarity matrixes (RDMs) – based on pair-wise decoding of all stimuli – derived from all time points in the EEG data and from searchlight voxels in the fMRI data to reveal the spatio-temporal dynamics of face processing.

2 Methods

2.1 Participants

18 volunteers were recruited to participate in the study. One participant dropped out before completing the first part. Thus, 17 people (7 males, mean age = 24, SD = 3.4) completed the study. All participants were right-handed, had normal or corrected to normal vision and had no (self-reported) diagnosed difficulties in recognizing faces (e.g. prosopagnosia). They were recruited through the students' mailing list of the Faculty of Behavioral Sciences at the University of Helsinki, and received monetary compensation for the participation.

2.2 Ethics

The present study was conducted in accordance with the declaration of Helsinki and was reviewed and approved by Ethics Review Board in the Humanities and Social and Behavioural Sciences of the University of Helsinki. All participants gave informed consent and were screened to be suitable for fMRI-scanning with the standard procedure of the AMI-centre of Aalto University.

2.3 Stimuli

The stimuli consisted of 112 full-colored images of faces. The root-mean-square contrast was 0.2, and width and height were approximately 8° and 11°, respectively. The images contained four identities (2 females, 2 males) from the Faces database (Ebner, Riediger, & Lindenberger, 2010), and four different expressions from each identity (neutral, angry, happy and fearful; see https://faces.mpdl.mpg.de/imeji/ for example images). These 16 (4*4) original images were first preprocessed with Corel Paintshop Pro X7. The images were 1) straightened horizontally, using the centers of the eyes as a reference line, 2) resized to have same interocular distance, and 3) centered in a way that the centers of the eyes were approximately in the same spot. This was done to control for irrelevant visual variability in the pictures, for example size or tilt of the face, and to improve face morphing.

After the preprocessing of the original faces, they were morphed with FantaMorph (version 5.4.6). Each original identity having a certain expression were morphed to every other identity with the same expression. From these morph dimensions, two morphed images, 1/3 of identity X and 2/3 of identity Y, and vice versa, were used in the study. As result, there were 12 mock identities in addition to the original four. These 16 identities were then morphed within an identity from neutral to the three other expressions, and the 50 % morphs were used in the study (Figure 4). Thus, the total stimuli contained 112 images in a 16 identities (4 original, 12 morphed) x 7 emotions (neutral, 3 original-100%, 3

morphed-50%) design.

2.4 Study design

The study contained two parts, fMRI-measurement and EEG-measurement, which were conducted on different days. Nine of the participants completed first the fMRI part, and eight completed first the EEG part. The time between the measurements were 1-40 days on different participants. Both sessions lasted approximately 1.5 hours, including preparation.

The primary task and stimuli were identical in both measurements. Participants saw a face on a screen (500 ms), and they were instructed to answer, by pressing one of two buttons, whether the face on the screen was a female or a male. The gender identification task was used in order to ensure participants' attentional focus on all facial features. The inter stimulus interval was 2500ms. The study was divided in three runs. Each stimuli was shown twice in each run, once (in random order) before a 30 sec rest period and second time (in random order) after the rest period. In total, each run contained 224 trials and the duration was approximately 12 minutes. Between the runs the participants were contacted through microphone and were given a rest period if necessary.

The fMRI-measurement began with an anatomical scan (6 min), and ended with a functional face area localizer run. In the localizer scan, intact faces, phase scrambled faces, and checkerboard stimuli were shown while the participant were instructed to fixate at the center of the screen. In the EEG-measurement, video clips of facial expressions were shown in the end of the experiment. The results of these are not, however, discussed here.

2.5 Acquisition and preprocessing of EEG

EEG data were recorded using a 64-channel cap and 6 additional electrodes (mastoids, HEOGs from both eyes and two VEOGs from left eye) with Biosemi Actiview. EEG data were sampled at 1024 Hz and referenced online to Common Mode Sense (CMS) electrode at PO1.

EEG preprocessing were done with EEGLAB v13.6.5b running in Matlab R2015a. The data were first downsampled to 200Hz, and bandpass-filtered between 1-51Hz (- 6db cutoff points). Line noise were removed with Cleanline-plugin (Mullen, 2012) and bad channels detected and removed with clean_rawdata v0.31-plugin (Kothe, 2013). The data were then re-referenced to the average of the 64 cap electrodes, the removed channels were interpolated and all the non-cap electrodes were removed. Epochs of -300 – 1000 ms from stimuli onset were created, and epochs with the same stimuli were averaged. Thus, a four dimensional EEG-data-matrix were created, the dimensions being the 17 participants, 112

different stimuli type, 64 channels and 260 timepoints (5ms each).

2.6 Acquisition and preprocessing of fMRI

FMRI were recorded using 3T MAGNETOM Skyra scanner (Siemens Healthcare, Erlangen, Germany) at the AMI-Centre of Aalto University School of Science. A 30-channel head coil was used. The functional echo planar (EPI) images were acquired with an imaging area consisting of 43 contiguous slices (TR 2.4 s, TE 32 ms, voxel matrix 64 · 64, field of view 20 cm, slice thickness 3.0 mm). Three functional runs of 275 volumes (including 4 initial dummy volumes) were measured, each lasting approximately 12 minutes. In preprocessing, slice timing and motion correction (but no smoothing) were applied, and all images were co-registered to T1 anatomical image (MPRAGE). Finally, images from each subject were normalized to a standardized head space. The fMRI data was modelled with event-related GLM analysis containing separate regressors for each stimulus (112 in total), and 6 nuisance regressors for motion. As a result, there were 112 (stimuli) x 3 (runs) whole-brain beta images for each subject. fMRI analyses were conducted with SPM12 toolbox for Matlab.

2.7 Decoding

We used leave-one out support vector machine (SVM; Cortes & Vapnik, 1995) decoding with the non-decoded stimulus dimension (expression or identity) as examples of the training set. Each expression/identity was decoded versus each other expression/identity. In EEG, this resulted in 21 decodings, 16-fold each, for the expressions, and 120 decodings, 7-fold each, for the identities. In fMRI, we divided the data between runs, using two runs as the training and the remaining run as the test set, leaving 1/3 of the data as the test set for each fold. All decodings were performed separately within each subject.

FMRI-data were decoded with The Decoding Toolbox (TDT; Hebart, Görgen, & Haynes, 2015), using beta-images in normalized head space. We used searchlight-based (Kriegeskorte, Goebel, & Bandettini, 2006) decoding with a radius of 10 mm, and with default settings of TDT, with L2-norm SVM running in LIBSVM (Chang & Lin, 2011). For EEG-data, we used DDTBOX (Bode, Feuerriegel, Bennett, & Alday, 2018). The EEG data were averaged over trials beforehand to lower the level of noise. We used spatiotemporal decoding with 2 consecutive timepoints, resulting in non-overlapping windows of 10 ms, and all 64 channels as features. Otherwise, default settings of DDTBOX were used, with C-SVM with a regularizing parameter C = 1 running in LIBSVM. All decoding analyses were run in Matlab 2018b.

2.8 Statistical analyses

To calculate a measure for the total amount of information of expressions and of identities in the brain, we took the within subject averages of all the expression decodings, and of all the identity decodings, in each timepoint in EEG, and in each voxel in fMRI. One sample t-tests over subjects were used to test whether, and which, timepoints and voxels contained information on which stimuli dimension. In fMRI, statistical significance was defined with cluster-based test using SPM12, with voxel-wise criterion of $t > 3.0$, clusterwise criterion of $p(FDR) < .05$, and minimum cluster size $k > 50$. In EEG, FDR-correction over timepoints were used, with criteria of $p < .05$. To compare different expressions, same analyses were conducted using only the decodings of 100% expression (happy, fearful and angry) versus neutral faces.

Representational similarity analysis was used to compare the information representations of expressions in fMRI, EEG and models (See *3.4. Models for expression processing*). *EEG-fMRI-analysis:* A representational dissimilarity matrix (RDM) was calculated for each voxel in fMRI, averaged over subjects, and for each timepoint in EEG, separately for each subject. Each cell *ij* in the RDM was the decoding accuracy between the two stimuli types of row *i* and column *j*. These RDMs were then correlated using Spearman correlation between each fMRI voxel and each EEG timepoint, and correlations from each subject (EEG) were then averaged. To calculate criterion for statistical significance, permutations were used as follows. 1) The highest voxel/timepoint correlation from each subject were selected, 2) the EEG RDM from that point were randomly permuted and correlated with the (real) fMRI RDM, 3) permuted correlations were averaged over subjects. These steps were iterated 5040 times (all possible permutations for 7x7 RDM), and the permuted correlation at the cutpoint for highest 1% in the iterations were used to define $p<.01$ ($r = .149$). *Model-fMRI comparisons:* RDMs averaged over subjects from each voxel were compared to the models, and significance were defined by permutation analysis from the voxel with highest correlation to the model in question. *Model-EEG, Model-ROI and ROI-EEG comparisons*: significance were defined by t-tests over subjects.

2.9 Data and code availability

Data and code are available upon request.

3 Results

3.1 Overall information related to facial expressions and identities

To investigate the processing of facial expression and identities, we showed participants a total of 112 face images in which both emotional expression (neutral, 50% and 100% happy, fearful and angry) and facial identity (4 original, 12

morphed (33/67%) mock identities) varied dimensionally. In order to reveal information related to expressions and identities, we used support vector machine (SVM) decoding to produce pairwise classification accuracies between all stimuli. This was done separately for each voxel in fMRI with searchlight analysis (Kriegeskorte et al., 2006), and for each timepoint in EEG.

To calculate a measure of overall expression processing, we took the mean of all pairwise expression decoding accuracies (henceforth called "total information"). In fMRI, this resulted in a broad area containing the early visual cortex (EVC), occipital face area (OFA), right fusiform face area (FFA) and right posterior superior temporal sulcus (pSTS; Figure 1A). This was in concordance with earlier studies, which have shown facial expression processing to be right-lateralized, and occur in the STS, FFA, OFA and EVC (for a review, see Duchaine & Yovel, 2015). The EEG results revealed coding of expressions to start at ~90 ms after stimulus onset, and continue until ~750 ms (Figure 1B) with highest peak at 140 ms (mean decoding accuracy 56%, p(FDR)<.001). Previous studies using decoding in MEG, (Dima et al., 2018) and single-cell recordings in monkeys (Sugase et al., 1999) have also found face expression coding to start at ~100 ms and peaking at ~150 ms.

For face identities, we performed similar analyses. Our preliminary analyses suggested that the motor responses of the task of gender identification were obscuring the identity results. To account for this, we performed analyses using only the within-gender identity pairs in which participants answered correctly (using the same finger) over 90% of the trials. Only 17/120 comparison pairs fulfilled the criteria. The results showed significant clusters in the left temporo-parietal junction and anterior STS, but not in face identity specific regions in ventral temporal or occipital cortexes (Figure 1A). As the results still appeared to be contaminated by the motoric responses (correlation of time trajectories for identities and responses in EEG were .70; Figure 1B), and no expected face patches appeared significant, no further analyses for facial identities were performed.

3.2 Different temporal and spatial responses to different facial expressions

To compare different expressions, we examined the decodings of neutral expression compared separately to (100%) happy, fearful, and angry faces. Happy faces were best decoded from early visual cortex in fMRI (Figure 1C), with highest peak in EEG at 130 ms (60.6%, p(FDR)=.007; Figure 1D), while angry faces peaked later, at 200 ms (61.7%, p(FDR)= .012; Figure 1D), and were best decoded in more anterior areas of pSTS and FFA (Figure 1C). Fearful faces fell in between in EEG with highest peak at 180 ms (62.9%, p(FDR)=.006), while no significant decoding in fMRI was found. To formally test temporal differences in the three expressions, we compared the latencies of the highest decoding

accuracies of each expression from each subject between 50 – 250 ms (Wilcoxon's signed rank test). Decoding of happy faces peaked significantly earlier than both fearful (p=.023) and angry (p=.001) faces, while no statistically significant difference was found between fearful and angry expressions (p=.09). Although reaching maximal decoding later than happy or fearful faces, a short period of significant coding of angry faces was also found at 80 – 90 ms (53.8%, p(FDR)=.047).

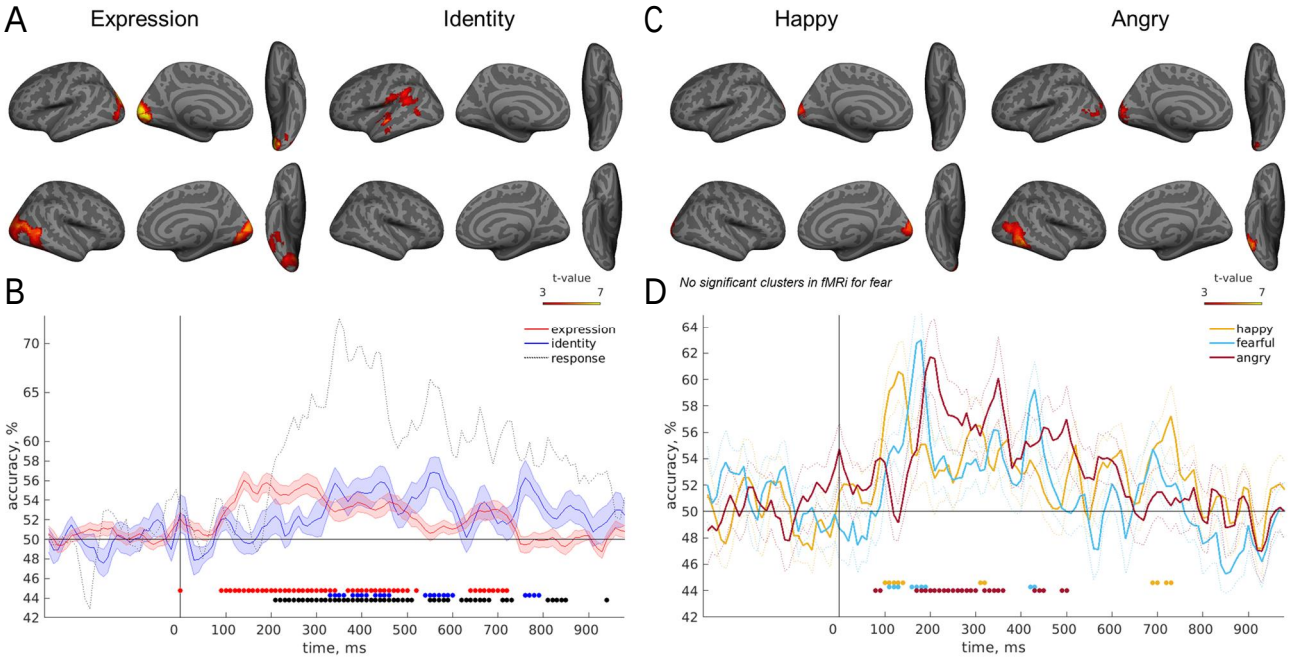

Figure 1. *Means of pair-wise decoding accuracies of expressions and identities.* ***A)*** *fMRI shows wide cluster for expressions in bilateral occipital cortex and right temporal lobe, including OFA, FFA and pSTS, and clusters for identities in left temporo-parietal junction and anterior STS.* ***B)*** *Total information from EEG. Decoding of responses is shown as they correlated and thus obscured the identity coding results.* ***C)*** *fMRI results show significant coding of happy (vs. neutral) expressions in EVC, and angry faces in EVC, right pSTS and right FFA. No significant clusters for fear were found.* ***D)*** *In EEG, the decoding accuracy for different expressions peaked between 130 – 200 ms after stimulus onset. fMRI thresholded at t>3.0 (voxel-level), p(FDR)<.05 (cluster-level), and cluster size k>50. Dots in EEG show significant timepoints (p(FDR)<.05), and shaded areas show SEMs.*

3.3. Combined EEG-fMRI reveals spatiotemporal pattern of face processing

We used representational similarity analysis (Kriegeskorte et al., 2008) to compare the information structures related to the processing of facial expressions in fMRI and EEG. We correlated representational dissimilarity matrixes (RDMs) from each voxel in fMRI and each timepoint in EEG. Each cell of the RDMs was the decoding accuracy between two

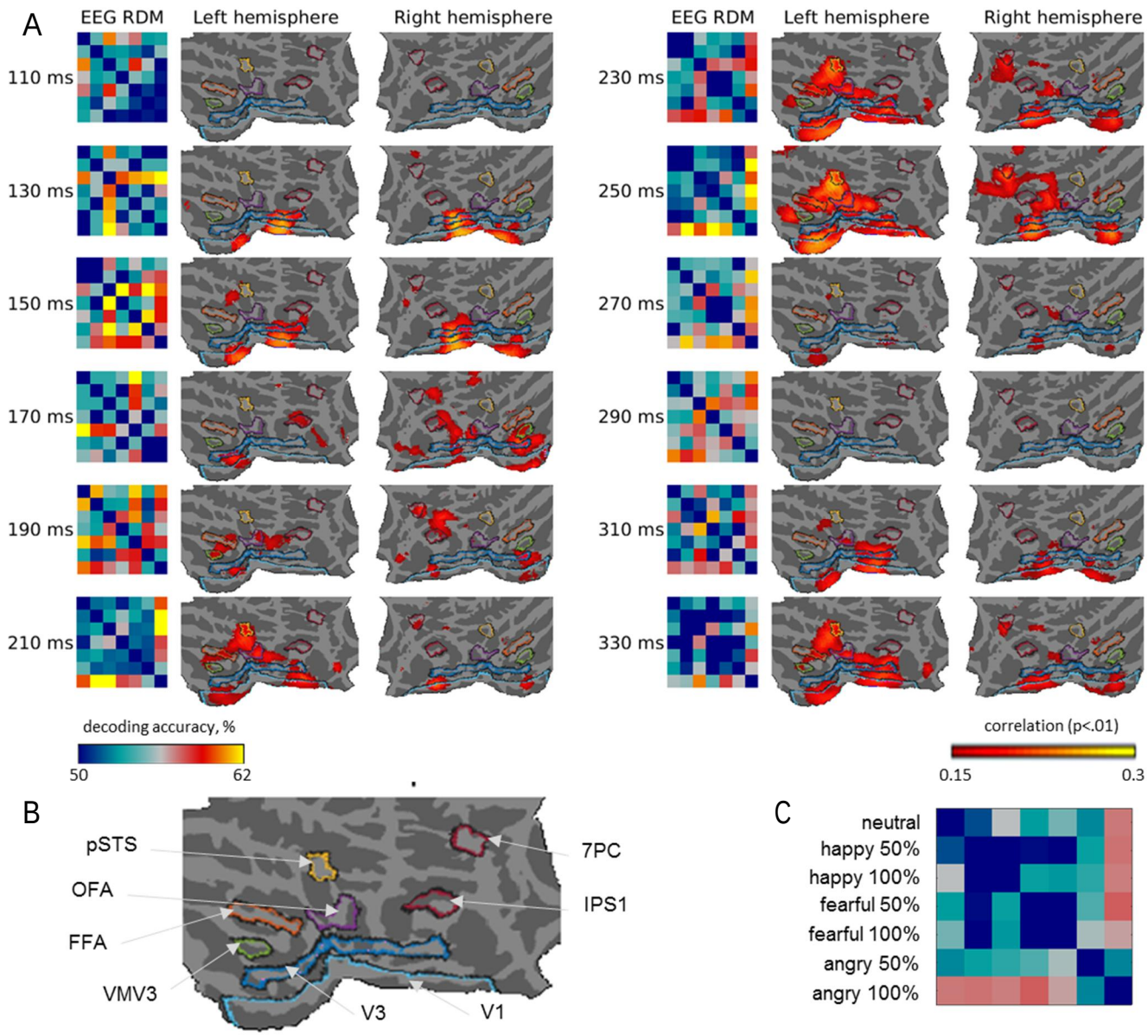

Figure 2. *EEG-fMRI searchlight correlations.* ***A)*** *At each timepoint, from left to right: EEG RDM, correlation between EEG and fMRI RDMs in each voxel in left and right hemisphere (permuted p<.01). The fMRI RDMs are same in each timepoint. The flatmaps contain approximately the posterior half of the brain.* ***B)*** *Labels for marked flatmap brain regions.* ***C)*** *Labels for RDM rows/columns.*

stimuli in a certain voxel or at certain timepoint. We found that information in EEG correlated with early visual cortex at 120-130 ms (Figure 2), enhancing within the EVC until 160 ms, after which it correlated also with an area posterior to right FFA as well as bilateral parietal areas at 170 ms. At 190 – 200 ms, correlations were found on parietal areas, at 210 – 250 ms in EVC as well as left FFA, OFA and part of left STS, and at 310 ms again with EVC (Figure 2). These results are consistent with temporal progress of information processing from posterior to more anterior sites in visual object

categorization (Cichy et al., 2016).

To further compare the fMRI-EEG –correlations, we performed ROI-analyses for eight selected areas (Figure 3B) from the BALSA parcellation map (Glasser et al., 2016): V1, fusiform (face) area (FFC in BALSA), lateral occipital (face) area (LO1 in BALSA), and posterior superior temporal sulcus (pSTS; TPOJ2 in BALSA), all from both hemispheres. The areas were selected based on their known role in face processing (Duchaine & Yovel, 2015). From each ROI, 5 voxels having the highest total information across all 21 expression conditions were selected, and the RDMs from each voxel were averaged for analysis. Thus, the selected ROIs represent the best information available in a given area. Note, however, that since searchlight analysis was used in pair-wise decodings, these 5 voxels contain information from the surrounding voxels as well (Figure 3B).

To compare the information structures between different fMRI ROIs and EEG, we calculated RDMs for each ROI (Figure 3A), and correlated them to the EEG data (average of three consecutive timepoints). Similarly to the whole brain searchlight analysis, V1 had highest correlation with earliest timepoints (~120 ms), staying significant for most of the time until around 500 ms (Figure 3C). Right OFA peaked a bit later at 170 - 190 ms, and again simultaneously with right FFA and left OFA, FFA and pSTS at ~250 ms. Finally, there were sustained correlations between ~300 – 500 ms with bilateral V1 and all areas from left hemisphere. No significant correlations for right pSTS were found. The overall spatio-temporal structure found here was quite similar to the one reported earlier in a face identity MEG-study (Vida et al., 2017), with accurate decodings at left V1 around 150 ms, and relatively higher for right LO and right FG around 250 ms. However, that study did not find a similar relative peak around 170 ms for OFA as we did.

3.4. Models for expression processing

Our experimental setup allows comparison of different models of expression encoding since we varied both expression intensities and categories. We compared our data with three model RDMs (Figure 4A). First, *Distance Model*, represented balanced processing of all emotions with neutral expression as a middle point. Thus, the representational distances between all 100% expressions to each other were given the maximum value (1), and distance from a given full expression (e.g. happy 100%) to intermediate intensity of the same expression (e.g. happy 50%) were ¼, to neutral face ½, and to intermediate intensities of other expressions (e.g. angry 50%) ¾. Intermediate expressions' (e.g. 50% happy to 50% angry) distance to each other were ½. In the second model (*Intensity Model*), no differences between different expressions were assumed; only differences between neutral, half morphs and expressions with full intensities were modelled. Third, *Category Model* represented categorical processing of emotions, with the intermediate and full expressions assumed to

be similar. The three models were not orthogonal, with correlations of distance model with intensity and category models being -.37 and .51, respectively, and with intensity and category model -.09.

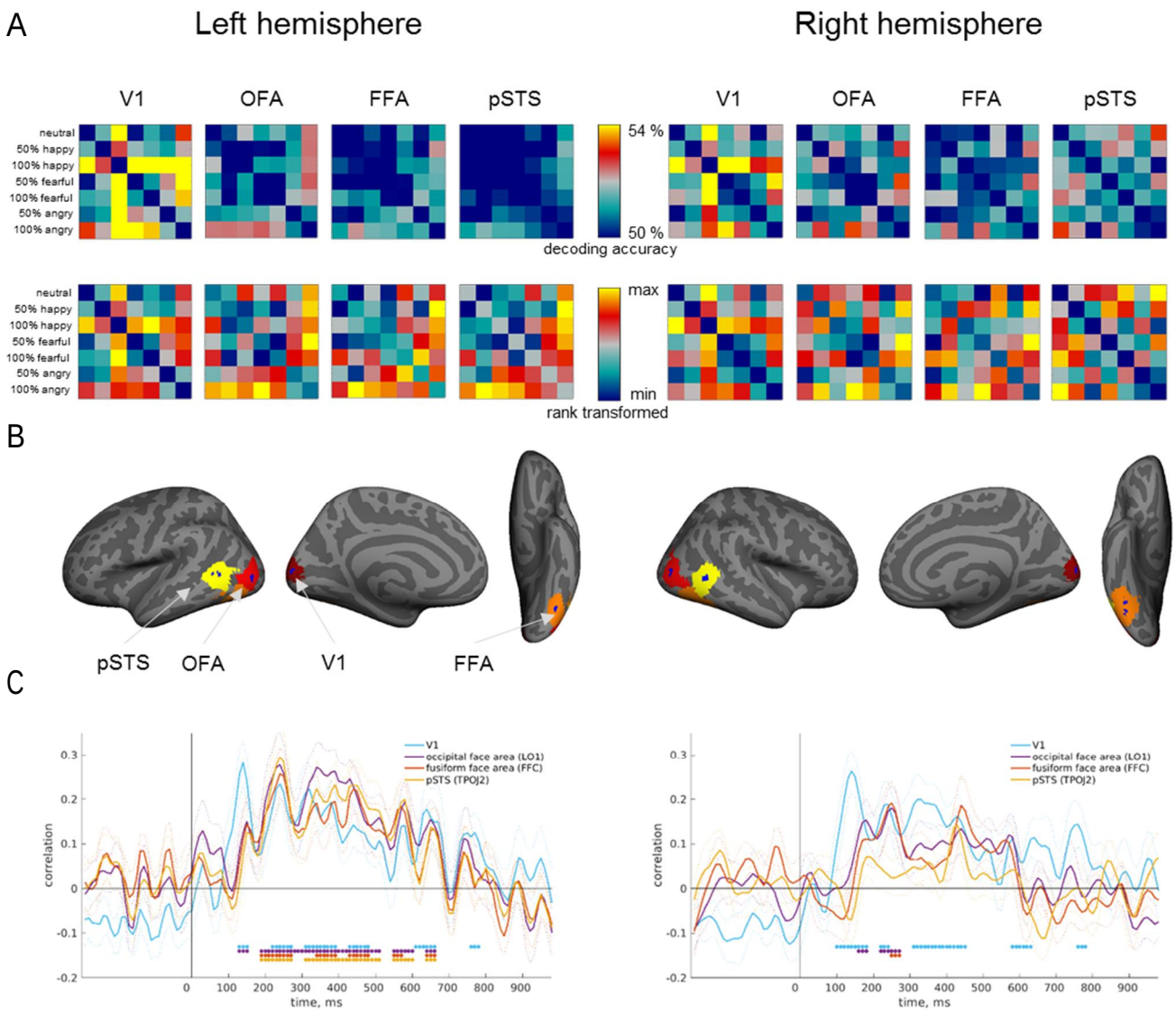

Figure 3. *ROI-analysis.* ***A)*** *RDMs from selected ROIs. In top row, decoding accuracies. In bottom row, rank transformed RDMs showing the relative weight of each decoding pair within each ROI.* ***B)*** *Locations of ROIs. Every ROI is the 5 voxels within a given BALSA-area that had highest total information over all decoding pairs. Voxel locations shown in dark blue, and the total area of their searchlight in yellow, orange and red.* ***C)*** *Correlations between each ROI and EEG from both hemispheres. Dots are significant timepoints (p(FDR)<.05, averages of three consecutive timepoints), and shaded lines are SEMs.*

Similarly to the EEG-fMRI –correlations, we correlated each voxel from fMRI (Figure 4B), and each timepoint from EEG (Figure 4C), with each of the models. The *Distance Model* correlated with EVC in fMRI, and with EEG between 130 – 520 ms. As the model is arguably quite similar to a low-level visual feature model, separating visually different pictures (different expressions), this could explain these findings. The *Intensity Model*, on the other hand, correlated most pronouncedly with an area near right pSTS. In EEG, the intensity model did not correlate significantly with classification accuracy at any timepoint (Figure 4C). This is in conjunction with the EEG-fMRI-analysis neither finding any correlations with right STS and EEG. It appears that STS codes especially the intensity of expressions, being less concerned with the lower-level visual features, and this coding is not temporally separable in the EEG data. This could be because the expression intensity is processed concurrently with other face features, which shadow the intensity coding in EEG measurements. Third, the *Category Model* had weak fMRI correlations with for example a right frontal patch near the inferior frontal gyrus face area (Figure 4B), and with EEG at 180 - 200 ms and 330 - 420 ms (Figure 4C). In addition, several small significant patches around the brain were found for both intensity and category models. However, as there is little evidence for face processing in these areas in the literature and our other results, no further conclusion are drawn from these findings.

Last, we correlated the above defined ROIs with the models to compare the models to the voxels with maximal information from each area. As seen in Figure 4D, these results were mainly similar to the results of the whole brain correlations. The distance model correlated with all ROIs from left hemisphere and V1 from right hemisphere, the category model correlated bilateral V1, left OFA, and right pSTS, and the intensity model with right FFA and STS. While these results were mainly consistent with the results of the whole brain model-searchlight, they also revealed the coding of intensities in right FFA and emotion category in right pSTS, not apparent in the whole brain results.

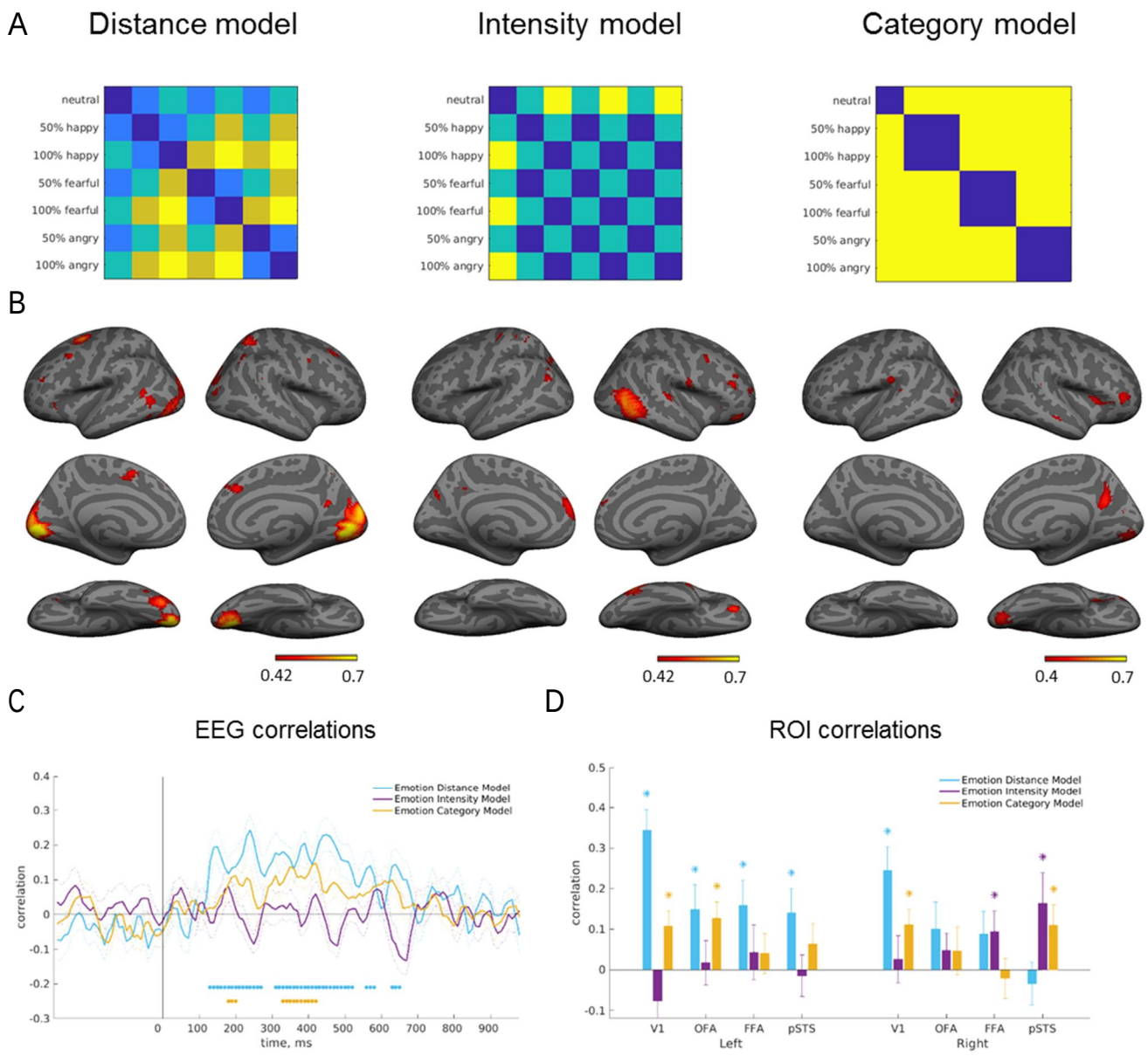

Figure 4. *Model analysis.* ***A)*** *RDMs for the three models.* ***B)*** *fMRI correlations with models (permuted p<.05).* ***C)*** *EEG correlations with the models. Dots are significant timepoints (p(FDR)<.05), and shaded lines are SEMs.* ***D)*** *correlations between selected fMRI ROIs and models. Dots indicate significant correlations (p<.05, uncorrected), and error bars are SEMs.*

4 Discussion

We investigated the processing of facial expressions and identities in the brain with EEG and fMRI. We used SVM decoding to measure the amount of information between each stimuli, and compared the information in the fMRI voxels

and EEG timepoints by correlating dissimilarity matrices obtained with the different methods. Our results showed gradual spatio-temporal spreading of expression information, starting from EVC at around 120 ms, enhancing in the early visual areas until 170 ms, when correlations were found near right FFA and in parietal areas. The fMRI data from left FFA, as well as left pSTS and OFA, correlated with the EEG data from 210 ms onwards. These results mirror and add up to a recent EEG-fMRI study of facial identity reconstruction showing EEG and fMRI correlations in pFG/aFG at 170-650 ms (Nemrodov et al., 2019), and a MEG-fMRI study showing progress of information related to visual object categories (Cichy et al., 2016). Comparisons between models and the fMRI data suggested expression distance/low-level visual feature-based processing in the early visual cortex, and expression intensity-based processing in the right pSTS, consistent with models of face perception. In contrast to previous M/EEG-fMRI studies investigating visual object categorization, our stimulus set contained only faces and was therefore visually less variable. Still, the multivariate data analysis methods, decoding and RSA, were able to reveal detailed spatio-temporal progress of facial information processing.

For identity information, our results were compromised by the motoric task of responding to the gender of the shown face. However, as we used searchlight analyses in the fMRI data, limiting the information used in each decoding to a small set of voxels, the motoric activations from the parietal and frontal lobes should not dilute possible findings in the visual and face processing areas most important for the aim of our study. Nevertheless, no significant decoding of identity information was found from these areas, possibly due to the smaller variance of visual information between the identities than the expressions. In EEG, although we found processing of identities from 330 ms onwards, the time trajectory of identity coding matched closely to that of the motoric responses. Thus, no further analyses or conclusions of the identity coding were made. To note, our decoding design required invariance to expressions in the identity decodings. Some previous studies with similar strict invariance requirement to viewpoints (Guntupalli et al., 2017) and generalization across face halves (Anzellotti & Caramazza, 2016), have found identity decoding in right inferior frontal and anterior temporal lobe, respectively, but not in FFA or OFA (although see Anzellotti, Fairhall, & Caramazza, 2014 for rotation-invariant decodings in OFA and FFA). However, as the area in frontal cortex is harder to disentangle from the motor responses, and our fMRI signal was weak in the anterior temporal lobe, our results cannot weight in on the role of these areas in identity processing.

Going deeper to our expression results, we first looked for the areas and timing of total information of the expressions. In EEG, expressions were decoded significantly between ~90 – 750 ms, replicating earlier findings of expression information starting around 70 - 110 ms (Dima et al., 2018; Müller-Bardorff et al., 2018; Sugase et al., 1999). While there

are multiple studies not finding as early an effect of expression processing (e.g., Blau, Maurer, Tottenham, & Mccandliss, 2007; Müller-Bardorff et al., 2016), this can be explained with the less sensitive methods mainly used in those studies. In fMRI, significant decoding was found, as expected, from early visual areas and the face processing areas in fusiform gyrus (FFA), lateral occipital (OFA) and near posterior superior temporal sulcus (pSTS). For the separate expressions, happy faces were decoded from early visual areas and most pronouncedly at early time points (~130 ms), while information on angry faces was found predominantly in right pSTS and FFA and mainly later timings (~200 ms), with fearful faces falling in between in EEG and showing no significant decodings in fMRI. These matching results of earlier/later timings in EEG with earlier visual areas/more anterior areas in fMRI for happy/angry faces anticipate the findings in EEG-fMRI analysis, discussed below. However, we also found even earlier coding of angry faces at 80 - 90 ms. Although this finding was quite weak, it mirrors an earlier study finding very early (90 ms) coding of angry faces versus neutral and happy faces, but not between happy and neutral faces (Dima et al., 2018). Besides these differences in peak latencies, the coding of angry faces remained statistically significant longer than other expressions, possibly showing the evolutionary importance of paying attention to angry faces, longer-lasting processing of eye than mouth information, or both.

We used representational similarity analysis to compare the information structure in EEG and in fMRI. One of our main findings was a gradually spreading information similarity with earlier time points (120 – 150 ms) and early visual cortex, and later time points (170 – 250 ms) and more anterior areas in (ventral) temporal lobe. The face models (Duchaine & Yovel, 2015; Haxby et al., 2000) postulate spatio-temporal progress in the processing of faces. However, as every non-invasive method (fMRI, M/EEG) have rather weak spatial or temporal resolution, few studies have been able to focus on both dimensions at once. Thus, while there is an exhaustive literature on both spatial areas of face processing from fMRI-studies**,** and temporal components found using M/EEG, the linkage between these results is much less understood. Our study takes a step into that direction, and shows, for example information in EEG at around 170 ms of the well-studied N170-component correlating with information near rFFA in fMRI. Our ROI-analyses similarly showed early peak of fMRI-measured V1 correlating with EEG, while OFA and FFA peaked later. Our time trajectories are similar to ones reported in a MEG-study (Vida et al., 2017), which found first peaks for V1, OFA and FFA at around 150 ms, when information in V1 clearly dominates, and later peaks ~250 ms, when the information from OFA and FFA are relatively higher, though mainly not surpassing correlations with V1. That study did not, however, find similar peak for rOFA around 170 ms as seen in our data. The different stimuli and methods (identities instead of expressions, unimodal-MEG, and long (60-ms) sliding window) could explain this; especially the long sliding window could miss the very short peak. Indeed, it is interesting that, despite the methodological differences, the results were quite similar. In relation to the debate

whether OFA and FFA process information sequentially or simultaneously, our ROI-analyses show weak evidence for earlier processing in OFA, although this was not supported by the whole brain searchlight and were not formally tested.

We compared our data to three models depicting emotional distance, emotional intensity, and emotion category information. The emotional distance, which most directly represented the low-level differences in the faces, matched with the information in the early visual cortex and EEG data from ~150 ms to ~500 ms. Intensity model, on the other hand, correlated with right pSTS and, to lesser extent, FFA, and not at all with EEG. As we did not find EEG-fMRI correlations from right pSTS, these model correlations provide an explanation for that: right pSTS seemed to code emotional intensity with less sensitivity to the lower level features, and this coding was not separable in EEG data. This is in line with an earlier study reporting STS to separate neutral faces from emotional faces, while being less sensitive to differences between different expressions (Zhang et al., 2016). Finally, the category model correlated with EEG at around 200 ms and 350 – 400 ms. For spatial locations of emotion category information, our results would suggest right ventral occipital areas, anterior STS and an anterior frontal patch near inferior frontal gyrus face area. However, as these areas, as well as a frontal patch correlating with the intensity model, were not found from any other analyses and had quite weak correlations with the models, we will not draw further conclusions concerning them.

It is worth underlining what the RDM correlations between EEG and fMRI exactly measure. The RDM correlations are insensitive to the *amount* of information, and measure only the information structure, namely the *relative* information between the different expressions. This is contrary to the single-modality decodings, which show the amount of information, but not the relative information. Thus, it is entirely possible there to be an area in the brain having superior decoding of all the different expressions, but is hardly showing up in the model or EEG-fMRI correlations. One example of these inconsistencies in our data is the rFFA. While the fMRI decodings of angry faces were statistically significant in right but not left FFA, the EEG-fMRI analysis found left but not right FFA correlating with EEG at 250 ms, when the EEG RDM was defined mainly by accurate decodings of angry faces. This contradictory-seeming result could be explained by larger *amount* of information in right than left FFA, but *relatively* better decoding of angry than other expressions in left FFA. This example also shows how these two methods complement each other.

In conclusion, we showed a glimpse of the spatio-temporal progress of within expression coding in the brain. Starting from the early visual cortex, the decoding spreads to an area near right FFA at 170 ms, and to left OFA, FFA and pSTS after 200 ms. By comparing these to the unimodal results from EEG and fMRI, the happy faces appeared to drive the correlations in EVC, and fearful and especially angry faces in the face processing areas. The model correlations suggested that the processing of faces hierarchically changes from visual features in EVC to expression intensities in right FFA, and

intensities and categories in right pSTS. All in all, our results suggest that combining M/EEG and fMRI provides a useful new tool in the search of the spatio-temporal processing dynamics in the human brain.

Acknowledgements

This work was supported by the Academy of Finland grant to VRS (grant number 298329).

Conflicting interests' statement

Authors declare no conflicting interests.

References

Ambrus, G. G., Kaiser, D., Cichy, R. M., & Kovács, G. (2019). The Neural Dynamics of Familiar Face Recognition. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhz010

Anzellotti, S., & Caramazza, A. (2016). From Parts to Identity: Invariance and Sensitivity of Face Representations to Different Face Halves. *Cerebral Cortex*, 26(5), 1900–1909. http://doi.org/10.1093/cercor/bhu337

Anzellotti, S., Fairhall, S. L., & Caramazza, A. (2014). Decoding representations of face identity that are tolerant to rotation. *Cerebral Cortex*, 24(8), 1988-1995. http://doi.org/10.1093/cercor/bht046

Atkinson, A. P., & Adolphs, R. (2011). The neuropsychology of face perception: beyond simple dissociations and functional selectivity. *Philosophical Transactions of the Royal Society of London. Series B,* 366(1571), 1726–1738. http://doi.org/10.1098/rstb.2010.0349*Cerebral Cortex*, 24(8), 1988–1995.

Bernstein, M., & Yovel, G. (2015). Two neural pathways of face processing: A critical evaluation of current models. *Neuroscience and Biobehavioral Reviews*, 55, 536–546. http://doi.org/10.1016/j.neubiorev.2015.06.010

Blau, V. C., Maurer, U., Tottenham, N., & Mccandliss, B. D. (2007). The face-specific N170 component is modulated by emotional facial expression. *Behavioral and Brain Functions*, 13, 1–13. http://doi.org/10.1186/1744-9081-3-7

Bode, S., Feuerriegel, D., Bennett, D., & Alday, P. M. (2018). The Decision Decoding ToolBOX (DDTBOX) – A Multivariate Pattern Analysis Toolbox for Event-Related Potentials. *Neuroinformatics*, 1–16. http://doi.org/10.1007/s12021-018-9375-z

Bushnell, I. W. R. (2001). Mother's Face Recognition in Newborn Infants: Learning and Memory. *Infant and Child Development*, 10, 67–74. http://doi.org/10.1002/icd.248

Bushnell, I. W. R., Sai, F., & Mullin, J. T. (1989). Neonatal recognition of the mother's face. *British Journal of Developmental Psychology*, 7(1), 3–15. http://doi.org/10.1111/j.2044-835X.1989.tb00784.x

Carlin, J. D., & Kriegeskorte, N. (2017). Adjudicating between face-coding models with individual-face fMRI responses. *PLoS Computational Biology*, 13(7), 1–28. http://doi.org/10.1371/journal.pcbi.1005604

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3), 1–27. http://doi.org/10.1145/1961189.1961199

Chang, L., & Tsao, D. Y. (2017). The Code for Facial Identity in the Primate. *Cell*, 169(6), 1013–1020.e14. http://doi.org/10.1016/j.cell.2017.05.011

Cichy, R. M., & Pantazis, D. (2017). Multivariate pattern analysis of MEG and EEG: A comparison of representational structure in time and space. *NeuroImage*, 158, 441–454. http://doi.org/10.1016/J.NEUROIMAGE.2017.07.023

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, 17(3), 455-62. http://dx.doi.org/10.1038/nn.3635

Cichy, R. M., Pantazis, D., & Oliva, A. (2016). Similarity-Based Fusion of MEG and fMRI Reveals Spatio-Temporal Dynamics in Human Cortex During Visual Object Recognition. *Cerebral Cortex*, 26(8), 3563–3579. http://doi.org/10.1093/cercor/bhw135

Collins, J. A., & Olson, I. R. (2015). Beyond the FFA: The Role of the Ventral Anterior Temporal Lobes in Face Processing. *Neuropsychologia*, 65–79. http://doi.org/10.1016/j.neuropsychologia.2014.06.005.Beyond

Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 20(3), 273–297. http://doi.org/10.1007/BF00994018

Dima, D. C., Perry, G., Messaritaki, E., Zhang, J., & Singh, K. D. (2018). Spatiotemporal dynamics in human visual cortex rapidly encode the emotional content of faces. *Human Brain Mapping*. http://doi.org/10.1002/hbm.24226

Dobs, K., Schultz, J., Bülthoff, I., & Gardner, J. L. (2018). Task-dependent enhancement of facial expression and identity representations in human cortex. *NeuroImage*, 172(July 2017), 689–702. http://doi.org/10.1016/j.neuroimage.2018.02.013

Dobson, V., & Teller, D. Y. (1978). Visual acuity in human infants: A review and comparison of behavioral and electrophysiological studies. *Vision Research*, 18(11), 1469–1483. http://doi.org/10.1016/0042-6989(78)90001-9

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

Ebner, N. C., Riediger, M., & Lindenberger, U. (2010). FACES - A database of facial expressions in young, middle-aged, and older women and men: Development and validation. *Behavior Research Methods*, 42(1), 351–62. http://doi.org/10.3758/BRM.42.1.351

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., … Van Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615), 171–178. http://doi.org/10.1038/nature18933

Gold, J. M., Mundy, P. J., & Tjan, B. S. (2012). The Perception of a Face Is No More Than the Sum of Its Parts. *Psychological Science*, 23(4), 427–434. http://doi.org/10.1177/0956797611427407

Greening, S. G., Mitchell, D. G. V., & Smith, F. W. (2018). Spatially generalizable representations of facial expressions: Decoding across partial face samples. *Cortex*, 101, 31–43. http://doi.org/10.1016/j.cortex.2017.11.016

Guntupalli, J. S., Wheeler, K. G., & Gobbini, M. I. (2017). Disentangling the Representation of Identity from Head View Along the Human Face Processing Pathway. *Cerebral Cortex*, 27, 46–53. http://doi.org/10.1093/cercor/bhw344

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, 4(6), 223–233. http://doi.org/10.1016/S1364-6613(00)01482-0

Hebart, M. N., Bankson, B.B., Harel, A., Baker, C. I., & Cichyl, R. M. (2018). The representational dynamics of task and object processing in humans. *eLife*, 7, 1-21. https://doi.org/10.7554/eLife.32816

Hebart, M. N., Görgen, K., & Haynes, J.-D. (2015). The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. *Frontiers in Neuroinformatics*, 8, 88. http://doi.org/10.3389/fninf.2014.00088

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. http://doi.org/10.1016/j.cortex.2015.06.030

Hinojosa, J. A., Mercado, F., & Carretié, L. (2015). N170 sensitivity to facial expression: A meta-analysis. *Neuroscience & Biobehavioral Reviews*, 55, 498–509. http://doi.org/10.1016/J.NEUBIOREV.2015.06.002

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society of London. Series B*, 361(1476), 2109–28. http://doi.org/10.1098/rstb.2006.1934

Kothe, C. (2013). The artifact subspace reconstruction method. http://sccn.ucsd.edu/eeglab/plugins/ASR.pdf.

Kriegeskorte, N., Goebel, R., & Bandettini, P. a. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, 103(10), 3863–8. http://doi.org/10.1073/pnas.0600244103

Kriegeskorte, N., Mur, M., & Bandettini, P. a. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2, 4. http://doi.org/10.3389/neuro.06.004.2008

Mullen, T (2012). CleanLine EEGLAB plugin. San Diego, CA: Neuroimaging Informatics Toolsand Resources Clearinghouse (NITRC).

Müller-Bardorff, M., Bruchmann, M., Mothes-Lasch, M., Zwitserlood, P., Schlossmacher, I., Hofmann, D., Straube, T. (2018). Early brain responses to affective faces: A simultaneous EEG-fMRI study. *NeuroImage*, 178, 660–667. http://doi.org/10.1016/J.NEUROIMAGE.2018.05.081

Müller-Bardorff, M., Schulz, C., Peterburs, J., Bruchmann, M., Mothes-Lasch, M., Miltner, W., & Straube, T. (2016). Effects of emotional intensity under perceptual load: An event-related potentials (ERPs) study. *Biological Psychology*, 117, 141–149. http://doi.org/10.1016/j.biopsycho.2016.03.006

Nemrodov, D., Behrmann, M., Niemeier, M., Drobotenko, N., & Nestor, A. (2019). Multimodal evidence on shape and surface information in individual face processing. *NeuroImage*, 184, 813–825. http://doi.org/10.1101/299933

Pitcher, D., Walsh, V., & Duchaine, B. (2011). The role of the occipital face area in the cortical face perception network. *Experimental Brain Research*, 209(4), 481–493. http://doi.org/10.1007/s00221-011-2579-1

Reid, V. M., Dunn, K., Young, R. J., Amu, J., Donovan, T., & Reissland, N. (2017). The Human Fetus Preferentially Engages with Face-like Visual Stimuli. *Current Biology*, 27(12), 1825–1828. http://doi.org/10.1016/j.cub.2017.05.044

Richler, J. J., & Gauthier, I. (2014). A Meta-Analysis and Review of Holistic Face. *Psychological Bulletin*, 140(5), 1281–1302. http://doi.org/10.1037/a0037004.A

Rossion, B., & Jacques, C. (2011). The N170: understanding the time-course of face perception in the human brain. *The Oxford Handbook of ERP Components*, Oxford University Press. 115–142. http://doi.org/10.1093/oxfordhb/9780195374148.013.0064

Sadeh, B., Podlipsky, I., Zhdanov, A., & Yovel, G. (2010). Event-related potential and functional MRI measures of face-selectivity are highly correlated: A simultaneous ERP-fMRI investigation. *Human Brain Mapping*, 31(10), 1490–1501. http://doi.org/10.1002/hbm.20952

Said, C. P., Haxby, J. V., & Todorov, A. (2011). Brain systems for assessing the affective value of faces. *Philosophical Transactions of the Royal Society of London B*, 366(1571), 1660–1670. http://doi.org/10.1098/rstb.2010.0351

Salmela, V., Salo, E., Salmi, J., & Alho, K. (2018). Spatiotemporal Dynamics of Attention Networks Revealed by Representational Similarity Analysis of EEG and fMRI. *Cerebral Cortex*, 28(2), 549–560. http://doi.org/10.1093/cercor/bhw389

Schweinberger, S. R., & Neumann, M. F. (2016). Repetition effects in human ERPs to faces. *Cortex*, 80, 141–153. http://doi.org/10.1016/J.CORTEX.2015.11.001

Shen, J., & Palmeri, T. J. (2015). The Perception of a Face is Greater Than the Sum of Its Parts The perception of a face is no more than the sum of its parts. *Psychonomic Bulletin & Review*, 22, 710–716. http://doi.org/10.3758/s13423-014-0726-y

Sugase, Y., Yamane, S., Ueno, S., & Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature*, 400(6747), 869–873. http://doi.org/10.1038/23703

Tanaka, J. W., & Simonyi, D. (2016). The “parts and wholes” of face recognition: a review of the literature. *Quarterly Journal of Experimental Psychology*, 218, 1–14. http://doi.org/10.1080/17470218.2016.1146780

Taubert, J., Apthorp, D., Aagten-Murphy, D., & Alais, D. (2011). The role of holistic processing in face perception: Evidence from the face inversion effect. *Vision Research*, 51(11), 1273–1278. http://doi.org/10.1016/j.visres.2011.04.002

Valentine, T. (1988). Upside-down faces: a review of the effect of inversion upon face recognition. *British Journal of Psychology*, 79(4), 471–491. http://doi.org/10.1111/j.2044-8295.1988.tb02747.x

Vida, M. D., Nestor, A., Plaut, D. C., & Behrmann, M. (2017). Spatiotemporal dynamics of similarity-based neural representations of facial identity. *Proceedings of the National Academy of Sciences of the United States of America*, 114(2), 388–393. http://doi.org/10.1073/pnas.1614763114

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, 81(1), 141–145. http://doi.org/10.1037/h0027474

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. http://doi.org/10.1016/j.neuroimage.2016.01.045