THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Multiple adjoining word- and face-selective regions in ventral temporal cortex exhibit distinct dynamics

Citation for published version:

Boring, M, Silson, E, Ward, M, Richardson, M, Fiez, J, Baker, C & Ghuman, AS 2021, 'Multiple adjoining word- and face-selective regions in ventral temporal cortex exhibit distinct dynamics', *The Journal of Neuroscience*, vol. 41, no. 29, pp. 6314-6327. <https://doi.org/10.1523/JNEUROSCI.3234-20.202>

Digital Object Identifier (DOI):

[10.1523/JNEUROSCI.3234-20.202](https://doi.org/10.1523/JNEUROSCI.3234-20.202)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

The Journal of Neuroscience

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Research Articles: Behavioral/Cognitive

Multiple adjoining word- and face-selective regions in ventral temporal cortex exhibit distinct dynamics

<https://doi.org/10.1523/JNEUROSCI.3234-20.2021>

Cite as: J. Neurosci 2021; 10.1523/JNEUROSCI.3234-20.2021

Received: 29 December 2020

Revised: 26 May 2021

Accepted: 1 June 2021

This Early Release article has been peer-reviewed and accepted, but has not been through the composition and copyediting processes. The final version may differ slightly in style or formatting and will contain links to any extended data.

Alerts: Sign up at www.jneurosci.org/alerts to receive customized email alerts when the fully formatted version of this article is published.

1 **Title:** Multiple adjoining word- and face-selective regions in ventral temporal cortex exhibit
 2 distinct dynamics

3 **Abbreviated title:** Dynamic word- and face-selective networks

4 **Authors:** Matthew J. Boring^{*1,2,3}, Edward H. Silson^{4,5}, Michael J. Ward³, R. Mark
 5 Richardson^{3,6,7}, Julie A. Fiez^{1,2,8}, Chris I. Baker⁴, Avniel Singh Ghuman^{1,2,3,8}

6 1. Center for Neuroscience at the University of Pittsburgh, University of Pittsburgh, A210
 7 Langley Hall, Pittsburgh, PA 15260, USA.

8 2. Center for the Neural Basis of Cognition, 4400 Fifth Avenue, Pittsburgh, PA 15213.

9 3. Department of Neurological Surgery, University of Pittsburgh Medical Center, 200 Lothrop
 10 Street, Pittsburgh, PA 15213, USA.

11 4. National Institute of Mental Health, National Institutes of Health, Magnuson Clinical Center,
 12 Bethesda, MD 20814 USA;

13 5. School of Philosophy, Psychology and Language Sciences, University of Edinburgh, 7
 14 George Square, Edinburgh EH8 9JZ, UK;

15 6. Department of Neurosurgery, Massachusetts General Hospital, 55 Fruit Street, Boston, MA
 16 02144, USA.

17 7. Harvard Medical School, 25 Shattuck St., Boston, MA 02115, USA.

18 8. Department of Psychology, University of Pittsburgh, 210 South Bouquet St, Pittsburgh, PA
 19 15260, USA.

20 **Corresponding author email address:** mjb200[at]pitt.edu

21 **Number of pages:** 52

22 **Number of Figures:** 8, **Number of tables:** 1

23 **Number of words** (total: 12,543, abstract: 250, introduction: 649, discussion: 1499)

24 **Conflict of interest statement:** The authors declare no competing financial interests.

25 **Acknowledgements:** We would like to thank the patients, their families, and the clinical staff at

26 the Epilepsy Monitoring Unit at the University of Pittsburgh Medical Center, without whom this

27 study would not be possible. We would also like to thank Marlene Behrmann, David Plaut,

28 Michael Tarr, and the anonymous reviewers for helpful suggestions regarding the analysis and

29 interpretations of the intracranial EEG results; Sean Walls, Ellyanna Kessler, Roma Konecky,

30 and Ashley Whiteman for assistance in intracranial EEG data collection. This study was

31 supported by the National Institutes of Health under R01MH107797 and R21EY030297 and

32 National Science Foundation under 1734907 to A.S.G., National Institutes of Health under

33 T32NS007433-20 to M.J.B., and National Institutes of Health under ZIAMH002909 to C.I.B.

34 and E.H.S.

35 **Abstract**

36 The map of category-selectivity in human ventral temporal cortex (VTC) provides organizational
37 constraints to models of object recognition. One important principle is lateral-medial response
38 biases to stimuli that are typically viewed in the center or periphery of the visual field. However,
39 little is known about the relative temporal dynamics and location of regions that respond
40 preferentially to stimulus classes that are centrally viewed, like the face- and word-processing
41 networks. Here, word- and face-selective regions within VTC were mapped using intracranial
42 recordings from 36 patients. Partially overlapping, but also anatomically dissociable patches of
43 face- and word-selectivity were found in VTC. In addition to canonical word-selective regions
44 along the left posterior occipitotemporal sulcus, selectivity was also located medial and anterior
45 to face-selective regions on the fusiform gyrus at the group level and within individual male and
46 female subjects. These regions were replicated using 7 Tesla fMRI in healthy subjects. Left
47 hemisphere word-selective regions preceded right hemisphere responses by 125 ms, potentially
48 reflecting the left hemisphere bias for language; with no hemispheric difference in face-selective
49 response latency. Word-selective regions along the posterior fusiform responded first, then
50 spread medially and laterally, then anteriorly. Face-selective responses were first seen in
51 posterior fusiform regions bilaterally, then proceeded anteriorly from there. For both words and
52 faces, the relative delay between regions was longer than would be predicted by purely
53 feedforward models of visual processing. The distinct time-courses of responses across these
54 regions, and between hemispheres, suggest a complex and dynamic functional circuit supports
55 face and word perception.

56 **Significance Statement:** Representations of visual objects in the human brain have been shown
57 to be organized by several principles, including whether those objects tend to be viewed centrally
58 or peripherally in the visual field. However, it remains unclear how regions that process objects
59 that are viewed centrally, like words and faces, are organized relative to one another. Here,
60 invasive and non-invasive neuroimaging suggests there is a mosaic of regions in ventral temporal
61 cortex that respond selectively to either words or faces. These regions display differences in the
62 strength and timing of their responses, both within and between brain hemispheres, suggesting
63 they play different roles in perception. These results illuminate extended, bilateral, and dynamic
64 brain pathways that support face perception and reading.

65 **Introduction**

66 Investigations into the spatial organization of category-selectivity in ventral temporal cortex
 67 (VTC) have been instrumental in establishing several organizational principles of the visual
 68 system. Functional magnetic resonance imaging (fMRI) studies have helped identify lateral-
 69 medial biases in ventral stream responses to objects depending on where they typically appear in
 70 the visual field (retinotopic eccentricity) (Hasson et al., 2002; Konkle and Caramazza, 2013;
 71 Grill-Spector and Weiner, 2014). Specifically, lateral regions of VTC are selective for objects
 72 that tend to be viewed centrally (foveated), like words and faces, whereas more medial regions
 73 are selective for objects that tend to fall on the periphery of the retina, like navigationally
 74 relevant information such as buildings (Haxby et al., 1996; Aguirre et al., 1998; Cohen et al.,
 75 2000; Hasson et al., 2002). This broad principle of organization by eccentricity fails to inform us
 76 about how representations of different stimuli that are foveated, like words and faces, are
 77 organized in VTC relative to one another.

78 Despite sharing similar typical retinotopic eccentricity, word and face stimuli are highly
 79 distinct along several axes that are hypothesized to influence where they are processed in VTC
 80 (Op de Beeck et al., 2019). Word- and face-processing operate on very different low-level visual
 81 properties (Kay and Yeatman, 2017), follow different developmental trajectories (Saygin et al.,
 82 2016), and feed into distinct networks that support either language or social interactions (Stevens
 83 et al., 2015, 2017), respectively. Despite this, the cortical localizations for word- and face-
 84 processing in VTC are remarkably close together, and it remains debated whether or not there are
 85 regions in VTC that independently encode word or face information at all (Behrmann and Plaut,
 86 2013). However, electrical stimulation and lesion studies suggest that they are independent in
 87 VTC (Hirshorn et al., 2016; Sabsevitz et al., 2020).

88 Neuroimaging studies have separately mapped word- and face-processing networks in
 89 VTC. Printed word recognition is thought to be carried out in part by a network of regions along
 90 the left occipitotemporal sulcus, that differ in the complexity of their responses and are
 91 hierarchically organized (Halgren et al., 1994; Cohen et al., 2000; Vinckier et al., 2007; Dehaene
 92 and Cohen, 2011; Lerma-Usabiaga et al., 2018). Face-processing is thought to be carried out in
 93 part by a network of regions distributed bilaterally along the midfusiform sulcus (Tsao et al.,
 94 2008; Weiner and Grill-Spector, 2010). However, few studies have investigated VTC's
 95 responses to word and face stimuli within the same participants (Allison et al., 1994; Haxby et
 96 al., 1994; Puce et al., 1996; Matsuo et al., 2015; Harris et al., 2016). Those that have, have relied
 97 on low sample sizes or imaging modalities with differential sensitivity to different aspects of
 98 neural activity (like high and low-frequency neural activity (Engell et al., 2012; Jonas et al.,
 99 2016)). Therefore, much remains unknown about how visual word- and face-processing
 100 networks organize relative to one another, and to what degree they overlap (Haxby et al., 1994;
 101 Puce et al., 1996; Dehaene et al., 2010; Matsuo et al., 2015; Harris et al., 2016).

102 Further, it is unclear if the nodes within these processing networks differ in the temporal
 103 dynamics of their responses, although previous studies have suggested that different regions may
 104 contribute to distinct stages of word- and face-processing (Federmeier and Kutas, 1999; Vinckier
 105 et al., 2007; Li et al., 2018). Additionally, category-selective maps derived from BOLD
 106 responses may be incomplete due to BOLD's increased sensitivity to early stimulus evoked
 107 activity (100-300 ms after stimulus presentations) relative to later responses (Jacques et al.,
 108 2016; Ghuman and Martin, 2019) and greater correlation with high frequency broadband activity
 109 in invasive neural recordings compared to lower-frequency electrical potentials (Engell et al.,
 110 2012; Jacques et al., 2016).

111 In the present study, we characterized the spatial organization and functional dynamics of
 112 word- and face-processing networks within VTC using intracranial electroencephalography
 113 (iEEG) data collected from 36 patients with pharmacologically intractable epilepsy and 7 T
 114 fMRI data collected from eight healthy participants.

115 **Materials and Methods**

116 Intracranial EEG data collection and preprocessing

117 **Participants**

118 38 patients (14 males, ages 19-65, 32 righthanded) had intracranial surface and/or depth
 119 electrodes implanted for the treatment of pharmacologically intractable epilepsy. Depth
 120 electrodes were produced by Ad-Tech Medical and PMT Corporation and were 0.86 and 0.8 mm
 121 in diameter, respectively. Grid electrodes were produced by PMT Corporation and were 4 mm in
 122 diameter. Because depth electrode contacts are cylindrical, the surface area of the recording site
 123 was similar across grid and strip electrode contacts. To be concise, “electrode contacts” are
 124 referenced to as “electrodes” throughout the manuscript. No consistent differences in neural
 125 responses were observed between grid and depth electrodes. Only electrodes implanted in ventral
 126 temporal cortex, defined as below the inferior temporal gyrus and anterior to the posterior tip of
 127 the fusiform in the participant-centered space, were considered in this study. Two patients did
 128 not have any electrodes within this region of interest, therefore only data from 36 participants
 129 were analyzed for this study. Electrodes identified as belonging to the seizure onset zone based
 130 on the clinical report or showing epileptiform activity during the tasks were excluded from the
 131 analysis. All participants gave written informed consent. The study was approved by the

132 University of Pittsburgh Institutional Review Board. Patients were monetarily compensated for
133 their time.

134 Electrodes were localized via either post-operative magnetic resonance imaging (MRI) or
135 computed tomography scans co-registered to the pre-operative MRI using Brainstorm (Tadel et
136 al., 2011). Surface electrodes were projected to the nearest point on the pre-operative cortical
137 surface automatically parcellated via Freesurfer (Dale et al., 1999) to correct for brainshift
138 (Hermes et al., 2010). Electrode coordinates were then coregistered via surface-based
139 transformations to the fsaverage template using Freesurfer cortical reconstructions.

140 Experimental Design

141 All participants underwent a category localizer task where they viewed grayscale images
142 presented on a computer screen positioned two meters from their face. Images occupied
143 approximately 6 x 6 degrees of visual angle and were presented for 900 ms with 1500 ms inter-
144 stimulus interval with random 400 ms jitter. Participants were instructed to press a button every
145 time an image was presented twice in a row (1/6 of the trials). These repeat trials were excluded
146 from the analysis yielding 70 trials per stimulus category left for analysis. Several participants
147 underwent multiple runs of this task and therefore had 140-210 trials per stimulus category.

148 31 of the participants saw pictures of faces, words, bodies, hammers, houses, and phase-
149 scrambled faces. The remaining participants viewed a modified set of stimuli with the same
150 viewing parameters described above. One participant viewed pictures of consonant-strings and
151 pseudowords instead of hammers, two viewed shoes instead of words, one viewed consonant-
152 strings and pseudowords instead of hammers and houses, and one viewed general tools and
153 animals instead of hammers.

154 A subset of the participants that underwent the category localizer task also participated in
 155 word and/or face individuation tasks (Table 1). These tasks shared identical presentation
 156 parameters as the category-localizer task (i.e. inter-stimulus interval, stimulus-on time, and
 157 viewing angle) but contained different images. Twelve underwent a word individuation task that
 158 included pictures of real words, pseudowords, and consonant-strings or false-fonts. Participants
 159 again were instructed to respond if a given stimulus was repeated twice in a row. Every stimulus
 160 (i.e. individual word) was presented sixty times. Twenty underwent a face individuation task
 161 where they viewed individuals of varying identity and emotions. Participants were instructed to
 162 indicate if each face was male or female during this task. Each identity was repeated 60 times.

163 Local field potentials were recorded via a GrapeVine Neural Interface (Ripple, LLC)
 164 sampling at 1 kHz. Notch filters at 60/120/180 Hz were applied online. Data was subsequently
 165 filtered from 0.1-115 Hz to isolate single trial potentials (stP) or decomposed via Morlet wave
 166 convolution to determine the power from 40-100 Hz to isolate single trial high frequency broad-
 167 band activity (stHFBB). These stHFBB responses were then Z-scored based on the baseline
 168 period from 500-0 ms preceding stimulus onsets. It has been previously shown that these two
 169 aspects of the local-field potential, stP and stHFBB, contain complementary information (Miller
 170 et al., 2016), though also potentially arise from different neurophysiological generators (Engell et
 171 al., 2012; Hermes et al., 2012; Jacques et al., 2016; Leszczyński et al., 2020). Therefore, to
 172 assess the overall selectivity across VTC we use both as features in the classifiers described in
 173 *Multivariate temporal pattern analysis* (Figures 1B, 2-4, 6-8). We also investigated the
 174 independent contributions of these signal components to our category-selectivity maps (Figure
 175 6). Trials where the stHFBB or stP exceeded 5 standard deviations from the mean were thought
 176 to contain noise and therefore excluded from further analysis.

177 Determining Language Laterality

178 Records from preclinical magnetoencephalography (MEG) language mapping sessions were
 179 used to determine the laterality of language function for 30 of the 36 iEEG participants.
 180 Language mapping records for the remainder of the participants could not be located. The
 181 preclinical language mapping records contained laboratory technician notes indicating whether
 182 MEG activity during reading, listening, and word-repetition tasks was lateralized to the left or
 183 right hemisphere. The original data from these sessions was not available to conduct more
 184 precise analyses of language laterality for these participants.

185 Multivariate temporal pattern analysis

186 To determine which electrodes contained information about word and face categories, leave-one
 187 trial out cross-validated Gaussian Naïve Bayes classifiers were used to predict the category of
 188 object participants were viewing given a sliding 100 ms of neural activity from one iEEG
 189 electrode during the category-localizer task (six-way classification). Signals from stP and
 190 stHFBB were both fed in as features to a single classifier for the main selectivity maps. This
 191 procedure was repeated from 100 ms prior to 900 ms after stimulus onset with 10 ms time-step to
 192 derive a time-course of decoding at each VTC electrode. We also ran separate classifiers on only
 193 features from stP or stHFBB to investigate the independent sources of information contained
 194 within these signal components. We ensured the number of features fed into these two types of
 195 classifiers was consistent by averaging 10 ms bins of stP, since stHFBB was sampled only every
 196 10 ms, before classification.

197 Face-selective iEEG electrodes were defined as those that achieved a peak sensitivity (d')
 198 of decoding for faces greater than the chance at the $p < 0.05$ level, Bonferroni corrected for

multiple comparisons in time and across the total number of electrodes within a participant. Sensitivity (d') describes the separation between a classifier's noise and signal distributions and is defined as the inverse normal cumulative distribution function (Z') of the true positive rate (TPR) minus the inverse normal cumulative distribution function of the false positive rate (FPR),

$$Z'(TPR) - Z'(FPR).$$

The Bonferroni corrected d' sensitivity threshold was found by performing a binomial test on a null distribution of 1 million d' sensitivities that were obtained by randomly classifying permutations of the trial labels. A small number of electrodes responded to all categories *except* faces, which resulted in above-chance face classification, since the distribution of responses to faces was significantly different from the responses to other object categories. Therefore, we imposed an additional criterion to determine selectivity: face-selective channels had to demonstrate a maximum peak event-related potential or event-related broadband response to faces relative to the other object categories. An identical procedure was done to define word- and house-selective electrodes.

To determine the independence of word- and face-selectivity within electrodes, we repeated the above multivariate pattern analysis for word- and face-selective electrodes after removing trials from the category they were most selective to. Word-selective electrodes were determined to also be selective for face stimuli if, after removing trials when words were presented, we could reliably predict trials where faces were presented from the other object categories (d' sensitivity corresponding to $p < 0.05$, Bonferroni corrected for multiple temporal and electrode comparisons within participants using the same permutation test described above). Further, we stipulated that this d' for faces must be greater than the d' for all the remaining

220 object categories. An identical procedure was used to define face-selective electrodes that were
 221 also selective for words.

222 To determine if word- and face-selective electrodes contained exemplar-level information
 223 about either faces or words, we performed pairwise classification of the face and word
 224 individuation stimuli for the electrodes on which we had data (Table 1). Specifically, in the case
 225 of word individuation, we used three-fold cross-validated Gaussian Naïve Bayes classifiers to
 226 predict which of two real words a participant was viewing based on sliding 100 ms of data from
 227 the word-selective electrodes. Three-fold cross-validation was used instead of leave-one-out
 228 cross validation (which was used for assessing category-level selectivity) to save computational
 229 time as there were many more models (stimulus pairs) tested with the exemplar classifier. We
 230 repeated this procedure across all pairs of real-words of the same length and averaged the time-
 231 courses of this pairwise decoding (56 pairs of words). We determined the $p < 0.05$ chance-level
 232 of this average pairwise decoding by repeating this procedure 1,000 times on data with shuffled
 233 trial labels in a subset of the word-selective electrodes (Maris and Oostenveld, 2007). These
 234 global null distributions were similar across the randomly subsampled electrodes and therefore
 235 we chose a d' threshold corresponding to the highest $p < 0.05$ level obtained from this randomly
 236 chosen subset. We ran similar pairwise decoding and threshold definition on real-word versus
 237 pseudowords of the same length (36 pairs) and real-word versus false-font stimuli (136 pairs) to
 238 determine if electrodes that could not individuate real-words could perform these finer
 239 discriminations compared to those tested in the category localizer task.

240 Similarly, for face individuation we performed pairwise decoding of face stimuli during
 241 sliding 100 ms time-windows of face-selective electrode activity. We then averaged these time-
 242 courses across all 120 pairwise face classifications and calculated the $p < 0.05$ corrected level by

243 repeating the permutation analysis described for the word individuation task on a random subset
 244 of face-selective electrodes.

245 Spatiotemporal k-means clustering

246 We used a spatiotemporal variant of k-means clustering to determine if spatially contiguous
 247 word- or face-selective regions demonstrated distinct temporal dynamics. For word- and face-
 248 selective electrodes, we separately standardized the d' sensitivity time-courses derived from the
 249 category-level multivariate classifiers of left and right hemisphere electrodes from 100 to 600 ms
 250 post stimulus onset. We then concatenated this matrix with the electrodes' MNI-coordinate,
 251 which was multiplied by a constant (spatial weighting parameter) that modulated the weight of
 252 the spatial versus temporal components of the signal to the clustering algorithm. We then
 253 performed k-means clustering using Euclidean distances and 100 repeats with random
 254 initializations to determine clusters of nearby word- or face-selective electrodes within each
 255 hemisphere that demonstrated correlated dynamics. Because the d' time-courses were
 256 standardized, Euclidean distances were equivalent to correlation distance for the temporal data
 257 and Euclidean distance for the spatial data.

258 To determine the optimal weighting of spatial and temporal signal components and
 259 optimal number of clusters, we calculated the total spatial and temporal variance explained by
 260 the clustering solutions run with several spatial weighting parameters. This was performed for k
 261 = 1 to 10 clusters per hemisphere per faces or words. The elbow method was used to determine
 262 the optimal number of clusters per hemisphere. The optimal number of clusters was 4 for right
 263 hemisphere face-selective electrodes, 3 for right hemisphere word-selective electrodes, 3 for left
 264 hemisphere face-selective electrodes, and 4 for left hemisphere word-selective electrodes. We
 265 chose the spatial weighting parameter that explained the maximum amount of variance across k

266 = 3 to 4 clusters per hemisphere per category (spatial weight = 300). Small deviations in the
 267 spatiotemporal weighting parameter did not strongly affect the overall organization of
 268 spatiotemporal clusters. The dynamics of these electrode clusters were then determined by
 269 averaging the selectivity time courses (d' derived using *multivariate temporal pattern analysis*)
 270 across the electrodes belonging to each cluster.

271 Statistical analyses

272 Two sample t-tests were used to compare peak d' sensitivity, peak latency, and onset latency for
 273 right versus left word- and face-selective electrodes. Onset latency was defined as the first time
 274 point that the d' sensitivity reached a $p < 0.001$ threshold, which was non-parametrically defined
 275 using the d' sensitivities of all object-selective electrodes from 500-0 ms prior to stimulus onset.
 276 Spearman's rank-order correlations were used to test for relationships between peak d'
 277 sensitivities and latency. We used linear mixed effects models to compare face and real word
 278 individuation in the category-selective clusters identified by the spatiotemporal k-means
 279 algorithm. Linear mixed effects models allowed us to determine if there were differences in peak
 280 individuation d' or latency across these clusters while correcting for cross-subject differences.
 281 We only compared spatiotemporal clusters with greater than 10 electrodes with individuation
 282 data. The Satterthwaite approximation was used to estimate the degrees of freedom in these
 283 linear mixed effects models to compute the reported p-values. The time points corresponding to
 284 the leading edge of the classification window were used for all temporal statistical analyses.

285 fMRI data collection and preprocessing

286 Participants

287 Eight participants (six females, mean age 25 years) participated in the fMRI experiment. All
 288 participants were right-handed, had normal or corrected to normal vision and gave written
 289 informed consent. The National institutes of Health Institutional review Board approved the
 290 consent and protocol (protocol 93 M-0170, clinical trials #NCT00001360). Participants were
 291 monetarily compensated for their time.

292 fMRI scanning parameters

293 All fMRI scans were conducted on a 7 T Siemens Mangetom scanner at the Clinical Research
 294 Center on the National Institutes of Health campus. Partial volumes of the occipital and temporal
 295 cortices were acquired using a 32-channel head-coil (42 slices, 1.2x1.2x1.2 mm; 10% interslice
 296 gap; TR=2s, TE=27ms; matrix size=170x170).

297 Experimental Paradigm

298 Participants fixated centrally whilst images of words, faces and houses were presented in blocks
 299 (16s per block). These images were taken from the same category localizer task presented to
 300 iEEG patients. In each block 20 exemplar stimuli were presented (300ms with a 500ms ISI).
 301 Participants performed a one-back task, responding, via MRI compatible response box, whenever
 302 the same image appeared twice in a row. Participants completed 10 runs of the localizer.

303 fMRI data preprocessing

304 All data were analyzed using the Analysis of Functional NeuroImages (AFNI) software package
 305 (Cox, 1996). Prior to statistical analysis, all images were motion corrected to the first volume of
 306 the first run. Post motion-correction data were detrended.

307 Statistical analysis

308 To identify word-, face- and house-selective regions, we performed a general linear model
309 (GLM) analysis using the AFNI functions 3ddeconvolve and 3dREMLfit. The data at each time-
310 point were treated as the sum of all effects thought to be present at that time point and the time
311 series was compared against a Generalized Least Squares Regression model fit with REML
312 estimation of the temporal auto-correlation structure. Responses were modelled by convolving a
313 standard gamma function with a 16 second square wave for each condition (words, faces &
314 houses). Estimated motion parameters were included as additional regressors of no-interest and
315 fourth-order polynomials were included to account for any slow drifts in the MRI signal over
316 time. Significance was determined by comparing the beta estimates for each condition
317 (normalized by the grand mean of each voxel for each run) against baseline.

318 Split-half analysis

319 For each participant, the ten localizer runs were divided into odd and even splits. In each split,
320 we performed the same GLM analysis as described above and looked for significant voxels for
321 the contrast of words versus faces. Despite having only half of the data, we observed significant
322 word-selectivity that was medial of face-selectivity consistently across participants. In order to
323 quantify this selectivity in an independent manner, we first defined medial word-selective
324 regions within a split (e.g. odd) and then sampled the data from the other half (e.g. even). ROIs
325 were defined using data spatially smoothed with a 2 mm Gaussian kernel to generate spatially
326 contiguous clusters, whereas the test data was not spatially smoothed. To avoid any bias in node
327 selection, this process was then reversed and the average computed. Within each ROI we
328 calculated the average t-value for each condition versus baseline.

329 Results

330 From 1,396 intracranial electrode contacts implanted within or on the surface of VTC of 36
 331 patients, we isolated those implanted in regions that were highly selective for either faces, words,
 332 or houses. Highly face-selective electrodes were defined as those that had both (1) single-trial
 333 responses that could significantly discriminate face presentations from presentations of five other
 334 object categories (words, houses, bodies, hammers, and phase-scrambled objects; $p < 0.05$ level,
 335 Bonferroni corrected for multiple spatial and temporal comparisons within participant; see
 336 *Materials and Methods*) and (2) responded maximally to faces compared to all other object
 337 categories on average. This ensured that electrodes designated as highly “face-selective” were
 338 those that responded maximally and were significantly selective for faces compared to the five
 339 other object categories. An identical procedure was used to define word- and house-selective
 340 electrodes.

341 108 electrodes demonstrated primarily face-selective responses (80 in the left, 28 in the
 342 right), 87 demonstrated primarily word-selective responses (64 in the left, 23 in the right), and 85
 343 demonstrated primarily house-selective responses (44 in the left, and 41 in the right) (Figure 1).
 344 Figure 2 and Table 1 illustrate the distribution of object-selective electrodes across participants.
 345 The greater number of left versus right object-selective electrodes was comparable to the greater
 346 coverage of left VTC relative to right VTC in our patient population (883 electrodes implanted in
 347 the left, 513 in the right, Figure 1A). Although some word- and face-selective electrodes
 348 demonstrated partial selectivity for the other object category, there were several examples of
 349 electrodes that were strongly tuned to only words or faces (Figure 3). This suggests that the
 350 neural circuits responsible for processing words and faces are, at least, partially dissociable
 351 (Behrmann and Plaut, 2013; Susilo and Duchaine, 2013; Susilo et al., 2015).

352 To assess how word- and face-processing networks organize relative to one another, the
 353 spatial topography of word-, face-, and house-selective electrodes was examined. At the group
 354 level, selectivity to house stimuli was found primarily along the left and right parahippocampal
 355 gyrus, with some cases where selectivity extended into the collateral sulcus and medial fusiform
 356 gyrus. These patches were generally medial to word- and face-selective locations, consistent with
 357 previous fMRI and iEEG studies (Halgren et al., 1994; Haxby et al., 1996; Aguirre et al., 1998;
 358 Cohen et al., 2000; Kadipasaoglu et al., 2016). Face-selectivity was found primarily along the
 359 left and right fusiform gyrus with some face-selective regions within the lingual gyrus, and
 360 occipitotemporal sulcus (Figure 1B). Consistent with prior findings (Cohen et al., 2000), word-
 361 selective regions were found on the lateral bank of the fusiform and into the occipitotemporal
 362 sulcus in the left hemisphere. Word-selective regions were also found anterior to most prior
 363 reports from fMRI, in locations that generally have poor signal due to susceptibility artifacts
 364 (Devlin et al., 2000). In contrast to most maps of word- and face-selective regions obtained from
 365 fMRI (Allison et al., 1994; Haxby et al., 1994; Puce et al., 1996; Harris et al., 2016; Saygin et al.,
 366 2016; Dehaene-Lambertz et al., 2018; Gomez et al., 2018), a mosaic of word-selective regions
 367 were also found medial to face-selective regions, on the medial bank of the fusiform and into the
 368 collateral sulcus. Each of these face-, word-, and house-selective regions were found in multiple
 369 participants (Figure 2), demonstrating relatively consistent localization of these regions at a
 370 group level.

371 Interdigitation of word- and face-selective regions was seen in the left hemisphere of 5
 372 out of 9 participants with at least two word-selective electrodes and one face-selective electrode
 373 or vice-versa and in the right hemisphere of 3 out of 5 such participants (Table 1, see Figure 4 for
 374 examples). Word-selective regions were found strictly medial to face-selective regions in the left

hemisphere of 7 out of 10 participants with at least one word- and one face-selective electrode and in right hemisphere of 4 out of 5 participants (Table 1, see Figure 4 for an example). Thus, highly word-selective regions medial to face-selective regions was not simply a consequence of individual variability in a group-level map but instead was detected in the majority of participants that had coverage of both face- and word-selective VTC.

Because word-selective patches were found medial to face-selective patches in the iEEG data, which is generally not observed in 3 T fMRI studies (Haxby et al., 1994; Puce et al., 1996; Dehaene et al., 2010), we sought to determine if a similar organization existed in healthy participants using the higher resolution of 7 T fMRI. When contrasting responses to words and faces in eight participants, face-selectivity was primarily centered on the midfusiform sulcus while word-selectivity was greatest in the occipitotemporal sulcus (Figure 5). Consistent with the iEEG results, six of the eight participants demonstrated left word-selective regions medial to face-selective regions on the fusiform gyrus. In these medial word-selective patches, responses to words were significantly greater than responses to both face and house stimuli ($p < 0.001$, split-halves analysis). These medial word-selective regions were approximately $1/3^{\text{rd}}$ the size of more lateral word-selective regions (mean size of lateral word-selective regions: 398 voxels; std. error: 43 versus medial regions: 139 voxels; std. error: 29 voxels; $p < 0.01$). Also, 7 out of 8 of the healthy participants demonstrated word-selective patches near the anterior tip of the fusiform, despite susceptibility artifacts (Devlin et al., 2000), consistent with the iEEG data (Figure 1B). Altogether, the map of word- and face-selective regions of the left hemisphere derived from 7 T fMRI were consistent with those derived from iEEG, medial and anterior word-selective regions are not seen in most maps drawn from 3 T fMRI (Haxby et al., 1994; Puce et al., 1996; Dehaene et al., 2010).

398 The maps in Figures 1-3 were made by combining two key aspects of the iEEG signals,
 399 the single trial potentials (stP) and the single trial high frequency broadband activity (stHFBB),
 400 to examine the category-selectivity of the underlying VTC neural populations in aggregate across
 401 these signal components. Studies have shown that while category-selectivity demonstrated in stP
 402 and stHFBB often overlaps, they are not redundant (Engell and McCarthy, 2011; Engell et al.,
 403 2012; Miller et al., 2016), suggesting that stP and stHFBB have at least partially distinct
 404 physiological generators. To examine these signal components separately, we trained
 405 multivariate classifiers solely on stP or stHFBB and isolated electrodes that were selective in
 406 either signal component using the same criteria as before (single-trial discriminability and
 407 highest signal amplitude for words, faces, or houses). 58 electrodes showed significant
 408 selectivity in both stP and stHFBB (Figure 6A). Notably, the regions that demonstrated
 409 selectivity in both stP and stHFBB were those most often identified in canonical maps of
 410 category-selectivity based on fMRI (Cohen et al., 2000; Vinckier et al., 2007; Tsao et al., 2008;
 411 Weiner and Grill-Spector, 2010; Lerma-Usabiaga et al., 2018). Specifically, house-selectivity
 412 was restricted to the parahippocampal cortex, face-selectivity was primarily restricted to the
 413 fusiform bilaterally, and word-selectivity was restricted primarily to the left posterior-lateral
 414 fusiform and occipitotemporal sulcus. Regions that were less consistent with canonical fMRI
 415 maps tended to be those that were not significantly selective in both stP and stHFBB. For
 416 example, the medial word-selective patches were primarily seen in stP alone (Figure 6B),
 417 whereas anterior and right hemisphere word-selectivity was prevalent in either stP or stHFBB
 418 alone (Figures 6B and 6C). Broadly, more electrodes demonstrated selectivity in stP (232
 419 electrodes from 32 participants, Figure 6B) compared to stHFBB (115 electrodes from 24
 420 participants, Figure 6C). More widespread stP selectivity is consistent with a previous study

421 comparing stP and stHFBB responses for faces in VTC, though that study did not observe any
 422 cases where selectivity for faces was demonstrated in stHFBB but not stP (Engell and McCarthy,
 423 2011). The similarities and differences in selectivity demonstrated in stHFBB and stP are
 424 consistent with the hypothesis that these signals have different physiological generators
 425 (Lachaux et al., 2005), which may differ in their laminar distribution (Leszczyński et al., 2020)
 426 and spatial signal-to-noise falloff (Engell and McCarthy, 2011). Additionally, different category-
 427 selectivity across these iEEG signal components may also help explain differences between
 428 category-selectivity maps drawn from iEEG and fMRI, as some studies suggest fMRI has
 429 differential sensitivity to these aspects of the iEEG signal (Conner et al., 2011; Engell et al.,
 430 2012; Jacques et al., 2016).

431 One question is whether word- and face-selective regions identified using iEEG
 432 discriminate between individual face and word exemplars, respectively. Classifying at the
 433 exemplar level also can address the potential concern that the word- and face-selective regions
 434 identified using iEEG may be responding to low-level features that drastically differ between the
 435 sampled image categories. A subset of the iEEG participants underwent independent word and
 436 face individuation tasks (see *Materials and Methods*, Table 1). Activity from 85 out of 97
 437 sampled face-selective electrodes in 13 participants could be used to reliably predict the identity
 438 of a presented face. Similarly, activity from 40 out of 53 sampled word-selective electrodes from
 439 10 participants could be used to discriminate single words of the same length from one another.
 440 Of those 13 word-selective electrodes that could not reliably achieve word individuation, six
 441 could reliably discriminate pseudowords from real words of the same length, seven could
 442 reliably discriminate false-fonts from real words. Therefore, most of the word- and face-selective

443 regions mapped with iEEG contained reliable exemplar-level information specific to the
 444 categories they were selective to.

445 Peak word and face individuation was significantly correlated with peak category-
 446 selectivity in word and face-selective regions for which we had individuation data (word-
 447 selective: Spearman's $\rho(53) = 0.50$, $p < 0.0001$, face-selective: $\rho(97) = 0.48$, $p < 0.0001$). Note
 448 that correlations in peak category-selectivity and within-category individuation may arise due to
 449 similar differences in measurement noise across recording contacts (for example, due to the
 450 distance the electrode was placed from the underlying face or word selective neural populations),
 451 underlying neural/physiological factors, or some mix of both.

452 In addition to the medial band of word-selective regions, there were a high proportion of
 453 right word-selective electrodes in our iEEG population (Figure 1B, Table 1). Although this
 454 finding is consistent with some other fMRI (Ben-Shachar et al., 2007; White et al., 2019) and
 455 iEEG studies (Halgren et al., 1994; Lochy et al., 2018), right hemisphere word-selectivity is
 456 often not seen in neuroimaging (Cohen et al., 2000, 2002) and was not very strong in our 7 T
 457 fMRI data either (Figure 5). 23 word-selective electrodes were found across nine participants in
 458 right VTC, out of 21 participants with right VTC object-selectivity. This discrepancy between
 459 right word-selectivity observed in fMRI and iEEG was also not attributable to participant
 460 handedness, since no participant with right word-selective regions was lefthanded. Three out of
 461 nine of these participants demonstrated evidence for bilateral language function while the other
 462 six demonstrated left dominant language function determined by preclinical
 463 magnetoencephalography (MEG, see *Materials and Methods*). Across the entire participant
 464 population, seven out of 30 iEEG participants with preclinical MEG demonstrated bilateral
 465 language function, the others were considered left dominant. One participant with bilateral

466 language function and right hemisphere object-selectivity did not demonstrate right word-
 467 selectivity. Overall, neither participant handedness nor language dominance sufficiently explains
 468 the high proportion of word-selective regions found in right VTC.

469 While neither language laterality nor handedness explained right word-selectivity,
 470 substantial differences were seen in the dynamics of neural activity recorded from left versus
 471 right word-selective regions (Figure 7). Latency to word-selectivity onset and peak was shorter
 472 in left compared to right hemisphere word-selective regions (mean onset latency difference +/-
 473 95 % CI: -133 +/- 61 ms, $T(85) = -4.4$, $p < 0.0001$, mean peak latency difference: -138 +/- 63 ms,
 474 $T(85) = -4.3$, $p < 0.0001$, Figure 7). These relationships held when taking into account potential
 475 differences in posterior to anterior coordinate of word-selective regions across hemispheres
 476 (onset: $T(85) = -4.01$, $p = 0.0001$, peak: $T(85) = -3.97$, $p = 0.0002$). There was no significant
 477 difference between the latency to peak d' sensitivity or sensitivity onset for right and left face-
 478 selective regions (mean onset latency difference: -29 +/- 53 ms, $T(106) = -1.1$, $p = 0.28$, mean
 479 peak latency difference: 18 +/- 57 ms, $T(106) = 0.63$, $p = 0.53$, Figure 7). Additionally, the
 480 amplitude of peak d' sensitivity for words was significantly greater in the left compared to right
 481 hemisphere word-selective regions (mean peak d' sensitivity difference: 0.66 +/- 0.37, $T(85) =$
 482 3.5, $p = 0.0006$). The amplitude of peak d' sensitivity to faces was also significantly greater in
 483 the left compared to right hemisphere face-selective regions (mean peak d' sensitivity difference:
 484 0.58 +/- 0.39, $T(85) = 3.0$, $p = 0.0037$). There was a significant correlation between peak latency
 485 and peak magnitude within face-selective regions in the left ($\rho(80) = -0.61$, $p < 0.0001$) and right
 486 ($\rho(28) = -0.79$, $p < 0.0001$) hemisphere and word-selective regions in the left ($\rho(64) = -0.68$, $p <$
 487 0.0001), but not right ($\rho(23) = -0.15$, $p = 0.48$) hemisphere, suggesting that longer peak latencies
 488 were associated with smaller peak selectivity. These correlations were not significantly different

489 between face-selective regions in the left and right hemisphere ($T(85) = -1.56$, $p = 0.058$), but
 490 there was a greater correlation between peak latency and magnitude in left compared to right
 491 hemisphere word-selective regions ($T(85) = 2.63$, $p = 0.004$). Given that it was only true for
 492 word-selective electrodes, the relatively slower response of right versus left word-selective
 493 regions may potentially explain differences in word-selectivity maps derived from iEEG and
 494 fMRI and may reflect the left hemisphere bias for language.

495 Finally, using the iEEG data, we sought to determine if there were any differences in the
 496 temporal dynamics of neural responses across word or face-selective regions within the same
 497 hemisphere. We used a spatiotemporal k-means clustering algorithm to find spatially contiguous
 498 regions of left and right VTC which demonstrated correlated category-selective dynamics. After
 499 optimizing the algorithm to capture the most spatiotemporal variance with the optimal number of
 500 clusters (see *Materials and Methods*), we could compare the dynamics of distinct word- and
 501 face-selective clusters within VTC.

502 Word-selective regions were clustered into 4 distinct left hemisphere clusters and 3 right
 503 hemisphere clusters (Figure 8A). Word-selective regions on the left fusiform gyrus demonstrated
 504 the earliest and strongest selectivity, peaking around 200 ms (Figure 8B, gray). Left hemisphere
 505 medial word-selective regions and right hemisphere word-selective regions came next, peaking
 506 around 300 ms (Figure 8B, green and cyan) followed by lateral regions around 350 ms (Figure
 507 8B, red). Word-selective regions in left anterior VTC peaked around 400-450 ms (Figure 8B,
 508 blue); right more anterior regions peaked around 600 ms (Figure 8B, magenta). When
 509 considering word-selectivity dynamics exhibited independently in stP and stHFBB signal
 510 components, word-selective electrodes on the fusiform demonstrated strong selectivity in both

511 signal components, whereas other regions displayed distinct dynamics across these signal
512 components (Figures 7C-D).

513 Face-selective regions were organized into 3 distinct clusters in the left hemisphere and 4
514 distinct clusters in the right hemisphere (Figure 8E). Face-selective regions of the left and right
515 fusiform gyrus demonstrated the earliest and largest peak selectivity around 200-250 ms (Figure
516 8F, gray and cyan). More anterior right hemisphere regions and a cluster of electrodes in left
517 posteromedial VTC (Figure 8F, yellow and green) peaked around 300 ms. Finally, more anterior
518 face-selective electrodes in left and right VTC peaked around 400 ms (Figure 8F, blue, black,
519 and magenta). When considering face-selectivity dynamics exhibited independently in stP and
520 stHFBB signal components, electrodes on the fusiform demonstrated strong selectivity in both
521 components, whereas other regions displayed distinct dynamics across these signal components
522 (Figures 7G-H).

523 From electrodes sampled in the word individuation task, we observed stronger word
524 individuation in left word-selective regions on the fusiform compared to the more medial word-
525 selective cluster illustrated in Figure 8A (peak d' of fusiform minus medial regions: $T(30) =$
526 3.62 , $p = 0.001$, linear mixed-effects model). There was no significant difference between the
527 latency to peak word individuation across these clusters ($T(30) = 2.91$, $p = 0.68$). There were not
528 sufficient subjects with electrodes in the other word-selective clusters with word individuation
529 data to make comparisons between all clusters. Neither peak face individuation ($T(50) = 1.03$, p
530 $= 0.31$) nor latency to peak face individuation ($T(50) = -0.21$, $p = 0.84$) was significantly
531 different between face-selective regions along the left fusiform gyrus and the posteromedial face-
532 selective cluster observed in Figure 8E. There were not sufficient subjects with electrodes in the

533 other face-selective clusters with face individuation data to make comparisons between all
534 clusters.

535 Overall, for both faces and words, these results suggest a cascade of processing that
536 begins in the fusiform. Notably, the dynamics of these clusters suggest that they contribute to
537 distinct stages of face- and word-processing, since the latencies of their responses are far longer
538 than would be expected from feedforward visual transmission delays alone (Thorpe et al., 1996;
539 Kravitz et al., 2013), but not long enough to exclude them from being relevant to perceptual
540 behavior (Quiñero Quiroga et al., 2008; Tang et al., 2014) .

541 Discussion

542 In the current study, we found several VTC regions that demonstrated strong word-, face- and
543 house-selective responses. Although activity recorded from VTC electrodes often contained
544 information about multiple object categories, several selectively responded only to faces or
545 words (Figure 3). Electrodes which demonstrated preference to only words or faces suggests that
546 VTC word- and face-processing networks are not entirely overlapping (Behrmann and Plaut,
547 2013), but instead involve at least some independent nodes (Susilo and Duchaine, 2013; Susilo et
548 al., 2015), which is also supported by stimulation and lesion evidence (Hirshorn et al., 2016;
549 Sabsevitz et al., 2020).

550 In both the iEEG and fMRI data, strong face-selectivity along the fusiform gyrus was
551 adjoining with highly word-selective regions in and around the occipitotemporal and collateral
552 sulci. House-selective regions were found primarily along the parahippocampal gyrus. This
553 organization of house- versus word- and face-selective regions supports that typical retinotopic
554 eccentricity is an important organizing principle of VTC (Grill-Spector and Weiner, 2014). The

word-selective regions around the occipitotemporal sulcus are consistent with prior studies showing word-selectivity within lateral aspects of VTC (Dehaene et al., 2002; Price and Devlin, 2003). Due to sparse and variable sampling across participants, the data cannot address the question of whether there is a gradient of word-selectivity along the occipitotemporal sulcus (Vinckier et al., 2007) or distinct patches (Lerma-Usabiaga et al., 2018; White et al., 2019).

Despite some similarities with previous neuroimaging work, the iEEG and 7 T fMRI data here are inconsistent with a map of VTC wherein word-selective regions are strictly lateral to face-selective regions (Haxby et al., 1994; Puce et al., 1996; Dehaene et al., 2010). While there has been some mixed reporting of word-selectivity in anterior and medial VTC regions (Allison et al., 1994; Haxby et al., 1994; Puce et al., 1996; Harris et al., 2016; Saygin et al., 2016; Dehaene-Lambertz et al., 2018; Gomez et al., 2018), most models of orthographic-processing within VTC consider only the more lateral, traditional “visual word form area” (Dehaene et al., 2002; Price and Devlin, 2003). The disagreement between the observed organization of face- and word-processing networks in VTC and most previous maps drawn from fMRI may be the product of spatial smoothing commonly applied during fMRI data analysis (Geissler et al., 2005), signal dropout induced by susceptibility artifacts (Devlin et al., 2000), or the inferior sensitivity of 3 T fMRI relative to 7 T fMRI. Here, a mosaic of word-selective regions was found medial and anterior to face-selective regions within multiple iEEG patients and in 7 T fMRI in healthy individuals. This evidence makes it unlikely that our observations are the product of inter-participant variability or differences between healthy controls and patients with intractable epilepsy (see also (Matsuo et al., 2015; Jonas et al., 2016; Kadipasaoglu et al., 2016; Lochy et al., 2018)). This mosaic organization of visual word-selective regions is similar to the mosaic

organization of auditory language processing networks (Flinker et al., 2011), suggesting this pattern of organization may not be specific to the visual system.

The interdigitation of word- and face-selective regions along the mediolateral axis is not well captured solely by a rectilinear model of VTC, wherein more medial regions are more responsive to straight over curvy objects (Srihasam et al., 2014; Bao et al., 2020), or a retinotopic model. Instead, medial and lateral word-selective regions with distinct dynamics may indicate an interaction between multiple representational axes in VTC (Konkle and Caramazza, 2013; Grill-Spector and Weiner, 2014) and possibly competition between faces and words for cortical space (Behrmann and Plaut, 2020). Others have suggested that lateral word-selective regions are responsible for recognizing word forms while medial, perirhinal word-selective regions associate concrete words with the objects they refer to (Liuzzi et al., 2019).

Previous studies have used electrical stimulation to demonstrate that a large portion of VTC, sometimes termed the “basal temporal language area,” plays a role in language processing (Krauss et al., 1996; Mani et al., 2008; Fonseca et al., 2009; Enatsu et al., 2017). However, the relationship between reading deficits and VTC lesions outside of the visual word form area (Gaillard et al., 2006; Hirshorn et al., 2016) is unclear. A recent study reported differential language-related deficits during reading, repetition, and picture naming depending on the area of VTC stimulated (Forseth et al., 2018). Future studies are necessary to understand the precise relationship between medial, lateral, and anterior word-selective VTC dynamics and these regions’ functional contribution to reading and/or language processing.

Category-selective regions most consistent with prior fMRI studies were those that demonstrated selectivity in both stHFBB and stP iEEG signal components. In contrast, we found that medial word-selectivity was primarily demonstrated in stP rather than stHFBB. Previous

600 studies have suggested that fMRI BOLD have differential sensitivity to stHFBB versus stP
 601 (Hermes et al., 2012), with some suggesting greater sensitivity to stHFBB (Engell et al., 2012;
 602 Jacques et al., 2016). Differential sensitivity to stP and stHFBB may explain why previous fMRI
 603 studies have only inconsistently observed medial word-selective regions. Our 7 T fMRI data
 604 shows that, with adequate power, both lateral and medial word-selective regions are seen in the
 605 left hemisphere using BOLD within individual participants. Future studies are necessary to fully
 606 understand the functional characteristics and neurophysiological generators of stP and stHFBB
 607 iEEG components (Miller, 2010; Ray and Maunsell, 2011; Leszczyński et al., 2020) and how
 608 they relate to any differential roles that medial and lateral word-selective regions play in reading.

609 In addition to this complex organization of word- and face-selectivity within
 610 hemispheres, our iEEG analyses suggest that right word-selective regions demonstrate longer
 611 latencies and lower amplitudes of peak selectivity compared to left word-selective regions,
 612 which may reflect the primary role the left, language dominant, hemisphere plays in word-
 613 processing (Fiez and Petersen, 1998). Previous studies have demonstrated weaker correlations
 614 between object-selectivity measured with iEEG and fMRI correlations at later time windows
 615 (Jacques et al., 2016). This may explain why bilateral selectivity to words is inconsistent across
 616 neuroimaging studies.

617 It has previously been suggested that right word-selective regions (along with left
 618 posterior word-selective regions) are involved in relatively early visual processing of words and
 619 then this information flows to left anterior word-selective regions (White et al., 2019). However,
 620 the dynamics observed here do not support this hypothesis, because left word-selectivity
 621 substantially preceded right word-selectivity. Instead, the time-course of right hemisphere
 622 activation is coincident with P300 and N400 potentials observed during reading, suggesting that

623 right hemisphere word-selective regions may support the left hemisphere in later computations,
 624 such as those involving word syntax, memory encoding, and/or semantic processing (Friedman
 625 et al., 1975; Kutas and Hillyard, 1980; Federmeier and Kutas, 1999; Otten and Donchin, 2000;
 626 Arbel et al., 2011).

627 Word- and face-selective regions within hemispheres also demonstrated distinct
 628 dynamics. Word-selective regions on the left fusiform gyrus demonstrated the earliest and
 629 strongest word-selective responses. This was followed by word-selective activity in left
 630 occipitotemporal and collateral sulcus as well as right posterior word-selective regions. Finally,
 631 word-selective activity spread to anterior VTC between 400-600 ms. The relatively later
 632 responses of word-selective regions outside of the fusiform may contribute to differences in
 633 category-selective maps drawn from iEEG and fMRI (Jacques et al., 2016).

634 Face-selective responses were strongest and earliest on the fusiform gyrus bilaterally. A
 635 cluster of posteromedial face-selective electrodes was found in early visual cortex. The slower
 636 time-course of these regions compared to face-selective regions on the fusiform suggests this
 637 posterior face-selectivity is a result of top-down attentional effects previously reported during
 638 face-viewing (Mo et al., 2018). Following fusiform responses, face-selectivity was then seen in
 639 more anterior VTC.

640 While delays in processing along the posterior-to-anterior VTC axis for both faces and
 641 words is somewhat consistent with feedforward models of visual processing, the relative
 642 latencies are far longer than would be expected in these models (Thorpe et al., 1996; Kravitz et
 643 al., 2013). These results instead suggest more extended dynamics, perhaps governed by recurrent
 644 processes (Kravitz et al., 2013), with different category-selective regions contributing
 645 differentially to multiple, temporally extended stages of face- and word-processing (Ghuman et

al., 2014; Hirshorn et al., 2016; Li et al., 2018). Further studies are required to identify these stages and link them to different spatiotemporal patterns of VTC activity. It is important to acknowledge that when analyzing the data at this fine granularity, between-participant variability in neural organization may influence the differences observed in dynamics across regions (Zhen et al., 2015; Gao et al., 2018).

The high-resolution maps of category-selectivity within VTC provided here suggest that in addition to more extensively studied word-selective patches within the occipitotemporal sulcus, additional patches of word-selectivity exist along the mid and anterior fusiform gyrus. These patches of word-selectivity differ in their temporal dynamics from word-selective patches along the occipitotemporal sulcus, but still contain information about word identity. How these word-selective regions differentially contribute to reading and the factors that lead to the development of adjoining patches of word- and face-selective regions remain as important outstanding questions. Understanding this complex and dynamic map of selectivity in VTC is necessary to fully understand the organizational and computational principles governing object recognition.

References

- Aguirre GK, Zarahn E, D'Esposito M (1998) An area within human ventral cortex sensitive to “building” stimuli: Evidence and implications. *Neuron* 21:373–383.
- Allison T, McCarthy G, Nobre A, Puce A, Belger A (1994) Human extrastriate visual cortex and the perception of faces, words, numbers, and colors. *Cereb Cortex* 4:544–554.
- Arbel Y, Spencer KM, Donchin E (2011) The N400 and the P300 are not all that independent. *Psychophysiology* 48:861–875.

- 668 Bao P, She L, McGill M, Tsao DY (2020) A map of object space in primate inferotemporal
669 cortex. *Nature* 583:103–108.
- 670 Behrmann M, Plaut DC (2013) Distributed circuits, not circumscribed centers, mediate visual
671 recognition. *Trends Cogn Sci* 17:210–219.
- 672 Behrmann M, Plaut DC (2020) Hemispheric Organization for Visual Object Recognition: A
673 Theoretical Account and Empirical Evidence. *Perception* 49:373–404.
- 674 Ben-Shachar M, Dougherty RF, Deutsch GK, Wandell BA (2007) Differential Sensitivity to
675 Words and Shapes in Ventral Occipito-Temporal Cortex. *Cereb Cortex* 17:1604–1611.
- 676 Chen R, Hillis AE, Pawlak M, Herskovits EH (2008) Voxelwise Bayesian lesion-deficit analysis.
677 *Neuroimage* 40:1633–1642.
- 678 Cohen L, Dehaene S, Naccache L, Lehéricy S, Dehaene-Lambertz G, Hénaff MA, Michel F
679 (2000) The visual word form area. Spatial and temporal characterization of an initial stage
680 of reading in normal subjects and posterior split-brain patients. *Brain* 123:291–307.
- 681 Cohen L, Lehéricy S, Chochon F, Lemer C, Rivaud S, Dehaene S (2002) Language-specific
682 tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain*
683 125:1054–1069.
- 684 Conner CR, Ellmore TM, Pieters TA, di Sano MA, Tandon N (2011) Variability of the
685 relationship between electrophysiology and BOLD-fMRI across cortical regions in humans.
686 *J Neurosci* 31:12855–12865.
- 687 Cox RW (1996) AFNI: Software for Analysis and Visualization of Functional Magnetic
688 Resonance Neuroimages. *Comput Biomed Res* 29:162–173.

- 689 Dale AM, Fischl B, Sereno MI (1999) Cortical Surface-Based Analysis. *Neuroimage* 9:179–194.
- 690 Dehaene-Lambertz G, Monzalvo K, Dehaene S (2018) The emergence of the visual word form:
 691 Longitudinal evolution of category-specific ventral visual areas during reading acquisition.
 692 *PLoS Biol* 16:e2004103.
- 693 Dehaene S, Cohen L (2011) The unique role of the visual word form area in reading. *Trends*
 694 *Cogn Sci* 15:254–262.
- 695 Dehaene S, Pegado F, Braga LW, Ventura P, Nunes Filho G, Jobert A, Dehaene-Lambertz G,
 696 Kolinsky R, Morais J, Cohen L (2010) How learning to read changes the cortical networks
 697 for vision and language. *Science* (80-) 330:1359–1364.
- 698 Dehaene SC, Le Clec'H G, Poline J-B, Le Bihan D, Cohen L (2002) The visual word form area:
 699 a prelexical representation of visual words in the fusiform gyrus. *Neuroreport* 13:321–325.
- 700 Devlin JT, Russell RP, Davis MH, Price CJ, Wilson J, Moss HE, Matthews PM, Tyler LK
 701 (2000) Susceptibility-Induced Loss of Signal: Comparing PET and fMRI on a Semantic
 702 Task. *Neuroimage* 11:589–600.
- 703 Enatsu R, Kanno A, Ookawa S, Ochi S, Ishiai S, Nagamine T, Mikuni N (2017) Distribution and
 704 Network of Basal Temporal Language Areas: A Study of the Combination of Electric
 705 Cortical Stimulation and Diffusion Tensor Imaging. *World Neurosurg* 106:1–8.
- 706 Engell AD, Huetzel S, McCarthy G (2012) The fMRI BOLD signal tracks electrophysiological
 707 spectral perturbations, not event-related potentials. *Neuroimage* 59:2600–2606.
- 708 Engell AD, McCarthy G (2011) The relationship of gamma oscillations and face-specific ERPs
 709 recorded subdurally from occipitotemporal cortex. *Cereb Cortex* 21:1213–1221.

- 710 Federmeier KD, Kutas M (1999) Right words and left words: Electrophysiological evidence for
711 hemispheric differences in meaning processing. *Cogn Brain Res* 8:373–392.
- 712 Fiez JA, Petersen SE (1998) Neuroimaging studies of word reading. *Proc Natl Acad Sci U S A*
713 95:914–921.
- 714 Flinker A, Chang EF, Barbaro NM, Berger MS, Knight RT (2011) Sub-centimeter language
715 organization in the human temporal lobe. *Brain Lang* 117:103–109.
- 716 Fonseca AT Da, Bénar CG, Bartoloméi F, Régis J, Démonet JF, Chauvel P, Liégeois-Chauvel C
717 (2009) Electrophysiological study of the basal temporal language area: A convergence zone
718 between language perception and production networks. *Clin Neurophysiol* 120:539–550.
- 719 Forseth KJ, Kadipasaoglu CM, Conner CR, Hickok G, Knight RT, Tandon N (2018) A lexical
720 semantic hub for heteromodal naming in middle fusiform gyrus. *Brain* 141:2112–2126.
- 721 Friedman D, Simson R, Ritter W, Rapin I (1975) The late positive component (P300) and
722 information processing in sentences. *Electroencephalogr Clin Neurophysiol* 38:255–262.
- 723 Gaillard R, Naccache L, Pinel P, Clémenceau S, Volle E, Hasboun D, Dupont S, Baulac M,
724 Dehaene S, Adam C, Cohen L (2006) Direct Intracranial, fMRI, and Lesion Evidence for
725 the Causal Role of Left Inferotemporal Cortex in Reading. *Neuron* 50:191–204.
- 726 Gao X, Gentile F, Rossion B (2018) Fast periodic stimulation (FPS): a highly effective approach
727 in fMRI brain mapping. *Brain Struct Funct* 223:2433–2454.
- 728 Geissler A, Lanzenberger R, Barth M, Tahamtan AR, Milakara D, Gartus A, Beisteiner R (2005)
729 Influence of fMRI smoothing procedures on replicability of fine scale motor localization.
730 *Neuroimage* 24:323–331.

- 731 Ghuman AS, Brunet NM, Li Y, Konecky RO, Pyles JA, Walls SA, Destefino V, Wang W,
 732 Richardson RM (2014) Dynamic encoding of face information in the human fusiform gyrus.
 733 Nat Commun 5:5672.
- 734 Ghuman AS, Martin A (2019) Dynamic Neural Representations: An Inferential Challenge for
 735 fMRI. Trends Cogn Sci 23:534–536.
- 736 Gomez J, Natu V, Jeska B, Barnett M, Grill-Spector K (2018) Development differentially sculpts
 737 receptive fields across early and high-level human visual cortex. Nat Commun 9:788.
- 738 Grill-Spector K, Weiner KS (2014) The functional architecture of the ventral temporal cortex
 739 and its role in categorization. Nat Rev Neurosci 15:536–548.
- 740 Halgren E, Baudena P, Heit G, Clarke M, Marinkovic K (1994) Spatio-temporal stages in face
 741 and word processing. 1. Depth recorded potentials in the human occipital and parietal lobes.
 742 J Physiol - Paris 88:1–50.
- 743 Harris RJ, Rice GE, Young AW, Andrews TJ (2016) Distinct but overlapping patterns of
 744 response to words and faces in the fusiform gyrus. Cereb Cortex 26:3161–3168.
- 745 Hasson U, Levy I, Behrmann M, Hendler T, Malach R (2002) Eccentricity Bias as an Organizing
 746 Principle for Human High-Order Object Areas. Neuron 34:479–490.
- 747 Haxby J V, Horwitz B, Ungerleider LG, Maisog JM, Pietrini P, Grady CL (1994) The functional
 748 organization of human extrastriate cortex: a PET-rCBF study of selective attention to faces
 749 and locations. J Neurosci 14:6336–6353.
- 750 Haxby J V, Ungerleider LG, Horwitz B, Maisog JM, Rapoport SI, Grady CL (1996) Face
 751 encoding and recognition in the human brain. Proc Natl Acad Sci 93:922–927.

- 752 Hermes D, Miller KJ, Noordmans HJ, Vansteensel MJ, Ramsey NF (2010) Automated
 753 electrocorticographic electrode localization on individually rendered brain surfaces. *J*
 754 *Neurosci Methods* 185:293–298.
- 755 Hermes D, Miller KJ, Vansteensel MJ, Aarnoutse EJ, Leijten FSS, Ramsey NF (2012)
 756 Neurophysiologic correlates of fMRI in human motor cortex. *Hum Brain Mapp* 33:1689–
 757 1699.
- 758 Hirshorn EA, Li Y, Ward MJ, Richardson RM, Fiez JA, Ghuman AS (2016) Decoding and
 759 disrupting left midfusiform gyrus activity during word reading. *Proc Natl Acad Sci*
 760 113:8162–8167.
- 761 Jacques C, Witthoft N, Weiner KS, Foster BL, Rangarajan V, Hermes D, Miller KJ, Parvizi J,
 762 Grill-Spector K (2016) Corresponding ECoG and fMRI category-selective signals in human
 763 ventral temporal cortex. *Neuropsychologia* 83:14–28.
- 764 Jonas J, Jacques C, Liu-Shuang J, Brissart H, Colnat-Coulbois S, Maillard L, Rossion B (2016)
 765 A face-selective ventral occipito-temporal map of the human brain with intracerebral
 766 potentials. *Proc Natl Acad Sci* 113:E4088–E4097.
- 767 Kadipasaoglu CM, Conner CR, Whaley ML, Baboyan VG, Tandon N (2016) Category-
 768 Selectivity in Human Visual Cortex Follows Cortical Topology: A Grouped icEEG Study
 769 Ben Hamed S, ed. *PLoS One* 11:e0157109.
- 770 Kay KN, Yeatman JD (2017) Bottom-up and top-down computations in word- and face-selective
 771 cortex. *Elife* 6:e22341.
- 772 Konkle T, Caramazza A (2013) Tripartite organization of the ventral stream by animacy and

- 773 object size. *J Neurosci* 33:10235–10242.
- 774 Krauss GL, Fisher R, Plate C, Hart J, Uematsu S, Gordon B, Lesser RP (1996) Cognitive effects
775 of resecting basal temporal language areas. *Epilepsia* 37:476–483.
- 776 Kravitz DJ, Saleem KS, Baker CI, Ungerleider LG, Mishkin M (2013) The ventral visual
777 pathway: an expanded neural framework for the processing of object quality. *Trends Cogn*
778 *Sci* 17:26–49.
- 779 Kutas M, Hillyard SA (1980) Reading senseless sentences: Brain potentials reflect semantic
780 incongruity. *Science* (80-) 207:203–205.
- 781 Lachaux JP, George N, Tallon-Baudry C, Martinerie J, Hugueville L, Minotti L, Kahane P,
782 Renault B (2005) The many faces of the gamma band response to complex visual stimuli.
783 *Neuroimage* 25:491–501.
- 784 Lerma-Usabiaga G, Carreiras M, Paz-Alonso PM (2018) Converging evidence for functional and
785 structural segregation within the left ventral occipitotemporal cortex in reading. *Proc Natl*
786 *Acad Sci U S A* 115:E9981–E9990.
- 787 Leszczyński M, Barczak A, Kajikawa Y, Ulbert I, Falchier AY, Tal I, Haegens S, Melloni L,
788 Knight RT, Schroeder CE (2020) Dissociation of broadband high-frequency activity and
789 neuronal firing in the neocortex. *Sci Adv* 6:977–989.
- 790 Li Y, Richardson RM, Ghuman AS (2018) Posterior Fusiform and Midfusiform Contribute to
791 Distinct Stages of Facial Expression Processing. *Cereb Cortex* 29:3209–3219.
- 792 Liuzzi AG, Bruffaerts R, Vandenberghe R (2019) The medial temporal written word processing
793 system. *Cortex* 119:287–300.

- 794 Lochy A, Jacques C, Maillard L, Colnat-Coulbois S, Rossion B, Jonas J (2018) Selective visual
 795 representation of letters and words in the left ventral occipito-temporal cortex with
 796 intracerebral recordings. *Proc Natl Acad Sci* 115:E7595–E7604.
- 797 Mani J, Diehl B, Piao Z, Schuele SS, Lapresto E, Liu P, Nair DR, Dinner DS, Lüders HO (2008)
 798 Evidence for a basal temporal visual language center: cortical stimulation producing pure
 799 alexia. *Neurology* 71:1621–1627.
- 800 Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. *J*
 801 *Neurosci Methods* 164:177–190.
- 802 Matsuo T, Kawasaki K, Kawai K, Majima K, Masuda H, Murakami H, Kunii N, Kamitani Y,
 803 Kameyama S, Saito N, Hasegawa I (2015) Alternating Zones Selective to Faces and Written
 804 Words in the Human Ventral Occipitotemporal Cortex. *Cereb Cortex* 25:1265–1277.
- 805 Miller KJ (2010) Broadband spectral change: Evidence for a macroscale correlate of population
 806 firing rate? *J Neurosci* 30:6477–6479.
- 807 Miller KJ, Schalk G, Hermes D, Ojemann JG, Rao RPN (2016) Spontaneous Decoding of the
 808 Timing and Content of Human Object Perception from Cortical Surface Recordings Reveals
 809 Complementary Information in the Event-Related Potential and Broadband Spectral Change
 810 Sporns O, ed. *PLOS Comput Biol* 12:e1004660.
- 811 Mo C, He D, Fang F (2018) Attention priority map of face images in human early visual cortex.
 812 *J Neurosci* 38:149–157.
- 813 Op de Beeck HP, Pillet I, Ritchie JB (2019) Factors Determining Where Category-Selective
 814 Areas Emerge in Visual Cortex. *Trends Cogn Sci* 23:784–797.

- 815 Otten LJ, Donchin E (2000) Relationship between P300 amplitude and subsequent recall for
 816 distinctive events: Dependence on type of distinctiveness attribute. *Psychophysiology*
 817 37:644–661.
- 818 Price CJ, Devlin JT (2003) The myth of the visual word form area. *Neuroimage* 19:473–481.
- 819 Puce A, Allison T, Asgari M, Gore JC, McCarthy G (1996) Differential Sensitivity of Human
 820 Visual Cortex to Faces, Letterstrings, and Textures: A Functional Magnetic Resonance
 821 Imaging Study. *J Neurosci* 16:5205–5215.
- 822 Quiñ Quiroga R, Mukamel R, Isham EA, Malach R, Fried I (2008) Human single-neuron
 823 responses at the threshold of conscious recognition. *Proc Natl Acad Sci U S A* 105:3599–
 824 3604.
- 825 Ray S, Maunsell JHR (2011) Different Origins of Gamma Rhythm and High-Gamma Activity in
 826 Macaque Visual Cortex Ungerleider L, ed. *PLoS Biol* 9:e1000610.
- 827 Sabsevitz DS, Middlebrooks EH, Tatum W, Grewal SS, Wharen R, Ritaccio AL (2020)
 828 Examining the function of the visual word form area with stereo EEG electrical stimulation:
 829 A case report of pure alexia. *Cortex* 129:112–118.
- 830 Saygin ZM, Osher DE, Norton ES, Yousoufian DA, Beach SD, Feather J, Gaab N, Gabrieli
 831 JDE, Kanwisher N (2016) Connectivity precedes function in the development of the visual
 832 word form area. *Nat Neurosci* 19:1250–1255.
- 833 Srihasam K, Vincent JL, Livingstone MS (2014) Novel domain formation reveals proto-
 834 architecture in inferotemporal cortex. *Nat Neurosci* 17:1776–1783.
- 835 Stevens WD, Kravitz DJ, Peng CS, Tessler MH, Martin A (2017) Privileged functional

- connectivity between the visual word form area and the language system. *J Neurosci*
37:5288–5297.
- Stevens WD, Tessler MH, Peng CS, Martin A (2015) Functional connectivity constrains the
category-related organization of human ventral occipitotemporal cortex. *Hum Brain Mapp*
36:2187–2206.
- Susilo T, Duchaine B (2013) Dissociations between faces and words: Comment on Behrmann
and Plaut. *Trends Cogn Sci* 17:545.
- Susilo T, Wright V, Tree JJ, Duchaine B (2015) Acquired prosopagnosia without word
recognition deficits. *Cogn Neuropsychol* 32:321–339.
- Tadel F, Baillet S, Mosher JC, Pantazis D, Leahy RM (2011) Brainstorm: a user-friendly
application for MEG/EEG analysis. *Comput Intell Neurosci* 2011:879716.
- Tang H, Buia C, Madhavan R, Crone NE, Madsen JR, Anderson WS, Kreiman G (2014)
Spatiotemporal Dynamics Underlying Object Completion in Human Ventral Visual Cortex.
Neuron 83:736–748.
- Thorpe S, Fize D, Marlot C (1996) Speed of processing in the human visual system. *Nature*
381:520–522.
- Tsao DY, Moeller S, Freiwald WA (2008) Comparing face patch systems in macaques and
humans. *Proc Natl Acad Sci U S A* 105:19514–19519.
- Vinckier F, Dehaene S, Jobert A, Dubus JP, Sigman M, Cohen L (2007) Hierarchical Coding of
Letter Strings in the Ventral Stream: Dissecting the Inner Organization of the Visual Word-
Form System. *Neuron* 55:143–156.

- 857 Weiner KS, Grill-Spector K (2010) Sparsely-distributed organization of face and limb
858 activations in human ventral temporal cortex. *Neuroimage* 52:1559–1573.
- 859 White AL, Palmer J, Boynton GM, Yeatman JD (2019) Parallel spatial channels converge at a
860 bottleneck in anterior word-selective cortex. *Proc Natl Acad Sci U S A* 116:10087–10096.
- 861 Zhen Z, Yang Z, Huang L, Kong X zhen, Wang X, Dang X, Huang Y, Song Y, Liu J (2015)
862 Quantifying interindividual variability and asymmetry of face-selective regions: A
863 probabilistic functional atlas. *Neuroimage* 113:13–25.
- 864

865 Table 1. iEEG participant coverage

Number	Tasks completed	Electrodes in VTC	Face-selective	Word-selective	House-selective	Word medial to face-selective	Alternating word- and face-selective
1	CL	L: 6	0	0	0	N/A	N/A
2	CL	L: 11	0	0	0	N/A	N/A
3	CL	L: 34, R: 18	0	0	L: 2, R: 2	N/A	N/A
4	CL	L: 20, R: 14	0	0	R: 2	N/A	N/A
5	CL	R: 18	R: 2	0	0	N/A	N/A
6 (Fig. 4B)	CL	L: 11	L: 1	L: 1	0	Yes	N/A
7	CL, WID	L: 17	L: 2	L: 1	0	No	No
8	CL	R: 9	0	0	R: 2	N/A	N/A
9	CL, WID	R: 21	0	R: 1	0	N/A	N/A
10 (Fig. 4B)	CL, WID, FID	L: 25, R: 16	L: 2, R: 1	L: 2	0	Yes	Yes
11	CL, FID	L: 4, R: 23	R: 5	L: 1, R: 1	R: 5	Yes	Yes
12	CL, FID	R: 42	R: 8	R: 4	R: 6	Yes	Yes
13	CL, FID	L: 38	0	L: 2	L: 2	N/A	N/A
14	CL, FID	L: 23, R: 24	L: 2, R: 1	0	L: 2, R: 2	N/A	N/A
15	CL, FID	L: 30	L: 1	0	L: 2	N/A	N/A
16	CL, FID	L: 23, R: 11	0	L: 1	R: 3	N/A	N/A
17 (Fig. 4B)	CL, WID, FID	L: 48	L: 6	L: 4	L: 2	Yes	Yes
18	CL, FID	L: 23	0	0	L: 7	N/A	N/A
19	CL	L: 4	0	L: 2	L: 2	N/A	N/A
20	CL	L: 23	0	0	0	N/A	N/A
21	CL	R: 11	0	0	R: 1	N/A	N/A
22	CL, WID, FID	R: 41	0	R: 6	0	N/A	N/A
23	CL	L: 10	L: 1	L: 2	0	No	No
24	CL, FID	L: 26, R: 25	L: 3, R: 1	R: 2	R: 1	Yes	Yes
25	CL, WID, FID	L: 21, R: 19	0	L: 6, R: 1	0	N/A	N/A
26	CL	L: 21, R: 28	L: 2	L: 3	R: 3	No	No
27	CL, FID	L: 5, R: 18	0	L: 1, R: 5	R: 3	N/A	N/A
28 (Fig. 4A)	CL, WID, FID	L: 55	L: 6	L: 4	0	Yes	Yes
29	CL, FID	L: 42	L: 2	L: 2	0	Yes	No
30	CL, FID	L: 26, R: 28	L: 1, R: 2	R: 1	L: 2, R: 1	Yes	No
31	CL, WID, FID	L: 19, R: 36	L: 1, R: 6	0	R: 2	N/A	N/A
32	CL, WID	L: 10, R: 34	L: 1	0	L: 1, R: 3	N/A	N/A

33	CL, WID, FID	L 39, R 50	0	L: 4	L: 3, R:2	N/A	N/A
34	CL, WID	L 24, R: 27	R: 2	L: 5, R: 2	L: 1, R: 3	No	No
35	CL, FID	L: 116	L: 16	L: 8	L: 6	Yes	Yes
36	CL, WID, FID	L: 129	L: 33	L: 15	L: 12	Yes	Yes
Total: 36	CL: 32, WID: 12, FID: 20	L 883, R: 513	L: 80, R: 28	L: 64, R: 23	L: 44, R: 41	L: 7/10, R: 4/5	L: 5/9, R: 3/5

866 Number of word-, face-, and house-selective electrodes in the left (L) and right (R) hemisphere
867 of each iEEG participant out of the total number of implanted VTC electrodes. All participants
868 underwent a category localizer task (CL) from which word, face, and house-selectivity was
869 determined by comparing electrode responses to six categories of objects (see *Materials and*
870 *Methods* and Figure 1B). The table indicates whether any word-selective electrodes were medial
871 to any face-selective electrodes in participants that had at least one word- and one face-selective
872 electrode within the same hemisphere. The table also indicates whether there was alternation of
873 word- and face-selective electrodes along the medial to lateral axis within participants that had at
874 least two word-selective electrodes and one face-selective electrode within the same hemisphere
875 or vice-versa. Participants with insufficient coverage of word or face-selective regions to
876 determine their relative anatomical location are listed as not available (N/A). A subset of
877 participants also participated in a face individuation task (FID) or word individuation task (WID)
878 from which the individuation capacity of word- and face-selective electrodes was tested.
879 Participants illustrated in figures are noted next to the corresponding participant number.

880 Figure 1. Spatial topography of word- and face-selective iEEG electrodes

881

882 A) Heat map of electrode coverage (both category-selective and non-selective) across 36 iEEG
883 participants. Electrodes below the inferior temporal sulcus and anterior to the posterior edge of
884 the fusiform gyrus on the participant's native space were considered VTC. There was a greater
885 proportion of left hemisphere coverage relative to right hemisphere coverage. B) Electrodes that
886 responded preferentially to words, faces, or houses and could significantly discriminate these
887 stimuli from all others using six-way Naïve Bayes classification ($p < 0.05$, Bonferroni corrected
888 within participant). House-selective electrodes are primarily medial to word- and face-selective
889 electrodes. Multiple adjoining word- and face-selective patches are found along the medio-lateral
890 axis of ventral temporal cortex, bilaterally. Stereotactic EEG electrodes have been brought to the
891 ventral surface for clarity.

892 Figure 2. Distribution of face-selective and word-selective electrodes by participant

893

894 Distribution of highly face-selective (left) and word-selective (right) electrodes by participant.

895 Color represents individual participants and corresponds across figure panels. Each group-level

896 cluster of word- and face-selective electrodes is represented by data from multiple participants.

897 Figure 3. Independence of word- and face-processing networks

898

899 A) Average decoding time-course for word- (top, $n = 39$) and face- (bottom, $n = 75$) selective
 900 electrodes that contained significant amounts of information about the other object category. 21
 901 out of 28 (75 %) face-selective electrodes in the right hemisphere and 54 out of 80 (66 %) in the
 902 left hemisphere could significantly discriminate words from the other object categories excluding
 903 faces (e.g. d' sensitivity for words was above chance for 5-way classification of the non-face
 904 object categories) at the $p < 0.05$ level (Bonferroni corrected for multiple comparisons in time
 905 and electrodes within participant, see *Materials and Methods*). 9 out of 23 (39 %) word-selective
 906 electrodes in the right hemisphere and 30 out of 64 (47 %) in the left hemisphere could
 907 discriminate faces from the other object categories excluding words. Error bars indicate standard
 908 error from the mean across electrodes. Colored bars under the curves indicate times where the
 909 average selectivity is above chance ($p < 0.001$ corrected for temporal comparisons). B) Average
 910 decoding time-course for word- (top, $n = 48$) and face- (bottom, $n = 33$) selective electrodes that
 911 did not contain above chance information for the other object category. Although decoding
 912 accuracy was above chance at later time points for the non-preferred category across the
 913 population of electrodes, decoding accuracy was much smaller for the non-preferred compared to
 914 preferred category. C) Example decoding time courses from three highly word-selective
 915 electrodes that did not display face-selectivity. D) Decoding time courses of three highly face-
 916 selective electrodes that did not display word-selectivity. The patient from which the middle
 917 recording was obtained was not presented with pictures of hammers.

918 Figure 4. Multiple adjoining word- and face-selective patches in individual participants

919

920 A) Representative single participant demonstrated alternating bands of word- and face-selectivity
921 along the left fusiform gyrus. Major VTC sulci (collateral sulcus [COS], midfusiform sulcus
922 [MFS], and occipitotemporal sulcus [OTS]) have been outlined for clarity. Shaded electrodes are
923 those selective to words (yellow) and faces (blue). Non-filled circles represent ventral temporal
924 electrodes that did not reach the selectivity criterion for either of these categories. Raw event-
925 related broadband activity is shown for each of the numbered electrodes in the right panel.
926 Moving from medial to lateral, electrodes demonstrate a preferential response to words, a
927 preferential response to faces, a mixed response to both words and faces, then preferential
928 response to words. B) Three additional examples of patients with multiple adjoining word- and
929 face-selective regions or word-selectivity medial to face-selectivity in VTC. Major VTC sulci
930 have been labeled for clarity, as in A.

931 Figure 5. Interdigitation of BOLD responses to words and faces across eight healthy participants

932

933 Eight healthy participants that underwent an identical category localizer task as the iEEG
934 participants demonstrated similar category selectivity. A) Individual maps demonstrate word
935 versus face-selectivity across left VTC. In six out of eight of these participants there was strong
936 word-selectivity medial to face-selectivity along the midfusiform sulcus. The bar graphs below
937 each participant indicates the selectivity of these word-selective regions when defining them
938 based on one half of the data and testing on the other half of the data. Word-selective responses
939 were less consistent in the right hemisphere across participants. B) Bar graph of word-selectivity
940 in left hemisphere medial word-selective regions across participants for the left-out half of the
941 data. Symbols: ** $p < 0.01$, *** $p < 0.001$.

942

943 Figure 6. Comparing category-selectivity in single-trial potentials and high-frequency broadband
944
945 Differing spatial distribution of electrodes that demonstrated selectivity in single-trial potentials
946 (stP) and single-trial high-frequency broadband activity (stHFBB). A) Electrodes that
947 demonstrated selectivity in both stP and stHFBB were clustered around the fusiform and
948 parahippocampal gyri. B) Electrodes selective in only stP were much more widely distributed
949 and included medial and anterior word-selective regions not typically seen in fMRI. C)
950 Electrodes that were only selective in stHFBB were less prevalent than those only selective in
951 stP, but also extended outside of the fusiform and parahippocampal gyri.

952 Figure 7. Differences in the dynamics of left versus right word- and face-selective regions

953

954 Latency of word (yellow) and face (blue) sensitivity onset, latency of peak sensitivity, and
955 magnitude of peak sensitivity across hemispheres. Latency of sensitivity onset is defined as the
956 first timepoint that reached a d' corresponding to $p < 0.001$ non-parametrically defined by the
957 pre-stimulus baseline period. All time points reference the leading edge of the classification
958 window. Box plots represent median, 25th and 75th percentiles. Summary statistics of each box
959 plot are presented in the table. Abbreviations: confidence interval (CI), degrees of freedom (d.f.),
960 left (L), right (R). Symbols: n.s. $p > 0.05$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

961 Figure 8. Spatiotemporal clustering of word- and face-selective regions

962

963 A) Spatiotemporal clustering of word-selective VTC electrodes. The illustrated clustering
 964 solution was robust to different weightings of spatial and temporal information. Left hemisphere
 965 word-selective electrodes were clustered into four spatial clusters. A cluster was found on the
 966 fusiform gyrus (gray, 21 electrodes from 5 participants), as well as medial (green, 20 electrodes
 967 from 10 participants), anterior (blue, 11 electrodes from 5 participants), and lateral (red, 12
 968 electrodes from 7 participants) to the fusiform gyrus. Right hemisphere word-selective regions
 969 had later onsets and were more clearly separated along the posterior to anterior axis (posterior:
 970 cyan; 8 electrodes from 4 participants, mid: yellow; 8 electrodes from 3 participants, anterior:
 971 magenta; 7 electrodes from 6 participants). B) Average d' timecourse of each group of electrodes
 972 in A when jointly classifying stP and stHFBB. Error bars represent standard error across
 973 electrodes. C) Average d' time course of each group of electrodes when classifying only stP. D)
 974 Average d' time course of each electrodes when classifying only stBB. Word-selective
 975 electrodes on the fusiform demonstrate strong selectivity in both stP and stHFBB, whereas other
 976 regions display distinct dynamics across these signal components. E) Spatiotemporal clustering
 977 of face-selective VTC electrodes. Left hemisphere electrodes were clustered into three spatial
 978 clusters roughly posterior to (green, 21 electrodes from 3 participants), on (gray, 46 electrodes
 979 from 12 participants), and anterior to the fusiform gyrus (blue, 13 electrodes from 7 participants).
 980 Right hemisphere, face-selective electrodes were primarily clustered along the posterior to
 981 anterior VTC axis into four clusters (posterior: cyan; 9 electrodes from 5 participants, mid:
 982 yellow; 13 electrodes from 6 participants and black; 3 electrodes from 2 participants, anterior:
 983 magenta; $n = 3$ electrodes from 3 participants). F) Average d' time course of each group of

984 electrodes illustrated in E when jointly classifying stP and stHFBB. G) Average d' timecourse of
985 each group of electrodes when classifying only stP. H) Average d' time course of each group of
986 electrodes when classifying only stBB. Face-selective electrodes on the fusiform demonstrate
987 strong selectivity in both stP and stHFBB, whereas other regions display distinct dynamics
988 across these signal components.

face-selective electrodes

word-selective electrodes

	Comparison	mean	CI	Test statistic	d.f.	p-value
Onset Latency	L - R Face	-29 ms	53 ms	-1.1	106	0.28
	L - R Word	-133 ms	61 ms	-4.4	85	0.0001
	L Face - L Word	-37 ms	38 ms	-1.9	142	0.056
	R Face - R Word	-141 ms	81 ms	-3.5	49	0.001
Peak Latency	L - R Face	18 ms	57 ms	0.6	106	0.53
	L - R Word	-138 ms	63 ms	-4.3	85	0.0001
	L Face - L Word	-11 ms	41 ms	-0.5	142	0.6
	R Face - R Word	-167 ms	85 ms	-3.9	49	0.0003
Peak Magnitude	L - R Face	0.58	0.39	3.0	106	0.0037
	L - R Word	0.66	0.37	3.5	85	0.0006
	L Face - L Word	0.19	0.30	1.3	142	0.21
	R Face - R Word	0.27	0.29	1.9	49	0.069