Behavioral/Cognitive

# Multiple Adjoining Word- and Face-Selective Regions in Ventral Temporal Cortex Exhibit Distinct Dynamics

**Matthew J. Boring,**[1,2,3] **Edward H. Silson,**[4,5] **Michael J. Ward,**[3] **R. Mark Richardson,**[3,6,7] **Julie A. Fiez,**[1,2,8] **Chris I. Baker,**[4] and **Avniel Singh Ghuman**[1,2,3,8]

[1]Center for Neuroscience at the University of Pittsburgh, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, [2]Center for the Neural Basis of Cognition, Pittsburgh, Pennsylvania 15213, [3]Department of Neurological Surgery, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania 15213, [4]National Institute of Mental Health, National Institutes of Health, Magnuson Clinical Center, Bethesda, Maryland 20814, [5]School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh, EH8 9JZ, United Kingdom, [6]Department of Neurosurgery, Massachusetts General Hospital, Boston, Massachusetts 02144, [7]Harvard Medical School, Boston, Massachusetts 02115, and [8]Department of Psychology, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

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

*Key words:* **electrocorticography; fMRI; fusiform-face area; intracranial electroencephalography; ventral stream; visual word-form area**

**Significance Statement**

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

Received Dec. 29, 2020; revised May 26, 2021; accepted June 1, 2021.

Author contributions: M.J.B., E.H.S., J.A.F., C.I.B., and A.S.G. designed research; M.J.B., E.H.S., M.J.W., and R.M.R. performed research; M.J.B. and E.H.S. analyzed data; M.J.B. and A.S.G. wrote the first draft of the paper; M.J.B., E.H.S., R.M.R., J.A.F., C.I.B., and A.S.G. edited the paper; M.J.B. and A.S.G. wrote the paper.

This work was supported by National Institutes of Health R01MH107797 and R21EY030297 and National Science Foundation 1734907 to A.S.G.; National Institutes of Health T32NS007433-20 to M.J.B.; and National Institutes of Health ZIAMH002909 to C.I.B. and E.H.S. We thank the patients, their families, and the clinical staff at the Epilepsy Monitoring Unit at the University of Pittsburgh Medical Center, without whom this study would not be possible; Marlene Behrmann, David Plaut, Michael Tarr, and the anonymous reviewers for helpful suggestions regarding the analysis and interpretations of the intracranial EEG results; and Sean Walls, Ellyanna Kessler, Roma Konecky, and Ashley Whiteman for assistance in intracranial EEG data collection.

The authors declare no competing financial interests.

Correspondence should be addressed to Matthew J. Boring at mjb200@pitt.edu.

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

## Introduction

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

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

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

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

In the present study, we characterized the spatial organization and functional dynamics of word- and face-processing networks within VTC using intracranial EEG (iEEG) data collected from 36 patients with pharmacologically intractable epilepsy and 7 T fMRI data collected from 8 healthy participants.

## Materials and Methods

### Intracranial EEG data collection and preprocessing

*Participants*

Thirty-eight patients (14 males, ages 19-65 years, 32 righthanded) had intracranial surface and/or depth electrodes implanted for the treatment of pharmacologically intractable epilepsy. Depth electrodes were produced by Ad-Tech Medical and PMT and were 0.86 and 0.8 mm in diameter, respectively. Grid electrodes were produced by PMT and were 4 mm in diameter. Because depth electrode contacts are cylindrical, the surface area of the recording site was similar across grid and strip electrode contacts. To be concise, "electrode contacts" are referenced to as "electrodes" throughout the manuscript. No consistent differences in neural responses were observed between grid and depth electrodes. Only electrodes implanted in VTC, defined as below the inferior temporal gyrus and anterior to the posterior tip of the fusiform in the participant-centered space, were considered in this study. Two patients did not have any electrodes within this ROI; therefore, only data from 36 participants were analyzed for this study. Electrodes identified as belonging to the seizure onset zone based on the clinical report or showing epileptiform activity during the tasks were excluded from the analysis. All participants gave written informed consent. The study was approved by the University of Pittsburgh Institutional Review Board. Patients were monetarily compensated for their time.

Electrodes were localized via either postoperative MRI or CT scans coregistered to the preoperative MRI using Brainstorm (Tadel et al., 2011). Surface electrodes were projected to the nearest point on the preoperative cortical surface automatically parcellated via Freesurfer (Dale et al., 1999) to correct for brainshift (Hermes et al., 2010). Electrode coordinates were then coregistered via surface-based transformations to the fsaverage template using Freesurfer cortical reconstructions.

*Experimental design*

All participants underwent a category localizer task where they viewed grayscale images presented on a computer screen positioned 2 m from their face. Images occupied $\sim 6 \times 6$ degrees of visual angle and were presented for 900 ms with 1500 ms interstimulus interval with random 400 ms jitter. Participants were instructed to press a button every time an image was presented twice in a row (1/6 of the trials). These repeat trials were excluded from the analysis yielding 70 trials per stimulus category left for analysis. Several participants underwent multiple runs of this task and therefore had 140-210 trials per stimulus category.

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

A subset of the participants that underwent the category localizer task also participated in word and/or face individuation tasks (Table 1). These tasks shared identical presentation parameters as the category-localizer task (i.e., interstimulus interval, stimulus-on time, and viewing angle) but contained different images. Twelve underwent a word individuation task that included pictures of real words, pseudowords, and consonant-strings or false-fonts. Participants again were instructed to respond if a given stimulus was repeated twice in a row. Every stimulus

**Table 1. iEEG participant coverage[a]**

| 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 | NA | NA |
| 2 | CL | L: 11 | 0 | 0 | 0 | NA | NA |
| 3 | CL | L: 34, R: 18 | 0 | 0 | L: 2, R: 2 | NA | NA |
| 4 | CL | L: 20, R: 14 | 0 | 0 | R: 2 | NA | NA |
| 5 | CL | R: 18 | R: 2 | 0 | 0 | NA | NA |
| 6 (Fig. 4*B*) | CL | L: 11 | L: 1 | L: 1 | 0 | Yes | NA |
| 7 | CL, WID | L: 17 | L: 2 | L: 1 | 0 | No | No |
| 8 | CL | R: 9 | 0 | 0 | R: 2 | NA | NA |
| 9 | CL, WID | R: 21 | 0 | R: 1 | 0 | NA | NA |
| 10 (Fig. 4*B*) | 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 | NA | NA |
| 14 | CL, FID | L: 23, R: 24 | L:2, R: 1 | 0 | L: 2, R: 2 | NA | NA |
| 15 | CL, FID | L: 30 | L: 1 | 0 | L: 2 | NA | NA |
| 16 | CL, FID | L: 23, R: 11 | 0 | L: 1 | R: 3 | NA | NA |
| 17 (Fig. 4*B*) | CL, WID, FID | L: 48 | L: 6 | L: 4 | L: 2 | Yes | Yes |
| 18 | CL, FID | L: 23 | 0 | 0 | L: 7 | NA | NA |
| 19 | CL | L: 4 | 0 | L: 2 | L: 2 | NA | NA |
| 20 | CL | L: 23 | 0 | 0 | 0 | NA | NA |
| 21 | CL | R: 11 | 0 | 0 | R: 1 | NA | NA |
| 22 | CL, WID, FID | R: 41 | 0 | R: 6 | 0 | NA | NA |
| 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 | NA | NA |
| 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 | NA | NA |
| 28 (Fig. 4*A*) | 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 | NA | NA |
| 32 | CL, WID | L: 10, R: 34 | L: 1 | 0 | L: 1, R: 3 | NA | NA |
| 33 | CL, WID, FID | L 39, R 50 | 0 | L: 4 | L: 3, R:2 | NA | NA |
| 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 |

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

(i.e., individual word) was presented 60 times. Twenty underwent a face individuation task where they viewed individuals of varying identity and emotions. Participants were instructed to indicate whether each face was male or female during this task. Each identity was repeated 60 times.

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

*Determining language laterality*

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

*Multivariate temporal pattern analysis*

To determine which electrodes contained information about word and face categories, leave-one trial out cross-validated Gaussian Naive Bayes classifiers were used to predict the category of object participants were viewing given a sliding 100 ms of neural activity from one iEEG electrode during the category-localizer task (six-way classification). Signals from stP and stHFBB were both fed in as features to a single classifier for the main selectivity maps. This procedure was repeated from 100 ms before 900 ms after stimulus onset with 10 ms time step to derive a time course of decoding at each VTC electrode. We also ran separate classifiers on only features from stP or stHFBB to investigate the independent sources of information contained within these signal components. We

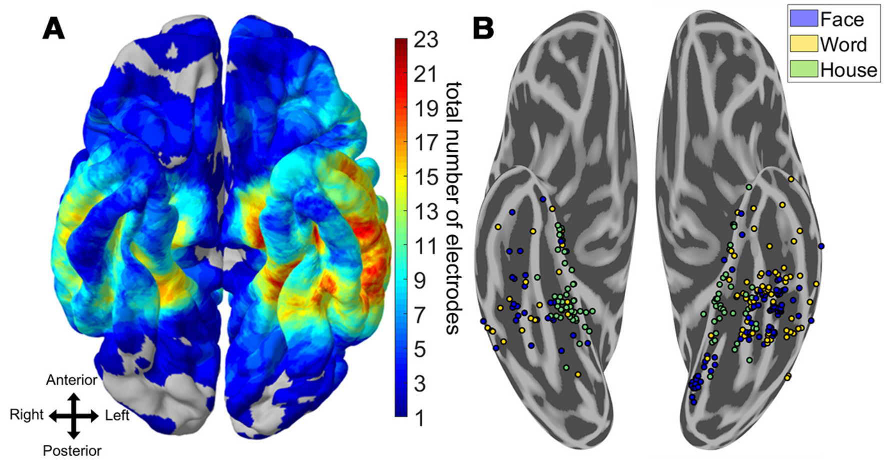



**Figure 1.** Spatial topography of word- and face-selective iEEG electrodes. ***A***, Heat map of electrode coverage (both category-selective and nonselective) across 36 iEEG participants. Electrodes below the inferior temporal sulcus and anterior to the posterior edge of the fusiform gyrus on the participant's native space were considered VTC. There was a greater proportion of left hemisphere coverage relative to right hemisphere coverage. ***B***, Electrodes that responded preferentially to words, faces, or houses and could significantly discriminate these stimuli from all others using six-way Naive Bayes classification ($p < 0.05$, Bonferroni-corrected within participant). House-selective electrodes are primarily medial to word- and face-selective electrodes. Multiple adjoining word- and face-selective patches are found along the mediolateral axis of VTC, bilaterally. Stereotactic EEG electrodes have been brought to the ventral surface for clarity.

ensured the number of features fed into these two types of classifiers was consistent by averaging 10 ms bins of stP, since stHFBB was sampled only every 10 ms, before classification.

Face-selective iEEG electrodes were defined as those that achieved a peak sensitivity ($d'$) 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), as follows:

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

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

To determine whether word- and face-selective electrodes contained exemplar-level information about either faces or words, we performed pairwise classification of the face and word individuation stimuli for the electrodes on which we had data (Table 1). Specifically, in the case of word individuation, we used threefold cross-validated Gaussian Naive Bayes classifiers to predict which of two real words a participant was viewing based on sliding 100 ms of data from the word-selective electrodes. Threefold cross-validation was used instead of leave-one-out cross validation (which was used for assessing category-level selectivity) to save computational time as there were many more models (stimulus pairs) tested with the exemplar classifier. We repeated this procedure across all pairs of real-words of the same length and averaged the time courses of this pairwise decoding (56 pairs of words). We determined the $p < 0.05$ chance-level of this average pairwise decoding by repeating this procedure 1000 times on data with shuffled trial labels in a subset of the word-selective electrodes (Maris and Oostenveld, 2007). These global null distributions were similar across the randomly subsampled electrodes; therefore, we chose a $d'$ threshold corresponding to the highest $p < 0.05$ level obtained from this randomly chosen subset. We ran similar pairwise decoding and threshold definition on real-word versus pseudowords of the same length (36 pairs) and real-word versus false-font stimuli (136 pairs) to determine whether electrodes that could not individuate real-words could perform these finer discriminations compared with those tested in the category localizer task.

Similarly, for face individuation, we performed pairwise decoding of face stimuli during sliding 100 ms time windows of face-selective electrode activity. We then averaged these time courses across all 120 pairwise face classifications and calculated the $p < 0.05$ corrected level by repeating the permutation analysis described for the word individuation task on a random subset of face-selective electrodes.

**Spatiotemporal *k*-means clustering**

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

To determine the optimal weighting of spatial and temporal signal components and optimal number of clusters, we calculated the total spatial and temporal variance explained by the clustering solutions run with several spatial weighting parameters. This was performed for $k$ = 1-10 clusters per hemisphere per faces or words. The elbow method was used to determine the optimal number of clusters per hemisphere. The optimal number of clusters was 4 for right hemisphere face-selective electrodes, 3 for right hemisphere word-selective electrodes, 3 for left hemisphere face-selective electrodes, and 4 for left hemisphere word-selective electrodes. We chose the spatial weighting parameter that explained the maximum amount of variance across $k$ = 3 or 4 clusters per hemisphere per category (spatial weight = 300). Small deviations in the spatiotemporal weighting parameter did not strongly affect the overall organization of spatiotemporal clusters. The dynamics of these electrode clusters were then determined by averaging the selectivity time courses ($d'$ derived using multivariate temporal pattern analysis) across the electrodes belonging to each cluster.

**Statistical analyses**

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

**fMRI data collection and preprocessing**

*Participants*

Eight participants (6 females, mean age 25 years) participated in the fMRI experiment. All participants were right-handed, had normal or corrected-to-normal vision, and gave written informed consent. The National institutes of Health Institutional Review Board approved the consent and protocol (protocol 93 M-0170, clinical trials #NCT00001360). Participants were monetarily compensated for their time.

*fMRI scanning parameters*

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

*Experimental paradigm*

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

*fMRI data preprocessing*

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

*Statistical analysis*

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

*Split-half analysis*

For each participant, the 10 localizer runs were divided into odd and even splits. In each split, we performed the same GLM analysis as described above and looked for significant voxels for the contrast of words versus faces. Despite having only half of the data, we observed significant word-selectivity that was medial of face-selectivity consistently across participants. In order to quantify this selectivity in an independent manner, we first defined medial word-selective regions within a split (e.g., odd) and then sampled the data from the other half (e.g., even). ROIs were defined using data spatially smoothed with a 2 mm Gaussian kernel to generate spatially contiguous clusters, whereas the test data were not spatially smoothed. To avoid any bias in node selection, this process was then reversed and the average computed. Within each ROI, we calculated the average $t$ value for each condition versus baseline.

## Results

From 1396 intracranial electrode contacts implanted within or on the surface of VTC of 36 patients, we isolated those implanted in regions that were highly selective for faces, words, or houses. Highly face-selective electrodes were defined as those that had both (1) single-trial responses that could significantly discriminate face presentations from presentations of five other object categories (words, houses, bodies, hammers, and phase-scrambled objects; $p < 0.05$ level, Bonferroni-corrected for multiple spatial and temporal comparisons within participant; see Materials and Methods) and (2) responded maximally to faces compared with all other object categories on average. This ensured that electrodes designated as highly "face-selective" were those that responded maximally and were significantly selective for faces compared with the five other object categories. An identical procedure was used to define word- and house-selective electrodes.

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

To assess how word- and face-processing networks organize relative to one another, the spatial topography of word-, face-, and house-selective electrodes was examined. At the group level, selectivity to house stimuli was found primarily along the left and right parahippocampal gyrus, with some cases where selectivity extended into the collateral sulcus and medial fusiform gyrus. These patches were generally medial to word- and face-selective locations, consistent with previous fMRI and iEEG studies (Halgren et al., 1994; Haxby et al., 1996; Aguirre et al., 1998;

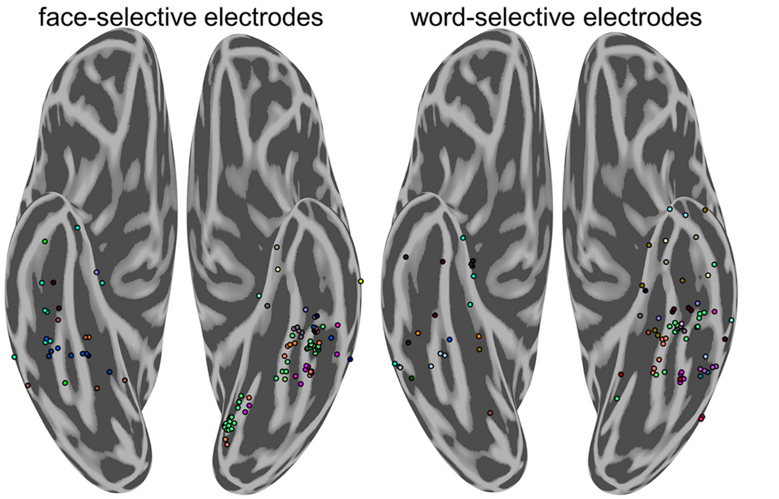



**Figure 2.** Distribution of face-selective and word-selective electrodes by participant. Distribution of highly face-selective (left) and word-selective (right) electrodes by participant. Color represents individual participants and corresponds across figure panels. Each group-level cluster of word- and face-selective electrodes is represented by data from multiple participants.

Cohen et al., 2000; Kadipasaoglu et al., 2016). Face-selectivity was found primarily along the left and right fusiform gyrus with some face-selective regions within the lingual gyrus, and occipitotemporal sulcus (Fig. 1B). Consistent with prior findings (Cohen et al., 2000), word-selective regions were found on the lateral bank of the fusiform and into the occipitotemporal sulcus in the left hemisphere. Word-selective regions were also found anterior to most prior reports from fMRI, in locations that generally have poor signal because of susceptibility artifacts (Devlin et al., 2000). In contrast to most maps of word- and face-selective regions obtained from fMRI (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), a mosaic of word-selective regions were also found medial to face-selective regions, on the medial bank of the fusiform and into the collateral sulcus. Each of these face-, word-, and house-selective regions were found in multiple participants (Fig. 2), demonstrating relatively consistent localization of these regions at a group level.

Interdigitation of word- and face-selective regions was seen in the left hemisphere of 5 of 9 participants with at least two word-selective electrodes and one face-selective electrode or vice-versa and in the right hemisphere of 3 of 5 such participants (Table 1; for examples, see Fig. 4). Word-selective regions were found strictly medial to face-selective regions in the left hemisphere of 7 of 10 participants with at least one word- and one face-selective electrode and in right hemisphere of 4 of 5 participants (Table 1; for an example, see Fig. 4). Thus, highly word-selective regions medial to face-selective regions were not simply a consequence of individual variability in a group-level map but instead were 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 whether a similar organization existed in healthy participants using the higher resolution of 7 T fMRI. When contrasting responses to words and faces in 8 participants, face-selectivity was primarily centered on the midfusiform sulcus while word-selectivity was greatest in the occipitotemporal sulcus (Fig. 5). Consistent with the iEEG results, 6 of the 8 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 one-third the size of more lateral word-selective regions (mean size of lateral word-selective regions: 398 voxels; standard error (SE): 43 vs medial regions: 139 voxels; SE: 29 voxels; $p < 0.01$). Also, 7 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 (Fig. 1B). Together, 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).

The maps in Figures 1-3 were made by combining two key aspects of the iEEG signals, the stP and the stHFBB, to examine the category-selectivity of the underlying VTC neural populations in aggregate across these signal components. Studies have shown that, while category-selectivity demonstrated in stP and stHFBB often overlap, they are not redundant (Engell and McCarthy, 2011; Engell et al., 2012; Miller et al., 2016), suggesting that stP and stHFBB have at least partially distinct physiological generators. To examine these signal components separately, we trained multivariate classifiers solely on stP or stHFBB and isolated electrodes that were selective in either signal component using the same criteria as before (single-trial discriminability and highest signal amplitude for words, faces, or houses). Fifty-eight electrodes showed significant selectivity in both stP and stHFBB (Fig. 6A). Notably, the regions that demonstrated selectivity in both stP and stHFBB were those most often identified in canonical maps of category-selectivity based on fMRI (Cohen et al., 2000; Vinckier et al., 2007; Tsao et al., 2008; Weiner and Grill-Spector, 2010; Lerma-Usabiaga et al., 2018). Specifically, house-selectivity was restricted to the parahippocampal cortex, face-selectivity was primarily restricted to the fusiform bilaterally, and word-selectivity was restricted primarily to the left posterior-lateral fusiform and occipitotemporal sulcus. Regions that were less consistent with canonical fMRI maps tended to be those that were not significantly selective in both stP and stHFBB. For example, the medial word-selective patches were primarily seen in stP alone (Fig. 6B), whereas anterior and right hemisphere word-selectivity was prevalent in either stP or stHFBB alone (Fig. 6B,C). Broadly, more electrodes demonstrated selectivity in stP (232 electrodes from 32 participants; Fig. 6B) compared with stHFBB (115 electrodes from 24 participants; Fig. 6C). More widespread stP selectivity is consistent with a previous study comparing stP and stHFBB responses for faces in VTC, although that study did not observe any cases where selectivity for faces was demonstrated in stHFBB but not stP (Engell and McCarthy, 2011). The similarities and differences in selectivity demonstrated in stHFBB and stP are consistent with the

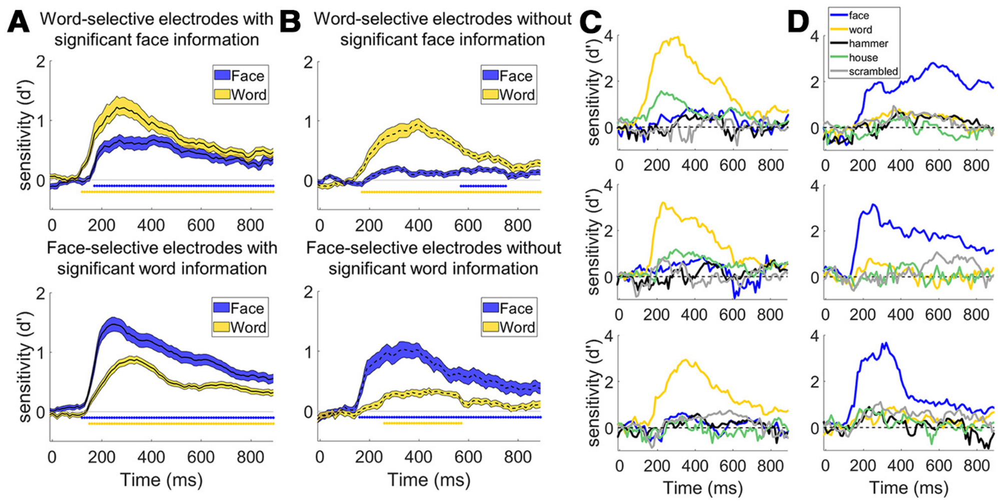



**Figure 3.** Independence of word- and face-processing networks. ***A***, Average decoding time course for word-selective (top, $n = 39$) and face-selective (bottom, $n = 75$) electrodes that contained significant amounts of information about the other object category. Twenty-one of 28 (75%) face-selective electrodes in the right hemisphere and 54 of 80 (66%) in the left hemisphere could significantly discriminate words from the other object categories excluding faces (e.g., $d'$ sensitivity for words was above chance for five-way classification of the nonface object categories) at the $p < 0.05$ level (Bonferroni-corrected for multiple comparisons in time and electrodes within participant; see Materials and Methods). Nine of 23 (39%) word-selective electrodes in the right hemisphere and 30 of 64 (47%) in the left hemisphere could discriminate faces from the other object categories excluding words. Error bars indicate SE across electrodes. Colored bars under the curves represent times where the average selectivity is above chance ($p < 0.001$ corrected for temporal comparisons). ***B***, Average decoding time course for word-selective (top, $n = 48$) and face-selective (bottom, $n = 33$) electrodes that did not contain above chance information for the other object category. Although decoding accuracy was above chance at later time points for the nonpreferred category across the population of electrodes, decoding accuracy was much smaller for the nonpreferred compared with preferred category. ***C***, Example decoding time courses from three highly word-selective electrodes that did not display face-selectivity. ***D***, Decoding time courses of three highly face-selective electrodes that did not display word-selectivity. The patient from which the middle recording was obtained was not presented with pictures of hammers.

hypothesis that these signals have different physiological generators (Lachaux et al., 2005), which may differ in their laminar distribution (Leszczyński et al., 2020) and spatial signal-to-noise falloff (Engell and McCarthy, 2011). Additionally, different category-selectivity across these iEEG signal components may also help explain differences between category-selectivity maps drawn from iEEG and fMRI, as some studies suggest that fMRI has differential sensitivity to these aspects of the iEEG signal (Conner et al., 2011; Engell et al., 2012; Jacques et al., 2016).

One question is whether word- and face-selective regions identified using iEEG discriminate between individual face and word exemplars, respectively. Classifying at the exemplar level also can address the potential concern that the word- and face-selective regions identified using iEEG may be responding to low-level features that drastically differ between the sampled image categories. A subset of the iEEG participants underwent independent word and face individuation tasks (see Materials and Methods; Table 1). Activity from 85 of 97 sampled face-selective electrodes in 13 participants could be used to reliably predict the identity of a presented face ($p<0.05$, permutation test). Similarly, activity from 40 of 53 sampled word-selective electrodes from 10 participants could be used to discriminate single words of the same length from one another. Of those 13 word-selective electrodes that could not reliably achieve word individuation, 6 could reliably discriminate pseudowords from real words of the same length, 7 could reliably discriminate false-fonts from real words. Therefore, most of the word- and face-selective regions mapped with iEEG contained reliable exemplar-level information specific to the categories they were selective to.

Peak word and face individuation was significantly correlated with peak category-selectivity in word and face-selective regions for which we had individuation data (word-selective: Spearman's $\rho_{(53)} = 0.50$, $p < 0.0001$, face-selective: $\rho_{(97)} = 0.48$, $p < 0.0001$). Correlations in peak category-selectivity and within-category individuation may arise because of similar differences in measurement noise across recording contacts (e.g., because of the distance the electrode was placed from the underlying face- or word-selective neural populations), underlying neural/physiological factors, or some mix of both.

In addition to the medial band of word-selective regions, there were a high proportion of right word-selective electrodes in our iEEG population (Fig. 1*B*; Table 1). Although this finding is consistent with some other fMRI (Ben-Shachar et al., 2007; White et al., 2019) and iEEG studies (Halgren et al., 1994; Lochy et al., 2018), right hemisphere word-selectivity is often not seen in neuroimaging (Cohen et al., 2000, 2002) and was not very strong in our 7 T fMRI data either (Fig. 5). Twenty-three word-selective electrodes were found across 9 participants in right VTC, of 21 participants with right VTC object-selectivity. This discrepancy between right word-selectivity observed in fMRI and iEEG was also not attributable to participant handedness, since no participant with right word-selective regions was left-handed. Three of 9 of these participants demonstrated evidence for bilateral language function while the other 6 demonstrated left dominant language function determined by preclinical MEG (see Materials and Methods). Across the entire participant population, 7 of 30 iEEG participants with preclinical MEG demonstrated bilateral language function; the others were considered left dominant. One participant with bilateral language function and right hemisphere object-selectivity did not demonstrate right word-selectivity. Overall, neither participant handedness nor language dominance sufficiently explains the high proportion of word-selective regions found in right VTC.

While neither language laterality nor handedness explained right word-selectivity, substantial differences were seen in the dynamics of neural activity recorded from left versus right word-selective regions (Fig. 7). Latency to word-selectivity onset and peak was shorter in left compared with right hemisphere word-selective regions (mean onset latency difference ± 95% CI:

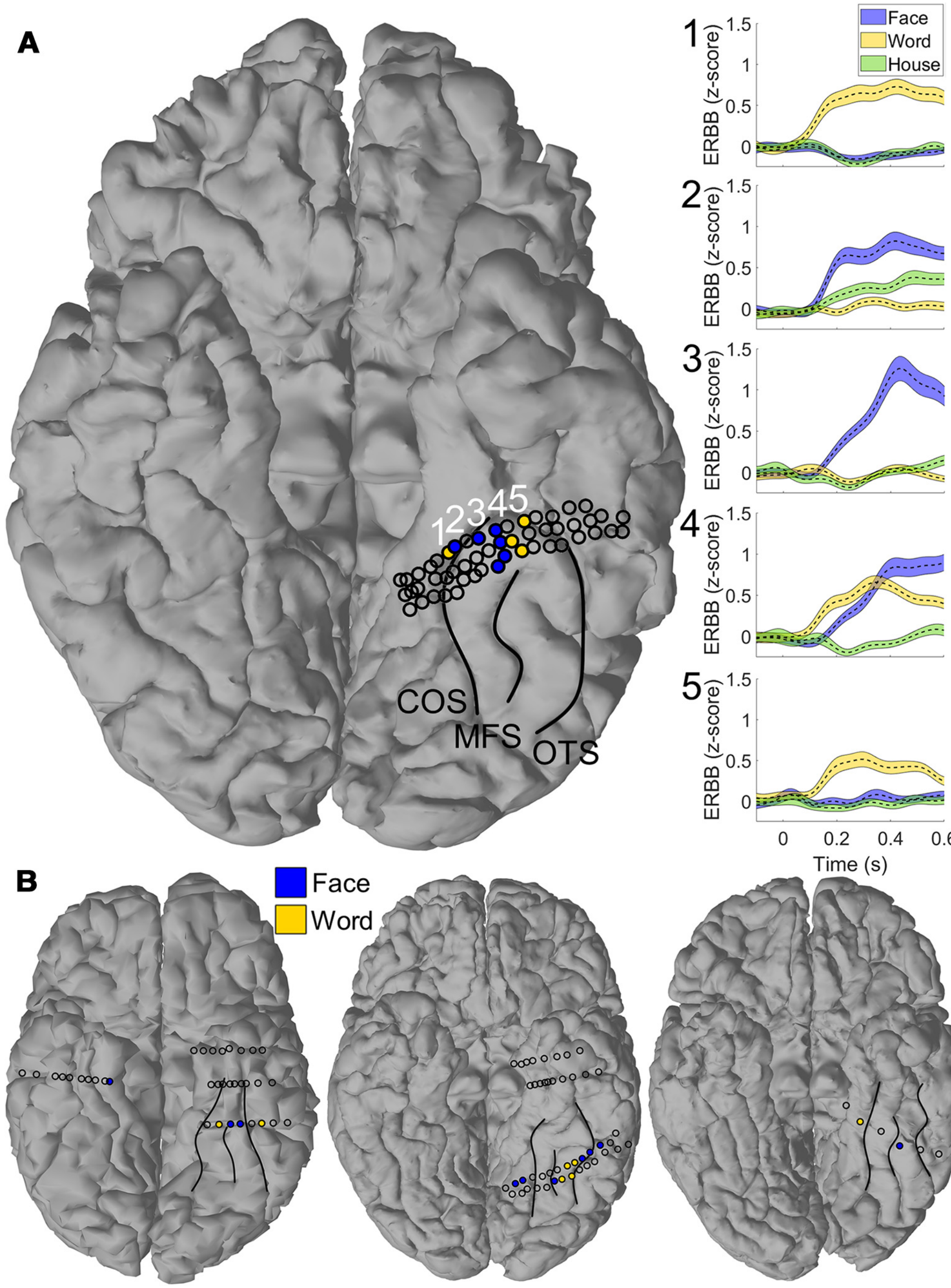



**Figure 4.** Multiple adjoining word- and face-selective patches in individual participants. ***A***, Representative single participant demonstrated alternating bands of word- and face-selectivity along the left fusiform gyrus. Major VTC sulci (collateral sulcus [COS], midfusiform sulcus [MFS], and occipitotemporal sulcus [OTS]) have been outlined for clarity. Shaded electrodes are those selective to words (yellow) and faces (blue). Open circles represent ventral temporal electrodes that did not reach the selectivity criterion for either of these categories. Raw event-related broadband activity (ERBB) is shown for each of the numbered electrodes in the right panel. Error bars represent SE across trials. Moving from medial to lateral, electrodes demonstrate a preferential response to words, a preferential response to faces, a mixed response to both words and faces, then preferential response to words. ***B***, Three additional examples of patients with multiple adjoining word- and face-selective regions or word-selectivity medial to face-selectivity in VTC. Major VTC sulci have been labeled for clarity, as in ***A***.

$-133 \pm 61$ ms, $T_{(85)} = -4.4$, $p < 0.0001$, mean peak latency difference: $-138 \pm 63$ ms, $T_{(85)} = -4.3$, $p < 0.0001$; Fig. 7). These relationships held when taking into account potential differences in posterior to anterior coordinate of word-selective regions across hemispheres (onset: $T_{(85)} = -4.01$, $p = 0.0001$, peak: $T_{(85)} = -3.97$, $p = 0.0002$). There was no significant difference between the latency to peak $d'$ sensitivity or sensitivity onset for right and left face-selective regions (mean onset latency difference: $-29 \pm 53$ ms, $T_{(106)} = -1.1$, $p = 0.28$, mean peak latency difference: $18 \pm 57$ ms, $T_{(106)} = 0.63$, $p = 0.53$; Fig. 7). Additionally, the amplitude of peak $d'$ sensitivity for words was significantly greater in the left compared with right hemisphere word-selective regions (mean peak $d'$ sensitivity difference: $0.66 \pm 0.37$, $T_{(85)} = 3.5$, $p = 0.0006$). The amplitude of peak $d'$ sensitivity to faces was also significantly greater in the left compared with right hemisphere face-selective regions (mean peak $d'$ sensitivity difference: $0.58 \pm 0.39$, $T_{(85)} = 3.0$, $p = 0.0037$). There was a significant correlation between peak latency and peak magnitude within face-selective regions in the left ($\rho_{(80)} = -0.61$, $p < 0.0001$) and right ($\rho_{(28)} = -0.79$,

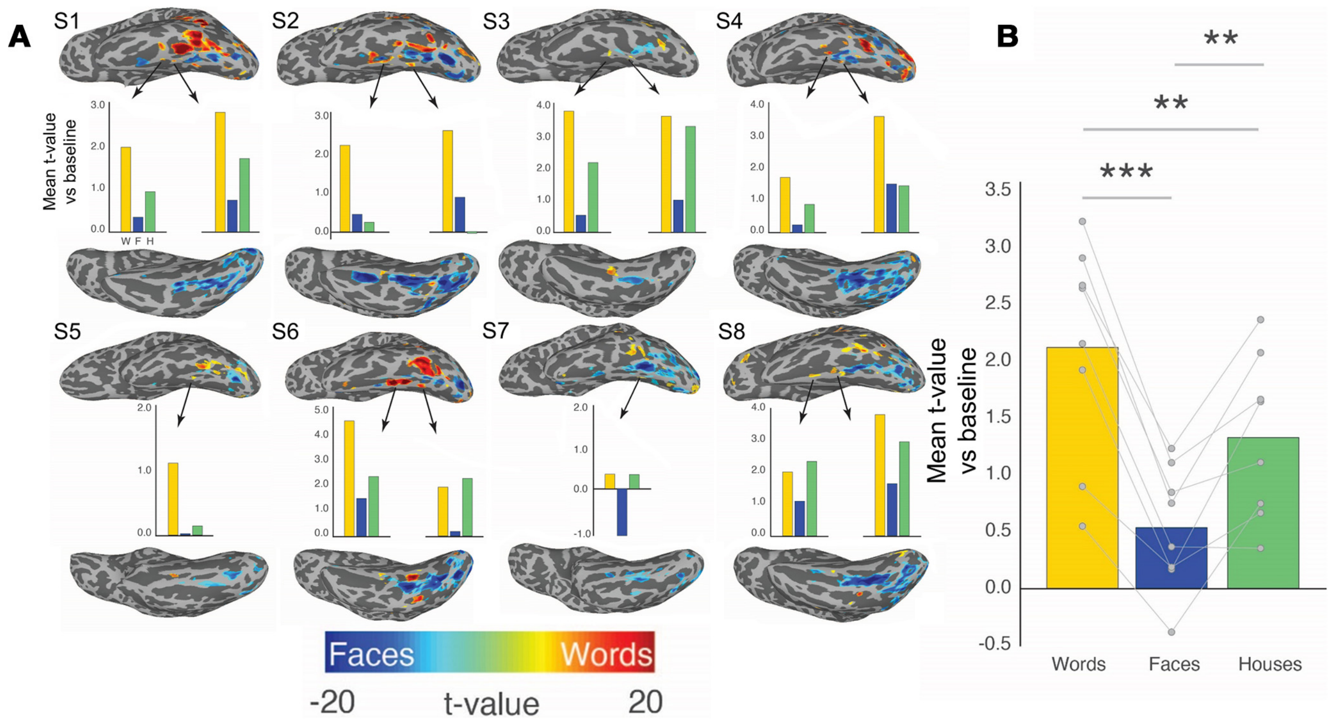



**Figure 5.** Interdigitation of BOLD responses to words and faces across 8 healthy participants. Eight healthy participants that underwent an identical category localizer task as the iEEG participants demonstrated similar category selectivity. ***A***, Individual maps demonstrate word versus face-selectivity across left VTC. In 6 of 8 of these participants, there was strong word-selectivity medial to face-selectivity along the midfusiform sulcus. Bar graphs below each participant represent the selectivity of these word-selective regions when defining them based on one-half of the data and testing on the other half of the data. Word-selective responses were less consistent in the right hemisphere across participants. ***B***, Bar graph of word-selectivity in left hemisphere medial word-selective regions across participants for the left-out half of the data. $^{**}p < 0.01$. $^{***}p < 0.001$.

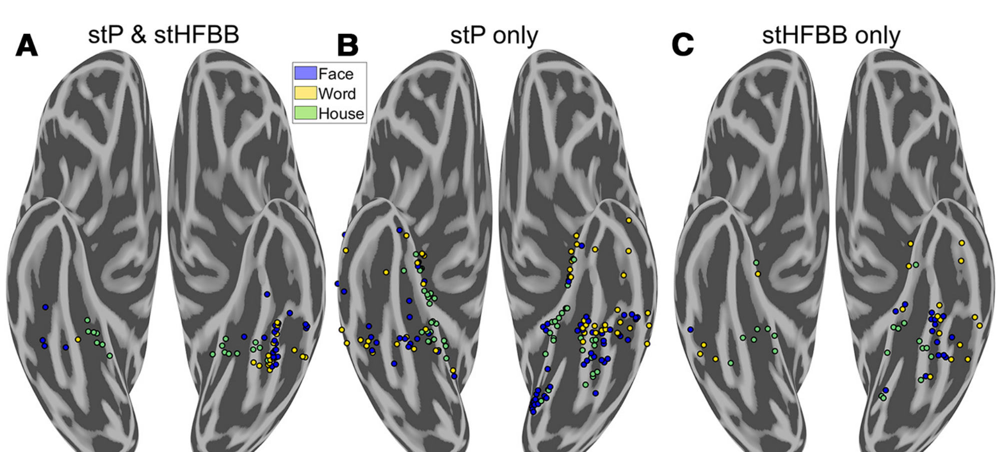
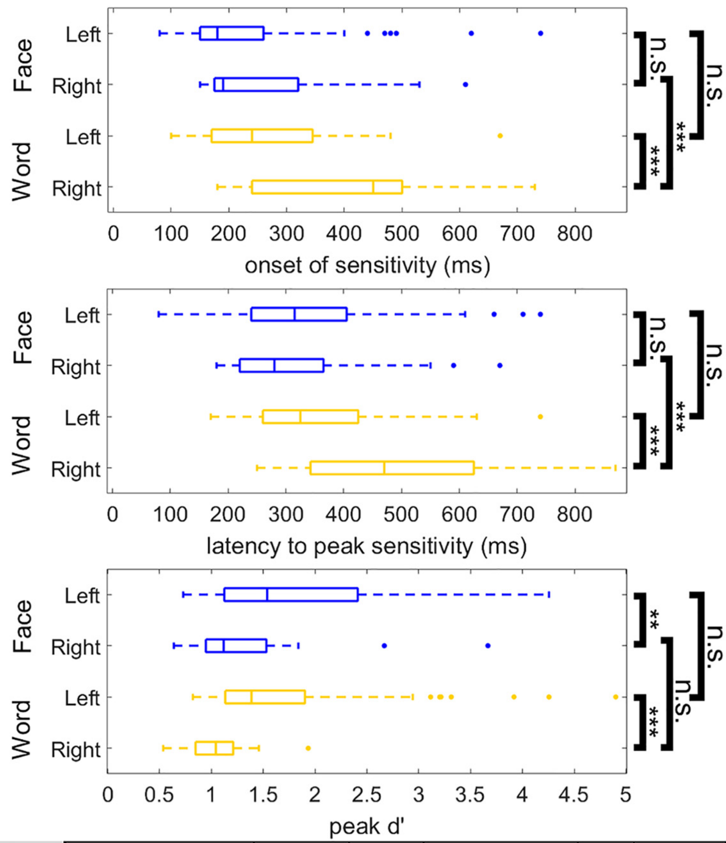



**Figure 6.** Comparing category-selectivity in single-trial potentials (stP) and high-frequency broadband (stHFBB). There were different spatial distributions of electrodes that demonstrated selectivity in stP and stHFBB. ***A***, Electrodes that demonstrated selectivity in both stP and stHFBB were clustered around the fusiform and parahippocampal gyri. ***B***, Electrodes selective in only stP were much more widely distributed and included medial and anterior word-selective regions not typically seen in fMRI. ***C***, Electrodes that were only selective in stHFBB were less prevalent than those only selective in stP, but also extended outside of the fusiform and parahippocampal gyri.

$p < 0.0001$) hemisphere and word-selective regions in the left ($\rho_{(64)} = -0.68$, $p < 0.0001$), but not right ($\rho_{(23)} = -0.15$, $p$ = 0.48) hemisphere, suggesting that longer peak latencies were associated with smaller peak selectivity. These correlations were not significantly different between face-selective regions in the left and right hemisphere ($T_{(85)} = -1.56$, $p = 0.058$), but there was a greater correlation between peak latency and magnitude in left compared with right hemisphere word-selective regions ($T_{(85)} = 2.63$, $p = 0.004$). Given that it was only true for word-selective electrodes, the relatively slower response of right versus left word-selective regions may potentially explain differences in word-selectivity maps derived from iEEG and fMRI and may reflect the left hemisphere bias for language.

Finally, using the iEEG data, we sought to determine whether there were any differences in the temporal dynamics of neural responses across word or face-selective regions within the same hemisphere. We used a spatiotemporal $k$-means clustering algorithm to find spatially contiguous regions of left and right VTC which demonstrated correlated category-selective dynamics. After optimizing the algorithm to capture the most




| | 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 |

**Figure 7.** Differences in the dynamics of left versus right word- and face-selective regions. Latency of word (yellow) and face (blue) sensitivity onset, latency of peak sensitivity, and magnitude of peak sensitivity across hemispheres. Latency of sensitivity onset is defined as the first time point that reached a $d'$ corresponding to $p < 0.001$ nonparametrically defined by the prestimulus baseline period. All time points reference the leading edge of the classification window. Box plots represent median, 25th and 75th percentiles. Summary statistics of each box plot are presented in the table. n.s., Not significant, $p > 0.05$. *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

spatiotemporal variance with the optimal number of clusters (see Materials and Methods), we could compare the dynamics of different word- and face-selective clusters within VTC.

Word-selective regions were clustered into four left hemisphere clusters and three right hemisphere clusters (Fig. 8*A*). Word-selective regions on the left fusiform gyrus demonstrated the earliest and strongest selectivity, peaking at ~200 ms (Fig. 8*B*, gray). Left hemisphere medial word-selective regions and right hemisphere word-selective regions came next, peaking at ~300 ms (Fig. 8*B*, green and cyan) followed by lateral regions at ~350 ms (Fig. 8*B*, red). Word-selective regions in left anterior VTC peaked at ~400-450 ms (Fig. 8*B*, blue); right more anterior regions peaked at ~600 ms (Fig. 8*B*, magenta). When considering word-selectivity dynamics exhibited independently in stP and stHFBB signal components, word-selective electrodes on the fusiform demonstrated strong selectivity in both signal components, whereas other regions displayed distinct dynamics across these signal components (Fig. 8*C*,*D*).

Face-selective regions were organized into three clusters in the left hemisphere and four clusters in the right hemisphere (Fig. 8*E*). Face-selective regions of the left and right fusiform gyrus demonstrated the earliest and largest peak selectivity at ~200-250 ms (Fig. 8*F*, gray and cyan). More anterior right hemisphere regions and a cluster of electrodes in left posteromedial VTC (Fig. 8*F*, yellow and green) peaked at ~300 ms. Finally, more anterior face-selective electrodes in left and right VTC peaked at ~400 ms (Fig. 8*F*, blue, black, and magenta). When considering face-selectivity dynamics exhibited independently in stP and stHFBB signal components, electrodes on the fusiform demonstrated strong selectivity in both components, whereas other regions displayed distinct dynamics across these signal components (Fig. 8*G*,*H*).

From electrodes sampled in the word individuation task, we observed stronger word individuation in left word-selective regions on the fusiform compared with the more medial word-selective cluster illustrated in Figure 8*A* (peak $d'$ of fusiform minus medial regions: $T_{(30)} = 3.62$, $p = 0.001$, linear mixed-effects model). There was no significant difference between the latency to peak word individuation across these clusters ($T_{(30)} = 0.41$, $p = 0.68$). There were not sufficient subjects with electrodes in the other word-selective clusters with word individuation data to make comparisons between all clusters. Neither peak face individuation ($T_{(50)} = 1.03$, $p = 0.31$) nor latency to peak face individuation ($T_{(50)} = -0.21$, $p = 0.84$) was significantly different between face-selective regions along the left fusiform gyrus and the posteromedial face-selective cluster observed in Figure 8*E*. There were not sufficient subjects with electrodes in the other face-selective clusters with face individuation data to make comparisons between all clusters.

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

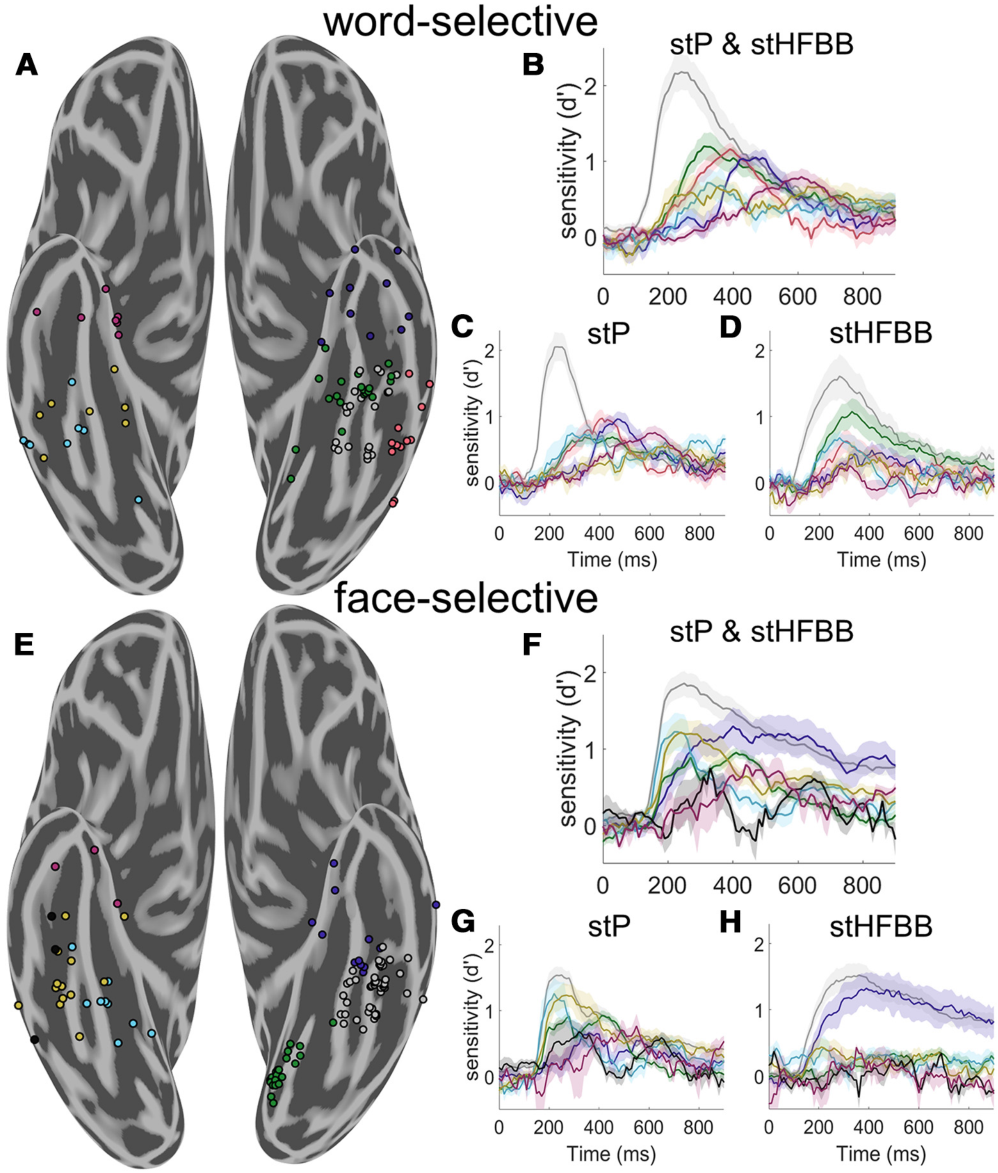



**Figure 8.** Spatiotemporal clustering of word- and face-selective regions. ***A***, Spatiotemporal clustering of word-selective VTC electrodes. The illustrated clustering solution was robust to different weightings of spatial and temporal information. Left hemisphere word-selective electrodes were clustered into four spatial clusters. A cluster was found on the fusiform gyrus (gray, 21 electrodes from 5 participants), as well as medial (green, 20 electrodes from 10 participants), anterior (blue, 11 electrodes from 5 participants), and lateral (red, 12 electrodes from 7 participants) to the fusiform gyrus. Right hemisphere word-selective regions had later onsets and were more clearly separated along the posterior to anterior axis (posterior: cyan; 8 electrodes from 4 participants, mid: yellow; 8 electrodes from 3 participants, anterior: magenta; 7 electrodes from 6 participants). ***B***, Average $d'$ time course of each group of electrodes in ***A*** when jointly classifying stP and stHFBB. Error bars indicate SE across electrodes. ***C***, Average $d'$ time course of each group of electrodes when classifying only stP. ***D***, Average $d'$ time course of each electrode when classifying only stBB. Word-selective electrodes on the fusiform demonstrate strong selectivity in both stP and stHFBB, whereas other regions display distinct dynamics across these signal components. ***E***, Spatiotemporal clustering of face-selective VTC electrodes. Left hemisphere electrodes were clustered into three spatial clusters roughly posterior to (green, 21 electrodes from 3 participants), on (gray, 46 electrodes from 12 participants), and anterior to the fusiform gyrus (blue, 13 electrodes from 7 participants). Right hemisphere, face-selective electrodes were primarily clustered along the posterior to anterior VTC axis into four clusters (posterior: cyan; 9 electrodes from 5 participants, mid: yellow; 13 electrodes from 6 participants, and black; 3 electrodes from 2 participants, anterior: magenta; $n = 3$ electrodes from 3 participants). ***F***, Average $d'$ time course of each group of electrodes illustrated in ***E*** when jointly classifying stP and stHFBB. ***G***, Average $d'$ time course of each group of electrodes when classifying only stP. ***H***, Average $d'$ time course of each group of electrodes when classifying only stBB. Face-selective electrodes on the fusiform demonstrate strong selectivity in both stP and stHFBB, whereas other regions display distinct dynamics across these signal components.

## Discussion

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

In both the iEEG and fMRI data, strong face-selectivity along the fusiform gyrus was adjoining with highly word-selective regions in and around the occipitotemporal and collateral sulci. House-selective regions were found primarily along the parahippocampal gyrus. This organization of house- versus word- and face-selective regions supports that typical retinotopic 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). Because of 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 interparticipant 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 that 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) Others have suggested that lateral word-selective regions are responsible for recognizing word forms, whereas 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; Chen et al., 2008; 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 studies have suggested that fMRI BOLD have differential sensitivity to stHFBB versus stP (Hermes et al., 2012), with some suggesting greater sensitivity to stHFBB (Engell et al., 2012; Jacques et al., 2016). Differential sensitivity to stP and stHFBB may explain why previous fMRI studies have only inconsistently observed medial word-selective regions. Our 7 T fMRI data show that, with adequate power, both lateral and medial word-selective regions are seen in the left hemisphere using BOLD within individual participants. Future studies are necessary to fully understand the functional characteristics and neurophysiological generators of stP and stHFBB iEEG components (Miller, 2010; Ray and Maunsell, 2011; Leszczyński et al., 2020) and how they relate to any differential roles that medial and lateral word-selective regions play in reading.

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

It has previously been suggested that right word-selective regions (along with left posterior word-selective regions) are involved in relatively early visual processing of words, and then this information flows to left anterior word-selective regions (White et al., 2019). However, the dynamics observed here do not support this hypothesis, because left word-selectivity substantially preceded right word-selectivity. Instead, the time course of right hemisphere activation is coincident with P300 and N400 potentials observed during reading, suggesting that right hemisphere word-selective regions may support the left hemisphere in later computations, such as those involving word syntax, memory encoding, and/or semantic processing (Friedman et al., 1975; Kutas and Hillyard, 1980; Federmeier and Kutas, 1999; Otten and Donchin, 2000; Arbel et al., 2011).

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

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

While delays in processing along the posterior-to-anterior VTC axis for both faces and words are somewhat consistent with feedforward models of visual processing, the relative latencies are far longer than would be expected in these models (Thorpe et al., 1996; Kravitz et al., 2013). These results instead suggest more extended dynamics, perhaps governed by recurrent processes (Kravitz et al., 2013), with different category-selective regions contributing differentially to multiple, temporally extended stages of face- and word-processing (Ghuman et al., 2014; Hirshorn et al., 2016; Li et al., 2019). 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.

Bao P, She L, McGill M, Tsao DY (2020) A map of object space in primate inferotemporal cortex. Nature 583:103–108.

Behrmann M, Plaut DC (2013) Distributed circuits, not circumscribed centers, mediate visual recognition. Trends Cogn Sci 17:210–219.

Ben-Shachar M, Dougherty RF, Deutsch GK, Wandell BA (2007) Differential sensitivity to words and shapes in ventral occipito-temporal cortex. Cereb Cortex 17:1604–1611.

Chen R, Hillis AE, Pawlak M, Herskovits EH (2008) Voxelwise Bayesian lesion-deficit analysis. Neuroimage 40:1633–1642.

Cohen L, Dehaene S, Naccache L, Lehéricy S, Dehaene-Lambertz G, Hénaff MA, Michel F (2000) The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. Brain 123:291–307.

Cohen L, Lehéricy S, Chochon F, Lemer C, Rivaud S, Dehaene S (2002) Language-specific tuning of visual cortex? Functional properties of the visual word form area. Brain 125:1054–1069.

Conner CR, Ellmore TM, Pieters TA, di Sano MA, Tandon N (2011) Variability of the relationship between electrophysiology and BOLD-fMRI across cortical regions in humans. J Neurosci 31:12855–12865.

Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 29:162–173.

Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis. Neuroimage 9:179–194.

Dehaene-Lambertz G, Monzalvo K, Dehaene S (2018) The emergence of the visual word form: longitudinal evolution of category-specific ventral visual areas during reading acquisition. PLoS Biol 16:e2004103.

Dehaene S, Cohen L (2011) The unique role of the visual word form area in reading. Trends Cogn Sci 15:254–262.

Dehaene S, Le Clec'H G, Poline JB, Le Bihan D, Cohen L (2002) The visual word form area: a prelexical representation of visual words in the fusiform gyrus. Neuroreport 13:321–325.

Dehaene S, Pegado F, Braga LW, Ventura P, Nunes Filho G, Jobert A, Dehaene-Lambertz G, Kolinsky R, Morais J, Cohen L (2010) How learning to read changes the cortical networks for vision and language. Science 330:1359–1364.

Devlin JT, Russell RP, Davis MH, Price CJ, Wilson J, Moss HE, Matthews PM, Tyler LK (2000) Susceptibility-induced loss of signal: comparing PET and fMRI on a semantic task. Neuroimage 11:589–600.

Enatsu R, Kanno A, Ookawa S, Ochi S, Ishiai S, Nagamine T, Mikuni N (2017) Distribution and network of basal temporal language areas: a study of the combination of electric cortical stimulation and diffusion tensor imaging. World Neurosurg 106:1–8.

Engell AD, Huettel S, McCarthy G (2012) The fMRI BOLD signal tracks electrophysiological spectral perturbations, not event-related potentials. Neuroimage 59:2600–2606.

Engell AD, McCarthy G (2011) The relationship of gamma oscillations and face-specific ERPs recorded subdurally from occipitotemporal cortex. Cereb Cortex 21:1213–1221.

Federmeier KD, Kutas M (1999) Right words and left words: electrophysiological evidence for hemispheric differences in meaning processing. Brain Res Cogn Brain Res 8:373–392.

Fiez JA, Petersen SE (1998) Neuroimaging studies of word reading. Proc Natl Acad Sci USA 95:914–921.

Flinker A, Chang EF, Barbaro NM, Berger MS, Knight RT (2011) Sub-centimeter language organization in the human temporal lobe. Brain Lang 117:103–109.

Fonseca AD, Bénar CG, Bartoloméi F, Régis J, Démonet JF, Chauvel P, Liégeois-Chauvel C (2009) Electrophysiological study of the basal temporal language area: a convergence zone between language perception and production networks. Clin Neurophysiol 120:539–550.

Forseth KJ, Kadipasaoglu CM, Conner CR, Hickok G, Knight RT, Tandon N (2018) A lexical semantic hub for heteromodal naming in middle fusiform gyrus. Brain 141:2112–2126.

Friedman D, Simson R, Ritter W, Rapin I (1975) The late positive component (P300) and information processing in sentences. Electroencephalogr Clin Neurophysiol 38:255–262.

Gaillard R, Naccache L, Pinel P, Clémenceau S, Volle E, Hasboun D, Dupont S, Baulac M, Dehaene S, Adam C, Cohen L (2006) Direct intracranial, fMRI, and lesion evidence for the causal role of left inferotemporal cortex in reading. Neuron 50:191–204.

Gao X, Gentile F, Rossion B (2018) Fast periodic stimulation (FPS): a highly effective approach in fMRI brain mapping. Brain Struct Funct 223:2433–2454.

Geissler A, Lanzenberger R, Barth M, Tahamtan AR, Milakara D, Gartus A, Beisteiner R (2005) Influence of fMRI smoothing procedures on replicability of fine scale motor localization. Neuroimage 24:323–331.

Ghuman AS, Brunet NM, Li Y, Konecky RO, Pyles JA, Walls SA, Destefino V, Wang W, Richardson RM (2014) Dynamic encoding of face information in the human fusiform gyrus. Nat Commun 5:5672.

Ghuman AS, Martin A (2019) Dynamic neural representations: an inferential challenge for fMRI. Trends Cogn Sci 23:534–536.

Gomez J, Natu V, Jeska B, Barnett M, Grill-Spector K (2018) Development differentially sculpts receptive fields across early and high-level human visual cortex. Nat Commun 9:788.

Grill-Spector K, Weiner KS (2014) The functional architecture of the ventral temporal cortex and its role in categorization. Nat Rev Neurosci 15:536–548.

Halgren E, Baudena P, Heit G, Clarke JM, Marinkovic K, Clarke M (1994) Spatio-temporal stages in face and word processing: 1. Depth recorded potentials in the human occipital and parietal lobes. J Physiol Paris 88:1–50.

Harris RJ, Rice GE, Young AW, Andrews TJ (2016) Distinct but overlapping patterns of response to words and faces in the fusiform gyrus. Cereb Cortex 26:3161–3168.

Hasson U, Levy I, Behrmann M, Hendler T, Malach R (2002) Eccentricity bias as an organizing principle for human high-order object areas. Neuron 34:479–490.

Haxby JV, Horwitz B, Ungerleider LG, Maisog JM, Pietrini P, Grady CL (1994) The functional organization of human extrastriate cortex: a PET-rCBF study of selective attention to faces and locations. J Neurosci 14:6336–6353.

Haxby JV, Ungerleider LG, Horwitz B, Maisog JM, Rapoport SI, Grady CL (1996) Face encoding and recognition in the human brain. Proc Natl Acad Sci USA 93:922–927.

Hermes D, Miller KJ, Noordmans HJ, Vansteensel MJ, Ramsey NF (2010) Automated electrocorticographic electrode localization on individually rendered brain surfaces. J Neurosci Methods 185:293–298.

Hermes D, Miller KJ, Vansteensel MJ, Aarnoutse EJ, Leijten FS, Ramsey NF (2012) Neurophysiologic correlates of fMRI in human motor cortex. Hum Brain Mapp 33:1689–1699.

Hirshorn EA, Li Y, Ward MJ, Richardson RM, Fiez JA, Ghuman AS (2016) Decoding and disrupting left midfusiform gyrus activity during word reading. Proc Natl Acad Sci USA 113:8162–8167.

Jacques C, Witthoft N, Weiner KS, Foster BL, Rangarajan V, Hermes D, Miller KJ, Parvizi J, Grill-Spector K (2016) Corresponding ECoG and fMRI category-selective signals in human ventral temporal cortex. Neuropsychologia 83:14–28.

Jonas J, Jacques C, Liu-Shuang J, Brissart H, Colnat-Coulbois S, Maillard L, Rossion B (2016) A face-selective ventral occipito-temporal map of the human brain with intracerebral potentials. Proc Natl Acad Sci USA 113: E4088–E4097.

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.

Kay KN, Yeatman JD (2017) Bottom-up and top-down computations in word- and face-selective cortex. Elife 6:e22341.

Konkle T, Caramazza A (2013) Tripartite organization of the ventral stream by animacy and object size. J Neurosci 33:10235–10242.

Krauss GL, Fisher R, Plate C, Hart J, Uematsu S, Gordon B, Lesser RP (1996) Cognitive effects of resecting basal temporal language areas. Epilepsia 37:476–483.

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.

Kutas M, Hillyard SA (1980) Reading senseless sentences: brain potentials reflect semantic incongruity. Science 207:203–205.

Lachaux JP, George N, Tallon-Baudry C, Martinerie J, Hugueville L, Minotti L, Kahane P, Renault B (2005) The many faces of the gamma band response to complex visual stimuli. Neuroimage 25:491–501.

Lerma-Usabiaga G, Carreiras M, Paz-Alonso PM (2018) Converging evidence for functional and structural segregation within the left ventral occipitotemporal cortex in reading. Proc Natl Acad Sci USA 115:E9981–E9990.

Leszczyński M, Barczak A, Kajikawa Y, Ulbert I, Falchier AY, Tal I, Haegens S, Melloni L, Knight RT, Schroeder CE (2020) Dissociation of broadband high-frequency activity and neuronal firing in the neocortex. Sci Adv 6:977–989.

Li Y, Richardson RM, Ghuman AS (2019) Posterior fusiform and midfusiform contribute to distinct stages of facial expression processing. Cereb Cortex 29:3209–3219.

Liuzzi AG, Bruffaerts R, Vandenberghe R (2019) The medial temporal written word processing system. Cortex 119:287–300.

Lochy A, Jacques C, Maillard L, Colnat-Coulbois S, Rossion B, Jonas J (2018) Selective visual representation of letters and words in the left ventral occipito-temporal cortex with intracerebral recordings. Proc Natl Acad Sci USA 115:E7595–E7604.

Mani J, Diehl B, Piao Z, Schuele SS, Lapresto E, Liu P, Nair DR, Dinner DS, Lüders HO (2008) Evidence for a basal temporal visual language center: cortical stimulation producing pure alexia. Neurology 71:1621–1627.

Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods 164:177–190.

Matsuo T, Kawasaki K, Kawai K, Majima K, Masuda H, Murakami H, Kunii N, Kamitani Y, Kameyama S, Saito N, Hasegawa I (2015) Alternating zones selective to faces and written words in the human ventral occipito-temporal cortex. Cereb Cortex 25:1265–1277.

Miller KJ (2010) Broadband spectral change: evidence for a macroscale correlate of population firing rate? J Neurosci 30:6477–6479.

Miller KJ, Schalk G, Hermes D, Ojemann JG, Rao RP (2016) Spontaneous decoding of the timing and content of human object perception from cortical surface recordings reveals complementary information in the event-related potential and broadband spectral change. PLoS Comput Biol 12:e1004660.

Mo C, He D, Fang F (2018) Attention priority map of face images in human early visual cortex. J Neurosci 38:149–157.

Op de Beeck HP, Pillet I, Ritchie JB (2019) Factors determining where category-selective areas emerge in visual cortex. Trends Cogn Sci 23:784–797.

Otten LJ, Donchin E (2000) Relationship between P300 amplitude and subsequent recall for distinctive events: dependence on type of distinctiveness attribute. Psychophysiology 37:644–661.

Price CJ, Devlin JT (2003) The myth of the visual word form area. Neuroimage 19:473–481.

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

Quian Quiroga R, Mukamel R, Isham EA, Malach R, Fried I (2008) Human single-neuron responses at the threshold of conscious recognition. Proc Natl Acad Sci USA 105:3599–3604.

Ray S, Maunsell JH (2011) Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. PLoS Biol 9:e1000610.

Sabsevitz DS, Middlebrooks EH, Tatum W, Grewal SS, Wharen R, Ritaccio AL (2020) Examining the function of the visual word form area with stereo EEG electrical stimulation: a case report of pure alexia. Cortex 129:112–118.

Saygin ZM, Osher DE, Norton ES, Youssoufian DA, Beach SD, Feather J, Gaab N, Gabrieli JD, Kanwisher N (2016) Connectivity precedes function in the development of the visual word form area. Nat Neurosci 19:1250–1255.

Srihasam K, Vincent JL, Livingstone MS (2014) Novel domain formation reveals proto-architecture in inferotemporal cortex. Nat Neurosci 17:1776–1783.

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.

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.

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 USA 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.

Weiner KS, Grill-Spector K (2010) Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. Neuroimage 52:1559–1573.

White AL, Palmer J, Boynton GM, Yeatman JD (2019) Parallel spatial channels converge at a bottleneck in anterior word-selective cortex. Proc Natl Acad Sci USA 116:10087–10096.

Zhen Z, Yang Z, Huang L, Kong X, Zhen Wang X, Dang X, Huang Y, Song Y, Liu J (2015) Quantifying interindividual variability and asymmetry of face-selective regions: a probabilistic functional atlas. Neuroimage 113:13–25.