Feed-forward, feed-back, and distributed feature representation during visual word recognition revealed by human intracranial neurophysiology

Laura Long
Columbia University
Minda Yang
Columbia University
Nikolaus Kriegeskorte
Columbia University https://orcid.org/0000-0001-7433-9005
Joshua Jacobs
Columbia University https://orcid.org/0000-0003-1807-6882
Robert Remez
Barnard College, Columbia University
Michael Sperling
Thomas Jefferson University https://orcid.org/0000-0003-0708-6006
Ashwini Sharan
Jefferson Medical College
Bradley Lega
UT Southwestern Medical Center
Alexis Burks
University of Texas Southwestern Medical Center
Gregory Worrell
Mayo Clinic
Robert Gross
Emory University
Barbara Jobst
Geisel School of Medicine at Dartmouth
Kathryn Davis
University of Pennsylvania
Kareem Zaghloul
National Institutes of Health https://orcid.org/0000-0001-8575-3578
Sameer Sheth
Department of Neurosurgery, Baylor College of Medicine
Joel Stein

Hospital of the University of Pennsylvania

Sandhitsu Das

Hospital of the University of Pennsylvania

Richard Gorniak

Thomas Jefferson University Hospital

Paul Wanda

University of Pennsylvania

Michael Kahana

University of Pennsylvania

Nima Mesgarani (✉ nima@ee.columbia.edu)

Columbia University

Article

Keywords: visual word recognition, human intracranial neurophysiology, feed-forward processing, feedback processing, anatomically distributed processing

DOI: https://doi.org/10.21203/rs.3.rs-95141/v1

Abstract

Scientists debate where, when, and how different visual, orthographic, lexical, and semantic features are involved in visual word recognition. In this study, we investigate intracranial neurophysiology data from 151 patients engaged in reading single words. Using representational similarity analysis, we characterize the neural representation of a hierarchy of word features across the entire cerebral cortex. We find evidence of both feed-forward and feedback processing, with early representation of visual and lexical information in lingual gyrus followed by lexical representation in fusiform gyrus and semantic sensitivity in inferior frontal gyrus, with letter representation emerging later in fusiform gyrus. Furthermore, we observed a variety of anatomically heterogeneous temporal response shapes, and these functional populations had significant feature sensitivity. Taken together, our results demonstrate the early influence of lexical, phonological, and semantic features in visual word recognition and reveal feed-forward, feed-back, and anatomically distributed processing mechanisms that contribute to visual word recognition.

Introduction

Visual word recognition, the process of mapping the written form of an individual word to its underlying linguistic item, is a building block of fluent reading and critical to successful written communication. Although reading has been a major topic of research in human neuroimaging studies for decades, the neural mechanisms of visual word recognition are not yet fully understood. Behavioral studies have demonstrated that many features of a word may be relevant to its recognition, including its low-level visual contours; the identity of its letters in an alphabetic language (orthography); knowledge of how it would be pronounced aloud (phonology); information about the word's frequency of occurrence, similar words, and composition (lexical features); and its semantic meaning [1]. Of particular debate is the extent to which the process is feed-forward and sequential, first decoding visual features into a specific written item before processing higher-order features [2–6], or connectionist, processing visual, lexical, semantic, and other information concurrently to reach a decision about word identity [7,8].

A majority of visual word recognition studies to date have used noninvasive neuroimaging methods in humans to identify when and where different word features are represented in the brain. Studies using the high spatial resolution of fMRI and PET have implicated dorsal regions including occipital lobe, supramarginal gyrus, angular gyrus, and pars opercularis in inferior frontal cortex, and ventral regions including left fusiform gyrus, middle and anterior temporal lobe, and pars triangularis in inferior frontal cortex [8,9]. Of particular debate is the role of left fusiform gyrus, which shows selectivity to words and pseudowords over false fonts and consonant strings [2,10–13]. It has been hypothesized to contain an abstract representation of visual word identity and named "visual word form area" (VWFA) [2,14–16]. Some studies suggest that it decodes this identity in bottom-up fashion from orthography alone [2–6,16–20]. However, mounting evidence suggests it receives additional top-down linguistic input as well [7,21–23].

Meanwhile, EEG and MEG studies have produced some evidence of the putative timecourse of visual word recognition. The event-related potentials N150 in EEG and M170 in MEG are sensitive to orthographic stimuli over symbols in normal readers [24–26] but not in dyslexic children [27–30], while N250 is modulated by orthographic similarity and the phonological content of letters [24,31–34], and N400 is sensitive to semantics [35].

Although the literature paints a detailed picture of visual word recognition in either anatomical space or over time, the lack of simultaneous spatiotemporal resolution has left the flow of information through the brain subject to debate. Furthermore, individual studies rarely manipulate or explore a full hierarchy of word features (visual, phonological, lexical, semantic) at the same time, so theories of how these hierarchical features interact are often patched together from separate experiments.

To address these gaps, high simultaneous spatial and temporal resolution is required. Intracranial neurophysiology provides a direct measure of neural activity at this resolution by recording from electrodes implanted in the awake behaving human brain, either on the surface under the dura (electrocorticography, ECoG) or through the depths of the brain (stereotactic electroencephalography, sEEG); together these techniques are referred to as intracranial electroencephalography (iEEG). Existing iEEG studies of reading have shown gamma band activation to visually presented words [10,11,19,36–38] and have begun to exploit its high spatiotemporal resolution to investigate interactions between brain areas[39,40]. However, small patient populations and sparse anatomical coverage are a major limitation of most intracranial neurophysiology studies, as access to patients is limited and electrode placement is determined by clinical rather than research factors.

In this study, we address these limitations by investigating a large dataset of human intracranial recordings of 151 patients with extensive surface and depth coverage across occipital, temporal, parietal, and frontal lobes as they read 300 individual words. Furthermore, we simultaneously investigate the encoding of visual, phonological, lexical, and semantic word features, enabling direct comparison of these features in a single experiment. With these approaches, we propose to characterize the brain network underlying visual word recognition at high resolution to clarify the flow of information through the brain.

Results

We recorded intracranial electroencephalography (iEEG) from 151 subjects as they engaged in a free recall task designed to study memory. During the encoding phase of the task, subjects viewed twenty-five 12-word lists for a total of 300 unique words. Each word was presented for 1600 ms, followed by a jittered interstimulus interval (ISI) between 750-1000 ms (1A). After each list, the subjects engaged in a distractor math task for 20 seconds, then had 30 seconds for free recall. All 151 subjects included in this study completed the entire experiment at least once. We excluded electrodes that showed pathological activity or were located outside of occipital, temporal, parietal, and frontal lobes, leaving 10,949 electrodes considered for analysis. A map of electrode coverage (1B, subject coverage S1A) shows broad coverage across all areas, with up to 45 subjects and 80 electrodes per 1cm^3 in high-density areas such as inferior temporal lobe.

To investigate the neural response during visual word recognition, we first extracted the high gamma power (70-150Hz) of each electrode and *z*-scored these responses to baseline (when the screen was blank). To test task responsiveness, we ran a *t*-test between high gamma activity during word presentation and a blank screen. Electrodes with p<1e-8 were considered "task responsive" and kept for further analysis; 2775/10949 electrodes (25.34%) fit this criterion (S1B/C). We then calculated the average magnitude *t*-value and percentage of electrodes above threshold by anatomical subregion, where areas with at least 50 task-responsive electrodes were considered for further analysis. We observed task-responsive electrodes across all four lobes, with the strongest average responses and highest proportion of electrodes over threshold in occipital lobe. Cuneus had the highest percentage of task-responsive electrodes (72.55±4.44%), followed by lingual gyrus (70.11±4.94%) and middle occipital gyrus (69.86±3.81%); these regions are significantly higher than all others (S1E/F, unpaired *t*-test with FDR correction q = 0.05), but not significantly different from each other. Our observations that a large percentage of occipital lobe electrodes are task-responsive while task response is sparse across temporal, parietal, and frontal lobes are consistent with a previous memory study using an overlapping portion of this dataset[41]. Middle occipital gyrus had the highest average *t*-value (39.37±3.97), followed by cuneus (36.20±4.44) and lingual gyrus (21.48±2.46; 1C/D). Next-highest was fusiform gyrus (10.70±0.86), followed by additional parietal, frontal, and temporal areas with significantly lower average *t*-values (S1D, unpaired *t*-test with FDR correction q = 0.05).

Having identified these 2775 task-responsive electrodes, we investigated their average response to word presentation. A vido (SV1) of the average response of each electrode projected on the cortical surface shows activity beginning in occipital areas before the rest of the brain. Notably, our task-responsive electrodes include both those that show increased high gamma power to word presentation, referred to as "enhanced" responses (red values), and those that show decreased high gamma power to word presentation, referred to as "suppressed" responses (blue values). The video shows primarily enhanced responses in occipital lobe, with mixed responses in the rest of cortex.

Quantitative analysis of electrode latency (measured as the point that the average word response reached its absolute maximum) corroborated our observation that occipital subregions respond fastest; middle occipital gyrus peaks first (243.27±17.34ms), followed by cuneus (304.72±26.56ms), fusiform gyrus (352.15±20.95) and lingual gyrus (433.46±51.67ms). Additional temporal, parietal, and frontal regions respond significantly later (S2B, unpaired *t*-test with FDR correction q = 0.05), with average latencies between 500-750 ms. This sequence of latencies, progressing roughly from posterior to anterior subregions, is again consistent with a previous memory study using an overlapping portion of this dataset[41]. We also quantified the percentage of electrodes with enhanced versus suppressed responses in each anatomical area. Lingual gyrus (94.23±3.26%), middle occipital gyrus (92.86±2.61%), and fusiform gyrus (92.36±2.22%) skewed overwhelmingly to enhanced responses (S2C, unpaired *t*-test with FDR correction q = 0.05). Cuneus, parietal, and temporal subregions follow, with frontal regions lowest.

Superior frontal gyrus is the only subregion favoring suppression at 35.27±3.33% enhanced, significantly lower than all others (3C/D).

While latency and peak analyses quantify major features of each electrode's response, they do not fully describe the shape of the response over time. To fully exploit the high resolution of iEEG, we wanted to qualitatively assess the diversity of neural responses beyond their sign and latency. To do so, we conducted agglomerative hierarchical clustering of our task-responsive electrodes' average word responses. This algorithm employs a bottom-up, data-driven approach, building a hierarchy by initially considering each electrode to be its own cluster, then progressively merging clusters until all observations have been united. We observe nine clusters with a remarkable diversity of temporal response shapes, shown in a raster of all electrodes sorted by cluster (3A), as well as the average response of each cluster (3B). Among this rich variety of temporal response shapes, we see onset responses, responses sustained for the duration of the word, and transient responses at a variety of latencies. We further investigated the distribution of response shapes within each anatomical area (Fig. 3C) and found that occipital responses are concentrated in the early, enhanced clusters 1 and 2, and 44.76% of fusiform gyrus electrodes belong to cluster 2. Additional temporal, parietal, and frontal subregions are heterogeneous across response shapes, with no more than 35% of electrodes belonging to a single cluster (3C, S3B). Further, each cluster is anatomically heterogeneous, spread across many subregions (S3A). Because clinical factors dictate electrode coverage, no single subject has electrodes in all subregions. However, we do observe heterogeneous response shapes across regions within individuals as well (S3C-F).

Previous iEEG studies have demonstrated language-related stimulus encoding, including representation of the phonetic features of speech in superior temporal gyrus [42] and of individual letters in occipital cortex [43]. To investigate how the words were encoded in different brain areas at different time points in our data, we used representational similarity analysis (RSA [44]) to interrogate stimulus representation in different anatomical areas over time. First, we collected a hierarchy of feature information about each word (see methods), including its bitmap (the image presented on screen), letters (a count of how many times each letter occurred in each word), phonemes (a count of how many times each phoneme occurred in each word), number of close orthographic neighbors (the number of words that can be created by changing a single letter, also known as orthographic neighborhood density), word frequency (how often it occurs in the English language), and semantic content (represented by the word2vec pre-trained word embedding) (4A-F). These features were selected to span visual, phonological, lexical, and semantic domains, and to minimize inter-feature correlations (S4). For each feature, we calculated the pairwise distances between stimuli to create a feature representational dissimilarity matrix (RDM). We also calculated neural RDMs across time for each anatomical area, taking pairwise distances across all task-sensitive electrode responses at each timestep. Lastly, we calculated the Spearman correlation between each neural RDM and each feature RDM; significant correlations suggest that the neural response contains information about the corresponding feature. Results are shown in Fig 4G-L; significant values are outlined in black (permutation with FDR correction q = 0.05).

This analysis allows us to evaluate feature sensitivity across the population of each brain area with high temporal resolution. Significant processing of the bitmap begins in lingual gyrus at 40 ms and is quickly followed by activation of orthographic neighborhood at 80 ms, earlier than previously reported (150 ms [45]). The next significant sensitivity to emerge is to word frequency at 120 ms in fusiform gyrus, in the same range as previously seen in ERPs [46–48], representing an early influence of lexical information. Next, semantic sensitivity begins in inferior frontal gyrus at 240 ms, later than the earliest reports of 160 ms [48–50], but compatible with the hypothesis that M250 could be an underlying component of the semantically-sensitive N400 [51]. From 280-360 ms, frequency information cascades through precentral, inferior frontal, and postcentral gyri (areas previously implicated in frequency processing [22]) while phoneme and letter information emerge in fusiform gyrus. After 400 ms, we observe a few distinct stages of processing: 1) late activation of low-level and orthographic visual features in middle occipital gyrus from 400-550 ms; 2) semantic representation in frontal lobe and middle temporal gyrus from 600-1000 ms; and 3) along the lateral sulcus, early frequency sensitivity is followed by letter, neighborhood, and bitmap information in postcentral gyrus, and by letter and phoneme information in precentral gyrus. This analysis provides a detailed account of feature encoding across anatomical areas in time. Notably, RSA allows us to interrogate the time course of feature sensitivity directly rather than relying on the latency of the high gamma response. For example, the average latency in fusiform gyrus is ~350 ms, but our results show that feature encoding begins much earlier for word frequency at ~120 ms.

Having observed a wide variety of temporal response shapes in our clustering analysis, we hypothesized that they could be functionally relevant to stimulus encoding. To test this hypothesis, we repeated our RSA analysis, this time segmenting electrode populations by their temporal response shape rather than their anatomical location (4M-R), again outlining significant values in black (permutation with FDR correction $q = 0.05$). We observe multi-feature representation in the fast, enhanced, occipital-heavy clusters. Cluster 1 represents phoneme at 60 ms, letter at 500 ms, and neighborhood at 820 ms. Cluster 2 displays significant sensitivity to bitmap, neighborhood, and phoneme within 100 ms of onset followed by frequency sensitivity at 200 ms, raising the possibility that it may serve as an early hub of multi-feature information. We also observe feature-specific sensitivity in the later, enhanced, transient clusters 4, 5, and 6: cluster 4 represents word frequency beginning at 160 ms; cluster 5 represents semantics at 180 ms; and finally, cluster 6 represents phoneme and letter more than a second after word onset. The suppressive clusters show only sparse significant representation, most notably to semantics in cluster 8 beginning at 400 ms. Interestingly, sustained clusters 3 and 9 show no significant sensitivity to any of our selected features. These results demonstrate that anatomically distributed electrode populations can carry stimulus information and suggest that distributed networks may play an important role in visual word recognition.

Discussion

Summary of findings

We found electrodes responsive to visual word presentation across occipital, temporal, parietal, and frontal lobes with a wide variety of response shapes, including both increases (enhancement) and decreases (suppression) in high gamma power. Responses in occipital lobe and fusiform gyrus were strongest, fastest, and had the highest percentage of enhanced responses, while frontal subregions had the slowest responses with the lowest percentage of enhanced responses. We further analyzed the diversity of temporal response shapes among electrodes using agglomerative clustering and found both sustained and transient responses at a variety of latencies. Last, we analyzed neural sensitivity to a hierarchy of word features including visual, phonological, lexical, and semantic features to provide a detailed account of stimulus encoding over time and space. Anatomically, we found early occipital representation of visual features, concurrent in lingual gyrus with sensitivity to word neighborhood size, followed shortly by sensitivity in fusiform gyrus to frequency, letter, and phoneme, sensitivity in inferior frontal gyrus to frequency and semantics, and late representation of several features along the lateral sulcus. Furthermore, we found that electrode populations with different temporal response shapes revealed in our clustering analysis showed significant encoding of stimulus features; we found a broadly-sensitive, enhanced cluster with early representation of visual, phonological, and lexical features; three separate mid-latency enhanced clusters with specialized sensitivity to frequency, semantics, and phonemes and letters, respectively; and sparse representation of semantics in a mid-latency suppressed cluster. Taken together, our results provide evidence of both feed-forward and feed-back processing during visual word recognition and demonstrate that stimulus encoding can be achieved by anatomically distributed networks.

Spatiotemporal feature sensitivity

Our results provide a detailed map of feature representation in time and space that builds upon an extensive literature of visual word recognition. To that end, it is useful to note how our observations compare with previous findings. Previous studies had implicated occipital lobe, fusiform gyrus, and the N150 in orthographic processing; we found low-level visual sensitivity occurring within 100 ms in lingual gyrus and in the early, enhanced functional cluster 2, which might underlie an early orthographic-specific ERP, though we do not see letter sensitivity in fusiform until 380 ms. Our observation of early phoneme sensitivity is plausibly consistent with the finding that the N250 is modulated by phonological content [24,31–34], while late phonological encoding in precentral gyrus is consistent with literature suggesting its involvement in phonological decoding [6,9,52]. Notably, we find that phonological sensitivity is similar but not precisely overlapping with letter sensitivity, and in fact precedes letter sensitivity in most areas. These similarities could be driven by the relatively high correlation between letter and phoneme information in English generally and in our stimulus set specifically (S3); a larger set of test words exhibiting contrasting attributes of each hierarchical type could better tease these factors apart in future study.

We observe substantial lexical sensitivity both for orthographic neighborhood size and word frequency. Neighborhood sensitivity largely overlaps with bitmap sensitivity in lingual gyrus and the broadly representative cluster 2, suggesting that visual processing may coincide with the activation of orthographically similar words, at higher spatial resolution and shorter latency than previously found (150 ms [45]). Our measure of frequency sensitivity is consistent with previous studies implicating word frequency in fusiform gyrus [22,53] and inferior frontal gyrus [22], and in the range of 100-200 ms [46–48]. We further demonstrate that frequency sensitivity is specifically represented by the anatomically distributed cluster 4 at 160 ms. Our results suggest that lexical features are represented early, robustly, and distributed across multiple brain areas.

Anatomically, semantic sensitivity is observed in areas that are reliably implicated in semantic processing, including inferior frontal gyrus, superior frontal gyrus, and middle temporal gyrus [9]. Further, we observe semantic sensitivity as early as 180 ms in the anatomically heterogeneous functional cluster 5, which could suggest an anatomically distributed semantic encoding as suggested by recent landmark neuroimaging studies [54,55]. This timing precedes the extensively-studied, semantically-sensitive N400 ERP [35], and is early enough to be compatible with the hypothesis that the N400 is the summation of M250 and M350 [51].

Interestingly, our analyses revealed several anatomical areas and a few clusters that did not show significant sensitivity to any of the selected features. There are several possible explanations for this. First, this may reflect a limitation of the experimental design; while the electrodes kept for analysis strongly respond to word presentation, our lack of a control task means that this activity may not be specific to word presentation, but may rather reflect more general visual or cognitive processing. Secondly, our analysis across time, space, and subjects with varying levels of noise may simply lack the statistical power to detect all effects in our data. Third, our feature list is not exhaustive, and it is possible that these areas are sensitive to aspects of the stimulus that were not tested here.

Feed-forward vs. connectionist

Evaluating the simultaneous spatiotemporal resolution of a hierarchy of features allows us to evaluate the flow of information through the brain and contribute to the debate between the feed-forward and connectionist accounts of visual word recognition. The strictest form of the feed-forward model posits a functional cascade in which visual features are first decoded into letters, then mapped to an item in the orthographic lexicon (likely in fusiform gyrus) before additional representations can be activated [2–6]. Notably, a study using an overlapping portion of this dataset found that the subsequent memory effect does appear to be distributed in a hierarchical, feed-forward stream[41]. However, our anatomical results are incompatible with such a strict feed-forward view of visual word recognition in several ways: 1) we observe phonological sensitivity as early as visual sensitivity; 2) sensitivity to orthographic neighborhood size occurs very shortly following bitmap sensitivity and preceding letter sensitivity; 3) fusiform gyrus shows frequency sensitivity at 180 ms, followed by sensitivity to phoneme at 280 ms and only then to letter at 380 ms. Additionally, in our analysis of cluster sensitivity, neighborhood and frequency sensitivity preceded letter sensitivity. Contrary to the strict feed-forward model, these results suggest an early role for lexical and phonological representations, especially in occipital cortex and fusiform gyrus. However, neither do our results suggest a completely integrated response, as we do observe some systematic separation of features in time and anatomical location. Anatomically, we see bitmap, neighborhood, and phoneme representation emerge within 100 ms, followed shortly by frequency encoding at 180 ms, semantic encoding at 240 ms, and letter encoding at 380 ms. Functionally, we observe clusters with early multi-feature representation, but also individual clusters with feature-specific sensitivity at middle latencies. Overall, these results suggest that the mechanisms underlying visual word recognition include both feed-forward and feed-back processing.

Role of fusiform gyrus

Fusiform gyrus has received special attention in the reading literature and as the center of two major debates: first, whether its function is specific to visual word processing and selective to word identities [2,8,11–15,56], and second, whether it serves as a strictly feed-forward hub of orthographic information [2,3,23,28,53,57,4–7,19–22]. In this study, we find significant fusiform sensitivity to frequency (consistent with previous studies[22,23,53]) as early as 120 ms, phoneme as early as 280 ms, and letter at 380 ms. Further study is needed to determine whether this lexical and phonological sensitivity arises from top-down or bottom-up influences on fusiform gyrus, but our results demonstrate that fusiform is not limited to encoding orthographic information. This result may be consistent with its role as sensitive to individual words. Further ECoG research using traditional false font paradigms could address this question more directly.

Anatomically distributed feature encoding

The spatiotemporal resolution of iEEG allowed us to decouple encoding networks in our data from anatomical boundaries. Using clustering analysis, we uncovered a variety of anatomically-heterogeneous temporal response shapes. Moreover, these functional populations had significant feature sensitivity, suggesting that anatomically-distributed networks can in fact be relevant to stimulus encoding. We find evidence for an early “hub” cluster which represents a combination of visual, phonological, and lexical features within 100 ms of word onset, which may feed forward to other functional groups. By 180 ms, both frequency and semantic information have been decoded in separate populations. Either by concurrent calculation or feedback, by 200 ms the hub cluster 2 has also represented word frequency. After this early, rapid processing, additional feature sensitivity emerges that could represent late-stage checking of the word identity [58]. Notably, we observed clusters that encode multiple features across different timepoints as well as clusters with sensitivity to specific features. The notion of distributed encoding has gained recent traction in the human language literature, especially in the semantic domain[54,55,59]. Further, distributed encoding could help explain the wide range of locations and timepoints implicated in our anatomical sensitivity analysis and in any literature that analyzes data by region of interest; if a particular feature is encoded by a network of similarly-behaving neurons spread across multiple areas, anatomical analyses may detect feature encoding in several involved areas.

Limitations & future work

Limitations of this work include nonuniform coverage across subjects and brain areas, limitations inherent to invasive neurophysiology in humans. Due to the limited time for testing and the experiment's dual role as a memory test, the range and quantity of stimuli were limited, and many subjects saw each word only one time; further work with repeated trials would enable useful decoding and reliability analyses. The task demands on the subject- to recall each list of words- may also impact the neural response; comparison with lexical decision or other reading tasks would allow assessment of the task impact on processing. Including useful controls such as false fonts, consonant strings, or pronounceable pseudowords would allow more fine-grained study of particular word features and neural mechanisms.

Conclusion

In this work, we studied visual word recognition in a large cohort of neurosurgical patients to characterize the flow of information in the brain. By analyzing electrode response properties and population sensitivity to a hierarchy of word features, we created a high-resolution map of stimulus encoding during single-word reading and showed evidence that anatomically distributed populations can encode word features. Our results suggest that feed-forward, feed-back, and distributed processing mechanisms contribute to visual word recognition.

Materials Methods

Intracranial recordings

Over a period of several years, patients undergoing chronic intracranial monitoring as treatment for pharmacologically intractable epilepsy were recruited to participate in the DARPA Restoring Active Memory program, a multi-center collaboration of neurology and neurosurgery departments across the United States. The research protocol was approved by the institutional review board at each hospital, and informed consent was obtained from all participants or their guardians. 151 subjects were included in this study. Clinical need determined the electrode number and placement for each subject; depending on patient need and technology specifications at each hospital, grid, strip, and depth electrodes were implanted and recorded with Bio-Logic, Natus, Nicolet, Grass Telefactor, or Nihon-Kohden EEG systems at 200, 256, 400, 500, 512, 1000, 1024, or 2000Hz. Electrodes were identified as hyperdense foci on the post-op CT scan and these points were then aligned to the MRI using Advanced Normalization Tools (ANTS, http://stnava.github.io/ANTs/). The cortex was parcellated using Freesurfer according to the Desikan–Kiliany atlas. To correct for post-operative brain shift, subdural electrode locations were projected to the cortical surface using the method of Dystra et al (https://pubmed.ncbi.nlm.nih.gov/22155045/). The automated anatomic parcellation and electrode localizations were confirmed by a neuroradiologist. We excluded electrodes located outside of brain tissue, showing pathological activity, or in limbic and sublobar regions. A total of 10949 eligible electrodes in occipital, temporal, parietal, and frontal lobes were considered for this study.

Behavioral task

Subjects completed a delayed free recall task. Each task session consisted of 25 12-word lists per session for a total of 300 words. Each word was presented in white, 70-point Verdana font on a black background for 1600 ms with a jittered 750-1000 ms ISI. After each list, subjects completed a distractor math task for 20 seconds, then had 30 seconds to verbally report as many words as they could remember from the last list. Patients who completed at least one complete task session in English were included in this study for a total of 151 subjects.

Neural data preprocessing

The amplitude of the high gamma (HG, 70-150Hz) response of each electrode was extracted using the Hilbert transform, then resampled to 100Hz. For subjects who completed the task more than once, data was averaged over repetitions after resampling. Finally, for clustering, latency, and peak calculation, each electrode's data were normalized by z-score to the response while viewing a blank screen; for sensitivity analyses, each electrode's data were normalized by z-score to its entire response.

Task responsiveness

At each electrode, we ran a *t*-test between the HG response during word presentation and the HG response when the screen was blank. We took the absolute value of all *t*-values to include electrodes with both enhanced and suppressed responses to word presentation; electrodes with $p < 1e\text{-}8$ (2775/10949) were considered "task responsive" and kept for further analysis. The percentage of task-responsive electrodes in each region was calculated as the ratio of electrodes with $p < 1e\text{-}8$ to the total number of electrodes in each region; subregions with at least 50 task-responsive electrodes were considered for further analysis. The average *t*-value was also calculated by subregion.

Brain plots

All brain plots were generated with Brainstorm [60], which is documented and freely available for download online under the GNU general public license (http://neuroimage.usc.edu/brainstorm). All plots used the Freesurfer average brain. For plots including all 10949 electrodes, electrode density was high enough that smoothing created a more intelligible plot (e.g. 1D vs S1B), so each electrode location was first converted from MNI to voxel space in nilearn [61] (https://nilearn.github.io/authors.html) before smoothing data over 1 cm^3 spans and converting back into MNI space. For coverage plots (1B, S1A), we counted the number of electrodes or subjects included in each 1 cm^3 and smoothed over .5 cm^3 before converting back to MNI space. The *t*-value plot (1D) was calculated in the same fashion, except values were averaged within each 1 cm^3. For remaining brain plots, only task-responsive electrodes were included, so voxel conversion was unnecessary. Instead, each individual electrode location was projected to the cortical surface of the average Freesurfer brain and plotted as a sphere using Brainstorm.

Latency and peak analysis

We calculated the peak amplitude of each electrode's average word response as the absolute maximum or minimum of the response, whichever was of greater magnitude. Latency was calculated as the sample after word onset where each electrode's average response reached its peak value. Histograms of latency and peak amplitude are plotted in S2A/C. Electrodes with positive peak values were classified as having an enhanced response, while those with negative peak values were classified as suppressive. We calculated the average latency (2A) and average enhancement vs. suppression ratio (2C) in each anatomical subregion, reporting standard error of the mean and conducting unpaired *t*-tests with FDR correction for statistical analysis (S2B/D). We plotted raw latencies and peak values on the brain using Brainstorm (2B/D).

Clustering analysis of temporal response shapes

To investigate the range of responses in the data, we averaged each electrode's HG response over words and conducted unsupervised agglomerative clustering (ward linkage, Euclidean distance) on the responses from 450 ms before word onset to 400 ms after word offset, smoothing with an 80 ms hanning window. Empirical inspection of different numbers of clusters revealed that using more than 9 clusters led to over clustering, with one or more clusters being further split into groups showing very similar response types. As gap, elbow, and other optimization methods attempted were unsuccessful, suggesting either 2 clusters or the maximum number tested, we chose to continue our analysis with 9 clusters. Clusters were first separated into enhanced and suppressive groups, then ordered by latency within each group. We plotted a raster of all electrodes organized by cluster (3A) as well as plots of each average cluster's waveform with standard deviation error bars (3B). We also calculated and plotted the percentage of electrodes in each anatomical lobe that belonged to each cluster, normalized by the total number of electrodes in each lobe (3C), as well as the percentage of electrodes in each cluster that belonged to each lobe, normalized by the total number of electrodes in each cluster (S3A). We plotted each electrode's cluster ID on the brain using Brainstorm, both for all subjects (S3B), as well as four example subjects with varying coverage, including right hemisphere (S3C), left hemisphere (S3D), bilateral (S3E), and occipital lobe implants (S3F).

Word feature catalog and feature RDMs

We collected feature data for all 300 words. Bitmaps were constructed by recreating exact images of each word that had been presented to the subjects (70-point white Verdana font on a black background, centered on the screen,) cropping to the smallest image where the longest word could still be fully captured, and resizing to a 6x29=174 pixel image. To create the feature RDM, each word's image was unrolled into a vector of length 174. Letter vectors were generated as a count of how many times each letter occurred in each word, deleting any letters that occurred fewer than 5 times across the entire stimulus set for a vector of length 23. Analogously, phoneme vectors were a count of how many times each phoneme occurred in each word as listed in the P2FA dictionary[62], deleting any phonemes that occurred fewer than 5 times across the entire stimulus set for a vector of length 34. Orthographic neighborhood size, also called number of neighbors, is defined as the "number of words that can be obtained by changing one letter while preserving the identity and positions of the other letters" and was collected from the English Lexicon Project[63]. Frequency data were also collected from the English Lexicon Project, using Hyperspace Analogue to Language (HAL) frequency norms based on approximately 131 million words[64]. Semantic content was represented by Google word2vec embeddings (https://code.google.com/archive/p/word2vec/ accessed using gensim https://radimrehurek.com/gensim/index.html), using 300-dimensional vectors pre-trained on Google News (100 billion tokens). These features were selected to span visual, phonological, lexical, and semantic domains, and to minimize inter-feature correlations.

For each feature, we calculated the pairwise distances between stimuli using cosine distance for multi-dimensional features (bitmap, letters, phonemes, semantics) and squared Euclidean distance for one-dimensional features (neighborhood, frequency) to create a feature representational dissimilarity matrix (RDM) with which to conduct representational similarity analysis (RSA [44]). We calculated the inter-feature correlations as the Spearman correlation between each pair of feature RDMs (S4).

Feature sensitivity

We conducted RSA using our feature RDMs for different electrode populations over time. To balance temporal resolution with computation time, responses were downsampled to 50Hz after smoothing with a 250 ms hanning window. For each anatomical area and timestep, we computed a neural RDM on the basis of all task-responsive electrodes in that area. The neural RDM for each timestep contains the response-pattern distances (Pearson correlation distance) for all pairs of stimuli. Next, we calculated the Spearman rank correlation between each neural RDM and each feature RDM. We used rank correlation to avoid the assumption of a linear relationship between the predicted and measured dissimilarities[44]. To conduct statistical inference, we permuted the stimulus labels 1000 times, recalculated neural RDMs with each permutation, and calculated the correlation between each permuted neural RDM and each intact feature RDM to obtain a permutation distribution of correlation values for every timestep, area, and feature. We used a generalized Pareto distribution to estimate the tail of each distribution to obtain a precise *p*-value for each comparison [65,66], then corrected for multiple comparisons by setting the false discovery rate across all areas, features, and timesteps to 0.05. Data were displayed by image, setting negative correlations to 0 (negative RSA correlations signify low representational similarity) and indicating significant values with a black box. Lastly, we repeated the entire procedure, this time segmenting electrodes into populations by temporal response shape (as assigned by the agglomerative clustering analysis) rather than by anatomical area.

Declarations

Author Contributions: LKL, MY, and NM conceived the study and analyzed data. LKL and NM wrote the manuscript. MS, AS, BL, AB, GW, RG, BJ, KD, KAZ, SS, JS, SD, RG, PW, and MK were part of the multi-hospital team involved in data collection, processing, and distribution. NK, JJ, and RR provided guidance and feedback on analysis approach and interpretation of results. All authors reviewed the manuscript.

Acknowledgments: This work was funded by the National Science Foundation Graduate Research Fellowship DGE 16-44869, the National Science Foundation Integrative Graduate Education and Research Traineeship: From Data to Solutions at Columbia University, and National Science Foundation Grant SBE 1827361.

References

1. Massaro, D. W. & Cohen, M. M. Visual, Orthographic, Phonological, and Lexical Influences in Reading. *J. Exp. Psychol. Hum. Percept. Perform.* **20**, 1107–1128 (1994).
2. Dehaene, S., Le Clec'H, G., Poline, J. B., Le Bihan, D. & Cohen, L. The visual word form area: A prelexical representation of visual words in the fusiform gyrus. *Neuroreport* **13**, 321–325 (2002).
3. Dehaene, S., Cohen, L., Sigman, M. & Vinckier, F. The neural code for written words: A proposal. *Trends Cogn. Sci.* **9**, 335–341 (2005).
4. Dehaene, S. & Cohen, L. The unique role of the visual word form area in reading. *Trends Cogn. Sci.* **15**, 254–262 (2011).
5. Jobard, G., Crivello, F. & Tzourio-Mazoyer, N. Evaluation of the dual route theory of reading: A metanalysis of 35 neuroimaging studies. *Neuroimage* **20**, 693–712 (2003).
6. Levy, J. *et al.* Testing for the dual-route cascade reading model in the brain: An fMRI effective connectivity account of an efficient reading style. *PLoS One* **4**, (2009).
7. Price, C. J. & Devlin, J. T. The Interactive Account of ventral occipitotemporal contributions to reading. *Trends in Cognitive Sciences* vol. 15 246–253 (2011).
8. Allison, T. *et al.* The neural code for written words: A proposal. *Neuroimage* **9**, 254–262 (2014).
9. Price, C. J. A review and synthesis of the first 20years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage* **62**, 816–847 (2012).
10. Allison, T., Mccarthy, G., Nobre, A., Puce, A. & Belger, A. Human extrastriate visual cortex and the perception of faces, words, numbers, and colors. *Cereb. Cortex* **4**, 544–554 (1994).
11. Nobre, A. C., Allison, T. & McCarthy, G. Word recognition in the human inferior temporal lobe. *Nature* **372**, 260–263 (1994).
12. Baker, C. I. *et al.* Visual word processing and experiential origins of functional selectivity in human extrastriate cortex. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 9087–9092 (2007).
13. Glezer, L. S., Jiang, X. & Riesenhuber, M. Evidence for Highly Selective Neuronal Tuning to Whole Words in the 'Visual Word Form Area'. *Neuron* **62**, 199–204 (2009).
14. Cohen, L. *et al.* 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 (2000).
15. Cohen, L. *et al.* Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain* **125**, 1054–1069 (2002).
16. Cohen, L. & Dehaene, S. Specialization within the ventral stream: The case for the visual word form area. *NeuroImage* vol. 22 466–476 (2004).

17. Norris, D., McQueen, J. M. & Cutler, A. Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences* vol. 23 (2000).

18. Solomyak, O. & Marantz, A. Evidence for early morphological decomposition in visual word recognition. *J. Cogn. Neurosci.* **22**, 2042–2057 (2010).

19. Sahin, N. T., Pinker, S., Cash, S. S., Schomer, D. & Halgren, E. Sequential processing of lexical, grammatical, and phonological information within broca's area. *Science (80-.).* **326**, 445–449 (2009).

20. Schurz, M. *et al.* Top-down and bottom-up influences on the left ventral occipito-temporal cortex during visual word recognition: An analysis of effective connectivity. *Hum. Brain Mapp.* **35**, 1668–1680 (2014).

21. Price, C. J. & Devlin, J. T. The myth of the visual word form area. *NeuroImage* vol. 19 473–481 (2003).

22. Kuo, W. J. *et al.* Frequency effects of Chinese character processing in the brain: An event-related fMRI study. *Neuroimage* **18**, 720–730 (2003).

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

24. Duñabeitia, J. A., Dimitropoulou, M., Grainger, J., Hernández, J. A. & Carreiras, M. Differential sensitivity of letters, numbers, and symbols to character transpositions. *J. Cogn. Neurosci.* **24**, 1610–1624 (2012).

25. Maurer, U., Brandeis, D. & McCandliss, B. D. Fast, visual specialization for reading in English revealed by the topography of the N170 ERP response. *Behav. Brain Funct.* **1**, (2005).

26. Tarkiainen, A., Helenius, P., Hansen, P. C., Cornelissen, P. L. & Salmelin, R. Dynamics of letter string perception in the human occipitotemporal cortex. *Brain* **122**, 2119–2131 (1999).

27. Helenius, P. *et al.* Cortical activation during spoken-word segmentation in nonreading-impaired and dyslexic adults. *J. Neurosci.* **22**, 2936–2944 (2002).

28. Simos, P. G. *et al.* Dyslexia-specific brain activation profile becomes normal following successful remedial training. *Neurology* **58**, 1203–1213 (2002).

29. Simos, P. G. *et al.* Early development of neurophysiological processes involved in normal reading and reading disability: A magnetic source imaging study. *Neuropsychology* **19**, 787–798 (2005).

30. Simos, P. G. *et al.* Altering the Brain Circuits for Reading Through Intervention: A Magnetic Source Imaging Study. *Neuropsychology* **21**, 485–496 (2007).

31. Carreiras, M., Vergara, M. & Perea, M. ERP correlates of transposed-letter priming effects: The role of vowels versus consonants. *Psychophysiology* **46**, 34–42 (2009).

32. Carreiras, M., Duñabeitia, J. A. & Molinaro, N. Consonants and vowels contribute differently to visual word recognition: ERPs of relative position priming. *Cereb. Cortex* **19**, 2659–2670 (2009).

33. Carreiras, M., Vergara, M. & Perea, M. ERP correlates of transposed-letter similarity effects: Are consonants processed differently from vowels? *Neurosci. Lett.* **419**, 219–224 (2007).

34. Holcomb, P. J. & Grainger, J. Exploring the temporal dynamics of visual word recognition in the masked repetition priming paradigm using event-related potentials. *Brain Res.* **1180**, 39–58 (2007).

35. Kutas, M. & Federmeier, K. D. Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annu. Rev. Psychol.* **62**, 621–647 (2011).

36. Crone, N. E. *et al.* Electrocorticographic gamma activity during word production in spoken and sign language. *Neurology* **57**, 2045–2053 (2001).

37. Mainy, N. *et al.* Cortical dynamics of word recognition. *Hum. Brain Mapp.* **29**, 1215–1230 (2008).

38. Lochy, A. *et al.* 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 (2018).

39. Vidal, J. R. *et al.* Long-distance amplitude correlations in the high gamma band reveal segregation and integration within the reading network. *J. Neurosci.* **32**, 6421–6434 (2012).

40. Whaley, M. L., Kadipasaoglu, C. M., Cox, S. J. & Tandon, N. Modulation of orthographic decoding by frontal cortex. *J. Neurosci.* **36**, 1173–1184 (2016).

41. Kucewicz, M. T. *et al.* Human verbal memory encoding is hierarchically distributed in a continuous processing stream. *eNeuro* **6**, (2019).

42. Mesgarani, N. *et al.* Phonetic feature encoding in human superior temporal gyrus. *Science* **343**, 1006–10 (2014).

43. Jacobs, J. & Kahana, M. J. Neural representations of individual stimuli in humans revealed by gamma-band electrocorticographic activity. *J. Neurosci.* **29**, 10203–10214 (2009).

44. Nili, H. *et al.* A Toolbox for Representational Similarity Analysis. *PLoS Comput. Biol.* **10**, (2014).

45. Taler, V. & Phillips, N. A. Event-related brain potential evidence for early effects of neighborhood density in word recognition. *Neuroreport* **18**, 1957–1961 (2007).

46. Assadollahi, R. & Pulvermüller, F. Early influences of word length and frequency: A group study using MEG. *Neuroreport* **14**, 1183–1187 (2003).

47. Dambacher, M., Kliegl, R., Hofmann, M. & Jacobs, A. M. Frequency and predictability effects on event-related potentials during reading. *Brain Res.* **1084**, 89–103 (2006).

48. Hauk, O., Davis, M. H., Ford, M., Pulvermüller, F. & Marslen-Wilson, W. D. The time course of visual word recognition as revealed by linear regression analysis of ERP data. *Neuroimage* **30**, 1383–1400 (2006).

49. Pammer, K. *et al.* Visual word recognition: The first half second. *Neuroimage* **22**, 1819–1825 (2004).

50. Wheat, K. L., Cornelissen, P. L., Frost, S. J. & Hansen, P. C. During visual word recognition, phonology is accessed within 100 ms and may be mediated by a speech production code: Evidence from magnetoencephalography. *J. Neurosci.* **30**, 5229–5233 (2010).

51. Pylkkänen, L. & Marantz, A. Tracking the time course of word recognition with MEG. *Trends Cogn. Sci.* **7**, 187–189 (2003).

52. Richardson, F. M., Seghier, M. L., Leff, A. P., Thomas, M. S. C. & Price, C. J. Multiple routes from occipital to temporal cortices during reading. *J. Neurosci.* **31**, 8239–8247 (2011).
53. Kronbichler, M. *et al.* Taxi vs. taksi: On orthographic word recognition in the left ventral occipitotemporal cortex. *J. Cogn. Neurosci.* **19**, 1584–1594 (2007).
54. Huth, A. G., De Heer, W. A., Griffiths, T. L., Theunissen, F. E. & Gallant, J. L. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* **532**, 453–458 (2016).
55. Deniz, F., Nunez-Elizalde, A. O., Huth, A. G. & Gallant, J. L. The Representation of Semantic Information Across Human Cerebral Cortex During Listening Versus Reading Is Invariant to Stimulus Modality. *J. Neurosci.* **39**, 7722–7736 (2019).
56. Binder, J. R., Medler, D. A., Westbury, C. F., Liebenthal, E. & Buchanan, L. Tuning of the human left fusiform gyrus to sublexical orthographic structure. *Neuroimage* **33**, 739–748 (2006).
57. Solomyak, O. & Marantz, A. *Lexical access in early stages of visual word processing: A single-trial correlational MEG study of heteronym recognition*. *Brain and Language* vol. 108 (2009).
58. Lim, S. W. H. The Influence of Orthographic Neighborhood Density and Word Frequency on Visual Word Recognition: Insights from RT Distributional Analyses. *Front. Psychol.* **7**, 401 (2016).
59. de Heer, W. A., Huth, A. G., Griffiths, T. L., Gallant, J. L. & Theunissen, F. E. The hierarchical cortical organization of human speech processing. *J. Neurosci.* **37**, 6539–6557 (2017).
60. Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D. & Leahy, R. M. Brainstorm: A user-friendly application for MEG/EEG analysis. *Comput. Intell. Neurosci.* **2011**, (2011).
61. Abraham, A. *et al.* Machine learning for neuroimaging with scikit-learn. *Front. Neuroinform.* **8**, 14 (2014).
62. Yuan, J., Yuan, J. & Liberman, M. Speaker identification on the SCOTUS corpus. *Proc. Acoust. 2008* (2008).
63. Balota, D. A. *et al.* The english lexicon project. *Behavior Research Methods* vol. 39 445–459 (2007).
64. Lund, K. & Burgess, C. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behav. Res. Methods, Instruments, Comput.* **28**, 203–208 (1996).
65. Knijnenburg, T. A., Wessels, L. F. A., Reinders, M. J. T. & Shmulevich, I. Fewer permutations, more accurate P-values. *Bioinformatics* **25**, i161-8 (2009).
66. Winkler, A. M., Ridgway, G. R., Douaud, G., Nichols, T. E. & Smith, S. M. Faster permutation inference in brain imaging. *Neuroimage* **141**, 502–516 (2016).