Attentional Weighting in the Face Processing Network: A Magnetic Response Image-guided Magnetoencephalography Study Using Multiple Cyclic Entrainments

Eelke de Vries[1,2] and Daniel Baldauf[1]

Abstract

■ We recorded magnetoencephalography using a neural entrainment paradigm with compound face stimuli that allowed for entraining the processing of various parts of a face (eyes, mouth) as well as changes in facial identity. Our magnetic response image-guided magnetoencephalography analyses revealed that different subnodes of the human face processing network were entrained differentially according to their functional specialization. Whereas the occipital face area was most responsive to the rate at which face parts (e.g., the mouth) changed, and face patches in the STS were mostly entrained by rhythmic changes in the eye region, the fusiform face area was the only subregion that was strongly entrained by the rhythmic changes in facial identity. Furthermore, top–down attention to the mouth, eyes, or identity of the face selectively modulated the neural processing in the respective area (i.e., occipital face area, STS, or fusiform face area), resembling behavioral cue validity effects observed in the participants' RT and detection rate data. Our results show the attentional weighting of the visual processing of different aspects and dimensions of a single face object, at various stages of the involved visual processing hierarchy. ■

INTRODUCTION

Neuroimaging and electrophysiology have revealed multiple cortical face-selective regions that are spatially and functionally separable (Issa, Papanastassiou, & DiCarlo, 2013; Freiwald & Tsao, 2010; Nichols, Betts, & Wilson, 2010; Tsao, Freiwald, Tootell, & Livingstone, 2006; Tsao, Freiwald, Knutsen, Mandeville, & Tootell, 2003; Haxby et al., 2001) and form a distributed cortical network specialized for face perception (Moeller, Freiwald, & Tsao, 2008; Tsao, Moeller, & Freiwald, 2008; Calder & Young, 2005; Haxby, Hoffman, & Gobbini, 2000). Three of the most studied face-selective regions are found along the occipital-temporal cortex: the occipital face area (OFA) in the inferior occipital gyrus (Jonas et al., 2012, 2014; Pitcher, Walsh, & Duchaine, 2011; Pitcher, Walsh, Yovel, & Duchaine, 2007; Gauthier et al., 2000; Puce, Allison, Asgari, Gore, & McCarthy, 1996), the fusiform face area (FFA) in the middle fusiform gyrus (Parvizi et al., 2012; Kanwisher & Yovel, 2006; Tsao et al., 2006; Kanwisher, McDermott, & Chun, 1997), and a region in the STS (Itier, Alain, Sedore, & McIntosh, 2007; Hoffman & Haxby, 2000; Perrett et al., 1985).

Whereas the OFA is hypothesized to be more responsive to local information, such as face parts (Pitcher et al., 2011; Liu, Harris, & Kanwisher, 2010), the FFA is often found to be more tuned for face identity (Grill-Spector, Knouf, & Kanwisher, 2004) or face categorization (Afraz, Boyden, & DiCarlo, 2015; Afraz, Kiani, & Esteky, 2006; Turk, Rosenblum, Gazzaniga, & Macrae, 2005; Liu, Harris, & Kanwisher, 2002). The areas in the STS seem to be closely related to the processing of eye gaze (Carlin & Calder, 2013; Carlin, Calder, Kriegeskorte, Nili, & Rowe, 2011).

Here, we tested the effects of visual selective attention on these three separate aspects of face processing. According to the biased competition theory (Desimone & Duncan, 1995), selective attention is the central mechanism that biases processing for behaviorally relevant stimuli by facilitating the processing of important information and, at the same time, filtering out or suppressing irrelevant information. On a behavioral level, it has been shown that visual attention can filter visual input on the basis of spatial location (Posner, 1980); on the basis of visual features, such as color (Wegener, Ehn, Aurich, Galashan, & Kreiter, 2008; Rossi & Paradiso, 1995); or on the basis of visual objects (Egly, Driver & Rafal, 1994; Duncan, 1984). Respective neuronal effects of visual attention have been found in populations of neurons specialized in the processing of topographic space (e.g., Sprague & Serences, 2013; Baldauf & Deubel, 2008, 2010; Gregoriou, Gotts, Zhou, & Desimone, 2009; Baldauf, Cui, & Andersen, 2008; Siegel, Donner, Oostenveld, Fries, & Engel, 2008;

[1]University of Trento, [2]Vrije Universiteit Amsterdam

Moore & Armstrong, 2003; Nobre, Gitelman, Dias, & Mesulam, 2000; Corbetta et al., 1998; Mangun & Hillyard, 1991), low-level visual features (Schwedhelm, Baldauf, & Treue, 2017; Bichot, Heard, DeGennaro, & Desimone, 2015; Serences & Boynton, 2007; Müller et al., 2006; Bichot, Rossi, & Desimone, 2005; Giesbrecht, Woldorff, Song, & Mangun, 2003; Saenz, Buracas, & Boynton, 2002; Treue & Maunsell, 1996), and object classes (Cohen & Tong, 2015; Baldauf & Desimone, 2014; Schoenfeld, Hopf, Merkel, Heinze, & Hillyard, 2014; Ciaramitaro, Mitchell, Stoner, Reynolds, & Boynton, 2011; Corbetta et al., 2005; Scholl, 2001; O'Craven, Downing, & Kanwisher, 1999). In this study, top–down attention to any one of the three face components (face identity, face parts, and eye gaze) was hypothesized to facilitate both the neural activity in the respective subnetwork related to processing of the attended stimulus (Spitzer, Desimone, & Moran, 1988) and the behavioral performance in a target detection task. By investigating the temporal dynamics of the complete occipital face processing network (OFA, FFA, and STS), we were able to dissect the functional compartmentalization of the system and the functional specialization of its components by demonstrating selective attentional modulation in each of the three regions.

METHODS

Participants

Ten healthy participants (five men, mean age = 26.3 years, SD = 3.59) took part in the study. All gave written informed consent, had normal or corrected-to-normal vision, and were naive regarding the aim of the study. Because of signal and movement artifacts, one participant was excluded from the magnetoencephalography (MEG) analyses. The entire session lasted approximately 2.5 hr including preparation time (1.5 in MEG).

Stimuli

The stimuli were created from a database of 128 pictures, which we created specifically for the purpose of this study (database available upon request). Eight individuals (two women, six men) posed for a total of 16 pictures, each with a specified facial expression (see Figure 1A). The outline of the face was cropped from each image, and the picture was then converted to gray scale, placed on a solid gray background (corresponding to the mean overall luminance of the image), and resized to 400 × 400 pixels. The luminance and spatial frequencies were equated using the SHINE toolbox (Willenbockel et al., 2010). Using the Random Image Structure Evolution (RISE) procedure (Sadr & Sinha, 2004), the image visibility was modulated. With this technique, the level of visible semantic content can be manipulated by partially randomizing the phase spectrum of the Fourier-transformed image while retaining the low-level visual features such as its original power spectrum, luminance, and contrast. This procedure results in a sequence of images, in which the visibility of the depicted face gradually emerges and
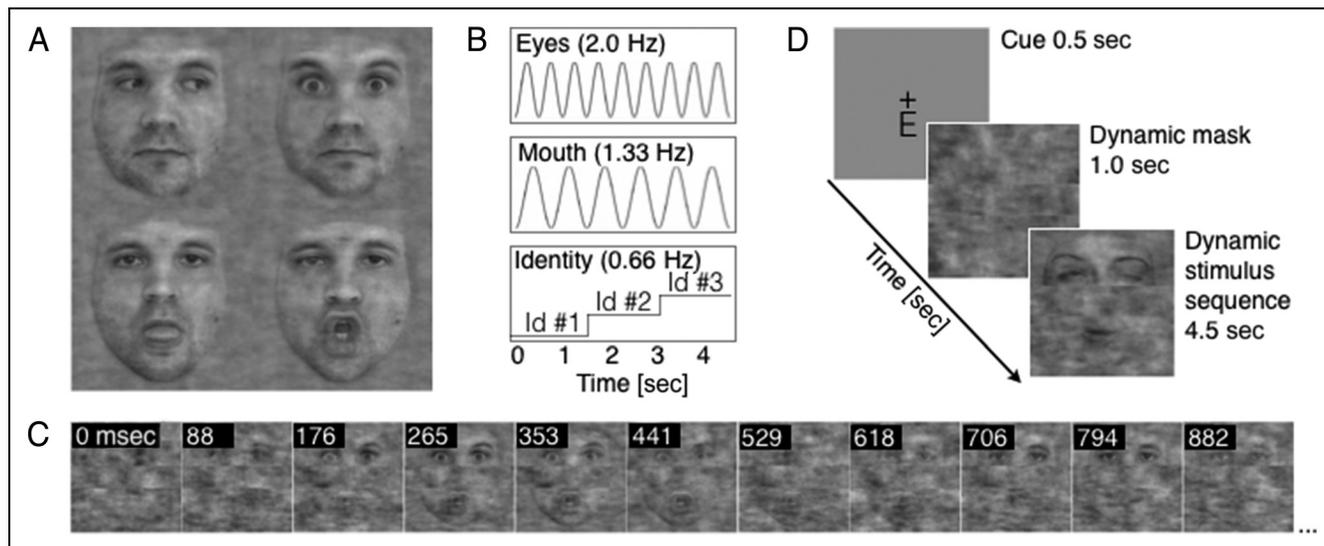

Figure 1. Stimuli and trial sequence. (A) Custom-made database of image stimuli with various identities and facial expressions. Eight volunteer models were asked to pose for 16 images each. For eight images, only the top part containing the eyes was used, and for the other eight, only the bottom part containing the mouth was used. Above are four examples: look toward the right (top left), eyes wide open (top right), stick out tongue (bottom left), and stick out lips (bottom right). Composite images were created with an eye sequence in the top section, a noise mask in the middle section, and a mouth image sequence in the bottom section. (B) Frequency tags: Three frequency tags were embedded in the stimuli; the visibility of the eyes and mouth images oscillated sinusoidally at 2 and 1.33 Hz, respectively, whereas the associated identity changed rhythmically at 0.66 Hz. (C) Several example frames taken from the dynamic sequence of stimuli images show that the visibility of the eyes and mouth oscillate at different frequencies. (D) A typical trial sequence: Trial onset was indicated by a change in color of the fixation cross and cue (from gray to black). This was followed by a short baseline period with a dynamic mask (containing no semantic information) and 4.5 sec of actual stimuli presentation.

disappears (i.e., sinusoidally oscillating), repeatedly at a steady rhythm. In addition, we also created a phase-scrambled RISE mask by applying the same procedure to a randomly selected image at minimum visibility. The eye and mouth regions from the resulting RISE sequences were extracted (upper and lower 160 pixels, respectively) and used to create new composite images, which consisted of three parts: a combination of an eye sequence oscillating at 2.00 Hz in the upper section (160 pixels), a RISE mask with superimposed fixation cross and cue indicator in the middle section (80 pixels), and a mouth image sequence in the bottom section (160 pixels), oscillating at 1.33 Hz (see Figure 1B). Therefore, the upper and lower image parts containing the eyes and mouth, respectively, were frequency tagged (at 2.00 and 1.33 Hz, respectively), and the associated identity changed rhythmically at 0.66 Hz. All the induced oscillations are not simply flicker on and off but are instead a gradually changing image sequence, in which the rates of image change are 0.66, 1.33, and 2.0 Hz, respectively. A movie showing an example stimulus is available at https://figshare.com/s/f8a1e2760937ca35c4f0 (Table 1).

Trial Sequence and Design

The experimental stimuli were presented on a PC, using PsychToolbox (Brainard, 1997) for MATLAB. Each trial lasted 6 sec starting with a 1-sec baseline period of masked images (i.e., RISE sequence with no visible objects), followed by 4.5 sec of stimuli presentation, and ended with another 0.5 sec of masked images to allow for late behavioral responses (see Figure 1D). Each trial was preceded by a fixed interval of 2.55 sec plus an interval that varied randomly between 0 and 100 msec. A cue below the central fixation cross indicated the target for that trial with 75% cue validity ("P" for person/identity, "E" for eyes, and "M" for mouth). For example, if participants were attending to the identity, then trials with changes to the formation of the mouth or eyes would be invalid trials. Throughout each trial, the cue and

Table 1. Face Expressions Used for the Stimuli

Eyes	Mouth
Neutral	Stick out tongue
Look right	Stick out lips
Look left	Say "fa"
Wide open	Mouth open
Left eye closed	Mouth open wide
Right eye closed	Suck in lips
Both eyes closed	Smile (no teeth)
Both eyes squeezed shut	Smile (with teeth)

fixation cross remained visible at the center of the stimuli display. Participants had to keep strict eye fixation throughout the trial, while covertly attending to the cued aspect of the face. Eye position was continuously monitored by an MEG-compatible eye-tracking device (Eyelink, SR-Research Ltd.). Participants had to respond by button press when detecting one of the three targets, which could be either an eye gaze toward the right, a tongue sticking out, or the appearance of a specific identity. All participants completed 450 trials, evenly distributed over five experimental blocks. Within each block, trials were grouped in random order in sets of 10 trials with a common attention cue and included a random number of trials (between 2 and 4) with an invalid cue. These cue groups were ordered in a (semi-)randomized fashion to minimize any repetition effects. There was a fixed interval of 12 sec between each group of trials. A new experimental block was started when the participant indicated to be ready to continue.

Behavioral Data Analysis

Trials with extreme RTs (i.e., < 200 msec) were excluded because they are likely to represent either guesses or inattentiveness (Whelan, 2008). For the RT analysis, all trials with outliers (exceeding 2.5 SDs based on each individual's mean) were excluded.

MEG Data Acquisition and Analysis

Whole-head MEG recordings were obtained at a sampling rate of 1000 Hz using a 306-channel (204 first-order planar gradiometers, 102 magnetometers) VectorView MEG system (Neuromag, Elekta Inc.) in a magnetically shielded room (AK3B, Vacuum Schmelze). For each participant, the individual head shape was digitized with a Polhemus Fastrak digitizer (Polhemus), including fiducial landmarks (nasion, preauricular points) and about 200 additional points on the scalp, all evenly spread out over the participant's head. Landmarks and head-position induction coils were digitized twice to ensure that their spatial accuracy was less than 1 mm. When positioning the participant, we ensured tight contact to the dewar. Participants were instructed to avoid any head, body, or limb movements and were asked to keep strict eye fixation and to avoid eye blinks as much as possible during stimulus presentation. The position of the head inside the dewar was measured by head-positioning coils (electromagnetic induction) before and after each recording block. In general, head movements did not exceed 1 cm within and between blocks. For three participants, the displacement between experimental blocks was >1 cm, and for those participants, source estimations were completed separately for each block and then averaged across blocks.

After visual inspection and exclusion of noisy recording channels, external noise was removed offline from the MEG recordings using MaxFilter software (tsss-filters; Taulu & Simola, 2006). The continuous data were first

visually inspected for system-related artifacts (e.g., SQUID jumps), and contaminated sensors were removed and interpolated (i.e., replaced by the averaged signal of neighboring sensors). A maximum of 12 sensors per experimental run had to be removed and interpolated. MEG data were then analyzed using Brainstorm (Tadel, Baillet, Mosher, Pantazis, & Leahy, 2011). The continuous recordings were segmented into epochs of 5.5 sec, starting 1 sec before stimulus onset and ending 4.5 sec after stimuli onset. The 500 msec before stimulus onset were used for baseline correction (DC subtraction). Each epoch was visually inspected, and those containing physiological artifacts (e.g., eye blinks) or other artifacts were discarded from further analyses (Gross et al., 2013). This resulted in an average of 23% of trials per participant that had to be discarded. To increase the signal-to-noise ratio of the experimentally induced frequency tags, the data for each participant were averaged for each condition in the time domain. We used minimum-norm estimates (Hämäläinen & Ilmoniemi, 1994) with overlapping spheres for the reconstruction of neuronal sources. The 3-D head model was based on an individual segmentation of the participant's MRI (see below). All source reconstructions were done in MATLAB with the Brainstorm toolbox. Our source space contained 15,000 vertices. To allow for interparticipant comparisons, the averaged source maps were normalized with respect to 500-msec baseline (z scores). The normalized averages from each ROI were then transformed into the frequency domain by means of a Fourier transformation. The signal-to-noise ratio was evaluated by dividing the amplitude at each tagging frequency by the average of their respective neighboring frequency bins.

MRI Acquisition, Analysis, and Coregistration

For each participant, high-resolution T1-weighted anatomical scans were acquired in a 4-T Bruker MedSpec Biospin MR scanner with an 8-channel birdcage head coil (magnetization prepared rapid gradient echo, $1 \times 1 \times 1$ mm, field of view = 256×224, 176 slices, repetition time = 2700 msec, echo time = 4.18 msec, inversion time = 1020 msec, 7° flip angle). The anatomical scans were then 3-D reconstructed using software (Dale, Fischl, & Sereno, 1999; Fischl, Sereno, & Dale, 1999) and used in the 3-D forward models of the MEG analyses.

ROIs

ROIs were defined for each participant based on local evoked responses of the initial presentation of the first face stimulus. Previous studies reported face-selective evoked responses occurring 100 and 170 msec, respectively, after stimulus onset (Alonso-Prieto, Belle, Liu-Shuang, Norcia, & Rossion, 2013). Liu et al. (2002) investigated the response patterns of the M100 and M170 and connected each to a different stage in the processing of face information. They showed that the M100 is

correlated with face categorization (i.e., the discrimination of faces vs. nonfaces), but not with face identification (i.e., the discrimination of individual faces), and that the M170 is correlated with face identification. The M100 and M170 also demonstrated opposite response patterns, such that the M100 showed a stronger response to face parts whereas the M170 showed a stronger response to configurations. Their findings suggest that local information (i.e., face parts) is extracted first and is used for face categorization, whereas global information (i.e., configuration) is extracted at a later stage and is used for discriminating between individual faces. In respect to the neural sources, the evoked responses at 100 msec have been localized in the OFA (Sadeh, Podlipsky, Zhdanov, & Yovel, 2010; Pitcher et al., 2007), whereas the evoked responses at 170 msec (Caharel, d'Arripe, Ramon, Jacques, & Rossion, 2009; Heisz, Watter, & Shedden, 2006; Jacques & Rossion, 2006) have been localized in the FFA (Gao et al., 2013; Deffke et al., 2007; Hoshiyama, Kakigi, Watanabe, Miki, & Takeshima, 2003; Halgren, Raij, Marinkovic, Jousmäki, & Hari, 2000) and STS (Dalrymple et al., 2011; Sadeh et al., 2010; Itier & Taylor, 2004). We therefore used the peak activations at about 100 and 170 msec for localizing the OFA and FFA/STS in the occipital-temporal cortex. To determine the exact spatial extent of the ROIs, the minimum-norm estimate maps were thresholded. If multiple separate peaks survived thresholding, those regions were combined in the same ROI. All ROIs were defined in each participant's individual MRI space, which was coregistered with the MEG Squid array to guide the reconstruction of neural sources within their individual anatomical frame
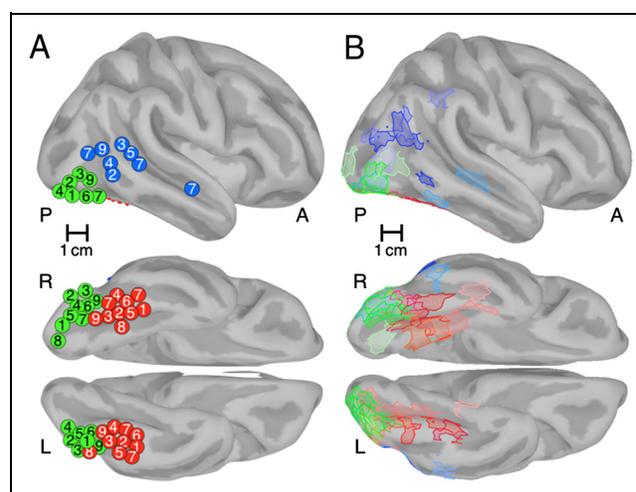

Figure 2. ROIs. Mean coordinates of the FFA (red), OFA (green), and STS (blue), (A) as reported in recent literature (see Arcurio, Gold, & James, 2012 [1]; Davies-Thompson, Gouws, & Andrews, 2009 [2]; Fairhall & Ishai, 2007 [3]; Hoffman & Haxby, 2000 [4]; Jiang et al., 2011 [5]; Nichols et al., 2010 [6]; Pinsk et al., 2009 [7]; Spiridon, Fischl, & Kanwisher, 2006 [8]; Weiner & Grill-Spector, 2013 [9]) and (B) as identified in each experimental participant of this study. The different shades of each color denote different individual participants.
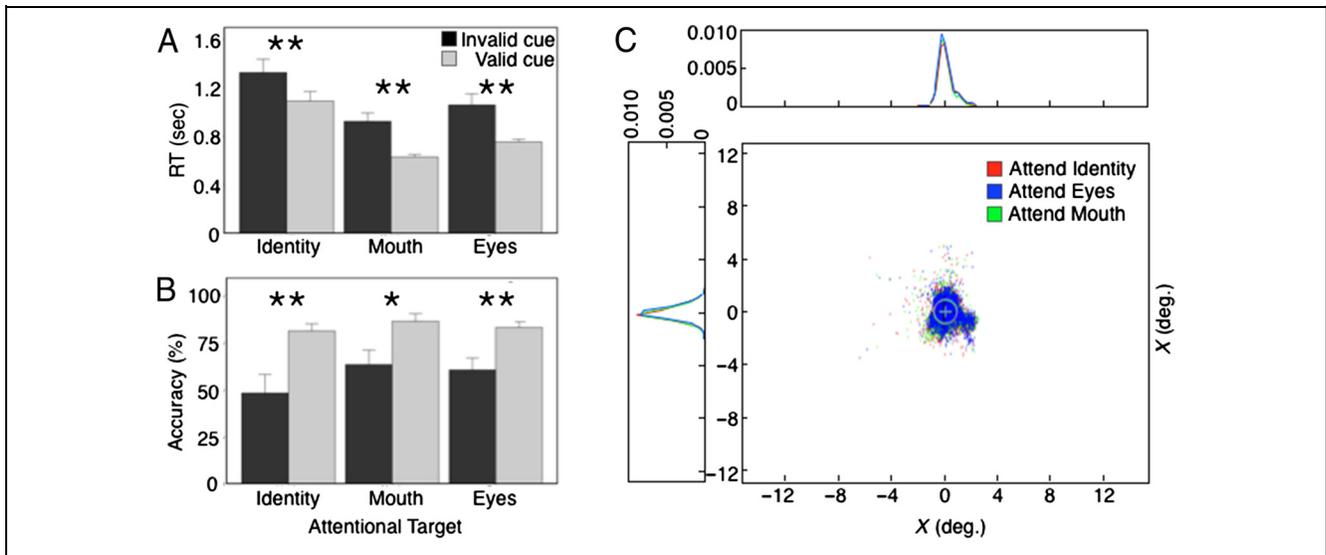
Figure 3. Behavioral results. Responses were significantly faster (A) and more accurate (B) for validly cued targets compared with invalidly cued targets. This indicates a cue validity effect for all three targets. Error bars represent the standard error of the mean. *$p < .01$, **$p < .005$. (C) Analysis of eye fixation behavior in all three experimental conditions.

of reference. Only for later illustration purposes (e.g., Figure 2), the individual participants' ROIs were transformed into Montreal Neurological Institute (MNI) space. For the subsequent spectro-analyses of the source space signals, the mean time series were extracted (average activity over all vertices within any given ROI) and subjected to a Fourier transformation.

The above-described definition of ROIs is advantageous because it is solely based on the evoked responses (i.e., the event-related field [ERF]) of the initial presentation of the first face stimulus, not on the periodic frequency-tagged activity. Therefore, such ERF-based definition is orthogonal and independent of the MEG signal we aim to analyze, namely, the frequency tags. However, the here defined ROIs may not perfectly overlap with the spatial locations at which the frequency tags are at their maximum, and it is conceivable that strong activations of neighboring regions affect the results in the selected ROIs. Therefore, we extended the ROI analyses to another set of ROIs that were selected around the most prominent activations in the steady-state responses (SSRs) in the frequency-tag spectrum of the MEG signal, namely, "FFA(SSR)," "OFA (SSR)," and "STS(SSR)."

RESULTS

Behavioral Data

After removing trials with incorrect answers (20.3%) and outliers (3.1%), the mean RTs (see Figure 3A) were subjected to a two-way repeated-measures ANOVA having two levels for the factor Cue validity (valid, invalid) and three levels for the factor Target (identity, eyes, mouth).

The main effect of Cue validity yielded an F ratio of $F(1, 9) = 20.55, p = .001$, indicating a significantly faster RT

for valid (825 msec) compared with invalid (1026 msec) trials. The main effect of Target yielded an F ratio of $F(2, 18) = 26, p < .001$. There was no significant interaction effect between target and cue validity, $F(2, 18) = 0.59, p = .56$. Tukey pairwise comparisons showed significantly slower RTs for identity ($M = 1208$ msec) compared with eye ($M = 907$ msec), $t(18) = 4.91, p < .001$, and mouth ($M = 776$ msec), $t(18) = 7.03, p < .001$, targets but no difference between eye and mouth targets, $t(18) = -2.12, p = .113$. Furthermore, pairwise comparisons showed a significant cue validity effect for all three targets. RT for identity targets was faster for valid ($M = 1092$ msec) compared with invalid (1325 msec) trials, $t(17) = 3.14, p = .005$. Valid "attend-eyes" trials ($M = 754$ msec) had faster RT than invalid trials (1059 msec), $t(17) = 4.13, p < .001$. Valid "attend-mouth" trials ($M = 629$ msec) had faster RT than invalid (924 msec) trials, $t(17) = 3.98, p < .001$.

Mean accuracy scores (see Figure 3B) were also subjected to a two-way repeated-measures ANOVA having again two levels of the factor Cue validity (valid, invalid) and three levels of the factor Target (identity, eyes, mouth). Here, the main effect of Target yielded an F ratio

Table 2. Mean MNI Coordinates from the Literature

ROI	Hemisphere	MNI (Mean)	MNI (SD)
FFA	Left	−39.5, −55.3, −21	4.2, 8.5, 4.5
FFA	Right	37.4, −53, −20.4	2.8, 6.3, 4.3
OFA	Left	−37.6, −78.4, −15.2	3.6, 4, 4.8
OFA	Right	35.1, −80, −13.8	8.4, 4.8, 5
STS	Left	−49.2, −49, 4.5	8.2, 25.1, 11.1
STS	Right	48.5, −50.6, 4.4	4.7, 19.1, 7.5

Table 3. Mean MNI Coordinates across Nine Participants

ROI	Hemisphere	MNI (Mean)	MNI (SD)	Vertices (Mean)	Vertices (SD)
FFA	Left	−31.3, −40.9, 34.9	8.7, 16.5, 6	49.6	8.7
FFA	Right	36, −20.9, 29.9	9.8, 16.7, 7.4	42.1	7.5
OFA	Left	−31.8, −61.3, 43.3	9.7, 6.4, 11	50.9	12.7
OFA	Right	34.7, −55.1, 40	5.4, 7.3, 10.5	50.9	14.4
STS	Left	−47.8, −37.7, 60.6	6, 11.2, 9	55.2	16.7
STS	Right	49.5, −31.5, 65.9	4, 15.8, 15.2	53.4	15.0

of $F(2, 18) = 4.10$, $p = .034$, indicating a significant difference in error rates between the three different types of targets. Tukey pairwise comparisons only showed significantly lower accuracy for identity (64.9%), compared with mouth (75.1%) targets, $t(18) = −2.79$, $p = .03$. The main effect of Cue validity yielded an F ratio of $F(1, 9) = 16.85$, $p = .003$, indicating significantly higher accuracy for valid (83.8%) than invalid (57.6%) trials. Again, there was no significant interaction effect between Target and Cue validity, $F(2, 18) = 2, 0$, $p = .164$. Tukey pairwise comparisons showed significantly higher accuracy scores for validly cued (81.4%) compared with invalidly cued (48.4%) identity targets, $t(14) = −4.56$, $p < .001$. In "attend-eyes" trials, accuracy was 83.2% for valid cues versus 60.8% for invalid cues, $t(14) = −3.10$, $p = .008$. In "attend-mouth" trials, valid cues resulted in 86.6% accuracy versus 63.6% for invalid cues, $t(14) = −3.17$, $p = .007$. Overall, valid cues significantly speeded up RTs and significantly increased detection accuracy in all three attentional conditions.

Next, we analyzed whether there were any systematic differences in eye-fixation behavior between the three experimental conditions (attend to eyes, mouth, or identity). Eye position samples during the stimulus period were analyzed by one-way repeated-measures ANOVAs with the factor Target (identity, eyes, mouth) for both the vertical and horizontal dimensions, respectively (see Figure 3C). The results showed that fixation behavior did not vary significantly between conditions, $F(2, 18) = 0.93$, $p > .41$, for the horizontal dimension, and $F(2, 18) = 2.18$, $p > .14$, for the vertical dimension.

MEG Data

For the MEG data, we first analyzed the ERFs in response to the initial appearance of a face stimulus in all trials. The evoked responses were mapped into source space for each participant separately, and the peak components of early (around 100 msec) and late (around 170 msec) face-specific processing stages were localized. This ERF analysis revealed activation peaks of the early component around (left OFA: $x = −30.8$, $y = −61$, $z = 42.7$; right OFA: $x = 34.9$, $y = −54.9$, $z = 39.8$; see Figure 2B, all coordinates refer to MNI space), not far from the average location of the OFA reported in previous studies (left OFA: $x = −37.6$, $y = −78.4$, $z = −15.2$; right OFA: $x = 35.1$, $y = −80$, $z = −13.8$; see Figure 2A for a review as well as Tables 3 and 4). The late components (around 170 msec; see Figure 2B) showed peak activations in both the inferior temporal (IT) cortex (left FFA: $x = −30.2$, $y = −35.5$, $z = 34$; right FFA: $x = 31.7$, $y = −21.1$, $z = 31.6$) and the superior temporal cortex (left STS: $x = −48.3$, $y = −36.8$, $z = 57$; right STS: $x = 48.2$,

Table 4. Mean MNI Coordinates for All Nine Participants

Participant	Left FFA	Right FFA	Left OFA	Right OFA	Left STS	Right STS
1	−26.5, −43.2, 38	23.5, −17.6, 39.1	−30.5, −67.7, 36.1	37.9, −56.9, 63.8	−43.2, −39.4, 68.9	46.1, −38.4, 71.4
2	−28.3, −19.5, 29.2	28.3, −25.8, 32	−45.4, −49.9, 53.7	37.1, −56.6, 34.6	−55.4, −26, 55.2	51.8, −30, 77.5
3	−17.5, −61.8, 37.5	44.3, 2.1, 21.2	−18.1, −65.8, 35.9	30.8, −50.4, 34.8	−45, −44.8, 65.9	46.9, −52.5, 48.9
4	−44.4, −51.2, 47	38.4, −42.3, 31.1	−42.5, −54.9, 66.3	26.3, −71.7. 51.3	−53.2, −35.9, 56.8	55, −19.5, 90.8
6	−37.5, −34.1, 29	33.7, −35.4, 31.2	−26.9, −57.9, 34.8	41.3, −55, 35	−47.3, −27, 62.4	49.5, −35, 53.8
7	−23.4, −57.1, 37.1	35.5, 9.1, 16.2	−35.9, −62.8, 35.3	30.4, −52.5, 35.9	−44.5, −44.7, 50.5	54.3, −21.6, 73.6
8	−30.8, −51.8, 34.8	38.8, −27.1, 27.3	−21, −64.4, 35.3	39.9, −49.2, 32.9	−37.2, −60.9, 50.5	43.6, −48.9, 67.2
9	−41.6, −34.6, 33.2	55.3, −28.7, 33.1	−40.3, −58.8, 48.8	29.8, −46.3, 35.8	−49.6, −30.6, 77.8	51.6, −1, 42
10	−31.4, −14.3, 28.2	26.3, −22.7, 38.1	−25.8, −69.4, 43.6	39.1, −57.2, 36.1	−54.7, −30.3, 57.5	46.2, −37, 68.3

Figure 4. Cortical maps of the SSR and general power distribution of the MEG minimum-norm estimates in the tagging frequency range. (A) The tagging frequency of the identity tag (0.66 Hz) was most prominently entrained in all inferior temporal cortex. The tagging frequency of the mouth tag (B) and eye tag (C) entrained neural populations in occipital and occipito-lateral cortices. The maps are an average across all participants; the colored lines represent the overall area of the three ROIs over all participants (red = FFA, green = OFA, and blue = STS).
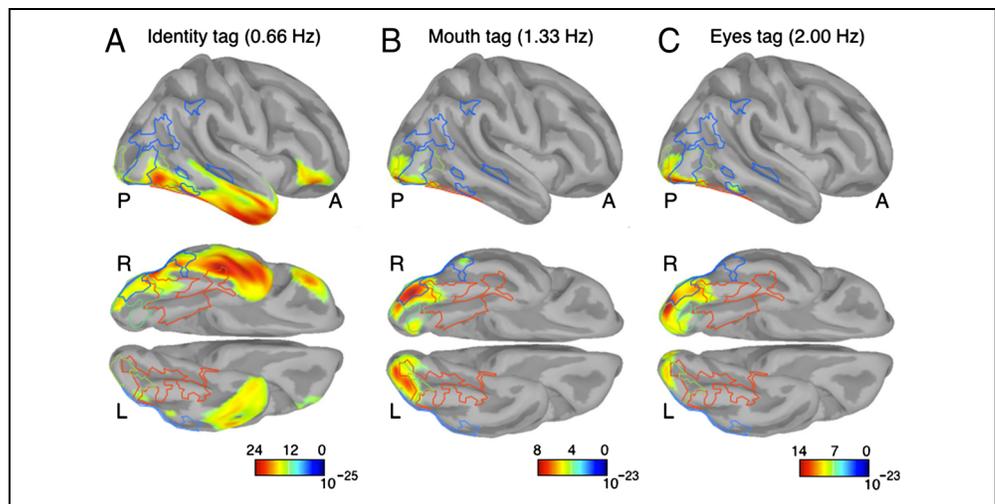
$y = -34.2, z = 58.2$; see Figure 2B), in close vicinity of the average localization of previous fMRI-based reports (left FFA: $x = -39.5$, $y = -55.3$, $z = -21$; right FFA: $x = 37.4$, $y = -53$, $z = -20.4$; left STS: $x = -49.2$, $y = -49$, $z = 4.5$; right STS: $x = 48.5$, $y = -50.6$, $z = 4.4$; see Figure 2A and Table 2 for a review). Hence, our ERF source-space analysis proofed to spatially replicate previous fMRI localization standards well, and we therefore used the results to define within-participant ROIs in the OFA, STS, and FFA, respectively.

In general, the stimulus frequencies entrained distinct neural populations in occipital and IT cortex (see Figure 4). Whereas the periodic updating of the mouth and eye parts of the face stimulus activated mostly posterior IT cortex and lateral occipital regions (Figure 4B and C), the rhythmic changes in face identity entrained, on average, more anterior areas of IT cortex (Figure 4A).

In the following, we Fourier-transformed the time series within the individually defined ROIs for each participant separately. The resulting power estimates clearly showed three distinct peaks in the spectrum, corresponding to the three respective presentation frequencies in our frequency-tagging paradigm (peaks at 0.66, 1.33, and 2.00 Hz; see Figure 5A–C). Figure 5 shows the entrainment over a wide range of frequencies. All dominant modulations of the spectrum were at the first harmonics of the presentation frequencies, with only very little modulation at higher harmonics.

To compensate for the $1/F$ characteristic of the spectrum, we applied normalization by the baseline spectrum (before stimulus onset) and extracted the individual participants' peaks in the power spectra at the three presentation frequencies (0.66, 1.33, and 2.00 Hz). Figure 6 shows the entrainment with the three stimulation frequencies in the various ROIs, both when top–down attention was directed to the respective stimulus part ("Attend IN") or somewhere else ("Attend OUT"). The identity tag (0.66 Hz) was strongly entrained only in the individual participants' FFAs, not in the OFA or STS

(Figure 6A). The activation of the mouth tag (1.33 Hz) was most strongly picked up in OFA and less so in FFA or STS. The activation of the eye tag, finally, was more equally distributed among the three areas with slightly stronger activation of FFA and STS compared with OFA (Figure 6C).
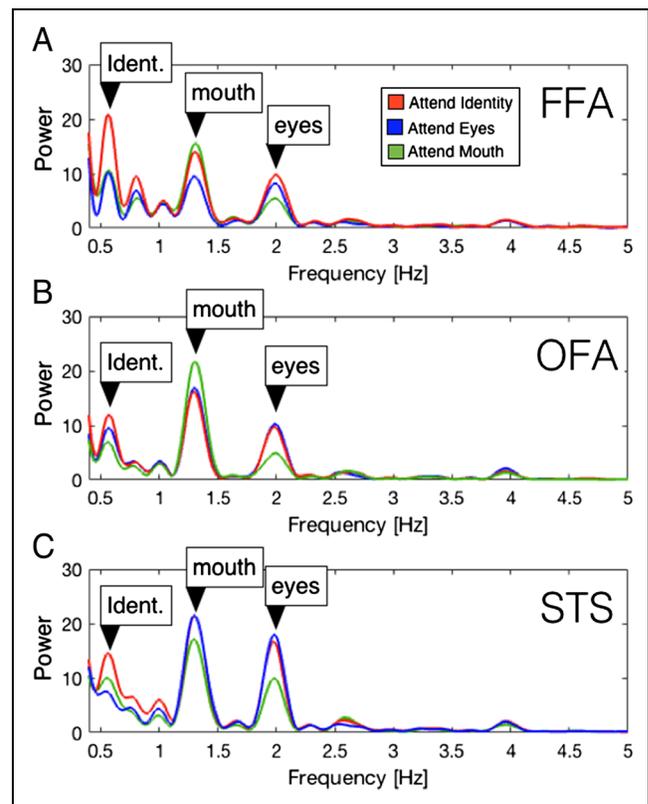

Figure 5. SSR of the MEG power of the minimum-norm estimates for each ROI and their attentional modulation in the three experimental conditions (red: "attend identity"; blue: "attend eyes"; green: "attend mouth"). The spectrograms of FFA (A), OFA (B), and STS (C) were dominated by the frequency tags of the identity (at 0.66 Hz), mouth (at 1.33 Hz), and eyes (at 2 Hz).
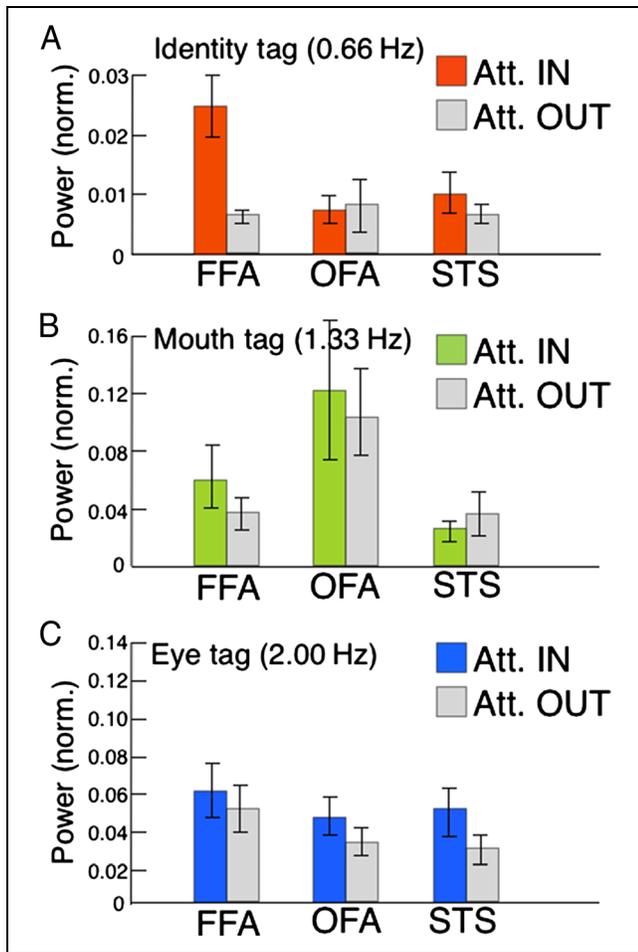
Figure 6. Overall entrainment of the ROIs by the tagging frequencies. Normalized power of the minimum-norm estimate at the tagging frequencies in the three ROIs, both when top–down attention was deployed to the respectively tagged stimulus component ("Attend IN," colored bars) and when not ("Attend OUT," gray bars). Error bars represent the standard error of the mean.

For statistical analyses, the power estimates at the tagging frequency bands were first normalized and then subjected to a four-way repeated-measures ANOVA having three levels for the factor ROI (OFA, FFA, STS), three levels for the factor Condition (attend identity, mouth, eyes), three levels for the factor Tagging frequency (0.66 Hz, 1.33 Hz, 2.00 Hz), and two levels for the factor Hemisphere (left, right; see Table 5). There was a significant main effect of Tagging frequency, $F(2, 16) = 10.9$, $p = .001$, but not ROI, $F(2, 16) = 1.56$, $p = .240$, Condition, $F(2, 16) = 0.78$, $p = .476$, or Hemisphere, $F(1, 8) = 0.89$, $p = .372$. There were significant interaction effects between ROI and Condition, $F(4, 32) = 3.79$, $p = .012$; Condition and Tagging frequency, $F(4, 32) = 4.83$, $p = .004$; and ROI, Condition, and Tagging frequency, $F(8, 64) = 2.42$, $p = .024$. None of the effects that included the factor Hemisphere reached significance (see Table 5), and therefore, power was averaged across hemispheres for all subsequent analyses.

To test whether these global trends were because of specific differences between conditions according to our hypotheses, we completed planned contrasts between the most important conditions in each ROI. In the FFA, pairwise comparisons (see Table 6, all with false discovery rate-adjusted p values) only showed a significant attention effect for identity (tagged at 0.66 Hz) compared with the eyes, $t(8) = 3.92$, $p = .018$, and mouth, $t(8) = 4.03$, $p = .018$, conditions. In the OFA, there was a significant attention effect for mouth (tagged at 1.33 Hz) compared with the eyes, $t(8) = 3.57$, $p = .022$, condition, and an attention effect for eyes (tagged at 2 Hz) compared with the mouth, $t(8) = -4.35$, $p = .018$, condition. In the STS (Table 6), there was a significant effect for eyes (tagged at 2 Hz) compared with the mouth, $t(8) = 3.24$, $p = .029$, condition.

Our ERF-based definition of the ROIs has the advantage that it is not based on the periodic MEG activations of the frequency tags themselves and therefore it is independent of the signal to be analyzed. However, the defined ROIs did not overlap completely with the observed peaks of the SSRs in the frequency-tag range (as

Table 5. Repeated-Measures ANOVA: MEG Power

Effect	F Ratio	df	p
ROI	1.56	2, 16	.240
Condition	0.78	2, 16	.476
Frequency	10.90	2, 16	.001**
Hemisphere	0.89	1, 8	.372
ROI × Condition	3.79	4, 32	.012*
ROI × Frequency	2.25	4, 32	.085
Condition × Frequency	4.83	4, 32	.004**
ROI × Hemisphere	1.91	2, 16	.180
Condition × Hemisphere	1.65	2, 16	.224
Frequency × Hemisphere	3.39	2, 16	.059
ROI × Condition × Frequency	2.42	8, 64	.024*
ROI × Condition × Hemisphere	0.06	4, 32	.993
ROI × Frequency × Hemisphere	1.00	4, 32	.420
Condition × Frequency × Hemisphere	1.05	4, 32	.399
ROI × Condition × Frequency × Hemisphere	0.34	8, 64	.947

* $p < .05$.

** $p < .01$.

Table 6. Pairwise Comparisons: MEG Power

| ROI	Frequency	Attend	df	t	p	p (FDR)
OFA	0.66 Hz	Mouth vs. identity	8	−0.56	.594	.648
OFA	0.66 Hz	Eyes vs. identity	8	0.41	.693	.693
OFA	1.33 Hz	Mouth vs. eyes	8	3.57	.007	.022*
OFA	2 Hz	Mouth vs. eyes	8	−4.35	.002	.018*
FFA	0.66 Hz	Identity vs. mouth	8	4.03	.004	.018*
FFA	0.66 Hz	Identity vs. eyes	8	3.92	.004	.018*
FFA	1.33 Hz	Mouth vs. eyes	8	2.26	.053	.107
FFA	2 Hz	Mouth vs. eyes	8	−1.98	.083	.142
STS	0.66 Hz	Eyes vs. identity	8	−1.84	.103	.154
STS	0.66 Hz	Mouth vs. identity	8	−0.89	.399	.479
STS	1.33 Hz	Eyes vs. mouth	8	1.36	.211	.281
STS	2 Hz	Eyes vs. mouth	8	3.24	.012	.029*

FDR = false discovery rate.

* $p < .05$.
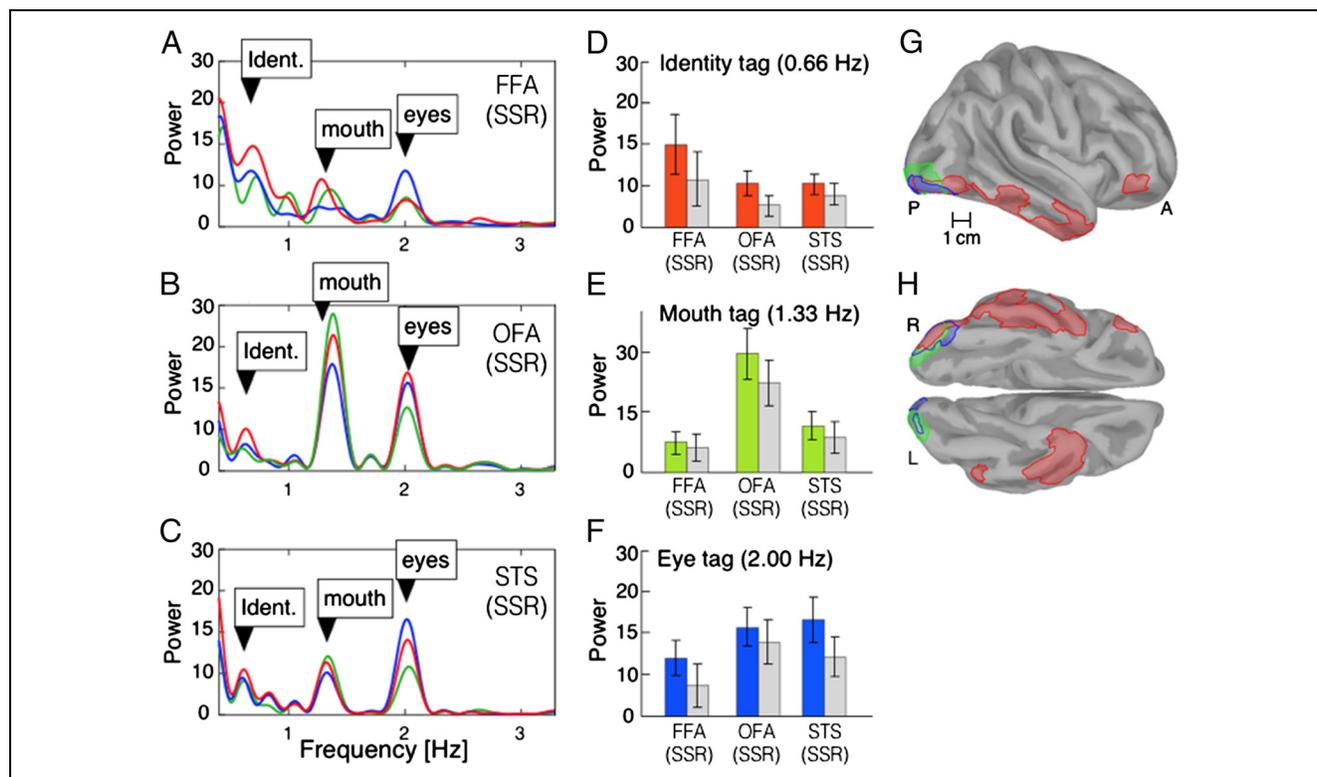
Figure 7. Analysis of MEG power in a set of control ROIs based on the peak activations of the SSRs. Mean MEG power of the minimum-norm estimates is shown for each ROI and their attentional modulation in the three experimental conditions (red: "attend identity"; blue: "attend eyes"; green: "attend mouth"). The spectrograms of FFA(SSR; A), OFA(SSR; B), and STS(SSR; C) were dominated by the frequency tags of the identity (at 0.66 Hz), mouth (at 1.33 Hz), and eyes (at 2 Hz). (D–F) Power of the minimum-norm estimate at the tagging frequencies in the three ROIs, both when top–down attention was deployed to the respectively tagged stimulus component ("Attend IN," colored bars) and when not ("Attend OUT," gray bars). Error bars represent the standard error of the mean. G and H show the alternative ROIs based on MEG peak activations of the SSRs in the tagging frequencies: FFA(SSR) in red, OFA(SSR) in green, and STS(SSR) in blue.

shown in Figure 4), and the activations of neighboring regions could have affected the results in the selected ROIs. Therefore, we repeated our ROI analyses in another set of ROIs that were selected directly on the basis of the most prominent activations of the SSRs in the frequency-tag spectrum (see spatial locations of peaks in Figure 4). As can be seen from Figure 7, the results obtained from these alternatives, that is, SSR-based ROIs, are qualitatively congruent with the previous analysis. Figure 7A–C shows the entrainment over the respective range of frequencies with three peaks at the three respective presentation frequencies. In the now peak-based ROIs, the identity tag (0.66 Hz) was again more strongly entrained in the individual participants' FFA(SSR), than in OFA(SSR) or STS (SSR; Figure 7D), particularly when top–down attention was directed to the respective stimulus part ("Attend IN" in red). Furthermore, the activation of the mouth tag (1.33 Hz) was most strongly picked up in the OFA(SSR) and less so in the FFA(SSR) or STS(SSR; Figure 7E), and the activation of the eye tag was strongest in the STS (SSR; Figure 7F).

DISCUSSION

The aim of this study was to investigate the functional specialization of the areas OFA, FFA, and STS by demonstrating the differential entrainment with the respective tagging frequencies of a face compound stimulus as well as to investigate the attentional modulation of the related neural activations at those specific frequencies. On a behavioral level, reflecting the cue validity effect, we expected faster RTs and higher accuracy rates for targets in validly cued trials compared with invalidly cued trials. On a neural level, specific response enhancements were expected reflecting the functional specialization of the three ROIs: When covertly attending to the eyes, an increased neural response was expected in the STS. In the OFA, enhanced responses were expected when attention was directed toward either the eyes or the mouth (i.e., all face parts); and in the FFA, when attending to face identity.

The behavioral results clearly showed the hypothesized cue validity effect. Faster and more accurate responses to targets after a valid cue indicated a significant facilitation effect by top–down attention on task performance. Similar endogenous cueing effects were observed behaviorally in tasks based on the Posner cueing paradigm (Baldauf, 2015, 2018; Bagherzadeh, Baldauf, Lu, Pantazis, & Desimone, 2017; Moore & Zirnsak, 2017; Voytek et al., 2017; Baldauf & Desimone, 2016; Baldauf, Grossman, Hu, Boyden, & Desimone, 2016; Baldauf & Deubel, 2009; Mangun & Hillyard, 1991; Posner, 1980). In a prototypical Posner cueing paradigm, participants are instructed to fixate at a central point on the screen and to attend covertly to either side of the fixation point to detect the temporal onset of a target stimulus. There are corresponding variants of the Posner cueing paradigm

for other, nonspatial attentional sets such as visual features (Störmer & Alvarez, 2014; Andersen, Fuchs, & Müller, 2011; Zhang & Luck, 2009; Liu, Stevens, & Carrasco, 2007; Maunsell & Treue, 2006; Müller et al., 2006; Hopf, Boelmans, Schoenfeld, Luck, & Heinze, 2004; Saenz, Buracas, & Boynton, 2003) and objects (Marinato & Baldauf, 2019; Kim, Tsai, Ojemann, & Verghese, 2017; Zhang, Mlynaryk, Japee, & Ungerleider, 2017; Liu, 2016; Baldauf & Desimone, 2014), all of which exhibit reliable attentional facilitation effects, that is, cue validity effects. The robust finding of such "cue validity effects" in our study indicates that attention was indeed covertly oriented to the cued aspects of the face stimulus. It is also noteworthy that the strongest cueing effects were found for identity as the attentional target, mostly because of the comparably low accuracy in invalid trials. This may reflect the fact that the representation and discrimination of a face's identity are presumably more complex than discriminating local features like the form of the mouth, and therefore in this condition, the target detection task is the hardest. In addition, the complex processes underlying the full representation of a face's identity may be more vulnerable, and consequently, the discrimination of identities may fail if attention is not directed to this dimension.

Much of our observed top–down attention is most likely based on the preferential processing of the respective visual features and object parts of the face stimuli. However, because of the inherent spatial arrangement of the facial components, such as eyes and mouth, within a human face, it is technically impossible to fully exclude contributions of spatial attention. We put extra care in experimentally minimizing such contributions of spatial attention in the first place by (a) presenting the relatively small face stimuli at the retina (which minimizes contributions of spatial attention network; see Baldauf & Desimone, 2014) and (b) instructing our participants to keep strict eye fixation at the central fixation cross. In addition, all participants' eye movement behavior was monitored with the highest possible accuracy standards (MEG-compatible binocular tracking at a 1-kHz sampling rate). In the following, our analysis pipeline disregarded any trials, in which saccades occurred, from the further analyses of the MEG data. However, there could still have existed a possibility of subtle but potentially systematic differences in fixation behavior between the various experimental conditions. For example, it could have been the case that, in the "attend eyes" condition, participants would have systematically tended to fixate slightly above the fixation cross—or below the fixation cross in the "attend mouth" condition. However, this was not the case. As our analysis of sample-by-sample eye position data revealed, there was no such systematic difference between conditions, in neither the vertical nor horizontal dimension (see Figure 3C). Therefore, there were no differences observed between the experimental conditions at least in terms of open spatial attention. If spatial

attention contributed to the task performance at all, it may have done so only in the form of covert spatial attention.

To investigate the attentional modulation on a neural level, we employed a cyclic entrainment paradigm (Lithari, Sánchez-García, Ruhnau, & Weisz, 2016; Norcia, Appelbaum, Ales, Cottereau, & Rossion, 2015; Baldauf & Desimone, 2014; Kaspar, Hassler, Martens, Trujillo-Barreto, & Gruber, 2010; Lakatos, Karmos, Mehta, Ulbert, & Schroeder, 2008; Parkkonen, Andersson, Hämäläinen, & Hari, 2008; Appelbaum, Wade, Vildavski, Pettet, & Norcia, 2006; Müller, Malinowski, Gruber, & Hillyard, 2003), in which the periodic modulations of certain parts of the visual stimulus generate electrophysiological responses with the same rhythmic modulation (see Regan, 1966). Such periodic modulations are strongest in brain areas that are tuned for the specific topographic location and/or a specific feature of the frequency-tagged stimulus. Here, we presented participants with compound face stimuli containing three different frequency tags (identity at 0.66 Hz, mouth at 1.33 Hz, eyes at 2 Hz) and modulated top–down attention to any of the three respective facial properties. By keeping visual stimulation constant, while modulating attention, we aimed to use the neural signatures of the attended stimulus to differentiate between brain areas with specialized processing.

In a recent MEG study, a similar frequency-tagging approach was used to study nonspatial, object-based attention (Baldauf & Desimone, 2014). Participants were presented with spatially overlapping face and house images that oscillated in visibility at different frequencies (1.5 and 2.0 Hz). A frequency analysis showed an enhanced response in the FFA (at 1.5 Hz) when attending to face images and an enhanced response in the parahippocampal place area (at 2.0 Hz) when attending to house images. In this study, in contrast, different frequency tags were not assigned to different objects, superimposed at the same spatial location. Rather, here, we implemented different frequency tags for subparts and/or aspects within the same face object. In other studies on face perception, using EEG recordings, frequency tagging has also been applied to images of whole faces while periodically altering some aspect of those faces, such as their identity, emotional expression, or orientation: Alonso-Prieto and colleagues, for example, systematically investigated different rhythmic presentation schedules (in a wide range of 1–16 Hz) for identical versus nonidentical faces (Alonso-Prieto et al., 2013; see also Rossion, 2014; Rossion & Boremanse, 2011). Recently, Zhu and colleagues periodically updated the facial expressions of a face stimulus at presentation frequencies in a range of 2–8 Hz and found increased EEG activity over the occipital-temporal electrodes at 5 Hz, compared with a baseline condition with no updated facial expressions (Zhu, Alonso-Prieto, Handy, & Barton, 2016; see also Mayes, Pipingas, Silberstein, & Johnston, 2009). A recent study by Boremanse and colleagues frequency-tagged the left and right sides of a face with different frequencies trying to dissociate between part-based and integrated EEG responses to faces (Boremanse, Norcia, & Rossion, 2014). However, none of the aforementioned EEG studies tried to target specific brain areas by applying inverse models to reconstruct the neural sources of the respectively observed frequency-tagged electrode activations. Furthermore, to our knowledge, no studies have used different frequency tags in combination with compound face stimuli to directly study the relative processing hierarchies of face parts, eye gaze, and facial identity in such an MRI-guided MEG paradigm.

To analyze the spectral modulations in the various subparts of the human face processing network, as well as the attentional modulation of those locally entrained frequency tags, we first identified the respective network components (ROIs) functionally with independent localizers. These functional localizers were also based on the MEG recordings, but instead of analyzing the rhythmic modulation of the signal during the complete, highly repetitive stimulus period, the functional localizers were solely based on the evoked magnetic fields (ERFs) in response to the very first appearance of a face stimulus at the beginning of each trial. The analysis of the ERFs in response to the initial presentation of a face stimulus revealed peak activations in occipital, IT, and superior temporal cortices, at systematically prolonged latencies, respectively. This provides further evidence that the time courses of the evoked responses in MEG contain rich information about the temporal sequence of various processing steps in high-level visual cortex (Isik, Meyers, Leibo, & Poggio, 2014) and that they can be successfully source-localized also for relatively deep cortical structures of origin (see Hadjikhani, Kveraga, Naik, & Ahlfors, 2009). Furthermore, these source localization results replicated functional nodes previously established in fMRI, such as OFA, STS, and FFA (Gao et al., 2013; Dalrymple et al., 2011; Sadeh et al., 2010; Deffke et al., 2007; Pitcher et al., 2007; Itier & Taylor, 2004; Hoshiyama et al., 2003; Halgren et al., 2000) and proved trustworthy to be used as a within-participant approach for functionally determining respective ROIs—independently of the experimentally frequency-modulated neural signature (steady-state visually evoked potentials) during later periods of stimulus presentation. Because we defined the ROIs for our main analyses independently of the frequency modulations, there was no complete overlap with the peaks of the frequency tags. Therefore, the effects we observed could have been affected by interference from nearby areas that responded more strongly to the presentation frequencies, given that the distributed MEG source estimates typically have a substantial spatial spread. We therefore repeated our analyses on a second set of ROIs that were directly selected from the activation peaks of the SSRs in the frequency tags themselves. These control analyses revealed very similar results with a congruent pattern of rhythmic entrainment in the various ROIs, namely, FFA(SSR), OFA (SSR), and STS(SSR).

The then following spectro-analyses of the time series in all three ROIs within the occipital-temporal face processing network showed that it is possible to entrain several functionally specialized neural populations with the presentation rhythms of respectively relevant information. The advantages of such a frequency-tagging approach are the increased signal-to-noise ratios because of the rapid and regular repetition of the stimuli (see Regan, 1966).

When attending to facial identity, the respective tag of identity changes was only picked up in FFA and in none of the other two regions of the face processing network. In addition, the neural populations in FFA showed a significant and highly selective modulation by top–down attention, at the identity tagging frequency range, in the sense that the identity tag only showed up in the spectrogram of FFA if attention was deployed to the identity of the stimulus face. Both the neural responses and signal-to-noise ratios were higher when attention was directed toward identity compared with eyes or mouth. Together, this indicates that the FFA is specialized in processing facial identity, which is in line with previous research findings (Afraz et al., 2006, 2015; Turk et al., 2005; Grill-Spector et al., 2004; Liu et al., 2002). Although the FFA was also strongly driven by the rhythmic changes of face parts, there was no attentional modulation of those. These results suggest that the core function of FFA indeed is the processing of facial identity and that the rhythmic updating of face parts also coactivates it, presumably because the respective face parts need to be combined and integrated into a representation of an identity. The FFA therefore seems to consist of neural populations capable of integrating face parts into a face gestalt (see also Harris & Aguirre, 2010).

In the OFA, on the other hand, there were significant responses and attention effects at the tagging frequencies of facial parts, such as mouth and eyes, but not at all at the identity frequency tag. This means that the OFA may not be directly involved in the representation of facial identity per se but may operate on an earlier level analyzing individual parts of the face object separately. This is in line with converging evidence that describe the OFA as an early node in the face processing network that represents face parts and that more complex facial features are subsequently processed in the FFA and STS (for a review, see Pitcher et al., 2011). Interestingly though, also attention to the facial identity boosted the processing of facial parts at the respective tagging frequencies. This is in line with the finding that top–down attention to a grouped object configuration can be fed back to hierarchically lower processing stages (Roelfsema, Lamme, & Spekreijse, 2004; Lamme, Supèr, Landman, Roelfsema, & Spekreijse, 2000), which are concerned with the processing of its features (Schoenfeld et al., 2014).

The STS was most responsive to the rhythmic modulation of the eye region and exhibited also the strongest top–down attentional modulation at this rhythm. These results confirm previous reports that the STS region is

closely related to the processing of the eye gaze in human face stimuli (see, e.g., Carlin & Calder, 2013; Carlin et al., 2011). Interestingly, the presentation rhythm of the eyes (2.00 Hz) was also significantly enhanced within the STS when participants were instructed to attend to the identity of the face stimuli, indicating the prominent role the eye region plays for determining a person's identity.

Interestingly, our results both from the ERF analysis and from the steady-state visually evoked potential analysis showed no significant difference between face-selective or face-part-selective activity in the left versus right hemisphere. This is surprising given the well-known right hemispheric dominance for face perception reported in both fMRI (e.g., Kanwisher et al., 1997; Sergent, Ohta, & Macdonald, 1992) and EEG (e.g., Bentin, Allison, Puce, Perez, & McCarthy, 1996) studies. However, our results are congruent with previous reports (Baldauf & Desimone, 2014) that showed, also in a frequency tagging paradigm, more balanced involvement of both hemispheres in occipital and IT regions. Taken together, whether or not the difference in activity between the left and right hemispheres reaches significance may depend on the stimulation protocol used and the respective signal-to-noise ratio.

In general, our MEG results are consistent with the view of an hierarchical organization of the three subnetworks: The network populations at an earlier level, such as OFA and, in part, also STS, are preferentially concerned with the analysis of crucial features or facial subcomponents and seem to feed those representations of object parts forward to higher level face processing networks in IT cortex, for example, FFA, where the information about facial features and subcomponents is integrated into a representation of facial identity. This relative arrangement in a hierarchical face processing network is also supported by the telling differences in response latencies both in our ERF results, as in previous reports (Liu et al., 2002). A limitation of our current experimental design is that the stimulation frequencies were not counterbalanced across conditions, simply because a fully counterbalanced design would not have been possible to record within the same MEG session because of the length of the individual trials (7 sec) and the amount of trials needed for stable averaging within each condition and frequency assignment. However, we do not believe that our results are affected by this choice. Although one might suspect that there are biases for slower stimulation rhythms to strongly entrain deep structures such as FFA, our data speak against such an explanation. In the spectrograms of Figure 7, it can be seen that the power of the respective frequency is only modulated by the state of attention and stimulus preference: in FFA(SSR; Figure 7A); for example, the frequencies of nonpreferred stimuli like mouth and eyes at 1.33 and 2.00 Hz, respectively, are equally strong. Therefore, there is no general bias in our data for slower or faster

frequencies to be entrained more (or less) in deeper or more superficial structures. Rather, the strength of entrainment reflects stimulus preference and attentional state.

In addition, the neural effects of top–down attention on various parts of this face processing network provided confirming evidence for its hierarchical organization: As previously described for more simplistic visual stimuli such as color patches or line segments, we also find in our current results that top–down attention to the higher level representation in FFA can lead to a spread or co-activation of lower level representations concerned with relevant facial features. For STS, the situation in this hierarchical organization seems to be more complex. With its preferential encoding of the eye's region (i.e., gaze), it still contributes crucial information about facial parts, which are then also integrated into the facial identity at the next processing level of FFA. Also for the representation of eye features in STS, we observed top–down coactivation—presumable channeled downstream through FFA—when attention was deployed to the facial identity. However, STS showed slightly longer response latencies in the ERF results, in comparison with OFA, and therefore seems to be at a slightly later processing stage. This is also consistent with the idea of STS being a comparably high-level representation in which encoding of the eye gaze is directly used by networks for social cognition.

Our results showed the preferential representation of different aspects and dimensions of a single face object in various face processing areas in occipital and IT cortex. Top–down attention resulted in an attentional weighting of the respective visual processing, at various stages of the involved visual processing hierarchy.

Acknowledgments

REFERENCES

Afraz, A., Boyden, E. S., & DiCarlo, J. J. (2015). Optogenetic and pharmacological suppression of spatial clusters of face neurons reveal their causal role in face gender discrimination. *Proceedings of the National Academy of Sciences, U.S.A.*, 112, 6730–6735.

Afraz, S.-R., Kiani, R., & Esteky, H. (2006). Microstimulation of inferotemporal cortex influences face categorization. *Nature*, 442, 692–695.

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

Andersen, S. K., Fuchs, S., & Müller, M. M. (2011). Effects of feature-selective and spatial attention at different stages of visual processing. *Journal of Cognitive Neuroscience*, 23, 238–246.

Appelbaum, L. G., Wade, A. R., Vildavski, V. Y., Pettet, M. W., & Norcia, A. M. (2006). Cue-invariant networks for figure and background processing in human visual cortex. *Journal of Neuroscience*, 26, 11695–11708.

Arcurio, L. R., Gold, J. M., & James, T. W. (2012). The response of face-selective cortex with single face parts and part combinations. *Neuropsychologia*, 50, 2454–2459.

Bagherzadeh, Y., Baldauf, D., Lu, B., Pantazis, D., & Desimone, R. (2017). Alpha and gamma neurofeedback reinforce control of spatial attention. *Journal of Vision*, 17, 385.

Baldauf, D. (2015). Top–down biasing signals of non-spatial, object-based attention. *Journal of Vision*, 15, 1395.

Baldauf, D. (2018). Visual selection of the future reach path in obstacle avoidance. *Journal of Cognitive Neuroscience*, 30, 1846–1857.

Baldauf, D., Cui, H., & Andersen, R. A. (2008). The posterior parietal cortex encodes in parallel both goals for double-reach sequences. *Journal of Neuroscience*, 28, 10081–10089.

Baldauf, D., & Desimone, R. (2014). Neural mechanisms of object-based attention. *Science*, 344, 424–427.

Baldauf, D., & Desimone, R. (2016). Mechanisms of spatial versus non-spatial, modality-based attention. In *Annual Meeting of the Society for Neuroscience*. San Diego, CA.

Baldauf, D., & Deubel, H. (2008). Properties of attentional selection during the preparation of sequential saccades. *Experimental Brain Research*, 184, 411–425.

Baldauf, D., & Deubel, H. (2009). Attentional selection of multiple goal positions before rapid hand movement sequences: An event-related potential study. *Journal of Cognitive Neuroscience*, 21, 18–29.

Baldauf, D., & Deubel, H. (2010). Attentional landscapes in reaching and grasping. *Vision Research*, 50, 999–1013.

Baldauf, D., Grossman, N., Hu, A.-M., Boyden, E., & Desimone, R. (2016). Transcranial alternating current stimulation (tACS) reveals causal role of brain oscillations in visual attention. *Journal of Vision*, 16, 937.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, 8, 551–565.

Bichot, N. P., Heard, M. T., DeGennaro, E. M., & Desimone, R. (2015). A source for feature-based attention in the prefrontal cortex. *Neuron*, 88, 832–844.

Bichot, N. P., Rossi, A. F., & Desimone, R. (2005). Parallel and serial neural mechanisms for visual search in macaque area V4. *Science*, 308, 529–534.

Boremanse, A., Norcia, A. M., & Rossion, B. (2014). Dissociation of part-based and integrated neural responses to faces by means of electroencephalographic frequency tagging. *European Journal of Neuroscience*, 40, 2987–2997.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433–436.

Caharel, S., d'Arripe, O., Ramon, M., Jacques, C., & Rossion, B. (2009). Early adaptation to repeated unfamiliar faces across viewpoint changes in the right hemisphere: Evidence from the N170 ERP component. *Neuropsychologia*, 47, 639–643.

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience*, 6, 641–651.

Carlin, J. D., & Calder, A. J. (2013). The neural basis of eye gaze processing. *Current Opinion in Neurobiology*, 23, 450–455.

Carlin, J. D., Calder, A. J., Kriegeskorte, N., Nili, H., & Rowe, J. B. (2011). A head view-invariant representation of gaze direction in anterior superior temporal sulcus. *Current Biology*, *21*, 1817–1821.

Ciaramitaro, V. M., Mitchell, J. F., Stoner, G. R., Reynolds, J. H., & Boynton, G. M. (2011). Object-based attention to one of two superimposed surfaces alters responses in human early visual cortex. *Journal of Neurophysiology*, *105*, 1258–1265.

Cohen, E. H., & Tong, F. (2015). Neural mechanisms of object-based attention. *Cerebral Cortex*, *25*, 1080–1092.

Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., et al. (1998). A common network of functional areas for attention and eye movements. *Neuron*, *21*, 761–773.

Corbetta, M., Tansy, A. P., Stanley, C. M., Astafiev, S. V., Snyder, A. Z., & Shulman, G. L. (2005). A functional MRI study of preparatory signals for spatial location and objects. *Neuropsychologia*, *43*, 2041–2056.

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage*, *9*, 179–194.

Dalrymple, K. A., Oruç, I., Duchaine, B., Pancaroglu, R., Fox, C. J., Iaria, G., et al. (2011). The anatomic basis of the right face-selective N170 IN acquired prosopagnosia: A combined ERP/fMRI study. *Neuropsychologia*, *49*, 2553–2563.

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.

Deffke, I., Sander, T., Heidenreich, J., Sommer, W., Curio, G., Trahms, L., et al. (2007). MEG/EEG sources of the 170-ms response to faces are co-localized in the fusiform gyrus. *Neuroimage*, *35*, 1495–1501.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, *113*, 501–517.

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General*, *123*, 161–177.

Fairhall, S. L., & Ishai, A. (2007). Effective connectivity within the distributed cortical network for face perception. *Cerebral Cortex*, *17*, 2400–2406.

Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis: II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage*, *9*, 195–207.

Freiwald, W. A., & Tsao, D. Y. (2010). Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science*, *330*, 845–851.

Gao, Z., Goldstein, A., Harpaz, Y., Hansel, M., Zion-Golumbic, E., & Bentin, S. (2013). A magnetoencephalographic study of face processing: M170, gamma-band oscillations and source localization. *Human Brain Mapping*, *34*, 1783–1795.

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

Giesbrecht, B., Woldorff, M. G., Song, A. W., & Mangun, G. R. (2003). Neural mechanisms of top–down control during spatial and feature attention. *Neuroimage*, *19*, 496–512.

Gregoriou, G. G., Gotts, S. J., Zhou, H., & Desimone, R. (2009). High-frequency, long-range coupling between prefrontal and visual cortex during attention. *Science*, *324*, 1207–1210.

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience*, *7*, 555–562.

Gross, J., Baillet, S., Barnes, G. R., Henson, R. N., Hillebrand, A., Jensen, O., et al. (2013). Good practice for conducting and reporting MEG research. *Neuroimage*, *65*, 349–363.

Hadjikhani, N., Kveraga, K., Naik, P., & Ahlfors, S. P. (2009). Early (N170) activation of face-specific cortex by face-like objects. *NeuroReport*, *20*, 403–407.

Halgren, E., Raij, T., Marinkovic, K., Jousmäki, V., & Hari, R. (2000). Cognitive response profile of the human fusiform face area as determined by MEG. *Cerebral Cortex*, *10*, 69–81.

Hämäläinen, M. S., & Ilmoniemi, R. J. (1994). Interpreting magnetic fields of the brain: Minimum norm estimates. *Medical & Biological Engineering & Computing*, *32*, 35–42.

Harris, A., & Aguirre, G. K. (2010). Neural tuning for face wholes and parts in human fusiform gyrus revealed by fMRI adaptation. *Journal of Neurophysiology*, *104*, 336–345.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*, 2425–2430.

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.

Heisz, J. J., Watter, S., & Shedden, J. M. (2006). Automatic face identity encoding at the N170. *Vision Research*, *46*, 4604–4614.

Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience*, *3*, 80–84.

Hopf, J.-M., Boelmans, K., Schoenfeld, M. A., Luck, S. J., & Heinze, H.-J. (2004). Attention to features precedes attention to locations in visual search: Evidence from electromagnetic brain responses in humans. *Journal of Neuroscience*, *24*, 1822–1832.

Hoshiyama, M., Kakigi, R., Watanabe, S., Miki, K., & Takeshima, Y. (2003). Brain responses for the subconscious recognition of faces. *Neuroscience Research*, *46*, 435–442.

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

Issa, E. B., Papanastassiou, A. M., & DiCarlo, J. J. (2013). Large-scale, high-resolution neurophysiological maps underlying fMRI of macaque temporal lobe. *Journal of Neuroscience*, *33*, 15207–15219.

Itier, R. J., Alain, C., Sedore, K., & McIntosh, A. R. (2007). Early face processing specificity: It's in the eyes! *Journal of Cognitive Neuroscience*, *19*, 1815–1826.

Itier, R. J., & Taylor, M. J. (2004). Source analysis of the N170 to faces and objects. *NeuroReport*, *15*, 1261–1265.

Jacques, C., & Rossion, B. (2006). The speed of individual face categorization. *Psychological Science*, *17*, 485–492.

Jiang, F., Dricot, L., Weber, J., Righi, G., Tarr, M. J., Goebel, R., et al. (2011). Face categorization in visual scenes may start in a higher order area of the right fusiform gyrus: Evidence from dynamic visual stimulation in neuroimaging. *Journal of Neurophysiology*, *106*, 2720–2736.

Jonas, J., Descoins, M., Koessler, L., Colnat-Coulbois, S., Sauvée, M., Guye, M., et al. (2012). Focal electrical intracerebral stimulation of a face-sensitive area causes transient prosopagnosia. *Neuroscience*, *222*, 281–288.

Jonas, J., Rossion, B., Krieg, J., Koessler, L., Colnat-Coulbois, S., Vespignani, H., et al. (2014). Intracerebral electrical stimulation of a face-selective area in the right inferior occipital cortex impairs individual face discrimination. *Neuroimage*, *99*, 487–497.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex

specialized for face perception. *Journal of Neuroscience*, 17, 4302–4311.

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, Biological Sciences*, 361, 2109–2128.

Kaspar, K., Hassler, U., Martens, U., Trujillo-Barreto, N., & Gruber, T. (2010). Steady-state visually evoked potential correlates of object recognition. *Brain Research*, 1343, 112–121.

Kim, Y.-J., Tsai, J. J., Ojemann, J., & Verghese, P. (2017). Attention to multiple objects facilitates their integration in prefrontal and parietal cortex. *Journal of Neuroscience*, 37, 4942–4953.

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science*, 320, 110–113.

Lamme, V. A. F., Supèr, H., Landman, R., Roelfsema, P. R., & Spekreijse, H. (2000). The role of primary visual cortex (V1) in visual awareness. *Vision Research*, 40, 1507–1521.

Lithari, C., Sánchez-García, C., Ruhnau, P., & Weisz, N. (2016). Large-scale network-level processes during entrainment. *Brain Research*, 1635, 143–152.

Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: An MEG study. *Nature Neuroscience*, 5, 910–916.

Liu, J., Harris, A., & Kanwisher, N. (2010). Perception of face parts and face configurations: An fMRI study. *Journal of Cognitive Neuroscience*, 22, 203–211.

Liu, T. (2016). Neural representation of object-specific attentional priority. *Neuroimage*, 129, 15–24.

Liu, T., Stevens, S. T., & Carrasco, M. (2007). Comparing the time course and efficacy of spatial and feature-based attention. *Vision Research*, 47, 108–113.

Mangun, G. R., & Hillyard, S. A. (1991). Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual-spatial priming. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 1057–1074.

Marinato, G., & Baldauf, D. (2019). Object-based attention in complex, naturalistic auditory streams. *Scientific Reports*, 9, 2854.

Maunsell, J. H. R., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences*, 29, 317–322.

Mayes, A. K., Pipingas, A., Silberstein, R. B., & Johnston, P. (2009). Steady state visually evoked potential correlates of static and dynamic emotional face processing. *Brain Topography*, 22, 145–157.

Moeller, S., Freiwald, W. A., & Tsao, D. Y. (2008). Patches with links: A unified system for processing faces in the macaque temporal lobe. *Science*, 320, 1355–1359.

Moore, T., & Armstrong, K. M. (2003). Selective gating of visual signals by microstimulation of frontal cortex. *Nature*, 421, 370–373.

Moore, T., & Zirnsak, M. (2017). Neural mechanisms of selective visual attention. *Annual Review of Psychology*, 68, 47–72.

Müller, M. M., Andersen, S., Trujillo, N. J., Valdés-Sosa, P., Malinowski, P., & Hillyard, S. A. (2006). Feature-selective attention enhances color signals in early visual areas of the human brain. *Proceedings of the National Academy of Sciences, U.S.A.*, 103, 14250–14254.

Müller, M. M., Malinowski, P., Gruber, T., & Hillyard, S. A. (2003). Sustained division of the attentional spotlight. *Nature*, 424, 309–312.

Nichols, D. F., Betts, L. R., & Wilson, H. R. (2010). Decoding of faces and face components in face-sensitive human visual cortex. *Frontiers in Psychology*, 1, 28.

Nobre, A. C., Gitelman, D. R., Dias, E. C., & Mesulam, M. M. (2000). Covert visual spatial orienting and saccades: Overlapping neural systems. *Neuroimage*, 11, 210–216.

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

O'Craven, K. M., Downing, P. E., & Kanwisher, N. (1999). fMRI evidence for objects as the units of attentional selection. *Nature*, 401, 584–587.

Parkkonen, L., Andersson, J., Hämäläinen, M., & Hari, R. (2008). Early visual brain areas reflect the percept of an ambiguous scene. *Proceedings of the National Academy of Sciences, U.S.A.*, 105, 20500–20504.

Parvizi, J., Jacques, C., Foster, B. L., Withoft, N., Rangarajan, V., Weiner, K. S., et al. (2012). Electrical stimulation of human fusiform face-selective regions distorts face perception. *Journal of Neuroscience*, 32, 14915–14920.

Perrett, D. I., Smith, P. A. J., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., et al. (1985). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proceedings of the Royal Society of London, Series B, Biological Sciences*, 223, 293–317.

Pinsk, M. A., Arcaro, M., Weiner, K. S., Kalkus, J. F., Inati, S. J., Gross, C. G., et al. (2009). Neural representations of faces and body parts in macaque and human cortex: A comparative fMRI study. *Journal of Neurophysiology*, 101, 2581–2600.

Pitcher, D., Walsh, V., & Duchaine, B. (2011). The role of the occipital face area in the cortical face perception network. *Experimental Brain Research*, 209, 481–493.

Pitcher, D., Walsh, V., Yovel, G., & Duchaine, B. (2007). TMS evidence for the involvement of the right occipital face area in early face processing. *Current Biology*, 17, 1568–1573.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32, 3–25.

Puce, A., Allison, T., Asgari, M., Gore, J. C., & McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: A functional magnetic resonance imaging study. *Journal of Neuroscience*, 16, 5205–5215.

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

Roelfsema, P. R., Lamme, V. A. F., & Spekreijse, H. (2004). Synchrony and covariation of firing rates in the primary visual cortex during contour grouping. *Nature Neuroscience*, 7, 982–991.

Rossi, A. F., & Paradiso, M. A. (1995). Feature-specific effects of selective visual attention. *Vision Research*, 35, 621–634.

Rossion, B. (2014). Understanding individual face discrimination by means of fast periodic visual stimulation. *Experimental Brain Research*, 232, 1599–1621.

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

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, 1490–1501.

Sadr, J., & Sinha, P. (2004). Object recognition and Random Image Structure Evolution. *Cognitive Science*, 28, 259–287.

Saenz, M., Buracas, G. T., & Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nature Neuroscience*, 5, 631–632.

Saenz, M., Buracas, G. T., & Boynton, G. M. (2003). Global feature-based attention for motion and color. *Vision Research*, 43, 629–637.

Schoenfeld, M. A., Hopf, J.-M., Merkel, C., Heinze, H.-J., & Hillyard, S. A. (2014). Object-based attention involves the sequential activation of feature-specific cortical modules. *Nature Neuroscience, 17,* 619–624.

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition, 80,* 1–46.

Schwedhelm, P., Baldauf, D., & Treue, S. (2017). Electrical stimulation of macaque lateral prefrontal cortex modulates oculomotor behavior indicative of a disruption of top–down attention. *Scientific Reports, 7,* 17715.

Serences, J. T., & Boynton, G. M. (2007). Feature-based attentional modulations in the absence of direct visual stimulation. *Neuron, 55,* 301–312.

Sergent, J., Ohta, S., & Macdonald, B. S. (1992). Functional neuroanatomy of face and object processing: A positron emission tomography study. *Brain, 115,* 15–36.

Siegel, M., Donner, T. H., Oostenveld, R., Fries, P., & Engel, A. K. (2008). Neuronal synchronization along the dorsal visual pathway reflects the focus of spatial attention. *Neuron, 60,* 709–719.

Spiridon, M., Fischl, B., & Kanwisher, N. (2006). Location and spatial profile of category-specific regions in human extrastriate cortex. *Human Brain Mapping, 27,* 77–89.

Spitzer, H., Desimone, R., & Moran, J. (1988). Increased attention enhances both behavioral and neuronal performance. *Science, 240,* 338–340.

Sprague, T. C., & Serences, J. T. (2013). Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices. *Nature Neuroscience, 16,* 1879–1887.

Störmer, V. S., & Alvarez, G. A. (2014). Feature-based attention elicits surround suppression in feature space. *Current Biology, 24,* 1985–1988.

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011). Brainstorm: A user-friendly application for MEG/EEG analysis. *Computational Intelligence and Neuroscience, 2011,* 879716.

Taulu, S., & Simola, J. (2006). Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. *Physics in Medicine and Biology, 51,* 1759–1768.

Treue, S., & Maunsell, J. H. R. (1996). Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature, 382,* 539–541.

Tsao, D. Y., Freiwald, W. A., Knutsen, T. A., Mandeville, J. B., & Tootell, R. B. H. (2003). Faces and objects in macaque cerebral cortex. *Nature Neuroscience, 6,* 989–995.

Tsao, D. Y., Freiwald, W. A., Tootell, R. B. H., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science, 311,* 670–674.

Tsao, D. Y., Moeller, S., & Freiwald, W. A. (2008). Comparing face patch systems in macaques and humans. *Proceedings of the National Academy of Sciences, U.S.A., 105,* 19514–19519.

Turk, D. J., Rosenblum, A. C., Gazzaniga, M. S., & Macrae, C. N. (2005). Seeing John Malkovich: The neural substrates of person categorization. *Neuroimage, 24,* 1147–1153.

Voytek, B., Samaha, J., Rolle, C. E., Greenberg, Z., Gill, N., Porat, S., et al. (2017). Preparatory encoding of the fine scale of human spatial attention. *Journal of Cognitive Neuroscience, 29,* 1302–1310.

Wegener, D., Ehn, F., Aurich, M. K., Galashan, F. O., & Kreiter, A. K. (2008). Feature-based attention and the suppression of non-relevant object features. *Vision Research, 48,* 2696–2707.

Weiner, K. S., & Grill-Spector, K. (2013). Neural representations of faces and limbs neighbor in human high-level visual cortex: Evidence for a new organization principle. *Psychological Research, 77,* 74–97.

Whelan, R. (2008). Effective analysis of reaction time data. *Psychological Record, 58,* 475–482.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods, 42,* 671–684.

Zhang, W., & Luck, S. J. (2009). Feature-based attention modulates feedforward visual processing. *Nature Neuroscience, 12,* 24–25.

Zhang, X., Mlynaryk, N., Japee, S., & Ungerleider, L. G. (2017). Attentional selection of multiple objects in the human visual system. *Neuroimage, 163,* 231–243.

Zhu, M., Alonso-Prieto, E., Handy, T., & Barton, J. (2016). The brain frequency tuning function for facial emotion discrimination: An ssVEP study. *Journal of Vision, 16,* 12.