scientific reports

Check for updates

OPEN

Distinct neural sources underlying visual word form processing as revealed by steady state visual evoked potentials (SSVEP)

Fang Wang[1], Blair Kaneshiro[1], C. Benjamin Strauber[1], Lindsey Hasak[1], Quynh Trang H. Nguyen[1], Alexandra Yakovleva[2,3], Vladimir Y. Vildavski[2,3], Anthony M. Norcia[2,3] & Bruce D. McCandliss[1,2,3]✉

EEG has been central to investigations of the time course of various neural functions underpinning visual word recognition. Recently the steady-state visual evoked potential (SSVEP) paradigm has been increasingly adopted for word recognition studies due to its high signal-to-noise ratio. Such studies, however, have been typically framed around a single source in the left ventral occipitotemporal cortex (vOT). Here, we combine SSVEP recorded from 16 adult native English speakers with a data-driven spatial filtering approach—Reliable Components Analysis (RCA)—to elucidate distinct functional sources with overlapping yet separable time courses and topographies that emerge when contrasting words with pseudofont visual controls. The first component topography was maximal over left vOT regions with a shorter latency (approximately 180 ms). A second component was maximal over more dorsal parietal regions with a longer latency (approximately 260 ms). Both components consistently emerged across a range of parameter manipulations including changes in the spatial overlap between successive stimuli, and changes in both base and deviation frequency. We then contrasted word-in-nonword and word-in-pseudoword to test the hierarchical processing mechanisms underlying visual word recognition. Results suggest that these hierarchical contrasts fail to evoke a unitary component that might be reasonably associated with lexical access.

Reading is a remarkable aspect of human cognitive development and is essential in everyday life. Through frequent exposure to printed words, visual specialization for letter strings is developed[1], and skilled readers can read around 250 words per minute[2]. Such high reading speed requires fast visual word recognition that is dependent on specialized visual processes and brain sources.

Functional magnetic resonance imaging (fMRI) studies have reliably localized an area of the left lateral ventral occipitotemporal cortex (vOT) that is particularly selective to printed words relative to other visual stimuli such as line drawings[3] and faces[4–6]. The sensitivity of the left vOT site to visual words is reproducible across different languages and fonts[7] and individuals[8,9], and is invariant to font, size, case, and even retinal location[10,11].

Numerous fMRI studies have proposed that the vOT follows a hierarchical posterior-to-anterior progression, with posterior regions being more selective to visual word form processing while anterior parts are more weighted to high-level word features[5,12]. The posterior-to-anterior gradient is accomplished by increasing neuron receptive fields from posterior occipital to anterior temporal regions, and as a result, the sensitivity of neurons hierarchically increases from letter fragments to individual letters, bigrams, trigrams, morphemes, and finally entire word forms[5]. A recent study[13] suggested that, in addition to sub-regions within the vOT, other regions of the language network (e.g., angular gyrus) are also involved in the rapid identification of word forms by transferring and integrating information from and towards the vOT. This study[13] further suggested that the posterior area of vOT is structurally connected to the intraparietal sulcus mostly through a bottom-up pathway while the middle/anterior area is connected to other language areas most likely through both feed-forward and -backward connections (see also Price & Devlin[14]).

[1]Graduate School of Education, Stanford University, Stanford, CA, USA. [2]Department of Psychology, Stanford University, Stanford, CA, USA. [3]Wu Tsai Neurosciences Institute, Stanford, CA, USA. ✉email: brucemc@stanford.edu

nature portfolio

1

In contrast to fMRI with its high spatial resolution, electroencephalography (EEG) can detect text-related brain electrical activity with high temporal resolution. Event-related potential (ERP) studies have characterized a component that peaks between 150 to 200 ms with an occipito-temporally negative and fronto-centrally positive topography, termed the visual word N170. This N170 component is typically larger for word and word-like stimuli than for visually controlled symbols[1,15,16]. During development, N170 sensitivity to printed words emerges when children learn print-speech sound correspondences, especially in alphabetic languages[17], within the first 2 years of formal education[16].

More recently, the steady-state visual evoked potential (SSVEP) paradigm has been used to investigate visual word recognition due to its high signal-to-noise (SNR) ratio. In contrast to typical ERP approaches demanding long inter-stimulus intervals, the SSVEP paradigm presents a sequence of stimuli at a fast periodic rate (e.g., 10 Hz, 100 ms per item). The presentation of temporally periodic stimuli elicits periodic responses at the predefined stimulation frequency and its harmonics (i.e., integer multiples of the stimulus frequency). Those periodic responses are referred to as SSVEPs because they are stable in amplitude and phase over time[18]. Importantly, the SSVEP paradigm can provide high SNR in only a few minutes of stimulation due to its small noise bandwidth. However, the SSVEP paradigm has long been limited to the field of low-level visual perception and attention (for a review, see[19]). Only recently, this paradigm has been extended to more complex visual stimuli processing, such as objects[20], faces[21–23], numerical quantities[24,25], text[26], letters[27], and words[28–31].

In this study, we seek to use this paradigm to address key, yet still unanswered, questions in neural studies of visual word recognition: How do we coordinate in real time the two very different types of neural computation involved in representing the visual forms of words versus their lexical-semantic representations? In particular, how do we study these conceptually distinct yet temporally overlapping levels of processing using EEG? How can we overcome limitations of approaches that rely on the assumption of "pure insertion"[32] to "subtract" visual feature processing from recognition of letter forms, and from processes of lexical-semantic access? One common approach used in previous work, including studies with fMRI, ERP, and SSVEP, is the use of hierarchical stimulus contrasts that attempt to isolate these levels of processing by relying on the assumption of "pure insertion". For example, a recent fMRI study[13] used "perceptual" contrasts (e.g., words–checkerboards, words–scrambled words) and "lexical" contrasts (e.g., words–pseudowords) to isolate brain activity associated with each distinct word recognition process. However, fMRI accumulates neural activity over the course of seconds and is unable to capture fast time-scale neural dynamics. ERP and SSVEP studies (e.g.,[28,33]) offering higher temporal resolution have also used similar hierarchical contrasts (words-pseudofonts, words-nonwords, and words-pseudowords). Yet this higher temporal resolution brings new challenges, as the processes of visual feature extraction, accessing learned representations of visual letter and word forms, and lexical semantic access unfold over overlapping time courses, producing topographic patterns that interact at the scalp level. For example, recent SSVEP studies[28] obtained responses to three contrasts (words–pseudofonts, words–nonwords, and words–pseudowords) with effects apparent in sensors located over the same left occipito-temporal region. Hence, resolving the question of how and when different types of information are processed and interact requires the ability to not only delineate the underlying brain sources, but also—and crucially—to characterize more clearly the information flow between them. The current study aimed to address these deeper challenges and gain a better understanding of how different levels of processing unfold during visual word recognition. Specifically, we first focused on the word-in-pseudofont contrast within an SSVEP paradigm. We then used a data-driven analysis method to distinguish two processes often proposed to be evoked by such a contrast—a visual form analysis of familiar letters and letters combinations within words, and a lexical-semantic process associated with recognition of familiar visual words. We then applied those same techniques to study contrasts of word-in-nonword and word-in-pseudoword to see if the same two sets of neural processes could be identified within the SSVEP data.

Additionally, previous SSVEP studies of higher-level processes have used different presentation paradigms including adaptation and "periodic deviant" approaches. The adaptation approach has been used to examine whether stimulus presentation locations affect perception, mainly in relation to holistic processing of faces, by comparing streams of upright or inverted faces with either the same or different identity[34]. In contrast, in the "periodic deviant" stimulation mode, a sequence of control stimuli are replaced by deviant stimuli at a predefined periodic rate. The deviant stimuli typically differ from the control stimuli in a particular aspect (e.g. word versus nonword, face versus object or intact versus scrambled images)[22,26]. The control and deviant stimuli can follow the "image alternation" mode, in which every other image in the periodic base rate is a deviant. For example, a base of 6 Hz with a 3-Hz deviant (6/2 = 3 Hz). Alternatively, the deviant stimuli can be embedded within a sub-multiple of the base stimulation rate that is greater than two[23,28–30]. For example, a base of 10 Hz with every 5th image (instead of every other image) being a deviant results in a deviant frequency at 2 Hz (base frequency 10 Hz divided by 5). However, to our knowledge, no study has directly compared the "image alternation" mode and the mode wherein deviant stimuli are presented at a sub-multiple greater than twice the base rate. Therefore, the current study compared these two modes to determine whether results converge and which elicits responses with a higher signal-to-noise ratio. Findings from this research may provide important considerations relevant to generalizability of findings and for designing future studies of early readers.

To date, most SSVEP work has focused on only a single source of word-related processing in the left hemisphere[28–30] by analyzing periodic responses from several pre-selected (literature-based) sensors. Of present interest, two SSVEP studies on text and letter recognition have revealed multiple underlying sources with different temporal dynamics and scalp topographies[26,27], either by defining different regions of interest[26] or employing a spatial filtering approach[27]. In contrast to data analyses of several pre-selected sensors that reduced the whole map of evoked data to a restricted and typically biased subset[35], spatial filtering approaches offer a purely data-driven alternative for selecting sensors. These methods compute weighted linear combinations across the full montage of sensors to capture and isolate different neural processes arising from distinct underlying cortical sources[36]. A number of linear spatial filters, such as Principal Components Analysis (PCA) and Common Spatial Patterns

	Condition 1	Condition 2	Condition 3	Condition 4	Condition 5
Stimuli	W in PF	W in PF	W in PF	W in NW	W in PW
Accuracy (% ± SD)	95.2 (± 12.7)	95.6 (± 11.4)	94.0 (± 12.7)	96.5 (± 10.4)	95.8 (± 12.0)
Reaction time (ms ± SD)	356.5 (± 45.1)	350.6 (± 40.7)	361.4 (± 38.6)	352.6(± 37.2)	356.1 (± 38.0)

Table 1. Accuracy and reaction time of responses to cross color change. Values are mean (±SD).

(CSP)[37], have been applied to SSVEP data, mainly in the brain–computer interface (BCI) field[38]. In cognitive neuroscience, a spatial filtering technique referred to as Reliable Components Analysis (RCA) has been increasingly used[27,39,40]. RCA derives a set of spatial components (i.e., spatial filters operationalized as topographic weights) that maximize across-trial or across-subject correlations ("reliability") while minimizing noise ("variance"). Specifically, RCA first discovers the optimal spatial filter weighting of the signal, then projects the data through this spatial filter to enable investigation of phase-locked topographic activities and to capture each temporal/topographical source of reliable signal across events and subjects (detailed information described in Refs.[39,40]).

Moreover, in conjunction with a recently developed RCA approach, Norcia et al.[41] for the first time estimated latencies of the SSVEP[41]. This was done by fitting a line through the phases at harmonics with significant responses; the slope of the line is interpreted as the response latency. Latency estimation of component(s) can provide insight into temporal dynamics of different processes located at different sources, extending on the majority of previous SSVEP studies, which have only focused on topographies and amplitudes[28–30].

Employing the spatial filtering component analysis (RCA) approach used in two recent studies[27,41], here we reproduced and extended a previous SSVEP study of French word processing[28], with periodic 4-letter English word versus pseudofont comparisons. Our central goal was to determine whether RCA would reveal multiple sources associated with word recognition, even though such sources might have overlapping time courses and topographies that might lead sensor-based investigations to posit a single source of activity reflecting a single process of word identification. A second goal was to explore whether such RCA-identified sources were specific to a narrow set of presentation parameters and frequencies, or whether such sources might emerge consistently and robustly across a range of experimental parameters (e.g., different stimulus presentation rates and jittered retinal locations). We were further interested in the potential RCA sources across a commonly used set of hierarchical stimulus contrasts that have often been thought to progressively isolate several subset processes underlying visual word recognition.

Specifically, the current study addresses these questions by presenting familiar words interspersed periodically among control stimuli (i.e., pseudofonts) in three conditions: (1) base stimulation frequency of 10 Hz consisting of a stream of pseudofont stimuli replaced by word deviants at 2 Hz; (2) same as in (1) but with stimulus presentation locations jittered around the center of the screen; and (3) same as in (1) yet with a base presentation frequency of 6 Hz and word deviants at 3 Hz. To investigate the hierarchy of visual word recognition, we included two additional stimulus contrasts: (4) word deviants embedded within a stream of nonwords, and (5) word deviants embedded in a stream of pseudowords. Each of these conditions is illustrated in Fig. 1.

Based on recent evidence[13,27], we hypothesized that word deviants among pseudofonts would elicit more than one neural discrimination source when subjected to RCA, producing at least two reliable signal sources with distinct temporal and topographical information. We sought to further investigate whether RCA component topographies and time-courses were specific to the particular experimental parameters or whether similar components could be elicited across a wider range of presentation rates and changes in stimulus locations. Finally, we wished to test a central assumption in previous reports[28] based on the notion that hierarchical aspects of visual word processing can be clearly isolated based on progressively specific contrasts of word-in-pseudofont, word-in-nonword, and word-in-pseudoword. Here, RCA provides a novel opportunity to first investigate distinct component topographies elicited from word-in-pseudofont contrasts, and then to investigate the hypothesis that word-in-nonword and word-in-pseudoword contrasts successfully isolate a subset of these sources.

Results

Behavioral results. For the color change detection task, the mean and standard deviation (SD) of accuracy and reaction time across all five conditions are summarized in Table 1. Separate one-way ANOVAs indicate that there was no significant difference across conditions in either accuracy ($F(4, 70) = 0.09$, $p = 0.98$) or reaction time ($F(4, 70) = 0.08$, $p = 0.99$). Thus, we concluded that participants were sufficiently engaged throughout the experiment. We did find one subject's accuracy was lower (53.8%) than others. But we still included this participant's data in our analysis after confirming that this participant showed a cerebral response to the base stimulation[25]. We observed that this participant's responses to the base rate were not an outlier (i.e., were within one standard deviation) compared with other participants.

Base analyses results. We performed RCA on responses at harmonics of the base frequency in order to investigate neural activity related to low-level visual processing; results are summarized in Fig. 2. Panel A displays topographic visualizations of the spatial filters for the maximally correlated components (RC1) for each condition (reliability explained: 37.5%). Here, all topographies show maximal weightings over medial occipital areas; correlations among components for conditions 1, 2, 4, and 5 are high ($r \geq 0.84$), and correlation for condition 3 is lower but still high ($r = 0.78$). The plots in Panel B present projected amplitudes (i.e., projecting data through the spatial filter) in bar plots, with statistically significant responses in the first harmonic ($p_{FDR} < 0.01$,
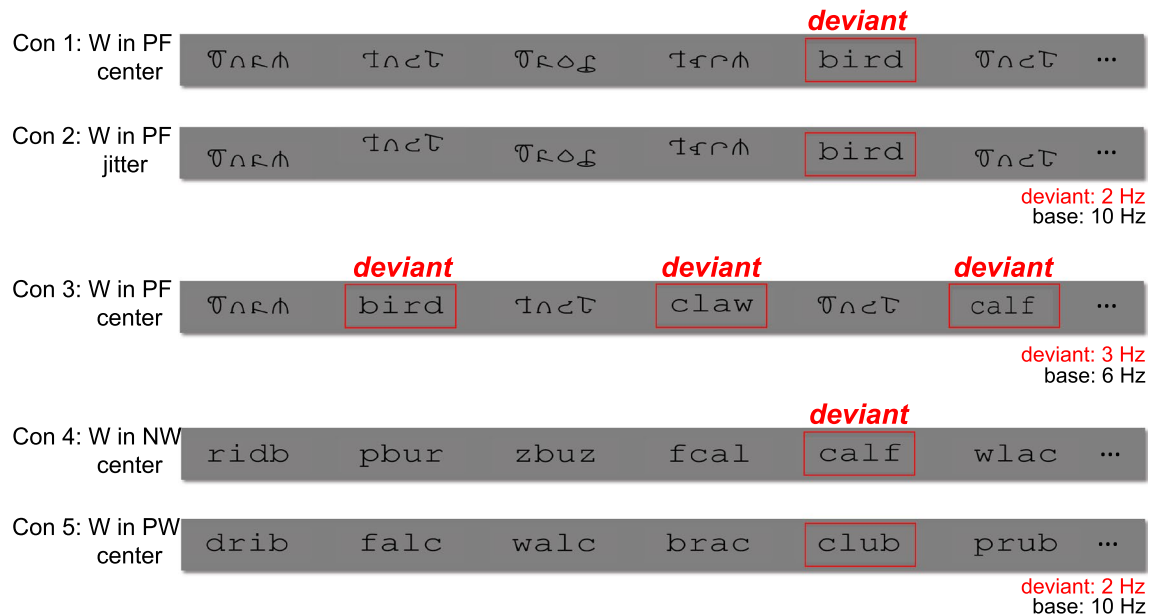
Figure 1. Examples of stimuli presented in the experiment. 2-Hz word deviants were embedded in a stream of pseudofont (W in PF) in conditions 1 and 2 to form a base stimulation rate of 10 Hz. Condition 2 used the same word and pseudofont stimuli as Condition 1, but spatially jittered their location on the monitor. Condition 3 presented the same stimuli used in Condition 1 but with 3-Hz deviant and 6-Hz base frequencies. 2-Hz word deviants were embedded in a stream of nonwords (W in NW) and pseudowords (W in PW) respectively, in conditions 4 and 5, to form a 10-Hz base stimulation frequency.
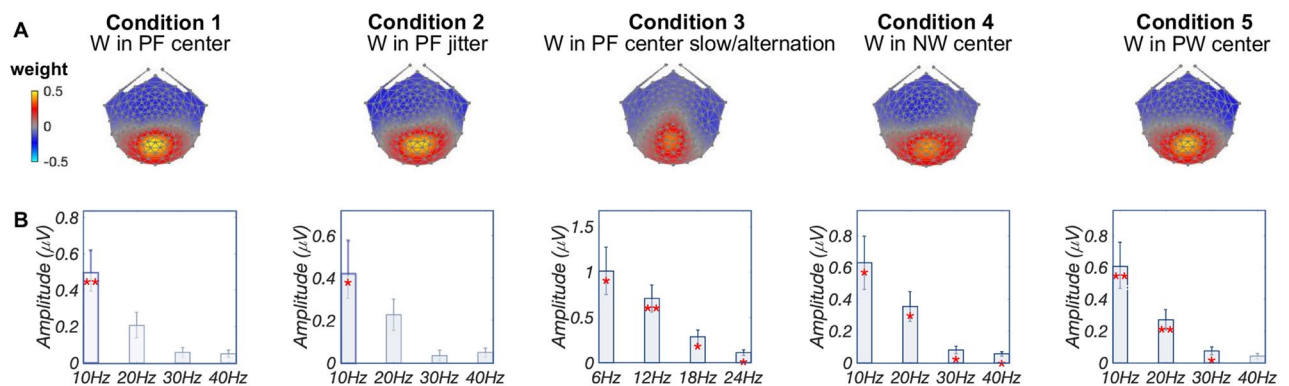

Figure 2. Base Analysis: low-level visual processing. (**A**) Topographic visualizations of the spatial filters for the maximally correlated component (RC1) for each condition. (**B**) Projected amplitude for each harmonic and condition in bar charts, $^*p_{FDR} < 0.05$, $^{**}p_{FDR} < 0.01$. Figures were created using publicly available MATLAB (R2020a) code: https://github.com/svndl/rcaExtra.

corrected for 4 comparisons) in condition 1 (word-in-pseudofont center); the first harmonic ($p_{FDR} < 0.05$) in condition 2 (word-in-pseudofont jitter); all four harmonics ($p_{FDR} < 0.05$) in conditions 3 (word-in-pseudofont center slower/alternation) and 4 (word-in-nonword center); and the first three harmonics ($p_{FDR} < 0.05$) in condition 5 (word-in-pseudoword center). Amplitude comparisons across conditions showed that there is no significant difference between conditions 1, 2, 4, and 5 ($F(3, 60) = 2.21$, $p = 0.09$), while amplitudes in condition 3 are significantly higher than other conditions ($p < 0.05$).

Deviant analyses results. Deviant analyses were conducted to investigate the mechanisms and sources underlying visual word processing. Given increasing evidence supporting segregations within vOT[13] and the discovery of distinct sources for text and letter recognition[27], we hypothesized that multiple sources should be captured during word-related processing (word-in-pseudofont). Due to previous observations of an apparent posterior-to-anterior gradient of responses to visually versus linguistically related sub-regions within vOT[12,13], we hypothesized that different condition manipulations (word-in-pseudofont, word-in-nonword, word-in-pseudoword) would evoke different response topographies and phases.
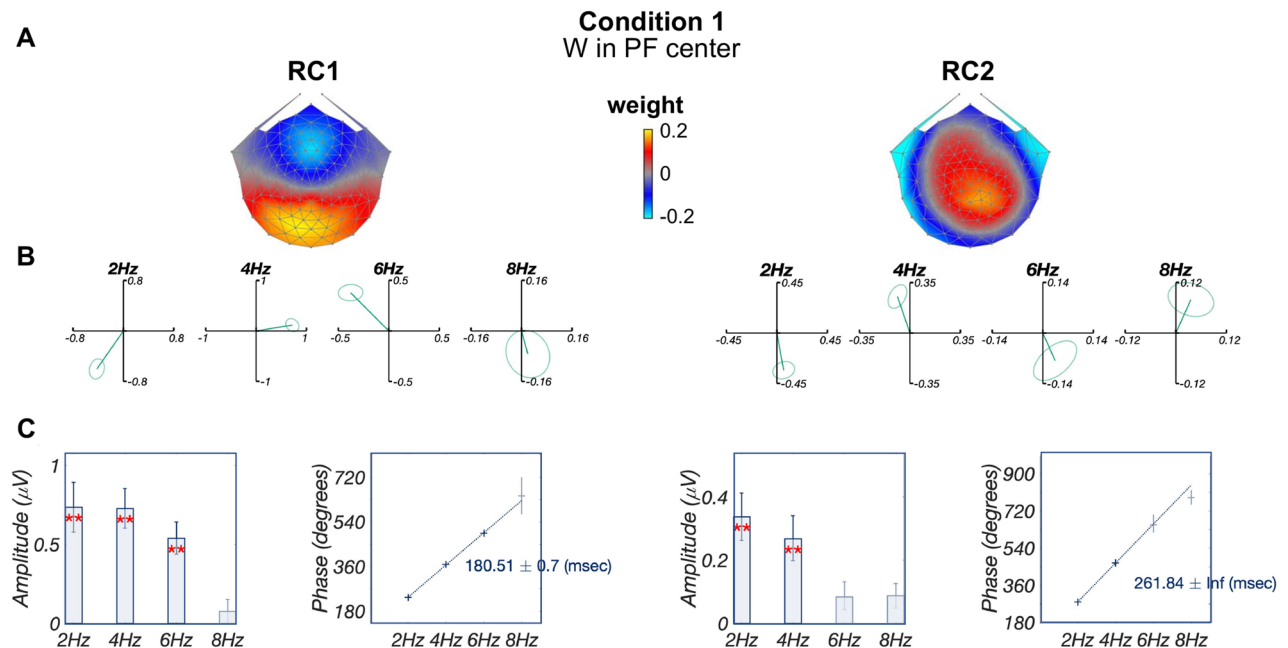
Figure 3. Deviant analysis: visual word form processing. (**A**) Topographic visualizations of the spatial filters for the first two components (RC1 and RC2); (**B**) Response data in the complex plane, where amplitude information is represented as the length of the vectors, and phase information in the angle of the vector relative to 0 radians (counterclockwise from 3 o'clock direction); ellipse indicates standard error of the mean for both amplitude and phase; (**C**) Projected amplitude (left) for each harmonic in bar charts, $**p_{FDR} < 0.01$, as well as latency estimations (right) across significant harmonics. Figures were created using publicly available MATLAB (R2020a) code: https://github.com/svndl/rcaExtra.

For word deviant responses appearing in a pseudofont context (word-in-pseudofont, condition 1), the first two reliable components explained the majority (RC1: 34.7%; RC2: 18.7%) of the reliability in the data. As shown in Fig. 3A, the topography of the first component (RC1) was maximal over left posterior vOT regions, while the second component (RC2) was distributed over more dorsal parietal regions. Significant signals were present in the first three harmonics of complex-valued data in RC1 and the first two harmonics of RC2 (Fig. 3B, see "Methods" section). More detailed amplitude and phase information are presented in Fig. 3C.

For RC1, the projected data contained significant responses in the first three harmonics (2 Hz, 4 Hz, and 6 Hz; $p_{FDR} < 0.01$, corrected for 8 comparisons), while the linear fit across phase distributions for these three harmonics produced a latency estimate of 180.51 ± 0.7 ms. Data projected through the RC2 spatial filter revealed two significant harmonics at 2 Hz and 4 Hz ($p_{FDR} < 0.01$) and a longer latency estimate of 261.84 ms (standard error is unavailable when there are two data points). Circular statistics of RC1 and RC2 phase comparisons showed that RC2 phases are significantly longer (mean latency difference at 2 Hz: 82.9 ms; at 4 Hz: 68.1 ms) than RC1 (circular t-test; $p_{FDR} < 0.01$ for both 2 Hz and 4 Hz, corrected for 2 comparisons).

As mentioned in "Methods" section, we found that when word-in-pseudofont stimuli were presented at jittered retinal locations (condition 2), the resulting RC topographies correlated highly with those computed from responses in the centered condition 1 (RC1: $r = 0.99$; RC2: $r = 0.95$). Therefore, in Fig. 4 we report RCA results of conditions 1 and 2 in a common component space. As expected, the RC1 and RC2 topographies (Panel A) are similar to those reported when training the spatial filters on condition 1 alone (as presented in Fig. 3A). The summary plots of the responses in the complex plane (Panel B), show overlapping amplitudes (vector lengths) and phases (vector angles) between these two conditions, especially at significant harmonics (first three in RC1 and RC2).

In Panel C response amplitudes did not differ significantly across four harmonics (2 Hz: $p = 0.70$; 4 Hz: $p = 0.96$; 6 Hz: $p = 0.91$; 8 Hz: $p = 0.65$), nor did the derived latency estimations (RC1: 180.42 ms and 173.72 ms for conditions 1 and 2 respectively; RC2: 260.16 ms and 233.32 ms for conditions 1 and 2 respectively). Thus, we did not find evidence that local adaptation is appreciable for these stimuli. Additionally, circular statistics of RC1 and RC2 phase comparisons showed that RC2 phases are significantly longer (mean latency difference at 2 Hz: 83.2 ms; at 4 Hz: 42.2 ms) than RC1 ($p_{FDR} < 0.05$ for both 2 Hz and 4 Hz, corrected for 2 comparisons).

Furthermore, these two RCs were also detected when presenting word-in-pseudofont at a slower alternation presentation rate, with 3-Hz deviant and 6-Hz base (condition 3, Fig. 5). Similar to conditions 1 and 2, topographies for conditions 1 and 3 were highly correlated (RC1: $r = 0.99$; RC2: $r = 0.91$). Although it was not possible to estimate the latency, as responses for condition 3 were significant only at the first harmonic (i.e., 3 Hz) for both components (RC1: $p_{FDR} < 0.001$; RC2: $p_{FDR} < 0.01$, Panel C), circular statistics showed that RC2 phase at 3 Hz is significantly longer (mean latency difference 82.0 ms) than RC1 ($p < 0.001$). The EEG spectra of representative sensors near the peak value of RC1 and RC2 are illustrated in supplementary materials (details in Fig. S1).
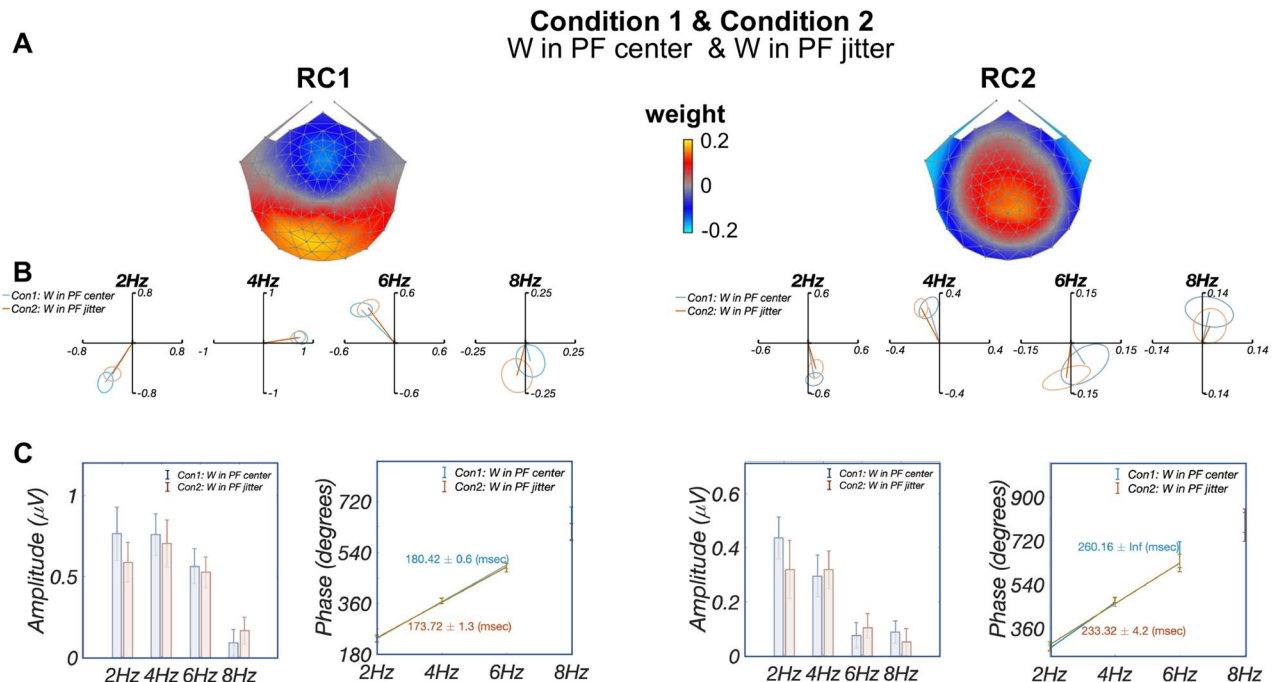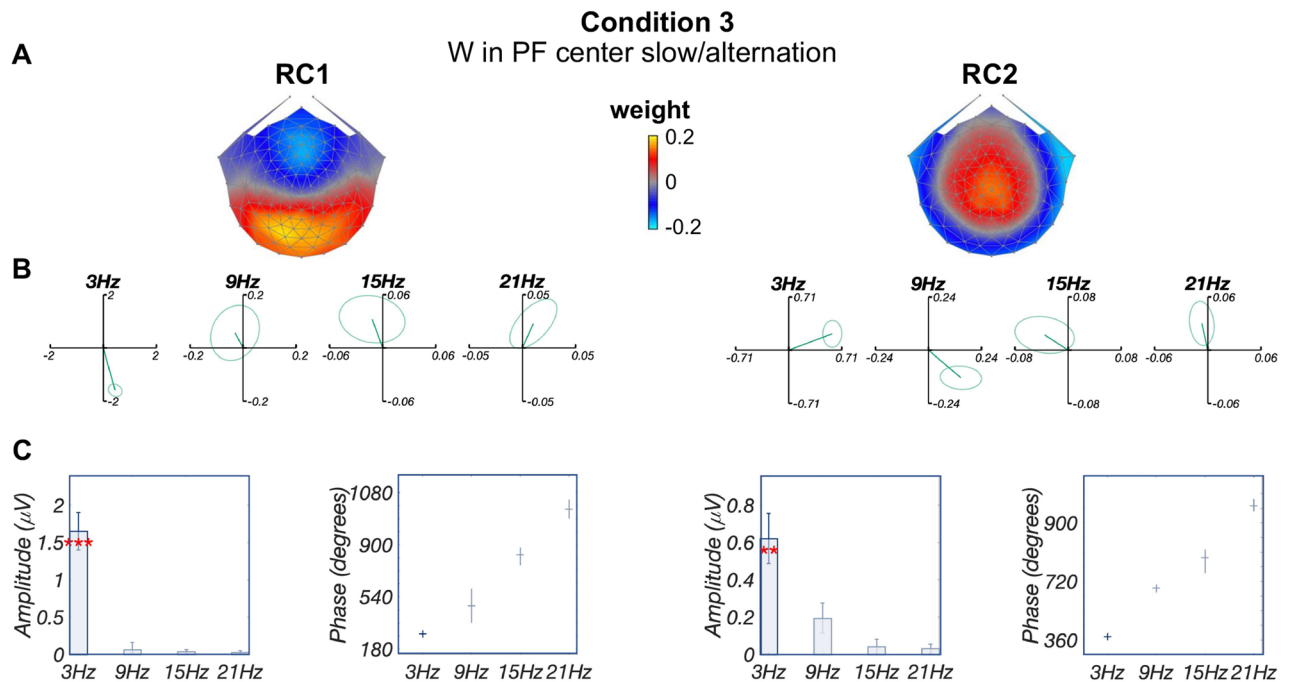
Figure 4. Deviant analysis: visual word form processing responses are similar irrespective of stimulus location. RCA was trained on response data from conditions 1 and 2 together as their RCs were highly similar when trained separately. (**A**) Topographic visualizations of the spatial filters for the first two components (RC1 and RC2); (**B**) Response data in the complex plane, with condition 1 (W in PF center) in blue and condition 2 (W in PF jitter) in red. Amplitude (vector length) and phase (vector angle, counterclockwise from 0 radians at 3 o'clock direction) overlap across conditions especially at significant harmonics (first three harmonics in RC1 and RC2). Ellipses indicate standard error of the mean; (**C**) Comparison of projected amplitude (left) and latency estimation (right) between two conditions. Response amplitudes did not differ significantly across the four harmonics ($p > 0.65$); latency estimations derived from phase slopes across significant harmonics were also similar between conditions (RC1: 180.42 ms and 173.72 ms for conditions 1 and 2 respectively; RC2: 260.16 ms and 233.32 ms for conditions 1 and 2 respectively). Figures were created using publicly available MATLAB (R2020a) code: https://github.com/svndl/rcaExtra.

Finally, for conditions 4 and 5, word deviants appearing in the other two contexts (word-in-nonword, word-in-pseudoword) produced components with weaker responses that were associated with distinct topographies, consistent with the hypothesis that each contrast was associated with overlapping yet distinct neural sources (Fig. 6). Of note, no more than one significant harmonic is observed for each component: for condition 4 (word-in-nonword), only the third harmonic of RC2 is significant (6 Hz, $p_{FDR} < 0.05$, corrected for 8 comparisons); for condition 5 (word-in-pseudoword) no harmonic is significant ($p_{FDR} > 0.14$, corrected for 8 comparisons). Due to the lack of significance at multiple harmonics, it was not possible to estimate latencies for these conditions.

Discussion

In this study, we examined the functional and temporal organization of brain sources involved in visual word recognition by employing a data-driven component analysis of steady-state visual evoked potential (SSVEP) data. We recorded SSVEP word deviant responses appearing in pseudofont contexts, projecting the multisensor EEG recordings onto single components using Reliable Components Analysis (RCA). Results at the first four harmonics of the base frequency revealed one component centered on medial occipital cortex. Results at harmonics of the deviant frequency revealed two distinct components, with the first component maximal earlier in time over left vOT regions, and the second maximal later in time over dorsal parietal regions. These two components were found to generalize across static versus jittered presentation locations as well as varying rates of base and deviant stimulation. In addition, distinct topographies were revealed during word deviant responses in the other two contexts (i.e., pseudowords, nonwords), which—compared with the word-in-pseudofont contrast—have different demands of distinguishing words from visual control stimuli.

RCA analyses of EEG responses at the first four harmonics of the base frequency revealed a maximally reliable component centered on medial occipital sensors across all conditions. This scalp topography corresponds to expected activations reported in fMRI literature[42–44] and in EEG source localization studies[45,46]. Medial occipital sensors are directly over early retinotopic visual areas known to support the first stages of visual processing, including of letter strings and objects[47]. Early stages of letter/word and object processing primarily involve low-level visual feature analysis, e.g., luminance, shape, contour, line junctions, and letter fragments[44,48,49]. In our study, basic low-level visual stimulus properties such as spatial frequency, spatial dimensions, and the sets of basic line-junction features[50] were well matched across pseudofont, nonword, and pseudoword base contexts,

Figure 5. Deviant analysis: visual word form processing evidenced by changing presentation rates. (**A**) Topographic visualizations of the spatial filters for the first two components (RC1 and RC2); (**B**) Response data in the complex plane, where amplitude information is represented as the length of the vectors, and phase information in the angle of the vector relative to 0 radians (counterclockwise from 3 o'clock direction); ellipse indicates standard error of the mean; (**C**) Projected amplitude for each harmonic in bar charts, $**p_{FDR} < 0.01$, $***p_{FDR} < 0.001$. Only one significant harmonic prevents us from estimating latency from the phase slope. Figures were created using publicly available MATLAB (R2020a) code: https://github.com/svndl/rcaExtra.

as well as between base and deviants within each contrast. This may explain comparable responses in terms of amplitudes and topographies across conditions with the same presentation rates. The amplitudes in condition 3 (word-in-pseudofont alternation presentation mode) are higher than other conditions, which may result from higher signal-to-noise ratio in one stimulation cycle of alternation presentation mode[26]. Nevertheless, consistent responses—in terms of underlying brain sources—across different contrasts, and at different image update rates and retinal locations, support the interpretation that current medial occipital activation reflects low-level visual feature processing rather than higher-level word-related processing.

RCA of word deviants in the word-in-pseudofont contrast produced two distinct components with differing latencies. The first component was maximal over ventral occipito-temporal (vOT) regions with slight left lateralization. Phase lag quantification of the first component revealed a linear fit of phases across successive harmonics, providing evidence of a latency estimation around 180 ms.

This 180 ms latency is consistent with the timing of the N170 component revealed in ERP data, which typically peaks around 140–180 ms especially in adults[16,33]. With its characteristic topography over the left occipitotemporal scalp, the N170 is considered to be an electrophysiological correlate of left vOT specialized activation for printed words in fMRI studies, as evidenced by magnetoencephalography (MEG)[51], EEG source localization[16,52], and simultaneous EEG-fMRI recordings[53].

Word deviant responses attributed over left-lateralized vOT (RC1) in the current study are most likely related to visual processes involved in extracting learned perceptual information associated with letters and words from visual stimuli. This interpretation is consistent with numerous fMRI studies showing that the left vOT plays a critical role in fast and efficient visual recognition of words[54,55], indicated by selective responses to visually presented words compared with novel symbols (e.g., pseudofonts)[6,56]. This left occipito-temporal activation has also been revealed in SSVEP studies of word recognition[28,29]. Additionally, using the same RCA approach as in the current study, Barzegaran et al.[27] also reported responses at left vOT cortex when viewing images containing sets of intact versus scrambled letters[27], which further supports the idea of the involvement of this first component in extraction of learned visual feature information associated with letters within word forms. Overall, consistent findings in the current study and in previous fMRI, ERP, and SSVEP studies support the reliable involvement of left vOT in processing of words and letters, especially word and letter form processing. This specificity of left vOT for words and letters over other visual stimuli[47] is an outcome of literacy acquisition[47] and emerges rapidly after a short term of reading training[17,53,57], a finding which has also been reported in different writing systems[58,59].

In addition to the first component, we also observed a second RC that was maximal over dorsal parietal regions. A linear fit across phases of successive harmonics demonstrated a latency estimation of around 260 ms, which is consistent with EEG and MEG studies showing phonological effects occurring 250–350 ms after the onset of a visual word[60–62]. We consider the delayed time course of RC2, and the distinct topography of this component compared with RC1, to be also consistent with a proposed function related to access of lexical-semantic
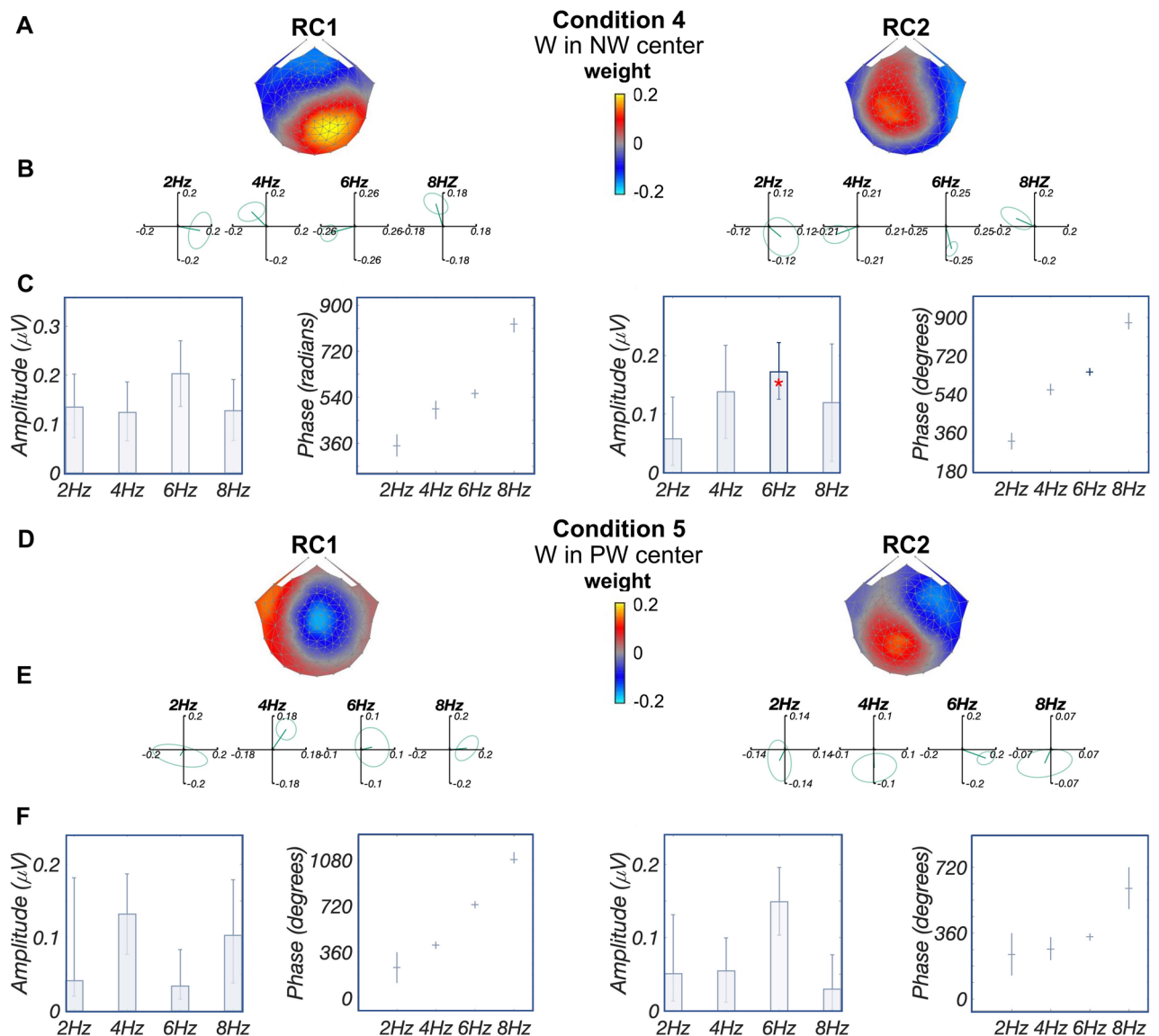
Figure 6. Deviant analysis: lexical-semantic processing. (**A,D**) Topographic visualizations of the spatial filters for the first two components (RC1 and RC2) for conditions 4 (W in NW center) and 5 (W in PW center), respectively; (**B,E**) Response data in the complex plane for conditions 4 and 5, respectively. Amplitude information is represented as the length of the vectors, and phase information in the angle of the vector relative to 0 radians (3 o'clock direction); ellipse indicates standard error of the mean. (**C,F**) Projected amplitude for each harmonic in bar charts, $*p_{FDR} < 0.05$, for condition 4, only the third harmonic in RC2 is significant; for condition 5, no harmonic is significant. Having a single significant harmonic prevents us from estimating the phase slope; thus, no latency estimates are provided. Figures were created using publicly available MATLAB (R2020a) code: https://github.com/svndl/rcaExtra.

representations. The finding of a second component is consistent with increasing evidence showing that visual word recognition is not limited to the left vOT. Instead, other higher-level linguistic representation areas also support final word recognition[13,14,63]. For example, a recent fMRI study showed that, connected with the middle and anterior vOT, other language areas (e.g., the supramarginal, angular gyrus, and inferior frontal gyrus) are responsible for lexical information processing especially for real words[13]. Additionally, anterior vOT is structurally connected to temporo-parietal regions, including the angular gyrus (AG[13,64,65]), the supramarginal gyri (SMG[66,67]) and the superior temporal gyrus (STS[68]). Those areas have been involved in mapping between orthographic, phonological and semantic representations for word recognition[69–72]. In relation to this, an intracranial EEG study suggested the left vOT is involved in at least two distinct stages of word processing: an early stage that is dedicated to gist-level word feature extraction, and a later stage involved in accessing more precise representations of a word by recurrent interactions with higher-level visual and nonvisual regions[51]. Therefore, the dorsal parietal brain sources in our study may reflect the later stage of integrating information between anterior vOT[13] and other language areas[73–75] to enable orthographic-lexical-semantic transformations[76].

A leading model, the Local Combination Detector (LCD) model[5], has been proposed to explain the neural mechanisms underlying visual specialization for words. This model suggests that the receptive fields of neurons (detectors), especially in the left vOT, progressively increase from posterior to anterior regions. As a result, the sensitivity of neurons increases from familiar letter fragments and fragments combinations to more complex letter strings (e.g., bigrams and morphemes) and whole words. However, given the functional and structural connections between vOT (especially anterior regions) and other language areas mentioned above (i.e., AG, SMG, and STS), a considerable number of studies support the interactive theory in which feed-forward and feed-back processing mechanisms together contribute to final recognition of a word[5,6,47,63]. Specifically, the forward pathway conveys bottom-up progression from early visual cortex (e.g., V4) to vOT, which accumulates inputs about the elementary forms of words[74], and continually from vOT to higher-level linguistic representation areas, which enables integration of orthographic stimuli with phonological and lexical representations[14]. Meanwhile, vOT also receives top-down modulations (backward pathways) from higher-level language regions, which provide (phonological and/or semantic) predictive feedback to the processing of visual attributes.

To summarize, two components were derived in a data-driven fashion from the SSVEP. The first component is located at left vOT, while the second component is more dorsal parietal. We speculate that the first component reflects more hierarchical bottom-up forward projections from early visual cortex to posterior vOT, while the second component represents both bottom-up and top-down integration of anterior vOT with other areas of the language network. However, these topographic interpretations are speculative without the evidence of source localization data. Thus, more direct evidence (e.g., source localization and functional connectivity) is needed for verification in future studies.

Our results partially replicated and extended previous work on multiple distinct brain sources involved in different stages of word processing. These findings were further corroborated by presenting stimuli at jittered locations. The response invariance we observe is in agreement with previous studies, in which left vOT activation was identical whether stimuli were presented in the right or left hemifield, suggesting that left vOT activation for words and readable pseudowords depends on language-dependent parameters and not visual features of stimuli[54,55]. In line with this, Maurer et al.[77] directly compared responses to words and faces under two contexts: blocks that alternated faces and words versus blocks of only faces or words. Results demonstrated that responses to words were consistently left-lateralized and were not influenced by context in skilled readers. In contrast, context[77] and spatial orientation[78] systematically influenced the degree to which a face is processed. Findings of this kind further suggest that responses to words depend more on lexical rather than contextual factors[79], which may be driven by the way that words are learned and read[77]. Anatomically, it has also been proposed that the visual word form system is homologous to inferotemporal areas in the monkey, where cells are selective to high-level features and invariant to size and position[54]. The similarity of the component structure (brain sources) under different presentation rates suggests that a fairly broad temporal filter is involved in word vs pseudofont discrimination. Of note, the amplitudes when slowing down the presentation rates are stronger than using higher rates, which sheds light on the choice of stimulus frequency for studies with children and for examining higher-level visual processes (see also[19]).

Given that the word-in-pseudofont contrast results above consistently identified and distinguished two components with distinct topographies and time courses—with the possibility of one being related to processing word visual forms and the other potentially related to higher-level integration with language regions—we went on to test whether the hierarchical stimulus contrast approach might also clearly isolate one of these components.

Neither word-in-nonword nor word-in-pseudoword sequences evoked component topographies that resembled either the early or late components of the word-in-pseudofont contrast. Specifically, the word-in-nonword contrast revealed two components over right occipito-temporal (RC1) and left centro-parietal (RC2) regions, while the word-in-pseudoword contrast revealed two components over right centro-frontal (RC1) and left occipito-temporal (RC2) regions. However, component-space neural activations for these two conditions were much weaker and less robust, so caution is needed in their interpretation. Such results favor the explanation that processes captured within the word-in-pseudofont condition may reflect a combination of letter-level perceptual expertise effect and higher-level lexico-semantic processing both of which only emerge when letter-level cues enable segmentation between unfamiliar visual forms. These RCA-based insights may suggest that previous claims (e.g.,[28]) to similar results across all three such contrasts (i.e., word-in-pseudofont, word-in-nonword, and word-in-pseudoword), based largely on single-channel findings, might fail when taking entire topographies into account using data-driven methods. Furthermore, such findings may underscore the importance of considering such SSVEP responses as being driven by the act of segregating deviant events out of well-controlled streams of stimuli, rather than reflecting inherent unique processes elicited by deviant events after common processes elicited by the base events are "subtracted" via the assumption of pure insertion.

Nevertheless, our challenges in resolving robust signals for word-in-nonword and word-in-pseudoword contrasts are consistent with a recent study[80]. Barnes et al.[80] also used a 4-letter English word-vs-pseudoword contrast, and the same frequency rates (2-Hz word deviants in a 10-Hz base) as in our study, finding that only 4 out of 40 data sets showed a reliable effect between words and pseudowords; while another study by Lochy et al.[28], which used 5-letter french words and pseudowords with the same 2-Hz/10-Hz rates, revealed response differences between words and pseudowords in 8 out 10 readers. Barnes et al. argued that the matching of bigram frequency between words and pseudowords (matched in their study but not in the Lochy et al. study[28]) could explain the disparity. Bigram frequency was also matched in our study, which further supports Barnes's argument. Additionally, other factors may also play a role in the disparity, such as number of letters (4 in ours and Barnes's study[80], 5 in Lochy's study[28]), number of syllables (monosyllabic in ours and Barnes's[80], monosyllabic and bisyllabic in Lochy's[28]), and the increased transparency of French relative to English.

Evoked response differences between words and pseudowords and/or nonwords have also been studied previously in ERP studies, but the results have been inconsistent (absent[81–83]; present[33,46,84,85]). Several reasons

for the mixed results have been proposed, including but not limited to language transparency, presentation modes and task demands. For example, it was found that the adult N170/left vOT for words is more sensitive to orthographic than lexical and/or semantic contrasts, especially during implicit reading[16,82]. Of note, an implicit color detection task was used in the present study, and the pseudowords and nonwords were created by reordering letters that appeared in the words; these factors, in combination, might have led to more difficult discrimination. In addition, words and pseudowords have elicited relatively smaller N170 amplitudes in less transparent English than in German[86]. Compared with more transparent languages (e.g., Italian, German and French), English has greater orthographic depth with inconsistent spelling-to-sound correspondence, leading to more ambiguous pronunciations. As a consequence of the inconsistency of mapping letters to sounds, lexical or semantic processing will be less automatic and more demanding[87], which is even more extreme when reading novel pseudowords and nonwords.

In addition, fast stimulus presentation rates (10 Hz, i.e., 100 ms per item) used in the current study may reduce the involvement of higher-level (e.g., semantic) processes[12,30] tapped by the word-in-nonword and word-in-pseudoword contrasts, especially during implicit tasks not requiring explicit pronunciation and semantic detection of the stimuli. Lower stimulation rates may be necessary to record SSVEP when discriminating higher-level lexical properties in word-in-nonword and word-in-pseudoword contrasts. By contrast, discrimination of words vs pseudofonts is robust over the range of presentation rates we examined. Future studies can examine these possibilities further by varying stimulation rates over broader ranges or by manipulating task demands.

In the present study, visual word recognition was examined using the SSVEP paradigm and a spatial filtering approach, which enabled the identification of robust neural sources supporting distinct levels of word processing within only several minutes' stimulation. This approach provided a unique extension of existing knowledge on word recognition regarding retinal location and stimulation rate and points to avenues for further investigation of important questions in reading.

First of all, given the short stimulation requirement and robust signal detection due to the high SNR of SSVEP and the spatial filtering approach, the current study points to new possibilities to study individual differences and developmental changes as children learn to read. Children's reading expertise develops dramatically, especially during the first year(s) of formal reading acquisition[1,33]. The high SNR paradigm used here may allow for more efficient (in terms of less time-consuming and fewer trials required) tracing of the developmental changes of brain circuitry due to increasing reading expertise in children at different stages of learning to read. Additionally, in the course of emerging specialization for printed words, neural component topographies would be expected to be increasingly left-lateralized[16]. The RCA approach of detecting multiple, distinct brain sources within the same signal response can increase understanding of how exactly such lateralization is best developed through reading training. Furthermore, high inter-subject variability of vOT print sensitivity was revealed in previous studies[53,68,88–90]. Therefore, exploring developmental changes—in terms of activity and topography—at the individual subject level will offer a chance to investigate this phenomenon more precisely and deeply. This approach would be especially relevant to early autistic and dyslexic readers, who face difficulties at the beginning of reading education[91] when intervention is thought to be most successful.

Another interesting direction to explore in future research rests upon the accumulating evidence that early visual-orthographic ("perceptual") and later lexico-semantic ("lexical") processing is located at segregated regions within vOT[12,13,92]. To our knowledge, different functional components of word recognition have been studied using general contrasts of words with pseudofonts and/or consonant strings[13], but not yet by directly manipulating orthographic regularity and lexical representation(s) separately. Additionally, studies have shown that perceptual tuning of posterior vOT to sublexical and lexical orthographic features develops when reading experience increases[93,94]. Therefore, isolating perceptual and lexical processing experimentally will be interesting to explore to understand how children's brains develop specialized visual-orthographic processing.

Last but not least, topographic lateralization and amplitudes of responses to text have been shown to differ when presenting stimuli with different temporal frequencies[26]. In addition, it has been proposed that lower stimulation rates may be necessary to record SSVEPs when studying higher-level visual processes (e.g., faces and words)[19]. The relatively high stimulation frequency (i.e., 10 Hz) used in the current study may underestimate effects especially associated with more complex "lexical" contrasts (herein word vs pseudoword and word vs nonword)[30], and this may have an even higher impact when it comes to children. Thus, an important question for future work is to see how children develop specialized visual-orthographic and lexical-semantic processing under lower stimulation rates.

Methods

Ethics statement.
This research was approved by the Institutional Review Board of Stanford University. All participants delivered written informed consent prior to the study after the experimental protocol was explained. The whole experiment was performed in accordance with relevant guidelines and regulations.

Participants.
Data from 16 right-handed, native English speakers (between 18.1 and 54.9 years old, median age 20.7 years, 7 males) were analyzed in this study. All participants had normal or corrected-to-normal vision and had no reading disabilities. Data from 5 additional non-native English speakers were recorded, but not analyzed here. After the study, each participant received cash compensation.

Stimuli.
The study involved four types of stimuli—words (W), pseudofonts (PF), nonwords (NW), and pseudowords (PW)—all comprising 4 elements (letters or pseudoletters). The English words were rendered in the Courier New font. Pseudofont letter strings were rendered from the Brussels Artificial Character Set font (BACS-2[95]), mapping between pseudofont glyphs and Courier New word glyphs. Nonwords and pseudowords

were also built on an item-by-item basis by reordering the letters of the words: nonwords were unpronounceable, statistically implausible letter string combinations, while pseudowords were pronounceable and well-matched for orthographic properties of intact words[96]. Bigram frequencies were matched between words (M (± SD) = 13664 (± 11007)), pseudowords (M (± SD) = 15177 (± 8549)), and nonwords (M(± SD) = 12775 (± 6065)) ($F(2, 87) < 1, p = 0.57$). Example stimuli are shown in Fig. 1. All words were common monosyllabic singular nouns. The initial and final letters in all words, pseudowords, and nonwords were consonants. Words were chosen to be frequent (average 97.7 per million) with limited orthographic neighbors (average 2.3, range from 0 to 4) according to the Children's Printed Word Database[97]. Words were also chosen with attention to feedforward consistency. All words were fully feedforward consistent based on rime according to the database provided by Ziegler et al.[98]. When averaging across consistency values for each word's onset, nucleus, and coda in the database provided by Chee et al.[99], words had an average token feedforward consistency of 0.79. All in all, there were 30 exemplars of each type of stimulus, for 120 exemplars total. All images were 600×160 pixels in size, spanning 7.5 (horizontal) by 2 (vertical) degrees of visual angle.

We investigated five experimental conditions. Conditions 1, 2, and 3 involved word deviants embedded in a stream of pseudofont stimuli. For conditions 1 and 2, word deviants were presented at a rate of 2 Hz and were embedded in a 10-Hz base stimuli rate. In order to explore the influence of presentation location on word processing, condition 1 stimuli were presented in the center of the screen, while in condition 2, stimulus positions were spatially jittered around the center of the monitor (8 pixels range, visual angle of 0.1 degrees). Moreover, to directly compare the effect of different base/deviant frequencies, in condition 3 we used a word deviant frequency of 3 Hz (based on the study of Yeatman and colleagues[26]) and a base frequency of 6 Hz, all presented at the center of the monitor. Finally, conditions 4 and 5 presented word-in-nonword and word-in-pseudoword contrasts, respectively (see also[28]). Like condition 1, word deviants were presented at the center of the monitor at 2 Hz, in a base stimulation rate of 10 Hz.

Experimental procedure. Participants were seated in a darkened room 1 m away from the computer monitor. Prior to the experiment, a brief practice session was held to familiarize the participant with the experimental procedure.

Each stimulation sequence started with a blank screen, the duration of which was jittered between 2500 ms and 3500 ms. Then, W deviant stimuli embedded in the stream of PF, NW, or PW stimuli were presented at a rate of 2 Hz (i.e., every 500 ms), with a base rate of 10 Hz (i.e, every 100 ms) in conditions 1, 2, 4, and 5; in condition 3, W deviant were presented at a frequency of 3 Hz (i.e, every 333 ms) with a base rate of 6 Hz (i.e, every 167 ms), during which deviant and base alternated with each other. Thus, a W deviant stimulus was presented every 5 item presentations in conditions 1, 2, 4, and 5, and every 2 items in condition 3. Each condition comprised four trials (each trial lasted for 12 s), and was repeated four times, resulting in 16 trials per condition, and 80 trials total in all conditions per participant. 20 trials (5 conditions × 4 trials of each) comprised a block, and the order of trials in each block and all 4 blocks were randomized.

In order to maintain participants' attention throughout the experiment, a fixation color change task was used. During the recording, the participant continuously fixated on a central cross, which was superimposed over the stimuli of interest, and pressed a button whenever they detected that the color of the fixation cross changed from blue to red (2 changes randomly timed per sequence/trial). The color change task was on a "staircase" mode, during which the time of color change flashes became faster as the accuracy increased, or became slower when the accuracy decreased. The whole experiment took around 30 minutes per participant, including breaks between blocks.

EEG recording and preprocessing. The 128-sensor EEG were collected with the Electrical Geodesics, Inc. (EGI) system[100], using a Net Amps 300 amplifier and geodesic sensor net. Data were acquired against Cz reference, at a sampling rate of 500 Hz. Impedances were kept below 50 kΩ. Stimuli were presented using in-house stimulus presentation software. Each recording was bandpass filtered offline (zero-phase filter, 0.3–50 Hz) using Net Station Waveform Tools. The data were then imported into in-house signal processing software for preprocessing. EEG data were re-sampled to 420 Hz to ensure an integer number of time samples per video frame at a frame rate of 60 Hz, as well as an integer number of frames per cycle for the present stimulation frequencies. EEG sensors with more than 15% of samples exceeding a 30-μV amplitude threshold were replaced by an averaged value from six neighboring sensors. The continuous EEG was then re-referenced to average reference[101] and segmented into 1-second epochs. Epochs with more than 10% of time samples exceeding a 30-μV noise threshold, or with any time sample exceeding an artifact threshold of 60 μV (e.g., eye blinks, eye movements, or body movements), were excluded from further analyses on a sensor-by-sensor basis. The EEG signals were filtered in the time domain using Recursive Least Squares (RLS) filters[102] tuned to each of the analysis frequencies and converted to complex amplitude values by means of the Fourier transform. Given 1-s data epochs, the resulting frequency resolution was 1 Hz. Complex-valued RLS outputs were decomposed into real and imaginary coefficients for input to the spatial filtering computations, as described below.

Analysis of behavioral data. Behavioral responses for the fixation cross color change task served to monitor participants' attention during EEG recording. We conducted one-way ANOVAs separately for reaction time and accuracy to determine whether participants were sufficiently engaged during the whole experiment.

Analysis of EEG data. *Reliable components analysis.* Reliable Components Analysis (RCA) is a matrix decomposition technique that derives a set of components that maximizes trial-to-trial covariance relative to within-trial covariance[39,40]. Matlab code is publicly available: https://github.com/svndl/rcaExtra. Since response

phases of SSVEP are constant over repeated stimulations, RCA uses this trial-to-trial reliability to decompose the entire 128-sensor array into a small number of reliable components (RCs), the activations of which reflect phase-locked activities. Moreover, RCA achieves higher output SNR with a low trial count compared to other spatial filtering approaches such as PCA and CSP[40].

Given a sensor-by-feature EEG data matrix (where features could represent e.g., time samples or spectral coefficients), RCA computes linear weightings of sensors—that is, linear spatial filters—through which the resulting projected data exhibit maximal Pearson Product Moment Correlation Coefficients[103] across neural response trials. The projection of EEG data matrices through spatial filter vectors transforms the data from sensor-by-feature matrices to component-by-feature matrices, with each component representing a linear combination of sensors. For the present study, EEG features are the real and imaginary Fourier coefficients at selected frequencies. As RCA is an eigenvalue decomposition[39], it returns multiple components, which are sorted according to "reliability" explained (i.e., the first component, RC1, explains the most reliability in the data). Forward-model projections of the eigenvectors (spatial filter vectors) can be visualized as scalp topographies[104]. As eigenvectors are known to receive arbitrary signs[105], we manually adjusted the signs of the spatial filters of interest based on the maximal correlation between raw sensor data and RCA data. Quantitative comparisons of topographies (e.g., across conditions) were made by correlating these projected weight vectors. Finally, we computed the percentage of reliability explained by individual components using the corresponding eigenvalues, as described in a recent study[40].

RCA calculations. In order to test whether low-level features were well matched across conditions, we first computed RCA at base frequencies only. Specifically, we input as features the real and imaginary frequency coefficients of the first four harmonics of the base (i.e, 10 Hz, 20 Hz, 30 Hz, 40 Hz for conditions 1, 2, 4 and 5; 6 Hz, 12 Hz, 18 Hz, 24 Hz for condition 3). RCA was computed separately for each stimulus condition.

We next computed RCA at deviant frequencies in order to investigate the processing differences between words and control stimuli (herein pseudofonts, nonwords, and pseudowords). For the deviant analyses, this involved real and imaginary coefficients at the first four harmonics (2 Hz, 4 Hz, 6 Hz, and 8 Hz) in conditions 1, 2, 4, and 5. To explore whether visual word processing can further be consistently detected under different presentation rates, we conducted RCA on data for condition 3. For this, we input frequency coefficients of odd harmonics of the deviant, excluding base harmonics (i.e., 3 Hz, 9 Hz, 15 Hz, 21 Hz—excluding 6 Hz, 12 Hz, 18 Hz, 24 Hz).

To assess the possible role of local adaptation to the stimulus presentation, we also measured responses to word-in-pseudofont using spatially jittered stimuli (condition 2). In comparing RCA results of conditions 1 (word-in-pseudofont with centered presentation location) and 2 (word-in-pseudofont with jittered presentation location), we found the RC topographies of conditions 1 and 2 to be highly correlated (RC1: $r = 0.99$; RC2: $r = 0.95$). Therefore, we subsequently computed RCA on these two conditions together to enable direct quantitative comparison of the projected data in a shared component space.

Analysis of component-space data. For each deviant RCA analysis, we report spatial filter topographies and statistical analysis of the projected data for the first two components returned by RCA. For each component, we analyzed component-space responses at each harmonic input to the spatial filtering calculation. We first projected the sensor-space data through the spatial filter vectors for RCs 1 and 2. The data were averaged across epochs on a per-participant basis, and statistical analyses were performed across the distribution of participants. The distribution of real and imaginary coefficients together at each harmonic formed the basis of a Hotelling's two-sample t^2 test[106] to identify statistically significant responses. We corrected for multiple comparisons using False Discovery Rate (FDR[107]) across 8 comparisons (4 harmonics × 2 components per condition).

To test whether phase information was consistent with a single phase lag reflected systematically across harmonics, we fit linear functions through the corresponding phases of successive harmonics, as such a linear relationship would implicate a fixed group delay which can be interpreted as an estimated latency in the SSVEP[41]. At harmonics with significant responses for both RCs (condition 1, conditions 1 and 2 comparison, condition 3), we used the Circular Statistics toolbox[108] to compare distributions of RC1 and RC2 phases at those significant harmonics. The results were corrected using FDR across 2 comparisons (2 significant harmonics) for condition 1 and for comparisons between condition 1 and 2; condition 3 involved no multiple comparisons as only one harmonic was significant. For each of these harmonics, we additionally report mean RC2-RC1 phase difference in msec.

For each deviant RCA analysis, we present topographic maps of the spatial filtering components, and also visualize the projected data in three ways. First, mean responses are visualized as vectors in the 2D complex plane, with amplitude information represented as vector length, phase information in the angle of the vector relative to 0 radians (counterclockwise from the 3 o'clock direction), and standard errors of the mean as error ellipses. Second, we present bar plots of amplitudes (μV) across harmonics, with significant responses (according to adjusted p_{FDR} values of t^2 tests of the complex data) indicated with asterisks. Finally, we present phase values (radians) plotted as a function of harmonic; when responses are significant for at least two harmonics, this is accompanied by a line of best fit and slope (latency estimate).

For each base RCA analysis, we report spatial filter topographies and statistical analysis of the projected data for the first component returned by RCA. As with the deviant RCA analysis, we also corrected for multiple comparisons using FDR[107] across 4 comparisons (4 harmonics × 1 component per condition). In contrast to deviant RCA analysis, we visualize the projected data only in bar plots of amplitudes (μV) across harmonics, with significant responses (according to adjusted p_{FDR} values) indicated with asterisks. Phase information and latency estimation are not included here because temporal dynamics are less accurate and less interpretable, especially at high-frequency harmonics (e.g., 30 Hz and 40 Hz)[19,109].

References

1. Maurer, U. *et al.* Coarse neural tuning for print peaks when children learn to read. *Neuroimage* **33**, 749–758 (2006).
2. Rayner, K. Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* **124**, 372 (1998).
3. Centanni, T. M., King, L. W., Eddy, M. D., Whitfield-Gabrieli, S. & Gabrieli, J. D. Development of sensitivity versus specificity for print in the visual word form area. *Brain Lang.* **170**, 62–70 (2017).
4. Baker, C. I. *et al.* Visual word processing and experiential origins of functional selectivity in human extrastriate cortex. *Proc. Natl. Acad. Sci.* **104**, 9087–9092 (2007).
5. Dehaene, S., Cohen, L., Sigman, M. & Vinckier, F. The neural code for written words: A proposal. *Trends Cogn. Sci.* **9**, 335–341 (2005).
6. Dehaene, S. & Cohen, L. The unique role of the visual word form area in reading. *Trends Cogn. Sci.* **15**, 254–262 (2011).
7. Krafnick, A. J. *et al.* Chinese character and english word processing in children ventral occipitotemporal cortex: FMRI evidence for script invariance. *Neuroimage* **133**, 302–312 (2016).
8. Dehaene, S. *et al.* How learning to read changes the cortical networks for vision and language. *Science* **330**, 1359–1364 (2010).
9. McCandliss, B. D., Cohen, L. & Dehaene, S. The visual word form area: Expertise for reading in the fusiform gyrus. *Trends Cogn. Sci.* **7**, 293–299 (2003).
10. Dehaene, S. *et al.* Cerebral mechanisms of word masking and unconscious repetition priming. *Nat. Neurosci.* **4**, 752–758 (2001).
11. Dehaene, S. *et al.* Letter binding and invariant recognition of masked words: Behavioral and neuroimaging evidence. *Psychol. Sci.* **15**, 307–313 (2004).
12. Vinckier, F. *et al.* Hierarchical coding of letter strings in the ventral stream: Dissecting the inner organization of the visual word-form system. *Neuron* **55**, 143–156 (2007).
13. Lerma-Usabiaga, G., Carreiras, M. & Paz-Alonso, P. M. Converging evidence for functional and structural segregation within the left ventral occipitotemporal cortex in reading. *Proc. Natl. Acad. Sci.* **115**, E9981–E9990 (2018).
14. Price, C. J. & Devlin, J. T. The interactive account of ventral occipitotemporal contributions to reading. *Trends Cogn. Sci.* **15**, 246–253 (2011).
15. Brem, S. *et al.* Evidence for developmental changes in the visual word processing network beyond adolescence. *Neuroimage* **29**, 822–837 (2006).
16. Maurer, U., Brem, S., Bucher, K. & Brandeis, D. Emerging neurophysiological specialization for letter strings. *J. Cogn. Neurosci.* **17**, 1532–1552 (2005).
17. Brem, S. *et al.* Brain sensitivity to print emerges when children learn letter-speech sound correspondences. *Proc. Natl. Acad. Sci.* **107**, 7939–7944 (2010).
18. Regan, D. Some characteristics of average steady-state and transient responses evoked by modulated light. *Electroencephalogr. Clin. Neurophysiol.* **20**, 238–248 (1966).
19. Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottereau, B. R. & Rossion, B. The steady-state visual evoked potential in vision research: A review. *J. Vis.* **15**, 4 (2015).
20. Stothart, G., Quadflieg, S. & Milton, A. A fast and implicit measure of semantic categorisation using steady state visual evoked potentials. *Neuropsychologia* **102**, 11–18 (2017).
21. Alonso-Prieto, E., Van Belle, G., Liu-Shuang, J., Norcia, A. M. & Rossion, B. The 6 hz fundamental stimulation frequency rate for individual face discrimination in the right occipito-temporal cortex. *Neuropsychologia* **51**, 2863–2875 (2013).
22. Farzin, F., Hou, C. & Norcia, A. M. Piecing it together: Infant neural responses to face and object structure. *J. Vis.* **12**, 6 (2012).
23. Liu-Shuang, J., Norcia, A. M. & Rossion, B. An objective index of individual face discrimination in the right occipito-temporal cortex by means of fast periodic oddball stimulation. *Neuropsychologia* **52**, 57–72 (2014).
24. Guillaume, M., Mejias, S., Rossion, B., Dzhelyova, M. & Schiltz, C. A rapid, objective and implicit measure of visual quantity discrimination. *Neuropsychologia* **111**, 180–189 (2018).
25. Van Rinsveld, A. *et al.* The neural signature of numerosity by separating numerical and continuous magnitude extraction in visual cortex with frequency-tagged eeg. *Proc. Natl. Acad. Sci.* **117**, 5726–5732 (2020).
26. Yeatman, J. D. & Norcia, A. M. Temporal tuning of word-and face-selective cortex. *J. Cogn. Neurosci.* **28**, 1820–1827 (2016).
27. Barzegaran, E. & Norcia, A. M. Neural sources of letter and vernier acuity. *Sci. Rep.* **10**, 1–11 (2020).
28. Lochy, A., Van Belle, G. & Rossion, B. A robust index of lexical representation in the left occipito-temporal cortex as evidenced by eeg responses to fast periodic visual stimulation. *Neuropsychologia* **66**, 18–31 (2015).
29. Lochy, A., Van Reybroeck, M. & Rossion, B. Left cortical specialization for visual letter strings predicts rudimentary knowledge of letter-sound association in preschoolers. *Proc. Natl. Acad. Sci.* **113**, 8544–8549 (2016).
30. 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.* **115**, E7595–E7604 (2018).
31. Lochy, A., Schiltz, C. & Rossion, B. The right hemispheric dominance for face perception in preschool children depends on the visual discrimination level. *Dev. Sci.* **23**, e12914 (2020).
32. Friston, K. J. *et al.* The trouble with cognitive subtraction. *Neuroimage* **4**, 97–104 (1996).
33. Eberhard-Moscicka, A. K., Jost, L. B., Raith, M. & Maurer, U. Neurocognitive mechanisms of learning to read: Print tuning in beginning readers related to word-reading fluency and semantics but not phonology. *Dev. Sci.* **18**, 106–118 (2015).
34. Rossion, B., Prieto, E. A., Boremanse, A., Kuefner, D. & Van Belle, G. A steady-state visual evoked potential approach to individual face perception: Effect of inversion, contrast-reversal and temporal dynamics. *Neuroimage* **63**, 1585–1600 (2012).
35. Kilner, J. Bias in a common eeg and meg statistical analysis and how to avoid it. *Clin. Neurophysiol.* **124**, 2062–2063 (2013).
36. Cohen, M. X. Comparison of linear spatial filters for identifying oscillatory activity in multichannel data. *J. Neurosci. Methods* **278**, 1–12 (2017).
37. Blankertz, B. *et al.* The berlin brain-computer interface: Accurate performance from first-session in bci-naive subjects. *IEEE Trans. Biomed. Eng.* **55**, 2452–2462 (2008).
38. Mohanchandra, K., Saha, S. & Deshmukh, R. Twofold classification of motor imagery using common spatial pattern. In *2014 International Conference on Contemporary Computing and Informatics (IC3I)*, 434–439 (IEEE, 2014).
39. Dmochowski, J. P., Sajda, P., Dias, J. & Parra, L. C. Correlated components of ongoing eeg point to emotionally laden attention—A possible marker of engagement?. *Front. Hum. Neurosci.* **6**, 112 (2012).
40. Dmochowski, J. P., Greaves, A. S. & Norcia, A. M. Maximally reliable spatial filtering of steady state visual evoked potentials. *Neuroimage* **109**, 63–72 (2015).
41. Norcia, A. M., Yakovleva, A., Hung, B. & Goldberg, J. L. Dynamics of contrast decrement and increment responses in human visual cortex. *Transl. Vis. Sci. Technol.* **9**, 6 (2020).
42. López-Barroso, D. *et al.* Impact of literacy on the functional connectivity of vision and language related networks. *Neuroimage* **213**, 116–722 (2020).
43. Turkeltaub, P. E., Gareau, L., Flowers, D. L., Zeffiro, T. A. & Eden, G. F. Development of neural mechanisms for reading. *Nat. Neurosci.* **6**, 767–773 (2003).

44. Szwed, M. *et al.* Specialization for written words over objects in the visual cortex. *Neuroimage* **56**, 330–344 (2011).
45. Rossion, B., Joyce, C. A., Cottrell, G. W. & Tarr, M. J. Early lateralization and orientation tuning for face, word, and object processing in the visual cortex. *Neuroimage* **20**, 1609–1624 (2003).
46. Proverbio, A. & Adorni, R. C1 and p1 visual responses to words are enhanced by attention to orthographic vs lexical properties. *Neurosci. Lett.* **463**, 228–233 (2009).
47. Dehaene, S., Cohen, L., Morais, J. & Kolinsky, R. Illiterate to literate: Behavioural and cerebral changes induced by reading acquisition. *Nat. Rev. Neurosci.* **16**, 234–244 (2015).
48. Ben-Shachar, M., Dougherty, R. F., Deutsch, G. K. & Wandell, B. A. Differential sensitivity to words and shapes in ventral occipito-temporal cortex. *Cereb. Cortex* **17**, 1604–1611 (2007).
49. Ostwald, D., Lam, J. M., Li, S. & Kourtzi, Z. Neural coding of global form in the human visual cortex. *J. Neurophysiol.* **99**, 2456–2469 (2008).
50. Changizi, M. A., Zhang, Q., Ye, H. & Shimojo, S. The structures of letters and symbols throughout human history are selected to match those found in objects in natural scenes. *Am. Nat.* **167**, E117–E139 (2006).
51. Hirshorn, E. A. *et al.* Decoding and disrupting left midfusiform gyrus activity during word reading. *Proc. Natl. Acad. Sci.* **113**, 8162–8167 (2016).
52. Brem, S. *et al.* Tuning of the visual word processing system: Distinct developmental erp and fmri effects. *Hum. Brain Mapp.* **30**, 1833–1844 (2009).
53. Pleisch, G. *et al.* Simultaneous eeg and fmri reveals stronger sensitivity to orthographic strings in the left occipito-temporal cortex of typical versus poor beginning readers. *Dev. Cogn. Neurosci.* **40**, 100717 (2019).
54. Cohen, *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).
55. Cohen, *et al.* Language-specific tuning of visual cortex? Functional properties of the visual word form area. *Brain* **125**, 1054–1069 (2002).
56. Rauschecker, A. M., Bowen, R. F., Parvizi, J. & Wandell, B. A. Position sensitivity in the visual word form area. *Proc. Natl. Acad. Sci.* **109**, E1568–E1577 (2012).
57. Chyl, K. *et al.* Prereader to beginning reader: Changes induced by reading acquisition in print and speech brain networks. *J. Child Psychol. Psychiatry* **59**, 76–87 (2018).
58. Bolger, D. J., Perfetti, C. A. & Schneider, W. Cross-cultural effect on the brain revisited: Universal structures plus writing system variation. *Hum. Brain Mapp.* **25**, 92–104 (2005).
59. Szwed, M., Qiao, E., Jobert, A., Dehaene, S. & Cohen, L. Effects of literacy in early visual and occipitotemporal areas of chinese and french readers. *J. Cogn. Neurosci.* **26**, 459–475 (2014).
60. Ashby, J. & Martin, A. E. Prosodic phonological representations early in visual word recognition. *J. Exp. Psychol. Hum. Percept. Perform.* **34**, 224 (2008).
61. Grainger, J., Kiyonaga, K. & Holcomb, P. J. The time course of orthographic and phonological code activation. *Psychol. Sci.* **17**, 1021–1026 (2006).
62. Sliwinska, M. W. W., Khadilkar, M., Campbell-Ratcliffe, J., Quevenco, F. & Devlin, J. T. Early and sustained supramarginal gyrus contributions to phonological processing. *Front. Psychol.* **3**, 161 (2012).
63. Long, L. *et al.* Feed-forward, feed-back, and distributed feature representation during visual word recognition revealed by human intracranial neurophysiology. *ResearchSquare* (2020).
64. Booth, J. R. *et al.* Development of brain mechanisms for processing orthographic and phonologic representations. *J. Cogn. Neurosci.* **16**, 1234–1249 (2004).
65. Yeatman, J. D., Rauschecker, A. M. & Wandell, B. A. Anatomy of the visual word form area: Adjacent cortical circuits and long-range white matter connections. *Brain Lang.* **125**, 146–155 (2013).
66. Kawabata Duncan, K. J. *et al.* Inter-and intrahemispheric connectivity differences when reading Japanese Kanji and Hiragana. *Cereb. Cortex* **24**, 1601–1608 (2014).
67. Seghier, M. L. & Price, C. J. Dissociating frontal regions that co-lateralize with different ventral occipitotemporal regions during word processing. *Brain Lang.* **126**, 133–140 (2013).
68. Stevens, W. D., Kravitz, D. J., Peng, C. S., Tessler, M. H. & Martin, A. Privileged functional connectivity between the visual word form area and the language system. *J. Neurosci.* **37**, 5288–5297 (2017).
69. Church, J. A., Balota, D. A., Petersen, S. E. & Schlaggar, B. L. Manipulation of length and lexicality localizes the functional neuroanatomy of phonological processing in adult readers. *J. Cogn. Neurosci.* **23**, 1475–1493 (2011).
70. Raij, T., Uutela, K. & Hari, R. Audiovisual integration of letters in the human brain. *Neuron* **28**, 617–625 (2000).
71. Van Atteveldt, N., Formisano, E., Goebel, R. & Blomert, L. Integration of letters and speech sounds in the human brain. *Neuron* **43**, 271–282 (2004).
72. Vandermosten, M., Hoeft, F. & Norton, E. S. Integrating mri brain imaging studies of pre-reading children with current theories of developmental dyslexia: A review and quantitative meta-analysis. *Curr. Opin. Behav. Sci.* **10**, 155–161 (2016).
73. Kay, K. N. & Yeatman, J. D. Bottom-up and top-down computations in high-level visual cortex. *BioRxiv.* https://doi.org/10.1101/053595 (2016).
74. 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).
75. Woodhead, Z. *et al.* Reading front to back: Meg evidence for early feedback effects during word recognition. *Cereb. Cortex* **24**, 817–825 (2014).
76. Woolnough, O. *et al.* Spatiotemporal dynamics of orthographic and lexical processing in the ventral visual pathway. *Nat. Hum. Behav.* **5**, 389–398 (2021).
77. Maurer, U., Rossion, B. & McCandliss, B. D. Category specificity in early perception: Face and word n170 responses differ in both lateralization and habituation properties. *Front. Hum. Neurosci.* **2**, 18 (2008).
78. Jacques, C. & Rossion, B. Early electrophysiological responses to multiple face orientations correlate with individual discrimination performance in humans. *Neuroimage* **36**, 863–876 (2007).
79. Hauk, O. *et al.* [q:] when would you prefer a sossage? [a:] at about 100 ms. ERP correlates of orthographic typicality and lexicality in written word recognition. *J. Cogn. Neurosci.* **18**, 818–832 (2006).
80. Barnes, L., Petit, S., Badcock, N. A., Whyte, C. J. & Woolgar, A. Word detection in individual subjects is difficult to probe with fast periodic visual stimulation. *Front. Neurosci.* **15**, 182 (2021).
81. Araújo, S., Bramão, I., Faísca, L., Petersson, K. M. & Reis, A. Electrophysiological correlates of impaired reading in dyslexic pre-adolescent children. *Brain Cogn.* **79**, 79–88 (2012).
82. Bentin, S., Mouchetant-Rostaing, Y., Giard, M.-H., Echallier, J.-F. & Pernier, J. Erp manifestations of processing printed words at different psycholinguistic levels: Time course and scalp distribution. *J. Cogn. Neurosci.* **11**, 235–260 (1999).
83. Wydell, T. N., Vuorinen, T., Helenius, P. & Salmelin, R. Neural correlates of letter-string length and lexicality during reading in a regular orthography. *J. Cogn. Neurosci.* **15**, 1052–1062 (2003).
84. Kast, M., Elmer, S., Jancke, L. & Meyer, M. ERP differences of pre-lexical processing between dyslexic and non-dyslexic children. *Int. J. Psychophysiol.* **77**, 59–69 (2010).
85. McCandliss, B. D., Posner, M. I. & Givon, T. Brain plasticity in learning visual words. *Cogn. Psychol.* **33**, 88–110 (1997).

86. 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**, 1–12 (2005).
87. Nosarti, C., Mechelli, A., Green, D. W. & Price, C. J. The impact of second language learning on semantic and nonsemantic first language reading. *Cereb. Cortex* **20**, 315–327 (2010).
88. Dehaene-Lambertz, G., Monzalvo, K. & Dehaene, S. The emergence of the visual word form: Longitudinal evolution of category-specific ventral visual areas during reading acquisition. *PLoS Biol.* **16**, e2004103 (2018).
89. Glezer, L. S. & Riesenhuber, M. Individual variability in location impacts orthographic selectivity in the visual word form area. *J. Neurosci.* **33**, 11221–11226 (2013).
90. van de Walle de Ghelcke, A., Rossion, B., Schiltz, C. & Lochy, A. Developmental changes in neural letter-selectivity: A 1-year follow-up of beginning readers. *Dev. Sci.* **24**, e12999 (2020).
91. Frith, U. & Snowling, M. Reading for meaning and reading for sound in autistic and dyslexic children. *Br. J. Dev. Psychol.* **1**, 329–342 (1983).
92. Stigliani, A., Weiner, K. S. & Grill-Spector, K. Temporal processing capacity in high-level visual cortex is domain specific. *J. Neurosci.* **35**, 12412–12424 (2015).
93. 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).
94. Zhao, J., Maurer, U., He, S. & Weng, X. Development of neural specialization for print: Evidence for predictive coding in visual word recognition. *PLoS Biol.* **17**, e3000474 (2019).
95. Vidal, C. & Chetail, F. Bacs: The brussels artificial character sets for studies in cognitive psychology and neuroscience. *Behav. Res. Methods* **49**, 2093–2112 (2017).
96. Keuleers, E. & Brysbaert, M. Wuggy: A multilingual pseudoword generator. *Behav. Res. Methods* **42**, 627–633 (2010).
97. Masterson, J., Stuart, M., Dixon, M. & Lovejoy, S. Children Printed Word Database: Continuities and changes over time in children early reading vocabulary. *Br. J. Psychol.* **101**, 221–242 (2010).
98. Ziegler, J. C., Stone, G. O. & Jacobs, A. M. What is the pronunciation for -*ough* and the spelling for/u/? A database for computing feedforward and feedback consistency in English. *Behav. Res. Methods Instrum. Comput.* **29**, 600–618 (1997).
99. Chee, Q. W., Chow, K. J., Yap, M. J. & Goh, W. D. Consistency norms for 37,677 English words. *Behav. Res. Methods* **25**, 1–21 (2020).
100. Tucker, D. M. Spatial sampling of head electrical fields: The geodesic sensor net. *Electroencephalogr. Clin. Neurophysiol.* **87**, 154–163 (1993).
101. Lehmann, D. & Skrandies, W. Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalogr. Clin. Neurophysiol.* **48**, 609–621 (1980).
102. Tang, Y. & Norcia, A. M. An adaptive filter for steady-state evoked responses. *Electroencephalogr. Clin. Neurophysiol. Evok. Potentials Sect.* **96**, 268–277 (1995).
103. Pearson, K. Mathematical contributions to the theory of evolution-iii. Regression, heredity, and panmixia. *Philos. Trans. R. Soc. Lond.* **187**, 253–318 (1896).
104. Parra, L. C., Spence, C. D., Gerson, A. D. & Sajda, P. Recipes for the linear analysis of EEG. *Neuroimage* **28**, 326–341 (2005).
105. Bro, R., Acar, E. & Kolda, T. G. Resolving the sign ambiguity in the singular value decomposition. *J. Chemom.* **22**, 135–140 (2008).
106. Victor, J. D. & Mast, J. A new statistic for steady-state evoked potentials. *Electroencephalogr. Clin. Neurophysiol.* **78**, 378–388 (1991).
107. Benjamini, Y. & Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **29**, 1165–1188 (2001).
108. Berens, P. *et al.* Circstat: A Matlab toolbox for circular statistics. *J. Stat. Softw.* **31**, 1–21 (2009).
109. Cottereau, B. R., McKee, S. P., Ales, J. M. & Norcia, A. M. Disparity-tuned population responses from human visual cortex. *J. Neurosci.* **31**, 954–965 (2011).

Author contributions

B.D.M. and A.M.N. conceived the study. C.B.S., L.H., and Q.T.H.N. conducted the experiment(s). F.W. and B.K. analyzed the data. F.W. wrote the original draft, and B.K., A.M.N., and B.D.M. edited the draft. A.Y. and V.Y.V. contributed to the method. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.