Bottom-up and top-down contributions to the organization and dynamics of ventral temporal cortex

by

Matthew J. Boring

B.S., University of Pittsburgh, 2016

Submitted to the Graduate Faculty of the

School of Medicine in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2022

UNIVERSITY OF PITTSBURGH

SCHOOL OF MEDICINE

This dissertation was presented

by

Matthew J. Boring

It was defended on

March 15, 2022

and approved by

Dr. Marlene R. Cohen, Professor, Department of Neuroscience

Dr. Julie A. Fiez, Professor, Department of Psychology

Dr. Nancy Kanwisher, Professor, Department of Brain & Cognitive Sciences, MIT

Dr. Carl R. Olson, Professor, Center for the Neural Basis of Cognition

Dr. Tobias Teichert, Associate Professor, Department of Psychiatry

Dissertation Director: Dr. Avniel Singh Ghuman, Associate Professor, Department of Neurological Surgery

Bottom-up and top-down contributions to the organization and dynamics of ventral temporal cortex

Matthew J. Boring, PhD

University of Pittsburgh, 2022

Ventral visual cortex is a hierarchical recursive network that facilitates object recognition. Many questions regarding the balance between bottom-up and top-down constraints on the functional organization and response dynamics of ventral temporal cortex (VTC) remain unanswered. Here, we map spatial and temporal properties of category-selective neural populations in VTC and investigate long-range functional connectivity patterns that are associated with key neural dynamics in these populations. These maps demonstrate systematic changes in neural dynamics and functional connectivity patterns at successive stages of the ventral visual hierarchy, which are not well characterized by fast feedforward models. Regions within VTC that were highly selective for specific object categories demonstrated a complex organization, with multiple adjoining patches selective for words and faces, each with distinct dynamics.

To understand how bottom-up and top-down interactions influence local neural representations, we specifically examined the responses of one key region involved in reading, word-selective ventral occipitotemporal cortex (wVOT), sometimes referred to as the Visual Word Form Area. We replicate the finding that activity in this region demonstrates a dynamic shift in its representation 250 ms after viewing real words. Early activity from wVOT was sufficient to disambiguate visually dissimilar words, whereas later activity could disambiguate words sharing all but one letter. This transition was strongest for real words compared to pseudowords, consonant strings, and false fonts and was associated with increased functional connectivity with anterior

VTC and early visual cortices. This suggests that bottom-up information, potentially across multiple eye movements, and top-down information, like phonology and semantics, encourage dynamic shifts in wVOT representations.

Finally, we review recent and seminal findings of wVOT's development and response properties. In doing so, we arrive at a model wherein wVOT's localization is constrained by preexisting white matter pathways specialized for evolutionarily older functions, including visual to phonological transformations necessary for lip/speech reading and visual to semantic transformations necessary for object naming. This model help explains why wVOT responds to non-visual linguistic stimuli and why it has a consistent localization across individuals. Together, this work provides a systematic investigation into how bottom-up and top-down influences shape the organization and dynamics of category-selective VTC.

Table of Contents

List of Tables

List of Figures

Preface

The question is not what you look at, but what you see.

--Henry D. Thoreau--

This dissertation represents a journey in my intellectual development, enriched by the company of many individuals, most of whom I have failed to list here. I started this journey with only a basic understanding of the visual system and a few remarkable antidotes on disorders of visual cognition passed on to me during undergraduate cognitive neuroscience courses, like the guy with prosopagnosia who pointed to a mirror and asked the bartender, "Who is that man that keeps shooting me fleeting glances?" How strange it seemed to me, to see the world without "seeing" it. Six years later, and my understanding of the visual system seems even more incomplete. Rather than instilling in me a sense of understanding, this journey opened my eyes to the complexities I had previously failed to see. I look forward to the time that I have left to continue asking questions, and by answering those questions, realizing how many more questions there are to be asked.

I am particularly and eternally grateful for my advisor, Avniel Ghuman, who has been my greatest advocate and mentor in science. Choosing an advisor is the most influential decision you make in graduate school, and I got extremely lucky. The hours we spent deliberating on experimental designs, analyses, and framing our ideas has been instrumental in developing my skills in asking and answering questions precisely and efficiently. The constant encouragement you gave me to push myself to new heights helped me realize my potential. I am a much better scientist because of you.

I would also like to thank my lab mates and collaborators; Arish Alreja, Brett Bankson, Liz Hirshorn, Zach Jessen, Yuanning Li, Shahir Mowlaei, Ed Silson, Maxwell Wang, Mike Ward, Ashley Whiteman, all of the members of my class in the Center for Neuroscience at the University of Pittsburgh, and several student members of the Center for the Neural Basis of Cognition; for facilitating interesting conversations, allowing me to contribute to their projects, and contributing to mine with their feedback and expertise. I am privileged to have been in such a selfless environment and to have met so many bright early career researchers.

I would also like to thank the scientists that encouraged my growth and inspired my work through conversation and feedback during graduate school. They include Chris Baker, Marlene Behrmann, Marlene Cohen, Carol Colby, Julie Fiez, Max G'Sell, Tanya Jonker, Carl Olson, David Plaut, Mark Richardson, Karl Ridgeway, Charlie Schroeder, Michael Shvartsman, and Michael Tarr. Discussions with you helped identify new ways of approaching complex data, encouraged a deeper understanding of the nuances in interpreting that data, and made me a more confident neuroscientist. Your examples of good science will continue to inspire me well into the future.

Finally, I would like to thank my mom and dad for the sacrifices they made to provide me with the opportunity to excel academically. Throughout my life, you have cheered on my successes and encouraged me to push further in person and in spirit, even when the path ahead was unclear. I would not be the scientist I am without the values you instilled in me and the environment you afforded me. To my newest family member, my wife, Huijun, you have been by my side throughout this entire PhD journey. You kept me grounded and helped me see the bright side when things were bleak or overwhelming. Without that stability, it is unclear where I would be, but I would certainly not have the same appreciation for the power of perspective if not for you. I am excited for all the places we have yet to see together.

1.0 Introduction

Ventral temporal cortex (VTC) is a critical hub for object recognition in humans. In macaques, lesions to the functionally homologous inferotemporal cortex results in a marked deficit in the ability of animals to recognize visual objects, while preserving visuospatial cognition [1]. In humans, more localized damage to VTC causes category-specific deficits in object recognition, such as the specific inability to recognize tools, animals, faces or words [2–6]. Early observations of category-selective deficits in object recognition lead to the hypothesis that the visual system was organized into domain specific modules, damage to which leads to the observed category specific deficits [2,3,7]. More recently, advances in neuroimaging have confirmed the existence of areas in VTC that respond selectively to certain categories of objects, including words, faces, places, and tools [8–15]. Electrical brain stimulation of these category-selective regions, including those selective for faces [16–18] and words [19–25], has been shown to specifically disrupt processing of the category that region is selective for, without evoking deficits in the processing of other categories. The degree to which these results support domain specific processing in VTC or whether this selectivity arises from more general principles remains debated [26–32], in addition to many other open questions regarding the bottom-up and top-down factors that influence the organization and dynamics of category-selective activity in VTC.

1.1 Bottom-up contributions to VTC dynamics and organization

Single unit studies in macaques have been instrumental in elucidating the hierarchical organization of ventral visual cortex [33–37]. Neurons along the ventral visual stream demonstrate hierarchical responses, with neurons in more anterior regions pooling over neurons from more posterior regions to form increasingly complex representations over increasingly large portions of the visual field [34,38]. These hierarchical neural responses only take approximately 10 ms to propagate to the next processing layer [39] and these early responses are sufficient to decode what category of object macaques are viewing [40]. This rapid propagation of activity and the hierarchical nature of ventral visual responses has led to the influential perspective that object recognition is achieved through fast, feedforward, hierarchical computations [34,38,40–42]. This perspective is useful for explaining early VTC response dynamics but, as we will see in the next section, fails to capture important top-down influences on extended VTC processing dynamics.

In addition to explaining early neural responses collected throughout inferotemporal cortex, the feedforward view also offers potential mechanisms whereby ecological categories of objects tend to be clustered together in mature visual cortex. Primary visual cortex is organized according to a topographic map of retinotopic space, which is conserved through several layers of the ventral visual hierarchy [34,43–48]. This retinotopic map biases where patches of category-selective neurons are located, depending on where their preferred objects tend to fall on the retina [9,43,49,50]. For example, humans tend to foveate faces and words throughout visual development, which can help explain why face- and word-selective regions tend to be constrained to lateral aspects of VTC, which has foveal receptive fields [9,43,46,47,49,51]. On the other hand, navigationally relevant information, like places, are not typically foveated and are therefore constrained to be represented in more medial aspects of VTC [9,15,49,51,52]. Other low-level

visual features like curvature [53] and shape [54] have also been suggested to influence where category-selective regions are localized in VTC. The weight of each of these bottom-up features in determining category-selective VTC organization remains an open question.

In addition to being constrained by low-level visual features, some argue that category-selective regions in VTC require visual experience to develop [31,32,44,48,55–57]. Macaques without exposure to faces for the first year of life did not develop face-selective patches seen in macaques with normal experiences, but they did develop patches selective for other objects that they were exposed to, like hands [58]. Conversely, macaques that underwent extensive training to discriminate between abstract symbols including Latin letters or Arabic numerals developed patches selective for these stimuli, which were not seen in untrained animals [55,56]. These symbol-selective patches developed in similar locations across animals, which was argued to suggest that this patch of inferotemporal cortex is innately specialized to process foveal information with the low-level shape characteristics of the learned symbols [55,56]. Together, studies like these suggest that extensive visual experience with specific object categories is necessary for the formation of category-selective patches in VTC, the organization of which are constrained by the topographical projections of low-level visual features throughout the ventral visual hierarchy [27,44,55,56]. However, bottom-up models like these fail to capture the potential influence of top-down interactions, including connectivity between higher-order social/affective processing centers and VTC face-selective regions or semantic/phonological centers and VTC word-selective regions, on the organization and dynamics of VTC [29,59–65].

1.2 Top-down contributions to VTC dynamics and organization

Most feedforward models of ventral visual processing consist of homogenous hierarchical layers, which pool inputs solely from the previous layer and output solely to the next layer [37,38,41]. Units in these layers quickly evaluate incoming stimuli for the presence of their preferred feature and pass that output to the next layer, without further contributing to later stages of processing [37,38,41]. This feedforward architecture ignores the fact that the majority of anatomical projections in the ventral visual pathway are reciprocal or feedback [34,35]. Feedforward models also fail to capture shifts in representations that occur within VTC regions after the first 250 ms after seeing a stimulus [19,66–68], the increasing timescales over which regions process information when moving up the ventral visual hierarchy [69–77], and the differences exhibited in long range connectivity patterns across layers of VTC [70,78–80].

In humans and macaques, face processing regions in the ventral visual stream undergo dynamic shifts in the information they represent after the first 150-250 ms after being exposed to a face [66–68]. When macaques were shown images of faces, face-selective neurons demonstrated a coarse representation at 100 ms that could disambiguate faces from shapes and whether the face was human or macaque. 50 ms later, the same neurons coded much more fine-grained information like facial identity and expression [67]. Similarly in humans, early face-selective responses in ventral occipitotemporal cortex (VOT) can disambiguate between faces versus other categories, but become more refined after 250 ms, when they can discriminate between faces with different expressions [66]. A similar phenomenon has also been demonstrated in word-selective VOT, where early activity (within 250 ms after seeing a word) is sufficient to discriminate between visually dissimilar words, but is later refined (after 250 ms) to disambiguate words sharing all but one letter [19]. The interactions that facilitate the sharpening of these representations are not well

understood, but could include the accumulation of bottom-up information (e.g., across eye movements) and top-down information (e.g., from centers that specialize in emotional processing for faces [61,81], or semantic and phonological processing for words [62,82,83]).

In addition to failing to capture the extended processing dynamics of VTC, fast feedforward models also do not capture the heterogeneity in processing dynamics demonstrated across layers of the ventral stream [69–77]. Studies mapping out the dynamics of hierarchical brain networks have consistently identified systematic differences in the timescales over which regions integrate information [69,73,77]. Specifically, resting state neural dynamics slow when moving along the axis extending from primary sensory/motor to association cortices [69], which has been assumed to correlate with how these regions integrate information over time [73]. For example, higher order visual cortices demonstrate slower cortical fluctuations at rest [84] and have been shown to integrate dynamic visual information over relatively long timescales using fMRI [76]. On the other hand, primary visual cortices have much faster resting state dynamics [84] and integrate information over much shorter timescales [76]. Again, these differences in neural timescales are not well captured by fast feedforward models, which posit that visual representations are built through quick and automatic evaluation of stimuli for the presence of their preferred visual features, without any differences in processing dynamics across layers [37,38,41].

It has also been shown that functional connectivity patterns are heterogeneous across layers of the ventral visual hierarchy [70,78–80], which is also not captured by fast feedforward models. In humans, functional connectivity to visual areas systematically decreases along the hierarchical axis of VTC, whereas functional connectivity to heteromodal association cortices increases [78,80]. These gradients in long range functional connectivity are thought to interact with local neural dynamics, but evidence for this association from neuroimaging modalities with high

temporal resolution is lacking [85]. In silica, circuit models of the macaque brain have demonstrated potential mechanisms linking connectivity gradients and resting-state neural dynamics [70]. However, in traditional fast-feedforward models, the connections of each layer are relatively homogeneous, they link the previous layer with the next without recursion from higher layers [37,38,41].

Long-range bidirectional connectivity between VTC and higher-level cognitive centers likely also plays an important role in constraining the organization of category-selective regions [29,59–65]. Although bottom-up models capture the importance of low-level features in constraining where category-selective regions are located [44,48,55–57], they fail to explain how congenitally blind individuals demonstrate category-selective VTC responses to auditory stimuli like laughing, car sounds, and clapping in similar locations as face-, object-, and body-selective regions in sighted individuals [86,87]. Similarly, braille reading in blind individuals evokes responses from regions of VOT that selectively respond to printed words in sighted literate individuals [88,89]. On the other hand, a model whereby cortical organization is jointly constrained by bottom-up factors as well as connectivity to higher-order processing circuits is better-able to explain these results through top-down interactive processes [61–63,90–93].

Specifically, long-range bidirectional anatomical pathways underlying transformations between visual to semantic, phonological, navigational, social, and other representations may help constrain where category-selective patches end up in VTC, in addition to the low-level visual properties of stimuli [61–63]. For example, preexisting white matter pathways connecting VTC to frontal, lateral temporal, and parietal language regions may bias the localization of word-selective VOT [29,60,64,65]. This is supported by diffusion weighted imaging in preliterate children, which can predict where word-selective VOT will develop after children learned to read [65]. Anatomical

connectivity to language centers in lateral prefrontal cortex, lateral temporal cortex, and parietal cortex was greater for the region of VOT that would become word-selective after children learned to read compared to adjacent face-selective cortex [65]. In summary, a balance between bottom-up properties of visual stimuli and top-down influences from other cognitive systems likely constrains VTC organization and dynamics.

1.3 Overview of the contributions and structure of this dissertation

In this dissertation we will investigate how long-range cortical interactions influence the dynamics and organization of VTC at several levels, with a particular emphasis on word-selective circuits. In Chapter 2, we map local prestimulus dynamics, information processing dynamics, and long-range functional connectivity patterns that systematically change across layers of the ventral visual hierarchy. This allows us to investigate the relationships between prestimulus and information processing dynamics, and whether either of these are associated with differences in patterns of functional connectivity. Additionally, we tested whether any of these functional gradients were associated with a region's ability to predict patient response time to determine if these factors help constrain a region's role in cognitive behavior.

In Chapter 3, we focus our attention on the organization and dynamics of VTC networks responsible for face and word processing. Face and word processing offer an especially interesting comparison for investigating how ventral visual cortex is organized. Faces and words are very different along several low-level visual features like contrast, spatial frequency, curvilinearity, as well as several high-level properties like when expertise develops for these stimuli during childhood, their evolutionary age, and their roles in social and linguistic cognition. Despite these

differences, they share a remarkably close cortical localization in VTC that is highly consistent across individuals [94]. Their localization is so close that the degree to which they are represented by the same cortical areas is debated [60]. Here, we investigate the degree to which face and word processing networks in VTC are separable within individuals, how these networks are organized relative to one another, and how their functional dynamics compare.

In Chapter 4, we zoom in on the information processing dynamics of an important word-selective region in VOT, sometimes referred to as the Visual Word Form Area (VWFA). Previous studies from our group examining intracranial recordings from epilepsy patients identified a dynamic shift in the representations in this region from an early coarse-level, which was sufficient to disambiguate visually dissimilar words from one another, to a later (after 250 ms) fine-grained representation, which was sufficient to disambiguate words sharing all but one letter [19]. Here, we investigate whether this dynamic shift in representation is also exhibited in healthy individuals, whether it generalizes to other word-like stimuli that vary in their semantic and/or phonological associations (including false fonts, pseudowords, and consonant strings), and examine the functional interactions that occurred while this region sharpens its representation.

Finally in Chapter 5, we review recent and seminal findings regarding the anatomy of word-selective VOT and its role in reading. In reviewing these data, we present a model of how reading networks develop by adapting visual to phonological pathways involved in lip/speech reading and visual to semantic pathways involved in object naming to accomplish similar transformations for printed words. The long-range white matter projections facilitating these transformations, including the arcuate, inferior longitudinal, and frontooccipital fasciculi, in turn constrain where printed words are represented in VTC, in addition to the low-level visual properties of words. This model helps explain why, in addition to demonstrating selectivity for

printed words, word-selective VOT also responds to non-visual stimuli, like braille in the congenitally blind [88,89], through top-down interactions with other language regions. This model also explains how visual word recognition circuits develop in similar localizations across individuals despite literacy being in its evolutionary infancy, due to the consistency of these long-range white matter projections in neonates [64]. Ultimately, this model demonstrates joint contributions of bottom-up and top-down influences on the organization and dynamics of category-selective regions in VTC.

2.0 Interacting cortical gradients of neural timescales and functional connectivity and their relationship to perceptual behavior

We begin our investigation into the spatial and temporal organization of object processing in ventral visual cortex by mapping several aspects of local neural dynamics and long-range functional connectivity along the hierarchical axis of VTC. Several studies have identified increasing timescales of neural dynamics, decreasing connectivity to unimodal sensory regions, and increasing connectivity to heteromodal regions across sensory and motor hierarchies [69,74,80]. However, it is unclear how these properties interact with one another or how they influence perceptual behavior. In this chapter, we demonstrate several functional gradients in prestimulus dynamics, information processing, and functional connectivity extending along the hierarchical axis of VTC. Gradients in local prestimulus dynamics were associated with changes in functional connectivity beyond shared correlations with anatomical position. Prestimulus dynamics and connectivity to visually responsive regions were also associated with how well a region's activity predicts patient response time during a 1-back task. That suggests that these properties arise from shared neurophysiological mechanisms, which may constrain a region's role in perceptual behavior. This functional map of category-selective VTC provides characteristics for future hierarchical models of the ventral stream to consider, including increasing recursive connectivity and extended processing dynamics in higher layers. At the time of writing this dissertation, this chapter has not been peer-reviewed but was coauthored by Mark R. Richardson and Avniel Singh Ghuman.

2.1 Introduction

A neural population's functional properties, including its dynamics and its functional connectivity to other brain regions, are ultimately linked to that population's role in cognition and perception. Several gradients in functional properties have been shown to exist along the cortical axis spanning from primary sensory/motor areas to association cortices [79,80,95,96]. For example, gradients in the timescales over which neural populations endogenously fluctuate and process information are demonstrated along this axis, with longer timescales further along cortical hierarchies [69,73–75,97,98]. Gradients of functional connectivity are also seen along this axis, with decreasing unimodal connectivity and increasing transmodal connectivity along cortical hierarchies [78,80]. These network-level neural properties likely influence local timescales, other computational characteristics of neural populations, and these populations' relationship to behavior [79,80,85,96]. However, empirical evidence linking functional gradients in local dynamics with gradients in the long-range connectivity of neural populations is limited. Additionally, it is unknown to what degree these gradients relate to a neural population's role in behavior.

One prevalent functional gradient in cortex is the increasing timescales over which neural populations integrate information when moving from primary sensory/motor to association cortices [69,72,74,76,97,99]. For example, rapidly varying acoustic inputs represented in low-level auditory cortex are combined into more complex representations in higher order auditory cortex, which operates over longer timescales [98]. These neural timescales, or temporal receptive windows, are related to the rate of decay of representations, or autocorrelation, within neural populations [69,74,76,97], because longer decay rates allow for more pieces of information to be integrated into a single representation.

Another key aspect of neural dynamics, which is less well understood, is information processing dynamics, including the initial rate at which neural populations discriminate between stimuli (i.e., the rise time of discriminant information in neural activity). These information processing dynamics relate to the speed of cortical computation and thus, ultimately limit the speed of decision and action processes [40]. Despite the importance of a neural population's information processing dynamics in cognition and perception, the functional characteristics that are associated with neural populations that processes information more quickly or slowly remain unclear [69,97,99].

In addition to anatomical gradients in neural dynamics, opposing anatomical gradients in connectivity to association versus primary sensory/motor cortices have also been demonstrated in human cortex. Unimodal connectivity, primarily within sensorimotor regions, decreases when moving up cortical processing hierarchies while transmodal connectivity linking multiple sensory domains increases [80,95]. However, it is unclear how gradients in local dynamics interact with gradients in long-range functional connectivity. In silica, circuit models of cortical processing suggest that inter- and intra-areal connectivity patterns help constrain a neural population's timescale [70,79], which has received some support from low temporal resolution measures of brain activity [85].

Finally, the functional properties that constrain a neural population's dynamics and long-range cortical connectivity ultimately constrain how that population contributes to cognition and perception. However, it is unknown whether the degree to which a neural population's activity predicts behavior displays anatomical gradients and/or correlates with that population's neurodynamics and functional connectivity.

In the current study, category-selective neural populations in ventral temporal cortex (VTC) were used as a model to examine the relationship between anatomical gradients in local cortical processing and long-range cortical interactions. We also explored how information processing dynamics, endogenous timescales (i.e., neural dynamics not directly linked to the exogenous, stimulus-evoked response; which we estimate using the prestimulus period when no stimulus was being presented), and long-range cortical connectivity interact with each other beyond any shared anatomical gradients, and which of these gradients were associated with the ability of a population's activity to predict response time during single trials of a visual 1-back task.

2.2 Methods

2.2.1 Intracranial electroencephalography (iEEG) patients

Stereotactic depth and surface electrocorticography (ECoG) electrodes were implanted in ventral temporal cortex (VTC) of 41 patients (15 males, ages 19-65) for the localization of pharmacologically intractable epileptiform activity. Different aspects of these recordings from 38 of these patients were previously reported in [100]. All patients gave written informed consent under protocols approved by the University of Pittsburgh's Institutional Review Board. Electrode contacts that were identified as belonging to the seizure onset zone were not included in the analysis.

Electrodes were localized via postoperative CT scans or postoperative magnetic resonance images (MRI). Postoperative CT scans were co-registered to preoperative MRIs using Brainstorm

[101]. Surface electrode contacts were projected to the nearest reconstructed cortical voxel of the preoperative MRI scan to correct for brain-shift [102,103]. These electrode locations were then registered to the Montreal Neurological Institute (MNI) common space via patient-specific linear interpolations [104]. VTC was defined as grey matter below the inferior temporal gyrus spanning from the posterior edge of the fusiform gyrus to the anterior temporal lobe in MNI common space.

Cortical distance between each electrode contact and the patient's occipital pole was computed using the patient's native neural anatomy. The occipital pole was defined as the intersection of the calcarine sulcus, inferior occipital gyrus, and superior occipital gyrus. The geodesic (cortical) distance between this point and the cortical surface coordinate nearest to each VTC electrode contact was computed using custom MATLAB scripts [105].

2.2.2 Experimental paradigm

All patients underwent a category localizer task containing images occupying approximately 6° x 6° of visual angle at the center of a stimulus display monitor positioned 2 meters from the patient's eyes. Each stimulus was presented for 900 ms on a black background. Inter-stimulus intervals were 1500 ms with a random 0-400 ms jitter during which the patient saw a white fixation cross. Patients were instructed to press a button every time an image was presented twice in a row (1/6 of all trials). Repeat trials were excluded from further analysis. This left 70 trials per category to train and test the classifiers described in 2.2.4. Several patients underwent more than one run of this experiment and therefore had 140 or 210 trials per category. All experimental paradigms were presented via custom MATLAB scripts running the Psychophysics toolbox [106].

35 patients underwent a category localizer task consisting of pictures of bodies, faces, hammers, houses, words, and non-objects. Six patients underwent category localizer tasks with slightly different object categories but with identical stimulus parameters. One of these patients viewed pictures of bodies, faces, *shoes*, hammers, houses, and phase-scrambled objects. One viewed pictures of bodies, faces, *consonant-strings*, *pseudowords*, real words, houses, and phase-scrambled objects. One patient viewed pictures of faces, bodies, *consonant-strings*, words, hammers, and phase-scrambled objects. One viewed pictures of faces, bodies, words, *pseudowords*, houses, and phase-scrambled objects. One viewed pictures of faces, bodies, words, *tools*, *animals*, houses, and phase-scrambled objects. One viewed pictures of faces, bodies, words, *tools*, *animals*, *numbers*, houses, and phase-scrambled objects.

2.2.3 Intracranial recordings

Local field potentials were collected from iEEG electrodes via a GrapeVine Neural Interface (Ripple, LLC) sampling at 1 kHz. Notch filters at 60/120/180 Hz were applied online. Stimulus presentation was synchronized to the neural recordings via parallel port triggers sent from the stimulus displaying computer to the neural data acquisition computer. The signal was off-line filtered from 0.2-115 Hz using two-pass fourth order butter-worth filters via the FieldTrip toolbox [107]. In addition to analyzing these single trial potentials (stP), we also extracted and analyzed the single trial high frequency broadband (stHFBB) activity of these electrodes, since these two components of the local field potential have been shown to contain complimentary information [108].

StHFBB activity was extracted via Morlet wavelet decompositions from 70-150 Hz over 200 ms Hanning windows with 10 ms spacing. The resulting power spectral densities were then

averaged over these frequency components and normalized to a baseline period from 500 ms to 50 ms prior to stimulus onset to yield the stHFBB activity. Data was then epoched from -500 to 1500 ms around stimulus presentation. Trials during which the stP amplitude changed more than 25 microvolts across a 1 ms sample, or during which stPs exceeded an absolute value greater than 350 microvolts, or during which either the stHFBB or stPs deviated more than 3 standard deviations from the mean were all assumed to contain noise and were therefore excluded.

2.2.4 Multivariate temporal pattern analysis

Sliding, leave-one-out cross-validated, Gaussian Naïve Bayes classifiers were applied to 100 ms time windows with 10 ms stride to determine if stHFBB or stP recorded from individual VTC contacts contained category-discriminant information. The input to these classifiers was 100 ms (100 samples) of stP and 100 ms (10 samples) of stHFBB from a single electrode contact. The output of the classifier was the category of object presented during the corresponding trial. This procedure was repeated for all VTC contacts from time windows beginning at 100 ms prior to stimulus onset to 1000 ms after stimulus onset.

The category-discriminant information content within each neural population was estimated by computing the mutual information (*I(S',S)*) between the output of the Gaussian Naïve Bayes classifiers (predicted category labels, *S'*) for a given 100 ms time window of neural activity and the actual presented stimulus (S):

$$I(S', S) = P(S)\, log_2(\frac{P(S', S)}{P(S')P(S)}),$$

where *P(S',S)* is the joint probability of the classifier correctly predicting the stimulus category *S* when the category was *S*, *P(S')* is the proportion of times the classifier guessed a trial

was of stimulus *S*, and *P(S)* is the proportion of trials which the stimulus presented was *S*. This allowed us to estimate the category-discriminant information contained within 100 ms time windows without estimating a joint probability table of neural responses that was intractable [109,110]. It has been shown that this estimate of information, which relies on a *P(S',S)* derived by an external classifier and not the actual neural code, is an underestimate of the neural information content [111]. Therefore, our calculated information is a lower bound for the actual neural information content.

Information content was averaged across all stimulus categories presented to the patient so as not to preclude electrode contacts as being selective for only one object category [28]. A threshold for significant category-discriminant information was determined by randomly shuffling stimulus labels for a subset of VTC electrode contacts and repeating the same classification analysis 1,000 times for each electrode contact [112]. Electrode contacts with the same number of runs of the category-localizer task demonstrated very similar null distributions and therefore we applied the result of this permutation test to all VTC electrode contacts. The threshold was chosen such that none of the random permutations for any electrode contacts in the subset reached the threshold, which corresponds to $p < .001$, corrected for multiple temporal comparisons. Electrode contacts with peak category-discriminant neural information exceeding this threshold were defined as category-discriminant.

We performed a similar decoding analysis to determine the time-course of visual responses in individual VTC contacts. This was done by classifying single trial baseline periods (100 ms to 0 ms prestimulus presentation) of neural activity from these neural populations against sliding 100 ms time-windows from -90 to 1000 ms post-stimulus presentation for all object categories treated as one class. This yielded a time-course of visual responses in each sampled neural population. By

randomly permuting the label of the baseline versus evoked data and repeating the analysis in a subset of electrode contacts 1,000 times, we defined a threshold of visual information that no random permutation of the data achieved, corresponding to the $p < .001$, chance level, corrected for multiple temporal comparisons. We used this threshold to define visually responsive brain regions and those that were not, which were separated to calculate their differential contributions of functional connectivity to VTC electrode contacts with significant category-discriminant information.

2.2.5 Estimating the dynamic properties of neural information processing

To estimate properties of the information processing dynamics of neural populations across VTC, the information time-courses derived from the Naïve-Bayes classifiers were first smoothed with a running average filter (width 50 ms). Next, onset latency of category-discriminant information was defined as the last time point that an electrode contact was below 10 % of the maximum information prior to the peak information. The initial rise in category-discriminant information was defined as the time between the onset and the point where the information time-course first exceeded 90 % of the peak information. These cutoffs were chosen to ensure that small deviations from chance-level information and peak information did not affect the estimated quantities. Our main findings were robust to specific choices in threshold (Figure 29). Finally, we estimated the duration of information maintenance as the time between when the neural population first reached 90% of its peak information to when the neural population's information first fell below 50% of this maximum after peaking. Similar dynamic properties were also estimated for visual response time-courses (Figure 3) and information processing time-courses for specific object categories (Figure 26 & Table 4).

2.2.6 Defining category-selective VTC electrode contacts

To determine if neural populations with sensitivity to different object categories demonstrated differences in the gradients of their local dynamics or long-range functional connectivity, we isolated category-discriminant VTC neural populations that responded primarily to one object category. To do this we computed the event related potential and event related broadband responses to each category during the 1-back task. Next, any of the previously defined category-discriminant neural populations that contained maximum information to the same category that evoked the maximum response across either of these averages was classified as selective to that object category. We then characterized the information onset latency, slope, and connectivity of these neural populations using the procedures described above. For these analyses we used the category-specific information processing time-course derived from the Naïve Bayes classifiers prior to averaging over all categories in the main analysis.

2.2.7 Information processing simulations

Simulations were used to test if increases in information processing duration exhibited along the ventral visual hierarchy could be explained by differences in peak information magnitude. Specifically, information time-courses were approximated as normal probability density functions (PDFs) parameterized by a mean, standard deviation, and magnitude (constant scaling). Normally distributed noise with the same standard deviation as prestimulus (-400 to 0 ms) information in category-selective VTC electrode contacts was then added to these curves. 1000 simulated signals were computed for each different PDF magnitude and standard deviation.

Information processing duration was calculated using the same procedure described for the actual signal, by calculating the time between when the signal first reached 90 % of its maximum amplitude and the last time it was below 10 % of its maximum before that. We then calculated the Spearman correlation between information processing duration when varying the PDF's standard deviation (to mimic changes in slope of the information processing time-course) and when varying the information's peak amplitude. Peak amplitude was varied from the minimum to maximum peak information in category-selective VTC electrode contacts. During the simulation investigating the effect of slope on information processing duration, signal amplitude was fixed at the average peak information in category-selective VTC electrode contacts.

2.2.8 Characterizing endogenous neural timescales

The endogenous timescales of VTC populations were characterized by computing the autocorrelation of prestimulus (-500 ms to stimulus onset) stPs and stHFBB activity from 1-250 ms lags during each clean trial of the 1-back task. These prestimulus autocorrelation functions were then averaged over all trials. The average autocorrelation function for each electrode contact was then fit with a single exponential decay function:

$$ACF(t) = t_0 + N_0 e^{-t/\tau}$$

The neural timescale (tau), which measures the rate at which the autocorrelation function decays, was then correlated with several other functional properties of the neural population. This estimation of neural autocorrelations and computation of tau is similar to the procedure described in [74].

2.2.9 Functional connectivity

To determine the connectedness of VTC neural populations to the rest of the brain, phase-locking values (PLVs) were calculated between neural populations with above chance levels of category-discriminant information and all other electrodes within the same patient (regardless of category-discriminant information content). Electrode contacts within 1 cm of the category-discriminant electrode were not included in the analysis to rule out effects caused purely by volume conduction. PLVs measure instantaneous phase-coupling across different brain regions independent of differences in amplitude, unlike coherence metrics [113]. This makes PLVs more sensitive to detecting weakly coupled oscillators despite differences in amplitude [114]. This coupling of oscillations is thought to indicate event-related communication between electrode contacts.

The instantaneous phase of each electrode contact during all category-localizer trials was computed via convolution of the filtered neural activity (from 1-115 Hz) with Morlet wavelets of frequencies ranging from 1-60 Hz (width = 5). This convolution allowed the separation of signal phase from envelope at each frequency [115]. Next, the PLV was computed by taking the vector average of the phase difference between two electrode contacts at each time point. PLVs close to 1 indicate two electrode contacts have similar phase differences at this frequency and time point across all trials. Conversely, if this number is close to 0, the phase difference between these electrode contacts is random at this given frequency and time point.

A spectral window of interest was defined to capture the part of the PLV spectrogram that showed increased functional connectivity across all category-discriminant VTC neural populations. We chose to focus on the time windows from -450 to 0 ms before stimulus onset to capture prestimulus functional connectivity of the neural populations. Next, we determined which

frequency components demonstrated increased stimulus-evoked functional connectivity across VTC. To do this we averaged the PLVs from 50 to 500 ms and performed a paired t-test against the average PLV from -450 ms to 0 ms before stimulus presentation between the category-selective VTC electrode contacts and the rest of the electrode contacts in the same patient. This analysis revealed that frequency components between 1 and 22 Hz all had significantly greater phase-locking across all category-discriminant VTC electrode contacts relative to baseline on average from 50 to 500 ms after stimulus presentation ($p < .001$, corrected).

Therefore, we averaged the PLVs across electrode contacts from 2 to 22 Hz (discarding 1 Hz frequency band to increase the temporal precision of our estimated phase-locking), and -450 to 0 ms before stimulus onset to calculate the functional connectedness of these same regions. We separately averaged the connectivity of category-discriminant VTC neural populations with visually responsive regions (defined above) and those that were not to determine if there were connectivity differences across these neural populations. Average functional connectivity from -450 to 0 ms prestimulus and 50-500 ms after stimulus presentation were strongly correlated with one another ($\rho = .96$, $p < .001$). Thus, results do not substantially change if either the prestimulus or post stimulus PLV is used.

2.2.10 Predicting patient response time from category-selective VTC

To test for differences in the correlation between category-selective VTC population activity and behavior, patient RT was predicted using the neural activity from each category-selective contact. Specifically, a sliding window L2-regularized multiple regression (100 ms window, 10 ms stride) was used to predict patient RT from stP and stHFBB activity using a leave-one-trial-out cross-validation procedure. Only trials when the patient correctly reported that an

object was repeated twice in a row were included in the analysis. The maximum Spearman correlation between the patient's RTs and the sliding-window RT predictions from 0-1000 ms after stimulus presentation was considered as the neural population's correlation with behavior. This correlation was then correlated with that population's dynamics, connectivity, and anatomical location.

2.2.11 Statistics

Spearman rank-order correlations were used to calculate the correlations between anatomical position and aspects of the neural information time-courses calculated above. Spearman rank-order partial correlations were used to calculate the correlation between variables while correcting for correlations shared with other variables. Benjamini-Hochberg False Discovery Rate estimation which is valid for dependent hypothesis tests was used where noted [116]. Paired T-tests were used to determine if there were differences in the dynamics of processing different levels of information (visual vs. category-discriminant) in the same electrode contacts.

Rank-order mixed-effects models were used to control for random effects of cross-patient variability while examining the main effects of connectivity and anatomical position on information processing dynamics. We chose to fit these mixed-effects models with equal slopes but random intercepts across patients to ensure the models converged. Because observations in mixed-effects models are not independent, it is difficult to determine the appropriate degrees of freedom. This makes estimation of p-values impossible without appropriate approximation. Therefore, to derive p-values for the main effects of the mixed-effects models, we use the Satterthwaite approximation, which has been shown to produce acceptable Type 1 error rates with relatively few samples [117].

Linear multiple regression models were used to compare gradients of information processing in VTC neural populations that were selective for different object categories. We only included the categories that most patients saw (bodies, words, faces, hammers, houses, and phase-scrambled objects). Specifically, linear models were used to predict information onset latency, peak, processing duration, maintenance duration, and connectedness as a function of the category-selective neural populations' distance from the occipital pole with an added factor indicating which category the neural population was selective for (Figure 26). Linear mixed-effects models were initially used for this analysis to simultaneously control for random effects across patients. However, these models failed to converge, likely indicating an insufficient number of data points per category and patient to estimate these random effects. Because face-selective electrode contacts were most prevalent in our population we used this as our baseline and compared all other categories to face-selective electrode contacts (Table 4). Analysis of covariance was also used to determine if there was a significant difference in information processing gradients or connectedness across hemispheres (Appendix A.1.1).

2.3 Results

Activity was recorded from 1,955 VTC electrode contacts (out of a total of 4,090 intracranial electrode contacts distributed throughout the brain) in 35 patients with pharmacologically intractable epilepsy (Figure 23) as they viewed images of objects (face, body, word, hammer, house, or phase scrambled image) during a 1-back task. Multivariate Naïve Bayes classifiers were used to predict the category of object participants were viewing during individual trials of the task using sliding 100 ms windows of single trial potentials (stP) and single trial high

frequency broadband activity (stHFBB) recorded from individual electrode contacts. At this stage of the analysis, these signal components were combined since previous studies have suggested that they contain complementary information [108], though in further analyses they were examined separately. Out of the 1,955 VTC electrode contacts, activity recorded from 390 electrode contacts (mean = 11; SD = 14 electrode contacts per patient) could reliably predict ($p < .001$, corrected via permutation testing) which category participants were viewing during single trials of the task (Figure 1). The time-course of category-discriminant information processing in these category-discriminant neural populations was calculated by computing the mutual information (in bits) between the classifier outputs and the true category labels. The functional properties of these populations were computed to examine the relationship between these variables and anatomical axes of VTC (see Methods). Specifically, we examined gradients of, and interactions between, nine factors: two stimulus response timescales (factors 1 and 2): initial rise duration and maintanence of category-discriminant information (see Figure 2A for illustration); category-discriminant information onset time and peak magnitude (factors 3 and 4; see Figure 2A for illustration); two endogenous (prestimulus) timescales (factors 5 and 6): the timescale of decay, "tau", for the prestimulus stP and stHFBB autocorrelation functions (see Figure 4A for illustration); functional connectivity to visually responsive populations and to populations that were not significantly visually responsive (factors 7 and 8); and the accuracy of a neural population's activity for predicting behavioral response time (RT; factor 9).

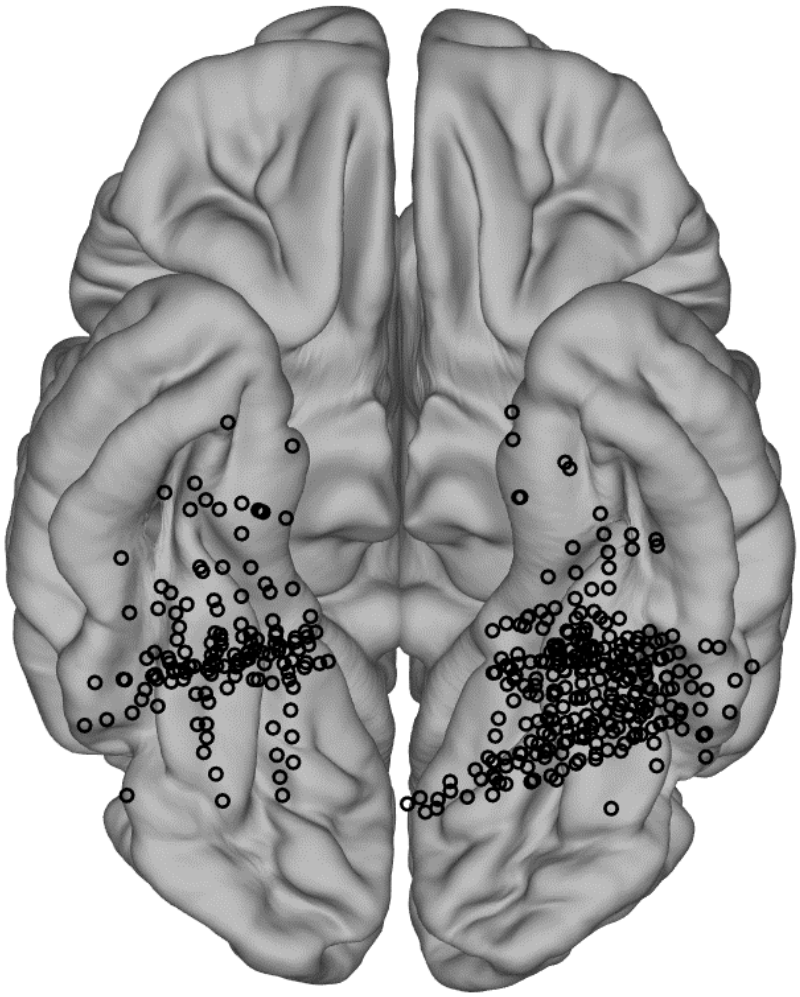


Figure 1. Category discriminant electrode contacts. Spatial topography of electrode contacts recording from neural populations that achieved peak category-discriminant information greater than chance at the $p < .001$ level corrected for multiple temporal comparisons. The proportion of left vs. right hemisphere category-discriminant contacts was comparable to the proportion of total left vs. right hemisphere VTC implants (see Appendix A.1).

The cortical distance from the occipital pole, which roughly corresponds to the fovea in primary visual cortex, was used to approximate the position of neural populations along the hierarchical axis of the ventral visual stream [34]. Distance along this axis was correlated with several aspects of information processing in these category-discriminant neural populations

(Figure 2*;* see Figure 24 for an example from a single subject). Along this axis, neural populations demonstrated increasing onset latencies and increasing durations of their initial rises in category-discriminant information. Additionally, neural populations maintained category-discriminant information longer after peaking, despite reaching smaller peak magnitudes, when moving along the visual hierarchy. See Figure 28 for simulations demonstrating the independence of peak magnitude and rise duration metrics.

In addition to examining discriminant information, we also examined the dynamics of the non-discriminant neural responses. Specifically, gradients in category-indiscriminant visual responses (discriminating all categories from baseline rather than categories from one another as in Figure 2) in the same neural populations were examined (Figure 3). Populations demonstrated increasing onset latencies and decreasing peak magnitudes of visual responsiveness when moving along the ventral visual hierarchy, like the gradients observed for category-discriminant information. However, there was no comparable increase in the duration of the initial rise in visual responsiveness along this axis and visual responsiveness was maintained for shorter durations in populations further along the visual hierarchy, which was opposite of the gradient observed for category-discriminant information. The contrast between visual response dynamics and category-discriminant information processing dynamics highlight differences in the neural encoding of these two levels of stimulus information [39,67].

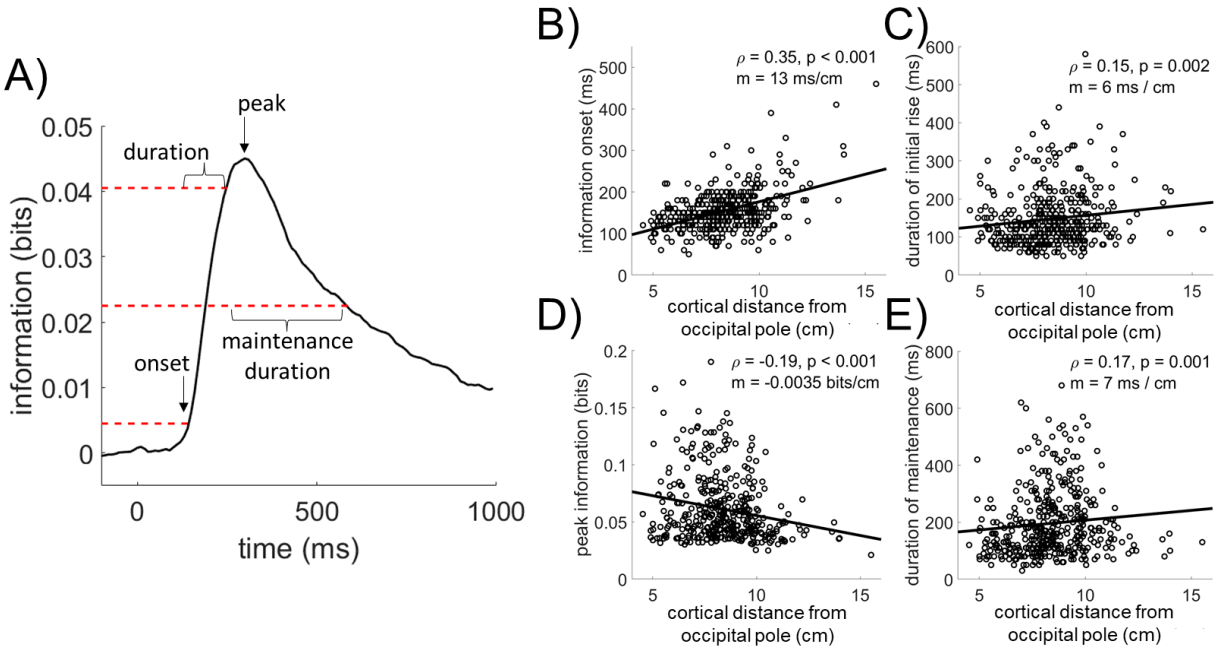


Figure 2. Category-discriminant information processing gradients along the ventral visual hierarchy. A) The time-course of category-discriminant information processing was computed for each neural population. The average time-course across category-discriminant VTC populations is illustrated here. From each neural population's information processing time-course, the information onset time (panel B), processing duration (C), peak magnitude (D), and maintenance duration (E) were computed. Simulations confirmed that decreases in information amplitude and information processing duration are independent using our methods (Figure 28). B) The onset of category-discriminant information, defined as the timepoint the information reached 10% of the maximum before peaking, was significantly correlated with the position of that neural population along the ventral visual hierarchy. The black line indicates the least-squares regression fit. Spearman's ρ and associated p-value shown on top right (n = 390). Spearman correlation was used because it is both more robust to outliers relative to Pearson's and is sensitive to non-linear monotonic relationships between variables, though this also means that the line drawn is not representative of the ρ. Slope of the least-squares regression line (m) indicated a 13 ms per centimeter increase in onset latency moving along VTC. Information onset was significantly associated with distance along the visual hierarchy even after correcting for cross-patient differences in onset latency (T(388) = 7.20, p < .001, tied-rank mixed-effects model). C) The duration of the initial rise in category-discriminant information, defined as the time between the onset of information and the time it took the population to reach 90% of its peak information, was negatively

correlated with distance along the visual hierarchy. The 90% threshold is used for the peak time because it better captures the initial rise in cases where there is a shallow peak among an extended plateau in the discriminant information time-course. *Note*: All correlations remain significant across a substantial range of the heuristic thresholds chosen to define them (Figure 29), thus the selection of 10% and 90% as thresholds for onset and peak time do not drive these effects. The slope of the least-squares regression line indicated a 6 ms increase in the duration of the initial rise of information per cm of VTC. This relationship did not reach $p < .05$ when correcting for random cross-patient effects ($T(388) = -1.55$, $p = .12$, tied-rank mixed-effects model). D) Peak category-discriminant information was negatively correlated with distance along the visual hierarchy, with a decrease of -0.0035 bits/cm. This relationship did not reach $p < .05$ when correcting for random cross-patient effects ($T(388) = -1.62$, $p = .11$, tied-rank mixed-effects model). E) Information maintenance duration, defined as the time between when the information first reached 90% and the time when it first decayed to 50% of the peak, was positively correlated with distance along the visual hierarchy. The slope of the least-squares regression line indicated a 7 ms increase in the duration of maintenance of information per cm of VTC. This relationship trended to $p < .05$ significance when correcting for random cross-patient effects ($T(388) = 1.87$, $p = .063$, tied-rank mixed-effects model).

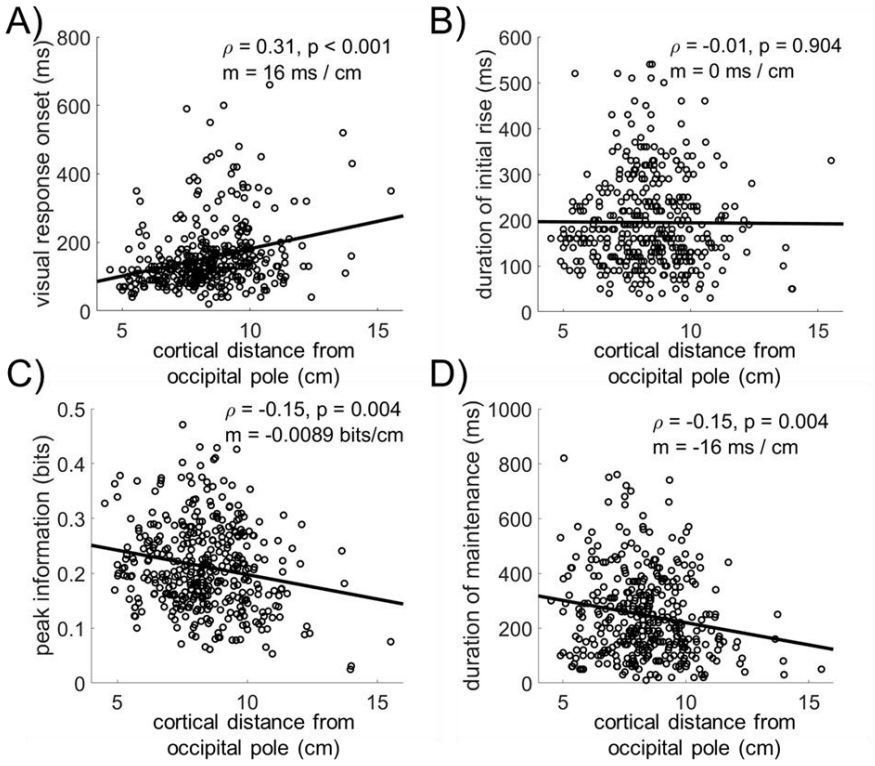


Figure 3. Visual response dynamics along the ventral visual hierarchy. Visual response dynamics were extracted by classifying all stimulus categories versus baseline with similar classifiers used to extract category-discriminant information (Figure 2). A) Onset of the visual response was positively correlated with a neural population's distance along the visual hierarchy. This effect held when correcting for random cross-patient effects (T(388) = 6.23, p < .001, tied-rank mixed-effects model). Onset latency of the visual response was not significantly different than the onset of category-discriminant information (T(389) = 0.11, p = .91, paired T-test). B) Duration of the initial increase in visual responsiveness was not significantly correlated with distance along the visual hierarchy, unlike the significant positive correlation observed for category-discriminant information (Figure 2C). C) Peak magnitude of visual responsiveness was negatively correlated with distance along the visual hierarchy. This effect held when correcting for random cross-patient effects (T(388) = -2.26, p < .001, tied-rank mixed-effects model). D) Visual response maintenance duration was also negatively correlated with distance along the visual hierarchy, which held when correcting for random cross-patient effects (T(388) = 5.45, p < .001, tied-rank mixed-effects model). This was opposite of the relationship between information maintenance duration and distance along the visual hierarchy observed for category-discriminant information (Figure 2E).

Next, the endogenous neural timescales of VTC populations were quantified by computing the autocorrelation of prestimulus activity at multiple temporal lags and modelling the resulting autocorrelation function with an exponential decay function (Figure 4). When moving along the visual hierarchy, neural populations demonstrated increasing time-constants of decay (tau) in the autocorrelation function of their prestimulus stP, indicating that their activity exhibited longer timescales/slower dynamics along this axis. This is consistent with previous studies observing slower timescales when moving up sensory processing hierarchies [74,76,77,80,96,99]. Conversely, neural populations demonstrated shorter timescales in their prestimulus stHFBB activity when moving along the ventral visual hierarchy. Time-constants across stP and stHFBB signal components were not significantly correlated with one another across electrode contacts ($\rho(390) = -0.05$, $p = .33$), highlighting the differentiability of these two aspects of the neural signal [100,119,120]. These results show that these components of the endogenous neural activity demonstrate distinct timescales that have opposite gradients along the ventral visual hierarchy.

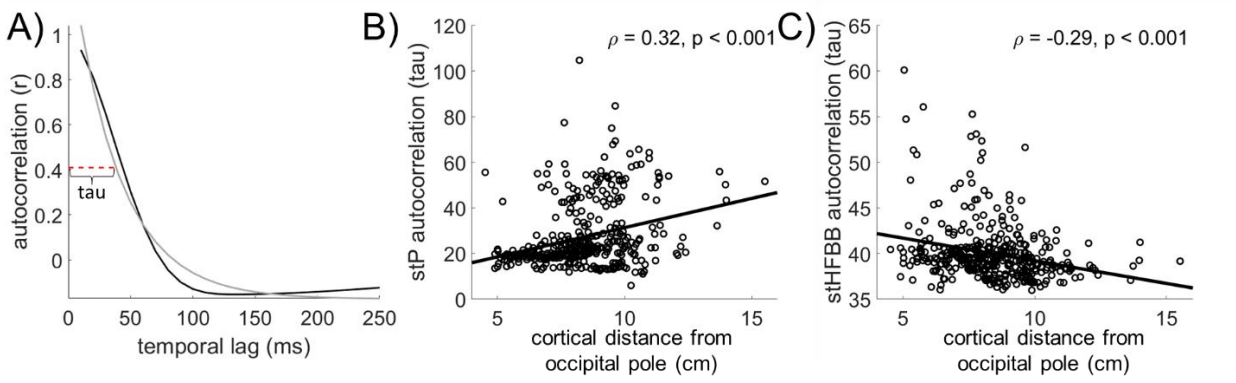


Figure 4. Prestimulus neural timescales along the ventral visual hierarchy. A) For each neural population, the autocorrelation function during the -500 to 0 ms prestimulus period was computed for temporal lags ranging from 1 to 250 ms, averaged across trials (black line is the average across populations), and fit with a single exponential decay function (gray line). The timescale (tau) indicates how fast the fitted exponential function decays (red dashed line; computed like those in [74]**) and was correlated with other functional properties of the category-discriminant neural populations' activity. B) The autocorrelation function of single trial potentials (stP) decayed more slowly when moving up the visual hierarchy, indicating that stP in more anterior VTC had higher autocorrelations at greater lags (longer timescales) relative to more posterior neural populations. This relationship held when correcting for random cross-patient effects (T(388) = 8.03, p < .001, tied-rank mixed-effects model). C) The autocorrelation function of single trial high frequency broadband (stHFBB) decayed more quickly when moving up the visual hierarchy, indicating that stHFBB in more anterior VTC had lower autocorrelations at greater lags (shorter timescales) relative to more posterior neural populations. This relationship also held when correcting for random cross-patient effects (T(388) = -5.32, p < .001, tied-rank mixed-effects model).**

Gradients in both information processing dynamics and neural timescales were present in individual patients (Figure 24) and several generalized across patients (linear mixed-effects models Figure 2 & Figure 3 captions). The gradient of information processing onset was stronger in the left compared to the right hemisphere (Appendix A.1.1). Notably, neural populations selective for individual categories demonstrated different gradients in neural dynamics relative to those

selective for other categories along the ventral visual hierarchy, with face-selective populations generally displaying shallower posterior-anterior gradients Appendix A.1.2, Figure 25, Table 4). Given the differences in prestimulus neural timescales exhibited in stP and stHFBB, we recomputed gradients in information processing dynamics from these signal components separately. With a few notable exceptions, stimulus related information processing dynamics demonstrated similar gradients for stP and stHFBB across these components when moving along the visual hierarchy (Appendix A.1.3 and Figure 26).

After examining gradients in information processing and endogenous timescales, we examined gradients in functional connectivity along the ventral visual hierarchy. Specifically, a measure of functional connectedness to the rest of the brain, the average prestimulus phase-locking value (PLV), was calculated between the 390 category-discriminant VTC electrode contacts and all other electrode contacts implanted within the same patient (on average 115 electrode contacts, SD = 41; note that none of the results reported below change substantially whether functional connectivity was calculated during the prestimulus or the poststimulus period as the Spearman correlation between the prestimulus and poststimulus connectivity metrics was > 0.95). These "other" electrode contacts were located across the entire brain, not only in VTC (Figure 23). Previous fMRI studies suggest opposite gradients in functional connectivity to unimodal sensory vs. association and transmodal areas when moving along sensory processing streams [80,95]. Therefore, we separately computed the functional connectivity of VTC category-selective contacts to visually responsive contacts ($p<0.001$, for visual response vs. baseline, corrected for multiple temporal comparisons) and to those that were not visually responsive. Also, given the wide variability of electrode coverage across patients, pooling connectivity across visually responsive

and not visually responsive contacts allowed us to partially overcome this cross-patient anatomical heterogeneity.

Connectivity between VTC electrode contacts and visually responsive contacts decreased when moving up the visual hierarchy. In contrast, the connectivity between VTC contacts and contacts that were not significantly visually responsive increased when moving up the visual hierarchy (Figure 5). Decreasing functional connectivity to visually responsive regions and increasing functional connectivity to regions that do not demonstrate strong visual responses is generally consistent with previous fMRI studies showing opposing anatomical gradients along VTC for functional connectivity to unimodal versus transmodal regions [80].

Additionally, gradients in local neural dynamics and long-range cortical interactions were examined to determine how they correlate to a neural population's role in visual perceptual behavior. This was done by predicting the RT of patients, using sliding windows of neural activity recorded at each category-selective VTC electrode contact, during trials of the 1-back task where patients correctly responded that an object was presented twice in a row. How predictive the activity in an electrode contact was of RT was used as a measure of how much the activity from that neural population contributed to perceptual behavior. When considering stP and stHFBB together, the ability of a VTC neural population's activity to predict RT was not significantly correlated with distance along the visual hierarchy ($\rho(390) = .02$, $p = .75$). However, when considering them separately, a neural population's ability to predict RT decreased along the visual hierarchy when looking at stHFBB but increased when looking at stP. These differences highlight nuances in large-scale neuroanatomical gradients when considering different aspects of the neural signal [100,119,120].

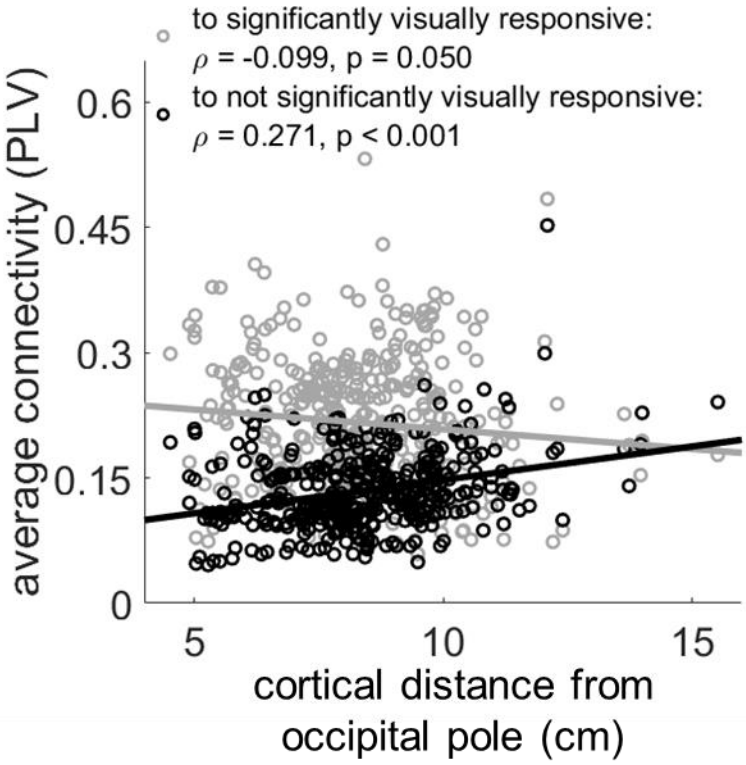


Figure 5. Gradients in long-range functional connectivity along the ventral visual hierarchy. The change in connectivity to visually responsive regions moving along VTC was opposite of the change in connectivity to populations that were not visually responsive. Connectivity to significantly visually responsive regions decreased along this axis, even when accounting for random cross-patient effects (T(388) = -4.42, p < .001, tied-rank mixed-effects model). On the other hand, connectivity to regions that were not significantly visually responsive increased along this axis, even when accounting for random cross-patient effects (T(388) = 3.98, p < .001, tied-rank mixed-effects model).

Given corresponding anatomical gradients in local dynamics and long-range cortical interactions, a key question is, to what degree these gradients are interrelated beyond shared anatomical axes. To explore this question, the partial correlations between these functional properties of category-selective VTC populations were calculated after removing the effects of distance along the visual hierarchy (Figure 6). Note that Spearman's partial correlation was used

to remove any monotonic relationship to distance along the visual hierarchy, not only linear relationships (see Figure 27 for the non-partialed correlations).

The negative partial correlation between a neural population's stP timescale and its functional connectivity to visually responsive populations throughout the brain was significant. This suggest that parts of VTC that communicate strongly with other visually responsive regions have shorter timescales. Furthermore, the negative partial correlations were significant between the magnitude of a neural population's peak category-discriminant information and both its connectivity to visually responsive regions and those that were not. This shows that neural populations with stronger connectivity, especially to non-visual areas have less category-discriminant activity.

None of the measures of endogenous or stimulus-response timescale (prestimulus stP and stHFBB tau, initial rise duration, and maintenance) were significantly correlated with one another, with or without removing the effects of distance along the visual hierarchy (Figure 6 and Figure 27). Thus, though there are gradients in neural timescales across VTC using each of these measures, neither these timescales nor their gradients are significantly correlated to one another even though they were measured from the same neural populations. This indicates that neural timescales are context dependent (prestimulus vs stimulus response, initial rise duration vs. maintenance, stP vs. stHFBB, etc. are all not significantly correlated) and measuring one type of timescale cannot be used to infer the general timescale of a neural population.

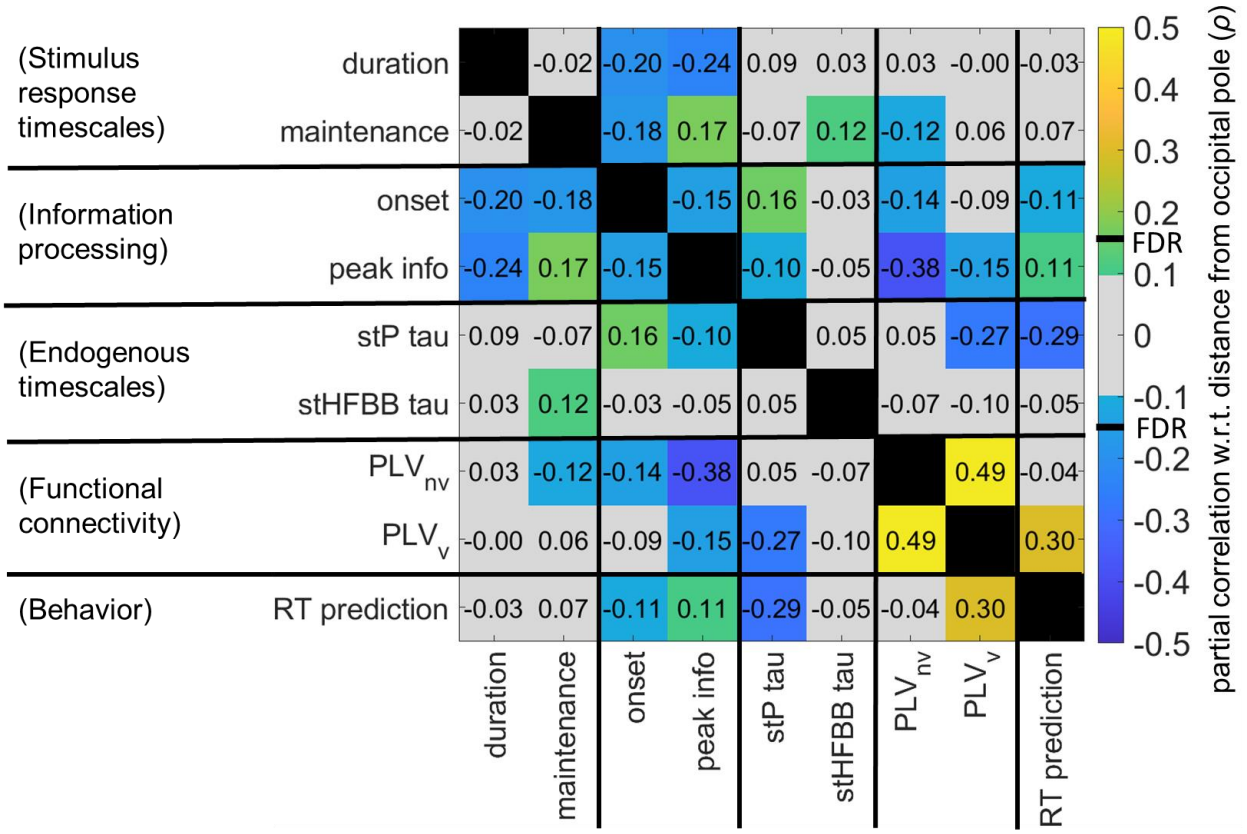


Figure 6. Interactions between local dynamics, long-range cortical interactions, and behavioral correlations. Partial correlation matrix between local dynamic properties and long-range cortical interactions after removing the effect of cortical distance along the visual hierarchy (see Figure 27 for non-partialed correlations). Colored squares are significant at the $p < .05$ level (uncorrected). The false-discovery rate adjusted critical value corresponds to $\rho = \pm .146$. Within each square is the partial Spearman correlation coefficient for the variables in the corresponding row and column. The matrix is symmetric across the diagonal. Several properties of the local information processing dynamics, including information onset, peak magnitude, duration of the initial rise, and the amount of time the information was maintained, were related to one another besides sharing a common anatomical gradient. The partial correlation between peak information and functional connectivity was also significant after removing the effect of distance along the visual hierarchy. The partial correlation between neural timescale (stP tau) and connectivity to visually responsive regions (PLV_v) was also significant as was the partial correlation between both connectivity to visual regions and stP timescale and a neural populations ability to predict patient response time (RT) during the 1-back task.

Partial correlations between nearly all the stimulus response variables (peak information, onset time, initial rise duration, and maintenance duration), other than the two timescales discussed in the previous paragraph (initial rise duration vs. maintenance duration), were significantly correlated with one another. This suggests that there are interactive factors driving these different aspects of the stimulus response.

The partial correlation between the ability of a neural population to predict RT and that neural population's connectivity to visually responsive brain regions and the partial correlation between a neural population's ability to predict RT and that neural population's prestimulus stP timescale after removing the effect of distance along the visual hierarchy were both significant (Figure 6). Thus, neural populations which integrate information over visual brain regions with short stP timescales were more predictive of behavior during the 1-back task observed here. This demonstrates that aspects of both local neural dynamics and long-range cortical interactions are intimately linked to a neural population's role in visual perceptual behavior.

2.4 Discussion

Taken together, these results illustrate interrelationships between a neural population's anatomical location, its local dynamics, and its long-range functional connectivity, which ultimately influence that population's role in perception. In the current study, progressing along the ventral visual hierarchy was associated with decreases in peak category-discriminant information, longer information onsets, longer durations of initial information processing, longer periods of information maintenance, longer prestimulus stP timescales but shorter prestimulus stHFBB timescales, and opposing changes in connectivity to visual and non-visual brain regions.

These results suggest that the anatomical and physiological gradients that exist along the visual hierarchy influence almost all aspects of prestimulus and information processing dynamics, which may constrain how these neural populations process information and their computational role in cognition. Indeed, a subset of these functional gradients were correlated with the ability of a neural population's activity to predict the speed of behavioral responses during a visual 1-back task. Furthermore, many aspects of stimulus response dynamics shared significant interrelationships with one another beyond any shared relationship with anatomical location. Functional connectivity was correlated to aspects of both the stimulus response and prestimulus timescales, demonstrating how long-distance interactions can influence local neurodynamics. However, prestimulus and poststimulus information processing timescales were not strongly correlated to one another, nor were the initial rise and maintenance of the visual response, suggesting that different aspects of neural dynamics arise through different processes and mechanisms.

Previous studies have observed that neural populations demonstrate longer timescales when moving from primary sensory and motor regions to association cortices [69,74,76,97,99]. The increasing endogenous timescales of stP activity along the ventral visual hierarchy observed here further support this organizing principle of cortex. Notably though, the endogenous stHFBB timescales demonstrated the opposite relationship along VTC, with shorter timescales in more anterior parts of VTC. Furthermore, the timescale of the stP and stHFBB were uncorrelated, demonstrating a dissociation between the dynamics of these two signal components recorded from the same neural population. This highlights a need to better understand the differences in the physiological origins of stP and stHFBB signal components [100,119,120].

The duration that category-discriminant neural populations initially process category-selective information increased along the ventral visual hierarchy, which may be the result of

increased computational demands involved in forming more complex and individuated representations in more anterior category-selective neural populations [121–124]. However, in traditional models of perception, neural units are passive visual feature detectors, that either fire or not depending on the presence or absence of their preferred features [125]. In these models, little difference should be seen in the speed that neural populations process information further downstream because these passive feature detectors, even if they are sensitive to complex features, should respond rapidly and automatically to the presence of that feature [125]. In this study, the duration of the initial rise in visual responsiveness did not change along the hierarchy, which fits with these traditional models. However, the divergence in the duration of category-discriminant versus visual response dynamics does not fit with these models. Instead, these results support a model of ventral visual representations that evolve through time, with information processing dynamics governed by interactions between the information being processed locally and globally through long-range connections, which reflect top-down and recurrent interactions [19,66–68].

Long-range functional connectivity demonstrated a crossover effect along the ventral visual hierarchy, with decreasing connectivity to visually responsive regions and increasing connectivity to those that were not, consistent with previous fMRI studies [80,95]. Some of these gradients in functional connectivity were also associated with gradients in neural timescales even after controlling for effects of distance along the visual hierarchy. Specifically, neural populations that were more strongly connected to visually responsive regions demonstrated shorter endogenous stP timescales. One potential explanation for this result is that neural populations which integrate primarily visual inputs have faster timescales compared to neural populations that have more diverse inputs so that they are prepared to rapidly process incoming visual information [34,69,80]. Notably, the partial correlation between connectivity to regions that did not demonstrate strong

visual responses and poststimulus stP timescale was not significant. Previous models have not investigated differential effects of long-range cortical interactions with visual versus non-visual regions on the timescale of neural populations [70]. This may be an important consideration for future models. Given the variable coverage of brain regions across patients in the current study, future studies are necessary to tease apart the impact that connectivity with specific brain regions has on local cortical dynamics.

Neural populations that demonstrated higher peak category-discriminant neural activity had earlier onsets, shorter durations of initial rise, and maintained that information longer. Our simulations demonstrated that our measures of peak and duration are independent, confirming that this correlation is physiological and not an artifact of the analysis (Figure 28). Longer initial rises in category-discriminant information with smaller peak information may reflect evidence accumulation over longer timescales in these neural populations [69]. Whereas partial correlations between local neural dynamics and long-range cortical interactions demonstrates that, in addition to sharing strong gradients along the primary axis of sensory processing systems, these properties of neural populations are closely linked to each other. These links between local dynamics and long-range cortical interactions are likely conferred in part by shared biochemical, microstructural, and macrostructural connectivity gradients that exist along the ventral visual axis beginning early in cortical development [79,80,95,126,127].

Functional gradients in VTC were also correlated with the degree to which a neural population's activity could predict perceptual behavior. In the current study, increased functional connectivity to visually responsive regions and shorter prestimulus stP timescales were associated with a greater ability for a neural population's activity to predict RT after removing the effect of distance along VTC. This suggests that these neural populations may play a larger role in the basic

visual discrimination task studied here. Behaviors involving more complex perceptual representations and/or more complex behavioral decisions may rely more heavily on neural populations with longer timescales and on higher order cortical regions [19,66,128–131]. Future studies are required to determine if finer level of visual discrimination involving longer response times [132] reflect contributions from neural populations with different information processing timescales and functional connectivity patterns compared to those involved in the 1-back task studied here.

There were not significant correlations between stimulus response timescales and endogenous timescales, or between onset dynamics and maintenance dynamics. Different aspects of task-evoked timescales were not closely linked to one another, suggesting the physiological drivers of initial information processing and maintenance may be independent. Additionally, endogenous neurodynamic timescales did not generalize to stimulus related information processing timescales. Notably, this is unlike functional connectivity patterns, which were highly correlated across the stimulus response and prestimulus periods. The lack of significant correlation highlights that endogenous neural timescales are not necessarily tightly related to task-evoked information processing dynamics [71,75,133,134]. Thus, inferences about a region's computational role in cognition, including its temporal integration and segregation [73] or temporal response windows [76,77], cannot be inferred from endogenous dynamics alone, as stimulus response and endogenous timescales are not necessarily strongly correlated. There is no single principle or process that governs a neural population's timescales, e.g. timescales are not a static and inherent property of a neural population [73]. Rather, these results suggest that different kinds of timescales are governed by different combinations of factors that can depend on cognitive and neural context.

The current study highlights how large-scale anatomical and functional gradients interact to constrain local neural processing dynamics and computation. The anatomical gradients of dynamics and connectivity demonstrated here impose important constraints for future neurobiological models of visual perception. This architecture may help the brain achieve abstract and conceptual representations seen in more anterior VTC neural populations [121,123,124]. While the present study examined these effects in visual processing, it is likely that similar principles apply to other hierarchically organized sensory and cognitive systems [80,97,98]. Indeed, gradients in physiological, and thus functional, organization are likely in part conferred by corresponding gradients in growth factors and, in turn, gene expression during and persisting after cortical development [80,96,126,127]. Interactions among response properties and functional connectivity patterns of neural populations suggest that shared neurophysiological mechanisms tie large-scale and local processing dynamics together. Distinctions among and between endogenous and stimulus response timescales suggest that these neurodynamics are caused by distinct neurobiological mechanisms and play different roles in the brain. These results highlight the mutual interrelationships between a neural population's position in the processing hierarchy, its functional connectivity, and its local dynamics, constraining its role in cognition.

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

Next, we sought to map out the VTC circuits responsible for face and word processing. Faces and words vary significantly along several axes, including their low-level visual properties, when expertise for these stimuli develops during childhood, and their evolutionary age [94]. However, their cortical localizations in VTC are remarkably close to one another, leading some to believe they are processed by the same visual circuits [60]. In this investigation, we isolated several neural populations highly selective for either words or faces that were not selective to the other category, demonstrating that the processing circuits for these objects are, at least, partially separable. Further, the maps drawn here illustrate a mosaic of word-selective regions across VTC which demonstrate distinct temporal dynamics, suggesting that they play different roles in word-processing. This highlights the importance of an extended basal temporal language system in reading. This chapter was published as MJ Boring, EH Silson, MJ Ward, RM Richardson, JA Fiez, CI Baker, AS Ghuman. Multiple adjoining word- and face-selective regions in ventral temporal cortex exhibit distinct dynamics. *J. Neurosci.* **41**, 6314–6327 (2021).

3.1 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. Functional magnetic resonance imaging (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) [9,49,135]. Specifically, lateral regions of VTC are selective for objects that tend to be viewed centrally (foveated), like words and faces, whereas more medial regions are selective for objects that tend to fall on the periphery of the retina, like navigationally relevant information such as buildings [49,136–138]. This broad principle of organization by eccentricity fails to inform us about how representations of different stimuli that are foveated, like 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 also hypothesized to influence where they are processed in VTC [94]. Word- and face-processing operate on very different low-level visual properties [139], follow different developmental trajectories [65], and feed into distinct networks that support either language or social interactions [140,141], 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 [28].

Neuroimaging studies have separately mapped word- and face-processing networks in VTC. Printed word recognition is thought to be carried out in part by a network of regions along the left occipitotemporal sulcus, that differ in the complexity of their responses and are thought to be hierarchically organized [122,136,142–144]. Face-processing is thought to be carried out in part by a network of regions distributed bilaterally along the midfusiform sulcus [145,146]. However, few studies have investigated VTC's responses to word and face stimuli within the same participants [8,147–150]. Those that have, have relied on low sample sizes or imaging modalities with differential sensitivity to different aspects of neural activity (like high and low-frequency neural activity [151,152]). Therefore, much remains unknown about how visual word- and face-

processing networks organize relative to one another, and to what degree they overlap [8,148–150,153].

Further, word- and face-selective regions have primarily been mapped using methods lacking high spatiotemporal resolution. Therefore, it is unclear if 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 [122,129,154]. Further, category-selective maps derived from BOLD responses may be incomplete due to BOLD's increased sensitivity to early stimulus evoked activity (100-300 ms after stimulus presentations) relative to later responses [128,155] and greater correlation with high frequency broadband activity in invasive neural recordings compared to lower-frequency electrical potentials [152,155].

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

3.2 Methods

3.2.1 Intracranial EEG data collection and preprocessing

3.2.1.1 Participants

38 patients (14 males, ages 19-65, 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 Corporation and were 0.86 and 0.8 mm in diameter, respectively. Grid electrodes were produced by PMT Corporation 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 ventral temporal cortex, 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 region of interest, 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 post-operative magnetic resonance imaging (MRI) or computed tomography scans co-registered to the pre-operative MRI using Brainstorm [101]. Surface electrodes were projected to the nearest point on the pre-operative cortical surface automatically parcellated via Freesurfer [102] to correct for brain shift [103]. Electrode coordinates were then co-registered via surface-based transformations to the fsaverage template using Freesurfer cortical reconstructions.

3.2.1.2 Experimental design

All participants underwent a category localizer task where they viewed grayscale images presented on a computer screen positioned two meters from their face. Images occupied approximately 6 x 6 degrees of visual angle and were presented for 900 ms with 1500 ms inter-

stimulus 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.

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

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 (Figure 10B)	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 (Figure 10B)	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 (Figure 10B)	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

Table 1 continued

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 (Figure 10A)	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

Number of word-, face-, and house-selective electrodes in the left (L) and right I 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 3.2.1.4 and Figure 7B). 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 word or face-selective regions to determine their relative anatomical location are listed as not available (N/A). 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.

Local field potentials were recorded via a GrapeVine Neural Interface (Ripple, LLC) sampling at 1 kHz. Notch filters at 60/120/180 Hz were applied online. Data was subsequently filtered from 0.1-115 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 broad-band activity (stHFBB). These stHFBB responses were then Z-scored based on the baseline period from 500-0 ms proceeding stimulus onsets. It has been previously shown that these two aspects of the local-field potential, stP and stHFBB, contain complementary information [108], though also potentially arise from different neurophysiological generators [120,152,155,156]. Therefore, to assess the overall selectivity across VTC we use both as features in the classifiers described in 3.2.1.4. We also investigated the independent contributions of these signal components to our category-selectivity maps (Figure 12). Trials where the stHFBB or stP exceeded 5 standard deviations from the mean were thought to contain noise and therefore excluded from further analysis.

3.2.1.3 Determining language laterality

Records from preclinical magnetoencephalography (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 was not available to conduct more precise analyses of language laterality for these participants.

3.2.1.4 Multivariate temporal pattern analysis

To determine which electrodes contained information about word and face categories, leave-one trial out cross-validated Gaussian Naïve 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 prior to 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 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 $p < .05$ 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 $p < .05$, 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 if 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 three-fold cross-validated Gaussian Naïve 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. Three-fold 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 $p < .05$ chance-level of this average pairwise decoding by repeating this procedure 1,000 times on data with shuffled trial labels in a subset of the word-selective electrodes [112]. These global null distributions were similar across the randomly subsampled electrodes and therefore we chose a d' threshold corresponding to the highest $p < .05$ 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 if electrodes that could not individuate real words could perform these finer discriminations compared to 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 $p < .05$ corrected level by repeating the permutation analysis described for the word individuation task on a random subset of face-selective electrodes.

3.2.1.5 Spatiotemporal k-means clustering

We used a spatiotemporal variant of k-means clustering to determine if 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 post 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 to 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 to 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.

3.2.1.6 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 $p < .001$ threshold, which was non-parametrically defined using the d' sensitivities of all object-selective electrodes from 500-0 ms prior to 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 if there were differences in peak individuation d' or latency across these clusters while correcting for cross-subject differences. We only compared spatiotemporal clusters with greater than 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.

3.2.2 fMRI data collection and preprocessing

3.2.2.1 Participants

Eight participants (six 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.

3.2.2.2 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, 1.2x1.2x1.2 mm; 10% interslice gap; TR = 2 s, TE = 27 ms; matrix size = 170x170).

3.2.2.3 Experimental paradigm

Participants fixated centrally whilst images of words, faces and houses were presented in blocks (16 seconds 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.

3.2.2.4 fMRI data preprocessing

All data were analyzed using the Analysis of Functional NeuroImages (AFNI) software package [157]. Prior to statistical analysis, all images were motion corrected to the first volume of the first run. Post motion-correction data were detrended.

3.2.2.5 Statistical analysis

To identify word-, face- and house-selective regions, we performed a general linear model (GLM) analysis using the AFNI 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 modelled by convolving a standard gamma function with a 16 second square wave for each condition (words, faces & 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.

3.2.2.6 Split-half analysis

For each participant, the ten 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. 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 was 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.

3.3 Results

From 1,396 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 either 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; $p < .05$ level, Bonferroni corrected for multiple spatial and temporal comparisons within participant; see Methods) and (2) responded maximally to faces compared to 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 to the five other object categories. An identical procedure was used to define word- and house-selective electrodes.

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) (Figure 7). Figure 8 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, Figure 7A). 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 (Figure 9). This suggests that the neural circuits responsible for processing words and faces are, at least, partially dissociable [4,28,158].

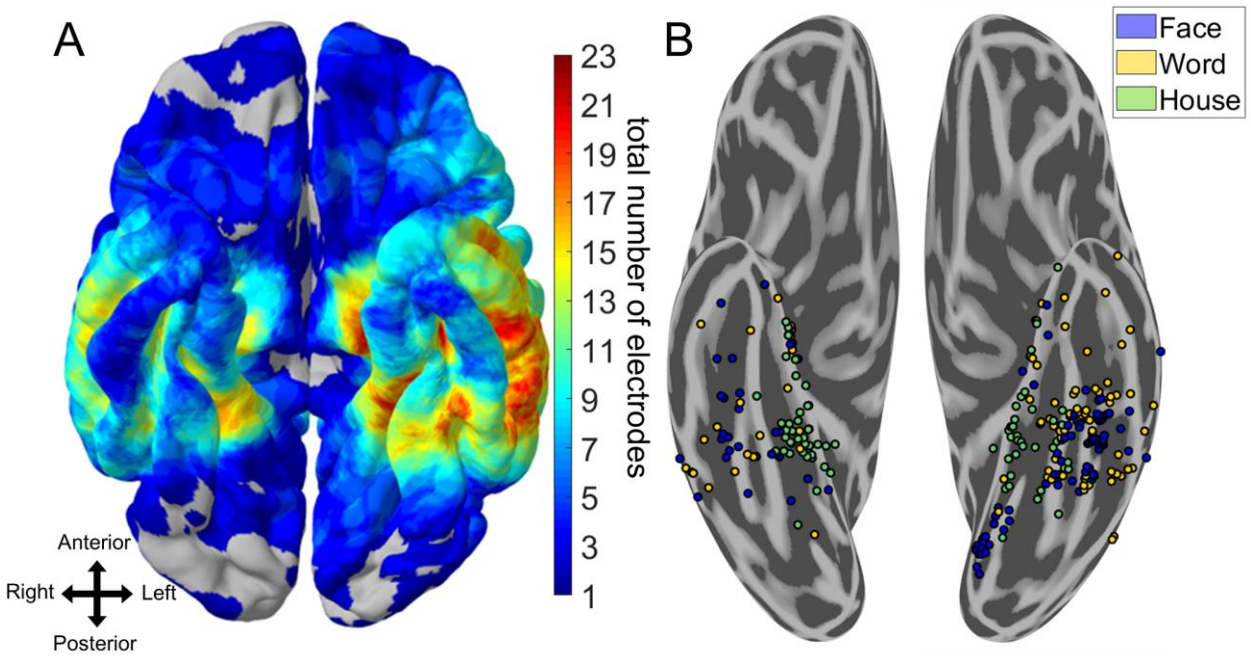


Figure 7. Spatial topography of word- and face-selective iEEG electrodes. A) Heat map of electrode coverage (both category-selective and non-selective) across 36 iEEG participants. Electrodes below the inferior temporal sulcus and anterior to the posterior edge of the fusiform gyrus on the participant's 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 Naïve Bayes classification ($p < .05$, 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 medio-lateral axis of ventral temporal cortex, bilaterally. Depth stereotactic EEG electrodes have been brought to the ventral surface for clarity.

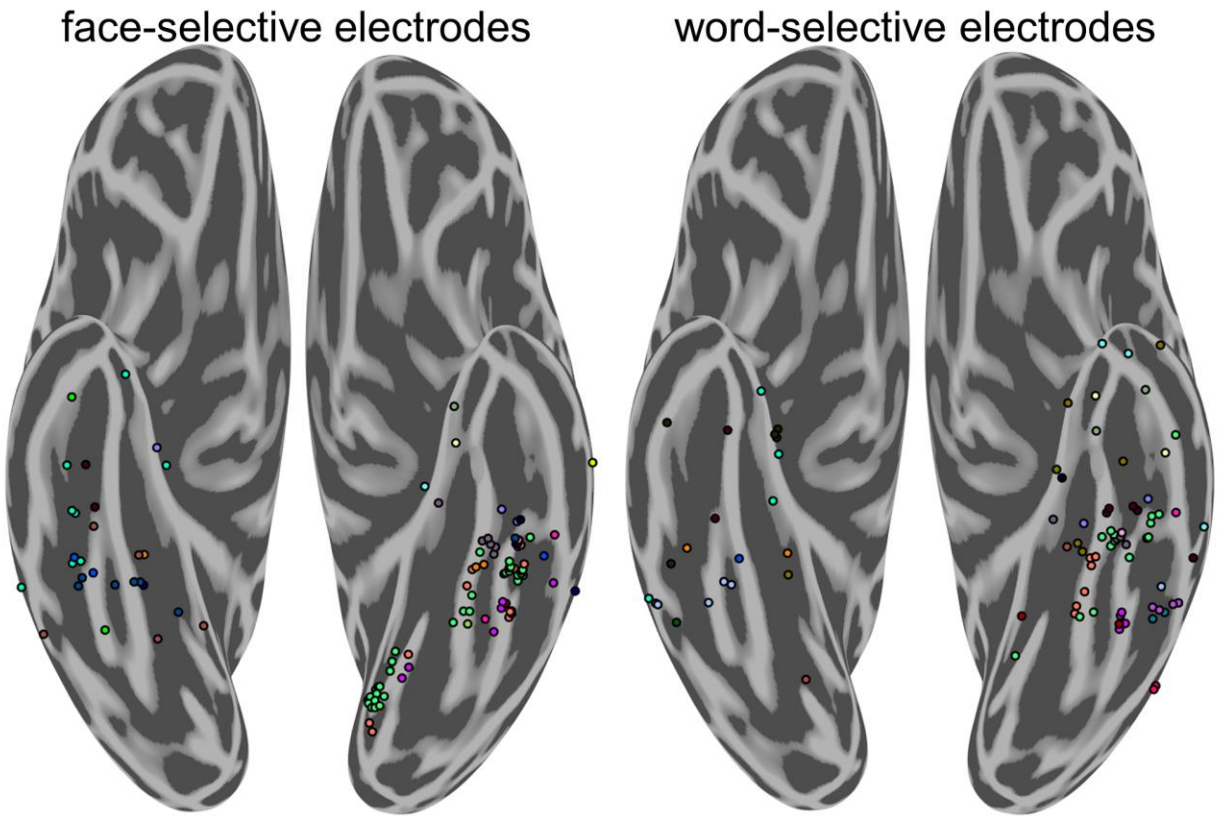


Figure 8. Distribution of face-selective and word-selective electrodes by participant. Distribution of highly face-selective (left) and word-selective (right) electrodes by participant. Color represents individual participants and corresponds across figure panels. Each group-level cluster of word- and face-selective electrodes is represented by data from multiple participants.

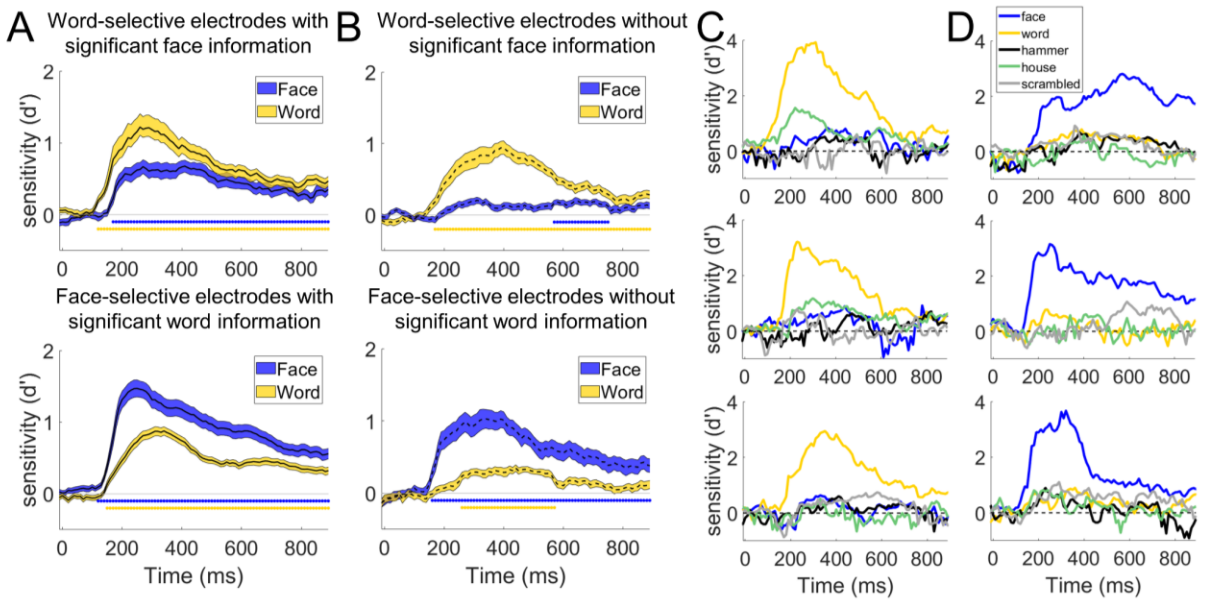


Figure 9. Independence of word- and face-processing networks. A) Average decoding time-course for word- (top, n = 39) and face- (bottom, n = 75) selective electrodes that contained significant amounts of information about the other object category. 21 out of 28 (75 %) face-selective electrodes in the right hemisphere and 54 out 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 5-way classification of the non-face object categories) at the p < .05 level (Bonferroni corrected for multiple comparisons in time and electrodes within participant, see Methods). 9 out of 23 (39 %) word-selective electrodes in the right hemisphere and 30 out of 64 (47 %) in the left hemisphere could discriminate faces from the other object categories excluding words. Error bars indicate standard error from the mean across electrodes. Colored bars under the curves indicate times where the average selectivity is above chance (p < .001 corrected for temporal comparisons). B) Average decoding time-course for word- (top, n = 48) and face- (bottom, n = 33) selective 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 non-preferred category across the population of electrodes, decoding accuracy was much smaller for the non-preferred compared to 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 electrode recording was obtained was not presented with pictures of hammers.

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 [136–138,143,159]. Face selectivity was found primarily along the left and right fusiform gyrus with some face-selective regions within the lingual gyrus, and occipitotemporal sulcus (Figure 7B). Consistent with prior findings [136], 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 due to susceptibility artifacts [160]. In contrast to most maps of word- and face-selective regions obtained from fMRI [8,65,147–149,161,162], 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 (Figure 8), 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 out 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 out of 5 such participants (Table 1, see Figure 10 for 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 10 for an example). 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 was detected in most 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 [8,148,153], 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 11). 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 ($p < .001$, split-halves analysis). These medial word-selective regions were approximately $1/3^{rd}$ 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; $p < .01$). Also, 7 out of 8 of the healthy participants demonstrated word-selective patches near the anterior tip of the fusiform, despite susceptibility artifacts [160], consistent with the iEEG data (Figure 7B). 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 [8,148,153].

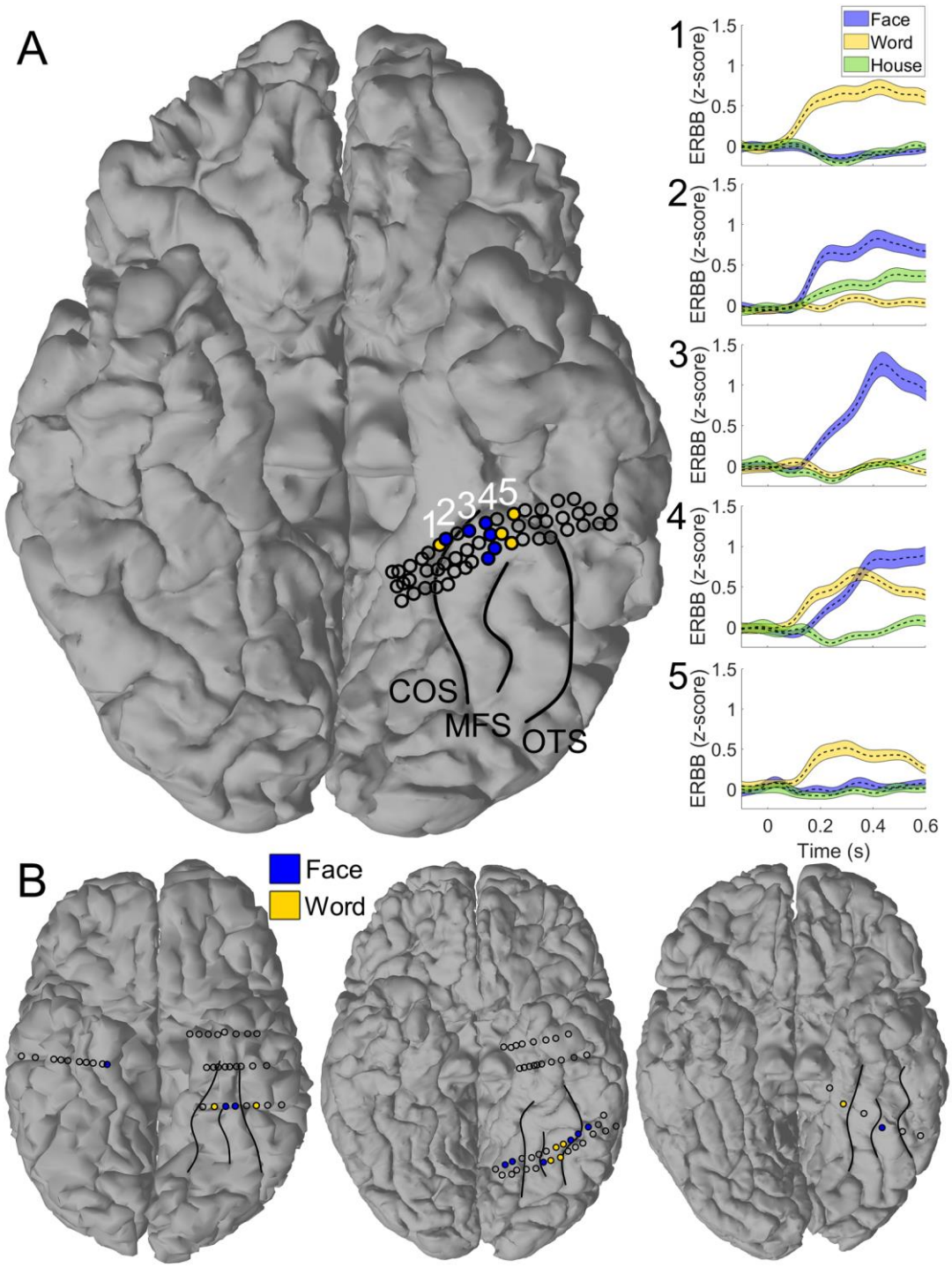


Figure 10. Multiple adjoining word- and face-selective patches in individual participants. A) Representative single participant demonstrated alternating bands of word- and face selectivity along the left fusiform gyrus. Shaded electrodes are those selective to words (yellow) and faces (blue). Non-filled circles represent ventral temporal electrodes that did not reach the selectivity criterion for either of these categories. Raw event-

related broadband activity is shown for each of the numbered electrodes in the right panel. Moving from medial to lateral, electrodes demonstrate a preferential response to words, mixed response to both words and faces, preferential response to faces then preferential response to words. Abbreviations: collateral sulcus (COS), midfusiform sulcus (MFS), occipitotemporal sulcus (OTS). 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.

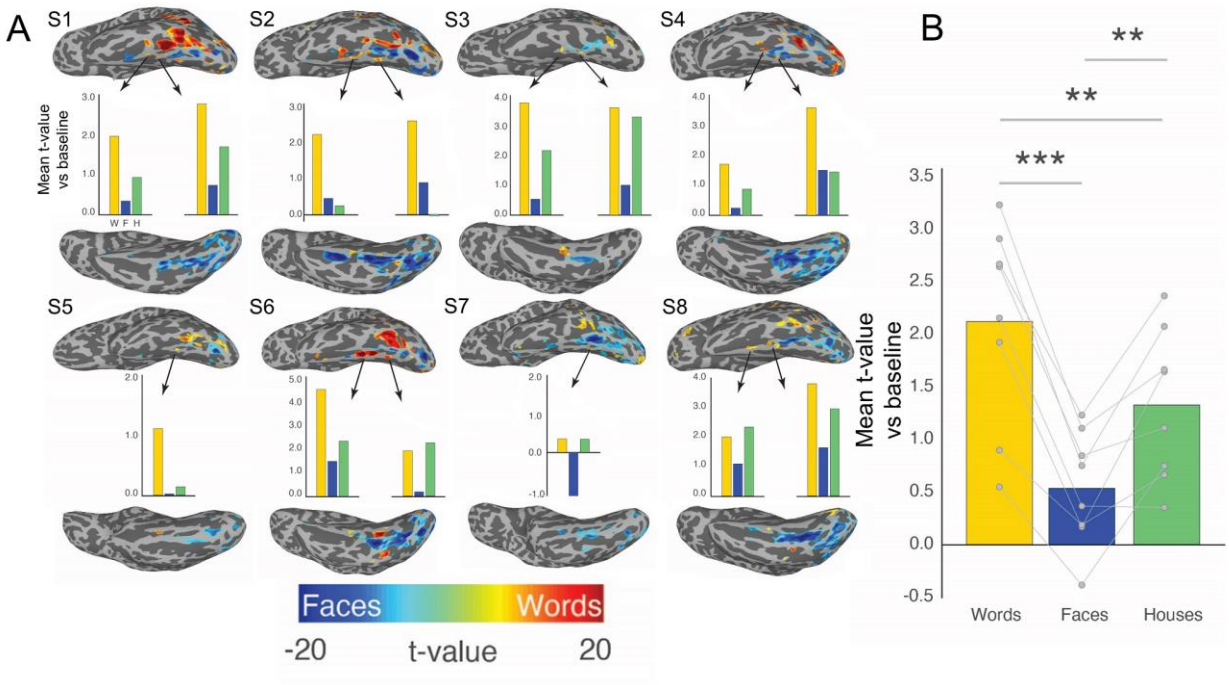


Figure 11. Interdigitation of BOLD responses to words and faces across eight healthy participants. Eight healthy participants that underwent an identical category localizer task as the iEEG participants demonstrated similar category selectivity. A) Individual maps demonstrate word versus face selectivity across left VTC. In six out of eight of these participants there was strong word selectivity medial to face selectivity along the midfusiform sulcus. The bar graphs below each participant indicates 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. Symbols: ** p < .01, *** p < .001.

The map in Figure 7 was made by combining two key aspects of the iEEG signals, the single trial potentials (stP) and the single trial high frequency broadband activity (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 overlaps, they are not redundant [108,119,152], 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). 58 electrodes showed significant selectivity in both stP and stHFBB (Figure 12A). 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 [122,136,142,145,146]. 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 (Figure 12B), whereas anterior and right hemisphere word selectivity was prevalent in either stP or stHFBB alone (Figure 12B and Figure 12C). Broadly, more electrodes demonstrated selectivity in stP (232 electrodes from 32 participants, Figure 12B) compared to stHFBB (115 electrodes from 24 participants, Figure 12C). More widespread stP selectivity is consistent with a previous study comparing stP and stHFBB responses for faces in VTC, though that study did not observe any cases where selectivity for faces was demonstrated in stHFBB but not stP [119]. The similarities and differences in selectivity demonstrated in stHFBB and stP are consistent with the hypothesis that these signals have different physiological generators [163], which may differ in their laminar distribution [120] and spatial signal-to-noise falloff [119]. 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 fMRI has differential sensitivity to these aspects of the iEEG signal [152,155,164].

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 Methods, Table 1). Activity from 85 out of 97 sampled face-selective electrodes in 13 participants could be used to reliably predict the identity of a presented face. Similarly, activity from 40 out 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, six could reliably discriminate pseudowords from real words of the same length, seven 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 $\rho(53) = .50$, $p < .0001$, face-selective: $\rho(97) = .48$, $p < .0001$). Note that correlations in peak category selectivity and within-category individuation may arise due to similar differences in measurement noise across recording contacts (for example, due to the distance the electrode was placed from the underlying face or word-selective neural populations), underlying neural/physiological factors, or some mix of both.

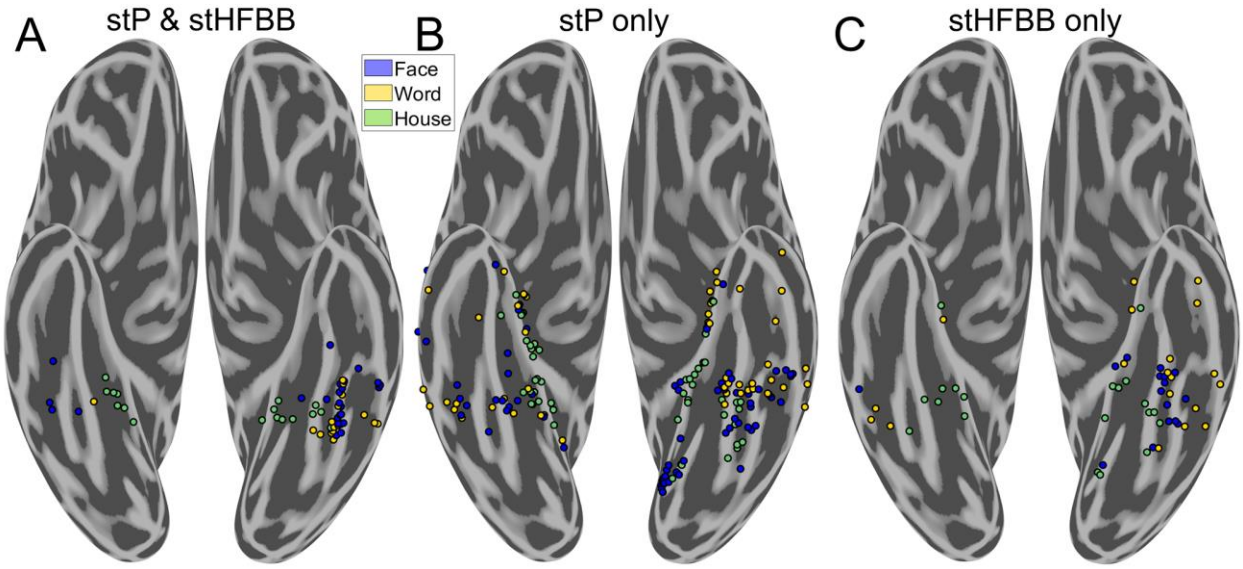


Figure 12. Comparing category-selectivity in single-trial potentials and high-frequency broadband. Differing spatial distribution of electrodes that demonstrated selectivity in single-trial potentials (stP) and single-trial high-frequency broadband activity (stHFBB). A) Electrodes that demonstrated selectivity 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.

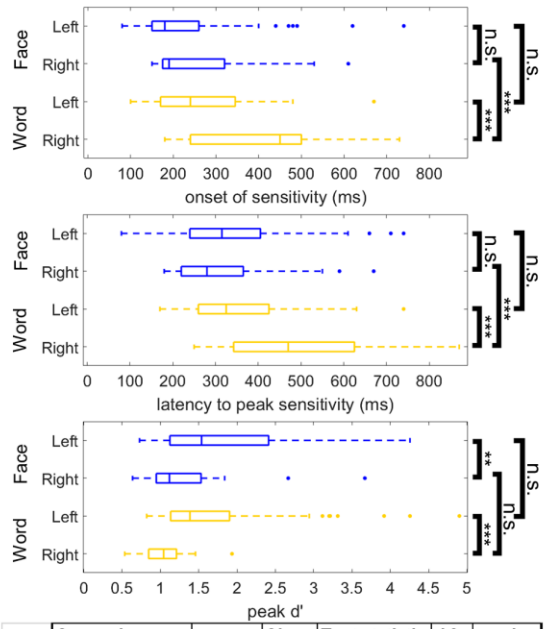
In addition to the medial band of word-selective regions, there were a high proportion of right word-selective electrodes in our iEEG population (Figure 7B, Table 1). Although this finding is consistent with some other fMRI [165,166] and iEEG studies [143,167], right hemisphere word selectivity is often not seen in neuroimaging [136,168] and was not very strong in our 7 T fMRI data either (Figure 11). 23 word-selective electrodes were found across nine participants in right VTC, out 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 lefthanded. Three out of nine of these

participants demonstrated evidence for bilateral language function while the other six demonstrated left dominant language function determined by preclinical magnetoencephalography (MEG, see Methods). Across the entire participant population, seven out 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 (Figure 13). Latency to word selectivity onset and peak was shorter in left compared to right hemisphere word-selective regions (mean onset latency difference +/- 95 % CI: -133 +/- 61 ms, $T(85) = -4.4$, $p < .0001$, mean peak latency difference: -138 +/- 63 ms, $T(85) = -4.3$, $p < .0001$, Figure 13). These relationships held when taking into account potential differences in posterior to anterior coordinate of word-selective regions across hemispheres (onset: $T(85) = -4.01$, $p = .0001$, peak: $T(85) = -3.97$, $p = .0002$). 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 +/- 53 ms, $T(106) = -1.1$, $p = .28$, mean peak latency difference: 18 +/- 57 ms, $T(106) = 0.63$, $p = .53$, Figure 13). Additionally, the amplitude of peak d' sensitivity for words was significantly greater in the left compared to right hemisphere word-selective regions (mean peak d' sensitivity difference: 0.66 +/- 0.37, $T(85) = 3.5$, $p = .0006$). The amplitude of peak d' sensitivity to faces was also significantly greater in the left compared to right hemisphere face-selective regions (mean peak d' sensitivity difference: 0.58 +/- 0.39, $T(85) = 3.0$, $p = .0037$). There was a significant correlation between peak latency and peak magnitude within face-selective

regions in the left ($\rho(80) = -0.61$, $p < .0001$) and right ($\rho(28) = -0.79$, $p < .0001$) hemisphere and word-selective regions in the left ($\rho(64) = -0.68$, $p < .0001$), but not right ($\rho(23) = -0.15$, $p = .48$) 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 ($T(85) = -1.56$, $p = .058$), but there was a greater correlation between peak latency and magnitude in left compared to right hemisphere word-selective regions ($T(85) = 2.63$, $p = .004$) 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 if 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 spatiotemporal variance with the optimal number of clusters (see Methods), we could compare the dynamics of distinct word- and face-selective clusters within VTC.



	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 13. 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 timepoint that reached a d' corresponding to p < .001 non-parametrically defined by the pre-stimulus 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. Abbreviations: confidence interval (CI), degrees of freedom (d.f.), left (L), right I. Symbols: n.s. p > 0.05, * p < .05, ** p < .01, *** p < .001.

Word-selective regions were clustered into 4 distinct left hemisphere clusters and 3 right hemisphere clusters (Figure 14A). Word-selective regions on the left fusiform gyrus demonstrated the earliest and strongest selectivity, peaking around 200 ms (Figure 14B, gray). Left hemisphere

medial word-selective regions and right hemisphere word-selective regions came next, peaking around 300 ms (Figure 14B, green and cyan) followed by lateral regions around 350 ms (Figure 14B, red). Word-selective regions in left anterior VTC peaked around 400-450 ms (Figure 14B, blue); right more anterior regions peaked around 600 ms (Figure 14B, 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 (Figure 14C-D).

Face-selective regions were organized into 3 distinct clusters in the left hemisphere and 4 distinct clusters in the right hemisphere (Figure 14E). Face-selective regions of the left and right fusiform gyrus demonstrated the earliest and largest peak selectivity around 200-250 ms (Figure 14F, gray and cyan). More anterior right hemisphere regions and a cluster of electrodes in left posteromedial VTC (Figure 14F, yellow and green) peaked around 300 ms. Finally, more anterior face-selective electrodes in left and right VTC peaked around 400 ms (Figure 14F, 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 (Figure 14G-H).

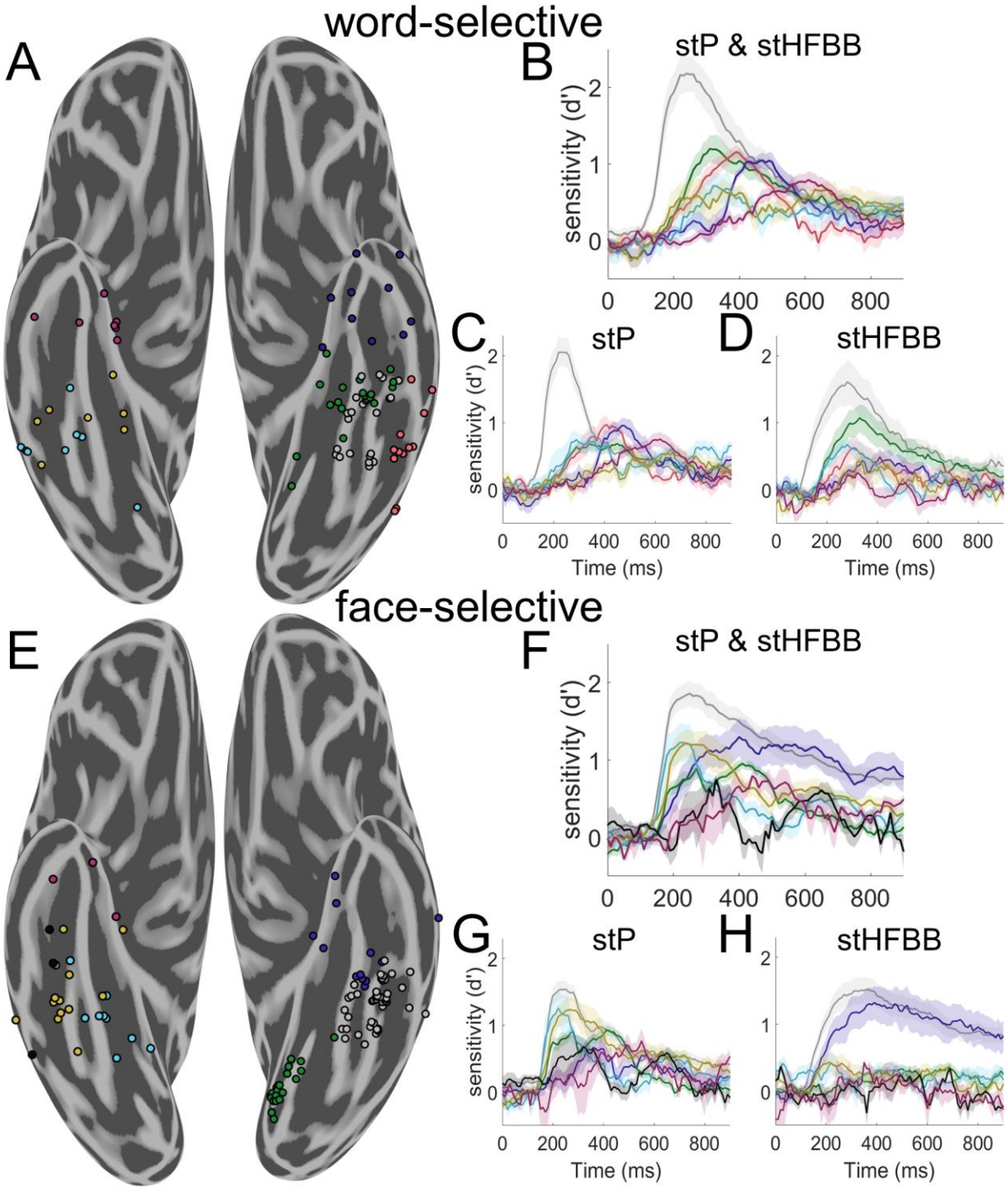


Figure 14. Spatiotemporal clustering of word- and face-selective regions. A) Spatiotemporal clustering of word-selective VTC electrodes. The illustrated clustering solution was robust to different weightings of spatial and temporal information. Left hemisphere word-selective electrodes were clustered 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 represent standard error across electrodes. C) Average d' time-course of each group of electrodes when classifying only stP. D) Average d' time-course of each electrodes 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.

From electrodes sampled in the word individuation task, we observed stronger word individuation in left word-selective regions on the fusiform compared to the more medial word-selective cluster illustrated in Figure 14A (peak d' of fusiform minus medial regions: $T(30) = 3.62$, $p = .001$, linear mixed-effects model). There was no significant difference between the latency to peak word individuation across these clusters ($T(30) = 2.91$, $p = .68$). 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 ($T(50) = 1.03$, $p = .31$) nor

latency to peak face individuation ($T(50) = -0.21$, $p = .84$) was significantly different between face-selective regions along the left fusiform gyrus and the posteromedial face-selective cluster observed in Figure 14E. 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 [34,40], but not long enough to exclude them from being relevant to perceptual behavior [130,131] .

3.4 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 (Figure 9). Electrodes which demonstrated preference to only words or faces suggests that VTC word- and face-processing networks are not entirely overlapping [28], but instead involve at least some independent nodes [4,158], which is also supported by stimulation and lesion evidence [19].

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 [9]. The word-selective regions around

the occipitotemporal sulcus are consistent with prior studies showing word selectivity within lateral aspects of VTC [169,170]. 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 [122] or distinct patches [142,165].

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 [8,148,153]. While there has been some mixed reporting of word selectivity in anterior and medial VTC regions [8,65,147–149,161,162], most models of orthographic-processing within VTC consider only the more lateral, traditional "visual word form area" [169,170]. 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 [171], signal dropout induced by susceptibility artifacts [160], 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 at the group level and 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 [150,151,159,167]). This mosaic organization of visual word-selective regions is similar to the mosaic organization of auditory language processing networks [172], suggesting this pattern of organization may not be specific to the visual system.

Medial word-selective regions may reflect differential mediolateral VTC selectivity to object rectilinearity, with more medial VTC being more responsive to straight over curvy objects [54,55]. However, the interdigitation of word- and face-selective regions along the mediolateral

axis is not well captured solely by this rectilinear model or the retinotopic model. Instead, medial and lateral word-selective regions with distinct dynamics may indicate an interaction between multiple representational axes in VTC [9,135] and possibly competition between faces and words for cortical space [60].

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 [22,23,173,174]. However, generalized language deficits after lesions of the basal temporal language area are relatively minor [173] and the relationship between reading deficits and VTC lesions, other than the visual word form area [19,25], is unclear. A recent study reported differential language-related deficits during reading, repetition, and picture naming depending on the area of VTC stimulated [21]. 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 [156], with some suggesting greater sensitivity to stHFBB [152,155]. 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 shows 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 [120,175,176] 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 to left word-selective regions, which may reflect the primary role the left, language dominant, hemisphere plays in word-processing [177]. Previous studies have demonstrated weaker correlations between object-selectivity measured with iEEG and fMRI correlations at later time windows [155]. 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 [165]. 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 [154,178–181].

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 400-600 ms. Further, the relatively later responses of word-

selective regions outside of the fusiform may also contribute to differences in category-selective maps drawn from iEEG and fMRI [155].

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 to face-selective regions on the fusiform suggests this posterior face selectivity is a result of top-down attentional effects previously reported during face-viewing [182]. 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 is somewhat consistent with feedforward models of visual processing, the relative latencies are far longer than would be expected in these models [34,40]. These results instead suggest more extended dynamics, perhaps governed by recurrent processes [34], with different category-selective regions contributing differentially to multiple, temporally extended stages of face- and word-processing [19,66,129]. 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 [183,184].

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.

4.0 Left word-selective ventral occipitotemporal cortex interacts with early visual cortex and the anterior temporal lobe to support word recognition

Last chapter we explored how face and word processing circuits in VTC are organized and their temporal dynamics. Next, we focus on how single word representations evolve in one important word-selective region in ventral occipitotemporal cortex (wVOT), sometimes referred to as the Visual Word Form Area. A previous iEEG study from our lab demonstrated that representations within wVOT undergo a dynamic shift over time, from a coarse level that can be used to discriminate between visually dissimilar words, and a later (after 250 ms) fine level representation that can differentiate between visually similar words [19]. Here, we replicate this finding in healthy individuals using magnetoencephalography (MEG) and demonstrate that this shift is stronger for real words compared to other word-like stimuli including pseudowords, false fonts, and consonant strings, which lack the semantic and/or phonological associations of real words. During this shift, increased connectivity to early visual cortex and anterior temporal lobe regions was observed. These results suggest that bottom-up and top-down interactions facilitate the disambiguation of word representations in wVOT. This chapter is available on the preprint server bioRxiv: MJ Boring, EA Hirshorn, Y Li, MJ Ward, RM Richardson, JA Fiez, AS Ghuman. Left mid-ventral temporal cortex interacts with early visual cortex and the anterior temporal lobe to support word individuation. *bioRxiv*, 411579 (2018).

4.1 Introduction

A word-selective region in the left ventral occipitotemporal cortex (wVOT), sometimes referred to as the "Visual Word Form Area (VWFA)," responds preferentially to words over other object categories [168,185] and is thought to play a key role in reading [25,62,144,168,185,186]. It has been shown that damage to [19,25,187,188] or stimulation of [19,22] the wVOT can cause pure alexia, and there is evidence that reading expertise shapes response properties of the wVOT, including differential activation to real words versus pseudowords (pronounceable but meaningless letter strings) [141,153,189–193]. However, there is still debate over the nature of orthographic representation in the wVOT. Specifically, does the wVOT encode whole-words [14,193,194], sub-lexical features [19,136,144,169,195,196], or purely visual statistics that are preferentially fed into higher-order language centers [62,170]?

A recent intracranial electroencephalography (iEEG) study demonstrated that early activity in the wVOT only allowed the decoding of words that were orthographically dissimilar (hint vs. dome). The activity then evolved in a way that also allowed orthographically similar real words (hint vs. lint) to be disambiguated after 200 ms [19]. These results suggest that representations in the wVOT are initially coarse but evolve over time, eventually allowing for the disambiguation of orthographically similar word forms. However, the degree to which this process is specific to known printed words, which have learned semantic and phonological associations in addition to their visual properties, has not yet been determined. Further, it is unknown if word individuation is achieved solely through hierarchical visual processing [194] or is instead driven by interactions between the wVOT and other parts of the language network that underpin phonological and/or semantic knowledge about words [141,177,197,198].

The current study probes these questions by examining the wVOT response to visually similar real words, pseudowords (pronounceable but meaningless letter strings), consonant strings (meaningless and unpronounceable letter strings), and false fonts (orthographic stimuli of an unfamiliar alphabet) using source-localized magnetoencephalography (MEG) and iEEG. We hypothesized that if the disambiguation of orthographic representations in the wVOT relies on learned semantic or phonological associations inherent to real words, then orthographically similar pseudowords and/or consonant strings would not be disambiguated by wVOT activity. In contrast, if the refinement of wVOT representation is independent of learned semantic or phonological knowledge, then we would expect to see a similar disambiguation for orthographically similar consonant strings and pseudowords. Additionally, we examined the functional connectivity between wVOT and the rest of the cortex during the transition between coarse and individuated representations to assess the extent to which the disambiguation of orthographic representations is a network-level process.

Our results demonstrate that wVOT responses to real words, pseudowords, consonant strings and false fonts allowed for reliable decoding of orthographically dissimilar stimuli from one another. However, while orthographically similar real words could be discriminated from wVOT activity from 200-350 ms, orthographically similar pseudowords, consonant strings, and false fonts could not. Functional connectivity analysis in MEG revealed increased phase-locking between the wVOT and left anterior temporal lobe and early visual cortex during the transition from these coarse to individuated real word representations. Taken together, these results support the idea of early, coarse wVOT representations that subsequently evolve through interactions with visual and semantic networks to allow for the disambiguation of orthographically similar real words.

4.2 Methods

4.2.1 MEG data collection and preprocessing

4.2.1.1 Participants

16 participants gave written informed consent to participate in the MEG portion of the experiment consistent with protocol approved by the University of Pittsburgh's Internal Review Board. One participant was removed from the analysis due to poor cortical surface reconstruction leaving 15 (5 males, ages 19-29) for the remaining analyses.

4.2.1.2 Experimental paradigm

First, a category localizer consisting of words, hammers, houses, and false fonts was administered to identify word-selective cortical sources and constrain the word-individuation analysis. Then, a word individuation task was administered to probe the dynamics of word representation across different types of orthographic stimuli. For both the category localizer and word-individuation tasks stimuli were presented via custom scripted code in Psychtoolbox [106] on a screen one meter in front of the participants. Stimuli occupied approximately 6 x 6° of visual angle and were shown for 300 ms with a variable inter-stimulus interval of approximately 1.5 s. One-sixth of the time a stimulus would be repeated, to which the participant responded with a button press. These trials were removed from the subsequent analyses. Three blocks of 140 trials each were performed for the category localizer and 5 blocks of 264 trials each were performed for the word individuation task. In total there were 90 trials per stimulus category in the category localizer and 30 trials per stimulus in the word-individuation task, after removing repeated trials.

In the word-individuation task, word and word-like stimuli consisted of four pairs of real words, pseudowords, false fonts (Old Hungarian alphabet) and consonant-strings each differing from each other in only one symbol or letter within pairs. All false fonts and consonant-strings had five symbols or letters, pseudowords had either four of five letters, and words had either three or four letters. Decoding analysis within stimulus categories were only performed across stimuli with the same number of letters and symbols to prevent length effects. Real word stimuli were selected to have similar log frequency, mean bigram frequency and bigram frequency by position across similar and dissimilar word pairs (measured using the English Lexicon Project [199]). Pseudowords were selected to have similar orthographic neighborhood size and bigram frequency by position across similar and dissimilar pseudoword pairs.

4.2.1.3 Structural MRI acquisition and preprocessing

T1 structural MRIs were used to constrain the cortical source estimates of the current study. Images were acquired with a Siemens 3T Tim Trio system scanner using a magnetization-prepared rapid acquisition with gradient echo sequence (TR = 2100 ms, T1 = 1050 ms, TE = 3.42 ms. 8° flip angle, 256x256x192 acquisition matrices, FOV = 256 mm, and 1 mm isotropic voxels). Cortical surface reconstructions were extracted via Freesurfer [102].

4.2.1.4 MEG acquisition, preprocessing, and source localization

MEG data were collected on an Elekta Neuromag VectorView MEG system (Elekta Oy, Helsinki, Finland) with 306 sensors (triplets of two orthogonal gradiometers and one magnetometer). Data were sampled at 1000 Hz with simultaneous recording of head position, electrooculogram, and electrocardiogram which were all corrected for off-line. The data were

processed with temporal signal-space separation [200], a 1-50 Hz bandpass filter, and down-sampled to 250 Hz for subsequent analyses.

Minimum norm estimate (MNE) software [201] was used to project the sensor data onto Freesurfer cortical reconstructions. Regions of interest were manually drawn around the left fusiform gyrus for each subject. Single compartment boundary-element models were calculated from the Freesurfer segmentation and used to compute forward solutions separately for each block, taking shifts in head position into account. Noise covariance matrices were computed from the inter-stimulus interval period, 500 to 30 ms prior to each stimulus presentation. Inverse operators were constructed using the computed noise covariance and forward solutions to obtain source estimates for approximately 7,600 vertices on the cortical surface reconstruction of each subject. Because magnetic sources originating from cortical neurons are typically normal to the cortical surface, tangential source components were scaled by a factor of .4 during the calculation of the inverse solution [202,203]. This procedure resulted in activity of 50-150 wVOT sources during the category localizer and word-individuation tasks for each subject.

4.2.1.5 Identification of word-sensitive wVOT sources

Sources in the wVOT were screened for word selectivity using four-way support vector machines [SVM] applied to 100 ms sliding time windows independently for each source. If the d' sensitivity index, defined as the inverse of the cumulative normal distribution for true positives for words minus the inverse of the cumulative normal distribution for false positives for words, exceeded chance with $p < .05$ (uncorrected) a particular source it was considered "word-selective" and belonging to the wVOT. This yielded a mean +/- standard deviation of 42.4 +/- 32.7 word-selective channels per subject. Only these sources were used for word-individuation decoding. Figure 15 shows the location of these word-sensitive sources across the group. Figure 30 shows

the mean event-related field of these word-selective sources to the different stimuli presented in the category localizer task.

4.2.2 Intracranial EEG data collection and preprocessing

4.2.2.1 Participants

Three right-handed patients (2 females, ages 38-64) with intractable epilepsy were included in the study. Inclusion was based on iEEG coverage in left mid-ventral temporal cortex that demonstrated selectivity to words over the other stimulus categories, as defined by the broadband gamma response, event-related potential amplitude, and d' sensitivity index in an independent category localizer containing words, faces, bodies, houses, hammers, and phase-scrambled objects. Figure 31 shows the d' sensitivity of each word-selective electrode. Figure 32 shows the event related potential or broadband response of each word-selective electrode to words and other object categories presented during the category localizer task. Figure 16 shows the average d' sensitivity of the seven word-sensitive electrodes identified across the three subjects. These electrodes were localized using either post-operative T1 structural MRI's or CT scans. Figure 15 illustrates the location of the word-sensitive electrodes in Montreal Neurological Institute (MNI) stereotaxic-space. Figure 33 illustrates the location of each electrode on the individual patient's anatomy. None of the electrodes presented here demonstrated ictal activity during the recording sessions, nor were they near the patient's seizure onset zone. One of the three patients (P1) was included in a previous study [19]; however, data from non-word orthographic stimuli in P1 were not previously reported. The other two participants from that previous study were not shown non-word stimuli, and therefore are not reported here. All patients gave written informed consent under protocols approved by the University of Pittsburgh Medical Center's Internal Review Board.

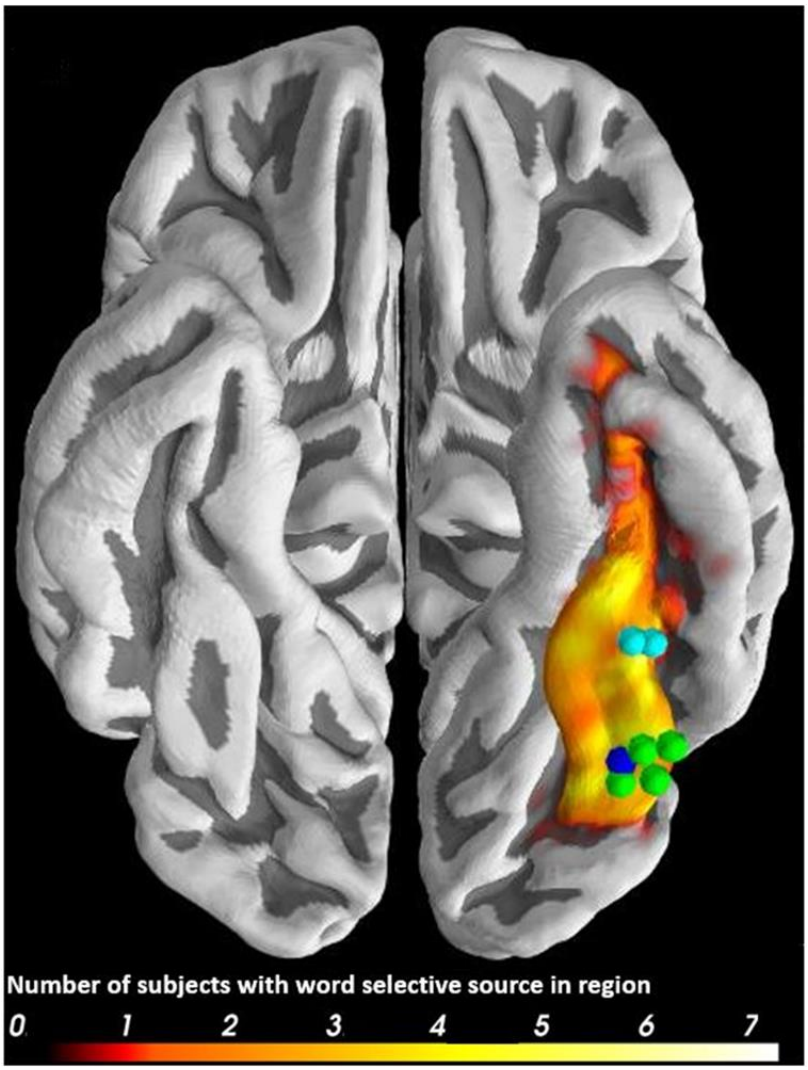


Figure 15. Word-selective MEG sources and iEEG electrodes in wVOT. Word-selective electrodes and sources in Montreal Neurological Institute common space. Dots are electrodes from the three iEEG patients (P1-green, P2-dark blue, P3-light blue). Number of MEG subjects with word-selective sources localized to a given region of the fusiform represented by color intensity. All sources are constrained to the left fusiform gyrus of the individual's anatomy.

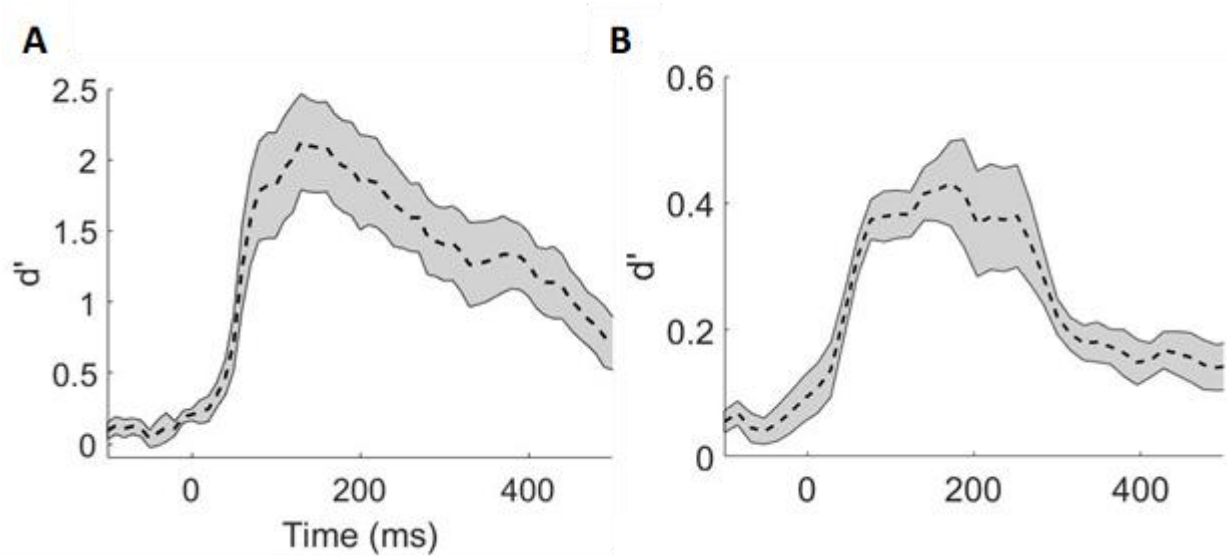


Figure 16. Average d' sensitivity of word-selective iEEG electrodes and MEG sources. A) Average sensitivity (norminv(true positive for words) – norminv(false positive for words)) of word-selective electrodes in a six-way SVM classifier across all three iEEG participants. Grey represents standard error from the mean across all electrodes. B) Average sensitivity across word-selective MEG sources in wVOT in a four-way SVM classifier. Grey represents standard error across all subjects.

4.2.2.2 Experimental paradigm

The experimental paradigm for intracranial subjects was the same as that of the MEG participants besides the following differences: The intracranial category localizer consisted of words, bodies, faces, hammers, houses, and phase scrambled objects. Stimulus on-times for both the category localizer and word-individuation task were increased to 900 ms with 1.8 s mean inter-stimulus interval. The word-individuation task contained the same stimuli as the MEG version; however, to maximize the number of trials per remaining stimuli, consonant-strings were dropped from the stimulus set. Overall, there were approximately 25, 45, and 30 trials per word-individuation stimulus for P1, P2 and P3 respectively, varying according to number of blocks of the task completed.

4.2.2.3 iEEG acquisition and preprocessing

Local field potentials were collected using a Grapevine Neural Interface Processor (Ripple, LLC) at 1000 Hz. Data was bandpass filtered offline from 0.2-115 Hz and notch filtered to exclude 60 Hz line noise using a fourth-order Butterworth filter implemented with FieldTrip [107]. In addition to this, broadband gamma amplitude, defined as the average increase in power from 40-100 Hz, was extracted and normalized to baseline (from 300-50 ms prior to stimulus presentation). Trials with peak amplitudes exceeding 5 standard deviations above or below the mean or exceeding 350 microvolts were eliminated to reduce potential artifacts.

4.2.3 Multivariate temporal pattern analysis

Data from MEG sources and iEEG electrodes identified as word-selective in the category-localizer task were used for all possible pairwise decoding of the word-individuation stimuli. For example, all word-selective sources in one subject were used as features to a two-class, 3-fold cross-validated SVM classification problem applied to two independent time windows to determine whether the participant was seeing stimulus A or B. Time windows were chosen based on the results from our previous study [19]: 50-200 ms and 200-350 ms for early and late stages of wVOT processing. LIBLINEAR [204] was used to implement the SVMs. This resulted in classification accuracy and d' sensitivity for each pairwise classification problem (44 x 43 / 2) across both time windows. We choose to report d' sensitivity here because it is normally distributed, unlike classification accuracy, which allows for parametric statistical testing across subjects. Additionally, d' sensitivity captures effect sizes on the same scale as Cohen's d, making it easily interpretable. Pairwise d' sensitivities in the early and late time windows are averaged according to the contrast of interest. For example, when determining the classifier sensitivity to

real words versus false fonts, all possible pairwise d' sensitivities between word and false font stimuli were averaged to create one average d' sensitivity per subject per time window.

Statistical significance of classification accuracy was determined via non-parametric permutation tests. Specifically, category labels were permuted randomly across each pairwise comparison then the two-class SVM was trained on data from the randomly permuted class labels. Classification accuracy for both time windows were computed for 1000 random permutations on the iEEG data and MEG data then averaged over the contrast of interest. Maximum classification accuracy across both time windows was used to construct the null-distribution of classification accuracy and then compared with the corresponding real-label time-course. For the MEG data, to obtain the statistical significance of group-wise average classification accuracy, the permuted time-courses were also averaged across subject before calculating the maximum accuracy for each of the 1000 random permutations.

4.2.4 Functional connectivity analysis

Functional connectivity analysis was carried out on the MEG data to evaluate the connectivity dynamics of the wVOT to the rest of cortex that facilitates real word individuation. Specifically, activity of word-selective sources in wVOT were averaged and phase-locking values (PLV) [113] were calculated between this activity and all other cortical sources during the word-individuation task. PLVs were normalized by taking their square root and standardizing relative to a baseline period from 500 to 0 ms prior to stimulus presentation [115].

Numerous previous non-invasive EEG studies have demonstrated functional connectivity differences in the delta, theta, alpha, and beta frequency bands related to various aspects of reading [205–207]. Therefore, we hypothesized communication between the wVOT and rest of the

language network would be most likely to occur in this frequency range. However, when calculating the phase of low frequency oscillations (i.e., delta and theta bands) using wavelets, these calculations have less temporal resolution than the higher frequency alpha and beta components. Therefore, transformed PLVs were averaged over only the canonical alpha and beta frequency bands (8-30 Hz) then co-registered to the MNI common brain.

To determine spatiotemporal clusters of sources whose PLV to the wVOT was significantly greater than baseline during the transition from coarse to fine representations (which occurred at approximately 175-225 ms post-stimulus presentation), cluster statistics were determined via a within-subjects permutation test. Specifically, T-statistics were computed for each source and clustered based on adjacency in time and cortical space. The sum of T-values within each cluster was then compared with the maximum cluster T-value of 10,000 randomly generated sign flipped matrices. This procedure has been shown to effectively correct for multiple spatiotemporal comparisons [112].

4.3 Results

4.3.1 Decoding real words from other orthographic stimuli

A support vector machine was trained to discriminate between pairs of word versus other orthographic stimuli using both word sensitive MEG sources and word sensitive iEEG electrodes in wVOT. Real words could be discriminated from false fonts, consonant-strings and pseudowords from MEG source activity during both the early (50-200 ms) and late (200-350 ms) time windows

($p < .001$). Comparable results were seen for word-selective iEEG electrodes in the wVOT besides null results for words versus pseudowords in both windows for patient 3 (Table 2).

Table 2. Decoding real words from other stimuli (mean pairwise d' sensitivity).

False fonts vs real words

	50-200 ms	200-350 ms
Patient 1	0.80***	1.5 ***
Patient 2	0.72 ***	1.3 ***
Patient 3	0.13***	0.11**
MEG	0.26***	0.40***

Pseudowords vs real words

	50-200 ms	200-350 ms
Patient 1	0.40***	1.1***
Patient 2	0.36***	0.54 ***
Patient 3	0.046	-0.034
MEG	0.051**	0.16***

Consonant-strings vs real words

	50-200 ms	200-350 ms
MEG	0.17***	0.28***

*p<.05 **p<.01 ***p<.001

A one-way ANOVA indicated a significant difference between mean MEG d' sensitivity for false fonts, consonant strings, and pseudowords versus real words in the early time window ($F = 5.29$, $p < .01$). A post-hoc T-test demonstrated that false fonts versus real words displayed higher d' sensitivity than pseudowords versus real words across MEG participants in the early time window($p < .01$, Bonferroni corrected). Sensitivity for consonant strings versus real words was not significantly different than either false fonts versus real words ($p > .2$, Bonferroni corrected) or pseudowords versus real words ($p > .5$) in the early time window. A one-way ANOVA on the late time window also revealed a significant difference between mean MEG d' sensitivity for false fonts, consonant strings, and pseudowords versus real words ($F = 3.47$, $p < .05$). A post-hoc T-test demonstrated that false fonts versus real words displayed higher d' sensitivity than pseudowords versus real words across MEG participants during the late time window ($p < .05$, Bonferroni corrected). However, there was no significant difference between the decoding accuracy of

consonant strings versus real words and false fonts versus real words ($p > .5$, Bonferroni corrected) or pseudowords versus real words ($p > .6$, Bonferroni corrected) in the late time window.

4.3.2 Decoding orthographically similar and dissimilar stimuli

Table 3. Decoding orthographically similar and dissimilar stimuli (mean pairwise d' sensitivity).

Orthographically similar real words		
	50-200 ms	200-350 ms
Patient 1	-0.0006	0.56**
Patient 2	-0.075	0.85***
Patient 3	-0.13	-0.26
MEG	-.048	0.10*

Orthographically dissimilar real words		
	50-200 ms	200-350 ms
Patient 1	0.41***	0.96***
Patient 2	0.62***	0.60***
Patient 3	-0.12	-0.028
MEG	0.091***	0.14***

Orthographically similar pseudowords		
	50-200 ms	200-350 ms
Patient 1	-0.049	0.040
Patient 2	-0.041	0.089
Patient 3	-0.0002	0.061
MEG	0.087	0.040

Orthographically dissimilar pseudowords		
	50-200 ms	200-350 ms
Patient 1	0.29**	0.72***
Patient 2	0.17	0.19*
Patient 3	0.48054	0.016
MEG	0.090**	0.13***

Orthographically similar consonant-strings		
	50-200 ms	200-350 ms
MEG ONLY	-0.079	0.048

Orthographically dissimilar consonant-strings		
	50-200 ms	200-350 ms
MEG ONLY	0.040	0.13***

Orthographically similar false fonts		
	50-200 ms	200-350 ms
Patient 1	0.029	0.43*
Patient 2	0.47*	-0.092
Patient 3	-0.049	-0.083
MEG	0.017	0.069

Orthographically dissimilar false fonts		
	50-200 ms	200-350 ms
Patient 1	0.13*	0.70***
Patient 2	0.11	0.35***
Patient 3	-0.018	0.19*
MEG	0.10***	0.18***

*p<.05 **p<.01 ***p<.001

First we determined if we could replicate our previous iEEG results regarding the dynamics of similar and dissimilar individual word decoding [19] using source localized MEG. Using activity evoked from word-selective MEG sources in wVOT, orthographically dissimilar real

words could be decoded from one another in the early and late time windows ($p < .001$ in both windows, see Table 3). However, orthographically similar real words could not be significantly decoded from each other until the late time window from 200-350 ms ($p < .05$). A similar pattern was seen in 4 of 5 iEEG subjects (2/3 reported here and 3/3 reported previously [19], with one subject shared between the two studies, see Methods). This pattern is consistent with our previous iEEG results regarding the dynamics of similar and dissimilar individual word decoding [19].

Like real words, orthographically dissimilar pseudowords ($p < .01$ early, $p < .001$ late), consonant-strings ($p < .001$ only in the late time window), and false fonts ($p < .001$ in both windows) could be significantly discriminated within wVOT activity using MEG and in most iEEG cases, particularly in the late time window (Table 3). In contrast to real words, orthographically similar consonant-strings, pseudowords and false fonts could not be consistently decoded with MEG or iEEG at either time window (Table 3).

4.3.3 Functional connectivity between the wVOT and rest of the brain during the disambiguation of real word stimuli

An open question is whether the individuation of real word stimuli is achieved locally within wVOT or involves a network level process that wVOT contributes to. To determine functional interactions that occur during the transition from coarse to fine orthographic representations in the wVOT, phase locking values (PLVs) were calculated between the average activity of word-selective sources localized to wVOT and the rest of the cortical sources from 175-225 ms post stimulus presentation. Normalizing these PLVs with respect to baseline and averaging over canonical alpha and beta frequency bands (see Methods) gave two clusters of sources that demonstrated above chance connectivity relative to pre-stimulus baseline at the cluster-level ($p <$

.05). Figure 17 illustrates the spatial locus of these clusters, which includes one that extends from right early visual cortex to right lateral occipital cortex and one encompassing a region anterior to the left fusiform gyrus, extending from the parahippocampal gyrus to the inferior temporal sulcus. No significant clusters were seen when directly contrasting the PLVs evoked by real words to those evoked by pseudowords, consonant strings, or false fonts.

4.4 Discussion

We found that wVOT representations of real words, pseudowords, consonant strings and false fonts were all initially coarse, allowing only for decoding of visually dissimilar stimuli from each other. Real word representations in the wVOT became disambiguated over time, allowing for reliable decoding of orthographically similar real words from wVOT activity after 200 ms post-stimulus presentation. However, decoding of visually similar pseudowords, letter strings and false fonts from wVOT activity did not rise above chance during either the early or late time window. Finally, in the transition period between coarse and individuated representations of real words, we observed significant functional connectivity between the wVOT and early visual cortex and between the wVOT and more anterior regions of the left temporal lobe.

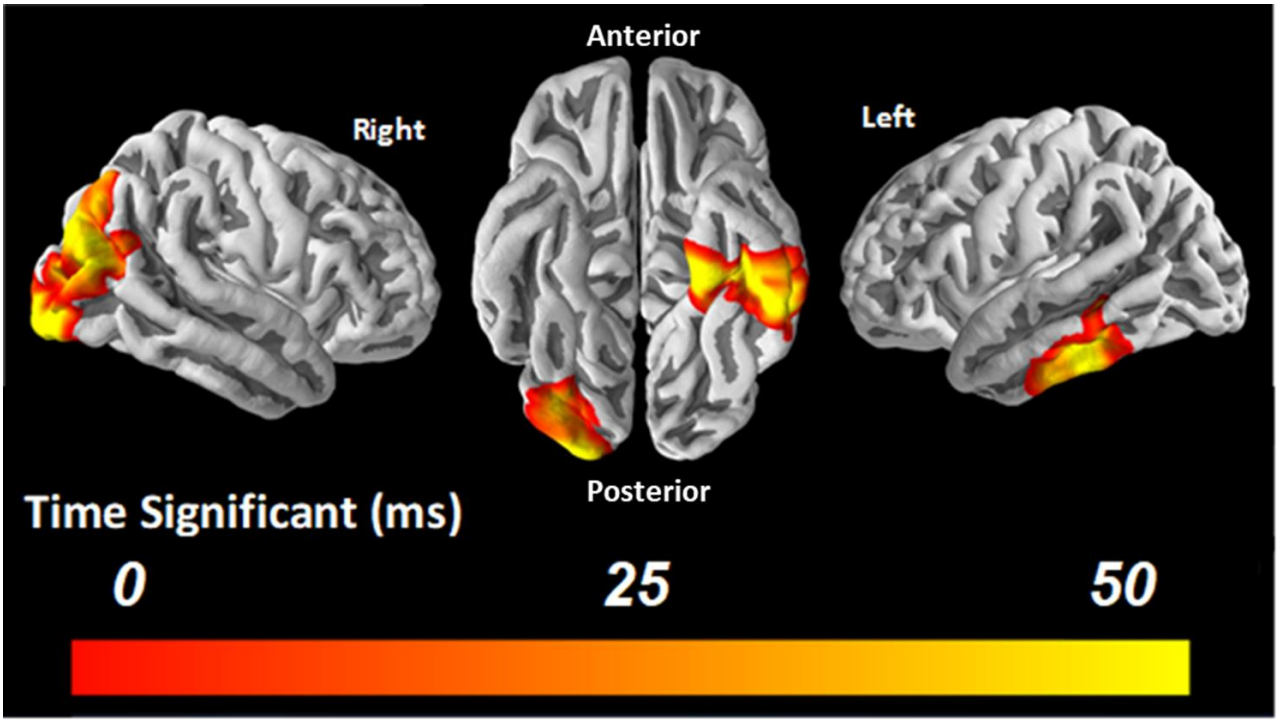


Figure 17. Significant clusters of phase-locking between the wVOT and rest of cortex 175-225 ms after real word presentation. Spatiotemporal clusters of significant phase-locking values (PLV) to the word-selective sources in wVOT during the individuation of real word representations (175-225 ms). Color intensity illustrates the duration each source had elevated PLV ($p < .01$, uncorrected) during real word trials versus baseline with 50 ms being the maximum possible duration. Significant clusters ($p < .05$, corrected) include one in right early visual cortex and in the left anterior temporal lobe.

Our decoding analyses for orthographically similar versus dissimilar real words replicate a previous finding from our group: real word representations in the wVOT are initially coarse but disambiguate with time to allow for the reliable representation of orthographically similar real words [19]. This replication, in addition to the high correspondence between the MEG and iEEG results presented here, provide an important cross-validation of iEEG and source-localized MEG. Specifically, these results demonstrate the sensitivity of MEG to subtle stimulus-induced changes in neural activity and source-localization's ability to approximately identify the neuroanatomical

origins of those neural signatures. Furthermore, the correspondence between source-localized MEG and iEEG validate iEEG results using data from a healthy population with larger sample size. Additionally, MEG supplements sensitive iEEG data with full brain coverage, facilitating analyses that require broader coverage, like the functional connectivity analysis presented here. Thus, combining iEEG and MEG with similar experiments is a potentially powerful paradigm to cross-validate, replicate, and extend findings by leveraging the respective strengths of these two recording techniques.

By combining MEG and iEEG in the current study, we were able to demonstrate that early, coarse coding in the wVOT exists not only for real words, but also false fonts, consonant-strings and pseudowords. This early representation may support rapid disambiguation of orthographically dissimilar stimuli. However, our data suggests that these coarse wVOT representations are subsequently disambiguated through interactions between visual and semantic networks, which allows the reliable individuation of orthographically similar real words. Given that we were only able to decode visually similar real words from late wVOT activity, and not visually similar stimuli from other orthographic categories, this suggests that wVOT representations of learned word forms are individuated to a greater degree than unfamiliar orthographic entities.

This conclusion is supported by previous studies observing decreases in wVOT BOLD responses for learned relative to unfamiliar orthographic entities [14,193], which may reflect more individuated wVOT representations for learned word forms in the later stages of processing. Further, the time-course of real word individuation in the wVOT is supported by scalp EEG evidence demonstrating lexical-semantic influences on visual word recognition that are observed approximately 250 ms post-stimulus presentation, as shown by studies involving transposed letter [208] and morphological primes [209]. However, the current study cannot rule out the possibility

that the wVOT has the capacity to individuate unfamiliar orthographically stimuli, either through purely bottom-up visual mechanisms or interactions with phonological processing networks, since care must be taken when interpreting null results. With that in mind, the results reported here do suggest that these stimuli are likely represented less robustly than known word forms.

Our results, which support early, coarse orthographic coding in the wVOT, contrasts with previous results obtained from rapid adaptation functional magnetic resonance imaging (fMRI) studies. [194] reported no effect of orthographic similarity on the on the adaptation of BOLD response to successively presented real words. Based on these results, the authors suggested that wVOT representations of real words are not coarse, but rather based on individuated whole word templates, hierarchically assembled from rapid, bottom-up visual information processing [194]. The results reported here show that the early response of wVOT is coarse, potentially reflecting an orthographic similarity space [196]. It has been suggested that the early response of an area reflects its intrinsic coding, since later activity is more susceptible to top-down and network-level influences [144]. Thus, these results suggest that wVOT's intrinsic code does not reflect individuated whole word templates. A potential explanation for the conflicting results is the difference in temporal resolution afforded by fMRI relative to MEG and iEEG. The sluggish hemodynamic response measured by fMRI may be disproportionately sensitive to the late stage of wVOT processing, when visually similar real words can be disambiguated from each other. Our results suggest that the early representations in the wVOT, which potentially arise from bottom-up visual processing, are consistent with coarse orthographic coding. Whole word representations emerge in the wVOT over time, however they likely require network interactions with semantic and visual regions to reliably disambiguate orthographically similar real words [197].

Functional connectivity analyses presented also support this conclusion. During real word trials, differences in phase-locking between two spatially distinct clusters, one centered on early visual areas (early visual cortex and right lateral occipital cortex) and the other on the left anterior temporal lobe, were present during the 50 ms transition from early to late decoding windows. This suggests that the individuation of real word representations in the wVOT takes place through recurrent interactions between regions both earlier in the ventral visual hierarchy and higher-level processing regions. Interactions with occipital regions may reflect continued accumulation of visual information over time, while anterior temporal regions may contribute learned information about real words that support the disambiguation of word forms through semantic properties which are largely orthogonal to the orthographic properties of printed words [197]. This role for the anterior temporal cortex in reading is supported by studies demonstrating increased BOLD activation of this region to pseudowords trained to have semantic associations [210] and studies of sentence comprehension [211].

However, no clusters of wVOT functional connectivity were found to be significantly different between real words and the other word-like stimuli at the group level. Thus, it may be that neural communication is shared among a similar set of regions regardless of the stimuli, but only supports individuated representations in the wVOT if there is useful stored information in a given node of the network. Notably, these functional connectivity results suggest that individuated representations are an emergent property of network interactions, with multiple nodes of the network contributing to and reflecting individuated representations. Thus, individuation of real words in the wVOT is unlikely to be a result of solely visual processing occurring in this region, but rather a local reflection of a network-level computation. A similar timing pattern has been reported for face individuation in the fusiform gyrus, where early activity in response to faces is

coarse and later activity is individuated [66] and may also reflect network-level interactions [212]. This suggests that a similar dynamic process is conserved between both word and face stimuli and may reflect a general principle of visual processing for other visual stimuli as well.

Taken together, our results support the idea of an early, coarse code in the wVOT that is sharpened through recurrent interactions between occipital and anterior temporal regions. First, a coarse level representation in the wVOT, built through bottom-up visual processing, allows for decoding of visually dissimilar stimuli within 200 ms of stimulus presentation. Next, interactions between anterior ventral temporal regions, possibly containing stored knowledge about words, and low-order visual regions assist in disambiguating real word representations over time. This information ultimately allows the individuation of visually similar real words. Further work investigating to what degree this process is sensitive to word context (i.e., when a word is presented in a meaningful sentence) and whether individuated word representations are conveyed throughout the language network is necessary to better understand the computations which facilitate expert reading.

5.0 How function, connectivity, and dynamics situate the ventral visual stream in neurocognitive models of reading and language

Finally, after investigating word-processing in the extended basal temporal language system and focusing on how long-range functional interactions may help refine representations in wVOT, we present a model for how visual word processing circuits develop. This model aims to reconcile apparently conflicting data demonstrating the specificity of wVOT responses to written words but also its responses during non-visual linguistic tasks, like reading braille in blind individuals [88]. Further, the model explains why wVOT demonstrates a consistent localization across individuals despite an apparent lack of evolutionary pressures on the development of literacy, which has only become widespread over the last 200 years. The model suggests that word-selective networks rely on preexisting circuits involved in visual to phonological transformations involved in speech reading and visual to semantic transformation involved in object naming. These pathways rely on prevalent bidirectional white matter projections that exist in neonates [64,65], which constrain the localization of wVOT and can activate this region through top-down interactions when the brain processes non-visual lexical information. At the time of writing this dissertation, this chapter has not been peer reviewed, but was coauthored by Julie A. Fiez and Avniel Singh Ghuman.

5.1 History and overview

Reading involves the visual identification of orthographic forms (letters up to words, sentences, and paragraphs) and mapping those forms onto their sounds and meaning. The fact that we hear words in our heads while we read has led to a nearly 150-year debate regarding the extent to which we represent visual words in our minds at all. This debate harkens back to Joseph Jules Dejeriné's report in the 1890s of a case of alexia, an inability to read words holistically, instead reading letter-by-letter, without agraphia (no disturbance in writing) [213]. Indeed, individuals with what has since been termed "pure alexia" cannot even read things they themselves have written. This led Dejeriné and Jean-Martin Charcot to posit that the brain had a center for the "optical images of words" in their models of how the brain processes language [213]. Carl Wernicke firmly rejected this notion, instead arguing that only letters are represented visually and are directly mapped onto their phonological forms, evident by the fact that we hear words in our head as we read. Therefore, he argued the brain has no visual center for whole words, only an auditory one [214].

In the century following the original reports by Dejeriné, studies of individuals with alexia probed the extent to which the disorder is specific to reading versus a more general deficit in visual recognition that reading happens to rely heavily upon. These studies examined whether alexia could be disentangled from object naming [215–217], non-specific visual simultagnosia [218,219] (a disorder where multiple objects cannot be perceived simultaneously), and other naming and/or visual deficits [219,220]. While some debate continued regarding whether sub-clinical deficits in non-reading visual and/or naming in alexic patients indicates that the disorder is more general, these studies ultimately led to the "visual word form hypothesis," which states that there is a

“visual word form system…which parses (multiple and in parallel) letter strings into ordered familiar units and categorizes these units visually. The components can range in size from graphemes, syllables, morphemes to whole words” [219].

With the advent of modern neuroimaging, allowing for precise in vivo measurements of brain damage and neural activity, much of the debate around the visual word form hypothesis has revolved around a word sensitive area of left ventral occipitotemporal cortex (wVOT, Figure 18), sometimes referred to as the Visual Word Form Area [136,221,222]. There is general agreement that wVOT responds more strongly to words than other visual stimuli [136,221,222], and damage to this region causes reading deficits [19,220,223]. However, there remains great debate as to whether wVOT specifically represents visual word forms, or whether this region plays a more general role in cognitive processing that reading relies particularly heavily upon [28,62,170], a continuation of the debate on how “pure” a deficit alexia is.

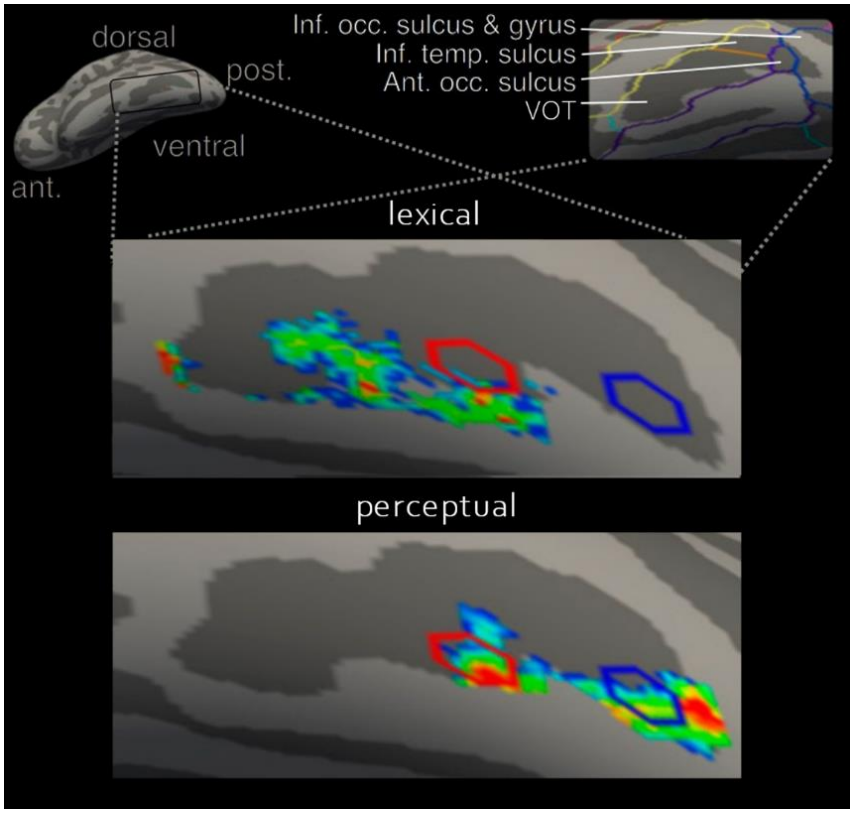


Figure 18. Localization of wVOT. A) Probabilistic maps for real words greater than pseudowords/consonant strings/false fonts (lexical contrast) and real words greater than checkerboards/scrambled words/phase scrambled words (perceptual contrast) in the occipitotemporal sulcus (OTS) across 66 participants adapted from [142]**. Red and blue hexagons illustrate clusters containing the peaks of the three lexical and perceptual contrasts across participants, respectively. WVOT contains multiple representational spaces for orthographic stimuli, with lower-level visual form analysis more prevalent in posterior aspects and lexical analysis constrained to more anterior areas.**

A key aspect of the debate around the existence of the visual word form system is that reading is a relatively recent phenomenon; written language was invented approximately five thousand years ago, and literacy only became widespread in the past couple hundred years. Therefore, there is no plausible mechanism by which a visual word form system could be evolutionarily/genetically predefined for reading [224]. This is distinct from other forms of visual

specialization, such as face recognition or visual navigation, that also occur in our evolutionary ancestors. Instead, reading expertise must be acquired during development and the circuit used for reading is developed and tuned by this extended learning process. The fact that this system is not predefined for reading by evolution, yet wVOT is located in a relatively consistent place in the brain across people, has led to debate around which neurocomputational constraints cause wVOT to become word sensitive and whether these constraints cause this region to become domain specific for reading or not. Here, we contend that the localization of wVOT is constrained by the confluence of two evolutionarily older pathways that have been adapted to process written language, one for visual to phonological transformations and another for visual to semantic transformations. Understanding the neural circuitry of reading, and wVOT's role in this circuit, strikes at the heart of the question of the balance between nature and nurture in the brain by illustrating how intrinsic functional brain architecture shapes, and is shaped by learning, skill acquisition, and expertise [225].

5.2 Is wVOT specifically tuned for reading?

Several hypotheses posit that wVOT ends up where it does because this area of the brain is specialized to perform computations that are particularly important for reading but are not specific to the domain of orthography. These proposals come in two different types: one which posits that wVOT is specialized for particular visual computations, like extracting particular shape combinations or high contrast information, that are particularly useful for, but not specific to,

processing words [55,226] and another type that posits that wVOT is involved in visual naming, which involves interactions between visual, phonological, and semantic circuits, but is also not specific to words [62].

The first model of non-specific wVOT function suggests that wVOT is specialized for processing specific visual features, like high contrast, complex, groupable shapes, which word-processing relies particularly heavily upon [226]. The consistent cortical location of word-selective patches across humans could be explained by the properties of the visual features themselves, i.e., they are sampled foveally and have a consistent curvilinear visual appearance across exemplars [55], which are important organizing principles across visual categories in ventral temporal cortex (VTC) [9]. Support for this hypothesis was gathered from macaques trained to associate English letters and Arabic numerals with varying levels of juice reward. After extensive training, juvenile macaques demonstrated domain-specific functional activations for the trained but not similar non-trained symbols. Further, the locus of cortical activation was consistent across seven monkeys [55]. The data were interpreted to suggest that wVOT responses in humans arise from extensive experience with stimuli with visual features typical of words, which leads to the development of patches of neurons that selectively process that domain of stimuli.

However, neuroimaging in children has demonstrated that the region which eventually becomes wVOT demonstrates increased connectivity to the broader language network prior to learning how to read (Figure 19) [64,65]. This suggests that the localization of wVOT is not strictly constrained by bottom-up constraints of visual processing, but also top-down connectivity to the broader language system. Existing studies of wVOT development in non-human primates are unable to speak to the importance of these top-down constraints, since macaques were not required to associate letter stimuli with phonological or semantic information, besides varying levels of

juice reward [55]. Additionally, several studies have demonstrated that in humans wVOT is insensitive to various low-level features of printed words including font and case [227,228].

Selective wVOT responses have also been demonstrated while deaf individuals read sign language or fingerspelling [231] and after individuals have been trained to associate artificial orthographies composed of objects represented elsewhere in the brain, such as faces and houses, with phonology and semantics (Figure 20A) [232,233]. These studies argue for a remarkable degree of local plasticity in the visual features wVOT can process, despite the specificity of its role in orthographic processing. Additionally, non-visual orthographic processing has been shown to recruit wVOT while sighted and blind individuals read braille (Figure 20B) [88,234]. Although it is unclear to what degree wVOT activation in blind individuals is a signature of more general language processes beyond orthography [235]. These studies demonstrate that wVOT is not strictly sensitive to the visual features of an orthographic system that it represents.

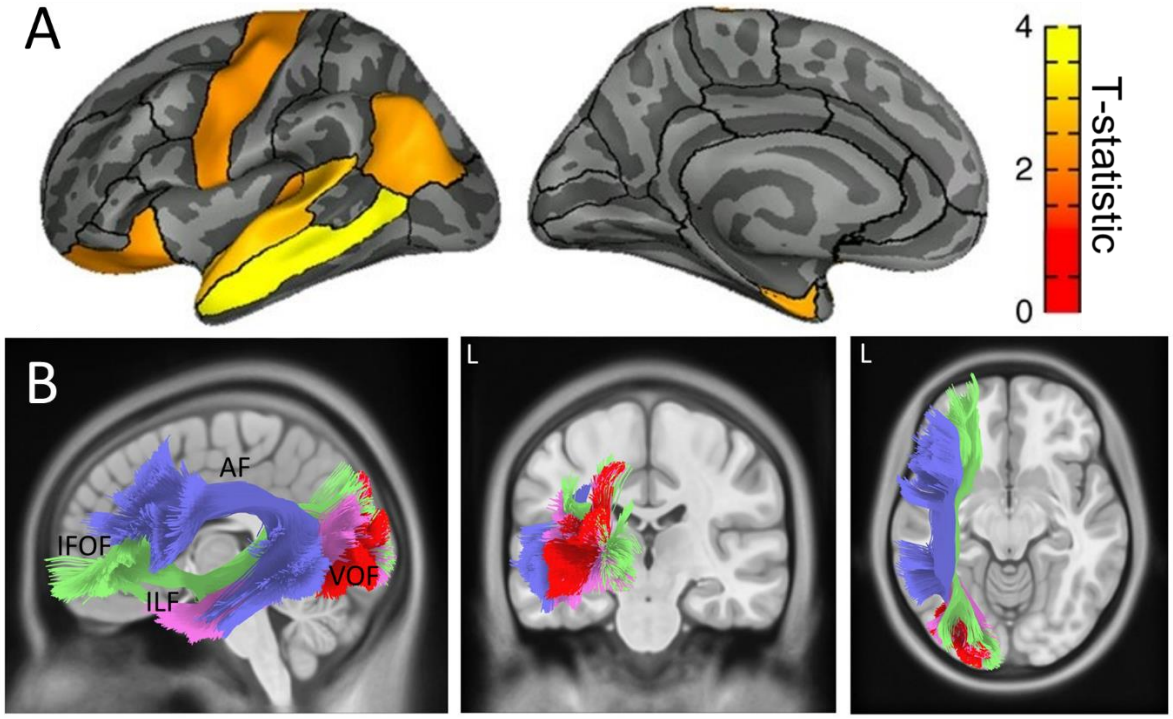


Figure 19. Preexisting connectivity to language regions constrains the localization of wVOT. A) Difference between the strength of diffusion weighted connectivity between the region that would become wVOT and left-lateralized brain regions compared to the corresponding connectivity of adjacent face-selective cortex in preliterate five-year-olds (from [65]**). The region that would become wVOT demonstrated increased connectivity relative to face-selective regions to several left-lateralized language regions prior to the acquisition of literacy. Notably, these regions lie along the prominent white matter tracts examined in panel B. B) Prominent white matter pathways link wVOT to other important language regions in the adult brain. The arcuate fasciculus (AF, blue) links wVOT with frontal and temporal regions important for orthographic-to-phonological conversion. The inferior frontooccipital fasciculus (IFOF, green) and inferior longitudinal fasciculus (ILF, pink) connects wVOT with anterior regions of ventral temporal cortex and frontal lobe, which likely play important roles in the lexical-to-semantic pathway. The vertical occipital fasciculus (VOF, red) connects wVOT with dorsal visual regions, which may play an important role in coordinating eye movements during reading. This illustration was built using DSI-studio** [229] **and the included ICBM 152 Atlas** [230]**.**

Another characterization of wVOT's role in reading highlights the relationship between visual word and face processing networks [28,60]. Although face and word stimuli are very different in terms of their visual appearance, their importance to our evolutionary ancestors, and feed into distinct brain networks supporting either social interactions or language, their cortical localizations are remarkably close together [94]. Some argue that this phenomenon is caused by both systems requiring input from regions with foveal receptive fields, which leads to competition between neurons that become selective for orthographic vs. facial stimuli [27]. Throughout the acquisition of literacy, this competition leads to a graded asymmetry in face and word perception, as wVOT neurons edge-out face-selective representations in the left hemisphere, due to its closer proximity with other left lateralized language areas [29]. After the acquisition of literacy, proponents of this account argue for roles of both the left and right VTC in face and word-processing, supported by graded deficits in both domains after lateralized lesions to VTC [60].

However, neuroimaging results in children beg the question, to what degree are connections between wVOT and the broader language network established purely as a result of visual experience with words or do they exist before children learn to read [64,65]. In particular, face-selective VOT in preliterate children has significantly less structural connectivity to lateral temporal and frontal language regions compared to the parts of VOT that become word-selective upon the acquisition of literacy [64,65]. Additionally, recent intracranial and stimulation evidence in literate adults demonstrate the insensitivity of neural responses recorded from wVOT to face stimuli, thus challenging the claim of graded specificity of wVOT [19,24,100]. In these studies, word-selective responses recorded from some left hemisphere regions demonstrated no degree of face selectivity [100] and cortical stimulation and lesioning of wVOT selectively impaired reading but not famous face recognition (Figure 20C & D) [19,24].

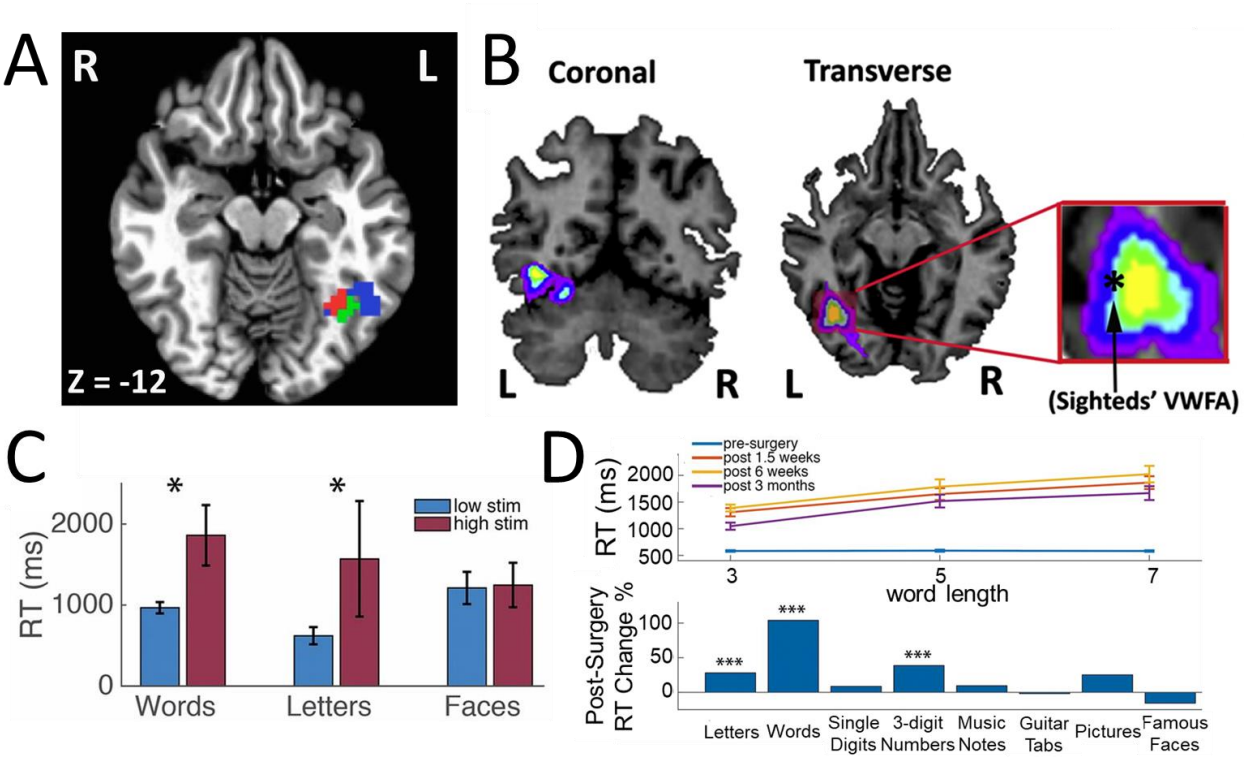


Figure 20. Functional properties of wVOT. A) Participants learning to read artificial orthographies composed of houses demonstrated significant changes in neural activation relative to untrained Korean orthography (blue) in areas overlapping (red) with wVOT (green) (adapted from [232]**). B) Parametric map of braille reading over nonsense braille reading in 8 blind individuals (from** [88]**). Peak BOLD activations in blind individuals reading braille was highly consistent with peak activations in sighted individuals performing an orthographic task (isolating words with descenders—letters extending below written baseline [g,j,p,q,y]) both at the group level and across individuals. C) Electrical brain stimulation through electrode contacts implanted in wVOT profoundly delayed word and letter naming but not famous face naming (adapted from** [19]**). D) Focal resection including wVOT caused profound and persistent alexia-like impairments in word naming which is exacerbated by word-length (adapted from** [19]**). Response time impairments were greatest for words, but also extended to letters and numbers but not faces, musical notation, or pictures of other objects.**

Regarding the proposal of graded hemispheric specialization for words, word-selective responses invasively recorded in the right hemisphere did not onset until approximately 400 ms after word presentation, whereas left hemisphere word-selective responses onset at approximately 200 ms (Figure 21C) [100]. These later, right hemisphere responses may be involved in integrating words into context, reflected in the N400 event-related potential component recorded using non-invasive EEG [154], but not necessarily the bottom-up processing of word forms [165]. Right hemisphere word sensitive regions have also been shown to be more sensitive to differences in font and case, unlike left hemisphere wVOT responses [228]. The results discussed here demonstrate that there are regions of VTC selective to words and not faces, and that the right and left hemisphere have qualitative, rather than graded, differences in their contributions to reading based on the differences in their responses and processing dynamics. While these results speak against truly joint circuits for face and word processing [59,60], we later discuss how the close proximity and analogous computations required to process faces and words provide clues as to the function and localization of reading pathways.

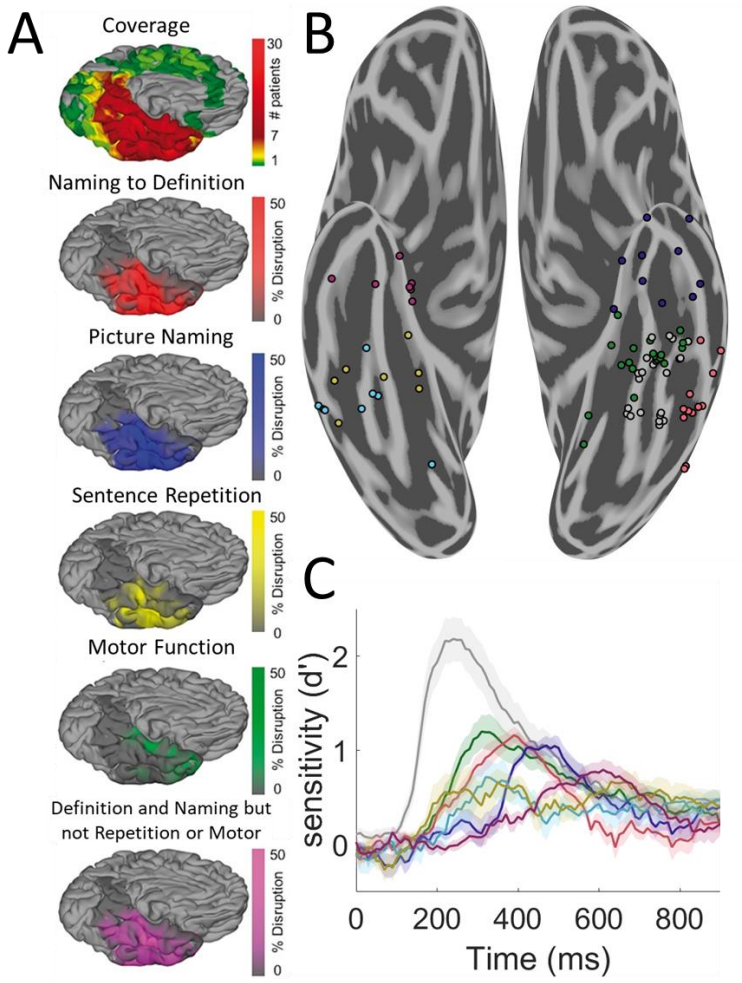


Figure 21. The dynamics and function of the basal temporal language system. A) Electrical brain stimulation across VTC of 70 patients elicited various semantic, phonological, and motor language deficits. The bottom panel shows regions that, when stimulated, evoked deficits in the ability to define objects based on their names and name objects in pictures without corresponding deficits to sentence repetition or motor function (adapted from [21]**). Large swaths of VTC, the basal temporal language system, participates in lexicosemantic processing. B) Ventral surface of inflated brain. Dots represent word-selective intracranial electrodes across 36 patients (taken from** [100]**). These contacts were clustered based on their spatial location and the time-course of their word-selective responses. Color represents the spatiotemporal cluster they belong to, the average time-course of which are illustrated in panel C. Word-selective contacts are found throughout VTC. C) Average time-course of word-selective responses recorded from intracranial contacts in panel B. Word-selective responses are strongest and earliest along the left posterior fusiform gyrus, before spreading**

medially and laterally from there. Finally, anterior word-selective responses peak around 400 ms after word presentation. Right hemisphere word-selective VOT responses did not peak until approximately 400 ms.

Finally, extending Wernicke's perspective of the importance of phonology in reading to the visual word form debate, the "interactive account" posits that wVOT does not contain visual word form representations, but instead contains "general purpose analyzers of visual forms" [62]. These non-specific, bottom-up feature detectors derive their observed selectivity for words through top-down interactions with brain regions that contain phonological and semantic associations that are established during the acquisition of language [62,170]. Proponents of this theory point to evidence that wVOT is active during various tasks that do not involve reading, like picture naming, auditory language tasks, or even thinking about the manipulability of objects [62,170]. The theory can also explain the consistent localization of wVOT across individuals due to its evolutionarily preserved role of linking visual inputs to the broader language system [62], and why prereading children already demonstrate preferential connectivity between wVOT and the broader language network [64,65].

However, disrupting wVOT activity through lesions or precise cortical stimulation can lead to selective deficits in reading but not picture naming (Figure 20C & D) [19,22,24,25]. Further, intracranial recordings collected from wVOT show selective responses to words but not to other nameable objects [19], making it implausible that wVOT responses represent non-specific interactions between bottom-up processing of complex visual forms and top-down contributions of semantics and phonology.

5.3 Preexisting anatomical pathways constrain wVOT's localization

Given that wVOT seems to contain domain-specific orthographic representations, the question then remains: why does wVOT end up in a similar location consistently from person-to-person? As previously mentioned, recent studies suggest that the area that becomes wVOT has preferential neural connectivity to the broader reading and language network even before a child learns to read (Figure 19) [64,65]. These studies used fMRI to show that in pre-reading children there was no region in VTC that responded preferentially to words versus other visual categories. However, when these same children were scanned after skilled reading had been achieved, there was an area in VTC that became selective for words. Looking at this region using diffusion weighted imaging of its structural connectivity during the first scan session, when the children were still pre-readers, it had preferential anatomical connectivity to the broader reading and language network before it became word-selective [65]. Furthermore, fMRI scans in young infants show that the region that is likely to become wVOT, based on the average location of this region in adults, showed functional connectivity to the broader reading and language network even at this young age [64]. These studies lend strong support to the hypothesis that the location of wVOT is consistent across people because this part of VTC has preexisting connections to the broader reading and language network.

Although studies examining connectivity during development provide important clues as to why wVOT ends up where it does in the brain, they have not addressed the question of why connectivity exists between this region and the broader reading and language networks prior to learning to read. Clues to answering this question may come from understanding parallels in the computations involved in reading versus evolutionarily older visual processes involved in language and communication.

Models of reading posit multiple processing routes linking orthography, phonology, and semantics: at least one pathway involves a rule based orthographic-to-phonological mapping (orthophonological pathway), which is in turn linked to a word's meaning, and another pathway that relies more heavily on a direct link from orthography to the name and/or meaning of the word (lexicosemantic pathway) [236]. Support for the existence of the orthophonological pathway includes the fact that we can pronounce non-words that follow linguistic rules (e.g. lerm) but that we have not experienced before, which presumably can be achieved through the rule-based orthographic-to-phonological mapping [236]. Disruption to these rule-based phonological mappings is thought to underly reading disorders like phonological dyslexia [237]. On the other hand, there are words we can pronounce that do not follow these rules (e.g., yacht). These exception words necessitate the existence of the lexicosemantic pathway, which is selectively impaired in patients with surface dyslexia. One feature of this disorder is "regularization" of exception words (pronouncing "yacht" as "ya-ch-it") [237], presumably because of dysfunctional lexicosemantic mappings while the orthophonological mapping remains intact. "Dual route" models treat these as two distinct paths (an indirect nonlexical orthophonological pathway and a direct lexical pathway) [238,239] whereas "Connectionist" models treat these as an interconnected set of paths between critical orthographic, phonological, and semantic processing systems [240,241]. The intricacies of the debate surrounding these competing models of reading are outside of the scope of this review, however, we argue that wVOT's response properties and anatomical location make it a prime candidate for containing the orthographic representations associated with either model.

Both the orthophonological and lexicosemantic mappings have parallels in evolutionarily older visual processes. Specifically, a system that maps visual forms to their corresponding sounds

is not only core to orthographic reading, but also to mapping the visual forms of lips, tongue, and teeth to their corresponding sounds during lip/speech reading [242]. Furthermore, a direct, non-systematic/arbitrary mapping from visual forms to words is seen not only for exception word reading, but also for object, place, and person naming [243]. Notably, these properties can also explain why face and word-selective areas are remarkably close together in VOT since speech reading involves extracting representations of different facial features and both orthographic and face processing benefit from exemplar-level recognition and naming. Below we detail a hypothesis that learning to read is constrained by preexisting neuroanatomical circuits for visual to phonological and visual to semantic transformations, with wVOT being a key hub with remarkable local plasticity in the visual forms it can learn to associate with these transformations, and present evidence for co-occurring neurocognitive deficits in reading and related visual processes when these networks are disrupted.

5.3.1 The anatomy of the orthographic-to-phonological pathway

Substantial evidence points to connections between wVOT and lateral temporal and frontal regions involved in phonological processing as the key pathway for orthophonological mapping [244–246], likely via the posterior arcuate fasciculus that has terminations in anterior wVOT [142,247]. Specifically, functional and anatomic connectivity studies implicate these regions in orthophonological mapping during reading [245,246,248], especially during pseudoword reading [249–251], which can only be achieved through orthophonological mapping. Furthermore, aberrant activity in these regions and abnormal connectivity in this pathway are associated with developmental and phonological dyslexia [245,252].

A few studies implicate a putatively similar pathway for speech reading [253,254], though evidence for colocalization of speech reading and word reading pathways within subjects is lacking. In particular, regions of the superior temporal sulcus that show connectivity to wVOT prior to learning to read (Figure 19) are similar to the regions strongly implicated in vocal audio-visual integration and the McGurk effect [254]. Studies have shown correlations between speech reading and word reading deficits, with some evidence implicating wVOT and the orthophonological pathway. Specifically, individuals with developmental and phonological dyslexia also have deficits in speech reading [255,256] and do not benefit as much as non-dyslexic individuals from viewing speakers while listening to their speech in noise [257]. One caveat being that these are correlational studies and therefore it is not certain that the relationship between them is due to dysfunction in a shared pathway. For example, poor phonological representations could impact both speech reading and word reading ability without the two sharing a common pathway. Additionally, neuropsychological studies provide evidence that a single lesion can cause both speech reading and word reading deficits [242,258]. In one of these studies, two patients were studied who had strokes in an artery that caused damage to VTC, though one had their stroke in the left hemisphere and the other in the right hemisphere. The patient with right hemisphere damage was prosopagnosic but had intact speech reading ability. The patient with left hemisphere damage was alexic and had substantial deficits in speech reading ability as well [242]. These results provide evidence for similar pathways for orthophonological mapping and speech reading, though it is unclear how this pathway develops when children learn to read. Specifically, it is unknown if the pathways responsible for speech reading and word reading are overlapping or if this pathway subdivides while individuals learn to read, leaving parallel but non-overlapping pathways for these skills. Functional and anatomical mapping studies of speech reading and word reading within the

same subjects, particularly in children learning to read, are required to assess whether these processes use joint or non-overlapping, but parallel pathways.

5.3.2 The anatomy of the lexical-to-semantic pathway

Some studies have implicated a pathway connecting wVOT to tempero-parietal and frontal regions in lexicosemantic processing [259,260], with particular emphasis on the angular gyrus [261,262]. However, several neurological, intracranial electrophysiology, and neuroimaging studies suggest that another pathway through left medial and anterior VTC via the inferior longitudinal fasciculus, that parallels the object naming pathway, is also a key part of the lexicosemantic pathway [21,259,263,264]. The neurological literature has long appreciated that stimulation of many places along VTC can cause speech and language dysfunction and has termed this area the "basal temporal language area." This "area" has been defined as spanning from 1 to 9 cm from the tip of the temporal lobe [259] and therefore likely consists of multiple areas along a basal temporal language system (BTLS), potentially including wVOT. Indeed, recent studies using intracranial recordings and stimulation have highlighted that BTLS consists of multiple areas contributing to multiple, temporally extended aspects of reading and language [21,100,265] (Figure 21). Furthermore, an area that shows connectivity to what becomes wVOT prior to learning to read is in anterior BTLS (Figure 19), suggesting that the connectivity between BTLS may help constrain the location of wVOT once reading expertise is acquired.

Despite the BTLS having been described using intracranial recordings and stimulation for at least 30 years [21,100,143,147,259,260,265], most models of the neural underpinnings of the reading pathways have not considered VTC regions medial and anterior to wVOT. Many of these models are built on meta-analyses of fMRI studies and may neglect BTLS due to susceptibility

artifacts in anterior temporal cortex [160] or differential sensitivity of the fMRI signal to different aspects of the electrophysiological response [155]. Additionally these other BTLS patches may be smaller and somewhat less word-selective than wVOT [100], or involved in reading processes, like semantic encoding, that not all imaging contrasts are sensitive to [21,266]. Indeed, a few studies of reading that used fMRI techniques designed to increase signal-to-noise ratio in VTC have observed the extended BTLS [100,249].

Evidence suggests that the medial-to-anterior BTLS plays a critical role in the lexicosemantic pathway. Neuroimaging studies that have used scan parameters associated with high SNR in VTC and anterior temporal cortex and intracranial recording studies have suggested that BTLS is involved in exception word processing [249,267,268]. Damage to BTLS and the inferior longitudinal fasciculus is associated with surface dyslexia [269–271]. Additionally, transcranial magnetic stimulation of anterior BTLS causes regularization errors of exception words [264] (e.g. temporary surface dyslexia) and a consistent early symptom of anterior temporal lobe degeneration in primary progressive aphasia (also called “semantic dementia”) is surface dyslexia [272]. Taken together, these studies strongly suggest that BTLS, particularly anterior BTLS, is a key part of the lexicosemantic pathway, along with temporoparietal cortex near or in angular gyrus [261,262] and inferior frontal cortex regions near or in pars triangularis [262,273], which are also associated with surface dyslexia and exception word reading.

It is well established that left VTC also plays a critical role in object and proper naming [274] and that objects, people, and places, like exception words, cannot be named using rule-based processes [243]. Thus, the lexicosemantic pathway may be parallel to, and/or overlapping with, the broader visual naming pathway. Much as developmental dyslexia and lip-reading deficits co-occur [255,256], surface dyslexia co-occurs with anomia in early stage primary progressive

aphasia [275], in developmental anomia [276], and in acquired anomia [277]. Furthermore, unlike stimulation of wVOT, which impairs reading, but not visual naming [19,22], disruptive stimulation of the rest of the BTLS, particularly anterior BTLS, impairs both reading and visual naming [21,264,278]. Additionally, surgical resection of the left anterior temporal lobes, including parts of the BTLS, can cause visual anomia with no auditory naming deficit [263,279], and there has been at least one report of surface dyslexia after this region was resected (data not yet published). One hypothesis from this model would be that those patients with the greatest visual anomia also display surface dyslexia. A key open question is whether exception word reading through anterior BTLS more closely resembles object naming per se, or specifically proper naming, because regions of anterior BTLS are associated with proper naming more so than general object naming [123] and these types of naming deficits may be dissociable [280]. A shared reliance on pathways specialized for proper naming could be another factor that drives the close proximity of face and word-selective regions in VOT.

The position of wVOT in these broader reading networks may be inferred both by its connectivity and its activity relative to the rest of the BTLS and VTC. WVOT lies at an opportune place along VTC, with access to the arcuate fasciculus, inferior frontal occipital fasciculus, vertical occipital fasciculus, and the inferior longitudinal fasciculus [247,281]. Additionally, results from intracranial recordings of VTC shows that wVOT has the strongest (by approximately a factor of 2) and earliest word selectivity in VTC, followed by lateral and medial VTC regions about 100 ms later, and anterior VTC ~150 ms after that (Figure 21C) [100]. This strong and early selectivity suggests that wVOT is a key hub for orthographic processing from which processing interactively proceeds medially, laterally, and anteriorly down the broader reading pathways. Indeed, the frank reading deficits that arise from lesions or stimulation to wVOT [19,22,25] supports the idea that it

is a key hub for reading. One open question is the role of posterior wVOT connectivity to dorsal visual processing regions via the vertical occipital fasciculus [281]. These regions have been associated with attentional deployment during reading [282] and therefore connections between wVOT and these areas may be key to saccade programming during natural reading. Regardless, the results reviewed here suggest wVOT is a key node of orthographic processing that is involved in multiple stages of the extended reading process through recurrent interactions with the broader language system.

5.4 Conclusion

Recent findings support a model in which wVOT is a key hub for orthographic processing that receives output from lower-level visual processing circuits and is a waypoint for reading to leverage preexisting higher-level visual processing circuitry (Figure 22). Specifically, the wVOT-lateral temporal and frontal cortex pathway underlying rule based, orthographic-to-phonological mappings for reading leverages an existing pathway for rule-based mapping between visual representation of lip, tongue, and teeth forms to phonology. Additionally, the wVOT-BTLS, lateral temporal, and frontal cortex pathway underlying the orthographic-to-lexicosemantic mapping leverages an existing pathway for visual naming. The degree to which these orthophonological and lexicosemantic pathways are innate or shaped by experience is an open question.

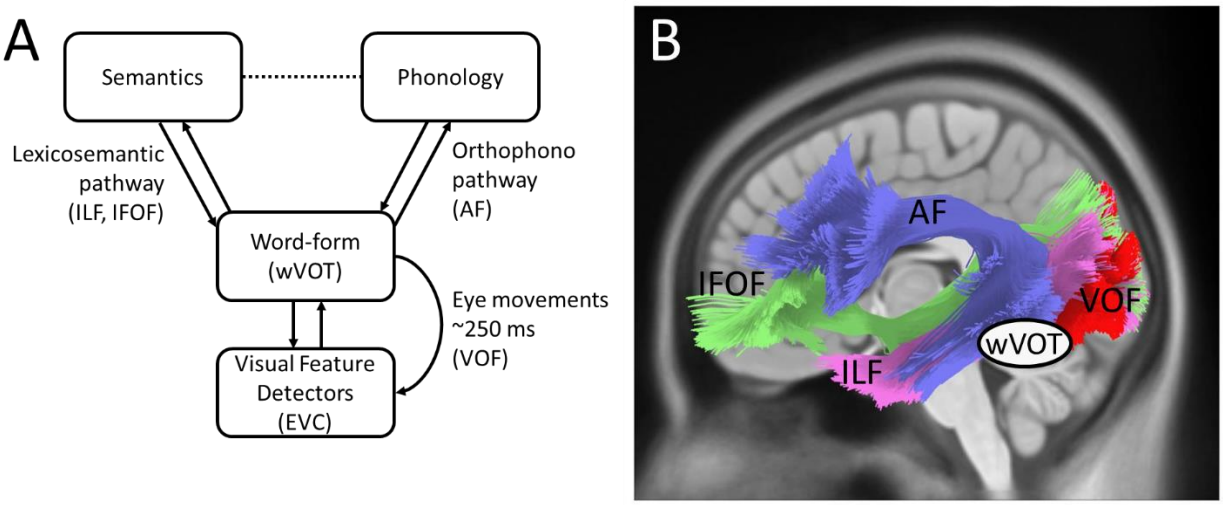


Figure 22. WVOT's position in the language system. A) Low-level visual feature detectors in early visual cortex interact with wVOT to form visual representations of words. These in turn feed into and interact with higher order centers for semantics and phonology. Potential interactions between phonological and semantic centers (dashed line) are beyond the scope of this review. Interactions between visual, semantic, and phonological systems can lead to differing amounts of ambiguity in word-form representations, which may be used to program later information gathering eye-movements. B) WVOT lies at a critical junction in visual, semantic, and phonological processing circuits in the brain. White matter pathways, which are prevalent before children learn to read, may help explain why wVOT shares a consistent localization across individuals, despite widespread literacy being an evolutionarily recent phenomenon. Abbreviations: AF-arcuate fasciculus, EVC-early visual cortex, IFOF-inferior frontooccipital fasciculus, ILF-inferior longitudinal fasciculus, VOF-vertical occipital fasciculus.

The connectivity in this model is fully bidirectional. This is supported both by neuroanatomy, where studies have shown very few cortical connections are unidirectional [34], and by functional studies suggesting recurrence. Specifically, the finding that the representation in wVOT shifts from being coarse to individuated over the period of hundreds of milliseconds [19] suggests that the mutual constraints applied by orthography, phonology, and lexicosemantics filter through the entire BTLS over time. This recurrent computation is in contrast with traditional feed-

forward models of visual processing, wherein areas complete their local computation and pass the information on to the next brain region and do not contribute to further visual processing [125]; a model that is not well supported by the anatomical connectivity of the visual stream [34]. Rather, the evidence presented here suggests that many parts of the network contribute to multiple stages of word recognition over time. Bidirectional connectivity also provides a pathway for top-down constraints to be applied to the reading process, for example by sentence context [283], and provides a pathway to explain how braille reading activates wVOT in the blind without visual input [88].

This model emphasizes the role of connectivity and the compatibility of the types of mapping between representations needed for reading, and de-emphasizes local computational compatibility, in constraining the location of wVOT. That said, local computational compatibility and bottom-up constraints likely also play a role in where wVOT ends up. This area of VTC is preferentially responsive to foveal parts of the visual field [43] and seems to be involved in holistic and configural processing [19,30]. This is not only true for words, but also for faces in neighboring face-selective regions [284], and other neighboring regions, which develop holistic and configural processing characteristics for stimuli of acquired visual expertise [30]. However, local computational compatibility lacks explanatory power for why braille reading in the blind, sign language in the deaf, or faces and houses in individuals trained to read artificial orthographies preferentially activate wVOT [88,231]. WVOT's positioning at the intersection of bidirectional projections via the inferior longitudinal fasciculus, inferior frontooccipital fasciculus, and arcuate fasciculus can explain these results. Furthermore, local computational compatibility cannot explain the co-occurrence of disorders such as developmental dyslexia and speech reading deficits [255,256], nor the co-occurrence of anomia and surface dyslexia [277]. This model, which

emphasizes the compatibility of representational transformations, rather than local visual computations, can explain these co-occurrences. A key open question is: what is the balance between local computational compatibility and the compatibility of representational transformations and connectivity in determining the location of wVOT?

More generally, reading may represent a special case of a broad principle for how acquiring expert skills leverages existing brain circuitry. Specifically, acquiring expertise involves adapting existing neural processing pathways to perform computationally similar transformations [285], whereas representations within brain regions have a remarkable degree of plasticity during skill acquisition [232,233,286]. This model bears some resemblance to the cultural recycling hypothesis [224], which hypothesizes that wVOT occupies a similar region across people due to that region's position in a circuit whose original function is sufficiently close to that required by reading. However, the cultural recycling hypothesis necessitates that those local cortical computations, "local combination detectors" in wVOT, are compatible with the skill being learned. The hypothesis presented here instead suggests that existing processing circuits within regions are less critical in defining function after acquiring expertise. For example, wVOT demonstrates remarkable local plasticity in adapting its computations for facial features important for speech reading to visual word forms. This model also bears some resemblance to theories of cortical specialization that emphasize the role of connectivity between areas [63,287], though these theories have not necessarily specified critical transformations between aspects of the neural representation or particular processes associated with acquired skills. The model for reading presented here suggests a merging of the recycling and connectivity hypotheses for expert skill acquisition, demonstrating a potential principle for how nurture creates neural changes that leverage the intrinsic nature of brain circuits.

6.0 Conclusion and future directions

Together, the works presented in this dissertation help illustrate ways in which bottom-up and top-down interactions jointly influence VTC structure and function. In Chapter 2, intracranial data across a large patient population helped identify functional principles organizing VTC, including gradients in local prestimulus dynamics, information processing dynamics, and long-range functional connectivity. In these data, the onset latency of category-selective information increased across VTC layers, consistent with fast feedforward models of the ventral stream [40,42]. However, there were also increases in the duration neural populations took to form their initial category-selective representations and increases in the length of time neural populations maintained those representations when moving along VTC. Also along this axis, there were systematic increases in connectivity to regions that were not visually responsive. Together, these results suggests that top-down interactions differentially impact hierarchical layers of VTC.

Next, we focused our attention on word- and face-selective neural populations in VTC. We identified several populations that responded preferentially to only words or faces, but not the other, arguing for, at least, partial separability of these networks. Word-selective regions were found medial, lateral, and anterior to face-selective regions on the fusiform gyrus, highlighting the existence of a large basal temporal language system [22,23,259,260], which we replicated in healthy participants using fMRI. These word-selective regions demonstrated distinct dynamics from one another, with strong early responses along the fusiform gyrus and later responses in medial, lateral, and anterior regions. These distinct time-courses suggest that these regions play different roles in reading.

Chapter 4 presented a case-study for how long-range interactions can influence extended processing dynamics and local representations in VTC using a key node in visual word processing, wVOT. In this study, we demonstrated that within the wVOT of healthy individuals, there is a shift from coarse to fine representations over the first 500 ms of word processing, similar to those originally identified using intracranial recordings in epilepsy patients [19]. This shift in representations was strongest for real word stimuli compared to other word-like stimuli including pseudowords, consonant-strings, and false fonts, which lack the semantic and/or phonological associations of real words. Further, the sharpening of real word representations was accompanied by increased connectivity between the wVOT and early visual cortex and anterior temporal lobe regions around the same time window, suggesting that bottom-up and top-down mechanisms help disambiguate representations in this region.

Finally, in Chapter 5 we presented a model for the development of wVOT in which the localization of this region is constrained based on long-range bidirectional white matter pathways underlying computational transformations necessary for reading. Specifically, the model suggests that wVOT is consistently localized to the intersection of several important white-matter pathways, including the arcuate fasciculus which facilitates visual-phonological transformations originally used for speech reading, inferior longitudinal and frontooccipital fasciculi which facilitate visual-semantic transformations originally used for object naming, and bottom-up pathways organized by the retinotopic projections of early visual cortical areas. This model helps explain why wVOT demonstrates selectivity for printed words in the visual domain, but also responds via top-down activation of non-visual stimuli like braille in the congenitally blind [88,89], and how reading can occupy similar circuits across individuals despite a lack of direct evolutionary constraints [144,169,185]. Additionally, the proximity of word and face processing networks in VTC can be

explained by wVOT's original function in speech reading and exemplar level naming, which requires visual and naming circuits similar to those involved in face recognition [94,243].

This model is in contrast to models of VTC organization that suggest bottom-up constraints imposed by topographical maps alone is sufficient for establishing a protoarchitecture of category-selective regions in VTC [44,48,55–58,288]. Additionally, the model argues for a tremendous amount of plasticity in local cortical representations, e.g. the appropriation of speech reading cortex to process word stimuli. This strong degree of local plasticity is supported by studies demonstrating that participants trained to associate artificial orthographies of faces and houses with semantic information demonstrated wVOT responses to faces and houses [232,233]. The congruence of high-order cortical transformations (orthographic to phonological or lexical to semantic in the case of written words) are more important in constraining the localization of acquired expertise compared to the specific visual features of the input stimulus.

The model may represent a general principle of how visual expertise develops and is constrained by adapting existing, computationally compatible information transformations that occur via the connections between brain regions. For example, in face processing, innate connectivity between VTC and higher order regions like medial prefrontal cortex may help orient infants to high-valence social interactions associated with faces [94,289,290]. Connectivity between these regions not only facilitates more experience with faces, and thus more bottom-up tuning of face-selective VOT, but also constrains the localization of face-selective VTC areas to those with direct cortical projections to other important face processing regions [50,61], in addition to the bottom-up constraint imposed by foveal receptive fields [43,49,51,61]. Similar principles may also apply to the development of tool-selective cortex, biased by bidirectional projections with sensorimotor regions associated with tool utility [63,291] and to the development of place-

selective cortex, biased by connectivity to dorsal spatial navigation networks [50,292,293]. Future studies examining the development of face patches in animals exposed to faces in the absense of meaningful social interactions, or animals trained to perform face/social computations on non-face stimuli may help further elucidate how top-down and bottom-up processes precisely constrain VTC organization [61], in addition to studies examining the specific constraints imposed on processing circuits for other categories of objects.

The results described here help identify ways in which we can extend current hierarchical models of ventral visual function. Future models should attempt to capture the diversity of dynamics and connectivity patterns exhibited across layers of the ventral visual hierarchy. Higher layers should demonstrate increasingly deliberative processing and increasing connectivity to units with non-visual response properties, like viewpoint invariance [121] or semantic similarity [124]. In fact, recursive connectivity in deep neural network models has been shown to better explain fMRI responses from anterior VTC regions compared to purely feedforward models [294]. Future models should also attempt to capture the dynamic shifts in representation exhibited within VTC regions, like the shifts from coarse to fine representations in face- and word-selective VOT [19,66–68]. In the future, we can combine insights derived from these computational models with studies investigating the effects of electrical or optogenetic stimulation of VTC layers to elucidate the specific contributions of bottom-up and top-down interactions on local information processing in the ventral visual stream.

These studies also highlight interesting differences in single trial potentials (stP) and high frequency broadband (stHFBB) activity recorded from the local field potentials of neural populations. For example, prestimulus neural dynamics in these two signal components were not not well correlated with one another across VTC and these components demonstrated spatial and

temporal differences in their word- and face-selective responses. Category-selective regions that demonstrated selectivity in both stP and stHFBB activity were those that are typically observed using fMRI [119,155,295], which has been shown to be differentially sensitive to different time periods and frequency components of the iEEG signal [155]. This is an important limitation to consider when interpreting maps of category-selectivity derived from fMRI [128]. StP and stHFBB are thought to be generated by different physiological generators, with stHFBB itself having multiple distinct components, including an early feedforward component and a later feedback component [120]. A deeper understanding of how these signal components are generated would enable a deeper understanding of the local computations that facilitate visual processing within category-selective regions.

More generally, the physiological properties of neural circuits that process category-selective information (i.e., their laminar organization, concentration of interneurons, and morphological properties of the neurons themselves) are also not well understood. Changes in cytoarchitecture, myelination, and neural growth factors are observed along hierarchical networks across the brain, which may influence long range functional connectivity patterns and local neural dynamics [80,95,96]. Furthermore, category-selective regions in VTC have been shown to have differences in their cytoarchitecture compared to one another [9,296–298]. Future studies are necessary to understand the cytoarchitectural properties that predispose regions of cortex to be specialized for specific visual computations and the degree to which local plasticity of VTC regions is constrained by the physiology of these local neural circuits.

Finally, much more work is necessary to understand how ventral visual cortex interacts with circuits involved in coordinating eye movements. The process of object recognition is accompanied by highly stereotypical patterns of eye movements which vary across object

categories [299–305]. Regions in parietal and frontal cortices have traditionally been implicated in guiding eye movements and allocating visual attention [282,306,307]. However, temporal cortex regions have recently been shown to be sensitive to various aspects of visual sampling behavior. For example, neurons in entorhinal cortex demonstrate grid-cell like firing patterns coding the position macaques are fixating in an image [308], BOLD activation in category-selective VTC has been shown to be sensitive to where objects are likely to be located in natural scenes [309], and face- and house-selective regions are activated when people mimic stereotypical eye movement patterns used to recognize these objects even when the objects themselves are not shown [305]. Further, damage to VTC resulting in agnosia also leads to less efficient eye movement patterns while searching for objects in natural scenes [310]. Better understanding the links between eye movements and object recognition may greatly increase our understanding of how top-down attentional effects interact with bottom-up visual information accumulation to achieve object recognition.

In summary, the work presented here suggests that non-feedforward interactions play critical roles in constraining the organization and extended information processing dynamics in VTC. Several open questions remain regarding the balance between top-down and bottom-up contributions to visual processing. But we have identified several specific ways in which we can improve the fast feedforward model of the ventral visual hierarchy.

Appendix A Supplementary materials for Chapter 2

Appendix A.1 Supplementary text

Appendix A.1.1 Interhemispheric differences in ventral stream information processing dynamics.

Given asymmetries in anatomical organization and response tuning across hemispheres [311], we sought to determine if there were differences in information processing or functional connectivity gradients in left versus right VTC. We found 263 electrode contacts with significant category-discriminant information in left VTC and 127 in the right. The disparity in left and right hemisphere contacts is proportional to the different proportion of total contacts implanted in the left versus right VTC across patients (1,258 in left, 698 in right). Using these category-discriminant electrode contacts, we ran an analysis of covariance to determine if there was an interaction between anatomical position and hemisphere when predicting information processing dynamics or functional connectedness. There was a significant interaction between cortical distance along the visual hierarchy and hemisphere when predicting onset latency, indicating a more dramatic gradient in the right compared to left hemisphere ($F(1) = 16.19$, $p < .001$), and connectivity to regions that were not visually responsive, indicating a more dramatic gradient in the left compared to right hemisphere ($F(1) = 11.76$, $p < .001$). There were no significant interactions between hemisphere and distance along the visual hierarchy in predicting peak information ($F(1) = 1.25$, $p = .26$), average PLV to visually responsive regions ($F(1) = 0.44$, $p = .51$), or duration of initial rise in information ($F(1) = 0.91$, $p = .34$). The interaction between hemisphere and distance along the

visual hierarchy was trending when predicting information maintenance duration ($F(1) = 4.49$, $p = .035$, uncorrected). This suggests that there were differences in anatomical gradients of information onset and connectivity with regions that were not visually responsive across hemispheres in our sampled neural populations.

Appendix A.1.2 Differences in functional anatomical gradients for different object categories

It has previously been suggested that VTC circuits responsible for processing different categories of objects exhibit different processing dynamics related to the rate at which those objects are encountered during natural vision [312]. To investigate differences in information dynamics between neural populations tuned primarily to single object categories, we identified category-discriminant VTC electrode contacts that demonstrated maximum evoked responses to the object category which also had the most information (see Methods). This procedure revealed that 246 of the 390 category-discriminant VTC neural populations were predominantly selective for a single object category: 66 were selective for faces, 50 for words, 47 for houses, 31 for phase-scrambled objects, 21 for bodies, 20 for hammers, 7 for tools, 3 for pseudowords, and 1 for consonant-strings (Figure 25). Next, we fit linear multiple regression models to explain information processing dynamics of the neural populations that were selective to the object categories that most patients saw (bodies, faces, words, hammers, houses, and phase-scrambled objects) as a function of the neural population's distance from the occipital pole and the category it was selective for (see Methods). This allowed us to compare the functional anatomical gradients specific to different categories of objects in VTC.

The main effects revealed by this procedure are contained in Table 4 and Figure 25. The information onset of face-selective neural populations increased at a rate of 11 ms per cm traveled along VTC. Neural populations selective for all other categories demonstrated faster increases in onset latency along this axis compared to face-selective neural populations. Peak face-selective information was not significantly different when moving along the ventral visual hierarchy. However, peak information for the other observed categories decreased faster along this axis. There was no significant change in the duration of the initial rise of face-selective information when moving up the visual hierarchy, nor was the gradient for any category significantly different from faces. There was also no significant change in the duration that face-selective information was maintained in neural populations further up the visual hierarchy, but information was maintained for shorter durations moving along this axis for all other object categories.

Functional connectivity to visually responsive regions did not significantly decrease along the ventral visual hierarchy when looking at face-selective information. However, visual connectivity decreased more quickly when moving along VTC in word-, hammer-, and house-selective populations compared to face-selective populations. Finally, there was not a significant gradient in the functional connectivity of face-selective populations to regions that were not visually responsive along the visual hierarchy, nor were there any significant differences observed for the other object categories.

In summary, at the level of individual categories, face-selective neural populations demonstrated faster onsets of information processing, larger peaks, longer durations, and increased connectivity to visually responsive neural populations compared to the other object categories. These differences in information processing dynamics across category-selective neural populations is consistent with previous fMRI studies demonstrating different preferential rates of

stimulus presentation for different categories of objects [312]. Together, these results suggest differences in information processing dynamics for different categories of objects, which may be related to the functional interactions that facilitate their recognition or how these stimuli are naturally encountered in the real world [312].

Appendix A.1.3 Comparing neuroanatomical gradients exhibited in single trial potentials versus high frequency broadband activity

Previous studies have identified differences in the information contained within single trial potentials (stP) and single trial high frequency broadband activity (stHFBB) [100,119] and others have suggested that these components of have different physiological generators [120,163]. To investigate the degree to which neuroanatomical gradients in neural dynamics and long-range functional interactions were consistent across these signal components, we re-ran the main analyses of this study using stP and stHFBB activity separately. We isolated 380 electrode contacts that demonstrated above-chance category-discriminant activity ($p < .001$, permutation test) in their stP and 150 contacts that were discriminant in their stHFBB activity (Figure 26).

In contacts selective in their stP, gradients in the onset, duration of initial rise, peak magnitude, and maintenance duration of category-selective information were consistent with those identified when jointly classifying stP and stHFBB activity. The gradients in stP timescale as well as connectivity to visually responsive regions and connectivity to regions that were not visually responsive were also consistent. Unlike the jointly classified data, electrodes selective in their stP demonstrated a decreasing ability to predict patient RT along the visual hierarchy ($\rho(380) = -0.11$, $p = .026$, uncorrected).

In contacts selective in their stHFBB activity, the correlations between local neural dynamics and long-range functional connectivity were smaller but were mostly consistent with those observed on the jointly classified data. However, connectivity to regions that were not visually responsive had the opposite relationship with distance along VTC compared to the jointly classified contacts ($\rho(150) = .30$, $p < .001$). Additionally, contacts that were sensitive to object category in their stHFBB activity demonstrated increasing ability to classify patient RT from the stHFBB activity when moving up the visual hierarchy ($\rho(150) = .24$, $p = .0026$). These results illustrate that functional anatomical gradients in information processing dynamics in ventral visual cortex are largely consistent across stP and stHFBB activity; however, these signal components do demonstrate differences in their anatomical gradients in timescales, connectivity to regions that are not visually response, and ability to predict patient response time during a visual 1-back task.

Appendix A.2 Supplementary figures

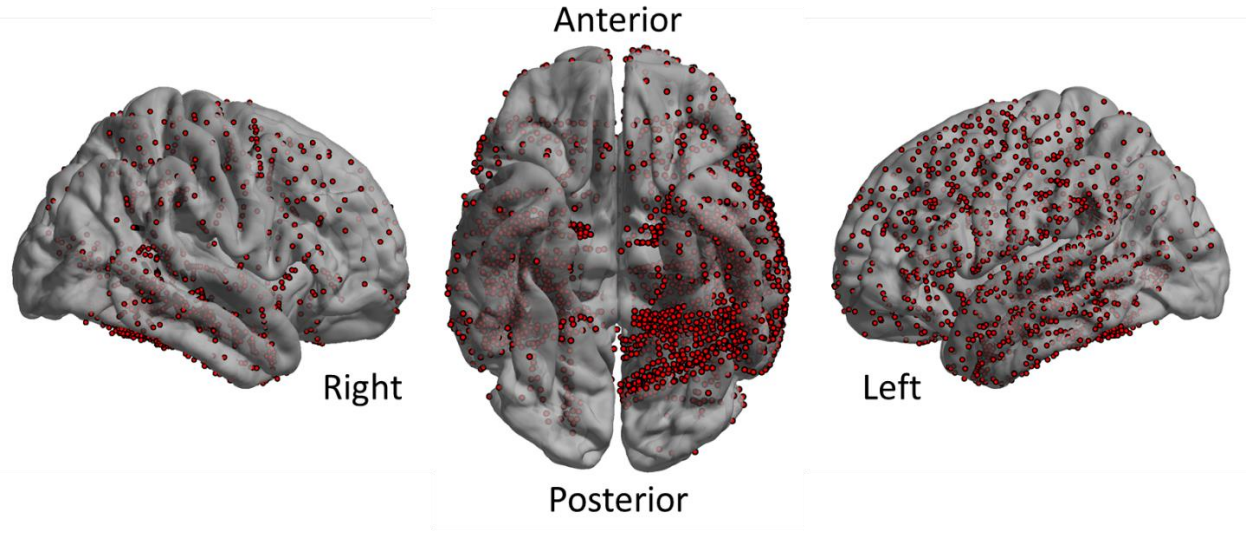


Figure 23. Intracranial electrode contact coverage of 35 patients with category-discriminant VTC electrode contacts. Contacts that appeared to be outside of the MNI standard brain due to differences in individual brain sizes were projected to the nearest MNI cortical vertex in this figure solely for illustrative purposes.

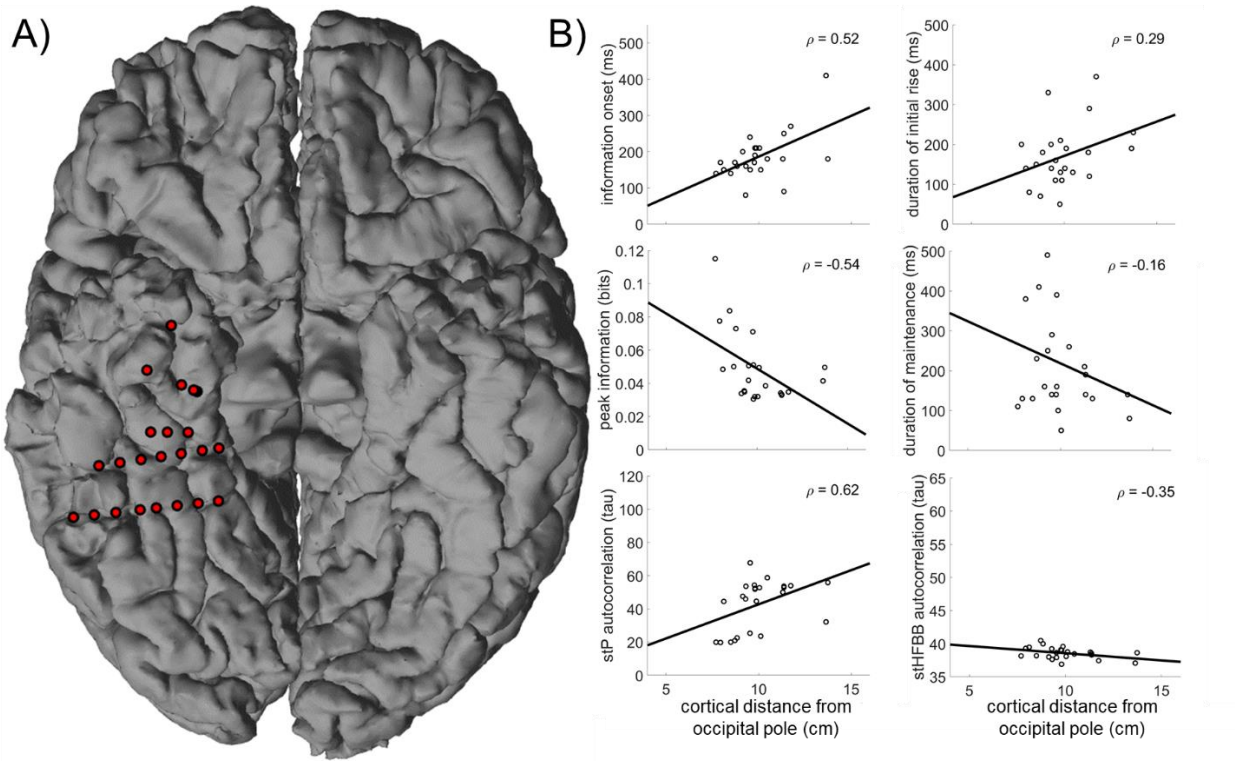


Figure 24. Information processing and neural timescales in VTC of one patient. A) Spatial topography of VTC electrode contacts with above chance ($p < .001$, corrected for multiple temporal comparisons) category-discriminant information on the individual's anatomy. B) Relationship between information processing dynamics and neural timescales examined in Figure 2 and Figure 3 and the cortical distance along VTC. Spearman correlation coefficient displayed in the top right of each panel. Line represents least-squares approximate fit. Information onset, peak information, and stP timescale (tau) were all significantly correlated with distance along VTC with the same direction as the group-level gradients. Other properties of the neural population's dynamics demonstrated similar gradients with distance along VTC as the group-level data but were not significant across these 24 electrodes.

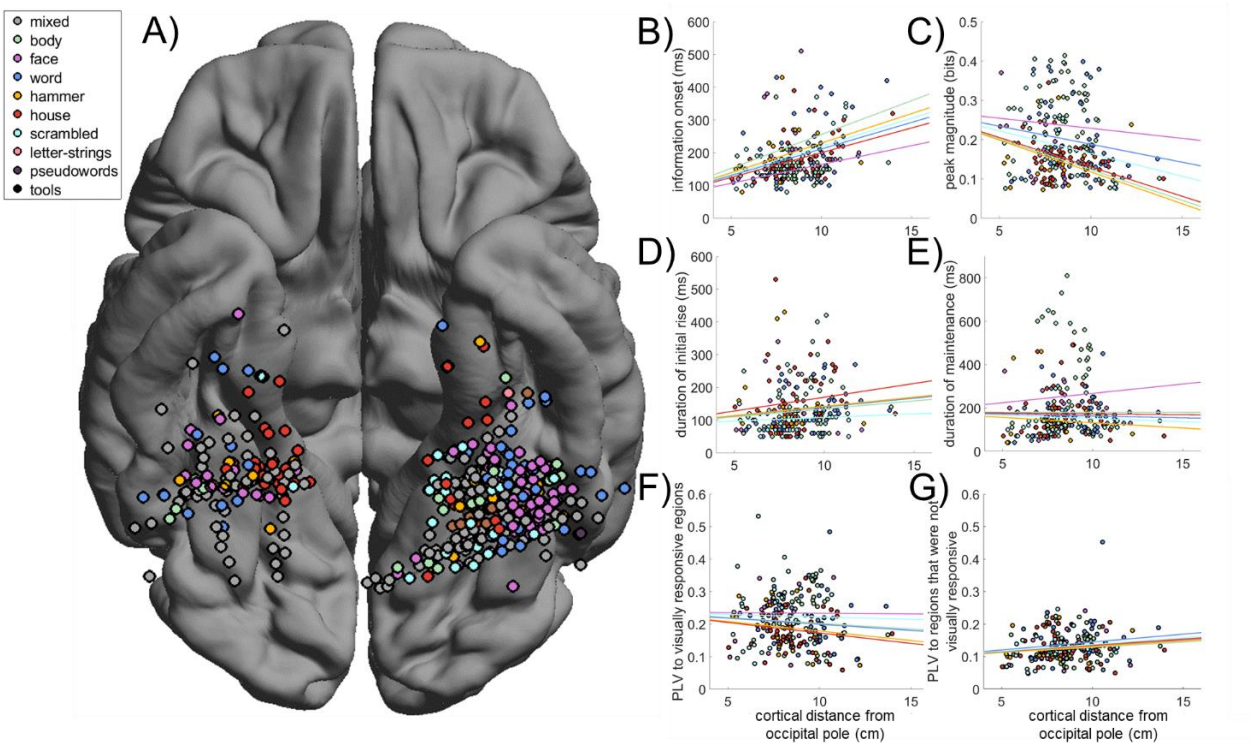


Figure 25. Differences in functional anatomical gradients across regions selective for different object categories. A) Spatial topography of electrode contacts predominantly selective for one object category. Grey electrode contacts contained a significant amount of category-discriminant neural information but were sensitive to more than one image category. B) Onset of category-discriminant information processing as a function of cortical distance along the ventral visual hierarchy for neural populations selective for faces, bodies, words, hammers, houses, and phase-scrambled objects. Lines are derived from multiple linear regression analyses. Corresponding statistics are contained in Table 4. Face-selective information demonstrated faster propagation along the visual hierarchy compared to the other object categories. C) Relationship between the duration of the initial rise in category-selective information versus distance along VTC. There was no significant difference in this gradient between neural populations selective for different object categories. D) Peak category-discriminant information as a function of distance along the visual hierarchy for VTC neural populations selective for different object categories. Face-selective information decayed less when moving down the ventral visual hierarchy compared to the other object categories. E) Maintenance duration of category-selective information as a function of distance along VTC. Face-selective neural populations demonstrated the greatest increase in information maintenance duration when moving

along VTC compared to the other object categories. F) Relationship between connectedness to visually responsive regions and distance along the visual hierarchy for VTC neural populations selective for different object categories. Word-, hammer-, and house-selective neural populations demonstrated greater decreases in connectivity to visually responsive regions when moving along VTC compared to face-selective neural populations. E) Relationship between connectedness to regions that were not visually responsive and position in the visual hierarchy for VTC neural populations selective for different object categories. There was no significant difference in this anatomical gradient across neural populations selective for different object categories.

Table 4. Statistically evaluating differences in functional anatomical gradients across processing hierarchies for different categories of objects.

	Estimate	St. Error	T-stat	Estimate	St. Error	T-stat
	Information onset (ms/cm)			**Duration of initial rise (ms/cm)**		
Faces	**11.46**	**2.83**	**4.04*** **	5.42	3.38	1.60
Bodies (difference from faces)	**9.15**	**1.98**	**4.61*** **	-0.56	2.37	-0.24
Words (difference from faces)	**4.72**	**1.37**	**3.46*** **	0.01	1.63	0.01
Hammers (difference from faces)	**6.52**	**2.06**	**3.16** **	0.22	2.46	0.09
Houses (difference from faces)	**3.60**	**1.43**	**2.51***	2.99	1.71	1.75
Scrambled (difference from faces)	**5.56**	**1.72**	**3.23** **	-3.18	2.06	-1.55
	Peak information magnitude (bits/cm)			**Duration of maintenance (ms/cm)**		
Faces	-0.0051	0.0035	-1.48	8.64	5.76	1.50
Bodies (difference from faces)	**-0.0105**	**0.0024**	**-4.32*** **	**-8.73**	**4.03**	**-2.17***
Words (difference from faces)	**-0.0041**	**0.0017**	**-2.419***	**-10.39**	**2.77**	**-3.74*** **
Hammers (difference from faces)	**-0.0111**	**0.0025**	**-4.39*** **	**-13.53**	**4.19**	**-3.23** **
Houses (difference from faces)	**-0.0098**	**0.0018**	**-5.59*** **	**-9.44**	**2.91**	**-3.24** **
Scrambled (difference from faces)	**-0.0064**	**0.0021**	**-3.04** **	**-11.57**	**3.50**	**-3.3017** **
	Connectivity to visually responsive regions (a.u.)			**Connectivity to regions that were not visually responsive (a.u.)**		
Faces	-0.0004	0.0033	-0.12	0.0036	0.0020	1.79
Bodies (difference from faces)	-0.0031	0.0023	-1.32	-0.0004	0.0014	-0.28
Words (difference from faces)	**-0.0034**	**0.0016**	**-2.09***	0.0012	0.0010	1.24
Hammers (difference from faces)	**-0.0055**	**0.0024**	**-2.24***	-0.0001	0.0015	-0.10
Houses (difference from faces)	**-0.0060**	**0.0017**	**-3.57*** **	0.0002	0.0010	0.20
Scrambled (difference from faces)	-0.0011	0.0020	-0.54	0.0005	0.0012	0.38

This table summarizes the statistical effects of the models illustrated in Figure 25**B-G. Contacts with mixed selectivity were not included in the models. Face-selective neural populations were used as the base-level of**

the analyses because they were the most prevalent. Therefore, all other rows indicate the difference between the slope of the gradient in face-selective versus other category-selective populations. There were significant differences in the anatomical gradients of information onset, peak, duration of maintenance, and connectivity to visual neural populations across neural populations selective for different object categories. Significant effects are highlighted in bold. (* p < .05, ** p < .01, *** p < .001)

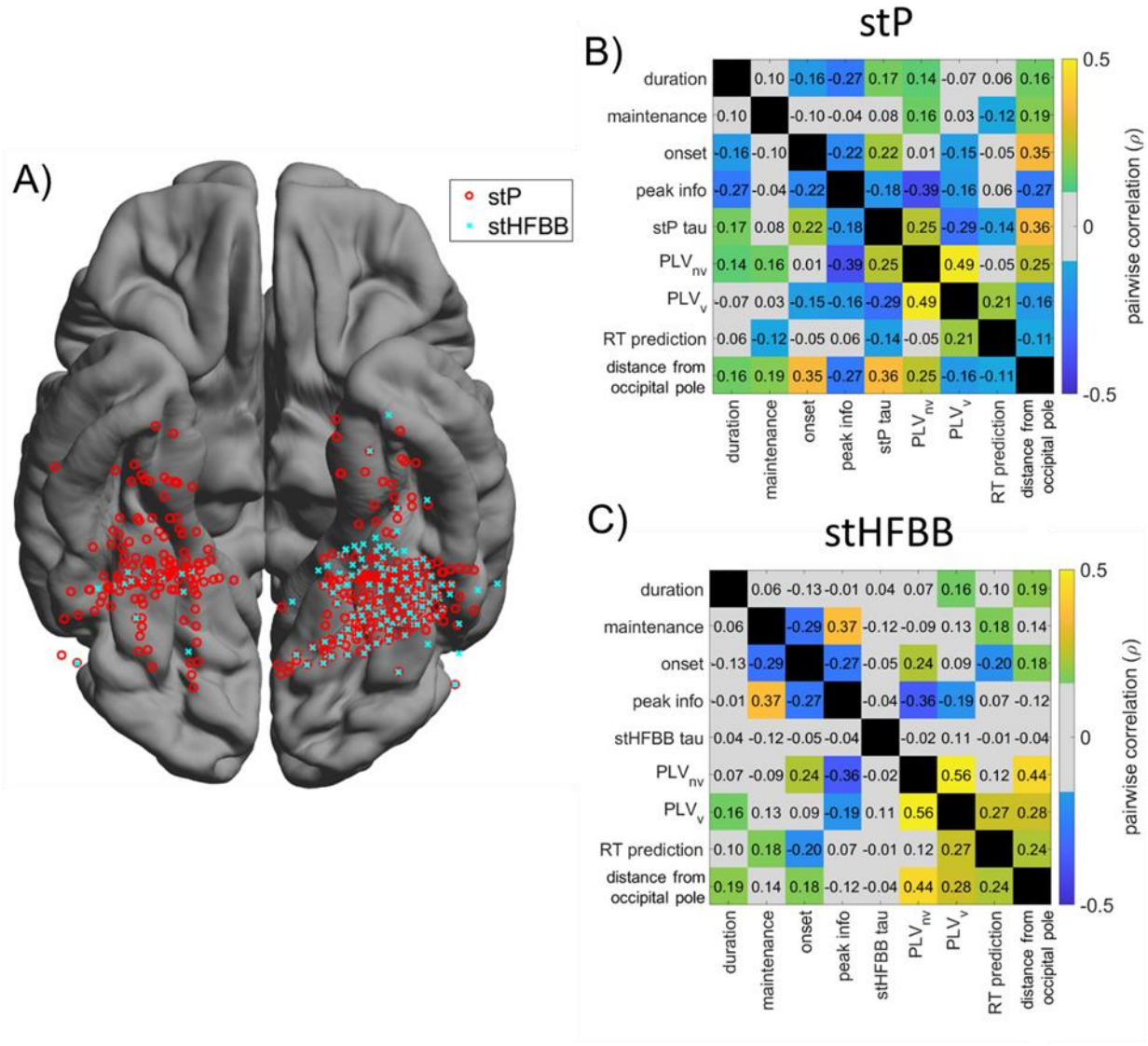


Figure 26. Differences in functional anatomical gradients across neural populations selective in their stP versus stHFBB. A) Electrode contacts demonstrating above-chance levels of category-discriminant information in single trial potentials (stP; n = 380) and single trial high frequency broadband (stHFBB; n = 150) when these signal components are classified separately. B) Pairwise Spearman correlations (ρ) between gradients in the corresponding row and column computed using electrode contacts selective in stP. Information processing metrics (onset, peak, maintenance, and rise durations) were computed for decoding time-courses derived from each signal separately. Shading indicates strength and direction of pairwise correlation. Grey squares indicate pairwise correlations that were not significant at the p < .05 level, uncorrected. The false-discovery rate adjusted critical value was estimated to be ρ = ± .138. C) Same as panel

B, but for contacts selective in their stHFBB activity. The false-discovery rate adjusted critical value was estimated to be ρ = ± .269. The pairwise correlation for jointly classified stP and stHFBB data is available in Figure 27.

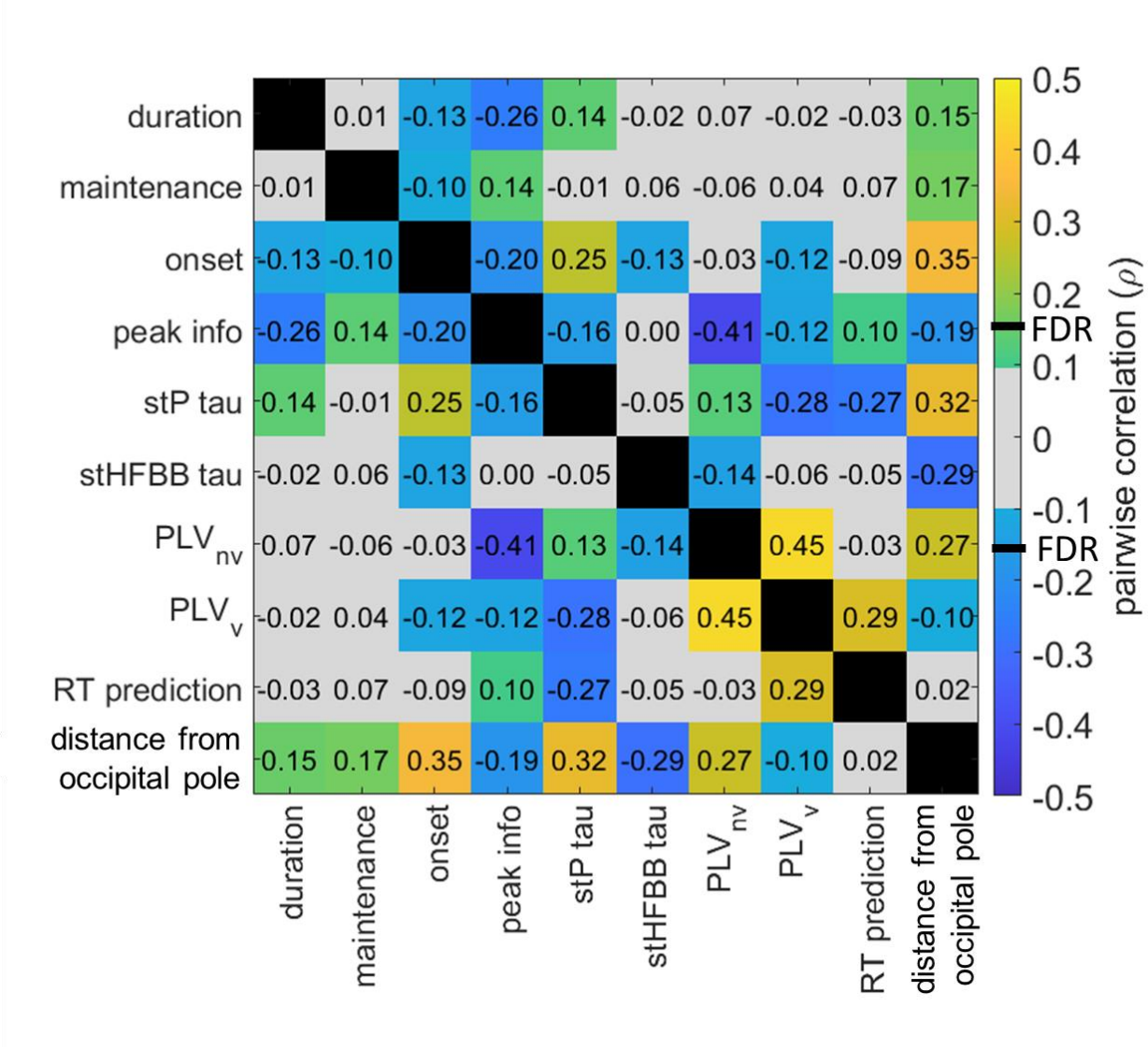


Figure 27. Pairwise correlations between dynamic and functional properties of VTC neural populations. Within each box is the pairwise correlation (ρ) between the variables in the corresponding row and column, like Figure 4, without removing the shared correlations with distance along VTC. Shading indicates strength and direction of pairwise correlation. Grey squares indicate pairwise correlations that were not significant at the $p < .05$ level, uncorrected. The false-discovery rate adjusted critical value was estimated to be $\rho = \pm .154$.

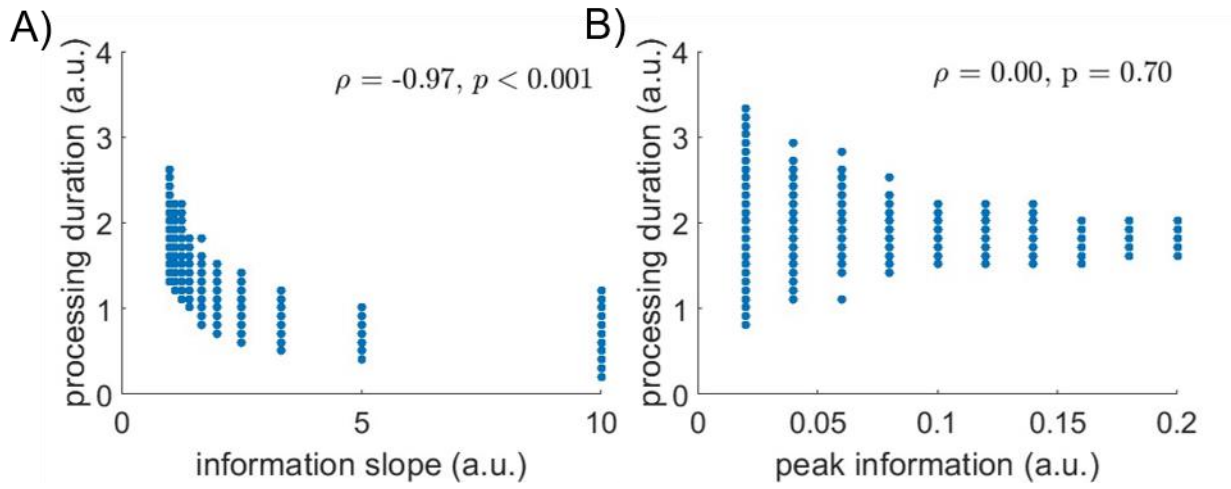


Figure 28. Simulating the effects of changes in information slope and peak amplitude on information processing duration. Information time-courses were simulated using a normal probability density function with similar signal to noise properties as category-selective information time-courses obtained from VTC. A) The slope of the information time-course was strongly correlated with the duration of information processing as expected. B) The duration of information processing was not correlated with changes in peak information. These simulations support that increases in information processing duration along the ventral visual hierarchy is not driven by differences in peak information amplitude.

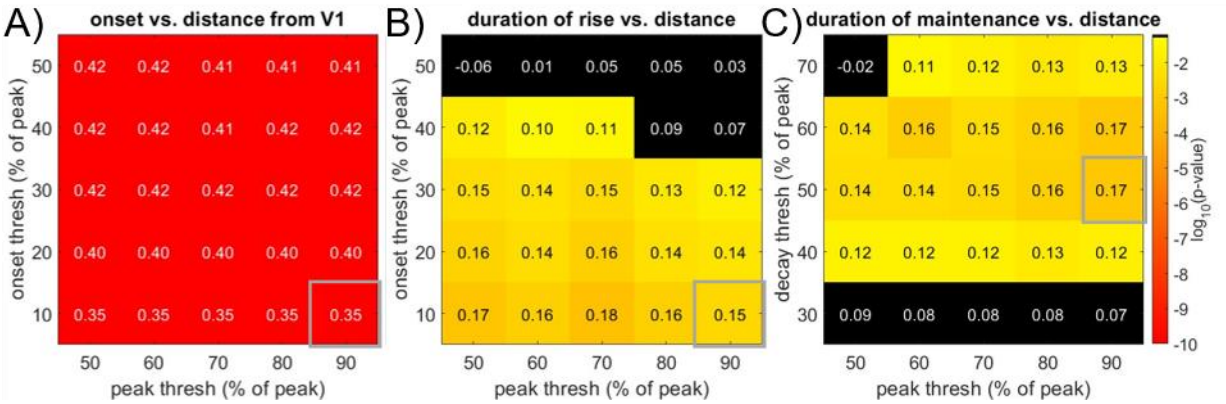


Figure 29. Anatomical gradients in information processing dynamics are robust to choices of onset and peak threshold. A) Spearman's correlation between a neural population's cortical distance along the visual hierarchy and its onset latency derived using different cutoff thresholds for onset and peak. Color of each square indicates p-value (color-bar on the right, log-scale). Inset of each square is the corresponding Spearman's ρ (n = 390). Gray square indicates result reported in the main text. Correlations derived from all criteria are very strong. B) Correlation between the duration of the initial rise in category-discriminant information and distance along the visual hierarchy with different onset and peak thresholds. Black squares indicate p-values greater than 0.05 uncorrected. C) Correlation between the duration of information maintenance and distance along VTC using different peak and decay thresholds. Decay was measured as the time between the reaching the peak threshold and the time the signal decayed below the decay threshold after reaching its absolute peak.

Appendix B Supplementary materials for Chapter 4

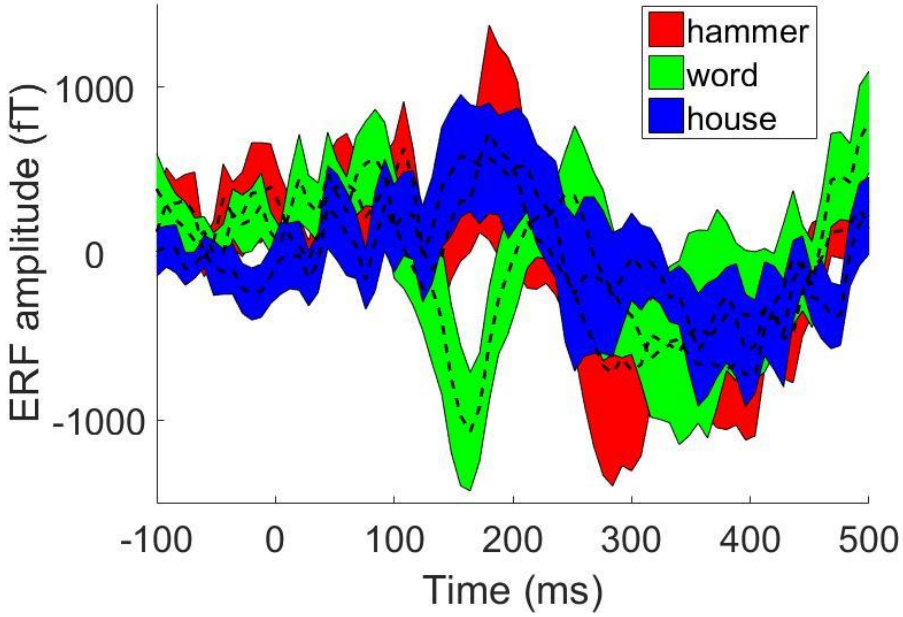


Figure 30. ERF of word-selective MEG wVOT sources. Average event-related field (ERF) of word-selective MEG wVOT sources across subjects. Error bars represent standard error across subjects.

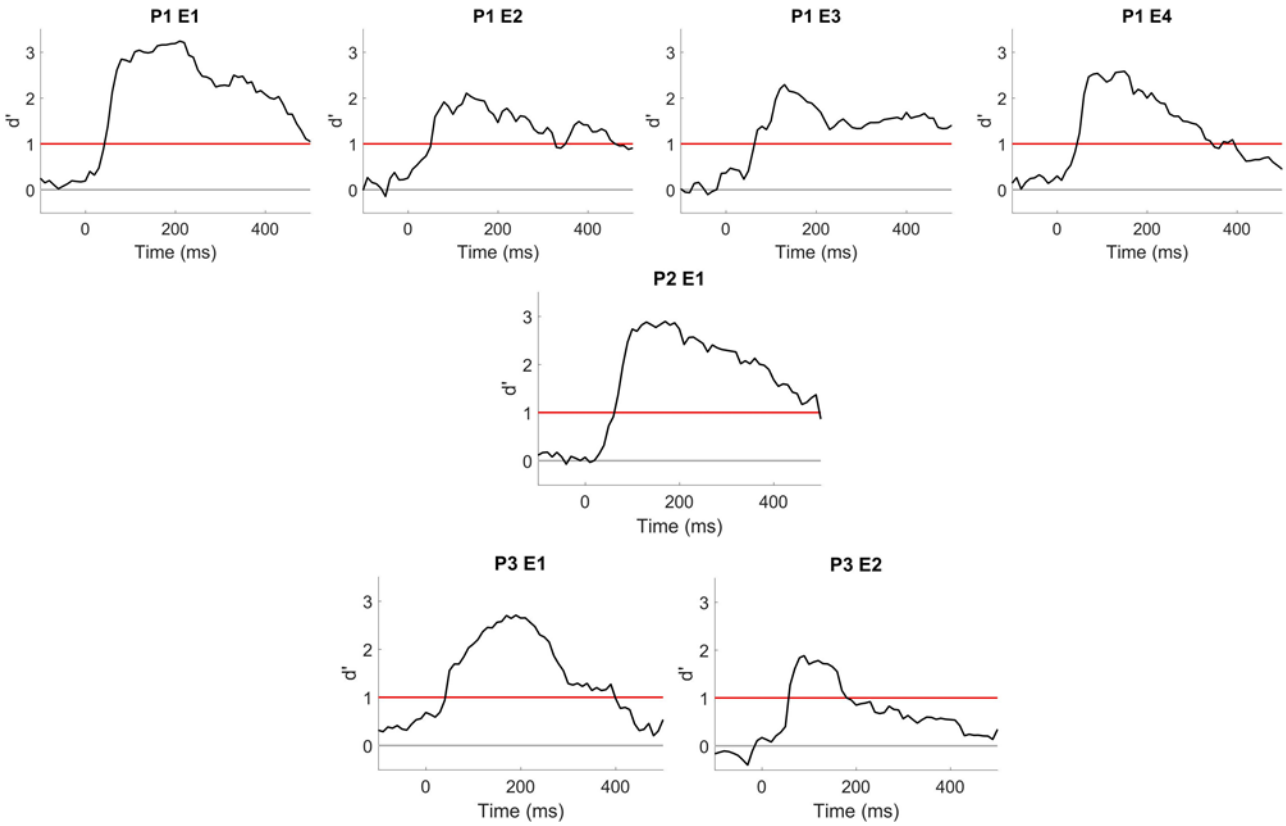


Figure 31. Sensitivity of word-selective iEEG electrodes. Sensitivity of six-way (words, bodies, hammers, houses, faces, phase-scrambled objects) decoding for the word-selective electrodes used in the word-individuation task. Electrodes with d' greater than one were considered word-selective.

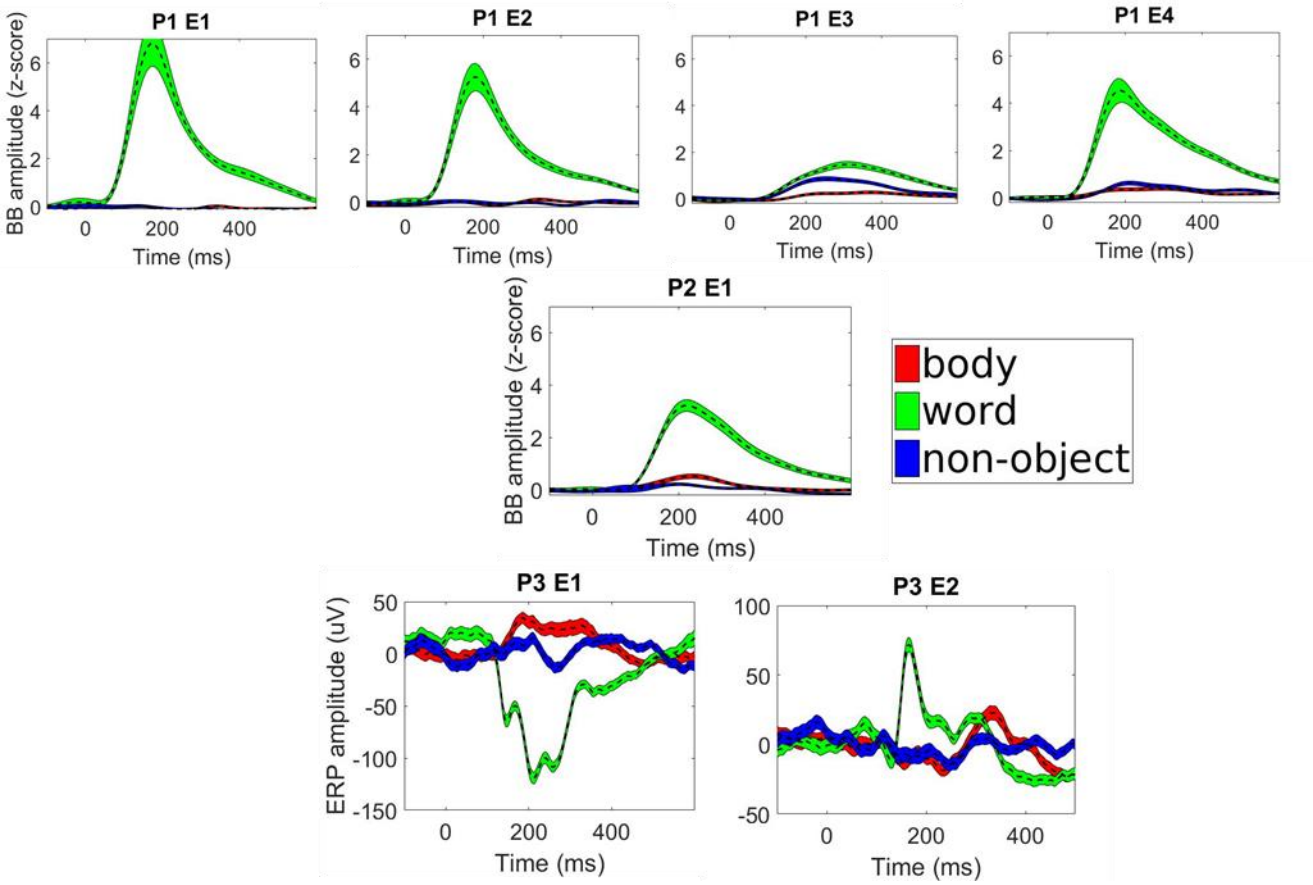


Figure 32. ERP/ERBB of word-selective iEEG electrodes. Event-related broad band (ERBB) or event-related potential (ERP) response to words, bodies and phase scrambled objects in word-selective electrodes used in the word-individuation task. For all electrodes either the ERP or ERBB for words was greater than for all other object categories. Error bars represent standard error across trials for each electrode.

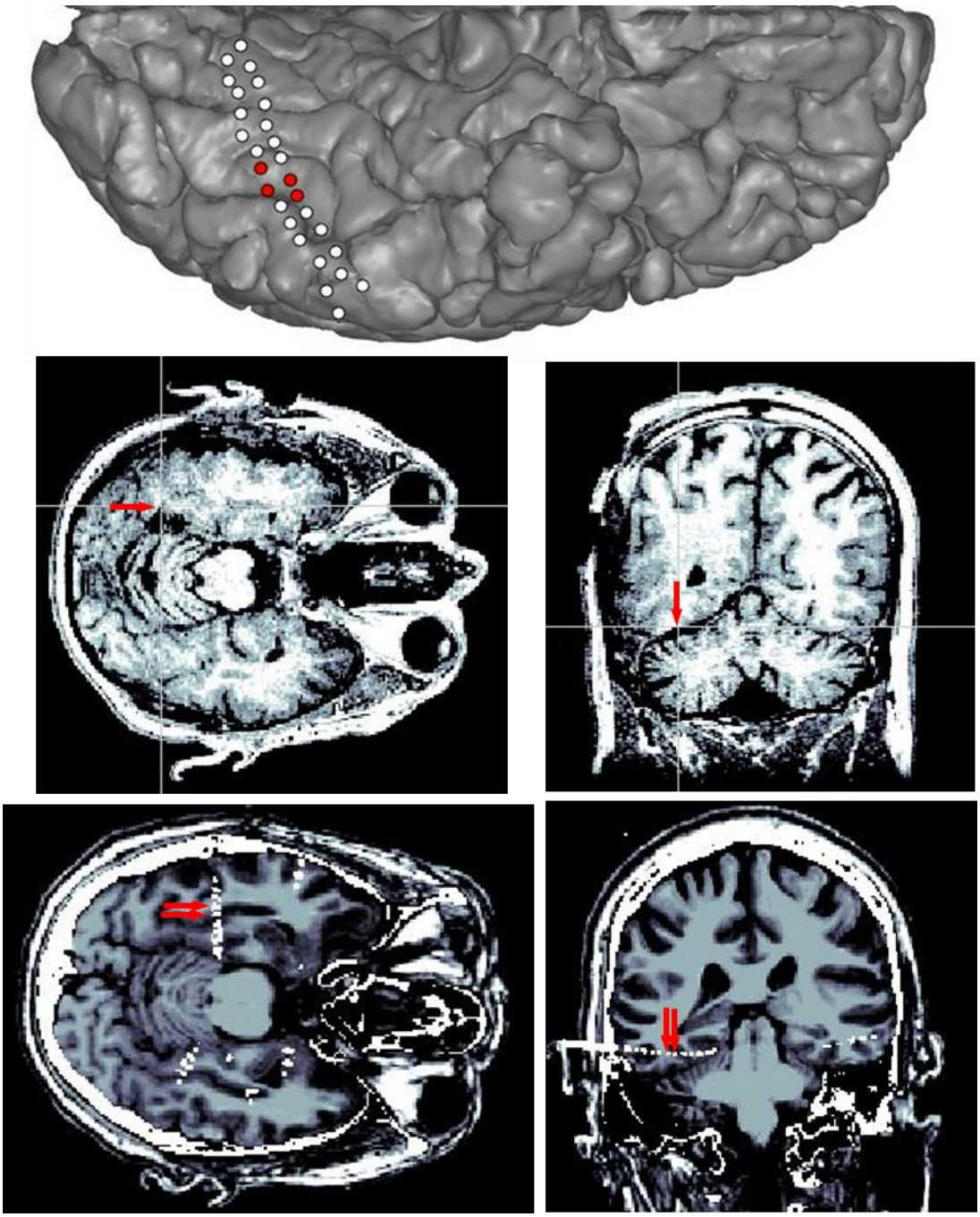


Figure 33. Localization of word-sensitive iEEG electrode localizations within patients. Word sensitive electrode localizations for the three iEEG participants. Red arrows indicate the contacts used in the word-individuation task.

Bibliography

1 Mishkin, M. *et al.* (1983) Object vision and spatial vision: two cortical pathways. *Trends Neurosci.* 6, 414–417

2 Warrington, E.K. and Shallice, T. (1984) Category specific semantic impairments. *Brain* 107, 829–853

3 Caramazza, A. and Mahon, B.Z. (2003) The organization of conceptual knowledge: the evidence from category-specific semantic deficits. *Trends Cogn. Sci.* 7, 354–361

4 Susilo, T. *et al.* (2015) Acquired prosopagnosia without word recognition deficits. *Cogn. Neuropsychol.* 32, 321–339

5 Farah, M.J. (2004) *Visual agnosia*, (2nd edn) MIT Press.

6 Moscovitch, M. *et al.* (1997) What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *J. Cogn. Neurosci.* 9, 555–604

7 Caramazza, A. and Shelton, J.R. (1998) Domain-Specific Knowledge Systems in the Brain: The Animate-Inanimate Distinction. *J. Cogn. Neurosci.* 10, 1–34

8 Puce, A. *et al.* (1996) Differential Sensitivity of Human Visual Cortex to Faces, Letterstrings, and Textures: A Functional Magnetic Resonance Imaging Study. *J. Neurosci.* 16, 5205–5215

9 Grill-Spector, K. and Weiner, K.S. (2014) The functional architecture of the ventral temporal cortex and its role in categorization. *Nat. Rev. Neurosci.* 15, 536–48

10 Kanwisher, N. *et al.* (1997) The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *J. Neurosci.* 17, 4302–4311

11 Downing, P.E. *et al.* (2001) A cortical area selective for visual processing of the human body. *Science.* 293, 2470–2473

12 Kanwisher, N. (2000) Domain specificity in face perception. *Nat. Neurosci. 2000 38* 3, 759–763

13 Downing, P.E. *et al.* (2006) Domain Specificity in Visual Cortex. *Cereb. Cortex* 16, 1453–1461

14 Kronbichler, M. *et al.* (2004) The visual word form area and the frequency with which words are encountered: evidence from a parametric fMRI study. *Neuroimage* 21, 946–953

15 Epstein, R. and Kanwisher, N. (1998) A cortical representation of the local visual environment. *Nat. 1998 3926676* 392, 598–601

16 Parvizi, J. *et al.* (2012) Electrical Stimulation of Human Fusiform Face-Selective Regions Distorts Face Perception. *J. Neurosci.* 32, 14915–14920

17 Rangarajan, V. *et al.* (2014) Electrical stimulation of the left and right human fusiform gyrus causes different effects in conscious face perception. *J. Neurosci.* 34, 12828–12836

18 Schrouff, J. *et al.* (2020) Fast temporal dynamics and causal relevance of face processing in the human temporal cortex. *Nat. Commun.* 11, 1–13

19 Hirshorn, E.A. *et al.* (2016) Decoding and disrupting left midfusiform gyrus activity during word reading. *Proc. Natl. Acad. Sci. U. S. A.* 113, 8162–8167

20 Parasuraman, R. (2011) Neuroergonomics: Brain, cognition, and performance at work. *Curr. Dir. Psychol. Sci.* 20, 181–186

21 Forseth, K.J. *et al.* (2018) A lexical semantic hub for heteromodal naming in middle fusiform gyrus. *Brain* 141, 2112–2126

22 Mani, J. *et al.* (2008) Evidence for a basal temporal visual language center: Cortical stimulation producing pure alexia. *Neurology* 71, 1621–1627

23 Enatsu, R. *et al.* (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

24 Sabsevitz, D.S. *et 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

25 Gaillard, R. *et al.* (2006) Direct Intracranial, fMRI, and Lesion Evidence for the Causal Role of Left Inferotemporal Cortex in Reading. *Neuron* 50, 191–204

26 Ishai, A. *et al.* (1999) Distributed representation of objects in the human ventral visual pathway. *Proc. Natl. Acad. Sci. U. S. A.* 96, 9379–9384

27 Blauch, N.M. *et al.* (2022) A connectivity-constrained computational account of topographic organization in primate high-level visual cortex. *Proc. Natl. Acad. Sci. U. S. A.* 119, e2112566119

28 Behrmann, M. and Plaut, D.C. (2013) Distributed circuits, not circumscribed centers, mediate visual recognition. *Trends Cogn. Sci.* 17, 210–219

29 Plaut, D.C. and Behrmann, M. (2011) Complementary neural representations for faces and words: a computational exploration. *Cogn. Neuropsychol.* 28, 251–275

30 Bukach, C.M. *et al.* (2006) Beyond faces and modularity: the power of an expertise

framework. *Trends Cogn. Sci.* 10, 159–166

31 Gauthier, I. and Tarr, M.J. (1997) Becoming a "Greeble" Expert: Exploring Mechanisms for Face Recognition. *Vision Res.* 37, 1673–1682

32 Gauthier, I. *et al.* (2000) Expertise for cars and birds recruits brain areas involved in face recognition. *Nat. Neurosci. 2000 32* 3, 191–197

33 Bruce, C. *et al.* (1981) Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J. Neurophysiol.* 46, 369–384

34 Kravitz, D.J. *et al.* (2013) The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends Cogn. Sci.* 17, 26–49

35 Felleman, D.J. and Van Essen, D.C. (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47

36 Schmolesky, M.T. *et al.* (1998) Signal timing access the macaque visual system. *J. Neurophysiol.* 79, 3272–3278

37 Yamins, D.L.K. *et al.* (2014) Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proc. Natl. Acad. Sci. U. S. A.* 111, 8619–8624

38 DiCarlo, J.J. *et al.* (2012) How Does the Brain Solve Visual Object Recognition? *Neuron* 73, 415–434

39 Lamme, V.A.F. and Roelfsema, P.R. (2000) The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* 23, 571–579

40 Thorpe, S. *et al.* (1996) Speed of processing in the human visual system. *Nature* 381, 520–522

41 Cadieu, C.F. *et al.* (2014) Deep Neural Networks Rival the Representation of Primate IT Cortex for Core Visual Object Recognition. *PLOS Comput. Biol.* 10, e1003963

42 Liu, H. *et al.* (2009) Timing, Timing, Timing: Fast Decoding of Object Information from Intracranial Field Potentials in Human Visual Cortex. *Neuron* 62, 281–290

43 Conway, B.R. (2018) The Organization and Operation of Inferior Temporal Cortex. *Annu. Rev. Vis. Sci.* 4, 381–402

44 Arcaro, M.J. and Livingstone, M.S. (2017) A hierarchical, retinotopic proto-organization of the primate visual system at birth. *Elife* 6, e26196

45 Grill-Spector, K. and Malach, R. The human visual cortex. , *Annual Review of Neuroscience*, 27. 06-Jul-(2004) , Annual Reviews, 649–677

46 De Beeck, H.O. and Vogels, R. (2000) Spatial sensitivity of macaque inferior temporal neurons. *J. Comp. Neurol.* 426, 505–518

47 Kravitz, D.J. *et al.* (2008) How position dependent is visual object recognition? *Trends Cogn. Sci.* 12, 114–122

48 Arcaro, M.J. and Livingstone, M.S. (2021) On the relationship between maps and domains in inferotemporal cortex. *Nat. Rev. Neurosci.* 22, 573–583

49 Hasson, U. *et al.* (2002) Eccentricity Bias as an Organizing Principle for Human High-Order Object Areas. *Neuron* 34, 479–490

50 Kamps, F.S. *et al.* (2020) Connectivity at the origins of domain specificity in the cortical face and place networks. *Proc. Natl. Acad. Sci. U. S. A.* 117, 6163–6169

51 Levy, I. *et al.* (2001) Center-periphery organization of human object areas. *Nat. Neurosci.* 4, 533–539

52 Silson, E.H. *et al.* (2015) A Retinotopic Basis for the Division of High-Level Scene Processing between Lateral and Ventral Human Occipitotemporal Cortex. *J. Neurosci.* 35, 11921–11935

53 Yue, X. *et al.* (2020) Curvature processing in human visual cortical areas. *Neuroimage* 222, 117295

54 Bao, P. *et al.* (2020) A map of object space in primate inferotemporal cortex. *Nature* 583, 103–108

55 Srihasam, K. *et al.* (2014) Novel domain formation reveals proto-architecture in inferotemporal cortex. *Nat. Neurosci.* 17, 1776–1783

56 Srihasam, K. *et al.* (2012) Behavioral and Anatomical Consequences of Early versus Late Symbol Training in Macaques. *Neuron* 73, 608–619

57 Arcaro, M.J. *et al.* (2019) Body map proto-organization in newborn macaques. *Proc. Natl. Acad. Sci. U. S. A.* 116, 24861–24871

58 Arcaro, M.J. *et al.* (2017) Seeing faces is necessary for face-patch formation. *Nat. Neurosci.* 20, 1404

59 Behrmann, M. and Plaut, D.C. (2014) Bilateral Hemispheric Processing of Words and Faces: Evidence from Word Impairments in Prosopagnosia and Face Impairments in Pure Alexia. *Cereb. Cortex* 24, 1102–1118

60 Behrmann, M. and Plaut, D.C. (2020) Hemispheric Organization for Visual Object Recognition: A Theoretical Account and Empirical Evidence. *Perception* 49, 373–404

61 Powell, L.J. *et al.* (2018) Social Origins of Cortical Face Areas. *Trends Cogn. Sci.* 22, 752

62 Price, C.J. and Devlin, J.T. (2011) The Interactive Account of ventral occipitotemporal contributions to reading. *Trends Cogn. Sci.* 15, 246–253

63 Mahon, B.Z. and Caramazza, A. (2011) What drives the organization of object knowledge in the brain? *Trends Cogn. Sci.* 15, 97–103

64 Li, J. *et al.* (2020) Innate connectivity patterns drive the development of the visual word form area. *Sci. Reports 2020 101* 10, 1–12

65 Saygin, Z.M. *et al.* (2016) Connectivity precedes function in the development of the visual word form area. *Nat. Neurosci.* 19, 1250–5

66 Ghuman, A.S. *et al.* (2014) Dynamic encoding of face information in the human fusiform gyrus. *Nat. Commun.* 5, 5672

67 Sugase, Y. *et al.* (1999) Global and fine information coded by single neurons in the temporal visual cortex. *Nature* 400, 869–873

68 Koyano, K.W. *et al.* (2021) Dynamic Suppression of Average Facial Structure Shapes Neural Tuning in Three Macaque Face Patches. *Curr. Biol.* 31, 1-12.e5

69 Honey, C.J. *et al.* (2012) Slow Cortical Dynamics and the Accumulation of Information over Long Timescales. *Neuron* 76, 423–434

70 Chaudhuri, R. *et al.* (2015) A Large-Scale Circuit Mechanism for Hierarchical Dynamical Processing in the Primate Cortex. *Neuron* 88, 419–431

71 Mellem, M.S. *et al.* (2017) Intrinsic frequency biases and profiles across human cortex. *J. Neurophysiol.* 118, 2853–2864

72 Cocchi, L. *et al.* (2016) A hierarchy of timescales explains distinct effects of local inhibition of primary visual cortex and frontal eye fields. *Elife* 5, e15252

73 Wolff, A. *et al.* (2022) Intrinsic neural timescales: temporal integration and segregation. *Trends Cogn. Sci.* 26, 159–173

74 Murray, J.D. *et al.* (2014) A hierarchy of intrinsic timescales across primate cortex. *Nat. Neurosci.* 17, 1661–1663

75 Spitmaan, M. *et al.* (2020) Multiple timescales of neural dynamics and integration of task-relevant signals across cortex. *Proc. Natl. Acad. Sci. U. S. A.* 117, 22522–22531

76 Hasson, U. *et al.* (2008) A Hierarchy of Temporal Receptive Windows in Human Cortex. *J. Neurosci.* 28, 2539–2550

77 Lerner, Y. *et al.* (2011) Topographic Mapping of a Hierarchy of Temporal Receptive Windows Using a Narrated Story. *J. Neurosci.* 31, 2906–2915

78 Margulies, D.S. *et al.* (2016) Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proc. Natl. Acad. Sci. U. S. A.* 113, 12574–12579

79 Demirtaş, M. *et al.* (2019) Hierarchical Heterogeneity across Human Cortex Shapes Large-

Scale Neural Dynamics. *Neuron* 101, 1181-1194.e13

80 Huntenburg, J.M. *et al.* Large-Scale Gradients in Human Cortical Organization. , *Trends in Cognitive Sciences*, 22. 01-Jan-(2018) , Elsevier Ltd, 21–31

81 Fan, X. *et al.* (2020) The bottom-up and top-down processing of faces in the human occipitotemporal cortex. *Elife* 9, e48764

82 Whaley, M.L. *et al.* (2016) Modulation of Orthographic Decoding by Frontal Cortex. *J. Neurosci.* 36, 1173–1184

83 McClelland, J.L. and Rumelhart, D.E. (1981) An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychol. Rev.* 88, 375–407

84 Raut, R. V. *et al.* (2020) Hierarchical dynamics as a macroscopic organizing principle of the human brain. *Proc. Natl. Acad. Sci. U. S. A.* 117, 20890–20897

85 Ito, T. *et al.* (2020) A cortical hierarchy of localized and distributed processes revealed via dissociation of task activations, connectivity changes, and intrinsic timescales. *Neuroimage* 221, 117141

86 Hurka, J. Vanden *et al.* (2017) Development of visual category selectivity in ventral visual cortex does not require visual experience. *Proc. Natl. Acad. Sci. U. S. A.* 114, E4501–E4510

87 Mahon, B.Z. *et al.* (2009) Category-Specific Organization in the Human Brain Does Not Require Visual Experience. *Neuron* 63, 397–405

88 Reich, L. *et al.* (2011) A Ventral Visual Stream Reading Center Independent of Visual Experience. *Curr. Biol.* 21, 363–368

89 Büchel, C. *et al.* (1998) A multimodal language region in the ventral visual pathway. *Nature* 394, 274–275

90 Bar, M. Visual objects in context. , *Nature Reviews Neuroscience*, 5. (2004) , Nature Publishing Group, 617–629

91 Kveraga, K. *et al.* (2007) Top-down predictions in the cognitive brain. *Brain Cogn.* 65, 145–168

92 Bar, M. *et al.* (2006) Top-down facilitation of visual recognition. *Proc. Natl. Acad. Sci. U. S. A.* 103, 449–454

93 Ghuman, A.S. *et al.* (2008) The effects of priming on frontal-temporal communication. *Proc. Natl. Acad. Sci. U. S. A.* 105, 8405–8409

94 Op de Beeck, H.P. *et al.* (2019) Factors Determining Where Category-Selective Areas Emerge in Visual Cortex. *Trends Cogn. Sci.* 23, 784–797

95 Paquola, C. *et al.* (2019) Microstructural and functional gradients are increasingly

dissociated in transmodal cortices. *PLoS Biol.* 17, e3000284

96 Gao, R. *et al.* (2020) Neuronal timescales are functionally dynamic and shaped by cortical microarchitecture. *Elife* 9, e61277

97 Runyan, C.A. *et al.* (2017) Distinct timescales of population coding across cortex. *Nature* 548, 92–96

98 Ding, N. *et al.* (2015) Cortical tracking of hierarchical linguistic structures in connected speech. *Nat. Neurosci.* 19, 158–164

99 Kiebel, S.J. *et al.* (2008) A hierarchy of time-scales and the brain. *PLoS Comput. Biol.* 4, e1000209

100 Boring, M.J. *et al.* (2021) Multiple adjoining word- and face-selective regions in ventral temporal cortex exhibit distinct dynamics. *J. Neurosci.* 41, 6314–6327

101 Tadel, F. *et al.* (2011) Brainstorm: a user-friendly application for MEG/EEG analysis. *Comput. Intell. Neurosci.* 2011, 879716

102 Dale, A.M. *et al.* (1999) Cortical Surface-Based Analysis. *Neuroimage* 9, 179–194

103 Hermes, D. *et al.* (2010) Automated electrocorticographic electrode localization on individually rendered brain surfaces. *J. Neurosci. Methods* 185, 293–298

104 Evans, A.C. *et al.* (1994) , 3D statistical neuroanatomical models from 305 MRI volumes. in *IEEE Nuclear Science Symposium & Medical Imaging Conference*, pp. 1813–1817

105 Kirsanov, D. (2021) Exact geodesic for triangular meshes. [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/18168-exact-geodesic-for-triangular-meshes. [Accessed: 30-Sep-2021]

106 Brainard, D.H. (1997) The Psychophysics Toolbox. *Spat. Vis.* 10, 433–436

107 Oostenveld, R. *et al.* (2011) FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011, 156869

108 Miller, K.J. *et al.* (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

109 Rolls, E.T. *et al.* (1997) The representational capacity of the distributed encoding of information provided by populations of neurons in primate temporal visual cortex. *Exp. Brain Res.* 114, 149–162

110 Rolls, E.T. *et al.* (1996) Representation of olfactory information in the primate orbitofrontal cortex. *J. Neurophysiol.* 75, 1982–1996

111 Samengo, I. and Inés (2002) Information Loss in an Optimal Maximum Likelihood Decoding. *Neural Comput.* 14, 771–779

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

113 Lachaux, J.-P. *et al.* (1999) Measuring Phase Synchrony in Brain Signals. *Hum Brain Mapp.* 8, 194–208

114 Lowet, E. *et al.* (2016) Quantifying Neural Oscillatory Synchronization: A Comparison between Spectral Coherence and Phase-Locking Value Approaches. *PLoS One* 11, e0146443

115 Ghuman, A.S. *et al.* (2011) A wavelet-based method for measuring the oscillatory dynamics of resting-state functional connectivity in MEG. *Neuroimage* 56, 69–77

116 Benjamini, Y. and Yekutieli, D. (2001) The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* 29, 1165–1188

117 Luke, S.G. (2017) Evaluating significance in linear mixed-effects models in R. *Behav. Res. Methods* 49, 1494–1502

118 Cavanagh, S.E. *et al.* (2016) Autocorrelation structure at rest predicts value correlates of single neurons during reward-guided choice. *Elife* 5, e18937

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

120 Leszczyński, M. *et al.* (2020) Dissociation of broadband high-frequency activity and neuronal firing in the neocortex. *Sci. Adv.* 6, 977–989

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

122 Vinckier, F. *et al.* (2007) Hierarchical Coding of Letter Strings in the Ventral Stream: Dissecting the Inner Organization of the Visual Word-Form System. *Neuron* 55, 143–156

123 Abel, T.J. *et al.* (2015) Direct physiologic evidence of a heteromodal convergence region for proper naming in human left anterior temporal lobe. *J. Neurosci.* 35, 1513–1520

124 Clarke, A. *et al.* (2011) The Evolution of Meaning: Spatio-temporal Dynamics of Visual Object Recognition. *J. Cogn. Neurosci.* 23, 1887–1899

125 Serre, T. *et al.* (2007) A feedforward architecture accounts for rapid categorization. *Proc. Natl. Acad. Sci. U. S. A.* 104, 6424–9

126 Sansom, S.N. and Livesey, F.J. (2009) Gradients in the brain: the control of the development

of form and function in the cerebral cortex. *Cold Spring Harb. Perspect. Biol.* 1, a002519

127 Burt, J.B. *et al.* (2018) Hierarchy of transcriptomic specialization across human cortex captured by structural neuroimaging topography. *Nat. Neurosci.* 21, 1251–1259

128 Ghuman, A.S. and Martin, A. (2019) Dynamic Neural Representations: An Inferential Challenge for fMRI. *Trends Cogn. Sci.* 23, 534–536

129 Li, Y. *et al.* (2018) Posterior Fusiform and Midfusiform Contribute to Distinct Stages of Facial Expression Processing. *Cereb. Cortex* 29, 3209–3219

130 Tang, H. *et al.* (2014) Spatiotemporal Dynamics Underlying Object Completion in Human Ventral Visual Cortex. *Neuron* 83, 736–748

131 Quian Quiroga, R. *et al.* (2008) Human single-neuron responses at the threshold of conscious recognition. *Proc. Natl. Acad. Sci. U. S. A.* 105, 3599–3604

132 Kampf, M. *et al.* (2002) A serial test of the laterality of familiar face recognition. *Brain Cogn.* 50, 35–50

133 Ghuman, A.S. *et al.* (2013) Interregional neural synchrony has similar dynamics during spontaneous and stimulus-driven states. *Sci. Rep.* 3, 1481

134 Churchland, M.M. *et al.* (2010) Stimulus onset quenches neural variability: A widespread cortical phenomenon. *Nat. Neurosci.* 13, 369–378

135 Konkle, T. and Caramazza, A. (2013) Tripartite organization of the ventral stream by animacy and object size. *J. Neurosci.* 33, 10235–42

136 Cohen, L. *et al.* (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

137 Haxby, J. V *et al.* (1996) Face encoding and recognition in the human brain. *Proc. Natl. Acad. Sci. U. S. A.* 93, 922–927

138 Aguirre, G.K. *et al.* (1998) An area within human ventral cortex sensitive to "building" stimuli: Evidence and implications. *Neuron* 21, 373–383

139 Kay, K.N. and Yeatman, J.D. (2017) Bottom-up and top-down computations in word- and face-selective cortex. *Elife* 6, e22341

140 Stevens, W.D. *et al.* (2015) Functional connectivity constrains the category-related organization of human ventral occipitotemporal cortex. *Hum. Brain Mapp.* 36, 2187–2206

141 Stevens, W.D. *et al.* (2017) Privileged functional connectivity between the visual word form area and the language system. *J. Neurosci.* 37, 5288–5297

142 Lerma-Usabiaga, G. *et al.* (2018) Converging evidence for functional and structural

segregation within the left ventral occipitotemporal cortex in reading. *Proc. Natl. Acad. Sci. U. S. A.* 115, E9981–E9990

143 Halgren, E. *et al.* (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

144 Dehaene, S. and Cohen, L. The unique role of the visual word form area in reading. , *Trends in Cognitive Sciences*, 15. (2011) , 254–262

145 Weiner, K.S. and Grill-Spector, K. (2010) Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. *Neuroimage* 52, 1559–73

146 Tsao, D.Y. *et al.* (2008) Comparing face patch systems in macaques and humans. *Proc. Natl. Acad. Sci. U. S. A.* 105, 19514–19519

147 Allison, T. *et al.* (1994) Human extrastriate visual cortex and the perception of faces, words, numbers, and colors. *Cereb. Cortex* 4, 544–554

148 Haxby, J. V *et al.* (1994) The functional organization of human extrastriate cortex: a PET-rCBF study of selective attention to faces and locations. *J. Neurosci.* 14, 6336–53

149 Harris, R.J. *et al.* (2016) Distinct but overlapping patterns of response to words and faces in the fusiform gyrus. *Cereb. Cortex* 26, 3161–3168

150 Matsuo, T. *et al.* (2015) Alternating Zones Selective to Faces and Written Words in the Human Ventral Occipitotemporal Cortex. *Cereb. Cortex* 25, 1265–1277

151 Jonas, J. *et al.* (2016) A face-selective ventral occipito-temporal map of the human brain with intracerebral potentials. *Proc. Natl. Acad. Sci. U. S. A.* 113, E4088–E4097

152 Engell, A.D. *et al.* (2012) The fMRI BOLD signal tracks electrophysiological spectral perturbations, not event-related potentials. *Neuroimage* 59, 2600–2606

153 Dehaene, S. *et al.* (2010) How learning to read changes the cortical networks for vision and language. *Science.* 330, 1359–1364

154 Federmeier, K.D. and Kutas, M. (1999) Right words and left words: Electrophysiological evidence for hemispheric differences in meaning processing. *Cogn. Brain Res.* 8, 373–392

155 Jacques, C. *et al.* (2016) Corresponding ECoG and fMRI category-selective signals in human ventral temporal cortex. *Neuropsychologia* 83, 14–28

156 Hermes, D. *et al.* (2012) Neurophysiologic correlates of fMRI in human motor cortex. *Hum. Brain Mapp.* 33, 1689–1699

157 Cox, R.W. (1996) AFNI: Software for Analysis and Visualization of Functional Magnetic Resonance Neuroimages. *Comput. Biomed. Res.* 29, 162–173

158 Susilo, T. and Duchaine, B. (2013) Dissociations between faces and words: Comment on

Behrmann and Plaut. *Trends Cogn. Sci.* 17, 545

159 Kadipasaoglu, C.M. *et al.* (2016) Category-Selectivity in Human Visual Cortex Follows Cortical Topology: A Grouped icEEG Study. *PLoS One* 11, e0157109

160 Devlin, J.T. *et al.* (2000) Susceptibility-Induced Loss of Signal: Comparing PET and fMRI on a Semantic Task. *Neuroimage* 11, 589–600

161 Gomez, J. *et al.* (2018) Development differentially sculpts receptive fields across early and high-level human visual cortex. *Nat. Commun.* 9, 788

162 Dehaene-Lambertz, G. *et al.* (2018) The emergence of the visual word form: Longitudinal evolution of category-specific ventral visual areas during reading acquisition. *PLoS Biol.* 16, e2004103

163 Lachaux, J.P. *et al.* (2005) The many faces of the gamma band response to complex visual stimuli. *Neuroimage* 25, 491–501

164 Conner, C.R. *et al.* (2011) Variability of the relationship between electrophysiology and BOLD-fMRI across cortical regions in humans. *J. Neurosci.* 31, 12855–12865

165 White, A.L. *et al.* (2019) Parallel spatial channels converge at a bottleneck in anterior word-selective cortex. *Proc. Natl. Acad. Sci. U. S. A.* 116, 10087–10096

166 Ben-Shachar, M. *et al.* (2007) Differential Sensitivity to Words and Shapes in Ventral Occipito-Temporal Cortex. *Cereb. Cortex* 17, 1604–1611

167 Lochy, A. *et al.* (2018) Selective visual representation of letters and words in the left ventral occipito-temporal cortex with intracerebral recordings. *Proc. Natl. Acad. Sci. U. S. A.* 115, E7595–E7604

168 Cohen, L. *et al.* (2002) Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain* 125, 1054–1069

169 Dehaene, S.C. *et al.* (2002) The visual word form area: a prelexical representation of visual words in the fusiform gyrus. *Neuroreport* 13, 321–325

170 Price, C.J. and Devlin, J.T. (2003) The myth of the visual word form area. *Neuroimage* 19, 473–481

171 Geissler, A. *et al.* (2005) Influence of fMRI smoothing procedures on replicability of fine scale motor localization. *Neuroimage* 24, 323–331

172 Flinker, A. *et al.* (2011) Sub-centimeter language organization in the human temporal lobe. *Brain Lang.* 117, 103–109

173 Krauss, G.L. *et al.* (1996) Cognitive effects of resecting basal temporal language areas. *Epilepsia* 37, 476–483

174 Fonseca, A.T. Da *et al.* (2009) Electrophysiological study of the basal temporal language area: A convergence zone between language perception and production networks. *Clin. Neurophysiol.* 120, 539–550

175 Ray, S. and Maunsell, J.H.R. (2011) Different Origins of Gamma Rhythm and High-Gamma Activity in Macaque Visual Cortex. 9, e1000610

176 Miller, K.J. Broadband spectral change: Evidence for a macroscale correlate of population firing rate? , *Journal of Neuroscience*, 30. (2010) , 6477–6479

177 Fiez, J.A. and Petersen, S.E. (1998) Neuroimaging studies of word reading. *Proc. Natl. Acad. Sci. U. S. A.* 95, 914–921

178 Kutas, M. and Hillyard, S.A. (1980) Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science.* 207, 203–205

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

180 Friedman, D. *et al.* (1975) The late positive component (P300) and information processing in sentences. *Electroencephalogr. Clin. Neurophysiol.* 38, 255–262

181 Arbel, Y. *et al.* (2011) The N400 and the P300 are not all that independent. *Psychophysiology* 48, 861–875

182 Mo, C. *et al.* (2018) Attention priority map of face images in human early visual cortex. *J. Neurosci.* 38, 149–157

183 Gao, X. *et al.* (2018) Fast periodic stimulation (FPS): a highly effective approach in fMRI brain mapping. *Brain Struct. Funct.* 223, 2433–2454

184 Zhen, Z. *et al.* (2015) Quantifying interindividual variability and asymmetry of face-selective regions: A probabilistic functional atlas. *Neuroimage* 113, 13–25

185 McCandliss, B.D. *et al.* (2003) The visual word form area: expertise for reading in the fusiform gyrus. *Trends Cogn. Sci.* 7, 293–299

186 Devlin, J.T. *et al.* (2006) The role of the posterior fusiform gyrus in reading. *J. Cogn. Neurosci.* 18, 911–22

187 Pflugshaupt, T. *et al.* (2009) About the role of visual field defects in pure alexia. *Brain* 132, 1907–1917

188 Leff, A.P. *et al.* (2001) The functional anatomy of single-word reading in patients with hemianopic and pure alexia. *Brain* 124, 510–21

189 Shaywitz, B.A. *et al.* (2004) Development of left occipitotemporal systems for skilled

reading in children after a phonologically- based intervention. *Biol. Psychiatry* 55, 926–933

190 Schlaggar, B.L. and McCandliss, B.D. (2007) Development of Neural Systems for Reading. *Annu. Rev. Neurosci.* 30, 475–503

191 Brem, S. *et al.* (2010) Brain sensitivity to print emerges when children learn letter-speech sound correspondences. *Proc. Natl. Acad. Sci. U. S. A.* 107, 7939–44

192 Ben-Shachar, M. *et al.* (2011) The Development of Cortical Sensitivity to Visual Word Forms. *J. Cogn. Neurosci.* 23, 2387–2399

193 Glezer, L.S. *et al.* (2015) Adding words to the brain's visual dictionary: novel word learning selectively sharpens orthographic representations in the VWFA. *J. Neurosci.* 35, 4965–72

194 Glezer, L.S. *et al.* (2009) Evidence for Highly Selective Neuronal Tuning to Whole Words in the "Visual Word Form Area." *Neuron* 62, 199–204

195 Binder, J.R. *et al.* (2006) Tuning of the human left fusiform gyrus to sublexical orthographic structure. *Neuroimage* 33, 739–748

196 Baeck, A. *et al.* (2015) Influence of lexical status and orthographic similarity on the multi-voxel response of the visual word form area. *Neuroimage* 111, 321–328

197 Plaut, D.C. *et al.* (1996) Understanding Normal and Impaired Word Reading: Computational Principles in Quasi-Regular Domains. *Psychol. Rev.* 103, 56–115

198 Price, C.J. and McCrory, E. (2008) Functional Brain Imaging Studies of Skilled Reading and Developmental Dyslexia. In *The Science of Reading: A Handbook* (1st edn) (Snowling, M. J. and Hulme, C., eds), pp. 473–496, Blackwell Publishing

199 Balota, D.A. *et al.* The english lexicon project. , *Behavior Research Methods*, 39. (2007) , 445–459

200 Taulu, S. and Hari, R. (2009) Removal of magnetoencephalographic artifacts with temporal signal-space separation: Demonstration with single-trial auditory-evoked responses. *Hum. Brain Mapp.* 30, 1524–1534

201 Gramfort, A. *et al.* (2014) MNE software for processing MEG and EEG data. *Neuroimage* 86, 446–460

202 Lin, F.-H. *et al.* (2006) Distributed current estimates using cortical orientation constraints. *Hum. Brain Mapp.* 27, 1–13

203 Lin, F.H. *et al.* (2006) Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. *Neuroimage* 31, 160–171

204 Fan, R.-E. *et al.* (2008) LIBLINEAR: A Library for Large Linear Classification. *J. Mach. Learn. Res.* 9, 1871–1874

205 Weiss, S. and Mueller, H.M. (2003) The contribution of EEG coherence to the investigation of language. *Brain Lang.* 85, 325–343

206 Spironelli, C. and Angrilli, A. (2010) Developmental aspects of language lateralization in delta, theta, alpha and beta EEG bands. *Biol. Psychol.* 85, 258–267

207 Weiss, S. and Mueller, H.M. (2012) "Too Many betas do not Spoil the Broth": The Role of Beta Brain Oscillations in Language Processing. *Front. Psychol.* 3, 201

208 Duñabeitia, J.A. *et al.* (2009) N250 effects for letter transpositions depend on lexicality: 'casual' or 'causal'? *Neuroreport* 20, 381–387

209 Morris, J. *et al.* (2007) Semantic transparency and masked morphological priming: An ERP investigation. *Psychophysiology* 44, 506–521

210 Malone, P.S. *et al.* (2016) Multivariate Pattern Analysis Reveals Category-Related Organization of Semantic Representations in Anterior Temporal Cortex. *J. Neurosci.* 36, 10089–96

211 Vandenberghe, R. *et al.* (2002) The response of left temporal cortex to sentences. *J. Cogn. Neurosci.* 14, 550–560

212 Li, Y. *et al.* (2017) Multi-Connection Pattern Analysis: Decoding the representational content of neural communication. *Neuroimage* 162, 32–44

213 Bub, D.N. *et al.* (1993) Jules Dejerine and His Interpretation of Pure Alexia. *Brain Lang.* 45, 531–559

214 Damasio, A.R. (1979) Wernicke's Works on Aphasia: A Sourcebook and Review. *Arch. Neurol.* 36, 324–324

215 Roberts, D.J. *et al.* (2013) Efficient Visual Object and Word Recognition Relies on High Spatial Frequency Coding in the Left Posterior Fusiform Gyrus: Evidence from a Case-Series of Patients with Ventral Occipito-Temporal Cortex Damage. *Cereb. Cortex* 23, 2568–2580

216 Yong, K.X.X. *et al.* (2013) Intact reading in patients with profound early visual dysfunction. *Cortex* 49, 2294–2306

217 Kherif, F. *et al.* (2011) Automatic Top-Down Processing Explains Common Left Occipito-Temporal Responses to Visual Words and Objects. *Cereb. Cortex* 21, 103–114

218 Kinsbourne, M. and Warrington, E.K. (1963) The localizing significance of limited simultaneous visual form perception. *Brain* 86, 697–701

219 Warrington, E.K. and Shallice, T. (1980) Word-form dyslexia. *Brain* 103, 99–112

220 Behrmann, M. *et al.* (1998) A literature review and new data supporting an interactive

account of letter-by-letter reading. *Cogn. Neuropsychol.* 15, 7–51

221 Petersen, S.E. *et al.* (1988) Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature* 331, 585–589

222 Mccarthy, G. *et al.* (1993) Echo-planar magnetic resonance imaging studies of frontal cortex activation during word generation in humans. *Proc. Natl. Acad. Sci. U. S. A.* 90, 4952–4956

223 Cloppenborg, T. *et al.* (2021) Reading and the visual word form area (VWFA) – Management and clinical experience at one epilepsy surgery center. *Epilepsy Behav.* 124, 108274

224 Dehaene, S. and Cohen, L. (2007) Cultural Recycling of Cortical Maps. *Neuron* 56, 384–398

225 Ghuman, A.S. and Fiez, J.A. (2018) Parcellating the structure and function of the reading circuit. *Proc. Natl. Acad. Sci. U. S. A.* 115, 10542–10544

226 Vogel, A.C. *et al.* (2012) The Left Occipitotemporal Cortex Does Not Show Preferential Activity for Words. *Cereb. Cortex* 22, 2715–2732

227 Polk, T.A. and Farah, M. (2002) Functional MRI evidence for an abstract, not perceptual, word-form area. *J. Exp. Psychol. Gen.* 131, 65–72

228 Barton, J.J.S. *et al.* (2010) Reading words, seeing style: The neuropsychology of word, font and handwriting perception. *Neuropsychologia* 48, 3868–3877

229 Yeh, F.C. *et al.* (2016) Connectometry: A statistical approach harnessing the analytical potential of the local connectome. *Neuroimage* 125, 162–171

230 Fonov, V. *et al.* (2011) Unbiased average age-appropriate atlases for pediatric studies. *Neuroimage* 54, 313–327

231 Waters, D. *et al.* (2007) Fingerspelling, signed language, text and picture processing in deaf native signers: The role of the mid-fusiform gyrus. *Neuroimage* 35, 1287–1302

232 Martin, L. *et al.* (2019) The VWFA is the home of orthographic learning when houses are used as letters. *eNeuro* 6, 0425–17

233 Moore, M.W. *et al.* (2014) Learning to Read an Alphabet of Human Faces Produces Left-lateralized Training Effects in the Fusiform Gyrus. *J. Cogn. Neurosci.* 26, 896–913

234 Siuda-Krzywicka, K. *et al.* (2016) Massive cortical reorganization in sighted braille readers. *Elife* 5, e10762

235 Kim, J.S. *et al.* (2017) Development of the visual word form area requires visual experience: Evidence from blind braille readers. *J. Neurosci.* 37, 11495–11504

236 Coltheart, M. (2006) Dual route and connectionist models of reading: an overview. *London*

Rev. Educ. 4, 5–17

237 Castles, A. and Coltheart, M. (1993) Varieties of developmental dyslexia. *Cognition* 47, 149–180

238 Hickok, G. and Poeppel, D. (2007) The cortical organization of speech processing. *Nat. Rev. Neurosci. 2007 85* 8, 393–402

239 Coltheart, M. *et al.* (2001) DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychol. Rev.* 108, 204–256

240 Rumelhart, D. (2013) Toward an Interactive Model of Reading. In *Theoretical Models and Processes of Reading* pp. 719–747

241 Seidenberg, M.S. (2005) Connectionist models of word reading. *Curr. Dir. Psychol. Sci.* 14, 238–242

242 Campbell, R. *et al.* (1986) Face recognition and lipreading: A neurological dissociation. *Brain* 109, 509–521

243 Brédart, S. (2017) The cognitive psychology and neuroscience of naming people. *Neurosci. Biobehav. Rev.* 83, 145–154

244 Gullick, M.M. and Booth, J.R. (2015) The direct segment of the arcuate fasciculus is predictive of longitudinal reading change. *Dev. Cogn. Neurosci.* 13, 68–74

245 Vandermosten, M. *et al.* (2012) A tractography study in dyslexia: neuroanatomic correlates of orthographic, phonological and speech processing. *Brain* 135, 935–948

246 Yeatman, J.D. *et al.* (2011) Anatomical Properties of the Arcuate Fasciculus Predict Phonological and Reading Skills in Children. *J. Cogn. Neurosci.* 23, 3304–3317

247 Bouhali, F. *et al.* (2014) Anatomical Connections of the Visual Word Form Area. *J. Neurosci.* 34, 15402–15414

248 Grotheer, M. *et al.* (2021) White matter fascicles and cortical microstructure predict reading-related responses in human ventral temporal cortex. *Neuroimage* 227, 117669

249 Hoffman, P. *et al.* (2015) Triangulation of the neurocomputational architecture underpinning reading aloud. *Proc. Natl. Acad. Sci. U. S. A.* 112, E3719–E3728

250 Vandermosten, M. *et al.* (2012) A qualitative and quantitative review of diffusion tensor imaging studies in reading and dyslexia. *Neurosci. Biobehav. Rev.* 36, 1532–1552

251 Richlan, F. *et al.* (2009) Functional abnormalities in the dyslexic brain: A quantitative meta-analysis of neuroimaging studies. *Hum. Brain Mapp.* 30, 3299–3308

252 Silani, G. *et al.* (2005) Brain abnormalities underlying altered activation in dyslexia: a voxel based morphometry study. *Brain* 128, 2453–2461

253 Michon, M. *et al.* (2019) Origin and evolution of human speech: Emergence from a trimodal auditory, visual and vocal network. In *Progress in Brain Research* 250pp. 345–371

254 Erickson, L.C. *et al.* (2014) Distinct cortical locations for integration of audiovisual speech and the McGurk effect. *Front. Psychol.* 0, 534

255 van Laarhoven, T. *et al.* (2018) Audio-visual speech in noise perception in dyslexia. *Dev. Sci.* 21, e12504

256 De Gelder, B. and Vroomen, J. (1998) Impaired speech perception in poor readers: Evidence from hearing and speech reading. *Brain Lang.* 64, 269–281

257 Francisco, A.A. *et al.* (2018) Adult dyslexic readers benefit less from visual input during audiovisual speech processing: fMRI evidence. *Neuropsychologia* 117, 454–471

258 Albonico, A. and Barton, J.J.S. (2017) Face perception in pure alexia: Complementary contributions of the left fusiform gyrus to facial identity and facial speech processing. *Cortex* 96, 59–72

259 Lüders, H. *et al.* (1991) Basal temporal language area. *Brain* 114, 743–754

260 Schäffler, L. *et al.* (1994) Anatomic Distribution of Cortical Language Sites in the Basal Temporal Language Area in Patients with Left Temporal Lobe Epilepsy. *Epilepsia* 35, 525–528

261 Humphries, C. *et al.* (2006) Syntactic and Semantic Modulation of Neural Activity during Auditory Sentence Comprehension. *J. Cogn. Neurosci.* 18, 665

262 Noonan, K.A. *et al.* (2013) Going beyond Inferior Prefrontal Involvement in Semantic Control: Evidence for the Additional Contribution of Dorsal Angular Gyrus and Posterior Middle Temporal Cortex. *J. Cogn. Neurosci.* 25, 1824–1850

263 Hermann, B.P. *et al.* (1999) Visual confrontation naming following left anterior temporal lobectomy: A comparison of surgical approaches. *Neuropsychology* 13, 3–9

264 Ueno, T. *et al.* (2018) The Ventral Anterior Temporal Lobe has a Necessary Role in Exception Word Reading. *Cereb. Cortex* 28, 3035–3045

265 Woolnough, O. *et al.* (2021) Spatiotemporal dynamics of orthographic and lexical processing in the ventral visual pathway. *Nat. Hum. Behav.* 5, 389–398

266 Chan, A.M. *et al.* (2011) First-pass selectivity for semantic categories in human anteroventral temporal lobe. *J. Neurosci.* 31, 18119–18129

267 Woolnough, O. *et al.* (2021) A Spatiotemporal Map of Reading Aloud. *bioRxiv* DOI: 10.1101/2021.05.23.445307

268 Wilson, M.A. *et al.* (2012) The role of the left anterior temporal lobe in exception word

reading: Reconciling patient and neuroimaging findings. *Neuroimage* 60, 2000–2007

269 Tomasino, B. *et al.* (2020) Phonological and surface dyslexia in individuals with brain tumors: Performance pre-, intra-, immediately post-surgery and at follow-up. *Hum. Brain Mapp.* 41, 5015–5031

270 Tailby, C. *et al.* (2014) Reading difficulty is associated with failure to lateralize temporooccipital function. *Epilepsia* 55, 746–753

271 Tomasino, B. *et al.* (2015) Double-letter processing in surface dyslexia and dysgraphia following a left temporal lesion: A multimodal neuroimaging study. *Cortex* 73, 112–130

272 Hodges, J.R. *et al.* (1992) Semantic dementia: Progressive fluent aphasia with temporal lobe atrophy. *Brain* 115, 1783–1806

273 Peters, F. *et al.* (2009) Neural substrates of phonological and lexicosemantic representations in Alzheimer's disease. *Hum. Brain Mapp.* 30, 185

274 Bédos Ulvin, L. *et al.* (2017) Intracerebral stimulation of left and right ventral temporal cortex during object naming. *Brain Lang.* 175, 71–76

275 Binney, R.J. *et al.* (2016) Reading words and other people: A comparison of exception word, familiar face and affect processing in the left and right temporal variants of primary progressive aphasia. *Cortex* 82, 147–163

276 Gvion, A. and Friedmann, N. (2016) A principled relation between reading and naming in acquired and developmental anomia: Surface dyslexia following impairment in the phonological output lexicon. *Front. Psychol.* 7, 340

277 Gvion, A. and Friedmann, N. (2013) Developmental and Acquired Surface Dyslexia and Anomia as a Result of a Shared Deficit in Phonological Output Lexicon. *Procedia - Soc. Behav. Sci.* 94, 205–206

278 Woollams, A.M. *et al.* (2017) Laterality of anterior temporal lobe repetitive transcranial magnetic stimulation determines the degree of disruption in picture naming. *Brain Struct. Funct.* 222, 3749–3759

279 Ives-Deliperi, V.L. and Butler, J.T. (2012) Naming outcomes of anterior temporal lobectomy in epilepsy patients: A systematic review of the literature. *Epilepsy Behav.* 24, 194–198

280 JR, H. and J, K. (1998) Proper name anomia and anomia for the names of people: functionally dissociable impairments? *Cortex.* 34, 155–158

281 Yeatman, J.D. *et al.* (2013) Anatomy of the visual word form area: adjacent cortical circuits and long-range white matter connections. *Brain Lang.* 125, 146–55

282 Komatsu, H. and Wurtz, R.H. (1988) Relation of cortical areas MT and MST to pursuit eye

movements. I. Localization and visual properties of neurons. *J. Neurophysiol.* 60, 580–603

283 Kutas, M. and Federmeier, K.D. (2011) Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621

284 Maurer, D. *et al.* (2002) The many faces of configural processing. *Trends Cogn. Sci.* 6, 255–260

285 Veit, M.J. *et al.* (2021) Temporal order of signal propagation within and across intrinsic brain networks. *Proc. Natl. Acad. Sci. U. S. A.* 118,

286 Riesenhuber, M. and Glezer, L.S. (2017) Evidence for rapid localist plasticity in the ventral visual stream: The example of words. *Lang. Cogn. Neurosci.* 32, 286

287 Martin, A. (2006) Shades of Déjerine—Forging a Causal Link between the Visual Word Form Area and Reading. *Neuron* 50, 173–175

288 Arcaro, M.J. *et al.* (2020) Anatomical correlates of face patches in macaque inferotemporal cortex. *Proc. Natl. Acad. Sci. U. S. A.* 117, 32667–32678

289 Minagawa-Kawai, Y. *et al.* (2009) Prefrontal Activation Associated with Social Attachment: Facial-Emotion Recognition in Mothers and Infants. *Cereb. Cortex* 19, 284–292

290 Wang, X. *et al.* (2016) The Hierarchical Structure of the Face Network Revealed by Its Functional Connectivity Pattern. *J. Neurosci.* 36, 890–900

291 Almeida, J. *et al.* (2013) Tool Manipulation Knowledge is Retrieved by way of the Ventral Visual Object Processing Pathway. *Cortex.* 49, 2334

292 He, C. *et al.* (2013) Selectivity for large nonmanipulable objects in scene-selective visual cortex does not require visual experience. *Neuroimage* 79, 1–9

293 Baldassano, C. *et al.* (2013) Differential connectivity within the Parahippocampal Place Area. *Neuroimage* 75, 228–237

294 Devereux, B.J. *et al.* (2018) Integrated deep visual and semantic attractor neural networks predict fMRI pattern-information along the ventral object processing pathway. *Sci. Rep.* 8, 10636

295 Grill-Spector, K. *et al.* (2017) The Functional Neuroanatomy of Human Face Perception. *Annu. Rev. Vis. Sci.* 3, 167–196

296 Caspers, J. *et al.* (2014) Functional characterization and differential coactivation patterns of two cytoarchitectonic visual areas on the human posterior fusiform gyrus. *Hum. Brain Mapp.* 35, 2754

297 Weiner, K.S. *et al.* (2017) The Cytoarchitecture of Domain-specific Regions in Human

High-level Visual Cortex. *Cereb. Cortex* 27, 146–161

298 Weiner, K.S. *et al.* (2014) The mid-fusiform sulcus: A landmark identifying both cytoarchitectonic and functional divisions of human ventral temporal cortex. *Neuroimage* 84, 453–465

299 Yarbus, A.L. (1967) *Eye Movements and Vision*, Springer US.

300 Or, C.C.F. *et al.* (2015) Initial eye movements during face identification are optimal and similar across cultures. *J. Vis.* 15, 12–12

301 Clifton, C. *et al.* (2007) Eye movements in reading words and sentences. In *Eye Movements* pp. 341–371, Elsevier

302 Rayner, K. (1977) Visual attention in reading: Eye movements reflect cognitive processes. *Mem. Cogn. 1977 54* 5, 443–448

303 Rayner, K. (1998) Eye Movements in Reading and Information Processing: 20 Years of Research. *Psychol. Bull.* 124, 372–422

304 Henderson, J.M. *et al.* (2005) Eye movements are functional during face learning. *Mem. Cogn. 2005 331* 33, 98–106

305 Wang, L. *et al.* (2019) Individual face- and house-related eye movement patterns distinctively activate FFA and PPA. *Nat. Commun. 2019 101* 10, 1–16

306 Shomstein, S. *et al.* (2010) Top-down and bottom-up attentional guidance: Investigating the role of the dorsal and ventral parietal cortices. *Exp. Brain Res.* 206, 197–208

307 Ptak, R. (2012) The frontoparietal attention network of the human brain: Action, saliency, and a priority map of the environment. *Neuroscientist* 18, 502–515

308 Killian, N.J. *et al.* (2012) A map of visual space in the primate entorhinal cortex. *Nature* 491, 761–764

309 Preston, T.J. *et al.* (2013) Neural representations of contextual guidance in visual search of real-world scenes. *J. Neurosci.* 33, 7846–55

310 Foulsham, T. *et al.* (2009) Fixation and saliency during search of natural scenes: The case of visual agnosia. *Neuropsychologia* 47, 1994–2003

311 Hutsler, J. and Galuske, R.A.W. (2003) Hemispheric asymmetries in cerebral cortical networks. *Trends Neurosci.* 26, 429–435

312 Stigliani, A. *et al.* (2015) Temporal processing capacity in high-level visual cortex is domain specific. *J. Neurosci.* 35, 12412–12424