Concurrent category-selective neural activity across the ventral occipito-temporal cortex supports a non-hierarchical view of human visual recognition

Corentin Jacques[1], Jacques Jonas[1,2] , Sophie Colnat-Coulbois[3], Bruno Rossion*[1,2]

1 Université de Lorraine, CNRS, IMoPA, F-54000 Nancy, France
2 Université de Lorraine, CHRU-Nancy, Service de Neurologie, F-54000 Nancy, France
3 Université de Lorraine, CHRU-Nancy, Service de Neurochirurgie, F-54000 Nancy, France

***Corresponding Author:**
Bruno Rossion
@mail: bruno.rossion@univ-lorraine.fr

IMoPA, UMR 7365, CNRS - Université de Lorraine
Pavillon Krug (1er étage - entrée CC-1), Hôpital Central
CHRU Nancy - University Hospital of Nancy
29 Avenue du Maréchal de Lattre de Tassigny
54000 NANCY FRANCE

Authors contribution: Corentin Jacques : Conceptualization, Methodology, Software, Formal analysis, Visualization, Writing - Original Draft, Writing - Review & Editing; **Jacques Jonas** : Conceptualization, Methodology, Investigation, Resources, Writing - Original Draft, Writing - Review & Editing; **Sophie Colnat-Coulbois** : Resources, Investigation; **Bruno Rossion** : Conceptualization, Methodology, Project administration, Supervision, Funding acquisition, Writing - Original Draft, Writing - Review & Editing .

Competing Interest Statement : The authors declare no competing interests.

Keywords: human face recognition; cortical hierarchy, SEEG; ventral occipito-temporal cortex; high-frequency broadband gamma.

ABSTRACT

Visual recognition is a fundamental human brain function, supported by a network of regions in the ventral occipito-temporal cortex (VOTC). This network is thought to be organized hierarchically, with definite processing stages increasing in invariance and time-course from posterior to anterior cortical regions. Here we provide a stringent test of this view by measuring category-selective neural activity to natural images of faces across the VOTC with electrophysiological intracerebral recordings in a large human sample (N=140; >11000 recording sites). Face-selective high frequency broadband (30-160 Hz) neural activity is distributed across the VOTC, with right-hemispheric dominance and regional peaks of activity. Crucially, while a progressive increase in degree of category-selectivity is found along the postero-anterior axis, neural activity occurs largely concurrently (~100ms onset – ~450ms offset) across all VOTC regions. These observations challenge the standard hierarchical view of neural organization of visual object recognition in the human association cortex, supporting alternative models of this key brain function.

INTRODUCTION

Recognition, i.e., reproducible discrimination, of signals from the sensory environment is a fundamental function of all central nervous systems(Edelman, 2004). Vision is considered the dominant sensory modality for most primates, who have high visual acuity and excellent binocular vision for recognizing their environment(Jacobs, 1999). In the primate order, visual object recognition is supported by a wide bilateral network of brain regions, including subcortical structures and the occipital and temporal cortices(Conway, 2018; DiCarlo et al., 2012; Ungerleider and Bell, 2011). Based primarily on non-human primates research, this network is generally conceived as being organized in a hierarchical manner, with definite processing stages increasing progressively in complexity of representation from posterior to anterior cortical regions, leading ultimately to rich invariant visual representations readily available for memory associations and behavior planning (Conway, 2018; DiCarlo et al., 2012; Freiwald, 2020; Grill-Spector et al., 2017; Hubel and Wiesel, 1968; Issa et al., 2018; Marr, 1982; Riesenhuber and Poggio, 1999; Tsao, 2014; Van Essen et al., 1992; Zeki and Shipp, 1988). In this hierarchical network, the selective response properties of neural populations in a given brain region are thought to stem from the combination of simpler responses from lower levels.

However, beyond the initial cascades of activities in early sensory brain structures (i.e., retina, lateral geniculate nucleus, primary visual cortices), whether visual object recognition is organized hierarchically remains largely unknown and disputed (Bullier, 2001; Eldridge et al., 2023; Hegdé and Felleman, 2007; Kravitz et al., 2013; Mumford, 1992; Rossion, 2022). This is particularly the case within the temporal association cortex, which is disproportionately enlarged in humans (Braunsdorf et al., 2021; Buckner and Krienen, 2013) and holds most category-selective ventral regions that are key for visual object recognition in our species (Bracci et al., 2017; Grill-Spector and Weiner, 2014; Rossion et al., 2018).

Here we provide important information to evaluate the hierarchical view of human visual object recognition through an extensive characterization of the time-course of category-selective brain activity along the human ventral occipito-temporal cortex (VOTC). To do so, we take advantage of the high spatial and temporal resolution provided by intracerebral electroencephalographic (iEEG) recordings in a particularly large sample of individual human brains (N=140) implanted from posterior to anterior regions of their VOTC

(>11000 intracerebral recording sites). With iEEG frequency-tagging, we isolate category-selective high frequency ('high frequency broadband', 30-160 Hz) neural activity to natural images of faces – arguably the most familiar and ecologically valid stimulus in the human environment-, to characterize its time-course across the whole VOTC. This allows testing two essential features of a hierarchical organization(Bullier and Nowak, 1995; Schmolesky et al., 1998): (1) an increase in representation complexity, or abstraction, along the hierarchy and (2) a progressive increase in the onset time of the earliest neural response at each level of the hierarchy, its inputs being driven by the output from the previous level.

Relative to natural images of various object shapes, we find consistent face-selective neural activity throughout the VOTC, with a progressive increase in face-selectivity, or abstraction, along the postero-anterior axis, reaching a high proportion of exclusive response to faces in the most anterior regions of the ventral temporal lobe. While these findings apparently support a hierarchical organization of visual recognition, face-selective neural activity occurs largely concurrently (~100ms onset – ~450ms offset latency) across the VOTC, challenging a hierarchical organization of visual object recognition in the human association cortex.

RESULTS

Neural signal was measured in 140 human participants implanted with intracerebral electrodes (**Figure 1A**) while they viewed sequences of variable natural images of objects presented at a rapid periodic rate (6 Hz, or 167ms SOA, one fixation/image). Highly variable natural images of faces appeared every 5 stimuli (i.e., 6/5 = 1.2 Hz; **Figure 1B-C**). This stimulation mode has been validated in tens of studies with multiple recording techniques (e.g., Electroencephalography (EEG) (Rossion et al., 2015) ; Magnetoencephalography (MEG) (Hauk et al., 2021); intracerebral recordings(Jacques et al., 2022; Jonas et al., 2016a); functional magnetic resonance imaging (fMRI)(Gao et al., 2018)), providing high signal-to-noise ratio (SNR) face-selective activity. In this stimulation mode, the periodic modulation of the high frequency broadband (HFB, 30-160 Hz, **Figure 1D**) signal allows to objectively (i.e., at pre-determined frequencies (Norcia et al., 2015; Regan, 1989)) identify and quantify both the neural activity common to faces and nonface objects at 6 Hz and harmonics, as well as the face-selective (i.e., differential) activity identified at the frequency of face stimulation (1.2 Hz) and specific harmonics (e.g., 2.4 Hz, 3.6 Hz,…; **Figure 1E**) in the frequency-

domain(Jacques et al., 2022; Retter and Rossion, 2016) ($z > 3.1$; $p < 0.001$). The use of HFB signal, considered as a local signal (relative to higher SNR phase-locked low frequency signals, (Jacques et al., 2022)) correlated with local population spiking activity (Manning et al., 2009) as well as BOLD signal (Jacques et al., 2016b), was intended to ensure maximal signal independence between VOTC regions investigated. Potential dependencies due to electrical field propagation were further limited by the use of local bipolar referencing.

Face-selective activity can also be characterized in the time-domain, which allows visualizing periodic activity common to faces and nonface objects (every 0.167 s) and any significant deviation from this activity (i.e., face-selective activity; every 0.833 s; **Figure 1F**). In the time-domain, the face-selective signal can be isolated by filtering-out the 6 Hz signal common to face and non-face objects (i.e., notch filter at 6 Hz and harmonics) (**Figure 1G**) (e.g., (Jacques et al., 2016a; Quek and Rossion, 2017; Retter and Rossion, 2016; Rossion et al., 2015)).

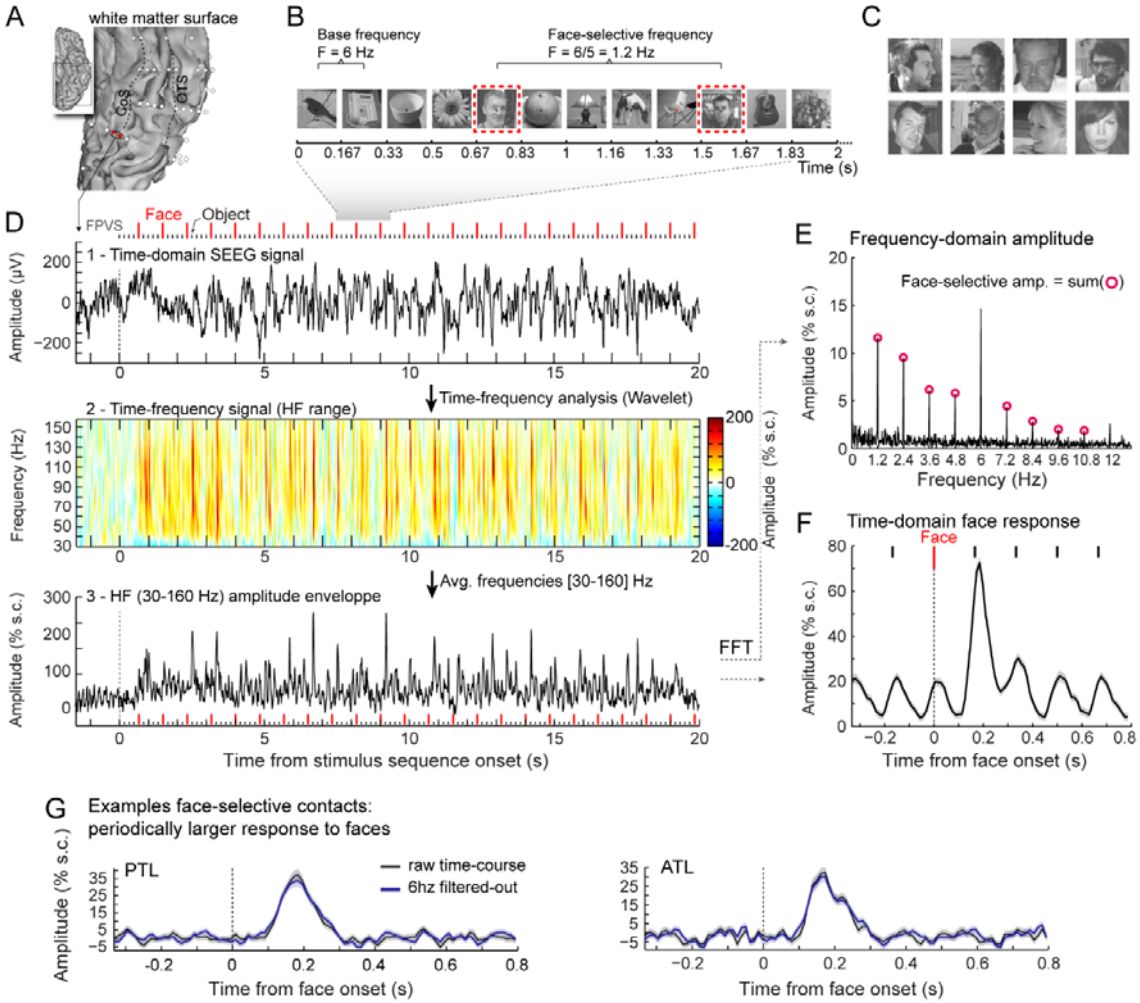


Figure 1. Recording and quantifying time-domain face-selective activity in the VOTC. **A.** SEEG (depth) electrode arrays (white circles) shown on the reconstructed white matter surface of one of the participants (ventral view of the left hemisphere). Electrodes penetrate both gyral and sulcal cortical tissues. **B.** The frequency-tagging paradigm to quantify face-selective neural activity: images of nonface objects appear at a rate of six stimuli per second (6 Hz) with variable face images presented every five stimuli (i.e., every 0.835s). Each stimulation sequence lasts for 65s (2s showed here). **C.** Representative examples of natural face images used in the study (actual images not shown for copyright reasons). **D**. Top: example raw intracranial EEG signal measured at the bipolar recording contacts shown in panel A (red). The signal is shown from -1.5 to 20 s relative to the onset of a stimulation sequence. The time-series displayed is an average of 2 sequences. Above the time-series, red vertical ticks indicate the appearance of face image every 0.835s and small black vertical ticks indicate the appearance of non-face objects every 0.167s. Middle: time by frequency representation of SEEG data in the HFB range (30-160 Hz). The plot shows the percent signal change at each frequency relative to a pre-stimulus baseline period (-1.6s to -0.3s). Periodic burst of HFB activity at the frequency of face stimulation (i.e., 1.2 Hz) are visible. Bottom: modulation of HFB amplitude over time obtained by averaging time-frequency signals across 30-160 Hz. **E.** HFB signal is transformed in the frequency domain to quantify face-selective amplitude as the sum of 12 face-selective frequency harmonics. **F.** Time-domain averaged HFB response to face images shows both the periodic response to non-face objects at 6 HZ (cycle duration = 0.167 s) and the larger face-selective response starting ~0.1 s after face onset. Shaded area around the curve is standard error of the mean across face trials. **G.** Mean face responses from two separate example recording contacts (in PTL and ATL) in which the 6 Hz response to non-face objects has (black traces) or has not (blue traces) been filtered-out.

Visualization of time-courses in contacts with significant face-selective activity reveals that face-selectivity is manifested either as a periodically larger response to faces compared to nonface objects (**Figure 1F,G, Figure 2**) or more rarely as a periodically smaller response to faces (**Figure S4**). Since category-selectivity is usually defined as a larger response for a given category compared to control categories in electrophysiological and neuroimaging studies, in the current study we focus on those larger activity for faces. Nevertheless, we provide illustrations and timing analyses for contacts showing response decrease in supplemental material **(Figures S5 and S7**). We identified 444 VOTC recording contacts in the gray matter and medial temporal lobe that showed a significant HFB face-selective response increase in 101 participants (among 8278 bipolar contacts located in gray matter or medial temporal lobe in the VOTC of 140 participants; 4680 in the left hemisphere and 3598 in the right hemisphere, **Table S1**). The proportion of face-selective contacts was significantly larger in the right (6.5%, 234/3598) than in the left hemisphere (4.5%, 210/4680, $p < 0.001$, 2-tailed permutation test).

Face-selective contacts displayed in the Talairach space for group visualization and analyses (**Figure 3A**), indicate that HFB face-selective activity was distributed across the VOTC, from the occipital lobe (OCC) to the anterior temporal lobe (ATL, 95% of contacts located in [-92 to -5] TAL y coordinate, **Figure 3A**), although not reaching the temporal pole (TP). Face-selective activity was also measured in subcortical structures of the medial

temporal lobe (MTL, amygdala- AMG). The bulk of the activity stretched from the IOG, through the FG and adjacent sulci (OTS and COS), up to the antFG+ (antFG and surrounding sulci: antOTS and antCOS). The highest number and proportion of face-selective contacts (**Figure 3B,C,** contacts labeled according to each participants individual anatomy as in(Jacques et al., 2022; Jonas et al., 2016a), **Figure S1**) were observed in bilateral latFG and antOTS as well as right IOG.

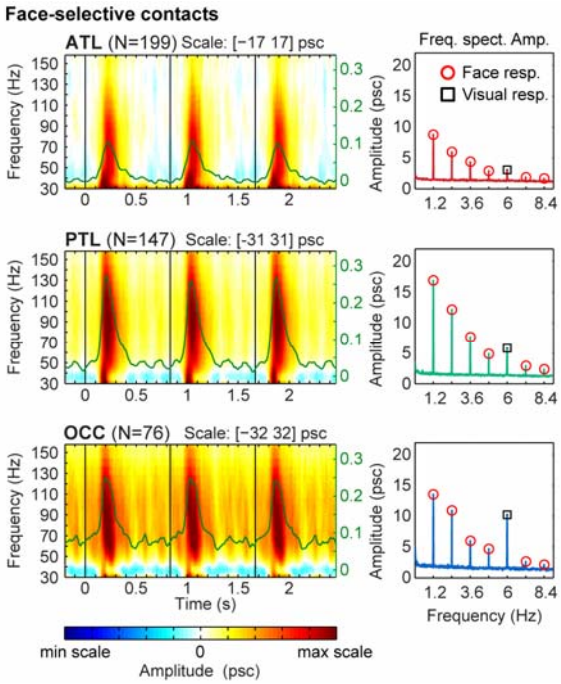


Figure 2: **Time-domain face-selective periodic increases.** Left column: Time by frequency response (percent signal change, psc; see scale value at the top of each plot and color scale at the bottom) in the HFB range (30-160 Hz) averaged over face-selective contacts in each main VOTC region (OCC: occipital, bottom; PTL: Posterior temporal lobe, middle; ATL: anterior temporal lobe, top). For each recording contact, the time-frequency data was segmented in epochs of about 3 face cycles (i.e. 3 x 0.833 s), averaged by contacts and then averaged over the three groups of contacts. Frequency axis is on the left. Green traces are the HFB amplitude envelope obtained by averaging over the 30-160 Hz range. Amplitude (psc) axis is on the right, show in green. Right column: Frequency spectra averaged over the corresponding groups of recording contacts and showing the face-selective response (red circles) at multiples of 1.2 Hz (i.e., face stimulation frequency) and visual response at 6 Hz (black square, other harmonics not shown). This highlights the sharp decrease in general visual response (i.e., 6Hz) relative to face-selective response from posterior to anterior VOTC.

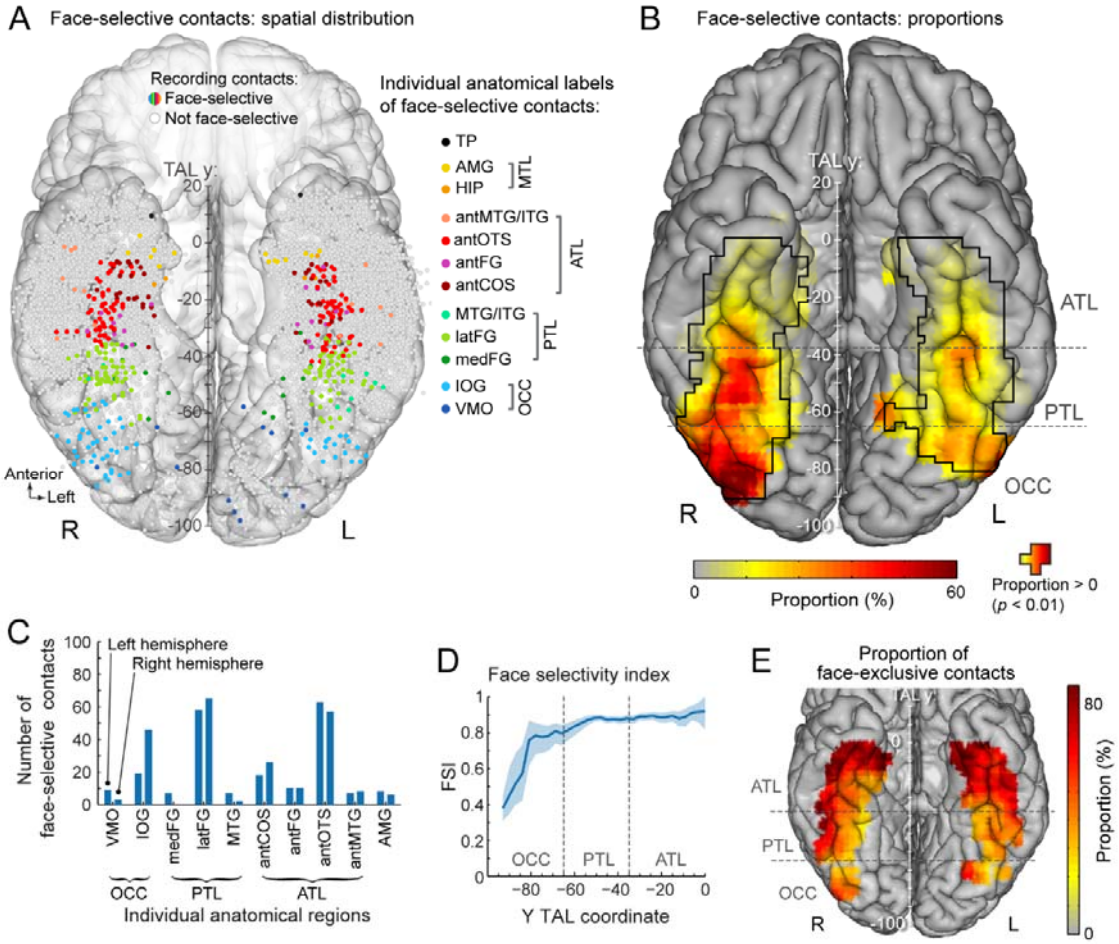


Figure 3: Spatial organization and increase in abstraction of face-selective HFB activity in VOTC. **A**. Map of all VOTC recording contacts across the 140 individual brains displayed in the Talairach space in a transparent reconstructed cortical surface of the Colin27 brain (ventral view). Each circle represents a single recording contact. Face-selective contacts are color-coded according to their anatomical location in the original individual anatomy. White-filled circles correspond to contacts without significant face-selective activity. Values along the y-axis of the Talairach coordinate system (postero-anterior) are shown near the interhemispheric fissure. **B**. VOTC maps of the local proportion of face-selective contacts relative to the number of recorded contacts. Black contour lines delineate local proportions significantly above zero ($p < 0.01$, percentile bootstrap). **C.** The number of face-selective contact is show for each anatomical region (region defined in each individual participant) and hemisphere. **D**. Face-selectivity index (FSI) along the postero-anterior axis collapsed along the X dimension (medio-lateral). The shaded area shows the 99% confidence interval computed using a percentile bootstrap. **E**. Map of the proportion of face-exclusive (i.e., face-selective without significant response to nonface-objects at 6Hz and harmonics) relative to face-selective contacts across VOTC.

Increase in selectivity ('abstraction') from posterior to anterior VOTC

Quantifying separately the amplitude of the face-selective response (i.e., 1.2Hz and harmonics, excluding harmonics of 6 Hz) and the amplitude the general visual response (i.e., manifested for all presented images, at 6Hz and harmonics) for each contact allows

173 computing an index of face-selectivity (face-selectivity index – FSI) that reflects the
174 magnitude of the face-selective response relative to the overall visual responsiveness of the
175 cortex around the recording contact. The FSI varies from 0 (no face-selective response) to 1
176 (only face-selective response, no general visual response) and can also be thought of as a
177 proxy for the level of abstraction exhibited by the local neural population (i.e., the degree to
178 which the neural population represents faces independently of generic visual input). FSI
179 increased from posterior VOTC to anterior VOTC: OCC = 0.72, 99% confidence interval: [0.65
180 0.79]; PTL = 0.88, [0.86 0.89]; ATL = 0.88, [0.86 0.9]; OCC vs. PTL: $p < 0.0002$, 2-tailed
181 bootstrap test; PTL vs. ATL: p = 0.45; OCC vs. ATL: $p < 0.0002$. Along the same lines,
182 computing the FSI as a function of the position along postero-anterior axis (Y Talairach
183 dimension, collapsing contacts over both hemispheres, **Figure 3D**) confirms a strong increase
184 in the FSI from posterior to anterior VOTC (from <0.4 to >0.9), most pronounced over the
185 posterior half of the VOTC. Additionally, we also quantified the proportion of face-selective
186 contacts that are 'face-exclusive' (i.e. face-selective contacts that do not display any
187 significant response to non-face objects measured at 6 Hz and harmonics with $z < 1.64$; $p >$
188 0.05). Face-exclusive contacts composed 38% of all face-selective contacts (169/444) with a
189 marked increase in proportion from OCC (18%), to PTL (26%) and ATL (51%), reaching ~80%
190 in the more anterior parts of ATL (**Figure 3E**).

191 **Concurrent face-selective activity across VOTC**

192 To achieve the most comprehensive and robust possible overview of the timing of
193 face-selective activity in VOTC, we started by examining the 3 main VOTC regions (i.e., OCC,
194 PTL, ATL, excluding TP and MTL for which there were only few significant contacts), including
195 all significant face-selective contacts showing a response increase (**Figure 3A**). Timing was
196 characterized by 4 parameters: (1) latency of response onset and (2) offset, as well as (3)
197 response overlap and (4) response correlation across VOTC regions (**Figure S2**). Analyses
198 focus on comparing these parameters (within hemispheres) to specifically test the
199 hierarchical processing hypothesis.

200 We first considered onset latencies of face-selective signals (**Figure 4**, see **Figure S3A**
201 for non-normalized time-series, **Figure S4** and supplementary results for contacts exhibiting
202 face-selective response decrease). These were overall similar across VOTC main regions,
203 ranging from 81 ms (95% confidence interval: [65-97] ms) in the right PTL region to 112 ms
204 ([100-122] ms) in the left PTL (**Figure 4**), with no clear pattern of inter-hemispheric or

205 postero-anterior differences. Statistically comparing regions within hemispheres revealed no
206 significant differences of onset latency (all *p*'s > 0.27, two-tailed permutation test, fdr-
207 corrected, **Figure 4B**). Second, the latency at which neural activity returned to baseline level
208 (i.e., offset latency) tended to slightly increase from OCC to ATL, ranging from 354 ms ([290-
209 415] ms) in left OCC to 532 ms ([454-620] ms) in right ATL (**Figure 4B**). However, due to a
210 large variability across contacts within each region, these differences in offset latency were
211 not significant (*p*'s > 0.1, fdr-corrected), except when comparing OCC to ATL in the right
212 hemisphere (393 ms, [349-425] ms vs. 532 ms ([454-620] ms for OCC and ATL respectively; *p*
213 = 0.034, two-tailed permutation test). Collapsing contacts across hemisphere confirmed the
214 significant increase in the offset latency from OCC (380 ms, [345-417] ms) to PTL (454 ms,
215 [407-483] ms; OCC vs PTL, *p* = 0.02, two-tailed permutation test, fdr-corrected) and from PTL
216 to ATL (522 ms, [497-618] ms; PTL vs. ATL., *p* = 0.0024). Together with the similar onset
217 latencies across main VOTC regions, this indicates that overall, response duration increases
218 as a function of postero-anterior VOTC location. Fourth, we examined the temporal overlap
219 between two given regions by computing the percentage of the area under the response
220 curve (AUC) of the overlap between region A and B (determined using onset and offset
221 latencies) relative to the total AUC of region A (see methods, **Figure S2**). Interestingly,
222 despite the differences in offset latencies across VOTC, on average when comparing two
223 regions, the AUC that falls within the temporal overlap between these regions represents
224 95% of the total AUC for each region, ranging from 87% (percentage of the right ATL AUC
225 occupied by the AUC of its temporal overlap with the right OCC, see **Figure 4B**) to 100% (e.g.,
226 percentage of the left OCC AUC occupied by the AUC of its temporal overlap with the left
227 ATL, **Figure 4B**). This indicates that, even in case of relatively large differences in offset
228 latency (e.g., 354ms for left OCC and 495ms for left ATL), the portion of the left ATL AUC
229 during non-overlapping window (e.g., between 354ms to 495ms for left OCC vs. ATL) is
230 negligible (12%) compared to the AUC observed during the temporal overlap (91 ms to 354
231 ms, 88% of the left ATL AUC occupied by the AUC of its temporal overlap with the left OCC).
232 Moreover, these observations indicate that the vast majority of the response in a region
233 located anteriorly (e.g., ATL) to another region (e.g., PTL) takes place during the time-
234 window of the posterior region.

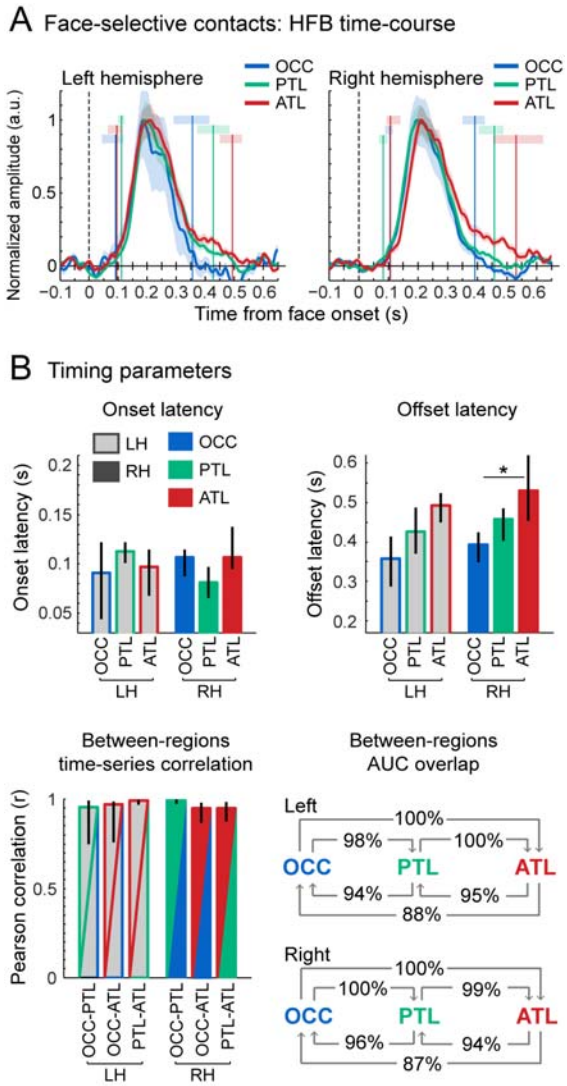


Figure 4: **Concurrent face-selective activity across VOTC**. **A.** Mean time-domain face-selective HFB activity in each VOTC region (OCC, PTL, ATL) and hemisphere. HFB time-series were filtered to remove the general visual response at 6 Hz and harmonics so only face-selective signal remains. The maximum amplitude of each averaged waveform was normalized to 1 for visualization purposes only (see Figure S4 for non-normalized waveforms). Shaded area represents the standard error of the mean between contacts. Colored vertical lines indicate onset and offset latencies for each VOTC region with shaded horizontal bar around each line representing a 95% confidence interval (percentile bootstrap). **B**. Timing parameters split by VOTC region and hemisphere: onset (top left) and offset (top right) latency of face-selective response, correlation of time-series between regions (bottom left) and between-region area under the curve (AUC) overlap (bottom right). For the AUC overlap, the arrow shows the directionality of the computed overlap. For instance, the arrow from ATL to OCC indicates that we are describing the percentage of the total AUC of ATL (measured between onset and offset latencies) occupied by the AUC of its temporal overlap (also determined using onset and offset) with OCC. Acronyms: LH: left hemisphere, RH: right hemisphere.

Fourth, as an additional measure of temporal overlap we considered the correlations between the mean time-series recorded in each region, which were highly correlated (**Figure 4B**), ranging from Pearson's r=0.94 when comparing right OCC to right ATL, to r=0.99 when

comparing right OCC to right PTL or left PTL to left ATL. This means that the time-series across regions share from 88% to 98% of their variance. Importantly, statistically comparing the correlations computed between-regions to correlations computed *within* each region revealed no significant difference (there were only trends for differences when comparing OCC to ATL and PTL to ATL in the right hemisphere (within OCC and within ATL: r=0.99 and 0.98 respectively, between OCC and ATL: 0.94, p = 0.06, one-tailed percentile boostrap test, fdr-corrected; within PTL and within ATL: r=0.99 and 0.98 respectively, between PTL and ATL: 0.94, p = 0.06). This indicates that the time-domain activity measured in signal+ contacts across different VOTC regions are not significantly different from the correlations expected within signals from a given region itself.

In a separate analysis (**Figure S5**), we also fully replicate the above observations but focusing on timing properties in 3 main face-selective regions in VOTC: IOG, latFG and antFG+ (i.e., antFG, antOTS, antCOS).

In addition to the group analyses described above, we performed a within-participants comparison of onset latencies in participants that had face-selective contacts in at least 2 different main VOTC regions (e.g., OCC and PTL). While within-participants statistics allow disregarding between-participants variability, it was done here at the expense of sample size since only a small subset of participants have face-selective contacts located in multiple face-selective regions. Onset latencies were computed for individual recording contacts and averaged by region within participants. These comparison did not reveal significant differences in onset latencies between main VOTC regions collapsed across hemispheres (OCC minus PTL: N=13, mean difference = -8 ms, 95% confidence interval: [-14 14]ms, p=0.27, 2-tailed permutation test; PTL minus ATL, N=19, mean difference = +14 ms, 95% CI: [-20 20]ms, p=0.18; OCC minus ATL, N=9, mean difference = +17 ms, 95% CI: [-27 27]ms, p=0.24) or between VOTC main face-selective regions (IOG minus latFG: N=9, mean difference = 0 ms, 95% CI: [-12 12]ms, p=0.88, 2-tailed permutation test; latFG minus antFG+, N=19, mean difference = +15 ms, 95% CI: [-20 20]ms, p=0.14; OCC minus ATL, N=7, mean difference = +8 ms, 95% CI: [-24 24]ms, p=0.52).

Concurrent functional connectivity between face-selective regions

Last, we estimated the functional connectivity between pairs of main face-selective regions (IOG, latFG, antFG+) by correlating single-trial face-selective amplitude measured in

285 different VOTC regions (lOG, latFG, antFG+), across time. Our use of local bipolar
286 referencing of electrophysiological signals ensures that responses in each region are
287 maximally local and that any significant correlation between regions cannot be attributed to
288 the reference itself. Directionality in connectivity (e.g. connectivity from lOG to latFG or the
289 reverse) was estimated by correlating face-selective amplitude measured at different time-
290 lags (-150 to 150ms) between the compared regions (Kadipasaoglu et al., 2017). The
291 properties of the functional connectivity patterns were largely similar across pairs or regions
292 (**Figure 5**), although correlations were weaker between lOG and antFG (**Figure 5C, F**;
293 especially in the left hemisphere where only 1 participant was included). Overall,
294 correlations were present before response onset, manifested by weaker scattered yet
295 largely symmetrical small correlation clusters. Stronger, more consistent and clustered
296 between-region functional connectivity appeared at response onset, i.e. around 80-100ms,
297 and was maximal at response peak between 150 and 200 ms (see Figure 4A), remaining
298 significant for the remaining of the response. Most importantly, correlations measured after
299 response onset were consistently centered on the diagonal (0 ms lag), further supporting
300 concurrent activity of the VOTC regions and suggesting that these regions receive common
301 correlated concurrent inputs possibly from lower-level visual cortex. Despite correlations
302 being centered on the diagonal, correlation coefficients symmetrically spread around the
303 diagonal, resulting either from reciprocal connections or from short (<30-40ms, likely due to
304 temporal smoothing caused by the time-frequency analyses) or longer (up to 150ms)
305 duration of temporal auto-correlation of the electrophysiological signal. These
306 broader/longer correlations around the diagonal, which were most pronounced between
307 left lOG-latFG and left/right latFG-antFG (Figure 5B, D-E), appeared after 200 ms and were
308 likely related to sustained and stable face-selective neural responses at these latencies
309 which are correlated in amplitude across regions.

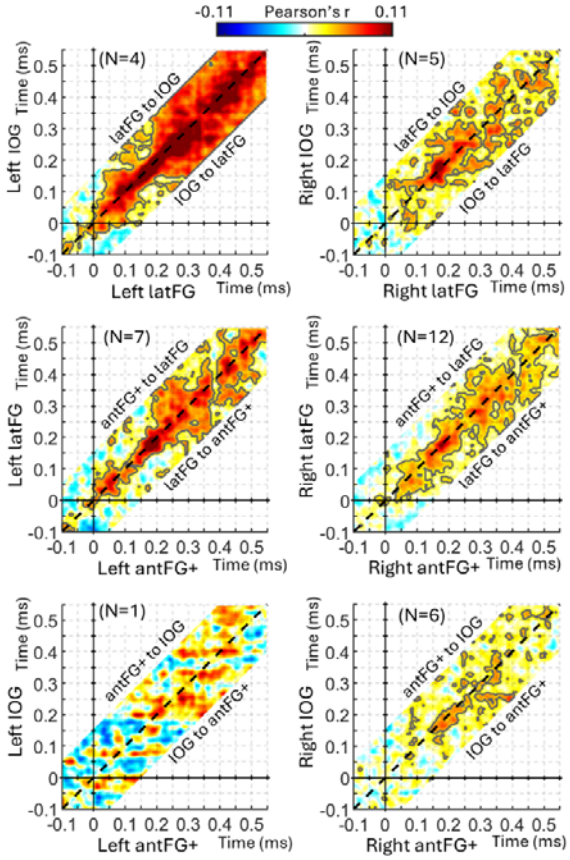


Figure 5: Concurrent functional connectivity between face-selective regions. (A) Group-averaged Pearson's correlations between single-trials face-selective amplitude measured at left lOG and left latFG computed across time, representing functional connectivity between these 2 regions. Black contour lines indicate significant positive correlations ($p < 0.01$, fdr-corrected). Correlations were computed across -150 to 150 ms lags between regions to infer direction of connectivity. The black dashed diagonal line represents a 0 ms time-lag between regions. Correlations centered above the diagonal would indicate that face-selective activity in latFG correlates with but precedes activity in lOG, suggesting information flow from latFG to lOG, and the reverse for correlations centered below the diagonal. B-F: functional connectivity between pairs of face-selective regions. Same conventions as in panel A.

Mapping face-selective onset latency in Talairach space

Next, we aimed to explore the onset latencies of face-selectivity HFB signal at a more local scale, taking advantage of the dense sampling of the VOTC to compute maps in the Talairach space. Due to the lower sample size in any given local measurement compared to when using larger anatomically-defined regions, and to limit noise in the latency estimates, we did not take into account contacts with low split-half correlation (SHC: r < 0.4; removing contacts with SHC < 0.4 had no effect when comparing timing properties across main VOTC regions). Moreover, we only included local VOTC volumes (i.e., 'voxels') with more than 5 contacts. For contacts showing face-selective response increase (contacts with response

decrease reported in **Figure S6**), face-selective response latency was lowest in regions exhibiting the most consistent face-selectivity: right IOG (right IOG: mean = 100 ms, range = [83 - 118ms], **Figure 6A**) and laFG (mean = 95 ms, range = [87 - 110ms]), as well as left latFG (mean = 103 ms, range: [99 - 112ms]), and the anterior portion of the right antFG+ area (mean = 113 ms, range: [95 - 132ms]). Between and around those regions, latency was slightly higher (e.g.: posterior FG: mean = 123 ms, range = [102 - 137ms]), especially in the antMTG region (mean = 148 ms, range = [132 - 172ms]). Examining the postero-anterior profile of response latency (**Figure 6B**) reveals relatively constant response latencies (90-110ms) in the OCC (Talairach Y coordinates: -90 to -65 mm]) and PTL regions (Talairach Y coordinates: -65 to -40 mm) up to the ATL region (Y > -40 mm) where the response latency increased, especially in the right hemisphere (i.e., 20-40 ms increase). Face-selective response onset latency in the right ATL (Y Talairach coordinates -30.5 to -27.5 and -18.5 to -15.5) was significantly above what would be expected by chance ($p < 0.05$, 2-tailed, fdr-corrected) if the antero-posterior position had no influence on face-selective response onset latency (**Figure 6B**, randomization test where the pairing between each recording contact location and its time-domain response was shuffled). Since the signal-to-noise ratio (SNR) and associated split-half reliability tends to be lower in the ATL(Jacques et al., 2022) and the estimated onset latency tends to increase at lower SNR, we computed the same analyses keeping only contacts with a high split-half correlation (SHC) above r = 0.7 (**Figure 6B, Figure S7**). As expected, due to the exclusion of lower SNR contacts, the estimated mean latency in the [-27.5 to -12.5 mm] Y coordinate interval was reduced by about 10-20 ms (compared to when using a threshold of SHC > 0.4) and was no longer significantly above chance. Additionally, given the relatively large distance between the ATL and posterior occipital regions (>50 mm), the slight increase in response onset latency from posterior to anterior VOTC can largely be accounted for by the predicted increase in response latency due to simple conduction delay (likely between 1.5 and 5 m/s(van Blooijs et al., 2023), **Figure 6B**), even with direct cortico-cortical connections from earlier visual cortical areas.

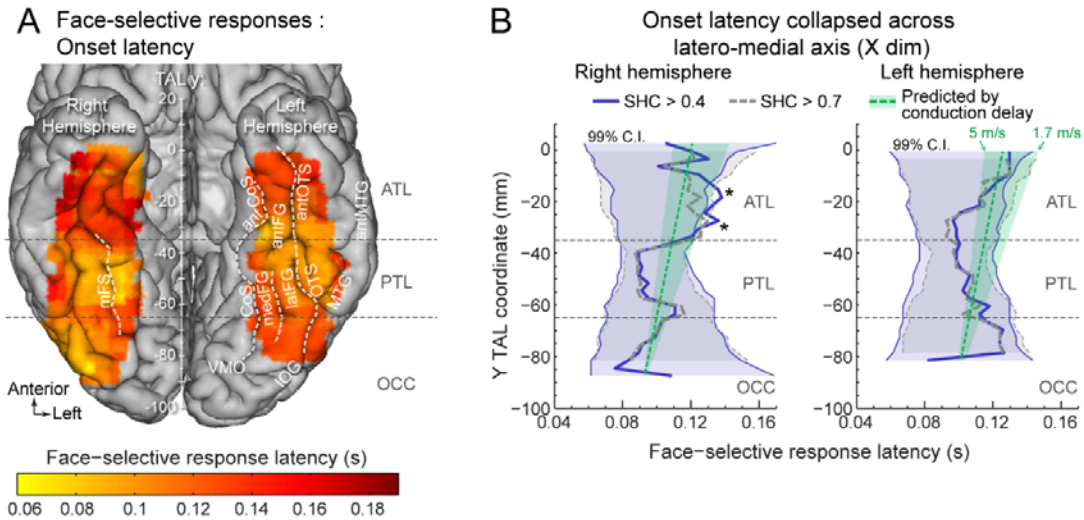


Figure 6: **Mapping concurrent face-selective response onset latency in VOTC**. **A**. Face-selective onset latency map across VOTC. **B.** Variation of face-selective response latency along the postero-anterior axis. Each data point represents the onset latency measured from the time-series averaged over contacts collapsed across the medio-lateral X dimension in each hemisphere within 20 mm segments (in the Y dimension). Thick lines are estimated onset latencies and shaded areas show the 99% confidence intervals expected under the null hypothesis that the postero-anterior location has no influence on the onset latency. Dark blue (full lines) is used for latencies computed when excluding contacts with SHC < 0.4 (as in panel A) and gray (dashed lines) are for latencies computed when excluding contacts with SHC < 0.7. Green line and shaded area show the expected increase in response onset latency based on simple conduction delays(van Blooijs et al., 2023) due to increasing distance from the occipital region, with reference to the latency averaged over the 2 most posterior face-selective response measured here. The dotted line is the mean expected conduction velocity for direct cortico-cortical connections (~3.2 m/s) and the shaded area represents the minimum (1.7 m/s) and maximum (5 m/s) expected conduction velocity (i.e., based on measurements from other brain regions (van Blooijs et al., 2023)).

Discussion

Here we provide a large-scale mapping of the time-course of neural activity supporting visual recognition in the human VOTC with electrophysiological intracerebral recordings. Faces are used as the key target category for visual recognition not only because of their highly significant social value for humans, but also due to their wide distribution of representation across the (bilateral) VOTC, allowing to compare the time-course of category-selectivity across multiple regions. In addition, the cortical network of face-selective regions is the most extensively studied within the occipital and temporal lobes and is thought to reflect the manifestation of a canonical set of operations that reveal general principles of how primate visual recognition works(Conway, 2018; DiCarlo et al., 2012; Freiwald, 2020; Grill-Spector et al., 2017).

We report two major findings. First, a progressive increase in face-selectivity along the postero-anterior axis. Second, a concurrent face-selective neural activity, with similar

385 onset latencies (80-100ms) and largely overlapping and correlated time courses across the
386 whole VOTC, spanning around 90 mm of cortical territory along the postero-anterior axis.
387 Our observations suggest that while the human VOTC may apparently be functionally
388 organized in a hierarchy of regions with increasingly more abstract representations, these
389 regions of the association cortex do not appear to be activated sequentially (or even in
390 cascade) but rather concurrently.

391 **1. Time course of VOTC face-selective activity and speed of face categorization**

392 The onset of category-selectivity for faces identified here in humans at around 80-
393 100 ms at the group level is early, supporting an initial fast feedforward sweep of activity for
394 rapid face categorization (Cauchoix et al., 2014; Crouzet and Thorpe, 2010; Retter et al.,
395 2020; Serre et al., 2007). This onset latency, with a peak of activity reached at about 200 ms,
396 is in line with multiple sources of evidence (speed of gaze to lateralized natural images of
397 faces(Crouzet and Thorpe, 2010); scalp EEG studies (Retter and Rossion, 2016; Rossion and
398 Caharel, 2011; Rousselet et al., 2008) ; intracranial EEG studies(Allison et al., 1999; Jacques
399 et al., 2016b; Kadipasaoglu et al., 2017; Liu et al., 2009); single neuron activity in human
400 face-selective fusiform regions (Laurent et al., 2025; Quian Quiroga et al., 2023)). Here we
401 extend these findings by providing comprehensive measurements of response onset
402 latencies across the whole VOTC. Importantly, we do not measure the absolute response
403 onset latency to face images(Cao et al., 2025; Schrouff et al., 2020) but the onset of
404 category-selectivity. Investigating category-selective activity is critical because it may be
405 temporally dissociated from general visual responses in the same brain areas(Jiang et al.,
406 2011) and is directly related to visual recognition function (Jonas and Rossion, 2021;
407 Rangarajan et al., 2014; Volfart et al., 2022).

408 Beyond onset latencies, we also report response durations that are in line with values
409 obtained from the scalp (EEG) with the same paradigm (about 420ms (Retter and Rossion,
410 2016; Rossion et al., 2015)). Despite a slight increase in duration of the sustained low
411 amplitude part of the time-series along postero-anterior VOTC axis, the overwhelming
412 majority of signal amplitude (87 to 100%) and variability (88 to 98%) occurs concurrently,
413 i.e., it temporally overlaps across brain regions. Given the timing similarity between the
414 present intracerebral data and scalp EEG data obtained in neurotypical individuals with the
415 same approach (Retter and Rossion, 2016; Rossion et al., 2015), there appear to be no delay

potentially attributable to the specific population tested here, i.e., patients with intractable refractory epilepsy (Allison et al., 1999). While we cannot exclude overall lower response amplitudes of face-selective activity in this specific population relative to normal individuals and/or potential shifts of hemispheric lateralization at the individual level depending on hemispheric seizure localization, this would not affect our conclusions. Finally, whereas face-selectivity is measured here during temporally fixed face stimulation (every 833 ms), observers perform an orthogonal task, are unaware of the periodicity of face stimulation and unable to determine it due to the rapid base stimulation rate. Most importantly, with this fast stimulation mode approach, both the amplitude and time course of face-selective activity are unaffected by the temporal distance between face stimuli/ratio of faces *vs.* objects (Retter & Rossion, 2016), or the temporal periodicity or predictability of face presentation (Quek and Rossion, 2017).

2. A progressive posterior to anterior gradient of category-selective abstraction in the VOTC

Our findings indicate a progressive increase in category-selectivity from posterior to anterior VOTC. More specifically, while category-selective populations of neurons in posterior VOTC regions tend to respond to all visual stimuli, as evidenced by high 6 Hz amplitude, this common visual response gradually decreases along the postero-anterior axis (**Figures 2, 3D**), leading to a high proportion of exclusive response to faces in the most anterior regions of the ventral temporal lobe (**Figure 3E**). The postero-anterior increase in face-selectivity is only partially in line with fMRI observations that have reported both larger face-selectivity in the fusiform gyrus ('Fusiform Face Area', FFA) than in the posteriorly located inferior occipital gyrus (IOG; 'Occipital Face Area', OFA)(Tsao et al., 2008; Weiner and Grill-Spector, 2010) , as well as the opposite finding (Rossion et al., 2012) (see also recent evidence(Chen et al., 2023) of larger face-selectivity in the posterior than middle fusiform gyrus), potentially depending on the general size/amplitude of activity in a given ROI and how selectivity indexes are computed. Most importantly, fMRI studies are limited to comprehensively explore the VOTC due to the large magnetic susceptibility artifacts (low SNR) in the ventral ATL (Ojemann et al., 1997; Rossion et al., 2024). To the best of our knowledge, previous intracranial studies have not described such a comprehensive pattern of variation in face-selectivity across the whole VOTC, due to limited sampling (spatially or

447 limited to the gyral surface), lack of mapping and/or face-selective response quantification,
448 or focus on medio-lateral variations of selectivity (Allison et al., 1999; Engell and McCarthy,
449 2014; Jacques et al., 2016b; Jonas et al., 2016a; Kadipasaoglu et al., 2017, 2016; Rangarajan
450 et al., 2014; Schrouff et al., 2020; Vidal et al., 2010).

451 This increase in selectivity for faces, as well as the increase in the proportion of face-
452 exclusive activity in the postero-anterior VOTC axis, cannot be attributed to a mere reduced
453 ability of anterior regions to generate responses at a fast presentation rate (6 Hz) (Jonas et
454 al., 2016a). Rather, these observations suggest an increase in 'abstraction' of visual neural
455 representations, with a large proportion of neuronal populations in anterior VOTC regions
456 exhibiting similar activity to different faces images independently of the context in which
457 they appear, while being much less (or not at all) responsive to other complex visual
458 stimulation. These findings are compatible with standard hierarchical models of visual
459 processing, such as proposed originally by Hubel and Wiesel (1962) (Hubel and Wiesel, 1962)
460 and extended to a hierarchy of stages for building increasingly complex invariant visual
461 object or face representations (Conway, 2018; DiCarlo et al., 2012; Duchaine and Yovel,
462 2015; Fairhall and Ishai, 2007; Freiwald, 2020; Grill-Spector et al., 2017; Issa et al., 2018;
463 Marr, 1982; Riesenhuber and Poggio, 1999; Tsao, 2014; Van Essen et al., 1992), as well as
464 with the view that this hierarchy is implemented anatomically along the postero-anterior
465 axis of the human VOTC.

466 3. Concurrent face-selectivity across VOTC: evidence for non-hierarchical 467 organization of visual association cortex

468 While the increase in category-selectivity from posterior to anterior VOTC as
469 described above is compatible with a hierarchical model of visual recognition, the
470 remarkably similar time-course of face-selective visual neural activity across human VOTC
471 regions is difficult to reconcile with the temporal properties implied by such model. In the
472 classical hierarchical view of visual processing (Conway, 2018; DiCarlo et al., 2012; Duchaine
473 and Yovel, 2015; Fairhall and Ishai, 2007; Freiwald, 2020; Grill-Spector et al., 2017; Issa et al.,
474 2018; Marr, 1982; Riesenhuber and Poggio, 1999; Tsao, 2014; Van Essen et al., 1992),
475 according to which representations at a given level of the hierarchy are built from
476 combination of inputs received from a lower level, implying serial ordering and thus a time

difference between successive levels (Bullier, 2001; Bullier and Nowak, 1995; DiCarlo et al., 2012; Schmolesky et al., 1998; Thorpe and Fabre-Thorpe, 2001).

Hierarchical models do not imply that a process should be completed at a given stage before the next one is initiated. Yet, even in a hierarchical cascade of cortical processes and representations (DiCarlo et al., 2012), and regardless of acknowledged feedback/reentrant loops (Edelman and Gally, 2013; Issa et al., 2018; Kar et al., 2019; Kravitz et al., 2013; Lamme and Roelfsema, 2000) there should be delays in onset times, and only partial temporal overlap between increasing processing level of the hierarchy(Bullier and Nowak, 1995; Issa et al., 2018; Schmolesky et al., 1998). Based on findings from electrophysiological as well as cortico-cortical evoked potential (CCEP) measurements, the delay between each level should be at least 10-30ms depending on (multi)synaptic transmission delays and conduction delays in short- or long-range connections(Conner et al., 2011; Nowak and Bullier, 1997; Schmolesky et al., 1998; van Blooijs et al., 2023). Importantly, this standard view is also how human, and non-human primate, cortical recognition of faces is generally conceived(Conway, 2018; Duchaine and Yovel, 2015; Fairhall and Ishai, 2007; Grill-Spector et al., 2017; Issa et al., 2018; Schweinberger and Neumann, 2016; Tsao, 2014), with 50-100 ms time difference postulated between stages (Sadeh et al., 2010; Schweinberger and Neumann, 2016).

In contrast to this hierarchical view, the present findings of concurrent face-selective activity across the human VOTC support a non-hierarchical organization of this recognition process. More specifically, our observations suggest that, rather than being successively activated along a posterior-axis, multiple face-selective populations of neurons spread across the whole VOTC receive direct and parallel sensory inputs from low-level visual areas (e.g., V1), through so-called bypass pathways. Such pathways have long been described in anatomical studies of the visual system of macaque monkeys (Conway, 2018; Distler et al., 1993; Eldridge et al., 2023; Van Essen and Maunsell, 1983; Zeki and Shipp, 1988;). In humans, specifically for face recognition, both lesion studies and time-resolved fMRI investigations have provided indirect evidence for bypass cortical pathways, reporting face-selective activity in the midFusiform gyrus (i.e. 'Fusiform face Area', FFA) in the absence of (lesion (Rossion et al., 2003; Steeves et al., 2006; Weiner et al., 2016)), or before(Jiang et al., 2011), any category-selective activity in the posteriorly located IOG. Moreover, while an intracranial recording study with a limited sample focusing largely on occipital and posterior

509 temporal face-selective regions has provided mixed outcomes regarding the onset of
510 selective response to faces (delayed onset in the left latFG relative to IOG in n=4; no delay in
511 the right hemisphere n=3), CCEP in the same study(Kadipasaoglu et al., 2017) also suggested
512 independent and parallel signal propagations between early visual areas and both face-
513 selective regions (i.e. latFG/FFA and IOG). Critically, this study did not report face-selective
514 onset latencies in the ATL, where ~200 face-selective contacts were measured in the current
515 study. Further evidence from Diffusion Tensor Imaging (DTI), suggest independent direct
516 connections from early visual cortex to face-selective regions in the IOG and post/mid-
517 fusiform gyrus(Bryant et al., 2019; Finzi et al., 2021; Kim et al., 2006; Wang et al., 2020;
518 Weiner et al., 2016). More generally, human V1 even appears to show (hominoid-specific)
519 connectivity with lateral, inferior, and anterior temporal regions, beyond the retinotopically
520 organized cortical areas(Bryant et al., 2019).

521 While these observations provide an anatomical basis for concurrent onset times in
522 face-selective regions of the human VOTC, the present electrophysiological data, collected
523 from a particularly large sample and with an objective face-selective measure, go beyond by
524 showing for the first time that (1) even the most anterior temporal population of neurons
525 (ATL) respond selectively to faces as early as posterior VOTC regions and (2) that all of these
526 regions' time-courses overlap in time remarkably. The current findings also support direct
527 connections between early visual cortex and face-selective regions around the anterior
528 fusiform gyrus, anterior OTS and anterior COS (antFG+) in the ATL (likely via the same fiber
529 bundles as for posterior face-selective regions: the inferior longitudinal fasciculus and
530 inferior fronto-occipital fasciculus) where DTI measurements are limited by magnetic
531 susceptibility artifacts. Importantly, while we report a slight overall increase in response
532 onset latency in the ATL compared to more posterior regions, this increase can be accounted
533 for by the predicted conduction delay given the large distance of the ATL from early visual
534 areas in the posterior occipital cortex and the likely reliance on long range white matter
535 fibers (~1.5 to 5 m/s or mm/ms(van Blooijs et al., 2023), **Figure 6B**). This further supports the
536 view that there is no need to postulate mandatory intermediate hierarchical stage/relays
537 between posterior and anterior regions. A number of factors can affect transmission speed,
538 as determined using CCEP in humans, among which the number of fibers connecting
539 different regions(Conner et al., 2011) and which is likely to vary between face-selective
540 regions(Finzi et al., 2021; Wang et al., 2020). Together with concurrent onset latencies, the

541 large overlap in time-courses suggests parallel category-selective processing, likely to be
542 initiated independently by sensory inputs from low-level visual cortex, as well as ongoing
543 reentrant interactions (i.e., dynamic recursive/recurrent exchange of signals (Edelman and
544 Gally, 2013; Kar et al., 2019; Lamme and Roelfsema, 2000)) between these regions for
545 ~300ms(Issa et al., 2018; Kadipasaoglu et al., 2017). Moreover, our finding of differences
546 across face-selective regions in the duration of the low amplitude sustained part of the
547 response may reflect different patterns of connectivity between these regions and other
548 (sub)cortical regions to which they are connected to. In particular, the longest duration in
549 the ATL may potentially be linked to the specific connectivity of this region with medial
550 temporal lobe and other structures involved in semantic or episodic memory formation or
551 retrieval (Persichetti et al., 2021).

552 **4. Implications for models of visual recognition**

553 How would sensory inputs be recognized as faces in such a non-hierarchical human
554 system? In a nutshell, through synaptic sensory inputs originating from low-level visual
555 cortex (e.g., V1) successfully triggering (post-synaptic) activity of sufficiently large face-
556 selective populations of neurons in parallel throughout the VOTC. Despite temporal overlap,
557 and concurrent process, there are genuine differences in degree of selectivity, as disclosed
558 here, as well as receptive fields sizes, foveal bias and ipsilateral sensitivity across face-
559 selective regions of the human VOTC (Finzi et al., 2021; Kay et al., 2015). These functional
560 differences between regions are advantageous for visual recognition since they increase the
561 probability of variable incoming sensory inputs from early visual cortices to successfully
562 trigger face-selective populations of neurons distributed along the VOTC. Importantly, in
563 such a non-hierarchical system, there is no requirement for normalization processes across
564 definite successive stages to build 'invariant' representations from inputs varying in size,
565 orientation, lighting, etc. Instead, through hebbian learning mechanisms, views that are
566 more commonly experienced (e.g., full-front faces at conversational distances) would be
567 represented by larger populations of neurons than rarely seen views, providing a faster
568 accumulation of category-selective neural activity (i.e., evidence accumulation) in this
569 network and a more efficient recognition function for sensory inputs corresponding to these
570 views(Perrett et al., 1998).

To be clear, such a non-hierarchical organization of the human ventral cortical face network corresponds fully to a feedforward/bottom-up view of visual recognition (at least for recognition of relatively clear views of faces) enriched by concurrent reentrant exchanges of signals within VOTC regions for a few hundreds of milliseconds. That is, putative 'top-down' signals from population of neurons in parietal or prefrontal cortices(Bar, 2003; Kar and DiCarlo, 2021) are not necessary for fast automatic recognition of faces. Yet, this simple view rests on a conceptual reinterpretation of functional organization of the human VOTC: rather than constituting a series of processing stages with increasing levels of complexity to derive visual representations (i.e., build the most veridical images of the world), face-selective populations of neurons are conceptualized as memory representations ('cortical memories'(Fuster, 1995)) distributed throughout the VOTC. That is, these populations of neurons, constrained by cytoarchitecture-specific white matter connections from birth (Kubota et al., 2025; Mahon, 2022), have learned (from early developmental stages(Kosakowski et al., 2022)) through Hebbian mechanisms to discharge selectively to faces, with temporal synchrony strengthening their connections. Recognition, which 'simply' involves successful – (often) concurrent - matching to these cortical memories in the human association cortex is achieved in a bottom-up fashion, with the 'up' component being critical in recognizing sometimes ambiguous or degraded sensory inputs as faces(Cavanagh, 1991).

Summary and conclusions

In summary, our large-scale human intracerebral investigation of category-selective activity in human VOTC reveals an increase in selectivity/abstraction in face-selective activity from posterior to anterior VOTC, in line with a hierarchical view of visual recognition ,yet with response timing indicating that category-selective neural processes occur largely concurrently across the whole VOTC. These findings suggest that visual recognition is supported by category-selective neural populations spread across VOTC receiving direct concurrent inputs from early visual cortex, without the need to rely on intermediate and successive hierarchical relays.

MATERIALS AND METHODS

Participants

The study included 140 participants (71 females, mean age: 33.0±9.2 years; 123 right-handed, 3 ambidextrous) undergoing clinical intracerebral evaluation with depth electrodes

602 (stereotactic electroencephalography or SEEG**, Figure 1A**) for refractory partial epilepsy.
603 Participants were included in the study if they had at least one intracerebral electrode
604 implanted in the ventral occipito-temporal cortex. All participants gave written consent to
605 participate to the study, which was approved by a national human investigation committee
606 certified by the French Ministry of Health (Institutional Review Board: IORG0009855)

607 **Fast periodic visual stimulation paradigm**

608 A well validated fast periodic visual stimulation (FPVS) paradigm with natural images
609 was used to elicit face-selective neural activity in iEEG with high signal to-noise ratio (SNR)
610 (see(Rossion et al., 2015) for the original description of the paradigm in EEG; see(Jonas et al.,
611 2016a; Rossion et al., 2018) for its validity in iEEG).

612 ***Stimuli.*** Two hundred grayscale natural images of various non-face objects (from 14 non-
613 face categories: cats, dogs, horses, birds, flowers, fruits, vegetables, houseplants, phones,
614 chairs, cameras, dishes, guitars, lamps) and 50 grayscale natural images of faces were
615 used(Jacques et al., 2022; Jonas et al., 2016a; Rossion et al., 2015). Each image contained an
616 unsegmented object or face near the center, these stimuli differing in terms of size,
617 viewpoint, lighting conditions and background. Images were equalized for mean pixel
618 luminance and contrast, but low-level visual cues associated with the faces and visual
619 objects remained highly variable, naturally eliminating the systematic contribution of low-
620 level visual cues to the recorded face-selective neural activity(Gao et al., 2018; Rossion et al.,
621 2015).

622 ***Procedure.*** Participants viewed continuous sequences of natural images of objects
623 presented at a fast rate of 6 Hz (i.e., stimulus onset asynchrony of 167ms) through sinusoidal
624 contrast modulation. This relatively fast rate allows only one fixation per stimulus and is
625 largely sufficient to elicit maximal face-selective activity(Retter and Rossion, 2016). Images of
626 faces appear periodically as every 5th stimulus, so that neural activity that is common to
627 faces and nonface stimuli is manifested at 6 Hz and harmonics, while differential (i.e.,
628 selective) reliable activity to faces are expressed at 1.2 Hz (i.e., 6 Hz/5) (see **Figure 1B,C**). All
629 images were randomly selected from their respective categories (face and nonface), with the
630 constrain that no image could be immediately repeated. A stimulation sequence lasted 70 s:
631 66 s of stimulation at full-contrast flanked by 2 s of fade-in and fade-out, where contrast

632 gradually increased or decreased, respectively. During a sequence, participants were
633 instructed to fixate a small black cross which was presented continuously at the center of
634 the stimuli and to detect brief (500 ms) color-changes (black to red) of this fixation-cross.
635 Among the 140 participants, participants viewed either 2 sequences (69 participants), 3
636 sequences (6 participants), 4 sequences (55 participants), 5 sequences (1 participant), 6
637 sequences (3 participants), 8 or more sequences (6 participants). No participant had seizures
638 in the 2 hours preceding the recordings.

639 **Intracerebral electrode implantation and SEEG recording**

640 Intracerebral electrodes (Dixi Medical, Besançon, France) were stereotactically
641 implanted within the participants' brains for clinical purposes, i.e., to delineate their seizure
642 onset zones (Talairach and Bancaud, 1973) and to functionally map the surrounding cortex
643 for potential epilepsy surgery. Each 0.8 mm diameter intracerebral electrode contains 5-15
644 independent recording contacts of 2 mm in length separated by 1.5 mm from edge to edge
645 (**Figure 1A**). A total of 997 electrode arrays were implanted in the VOTC of the 140
646 participants. These electrodes contained 11121 individual recording contacts in the VOTC
647 (i.e., in the gray/white matter or medial temporal lobe-MTL; 6295 and 4826 contacts in the
648 left and right hemisphere respectively). Intracerebral EEG was sampled at either 500 or 512
649 Hz and referenced to either a midline prefrontal scalp electrode or an intracerebral contact
650 in the white matter. SEEG signal was re-referenced offline to bipolar reference to limit
651 dependencies between neighboring contacts(Hagen et al., 2025). Specifically, the signal at a
652 given recording contact was computed as the signal measured at that contact (i.e., with the
653 recording reference) minus the signal at the directly adjacent contact located more medially
654 on the same SEEG electrode array. Since SEEG field potentials are computed using pairs of
655 adjacent contacts, each electrode array contains 1 contact less than in the original recording.
656 All subsequent analyses were performed on bipolar-referenced signal in the set of bipolar
657 contacts as described just above.

658 ***Contact localization in the individual anatomy***. The position of each contact relative to brain
659 anatomy was determined in each participant's own brain by coregistration of the post-
660 operative CT-scan with a T1-weighted MRI of the patient's head. Anatomical labels of bipolar
661 contacts were determined using the anatomical location of the 'active' contact. In cases
662 where the active contact was in the white matter and the 'reference' contact was in the gray

663 matter, the active contact was labeled according to the anatomical location of the reference
664 contact. Bipolar contacts in which both the active and reference contacts were in the white
665 matter were excluded from analyses. To accurately assign an anatomical label to each
666 contact, we used the same topographic parcellation of the VOTC as in(Jacques et al., 2022;
667 Jonas et al., 2016b) (**Table S1, Figure S1**).

668 **SEEG signal processing and analyses**

669 ***High frequency broadband (HFB) preprocessing.*** Segments of iEEG corresponding to
670 FPVS stimulation sequences were extracted (74-second segments, -2s to +72s, **Figure 1D**,
671 top) and notched-filtered to remove 50Hz line noise and 2 harmonics (100 and 150Hz).
672 Variation in signal amplitude as a function of time and frequency was estimated by a Morlet
673 wavelet transform applied on each SEEG 74-second segment from frequencies of 30 to 160
674 Hz, in 2 Hz increments (**Figure 1D**, middle). The number of cycles (i.e., central frequency) of
675 the wavelet was adapted as a function of frequency from 2 cycles at the lowest frequency to
676 9 cycles at the highest frequency. The temporal smoothing resulting from the wavelet
677 transform was minimal: wavelet of 20 ms of full width at half maximum (FWHM) across the
678 frequency range (i.e. median of FWHM computed at each frequency bin), ensuring that
679 timing information is accurate up to 10 ms which corresponds to half of the FWHM. The
680 wavelet transform was computed on each time-sample and the resulting amplitude
681 envelope was downsampled by a factor of 6 (i.e., to a 85.3 Hz sampling rate) to save disk
682 space and computation time. However, for timing analyses on face-selective contacts, the
683 original recording sampling rate (500 or 512 Hz) was used to preserve temporal resolution.
684 Resulting time-by-frequency SEEG segment were normalized to obtain, for each frequency
685 bin, the percentage of signal change (PSC) generated by the stimulus onset relative to the
686 mean amplitude in a pre-stimulus time-window (-1600 ms to -300 ms relative to the onset of
687 the stimulation sequence). The PSC was then averaged across frequencies (between 30 Hz
688 and 160 Hz) to obtain 74-seconds segments of time-varying HFB amplitude envelope (**Figure**
689 **1D**, bottom).

690 ***HFB frequency-domain processing.*** For each recording contact, 74s segments (i.e., 2-8 per
691 participant) of HFB PSC envelope were averaged in the time-domain and cropped to contain
692 an integer number of 1.2 Hz cycles beginning 2 seconds after the onset of the FPVS sequence
693 (right at the end of the fade-in period) until approximately 68 seconds, before stimulus fade-

694 out (79 face cycles ≈ 65.8 s). An FFT was then applied to the resulting cropped HFB envelope
695 to obtain the amplitude spectrum in the frequency domain (**Figure 1E**). The frequency-
696 tagging approach used here allows identifying and separating two types of neural activity:
697 (1) a general visual response occurring at the base stimulation frequency (6 Hz) and its
698 harmonics, as well as (2) a face-selective activity at 1.2 Hz and its harmonics(Jacques et al.,
699 2022; Jonas et al., 2016a; Rossion et al., 2018). Face-selective activity significantly above
700 noise level at the face stimulation frequency (1.2 Hz) and its harmonics (2.4, 3.6 Hz, etc.)
701 were determined as follows(Hagen et al., 2025; Jacques et al., 2022): (1) the FFT spectrum
702 was cut into 4 segments centered at the face frequency and harmonics, from the 1st until
703 the 4th (1.2 Hz until 4.8 Hz), and surrounded by 25 neighboring bins on each side; (2) the
704 amplitude values in these 4 segments of FFT spectra were summed; (3) the summed FFT
705 spectrum was transformed into a Z-score. Z-scores were computed as the difference
706 between the amplitude at the face frequency bin and the mean amplitude of surrounding
707 bins divided by the standard deviation of amplitudes in the surrounding bins. A contact was
708 considered as showing a face-selective response in HFB if the Z-score at the frequency bin of
709 face stimulation exceeded 3.1 (i.e., $p < 0.001$ one-tailed: signal>noise).

710 ***HFB time-domain preprocessing.*** Recording contacts with a significant face-selective
711 response (based on frequency-domain analyses) were further processed in the time-domain.
712 Starting from the 74-seconds HFB amplitude segments (see 'HFB preprocessing'), time-series
713 were processed in the following way: (1) an FFT notch filter (filter width = 0.07Hz) was
714 applied to remove the general visual response at 6Hz and 3 additional harmonics (i.e. 6, 12,
715 18, 24 Hz); (2) time-series were segmented in 1.17 s epochs centered on the onset of each
716 face (i.e. [-2 to + 5] 6Hz-cycles relative to face onset) in the FPVS sequences; (3) resulting
717 epochs were averaged (see **Figure 1F** for an example averaged time-course without the
718 notch filtering of the general visual response). Unless noted otherwise, averaged time-series
719 per contact were baseline-corrected by subtracting the mean amplitude in a [-0.166 to 0 s]
720 time-window relative to face onset. Considering time-series for each contact revealed that
721 face-selective activity manifest either as a periodically *larger* response to faces compared to
722 non-face objects, or as a periodically *smaller* response to faces (**Figure S4)**. Response
723 increase/decrease was defined as a function of whether the amplitude in the 0.15s to 0.35s
724 time-window was respectively larger or smaller than the amplitude in the time-window just

725 preceding the onset of the face (1 cycle at 6Hz, i.e. [-0.167s to 0s]). During the FPVS
726 sequences, images are presented using a sinusoidal modulation of contrast, rather than an
727 abrupt onset (**Figure 1A**). Hence, to account for the delay with which the images become
728 visible/perceivable in this setting, the face onset time was shifted forward by 33 ms, as
729 established from a direct comparison of sine wave to square wave/abrupt stimulus
730 presentation in a full EEG study (Retter and Rossion, 2016) and a subset of participants of
731 the present study. This corresponds to 4 frames at 120Hz screen refresh rate and to a face
732 contrast of about 35% of the maximal(Retter and Rossion, 2016).

733 ***Group visualization and analyses in Talairach space.*** For group mapping and visualization,
734 anatomical MRIs were spatially normalized to determine the Talairach (TAL) coordinates of
735 VOTC intracerebral contacts. The cortical surface used to display group contact locations and
736 maps was obtained from segmenting the Collin27 brain from AFNI(Cox, 1996), which is
737 aligned to the TAL space. We used TAL transformed coordinates to compute maps of the
738 local proportion of face-selective intracerebral contacts across the VOTC. Local proportion of
739 contacts was computed in volumes (i.e., 'voxels') of size 15 x 15 x 100 mm (respectively for
740 the X: left – right, Y: posterior – anterior, and Z: inferior – superior dimensions) by steps of 3
741 x 3 x 100 mm over the whole VOTC. A large voxel size in the Z dimension enabled collapsing
742 across contacts along the inferior-superior dimension. For each voxel, we extracted the
743 following information across all participants in our sample: (1) number of recorded contacts
744 located within the voxel across all participants; (2) number of significant face-selective
745 contacts. From these values, for each voxel we computed the proportion of face-selective
746 contacts as the number of significant contacts within the voxel divided by the total number
747 of recorded contacts in that voxel. Then, for each voxel we determined whether the
748 proportion of significant contacts was significantly above zero using a percentile bootstrap
749 procedure, as follows: (1) within each voxel, sample as many contacts as the number of
750 recorded contacts, with replacement; (2) for this bootstrap sample, determine the
751 proportion of significant contacts and store this value; (3) repeat steps (1) and (2) 2,000
752 times to generate a distribution of bootstrap proportions; and (4) estimate the *p*-value as
753 the fraction of bootstrap proportions equal to zero.

754 ***Quantification of response amplitudes.*** Amplitude quantification was performed on face-
755 selective contacts. We first computed baseline-subtracted amplitudes in the frequency

domain as the difference between the amplitude at each frequency bin and the average of 48 corresponding surrounding bins (up to 25 bins on each side, i.e., 50 bins, excluding the 2 bins directly adjacent to the bin of interest, i.e., 48 bins). Then, for each contact, face-selective amplitude was quantified as the sum of the baseline-subtracted amplitudes at the face frequency from the 1st until the 14th harmonic (1.2 Hz until 16.8 Hz), excluding the 5th and 10th harmonics (6 Hz and 12 Hz) that coincided with the base frequency(Jonas et al., 2016a). General visual response amplitude was quantified separately as the sum of the baseline-subtracted amplitudes at the base frequency from the 1st until the 3rd harmonic (6 Hz until 18 Hz).

Face-selectivity index. We computed an index of face-selectivity for face-selective contacts by taking the ratio of the face-selective amplitude (i.e., at 1.2Hz and harmonics, see above) to the sum of the face-selective and general visual amplitude (i.e., 6Hz and harmonics). This index provides an additional quantification of face-selectivity by taking into account the magnitude of response to non-face stimuli. It varies from 0 (no face-selective response) to 1 (face-selective response only, no general visual response). Face-selectivity indices were either computed at the scale of main VOTC regions or computed in Talairach space as a function of the posterior-anterior axis (Y TAL dimension). In the latter case, face-selectivity indices were computed along the Y dimension in segments of 15 mm and by steps of 3 mm, collapsing contacts across the X (lateral-medial, collapsed across both hemispheres) and Z (inferior-superior) dimensions. Face-selective and general visual activity were first averaged across contacts (i.e., within regions or voxels) before computing the index. This avoided obtaining indices outside the [0 - 1] range if, for instance, the denominator is smaller than 1.

HFB response timing parameters. In a first group timing analysis, the averaged HFB time-domain activity per recording contact (see '*HFB time-domain preprocessing',* **Figure 1G**) were grouped by main VOTC region (OCC, PTL, and ATL) and hemisphere. The timing of face-selective activity was characterized using 4 parameters: (1) latency of response onset and (2) offset, as well as (3) response overlap and (4) response correlation across VOTC regions (**Figure S2**). The first two parameters, onset and offset latencies of face-selective response, were computed per group of contacts using a bootstrapping approach in the following way (**Figure S2A)**: (1) randomly sample contacts with replacement; (2) averaging HFB time-series from sampled contacts; (3) converting this averaged HFB time-series to Z-score values by

subtracting the mean amplitude in the baseline window of the time-series (i.e. before face onset: [-0.166 to 0s]) and dividing by the standard deviation of the amplitude in the same baseline window; (4) converting the z-score values to *p*-values, apply an FDR correction and compute onset latency as the first time-point at which the $p < 0.05$ (two-tailed), for at least 30 ms. Offset latency was the first time bin (after the onset) at which the response was no longer significantly different from baseline; (5) store onset and offset latencies for the current bootstrap sample; (6) repeat steps (1) to (5) 4000 times; (7) use the obtained bootstrap distributions of onsets/offsets to compute the mean and 95% confidence interval for these 2 parameters. Onset and offset latencies were compared across regions within hemispheres and signal type using a permutation test, randomly shuffling (5000 times) the assignment of contacts in the two regions compared. Third, we estimated the overlap between time-series for pairs of regions (e.g., region A and B) within hemispheres (**Figure S2B)**. The overlap is asymmetrical and is calculated separately for region A and region B as the ratio between the area under the curve (AUC) of the overlap between regions A and region B (i.e., summing the amplitude between the maximum of onset A and B and the minimum of offset A and B) and the total AUC for region A (for overlap of region A to B) or for region B (for overlap of region B to A). Fourth, for each signal type we computed Pearson correlations between the time-series obtained in the three main VOTC regions, within each hemisphere (**Figure S2C)**. For each pair of regions, we determined whether the between-region correlation coefficient was significantly lower than within-regions correlations using a bootstrapping approach: (1) compute the *between-region correlation* by repeatedly (5000 times) randomly sampling half of the available time-series from each of the two regions, averaging the time-series for each region, calculate the Pearson correlation between these averaged time-series and take the mean over the obtained 5000 correlation coefficient; (2) compute the within-region correlation for each region in a similar manner by repeatedly (5000 times) randomly splitting the available time-series in two groups, average the time-series for each group, correlate time-series across the two groups and take the mean over the obtained 5000 correlation coefficient.

Functional connectivity between VOTC regions. Connectivity between regions was computed using participants that had face-selective contact(s) in at least 2 of the dominant face-selective regions in the same hemisphere: right/left IOG, latFG, antFG+. We used largely the same approach as in (Kadipasaoglu et al., 2017) , except for the statistics. Starting from

819 the single-trial HFB time-domain activity per recording contact, we performed the following
820 steps: (1) dowsampling of the time-series by a factor of 2 to save computation time; (2)
821 baseline-correction of each trial by subtracting the mean amplitude in the -166 to 0 ms; (3)
822 subtract from each single-trial, the mean across trials to obtain trial-by-trial variability
823 around the mean. The single-trial amplitudes were correlated between pairs of electrodes at
824 each time point from -100 to 550 ms to compute the instantaneous functional connectivity
825 between regions. To explore the directionality in the connectivity, we also computed the
826 across-trial correlations after introducing variable temporal lags between regions (from -150
827 to 150 ms). For each pair of contacts, we lagged the time-series at one of the contact before
828 computing the correlation. This allows determining directionality between regions if the
829 amplitude of face-selective activity at a given time point in one region correlates with
830 amplitude in the other region at an earlier or later time-point. Correlations were
831 summarized in lagged-correlation matrices (**Figure 5**). In each participant and each pair of
832 regions, correlations were computed across all possible pairs of face-selective contacts,
833 averaged across contacts and then averaged across participants to obtain group estimates.
834 Statistics were performed using a randomization test to determine the null distribution, by
835 running 1000 times the same analyses steps after having randomized the order of trials
836 between pairs of electrodes being correlated. Group-level correlations were tested against
837 the null distributions at $p < 0.01$ fdr-corrected.
838
839 ***HFB onset latency map.*** In addition to computing timing parameters across main VOTC
840 regions, we computed a group map of local face-selective response onset latency across
841 VOTC in Talairach space. We averaged HFB time-series from contiguous face-selective
842 contacts located within voxels of 20 x 20 x 100 mm (swept across VOTC in steps of 3 x 3 x
843 100 mm; see above '*Group visualization and analyses in Talairach space')* and computed
844 onset latency of face-selective response for the averaged time-series in the same way as for
845 the analyses in main VOTC regions (see '*HFB response timing parameters',* **Figure S2**). Due to
846 the lower sample size in any given local measurement (i.e., voxel) compared to when using
847 larger anatomically-defined regions and to limit noise in the latency estimates, we rejected
848 recording contacts with split-half correlation below $r = 0.4$. Split-half correlation was
849 computed for each face-selective contact by repeatedly (5000 times) randomly splitting the
850 available single face trials (i.e., short time-series evoked by individual face presentations) in 2

groups, averaging the time-series for each group, calculate the Pearson correlation between these averaged time-series and take the mean over the obtained 5000 correlation coefficients. In additions, we only considered voxels with more than 5 contacts. We also examined the poster-anterior variation of face-selective response latency by averaging HFB time-series from contacts located in voxels of 20 mm (along the Y dimension) collapsing contacts across the X (lateral-medial, separately for each hemisphere) and Z (inferior-superior) dimensions. Voxels were swept across the postero-anterior axis in steps of 3 mm. We used a randomization procedure to determine whether the measured postero-anterior profile of face-selective response onset latency deviate from what would be expected under the (null) hypothesis that the postero-anterior VOTC location has no influence on the onset latency. Specifically, we built a random null distribution by repeatedly (5000 times) computed the postero-anterior profile of onset latency after having randomly shuffled the association between contacts (and associated averaged time-series) and their TAL coordinates. P-values were determined for each voxel as the fraction of the null distribution that was higher/lower than the actual measured onset latencies and were corrected for multiple comparisons in each hemisphere using FDR(Benjamini and Hochberg, 1995).

Acknowledgments/Funding

This research was supported by ERC AdG HUMANFACE 101055175

References

Allison T, Puce A, Spencer DD, McCarthy G, Belger A. 1999. Electrophysiological studies of human face perception. I: Potential generated in occiptotemporal cortex by face and non-face stimuli. *Cerebral Cortex* **9**:415–430.

Bar M. 2003. A cortical mechanism for triggering top-down facilitation in visual object recognition. *J Cogn Neurosci* **15**:600–609. doi:10.1162/089892903321662976

Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological)* **57**:289–300.

Bracci S, Ritchie JB, de Beeck HO. 2017. On the partnership between neural representations of object categories and visual features in the ventral visual pathway. *Neuropsychologia* **105**:153–164. doi:10.1016/j.neuropsychologia.2017.06.010

Braunsdorf M, Blazquez Freches G, Roumazeilles L, Eichert N, Schurz M, Uithol S, Bryant KL, Mars RB. 2021. Does the temporal cortex make us human? A review of structural and functional diversity of the primate temporal lobe. *Neurosci Biobehav Rev*. doi:10.1016/j.neubiorev.2021.08.032

Bryant KL, Glasser MF, Li L, Bae JJ, Jacquez NJ, Preuss TM. 2019. Organization of extrastriate and temporal cortex in chimpanzees compared to humans and macaques. *Cortex* **118**:223–243. doi:10.1016/j.cortex.2019.02.010

Buckner RL, Krienen FM. 2013. The evolution of distributed association networks in the human brain. *Trends Cogn Sci*. doi:10.1016/j.tics.2013.09.017

Bullier J. 2001. Integrated model of visual processing. *Brain Res Brain Res Rev* **36**:96–107.

Bullier J, Nowak LG. 1995. Parallel Versus Serial Processing - New Vistas on the Distributed Organization of the Visual-System. *Curr Opin Neurobiol* **5**:497–503.

Cao R, Zhang J, Zheng J, Wang Y, Brunner P, Willie JT, Wang S. 2025. A neural computational framework for face processing in the human temporal lobe. *Current Biology*. doi:10.1016/j.cub.2025.02.063

Cauchoix M, Barragan-Jason G, Serre T, Barbeau EJ. 2014. The neural dynamics of face detection in the wild revealed by MVPA. *The Journal of Neuroscience* **34**:846–854. doi:10.1523/JNEUROSCI.3030-13.2014

Cavanagh P. 1991. What's up in top-down processing In: Gorea A, editor. Representations of Vision: Trends and Tacit Assumptions in Vision Research. Cambridge University Press.

Chen X, Liu X, Parker BJ, Zhen Z, Weiner KS. 2023. Functionally and structurally distinct fusiform face area(s) in over 1000 participants. *Neuroimage* **265**:119765. doi:10.1016/j.neuroimage.2022.119765

Conner CR, Ellmore TM, DiSano MA, Pieters TA, Potter AW, Tandon N. 2011. Anatomic and electro-physiologic connectivity of the language system: A combined DTI-CCEP study. *Comput Biol Med* **41**:1100–1109. doi:10.1016/j.compbiomed.2011.07.008

Conway BR. 2018. The organization and operation of inferior temporal cortex. *Annu Rev Vis Sci*. doi:10.1146/annurev-vision-091517-034202

Cox RW. 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* **29**:162–173. doi:10.1006/cbmr.1996.0014

Crouzet SM, Thorpe SJ. 2010. Fast saccades toward faces: Face detection in just 100 ms. *J Vis* **10**:(4) 16, 1-17. doi:10.1167/10.4.16.

DiCarlo JJ, Zoccolan D, Rust NC. 2012. How does the brain solve visual object recognition? *Neuron* **73**:415–434. doi:10.1016/j.neuron.2012.01.010

Duchaine B, Yovel G. 2015. A Revised Neural Framework for Face Processing. *Annu Rev Vis Sci* **1**:393–416. doi:10.1146/annurev-vision-082114-035518

Edelman GM. 2004. Biochemistry and the Sciences of Recognition. *Journal of Biological Chemistry*. doi:10.1074/jbc.X400001200

Edelman GM, Gally JA. 2013. Reentry: A key mechanism for integration of brain function. *Front Integr Neurosci*. doi:10.3389/fnint.2013.00063

Eldridge MAG, Pearl JE, Fomani GP, Masseau EC, Fredericks JM, Chen G, Richmond BJ. 2023. Visual recognition in rhesus monkeys requires area TE but not TEO. *Cerebral Cortex* **33**:3098–3106. doi:10.1093/cercor/bhac263

Engell AD, McCarthy G. 2014. Face, eye, and body selective responses in fusiform gyrus and adjacent cortex: an intracranial EEG study. *Front Hum Neurosci* **8**:642. doi:10.3389/fnhum.2014.00642

Fairhall SL, Ishai A. 2007. Effective connectivity within the distributed cortical network for face perception. *Cerebral Cortex* **17**:2400–2406. doi:10.1093/cercor/bhl148

931 Finzi D, Gomez J, Nordt M, Rezai AA, Poltoratski S, Grill-Spector K. 2021. Differential spatial
932 computations in ventral and lateral face-selective regions are scaffolded by structural
933 connections. *Nat Commun* **12**:2278. doi:10.1038/s41467-021-22524-2
934 Freiwald WA. 2020. The neural mechanisms of face processing: cells, areas, networks, and
935 models. *Curr Opin Neurobiol* **60**:184–191. doi:10.1016/j.conb.2019.12.007
936 Fuster J. 1995. Memory in the Cerebral Cortex: An Empirical Approach to Neural Networks in
937 the Human and Nonhuman Primates., MIT Press. Cambridge, MA: MIT Press.
938 Gao X, Gentile F, Rossion B. 2018. Fast periodic stimulation (FPS): a highly effective approach
939 in fMRI brain mapping. *Brain Struct Funct* **223**:2433–2454. doi:10.1007/s00429-018-
940 1630-4
941 Grill-Spector K, Weiner KS. 2014. The functional architecture of the ventral temporal cortex
942 and its role in categorization. *Nat Rev Neurosci* **15**:536–548. doi:doi:10.1038/nrn3747
943 Grill-Spector K, Weiner KS, Kay KN, Gomez J. 2017. The Functional Neuroanatomy of Human
944 Face Perception. *Annu Rev Vis Sci*. doi:10.1146/annurev-vision-102016-061214
945 Hagen S, Jacques C, Ranta R, Koessler L, Maillard L, Colnat-Coulbois S, Rossion B, Jonas J.
946 2025. Intracranial EEG referencing for large-scale category-selective mapping in the
947 human ventral occipito-temporal cortex. *Imaging Neuroscience* **3**.
948 doi:10.1162/imag_a_00479
949 Hauk O, Rice GE, Volfart A, Magnabosco F, Ralph MAL, Rossion B. 2021. Face-selective
950 responses in combined EEG/MEG recordings with fast periodic visual stimulation
951 (FPVS). *Neuroimage* **242**:118460. doi:10.1016/j.neuroimage.2021.118460
952 Hegdé J, Felleman DJ. 2007. Reappraising the functional implications of the primate visual
953 anatomical hierarchy. *Neuroscientist*. doi:10.1177/1073858407305201
954 Hubel DH, Wiesel TN. 1968. Receptive fields and functional architecture of monkey striate
955 cortex. *Journal of Physiology* **195**:215–243.
956 Hubel DH, Wiesel TN. 1962. Receptive Fields, Binocular Interaction and Functional
957 Architecture in the Cat's Visual Cortex. *J Physiol* **160**:106–154.
958 Issa EB, Cadieu CF, DiCarlo JJ. 2018. Neural dynamics at successive stages of the ventral
959 visual stream are consistent with hierarchical error signals. *Elife* **7**:e42870.
960 doi:10.7554/eLife.42870

961 Jacobs GH. 1999. Vision and Behavior in Primates In: Archer SN, Djamgoz MBA, Loew ER,
962 Partridge JC, Vallerga S, editors. Adaptive Mechanisms in the Ecology of Vision.
963 Dordrecht: Springer Netherlands. pp. 629–650. doi:10.1007/978-94-017-0619-3_21
964 Jacques C, Jonas J, Colnat-Coulbois S, Maillard L, Rossion B. 2022. Low and high frequency
965 intracranial neural signals match in the human associative cortex. *Elife* **11**:e76544.
966 doi:10.7554/eLife.76544
967 Jacques C, Retter TL, Rossion B. 2016a. A single glance at natural face images generate larger
968 and qualitatively different category-selective spatio-temporal signatures than other
969 ecologically-relevant categories in the human brain. *Neuroimage* **137**:21–33.
970 doi:10.1016/j.neuroimage.2016.04.045
971 Jacques C, Witthoft N, Weiner KS, Foster BL, Rangarajan V, Hermes D, Miller KJ, Parvizi J,
972 Grill-spector K. 2016b. Corresponding ECoG and fMRI category-selective signals in
973 human ventral temporal cortex. *Neuropsychologia* **83**:14–28.
974 doi:10.1016/j.neuropsychologia.2015.07.024
975 Jiang F, Dricot L, Weber J, Righi G, Tarr MJ, Goebel R, Rossion B. 2011. Face categorization in
976 visual scenes may start in a higher order area of the right fusiform gyrus: Evidence from
977 dynamic visual stimulation in neuroimaging. *J Neurophysiol* **106**:2720–2736.
978 doi:10.1152/jn.00672.2010
979 Jonas J, Jacques C, Liu-shuang J, Brissart H, Colnat-coulbois S, Maillard L, Rossion B. 2016a. A
980 face-selective ventral occipito-temporal map of the human brain with intracerebral
981 potentials. *Proc Natl Acad Sci U S A* **113**:E4088--E4097. doi:10.1073/pnas.1522033113
982 Jonas J, Jacques C, Liu-shuang J, Brissart H, Colnat-coulbois S, Maillard L, Rossion B. 2016b. A
983 face-selective ventral occipito-temporal map of the human brain with intracerebral
984 potentials. *Proc Natl Acad Sci U S A* **113**:E4088--E4097. doi:10.1073/pnas.1522033113
985 Jonas J, Rossion B. 2021. Intracerebral electrical stimulation to understand the neural basis
986 of human face identity recognition. *Eur J Neurosci* **54**:4197–4211.
987 doi:10.1111/ejn.15235
988 Kadipasaoglu CM, Conner CR, Baboyan VG, Rollo M, Pieters TA, Tandon N. 2017. Network
989 dynamics of human face perception. *PLoS One* **12**:e0188834.
990 doi:10.1371/journal.pone.0188834

Kadipasaoglu CM, Conner CR, Whaley ML, Baboyan VG, Tandon N. 2016. Category-Selectivity in Human Visual Cortex Follows Cortical Topology: A Grouped icEEG Study. *PLoS One* **11**:e0157109. doi:10.1371/journal.pone.0157109

Kar K, DiCarlo JJ. 2021. Fast Recurrent Processing via Ventrolateral Prefrontal Cortex Is Needed by the Primate Ventral Stream for Robust Core Visual Object Recognition. *Neuron* **109**:164-176.e5. doi:10.1016/j.neuron.2020.09.035

Kar K, Kubilius J, Schmidt K, Issa EB, DiCarlo JJ. 2019. Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nat Neurosci* **22**:974–983. doi:10.1038/s41593-019-0392-5

Kay KN, Weiner KS, Grill-Spector K. 2015. Attention reduces spatial uncertainty in human ventral temporal cortex. *Current Biology* **25**:595–600. doi:10.1016/j.cub.2014.12.050

Kim M, Ducros M, Carlson T, Ronen I, He S, Ugurbil K, Kim DS. 2006. Anatomical correlates of the functional organization in the human occipitotemporal cortex. *Magn Reson Imaging* **24**:583–590. doi:10.1016/j.mri.2005.12.005

Kosakowski HL, Cohen MA, Takahashi A, Keil B, Kanwisher N, Saxe R. 2022. Selective responses to faces, scenes, and bodies in the ventral visual pathway of infants. *Current Biology* **32**:265-274.e5. doi:10.1016/j.cub.2021.10.064

Kravitz DJ, Saleem KS, Baker CI, Ungerleider LG, Mishkin M. 2013. The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends Cogn Sci* **17**:26–49. doi:10.1016/j.tics.2012.10.011

Kubota E, Yan X, Tung S, Fascendini B, Tyagi C, Duhameau S, Ortiz D, Grotheer M, Natu VS, Keil B, Grill-Spector K. 2025. White matter connections of human ventral temporal cortex are organized by cytoarchitecture, eccentricity and category-selectivity from birth. *Nat Hum Behav*. doi:10.1038/s41562-025-02116-6

Lamme VAF, Roelfsema PR. 2000. The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci* **23**:571–579.

Laurent M-A, Jacques C, Yan X, Jurczynski P, Colnat-Coulbois S, Maillard L, Cam S Le, Ranta R, Cottereau BR, Koessler L, Jonas J, Rossion B. 2025. A tight relationship between BOLD fMRI activation/deactivation and increase/decrease in single neuron responses in human association cortex. doi:10.7554/elife.104779.1

Liu H, Agam Y, Madsen JR, Kreiman G. 2009. Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron* **62**:281–90. doi:10.1016/j.neuron.2009.02.025

Mahon BZ. 2022. Domain-specific connectivity drives the organization of object knowledge in the brainHandbook of Clinical Neurology. Elsevier B.V. pp. 221–244. doi:10.1016/B978-0-12-823493-8.00028-6

Manning JR, Jacobs J, Fried I, Kahana MJ. 2009. Broadband shifts in local field potential power spectra are correlated with single-neuron spiking in humans. *Journal of Neuroscience* **29**:13613–13620. doi:10.1523/JNEUROSCI.2041-09.2009

Marr D. 1982. Vision: A Computational Investigation into the Human Representation and Processing of Visual Information, W. H. Free. ed. San Francisco, CA, USA: W. H. Freeman and Co.

Mumford D. 1992. On the computational architecture of the neocortex II The role of cortico-cortical loops, Biol. Cybern.

Norcia AM, Appelbaum LG, Ales JM, Cottereau BR, Rossion B. 2015. The steady-state visual evoked potential in vision research: A review. *J Vis* **15**:(6) 4, 1--46. doi:10.1167/15.6.4

Nowak LG, Bullier J. 1997. The Timing of Information Transfer in the Visual System In: Rockland KS, Kaas JH, Peters A, editors. Extrastriate Cortex in Primates. Boston, MA: Springer US. pp. 205–241. doi:10.1007/978-1-4757-9625-4_5

Ojemann JG, Akbudak E, Snyder AZ, McKinstry RC, Raichle ME, Conturo TE. 1997. Anatomic localization and quantitative analysis of gradient refocused echo-planar fMRI susceptibility artifacts. *Neuroimage* **6**:156–167. doi:10.1006/nimg.1997.0289

Perrett DI, Oram MW, Ashbridge E. 1998. Evidence accumulation in cell populations responsive to faces: an account of generalisation of recognition without mental transformations. *Cognition* **67**:111–145.

Persichetti AS, Denning JM, Gotts SJ, Martin A. 2021. A data-driven functional mapping of the anterior temporal lobes. *Journal of Neuroscience* **41**:6038–6049. doi:10.1523/JNEUROSCI.0456-21.2021

Quek GL, Rossion B. 2017. Category-selective human brain processes elicited in fast periodic visual stimulation streams are immune to temporal predictability. *Neuropsychologia* **104**:182–200. doi:10.1016/j.neuropsychologia.2017.08.010

1052 Quian Quiroga R, Boscaglia M, Jonas J, Rey HG, Yan X, Maillard L, Colnat-Coulbois S, Koessler
1053 L, Rossion B. 2023. Single neuron responses underlying face recognition in the human
1054 midfusiform face-selective cortex. *Nat Commun* **14**:5661. doi:10.1038/s41467-023-
1055 41323-5
1056 Rangarajan V, Hermes D, Foster BL, Weiner KS, Jacques C, Grill-Spector K, Parvizi J. 2014.
1057 Electrical Stimulation of the Left and Right Human Fusiform Gyrus Causes Different
1058 Effects in Conscious Face Perception. *The Journal of Neuroscience* **34**:12828–12836.
1059 doi:10.1523/JNEUROSCI.0527-14.2014
1060 Regan D. 1989. Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic
1061 Fields in Science and Medicine, Elsevier. ed. New York, NY, USA.
1062 Retter TL, Jiang F, Webster MA, Rossion B. 2020. All-or-none face categorization in the
1063 human brain. *Neuroimage* **213**:116685. doi:10.1016/j.neuroimage.2020.116685
1064 Retter TL, Rossion B. 2016. Uncovering the neural magnitude and spatio-temporal dynamics
1065 of natural image categorization in a fast visual stream. *Neuropsychologia* **91**:9–28.
1066 doi:10.1016/j.neuropsychologia.2016.07.028
1067 Riesenhuber M, Poggio T. 1999. Hierarchical models of object recognition in cortex. *Nat*
1068 *Neurosci* **2**:1019–1025.
1069 Rossion B. 2022. Twenty years of investigation with the case of prosopagnosia PS to
1070 understand human face identity recognition.Part II: Neural basis. *Neuropsychologia*
1071 **173**:108279. doi:10.1016/j.neuropsychologia.2022.108279
1072 Rossion B, Caharel S. 2011. ERP evidence for the speed of face categorization in the human
1073 brain: Disentangling the contribution of low-level visual cues from face perception.
1074 *Vision Res* **51**:1297–1311. doi:10.1016/j.visres.2011.04.003
1075 Rossion B, Caldara R, Seghier M, Schuller AM, Lazeyras F, Mayer E. 2003. A network of
1076 occipito-temporal face-sensitive areas besides the right middle fusiform gyrus is
1077 necessary for normal face processing. *Brain* **126**:2381–2395. doi:10.1093/brain/awg241
1078 Rossion B, Hanseeuw B, Dricot L. 2012. Defining face perception areas in the human brain: a
1079 large-scale factorial fMRI face localizer analysis. *Brain Cogn* **79**:138–157.
1080 doi:10.1016/j.bandc.2012.01.001
1081 Rossion B, Jacques C, Jonas J. 2024. The anterior fusiform gyrus: The ghost in the cortical
1082 face machine. *Neurosci Biobehav Rev*.
1083 doi:https://doi.org/10.1016/j.neubiorev.2024.105535

1084 Rossion B, Jacques C, Jonas J. 2018. Mapping face categorization in the human ventral
1085 occipito-temporal cortex with direct neural intracranial recordings. *Ann N Y Acad Sci*
1086 **1426**:5–24. doi:10.1111/nyas.13596
1087 Rossion B, Jacques C, Liu-shuang J. 2015. Fast periodic presentation of natural images reveals
1088 a robust face-selective electrophysiological response in the human brain. *J Vis* **15**:(1) 18,
1089 1--18. doi:10/1167.15.1.18.
1090 Rousselet G, Husk JS, Bennett PJ, Sekuler AB. 2008. Time course and robustness of ERP
1091 object and face differences. *J Vis* **8**:(12) 3, 1-18. doi:10.1167/8.12.3.
1092 Sadeh B, Podlipsky I, Zhdanov A, Yovel G. 2010. Event-related potential and functional MRI
1093 measures of face-selectivity are highly correlated: a simultaneous ERP-fMRI
1094 investigation. *Hum Brain Mapp* **31**:1490–1501. doi:10.1002/hbm.20952
1095 Schmolesky MT, Wang YC, Hanes DP, Thompson KG, Leutgeb S, Schall JD, Leventhal AG.
1096 1998. Signal timing across the macaque visual system. *J Neurophysiol* **79**:3272–3278.
1097 Schrouff J, Raccah O, Baek S, Rangarajan V, Salehi S, Mourão-Miranda J, Helili Z, Daitch AL,
1098 Parvizi J. 2020. Fast temporal dynamics and causal relevance of face processing in the
1099 human temporal cortex. *Nat Commun* **11**:656. doi:10.1038/s41467-020-14432-8
1100 Schweinberger SR, Neumann MF. 2016. Repetition effects in human ERPs to faces. *Cortex*
1101 **80**:141–153. doi:10.1016/j.cortex.2015.11.001
1102 Serre T, Oliva A, Poggio T. 2007. A feedforward architecture accounts for rapid
1103 categorization. *Proceedings of the National Academy of Sciences* **104**:6424–6429.
1104 doi:10.1073/pnas.0700622104
1105 Steeves JK, Culham JC, Duchaine BC, Pratesi CC, Valyear KF, Schindler I, Humphrey GK, Milner
1106 AD, Goodale MA. 2006. The fusiform face area is not sufficient for face recognition:
1107 evidence from a patient with dense prosopagnosia and no occipital face area.
1108 *Neuropsychologia* **44**:594–609.
1109 Talairach J, Bancaud J. 1973. Stereotaxic approach to epilepsy. Methodology of anatomo-
1110 functional stereotaxic investigations. *Prog Neurol Surg* **5**:297–354.
1111 doi:https://doi.org/10.1159/000394343
1112 Thorpe SJ, Fabre-Thorpe M. 2001. Seeking Categories in the Brain. *Science (1979)* **291**:260–
1113 263. doi:10.1126/science.1058249
1114 Tsao D. 2014. The macaque face patch system: A window into object representation. *Cold*
1115 *Spring Harb Symp Quant Biol* **79**:109–114. doi:10.1101/sqb.2014.79.024950

Tsao DY, Moeller S, Freiwald W a. 2008. Comparing face patch systems in macaques and humans. *Proc Natl Acad Sci U S A* **105**:19514–19519. doi:10.1073/pnas.0809662105

Ungerleider LG, Bell AH. 2011. Uncovering the visual " alphabet": Advances in our understanding of object perception. *Vision Res*. doi:10.1016/j.visres.2010.10.002

van Blooijs D, van den Boom MA, van der Aar JF, Huiskamp GM, Castegnaro G, Demuru M, Zweiphenning WJEM, van Eijsden P, Miller KJ, Leijten FSS, Hermes D. 2023. Developmental trajectory of transmission speed in the human brain. *Nat Neurosci* **26**:537–541. doi:10.1038/s41593-023-01272-0

Van Essen DC, Anderson CH, Felleman DJ. 1992. Information processing in the primate visual system: an integrated systems perspective. *Science (1979)* **255**:419–23.

Van Essen DC, Maunsell JHR. 1983. Hierarchical organization and functional streams in the visual cortex. *Trends Neurosci* **6**:370–375. doi:https://doi.org/10.1016/0166-2236(83)90167-4

Vidal JR, Ossandón T, Jerbi K, Dalal SS, Minotti L, Ryvlin P, Kahane P, Lachaux J-P. 2010. Category-Specific Visual Responses: An Intracranial Study Comparing Gamma, Beta, Alpha, and ERP Response Selectivity. *Front Hum Neurosci* **4**:195. doi:10.3389/fnhum.2010.00195

Volfart A, Yan X, Maillard L, Colnat-Coulbois S, Hossu G, Rossion B, Jonas J. 2022. Intracerebral electrical stimulation of the right anterior fusiform gyrus impairs human face identity recognition. *Neuroimage* **250**:118932. doi:10.1016/j.neuroimage.2022.118932

Wang Y, Metoki A, Smith D V., Medaglia JD, Zang Y, Benear S, Popal H, Lin Y, Olson IR. 2020. Multimodal mapping of the face connectome. *Nat Hum Behav* **4**:397–411. doi:10.1038/s41562-019-0811-3

Weiner KS, Grill-Spector K. 2010. Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. *Neuroimage* **52**:1559–1573. doi:10.1016/j.neuroimage.2010.04.262

Weiner KS, Jonas J, Gomez J, Maillard L, Brissart H, Hossu G, Jacques C, Loftus D, Colnat-Coulbois S, Stigliani A, Barnett MA, Grill-Spector K, Rossion B. 2016. The Face-Processing Network Is Resilient to Focal Resection of Human Visual Cortex. *The Journal of neuroscience* **36**:8425–8440. doi:10.1523/JNEUROSCI.4509-15.2016

Zeki S, Shipp S. 1988. The functional logic of cortical connections. *Nature* **335**:311–317. doi:10.1038/335311a0

Supporting Information for :

Concurrent category-selective neural activity across the ventral occipito-temporal cortex supports a non-hierarchical view of human visual recognition

Corentin Jacques[1], Jacques Jonas[1,2] , Sophie Colnat-Coulbois[3], Bruno Rossion*[1,2]

Université de Lorraine, CNRS, IMoPA, F-54000 Nancy, France

Université de Lorraine, CHRU-Nancy, Service de Neurologie, F-54000 Nancy, France

Université de Lorraine, CHRU-Nancy, Service de Neurochirurgie, F-54000 Nancy, France

@mail: bruno.rossion@univ-lorraine.fr

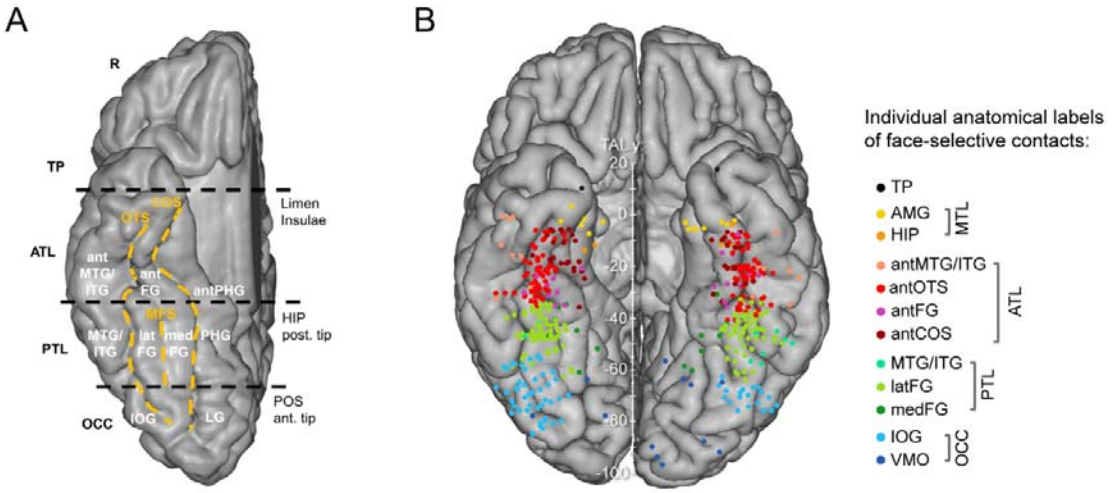
This file includes:

Figures S1 to S7

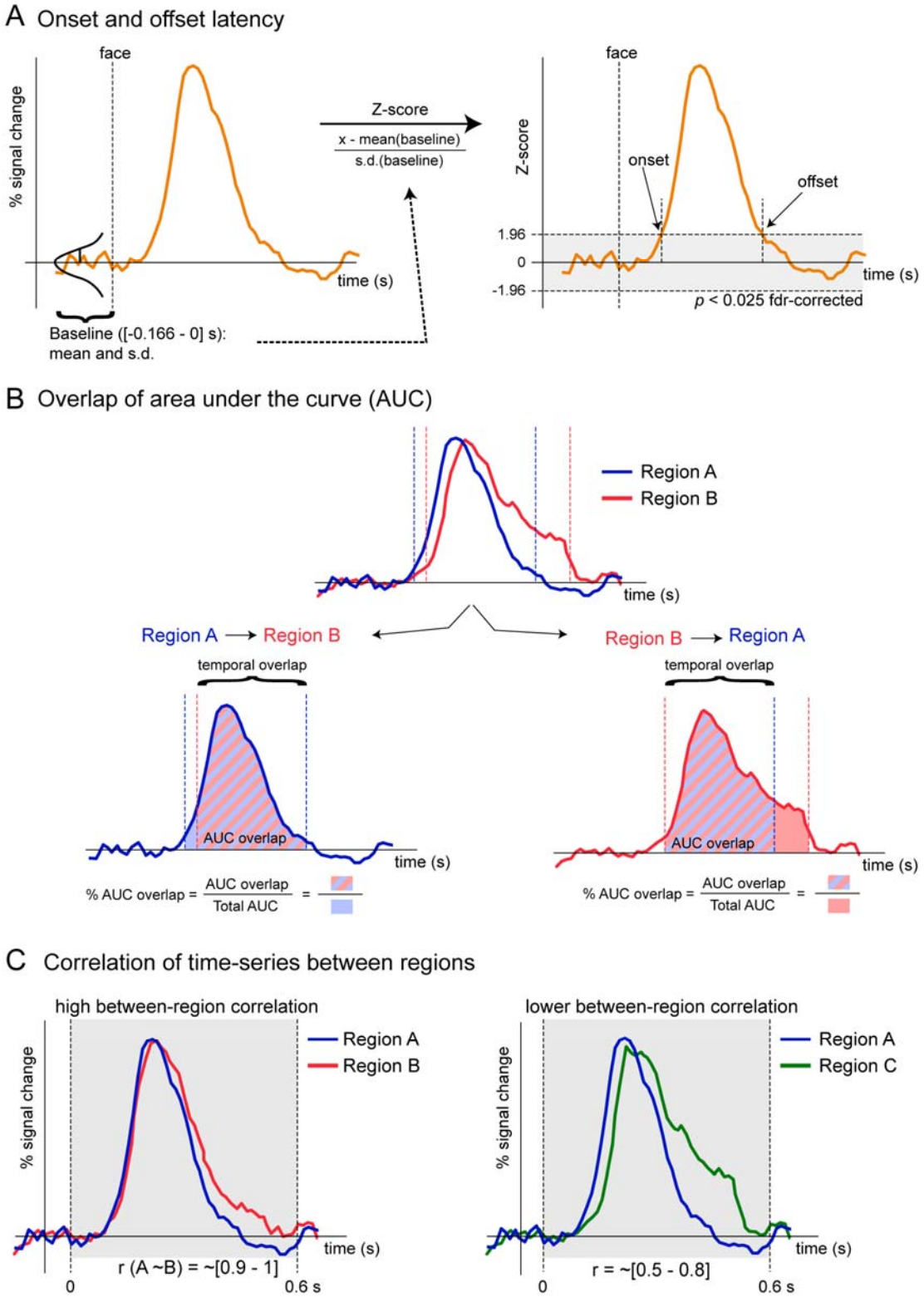
Tables S1

	Face-selective contacts : response increase	
	Left hemisphere	**Right hemisphere**
VMO	**9 (5)**	**3 (3)**
IOG	**19 (10)**	**46 (12)**
medFG	**7 (6)**	**7 (6)**
latFG	**58 (21)**	**65 (25)**
MTG/ITG	**7 (6)**	**2 (2)**
antCOS	**18 (13)**	**26 (16)**
antFG	**10 (9)**	**10 (8)**
antOTS	**63 (29)**	**57 (34)**
antMTG/ITG	**7 (6)**	**8 (6)**
AMG	**8 (6)**	**6 (6)**

Supplementary Table 1: Number of contacts showing significant face-selective response increase in each anatomical region. The corresponding number of participants in which these contacts were found is indicated in parenthesis. Acronyms: VMO: ventro-medial occipital cortex; IOG: inferior occipital gyrus; medFG: medial fusiform gyrus and collateral sulcus; latFG: lateral fusiform gyrus and occipito-temporal sulcus; MTG/ITG: the inferior and middle temporal gyri; antCoS: anterior collateral sulcus; antOTS: anterior occipito-temporal sulcus; antFG: anterior fusiform gyrus; antMTG/ITG: anterior middle and inferior temporal gyri.

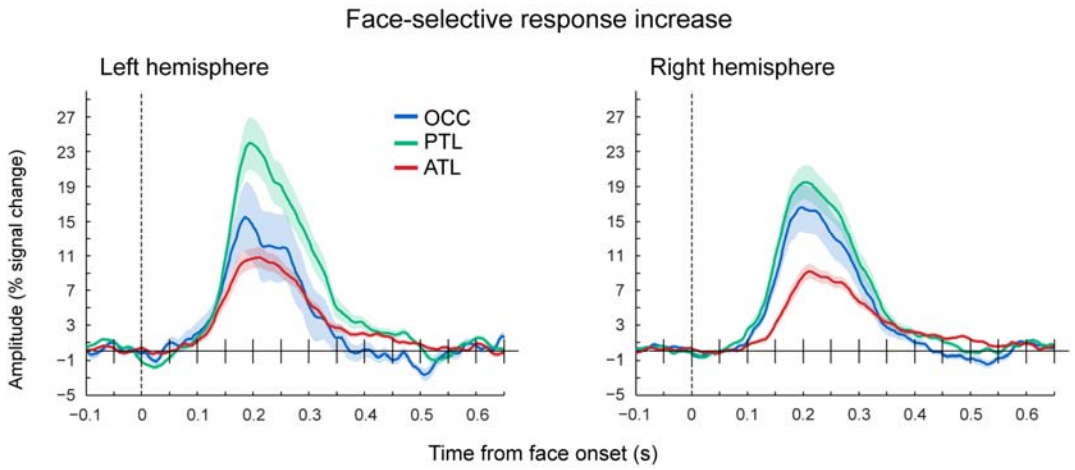


Supplementary Figure 1: VOTC anatomical parcellation scheme and face-selective recording contacts anatomical labels. A. Anatomical regions were defined in each individual hemisphere according to major anatomical landmarks. The ventral temporal sulci (COS, OTS, and midfusiform sulcus, i.e., MFS) serve as medial/lateral borders of regions, whereas three coronal reference planes containing anatomical landmarks (posterior tip of the hippocampus, i.e., HIP, anterior tip of the parieto-occipital sulcus, i.e., POS, limen insulae) serve as an anterior/posterior boundary for each region. We considered contacts in the ATL if they were located anteriorly to the posterior tip of the hippocamps and posteriorly to the limen insulae. The schematic locations of these anatomical structures are shown on a reconstructed cortical surface of the Colin27 brain. Acronyms: TP: temporal pole; ATL: anterior temporal lobe; PTL: posterior temporal lobe; OCC: occipital lobe; PHG: parahippocampal gyrus; CoS: collateral sulcus; FG: fusiform gyrus; ITG: inferior temporal gyrus; MTG: middle temporal gyrus; OTS: occipito-temporal sulcus; CS: calcarine sulcus; IOG: inferior occipital gyrus; LG: lingual gyrus; ant: anterior; lat: lateral; med: medial. **B.** Map of all face-selective recording contacts and displayed in the Talairach space. Each circle represents a single face-selective contact color-coded according to its anatomical location in the original individual anatomy (see legend on the right).

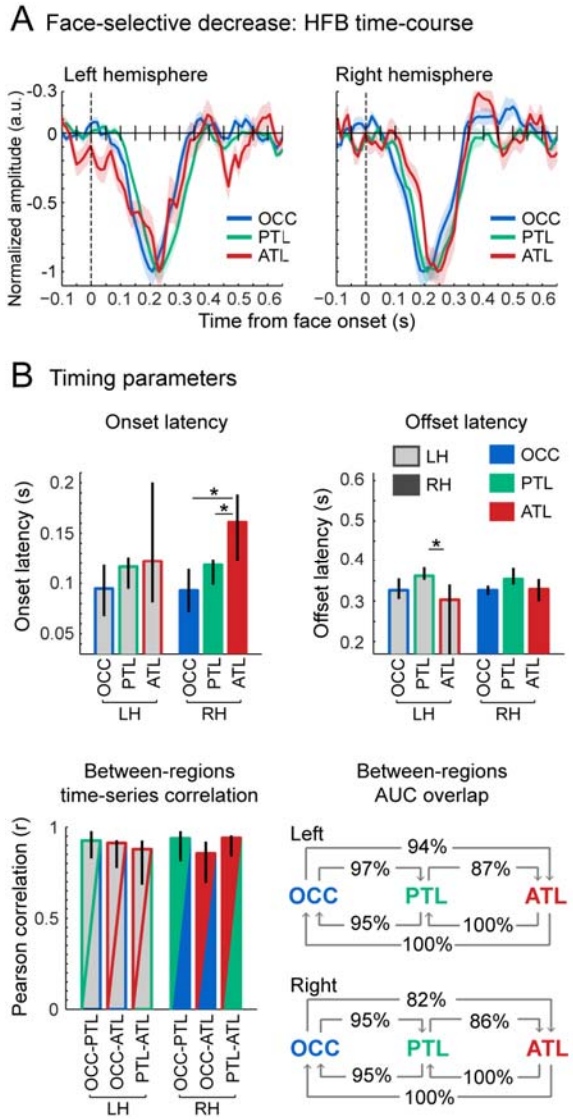


Supplementary Figure 2: Parameters for timing analyses. The timing of face-selective activity was characterized using 4 parameters: (1) latency of response onset and (2) offset, as well as (3) response overlap and (4) response correlation across VOTC regions. **A**. The first two parameters, onset and offset latencies of face-selective response, were computed on

HFB time-series averaged by groups of contacts (left). Averaged HFB time-series were converted to Z-score values by subtracting the mean amplitude in the baseline window of the time-series (i.e. before face onset: [-0.166 to 0s]) and dividing by the standard deviation of the amplitude in the same baseline window (right). Z-score values were further converted to *p*-values and corrected for multiple comparison using FDR. Onset latency was defined as the first time-point at which the $p < 0.05$ (two-tailed, corresponding to a z-score of 1.96), for at least 30 ms. Offset latency was the first time bin (after the onset) at which the response was no longer significantly different from baseline. **B**. A third parameter estimated the overlap between time-series for pairs of regions (e.g. regions A and B). The overlap is asymmetrical and is calculated separately for region A and region B as the ratio between the area under the curve (AUC) of the overlap between regions A and region B (i.e. summing the amplitude between the maximum of onset A and B and the minimum of offset A and B) and the total AUC for region A (for overlap of region A to B) or for region B (for overlap of region B to A). **C**. As the fourth timing parameter, we computed Pearson correlations between the time-series of pairs of regions using data between 0 and 0.6 s relative to face onset (gray shaded area). This provided an estimate of the similarity between the response function of the two regions compared. The shape of the response is more similar between region A and B than between region A and C. Between region correlations were compared against within-region correlations.

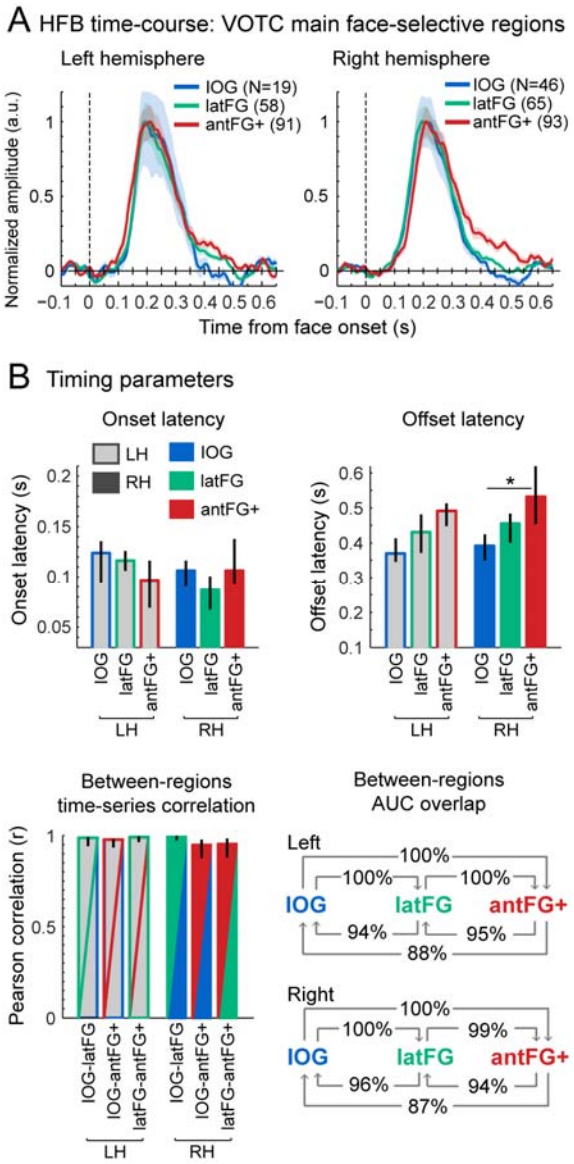


Supplementary Figure 3: non-normalized time-courses. Time-domain face-selective HFB activity in contacts showing response increase, averaged by main VOTC region (OCC, PTL, ATL) and hemisphere. HFB time-series were filtered to remove the general visual response at 6 Hz and harmonics so only face-selective signal remains. Time-series are original non-normalized versions of what appears in Figure 4 of the main paper. Shaded area represents the standard error of the mean between contacts.



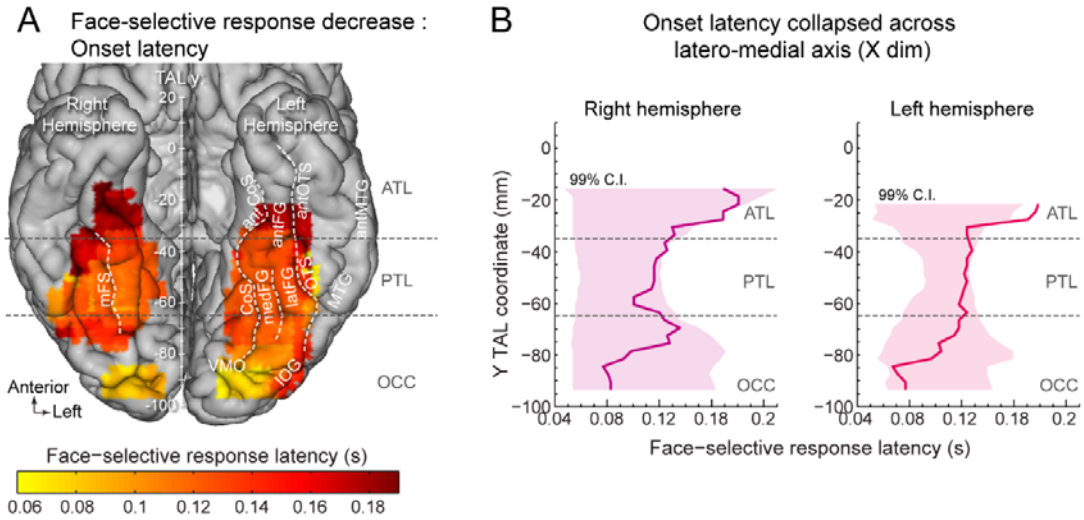
Supplementary Figure 4: Response timing in signal- contacts. A. Time-domain face-selective HFB activity in signal- contacts averaged by main VOTC region (OCC, PTL, ATL) and hemisphere. HFB time-series were filtered to remove the general visual response at 6 Hz and harmonics so only face-selective signal remains. The maximum amplitude of each averaged waveform was normalized to -1 for visualization purposes only (see Figure S3 for non-normalized waveforms). Shaded area represents the standard error of the mean between contacts. **B**. Timing parameters split by VOTC region and hemisphere: onset (top left) and offset (top right) latency of face-selective response, correlation of time-series between regions (bottom left) and between-region area under the curve (AUC) overlap (bottom right). For the AUC overlap, the arrow shows the directionality of the computed overlap. For instance, the arrow from ATL to OCC indicates that we are describing the percentage of the total AUC of ATL (measured between onset and offset latencies) occupied by the AUC of its temporal overlap (also determined using onset and offset) with OCC. Acronyms: LH: left hemisphere, RH: right hemisphere. Onset latencies were not statistically different across

VOTC main regions in the left hemisphere (OCC: 95 ms, [67-118] ms; PTL: 116 ms, [95-126] ms; ATL: 122 ms, [81-200] ms; OCC vs. PTL: p = 0.31; OCC vs. ATL: p = 0.48; PTL vs. ATL: p = 0.97, two-tailed permutation test, fdr-corrected), latencies in the right hemisphere were significantly higher in the ATL compared to PTL and OCC (OCC: 93 ms, [71-114] ms; PTL: 118 ms, [99-124] ms; ATL: 161 ms, [122-188] ms; OCC vs. PTL: p = 0.29; OCC vs. ATL: p = 0.005; PTL vs. ATL: p = 0.043, fdr-corrected). Offset latencies in the left hemisphere were slightly, but significantly, later in the PTL (362 ms, [352-384] ms) compared to ATL (304 ms, [155-341 ms], p = 0.045), and to OCC (325 ms, [306-356 ms], p = 0.046) . No difference in offset latencies were found across regions in the right hemisphere (p's > 0.11). The temporal overlap between regions computed using AUC was on average of 94%, ranging from 82% (right OCC AUC occupied by the AUC of its temporal overlap with the right ATL) to 100% (right and left ATL in PTL, right and left ATL in OCC). Correlating time-series across regions reveals that the lowest correlation was between right OCC to ATL (Pearson's r=0.85, also lowest overlap in the AUC measure) and the highest was between right PTL and ATL (r=0.95), indicating that time-series across regions share 72% to 90% of their variance. Comparing within- to between-region correlations revealed only a small borderline difference when comparing OCC to PTL in the left hemisphere (within OCC and within PTL: r=0.97 and 0.98 respectively, between OCC and PTL: 0.91, p = 0.05, one-tailed percentile boostrap test, fdr-corrected), and OCC to ATL in the right hemisphere (within OCC and within ATL: r=0.96 and 0.91 respectively, between OCC and ATL: r=0.83, p = 0.05).

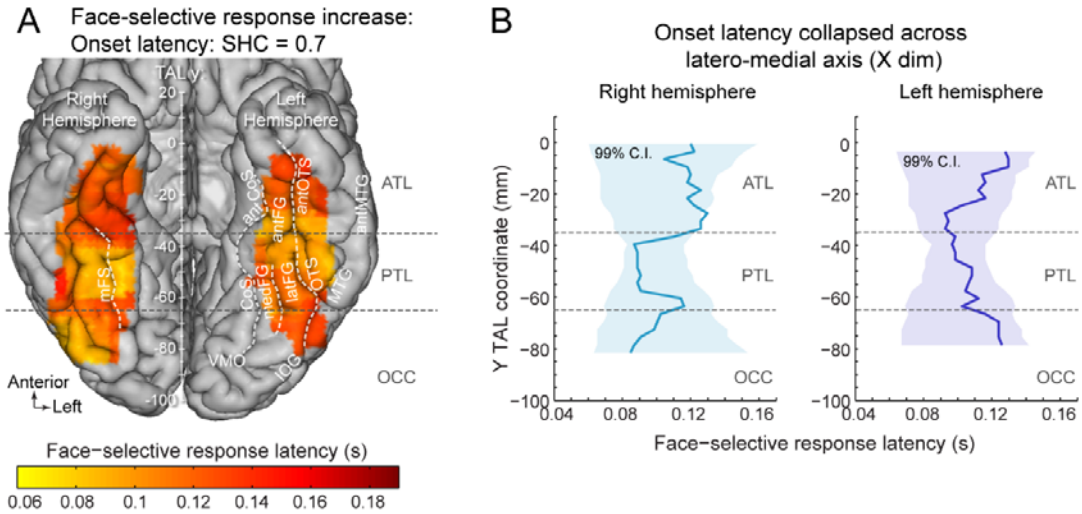


Supplementary Figure 5: Response timing in face-selective contacts in main VOTC face-selective regions. Response timing in face-selective contacts showing response increase in main VOTC face-selective regions: Inferior Occipital Gyrus (IOG, (N=65), lateral Fusiform Gyrus and OTS (latFG, N=123) and antFG+ (anterior fusiform Gyrus, Anterior Occipitotemporal Sulcus, Anterior Collateral Sulcus, N=184). **A.** Time-domain face-selective HFB activity averaged by main face-selective region and hemisphere. HFB time-series were filtered to remove the general visual response at 6 Hz and harmonics so only face-selective signal remains. The maximum amplitude of each averaged waveform was normalized to 1 for visualization purposes only. Shaded area represents the standard error of the mean between contacts. **B**. Timing parameters split by VOTC region and hemisphere: onset (top

left) and offset (top right) latency of face-selective response, correlation of time-series between regions (bottom left) and between-region area under the curve (AUC) overlap (bottom right). For the AUC overlap, the arrow shows the directionality of the computed overlap. For instance, the arrow from ATL to OCC indicates that we are describing the percentage of the total AUC of ATL (measured between onset and offset latencies) occupied by the AUC of its temporal overlap (also determined using onset and offset) with OCC. Acronyms: LH: left hemisphere, RH: right hemisphere. Onset latencies of face-selective activity were not significantly different across regions either in the left hemisphere (IOG: 124 ms, [96-136] ms; latFG: 116 ms, [104-126] ms; antFG+: 97 ms, [69-116] ms; *p*s > 0.43 for all comparisons, two-tailed permutation test, fdr-corrected, **Figure S5B**) or in the right hemisphere (IOG: 106 ms, [91-114] ms; latFG: 87 ms, [67-100] ms; antFG+: 106 ms, [93-138] ms; *p*s > 0.43 for all comparisons, two-tailed permutation test, fdr-corrected, **Figure S5A**). The lack of significant difference in onset latencies of face-selective activity was confirmed when collapsing responses across hemispheres, being virtually identical between core face-selective regions (IOG: 106 ms, [95-118] ms; latFG: 100 ms, [95-108] ms; antFG+: 99 ms, [81-110] ms; *p*s > 0.7 for all comparisons, two-tailed permutation test, fdr-corrected). Offset latencies increased from the IOG (left: 370 ms, 95% confidence interval: [347-413] ms; right: 391 ms, [349-425] ms), to the latFG (left: 431 ms, [372-481] ms; right: 456, [401-483] ms) and antFG+ (left: 491 ms, [448-514] ms; right: 532 ms, [456-620] ms) but were not significantly different, unlike when collapsing across the two hemispheres (IOG vs. latFG: p = 0.031; latFG vs. antFG+: p < 0.005; IOG vs. antFG+: p < 0.005). Despite the difference in offset latency, there was a large temporal overlap between regions (computed either using AUC, ranging from 87% of the right antFG+ AUC included within its temporal overlap with the IOG, to 99-100% of each region with more anterior regions, or using time-series correlations, ranging from r = 0.95 to 0.99) indicating that most of the neural response occurs simultaneously across these 3 face-selective regions.



Supplementary Figure 6: VOTC map of face-selective onset latency in contacts showing face-selective response decrease. A. Face-selective onset latency was computed in 20 x 20 mm voxels by averaging the time-series of the signal- contacts in each voxel before computing onset latency. To reduce noise in the latency estimates, contacts with split half correlation (SHC) lower than r=0.4 were rejected from this analysis. **B.** Variation of face-selective response latency in signal+ contacts along the postero-anterior axis. Each data point represent the onset latency measured from the time-series averaged over contacts collapsed across each hemisphere (i.e. collapsed across the medio-lateral x dimension) within 20 mm segments (in the y dimension). Thick lines are estimated onset latencies and shaded areas show the 99% confidence intervals expected under the null hypothesis that the postero-anterior location has no influence on the onset latency.



Supplementary Figure 7: VOTC map of face-selective onset latency in high SNR contacts showing face-selective response increase. VOTC map of face-selective onset latency in contacts showing face-selective response increase, computed when excluding contacts with split half correlation (SHC) lower than r=0.7 to further limit noise in onset latency estimates. **A**. Face-selective onset latency was computed in 20 x 20 mm voxels by averaging the time-series of the signal- contacts in each voxel before computing onset latency. To reduce noise in the latency estimates, contacts with split half correlation (SHC) lower than r=0.4 were rejected from this analysis. **B.** Variation of face-selective response latency in signal+ contacts along the postero-anterior axis. Each data point represent the onset latency measured from the time-series averaged over contacts collapsed across each hemisphere (i.e. collapsed across the medio-lateral x dimension) within 20 mm segments (in the y dimension). Thick lines are estimated onset latencies and shaded areas show the 99% confidence intervals expected under the null hypothesis that the postero-anterior location has no influence on the onset latency.