Starting to see a pattern: Longitudinal SSVEP signals reveal how growth in children's cortical sensitivity to visual word form structure is uniquely linked to two-year growth in reading fluency

Fang Wang[a,1], Blair Kaneshiro[a], Elizabeth Y. Toomarian[a,b], Radhika S. Gosavi[a,b], Lindsey R. Hasak[a], Suanna Moron[a], Quynh Trang H. Nguyen[a], Anthony M. Norcia[c,d], and Bruce D. McCandliss[a,1]

[a]Graduate School of Education, Stanford University, Stanford, CA, USA; [b]Synapse School, Menlo Park, CA, USA; [c]Department of Psychology, Stanford University, Stanford, CA, USA; [d]Wu Tsai Neurosciences Institute, Stanford, CA, USA

Learning to read depends on the ability to extract precise details of letter combinations, which convey critical information that distinguishes tens of thousands of visual word forms. To support fluent reading skill, one crucial neural developmental process is one's ability to rewire visual cortical circuitry to take advantage of statistical constraints inherent in combining letters into visual word forms. To test this idea in early readers, we tracked the impact of two years of schooling on within-student longitudinal changes in cortical responses to visual word forms and their growth in reading fluency. Three stimulus contrasts—words versus pseudofonts, words versus pseudowords, pseudowords versus nonwords—were presented while high-density EEG Steady-State Visual Evoked Potentials (SSVEPs, n=31) were recorded. Internalization of abstract visual word form structures over two years of reading experience resulted in a near doubling of SSVEP amplitude, with increasing left lateralization. Cortical entrainment by such word form structural information predicted the growth in reading fluency. However, no such changes were observed for whole word representation processing. Collectively, these findings indicate that sensitivity to visual word form structure develops rapidly through exposure to print and is uniquely linked to growth in reading fluency.

EEG–SSVEP | Visual word form structure | Reading growth predictor | Reliable Components Analysis (RCA) | Spatial filtering | Early readers

Visual word recognition is a rapid yet complex process involving multiple dimensions of analysis, including the specific shape of each letter form, the rules of letter combine within a word form, and whole word representation. Given this complexity, it has long been an ambition of developmental and educational cognitive neuroscience research to clarify how emerging expertise at different levels of word processing is reflected by changes in visual reading circuits and how these neural changes influence reading skills. Such work can help elucidate the neural basis of individual differences in reading acquisition and reading skills, and also help demonstrate within-individual changes due to specific educational activities.

A large body of literature (1–5) has investigated visual word recognition by contrasting responses to words versus visual controls (e.g., pseudofonts, strings combined with artificial character set font). Taken together, studies have consistently reported brain sensitivity to words compared to pseudofonts, a finding that is referred to as "coarse neural tuning" for print (4). During development, coarse neural tuning starts to emerge when children begin to read, following an inverted U-curve with an initial increase (6) and then a later decrease starting in second grade (4, 7).

The often-used coarse neural tuning contrast (words versus pseudofonts) is limited, however, in its connection to theory-driven insights into word recognition development, which stress the importance of multiple levels of word reading expertise (8). This limitation leads to ambiguity in recovering the unique contributions of specific properties of words to developmental changes in reading and brain circuitry, leaving open the question of whether coarse neural tuning reflects merely the aggregation of increasing word specific knowledge, or instead reflects insights into word form structures, perhaps as they map to phonological patterns (9).

Event-related potential (ERP) studies have attempted yet failed to isolate different levels of visual word processing in early readers (3, 10). For instance, Zhao and colleagues found brain sensitivity to visual word form structure only in 7-year-old children with high reading fluency, but failed to reveal sensitivity to whole visual word knowledge (10). At the same time, steady-state visual evoked potential (SSVEP) paradigms have been increasingly used, as this approach rapidly measures discrimination responses with high signal-to-noise ratio (SNR) in only a few minutes of stimulation (11) and has high test-retest reliability (12). Despite these advantages, SSVEP studies on word reading have not yet successfully differentiated effects of learning at the whole word level (such as recognizing a specific familiar word) from effects of word form structure learning in early readers (i.e., kindergartners (13) or first and second graders (6)). More importantly, no study so far has systematically tracked the developmental trajectories and neural dynamics of multiple levels of word reading expertise in early reading acquisition.

The current study aims to address this gap by isolating different levels of word reading expertise that emerge over the course of reading development, and by better connecting brain signals related to these different levels of word recognition with early reading fluency and reading growth.

In our recent SSVEP study (14), we isolated different levels of word-related information using multiple well-controlled contrasts and an adjusted SSVEP paradigm. Typically, SSVEP

paradigms involve presentation of two categories of stimuli that differ in a particular aspect (e.g., words versus pseudofonts) at two distinct, experimentally defined periodic rates. Previous SSVEP studies have used this "base/deviant" approach at faster stimulation rates, wherein deviant stimuli were embedded within a sub-multiple of the base rate that is greater than two, for example, 1.2-Hz deviant and 6-Hz base rates. We used an adjusted paradigm—which we refer to as "image alternation" mode—involving slower stimulation rates, where image exemplars from two categories of stimuli alternate at slower 1-Hz alternation and 2-Hz base rates. This slower image alternation mode has been shown to elicit responses with a higher SNR (15–17) than base/deviant approaches and may thus be more suitable for capturing reading-related brain signals in children (14).

We used three stimulus contrasts. First, the commonly used words–pseudofonts contrast was included to probe coarse neural tuning. Second, we used a pseudowords–nonwords contrast to investigate responses specific to orthographic structures within visual word forms while controlling for visual familiarity and whole word representation. In this contrast, well-structured orthographically reasonable letter combinations (pseudowords) were alternated with letter strings that violate orthographic constraints (nonwords). Notably, to avoid potential confounding of consonant-vowel distributions, orthographically illegal nonwords were created by reordering letters from consonant-vowel-consonant (CVC) pseudowords instead of pure consonant strings used in previous contrasts (10). Finally, responses to whole visual words was examined through a words–pseudowords contrast, which serves to contrast information on the level of a word unit while controlling for letter and letter combination structure experience.

In all, the stimulus contrasts and stimulation rates in the present sudy are those used in our recent non-longitudinal SSVEP study (14). Additionally and most importantly, longitudinal EEG data were recorded over a two-year interval in order to (1) track the developmental profile of different levels of word information, and (2) clarify to what extent sensitivity to visual word form structure and/or sensitivity to whole word knowledge relates to reading fluency and reading growth.

Results

Alternate RCA Results. Reliable component (RC) coefficients, which indicate the extent of across-trials covariance explained by a given component, dropped steeply after the first RC (see supplemental figure, SF 2), especially for words–pseudofonts and pseudowords–nonwords which have stronger signals. We therefore focused our alternate RCA results solely on the maximally correlated component, RC1.

Development of sensitivity to word form structure (pseudowords–nonwords). For responses to pseudowords–nonwords, the contrast designed to probe visual word form structure processing, RC1 is maximal over occipital electrodes on both hemispheres at the first testing time point, T1 (Figure 1A). Responses are statistically significant at the first two harmonics 1 Hz and 3 Hz (Hotelling's two-sample t^2 test, $p_{FDR} < 0.01$, corrected for 10 comparisons, Figure 1C). At the second testing time point, T2, the activation is more left lateralized (Figure 1B) with statistically significant responses at only the first harmonic, 1 Hz ($p_{FDR} < 0.001$, corrected for 10 comparisons, Figure 1C). Developments of lateralizaion of brain responses at T1 and T2 are shown in Figure S5B.

In order to investigate changes in response amplitude across testing time points, we computed paired sample t-tests of amplitude. We considered only the first two harmonics (1 Hz and 3 Hz), for which responses were significant at at least one testing time point. Paired t-test (two tailed) of RC1 amplitudes at 1 Hz showed that amplitudes at T2 are significantly higher than (nearly double) those at T1 ($t(1,30) = 2.74, p_{FDR} < 0.01$, corrected for two comparisons), while no significant difference between testing time points was observed at 3 Hz ($t(1,30) = 0.71, p_{FDR} = 0.48$, corrected for two comparisons). There was no significant difference in phase at 1 Hz (the only significant harmonic at both testing time points) between the testing time points (circular t-test; $p = 0.55$).

Development of whole word representation processing (words–pseudowords). The words–pseudowords contrast was designed to detect whole word representation processing. We first calculated RCA on each testing time point separately, as was done for the other two contrasts. Since EEG responses to this contrast were relatively weak, we further pooled the sensor-space data across testing time points to calculate the RCA filters.

As shown in Figure 2A&B, the topographies of RC1 trained on T1 and T2 separately include activation over occipital and temporal electrodes, with relatively weak and noisy responses. The topography of RC1 trained on the two time points together includes activation at more anterior left vOT electrodes (Figure 2C). Projected data were statistically significant at 1 Hz ($p_{FDR} < 0.01$) and 5 Hz ($p_{FDR} < 0.05$) at T1, and only at 1 Hz ($p_{FDR} < 0.01$) at T2 (Figure 2D). A paired-sample t-test of projected amplitude at the first harmonic (significant at both testing time points) showed no significant difference between T1 and T2 ($t(1,30) = 0.33, p = 0.74$). Phase comparisons between testing time points at 1 Hz (Figure 2E) were also not significant (circular t-test; $p = 0.17$).

Development of coarse print tuning (words–pseudofonts). For responses to the words–pseudofonts contrast, aimed at coarse print tuning, the topography of RC1 at T1 includes bilateral peaks at posterior vOT electrodes (Figure 3A). At T2, the topography is more lateralized and maximal at left vOT electrodes (Figure 3B). The projected data were statistically significant in the first three harmonics (1 Hz, 3 Hz, and 5 Hz, $p_{FDR} < 0.001$; corrected for 10 comparisons) at T1, and all five harmonics at T2 (1 Hz, 3 Hz, and 5 Hz, $p_{FDR} < 0.001$, 7 Hz and 9 Hz, $p_{FDR} < 0.05$; corrected for 10 comparisons, Figure 3C). A paired sample t-test (two-tailed) showed no significant difference in RC1 amplitude between T1 and T2 ($t(1,30) = 1.10, p = 0.28$). Phase comparisons (Figure 3D) between testing time points at 1 Hz, 3 Hz and 5 Hz showed no significant difference (circular t-test; $p_{FDR} > 0.37$, corrected for 3 comparisons); phases were not compared for 7 Hz and 9 Hz given that signals were significant at only one testing time point.

Brain-behavior relationships. Brain-behavior analyses were performed to assess the relationship between component-space EEG amplitudes and behavioral reading abilities: Rapid naming abilities (RANcolor, RANletter), word reading efficiency (TOWRE), and word decoding ability (WJ). See *Behavioral*

Assessments in *Materials and Methods* section and Table 1 for detailed information.

At T1, no clear relationships were found between response amplitudes of RC1 and reading scores for all three contrasts either before or after outlier removal (all $p_{FDR} > 0.17$, corrected for 4 comparisons).

However, brain-behavior relationships changed with 2 years of reading instruction in school (T2), as illustrated by significant linear relationships between response amplitudes to words–pseudofonts with WJ ($p_{FDR} < 0.05$, corrected for 4 comparisons, Figure 4A), as well as pseudowords–nonwords with RANletter ($p_{FDR} < 0.05$, corrected for 4 comparisons) and WJ ($p_{FDR} < 0.001$, corrected for 4 comparisons, Figure 4B), after removal of outliers. No significant relationships were found between response amplitudes to the words–pseudowords contrast and reading scores (all $p_{FDR} > 0.28$).

Even more interestingly, we found a strong correlation between brain signal *improvement* (Amplitude at 1 Hz_{T2} − Amplitude at 1 Hz_{T1}) for visual word form structure (pseudowords-nonwords) and letter reading speed *improvement* ($\text{RANletter}_{T2} - \text{RANletter}_{T1}$): ($p_{FDR} < 0.01$, after outlier removal and correction for five comparisons), see Figure 4C.

Behavioral Results. For the repetition detection task performed during EEG sessions, *d'* is summarized in Table S1. We conclude that participants were equally engaged throughout the three conditions of the experiment at each testing time point, and that older children performed better than younger children.

Discussion

Over two years of school reading instruction, visual word form structure processing demonstrated striking patterns of growth: EEG responses nearly doubled in amplitude with increased left lateralization (Figure 1). More importantly, cortical entrainment of visual word form structure was a good predictor of the growth in reading fluency. No such changes were observed for whole word representation processing, indicating that sensitivity to visual word form structure is uniquely linked to growth in reading fluency. Theoretically, this finding sheds new light on current theories of word reading development and provides implications for models of visual word recognition. Practically, it yields insights for educational interventions and activities to improve reading fluency.

Development of brain sensitivity to visual word form structure (pseudowords–nonwords) in early readers. Behavioral studies have shown that sensitivity to visual word form structure develops rapidly through exposure to print (18), even in fully unfamiliar and complex scripts (19). Here, we measured changes in *brain circuits* (in terms of response amplitude and topography) underpinning word form structure processing over two years of reading instruction. The doubled response amplitude likely reflects increased sensitivity to orthographically reasonable word form structure as a result of reading expertise due to repeated exposure to print (20).

The maximally reliable spatial component of word form structure processing was distributed over the occipito-temporal (OT) regions and became increasingly left lateralized over two years (Figure 1A&B). Presumptive activation over the left OT area here is consistent with previous studies showing its involvement in prelexical orthographic processing (21–23). For example, an fMRI study with adult participants found sensitivity of the left lateral fusiform (occipitotemporal) gyrus to familiar over unfamiliar letter sequences (as measured by mean positional bigram frequency), even when these sequences did not resemble words (21) .

Previous studies have not examined the lateralization changes of word form structure processing in early readers. Instead, most previous findings are centered on the lateralization changes of the coarse neural tuning effect and support the assumption that the degree of left lateralization for print is dependent on the level of reading expertise with a script (24, 25). Right hemisphere activation is usually detectable in the early stages of learning to read, and tends to reduce or disappear as development proceeds (26). This reduced involvement of right hemisphere areas over the course of reading development reflects the experience-driven functional refinement of the skill zone for reading, which manifests as greater recruitment of left hemispheric OT regions (26). The current study extends previous findings on lateralization changes of coarse print tuning to further reveal that left OT regions attune specifically to word form structure during early reading acquisition.

Correlations between brain sensitivity to visual word form structure and reading fluency. We also found that brain sensitivity to word form structure contributes to reading fluency. Children with larger response amplitudes to word form structure had better reading performance in rapid letter naming (RAN for letters) as well as word decoding (measured with letter-word identification using WJ) at T2. This finding aligns with the observation of brain sensitivity to visual word form structure in skilled adult readers but not in dyslexic individuals (27). Such deviation between typical and dyslexic readers suggests a connection between sufficient tuning to visual word form structure and typical reading skills. In relation to this, a developmental ERP study demonstrated neural sensitivity to visual word form structure in 7-year-old children with high reading ability, but not in those with low reading skill (10). Collectively, the current study and previous ERP studies provide evidence that the ability to detect and apply visual word form structure is a crucial factor promoting reading success (28).

Correlation between development of brain signal and growth in reading fluency. Remarkably, our results showed that the development of brain sensitivity to word form structure is significantly correlated with growth in early reading fluency. We found that early readers with larger changes in amplitudes to word form structure had higher improvement in RAN (letters not colors). A potential interpretation for this relationship is that beginning readers who are slow to identify individual letters may not activate the letters in memory close enough in time to encode the letter combinations that occur most frequently in print (29). This is consistent with the assumption that the ability to sufficiently automatize letter recognition interferes with letter string processing and growth of orthographic knowledge (30). For instance, Bowers and colleagues reported a deficit in word form structure learning in their RAN deficit group (31). Numerous studies have suggested that RAN is a strong predictor of reading fluency (for a meta-analysis, see (32); for a review, see (33)). Given that no significant cor-

relations were found with RAN for colors, the current study provides support for the idea that the growth of brain sensitivity to visual word form structure specifically interferes with reading fluency and reading development, instead of general articulation speed. To conclude, the development of visual word from structure encoding mechanisms is an important aspect of word recognition skill that allows readers to process letters and letter combinations rapidly and fluently.

Evidence for sensitivity to whole word representation in early readers. Brain responses to word-level representations have also been studied previously; inconsistent results are presumably due to this effect being less pronounced and more task- and developmental stage-dependent (4, 34). For example, past ERP studies (4, 10) reported larger amplitudes for words than for pseudowords in children, which was also demonstrated in adult ERP (35) and SSVEP (36) studies. In contrast, Brem et al. (37) (an ERP-fMRI study with children and adolescents) and Hauk et al. (38) (an ERP study with adults) reported stronger neural responses for pseudowords than for words, while yet other studies found null effects (3, 39–41).

The current study was able to capture brain signals related to representation of familiar words over occipital and temporal regions, potentially due to slower presentation rates in an SSVEP alternation paradigm with a repetition detection task. However, no correlations between response amplitudes and reading skills were found for this contrast in either of the two testing time points. The ambiguity of brain-behavioral relationships here might be due to the small size of this word representation effect (3). Future studies can examine this issue further by recruiting participants spanning a larger range of word representation processing and/or by employing more explicit tasks (e.g., lexical decision task), which might better engage this type of processing. An alternative explanation is that the words we used were all very high frequency, simple, and short (3 letters). Such stimuli do not require much attention on word decoding, resulting in less robust correlations with reading skills related to decoding (WJ) and automatized letter naming (RAN for letters). This speculation could be resolved through a further test of word form structure learning and reading development with more complex, longer stimuli with variable frequencies.

Implications for reading models and educational neuroscience research. Response amplitudes to coarse neural tuning were significantly correlated with amplitudes to word form structure, instead of whole word representation, at both testing time points (Figure S4). Moreover, similar correlations between brain signal and reading skill (mainly word decoding ability) were found for contrasts of coarse neural tuning and word form structure. These findings may indicate that, at least during early reading acquisition, the words-pseudofonts contrast (coarse neural tuning) often used in previous studies is mainly picking up information about word form structure rather than whole word processing. Word form structure processing plays a crucial role in reading fluency and is shown to be a good predictor for reading growth.

Sensitivity to word form structure has been rarely or not even taken into account in current models of visual word recognition (19, 42, 43). Most of these models, including Dual Route Cascaded (DRC) model (44) and triangle model (45), have focused mainly on properties of lexical-semantic and phonological representations and processes. It will be important for future research and reading models to implement visual word form structure in a manner analogous to what has been done with semantics and phonology. This extra word form structure component of the model would provide insights for bridging educational practice and neuroscience to improve early reading fluency.

Limitations. Limitations of the present study include the unbalanced condition order and small sample size. We cannot rule out the potential for serial order carryover effects from one condition to the next. However, the current condition order started with the strongest stimulus contrast (words–pseudofonts), followed by weaker contrast (words–pseudowords), and ended with the most subtle contrast (pseudowords–nonwords). As a consequence, any carryover effects could have reduced and underestimated the condition effect in the current study. The small sample sizes of the grade groups limit us from making conclusions about different brain-behavior relations based on different grades. Future studies with larger sample sizes in each age and/or grade group will enhance our ability to trace the developmental trajectories of brain-behavior relationships due to increasing reading expertise in children at different stages of learning to read.

Conclusion

This longitudinal study demonstrates developmental changes—in terms of response amplitudes and left lateralization—of visual word form structure processing in early readers. Moreover, the word form structure effect became stronger in faster readers, supporting its functional role in early reading fluency. No such changes were observed for responses to whole word representation. Taken together, our results suggest that word form structure processing, which is indispensable for further accessing the phonology and semantics of written words, may be a good predictor for reading ability growth. Such knowledge is crucial to better understand how children develop sensitivity to visual words and is insightful for models of visual word recognition.

Materials and Methods. Further methodological details are given in SI Methods.

Participant Sample. Thirty-three English-speaking children, with normal or corrected-to-normal vision and no reading disabilities, completed behavioral and EEG sessions at two time points (two years apart). Two participants were excluded due to data quality issues, resulting in $N = 31$ participants (16 males, 15 females). The T1 sample included 10 kindergarteners, 12 first graders, and 9 second graders (ages are summarized in Table 1).

Behavioral Assessments. All children at both time points were tested on handedness (Edinburgh Handedness Inventory, (46)), phonological awareness and rapid naming abilities (Rapid Automatized Naming (RAN) of colors and letters, (47)), word reading efficiency (Test of Word Reading Efficiency, Second Edition, TOWRE-II, (48)), and word decoding ability (Woodcock-Johnson Tests of Achievement, Fourth Edition, WJ-IV, (49)). Results of behavioral assessments are summarized in Table 1.

Table 1. Subject characteristics and behavioral assessments at T1 and T2. Values are mean(±SD).

	Group 1	Group 2	Group 3
Sex (female/male)	3/7	6/6	6/3
Handedness (right/left)	9/1	10/2	8/1
T1	Kindergarten	Grade 1	Grade 2
Age in years	5.88(±0.42)	6.81(±0.31)	7.74(±0.24)
Test of word reading efficiency (TOWRE)	10.55(±9.32)	63.33(±28.07)	79.22(±22.73)
Rapid automatic naming of colors (RANcolor)	42.80(±9.02)	37.50(±7.66)	36.89(±9.23)
Rapid automatic naming of letters (RANletter)	45.11(±15.00)	25.33(±6.34)	22.33(±2.78)
Word decoding ability (WJ)	15.50(±5.38)	43.83(±14.02)	51.44(±8.56)
T2	Grade 2	Grade 3	Grade 4
Age in years	7.90(±0.38)	8.80(±0.34)	9.73(±0.23)
Test of word reading efficiency	79.30(±21.11)	100.42(±13.46)	104.11(±19.41)
Rapid automatic naming of colors (RANcolor)	30.20(±4.96)	31.42(±5.99)	27.63(±7.40)
Rapid automatic naming of letters (RANletter)	21.09(±3.20)	19.03(±3.89)	18.49(±4.27)
Word decoding ability (WJ)	54.11(±7.77)	62.33(±5.10)	62.22(±3.80)

TOWRE: Number of real words and pronounceable nonwords read in 45 seconds. RAN: Time (ms) used to quickly and accurately name all stimuli (e.g., letters or colors) on a test form. WJ: Number of correctly named letters and words until getting 6 in a row wrong. T1: First testing time; T2: Second testing time (two years later).

Stimuli and Experimental Paradigm. Stimuli were categorized as words, pseudofonts, pseudowords, or nonwords. Using these four categories, three different contrast conditions were investigated: Coarse print tuning (words–pseudofonts alternation), whole word representation processing (words–pseudowords alternation), and sensitivity to visual word form structure (pseudowords–nonwords alternation). The three conditions were presented in the aforementioned fixed order (see Limitations). Each condition started with a blank screen shown for a random interval of 1500–2500 ms. Then, a given image was presented for 500 ms with a complete alternation cycle of 1000 ms; these periods denote the 2-Hz *base frequency* and a 1-Hz *alternate frequency*, respectively. Participants performed a repetition detection task, pressing a button on a response pad with their preferred hand after detection of a stimulus presented three times consecutively. Stimulus conditions and examples are shown in Figure 5.

Experimental Procedure and Data Acquisition. Participants sat in a dimly lit room 1 m away from a computer screen. Prior to EEG recording, participants were instructed on the repetition detection task and performed a brief practice session with corrective feedback. EEG sessions took approximately 45–50 minutes, including setup and between-trial breaks. As compensation, each child received a small token gift (e.g. stickers) after the study.

128-sensor EEG (Electrical Geodesics, EGI NA 400 amplifier) were recorded against Cz reference, sampled at 500 Hz. Impedances were kept below 50 kΩ. Recorded data were digitally low-pass (50 Hz) and high-pass (0.3 Hz) filtered offline using Net Station Waveform Tools. Filtered data were then imported into in-house signal processing software for preprocessing.

EEG Preprocessing and Analysis. Continuous EEG data were re-sampled to 420 Hz, and sensors with more than 15% of samples exceeding a ± 60 μV amplitude threshold were interpolated with an average from six neighboring sensors. EEG were then re-referenced to average reference (50) and segmented into 1-second epochs. Epochs with more than 10% of time samples exceeding ± 60 μV, or with any time sample exceeding ± 70 μV were excluded from further analyses.

We spatially filtered sensor-space data using Reliable Components Analysis (RCA), which computes optimal weightings of sensors to maximize between-trials covariance relative to within-trials covariance (51, 52). RCA has been shown to capture and isolate different neural processes arising from different underlying sources (51), and topographies of corresponding spatial filters can strongly resemble lead fields generating observed SSVEPs (52).

For each RCA calculation, we computed the first five maximally reliable components (RCs) over 1-second epochs of sensor-by-feature (i.e., real and imaginary Fourier coefficients at selected frequencies of interest) data.

For the alternate stimulation frequency we computed RCA over the first five harmonics (1 Hz, 3 Hz, 5 Hz, 7 Hz, and 9 Hz), excluding base harmonics (2 Hz, 4 Hz, 6 Hz, 8 Hz, and 10 Hz). We computed RCA separately for each condition and testing time point to investigate different neural mechanisms underpinning multiple levels of information processing that were evoked by different stimulus contrast conditions over developmental stages. Similar computations for base simulation frequency harmonics are provided in the Supplement.

Statistical Analyses of RC Data. For each RC, a 1-second epoch of frequency-domain component-space data contained 10 data points (5 harmonics times 2 real and imaginary coefficients). Component-space data were first averaged across 1-second epochs on a per-participant basis. Following this, statistical analyses were performed across distributions of participants. We performed Hotelling's two-sample t^2 tests (53) on distributions of real and imaginary Fourier coefficients on a per-harmonic, per-component basis to identify responses that differed significantly from zero in the complex plane, correcting for multiple comparisons using False Discovery Rate (FDR, (54)). In addition, statistical significance of RCA coefficients was assessed via permutation test over 500 surrogate data iterations using the procedure described in (14).

To identify and compare overall response amplitude *across* harmonics, we computed the root sum of squares (henceforth referred to as RSS projected amplitude) by projecting each complex-valued data point onto the mean vector (55, 56), and then computing the square root of summed squared amplitudes across multiple harmonics (57, 58).

We conducted repeated-measures analyses to compare results across testing time points. For alternate harmonics (RCA trained on each condition separately), we performed paired-sample, two-tailed t tests on RSS projected amplitudes across testing time points for RC separately in each condition; for words–pseudowords and pseudowords–nonwords, where only the first harmonic (1 Hz) was significant, we further performed paired-sample t tests (two tailed) on projected amplitude at 1 Hz.

We used the Circular Statistics toolbox (59) to compare distributions of RC phases for harmonics with significant responses at both testing time points. For alternate harmonics, we computed circular t tests at a given harmonic between two testing points; here, words–pseudofonts results were corrected for 3 comparisons, while words–pseudowords and pseudowords-nonwords involved no multiple comparisons as only the first harmonic was significant. RSS amplitude comparisons and circular statistics of RC phases for base harmonics are provided

in the Supplement.

Analysis of brain-behavior relationships. We explored possible brain-behavior relationships using RSS projected amplitudes from alternate RCA output for words–pseudofonts and 1-Hz amplitude for words–pseudowords and pseudowords–nonwords in conjunction with reading scores. To test the association between brain signals and reading performance at different developmental stages, we performed linear regressions of response amplitude with each of the reading scores (i.e., RANcolor, RANletter, TOWRE, and WJ), separately for T1 and T2. Next, to relate the development of brain responses to changes in reading performance, we computed linear regression of the change in response amplitude from T1 to T2 ($RSS_{T2} - RSS_{T1}$ for words–pseudofonts, amplitude at 1 Hz_{T2} − amplitude at 1 Hz_{T1} for words–pseudowords and pseudowords–nonwords), with corresponding reading performance improvements (i.e., $\mathrm{RANcolor}_{T2} - \mathrm{RANcolor}_{T1}$, $\mathrm{RANletter}_{T2} - \mathrm{RANletter}_{T1}$, $\mathrm{TOWRE}_{T2} - \mathrm{TOWRE}_{T1}$, and $\mathrm{WJ}_{T2} - \mathrm{WJ}_{T1}$). Outliers determined by Cook's Distance (60) based on the regression model were removed if they exceeded the $4/n$ threshold (n total data points).

Behavioral Analysis. For the repetition detection task performed during EEG sessions, we computed d' based on the z-transformed probabilities of hits and false alarms (61). A two-way repeated-measures ANOVA with within-subjects factors of condition and testing time point was computed on d' across conditions and time points.

ACKNOWLEDGMENTS. We thank the students, their families, and teachers for participating. We also thank Angie Wang and Rachana Pillai for their help with data collection.

1. S Bentin, Y Mouchetant-Rostaing, MH Giard, JF Echallier, J Pernier, ERP manifestations of processing printed words at different psycholinguistic levels: time course and scalp distribution. *J. Cogn. Neurosci.* **11**, 235–260 (1999).
2. S Brem, et al., Neurophysiological signs of rapidly emerging visual expertise for symbol strings. *Neuroreport* **16**, 45–48 (2005).
3. AK Eberhard-Moscicka, LB Jost, M Raith, U Maurer, Neurocognitive mechanisms of learning to read: print tuning in beginning readers related to word-reading fluency and semantics but not phonology. *Dev. science* **18**, 106–118 (2015).
4. U Maurer, et al., Coarse neural tuning for print peaks when children learn to read. *Neuroimage* **33**, 749–758 (2006).
5. F Wang, U Maurer, Top-down modulation of early print-tuned neural activity in reading. *Neuropsychologia* **102**, 29–38 (2017).
6. A van de Walle de Ghelcke, B Rossion, C Schiltz, A Lochy, Developmental changes in neural letter-selectivity: A 1-year follow-up of beginning readers. *Dev. science* **24**, e12999 (2021).
7. U Maurer, VC Blau, YN Yoncheva, BD McCandliss, Development of visual expertise for reading: rapid emergence of visual familiarity for an artificial script. *Dev. neuropsychology* **35**, 404–422 (2010).
8. M Carreiras, BC Armstrong, M Perea, R Frost, The what, when, where, and how of visual word recognition. *Trends cognitive sciences* **18**, 90–98 (2014).
9. U Maurer, BD McCandliss, The development of visual expertise for words: the contribution of electrophysiology in *Single-Word Reading.* (Psychology Press), pp. 57–77 (2007).
10. J Zhao, et al., Fine neural tuning for orthographic properties of words emerges early in children reading alphabetic script. *J. Cogn. Neurosci.* **26**, 2431–2442 (2014).
11. AM Norcia, LG Appelbaum, JM Ales, BR Cottereau, B Rossion, The steady-state visual evoked potential in vision research: A review. *J. vision* **15**, 4–4 (2015).
12. M Dzhelyova, et al., High test-retest reliability of a neural index of rapid automatic discrimination of unfamiliar individual faces. *Vis. Cogn.* **27**, 127–141 (2019).
13. A Lochy, M Van Reybroeck, B Rossion, Left cortical specialization for visual letter strings predicts rudimentary knowledge of letter-sound association in preschoolers. *Proc. Natl. Acad. Sci.* **113**, 8544–8549 (2016).
14. F Wang, et al., Early detection of sublexical and lexical processing in beginning readers: evidence from steady-state visual evoked potentials (ssveps). *bioRxiv* (2021).
15. JD Yeatman, AM Norcia, Temporal tuning of word- and face-selective cortex. *J. cognitive neuroscience* **28**, 1820–1827 (2016).
16. E Barzegaran, AM Norcia, Neural sources of letter and vernier acuity. *Sci. reports* **10**, 1–11 (2020).
17. F Wang, et al., Distinct neural sources underlying visual word form processing as revealed by steady state visual evoked potentials (SSVEP). *Sci. reports* **11**, 1–15 (2021).
18. SH Deacon, E Commissaire, X Chen, A Pasquarella, Learning about print: The development of orthographic processing and its relationship to word reading in first grade children in french immersion. *Read. Writ.* **26**, 1087–1109 (2013).
19. F Chetail, What do we do with what we learn? statistical learning of orthographic regularities impacts written word processing. *Cognition* **163**, 103–120 (2017).
20. DL Compton, Modeling the growth of decoding skills in first-grade children. *Sci. studies reading* **4**, 219–259 (2000).
21. JR Binder, DA Medler, CF Westbury, E Liebenthal, L Buchanan, Tuning of the human left fusiform gyrus to sublexical orthographic structure. *Neuroimage* **33**, 739–748 (2006).
22. L Cohen, et al., Language-specific tuning of visual cortex? functional properties of the visual word form area. *Brain* **125**, 1054–1069 (2002).
23. BD McCandliss, KG Noble, The development of reading impairment: a cognitive neuroscience model. *Mental retardation developmental disabilities research reviews* **9**, 196–205 (2003).
24. U Maurer, JD Zevin, BD McCandliss, Left-lateralized N170 effects of visual expertise in reading: evidence from Japanese syllabic and logographic scripts. *J. Cogn. Neurosci.* **20**, 1878–1891 (2008).
25. C Spironelli, A Angrilli, Developmental aspects of automatic word processing: language lateralization of early erp components in children, young adults and middle-aged subjects. *Biol. psychology* **80**, 35–45 (2009).
26. PE Turkeltaub, L Gareau, DL Flowers, TA Zeffiro, GF Eden, Development of neural mechanisms for reading. *Nat. neuroscience* **6**, 767–773 (2003).
27. S Araújo, L Faísca, I Bramão, A Reis, KM Petersson, Lexical and sublexical orthographic processing: An ERP study with skilled and dyslexic adult readers. *Brain Lang.* **141**, 16–27 (2015).
28. NJ Conrad, N Harris, J Williams, Individual differences in children's literacy development: The contribution of orthographic knowledge. *Read. Writ.* **26**, 1223–1239 (2013).
29. PG Bowers, J Golden, A Kennedy, A Young, Limits upon orthographic knowledge due to processes indexed bynaming speed. in vw berninger (ed.), the varieties of orthographic knowledge. vol. i: Theoretical and developmental issues (pp. 173-218) (1994).
30. FR Manis, LM Doi, B Bhadha, Naming speed, phonological awareness, and orthographic knowledge in second graders. *J. learning disabilities* **33**, 325–333 (2000).
31. PG Bowers, K Sunseth, J Golden, The route between rapid naming and reading progress. *Sci. Stud. Read.* **3**, 31–53 (1999).
32. S Araújo, A Reis, KM Petersson, L Faísca, Rapid automatized naming and reading performance: a meta-analysis. *J. Educ. Psychol.* **107**, 868 (2015).
33. GK Georgiou, R Parrila, Y Cui, TC Papadopoulos, Why is rapid automatized naming related to reading? *J. experimental child psychology* **115**, 218–225 (2013).
34. G Xue, T Jiang, C Chen, Q Dong, Language experience shapes early electrophysiological responses to visual stimuli: the effects of writing system, stimulus length, and presentation duration. *NeuroImage* **39**, 2025–2037 (2008).
35. O Hauk, C Coutout, A Holden, Y Chen, The time-course of single-word reading: evidence from fast behavioral and brain responses. *NeuroImage* **60**, 1462–1477 (2012).
36. A Lochy, G Van Belle, B Rossion, 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).
37. S Brem, et al., Tuning of the visual word processing system: Distinct developmental ERP and fMRI effects. *Hum. Brain Mapp.* **30**, 1833—-1844 (2009).
38. O Hauk, et al., [Q:] when would you prefer a SOSSAGE to a SAUSAGE? [A:] at about 100 msec. ERP correlates of orthographic typicality and lexicality in written word recognition. *J. Cogn. Neurosci.* **18**, 818–832 (2006).
39. S Araújo, I Bramão, L Faísca, KM Petersson, A Reis, Electrophysiological correlates of impaired reading in dyslexic pre-adolescent children. *Brain cognition* **79**, 79–88 (2012).
40. U Maurer, D Brandeis, BD McCandliss, Fast, visual specialization for reading in English revealed by the topography of the N170 ERP response. *Behav. brain functions* **1**, 1–12 (2005).
41. AM Proverbio, L Vecchi, A Zani, From orthography to phonetics: ERP measures of grapheme-to-phoneme conversion mechanisms in reading. *J. cognitive neuroscience* **16**, 301–317 (2004).
42. F Chetail, Reconsidering the role of orthographic redundancy in visual word recognition. *Front. Psychol.* **6**, 645 (2015).
43. MW Harm, MS Seidenberg, Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. *Psychol. review* **111**, 662 (2004).
44. M Coltheart, K Rastle, C Perry, R Langdon, J Ziegler, Drc: a dual route cascaded model of visual word recognition and reading aloud. *Psychol. review* **108**, 204 (2001).
45. MW Harm, MS Seidenberg, Phonology, reading acquisition, and dyslexia: insights from connectionist models. *Psychol. review* **106**, 491 (1999).
46. RC Oldfield, The assessment and analysis of handedness: the edinburgh inventory. *Neuropsychologia* **9**, 97–113 (1971).
47. R Wagner, J Torgesen, C Rashotte, N Pearson, Comprehensive test of phonological processing–second edition (2013).
48. JK Torgesen, RK Wagner, CA Rashotte, TOWRE-2 examiner's manual. *Austin, TX: Pro-Ed* (2012).
49. FA Schrank, KS McGrew, N Mather, BJ Wendling, EM LaForte, *Woodcock-Johnson IV tests of cognitive abilities.* (Riverside), (2014).
50. D Lehmann, W Skrandies, Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalogr. clinical neurophysiology* **48**, 609–621 (1980).
51. JP Dmochowski, P Sajda, J Dias, LC Parra, Correlated components of ongoing EEG point to emotionally laden attention–a possible marker of engagement? *Front. human neuroscience* **6**, 112 (2012).
52. JP Dmochowski, AS Greaves, AM Norcia, Maximally reliable spatial filtering of steady state visual evoked potentials. *NeuroImage* **109**, 63–72 (2015).
53. JD Victor, J Mast, A new statistic for steady-state evoked potentials. *Electroencephalogr. Clin. Neurophysiol.* **78**, 378 – 388 (1991).
54. Y Benjamini, D Yekutieli, The control of the false discovery rate in multiple testing under

dependency. *The Annals Stat.* **29**, 1165–1188 (2001).
55. C Hou, R Gilmore, M Pettet, A Norcia, Spatio-temporal tuning of coherent motion evoked responses in 4–6 month old infants and adults. *Vis. research* **49**, 2509–2517 (2009).
56. PJ Kohler, E Barzegaran, AM Norcia, BD McCandliss, Parietal contributions to abstract numerosity measured with steady state visual evoked potentials. *bioRxiv* (2020).
57. L Appelbaum, J Ales, B Cottereau, A Norcia, Configural specificity of the lateral occipital cortex. *Neuropsychologia* **48**, 3323–3328 (2010).
58. AI Tlumak, JD Durrant, RE Delgado, JR Boston, Steady-state analysis of auditory evoked potentials over a wide range of stimulus repetition rates: Profile in adults. *Int. J. Audiol.* **50**, 448–458 (2011).
59. P Berens, , et al., Circstat: a matlab toolbox for circular statistics. *J Stat Softw* **31**, 1–21 (2009).
60. RD Cook, Detection of influential observation in linear regression. *Technometrics* **19**, 15–18 (1977).
61. NA Macmillan, CD Creelman, *Detection theory: A user's guide.* (Psychology press), (2004).

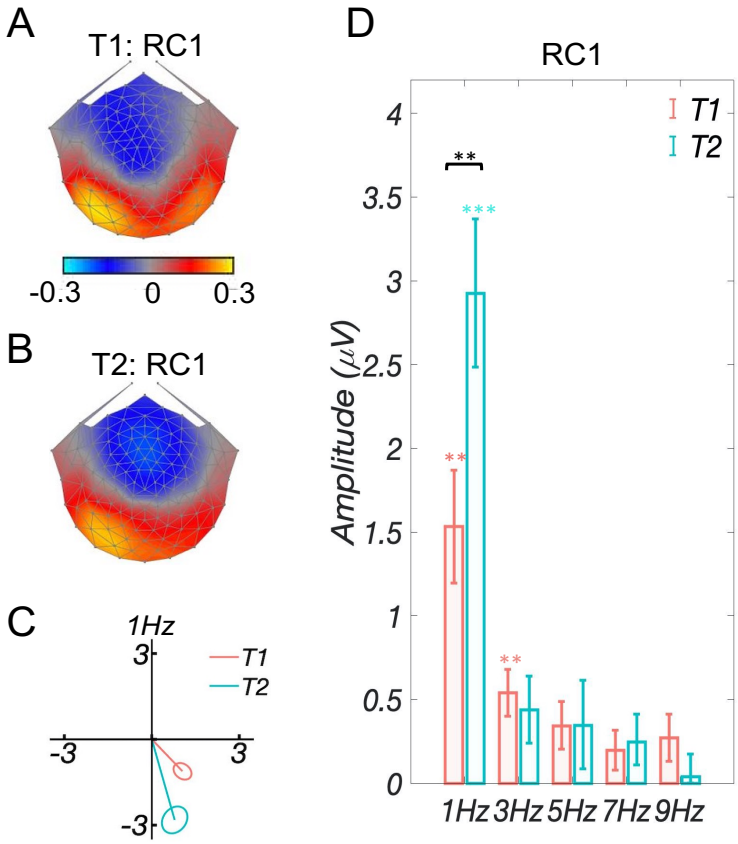

Fig. 1. Alternate Analysis: Developmental changes of visual word form structure tuning (pseudowords–nonwords) in early readers. A: Topographic visualization of the spatial filter (bilateral OT) for the maximally reliable component (RC1) at T1; B: Topographic visualization of the spatial filter (left OT) for RC1 at T2; C: Response data at the first harmonic (1 Hz, the only significant harmonic at both testing time points) presented in the complex plane: Amplitude information is represented as the length of the vector, and phase information in the angle of the vector relative to 0 degree (counterclockwise from 3 o'clock direction), ellipses indicate standard error of the mean (SEM). No significant phase difference ($p = 0.55$) was found between T1 and T2; D: Projected amplitude for each harmonic at T1 (orange) and T2 (blue), respectively. The first two harmonics were significant at T1, while only the first harmonic was significant at T2. A significant difference ($p < 0.01$) was found at 1 Hz between T1 and T2. No significant differences were revealed at other harmonics. **: $p_{FDR} < 0.01$, ***: $p_{FDR} < 0.001$.

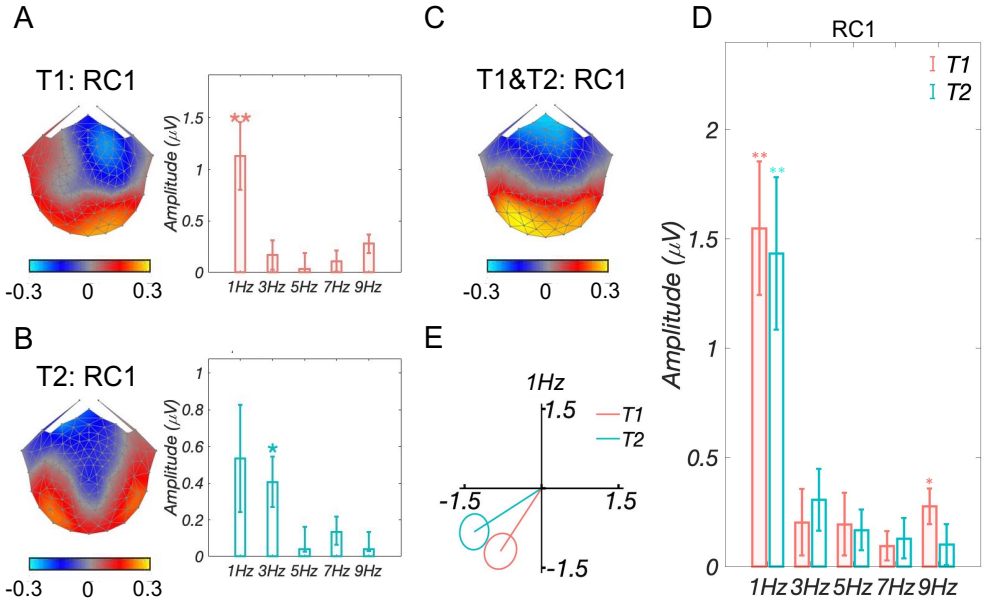

Fig. 2. Alternate Analysis: Sensitivity to whole word representation (words–pseudowords) in early readers. A&B: Topographic visualizations of spatial filters and responses amplitudes for the maximally reliable component (RC1) trained on T1 and T2 separately; C: Topographic visualization of the spatial filter (more anterior left vOT) for RC1 trained on pooled T1 and T2 samples; D: Projected amplitude of RC1 (trained on T1 and T2 together) for each harmonic at T1 (orange) and T2 (blue), respectively. At T1, 1-Hz and 9-Hz harmonics were significant, while only the 1-Hz harmonic was significant at T2. There was no significant RSS amplitude difference ($p = 0.07$) between T1 and T2. Projected amplitude at the first harmonic showed no significant difference ($p = 0.74$) either between T1 and T2; E: Response data at the first harmonic (1 Hz, the only significant harmonic at both testing time points) presented in the complex plane, where amplitude information is represented as the length of the vector, and phase information in the angle of the vector relative to 0 degree (counterclockwise from 3 o'clock direction), ellipse indicates standard error of the mean (SEM). No significant phase difference ($p = 0.17$) was found between T1 and T2; *: $p_{FDR} < 0.05$, **: $p_{FDR} < 0.01$.

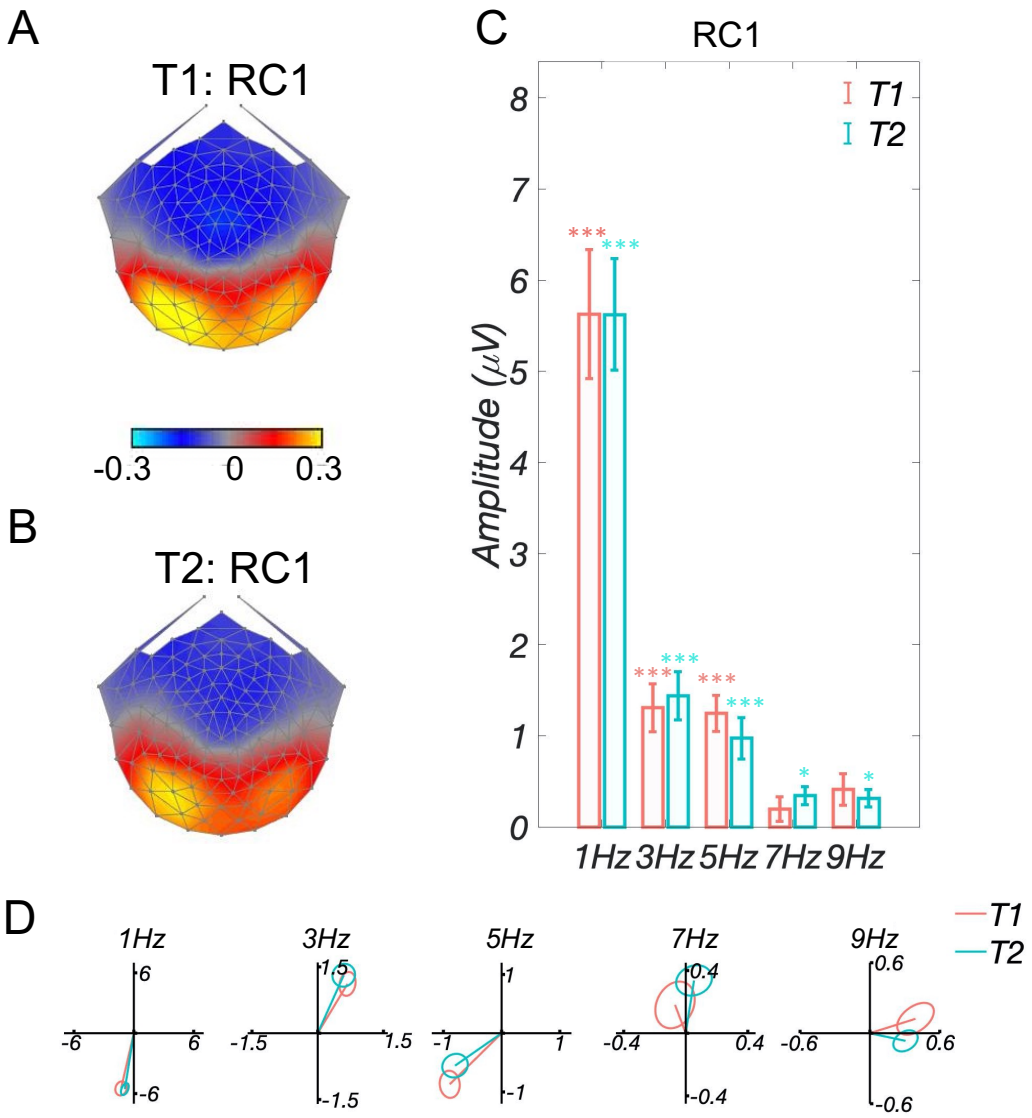

Fig. 3. Alternate Analysis: development of coarse print tuning (words–pseudofonts) in early readers. A: Topographic visualization of the spatial filter (less left lateralized) for the maximally reliable component (RC1) at T1; B: Topographic visualization of the spatial filter (more left lateralized) for RC1 at T2; C: Projected amplitude of RC1 for each harmonic at T1 (orange) and T2 (blue), respectively. The first three harmonics were significant at T1, and all five harmonics were significant at T2. RSS amplitude did not differ significantly ($p = 0.28$) between T1 and T2. D: Response data in the complex plane, with T1 in orange and T2 in blue. Amplitude (vector length) and phase (vector angle, counterclockwise from 0 degrees at 3 o'clock direction) overlap between two testing time points especially for significant harmonics (first three harmonics). Ellipses indicate standard error of the mean. No significant phase differences (1 Hz: $p_{FDR} = 0.61$, 3 Hz: $p_{FDR} = 0.61$, 5 Hz: $p_{FDR} = 0.61$) were found at first three significant harmonics between T1 and T2. *: $p_{FDR} < 0.05$, ***: $p_{FDR} < 0.001$.

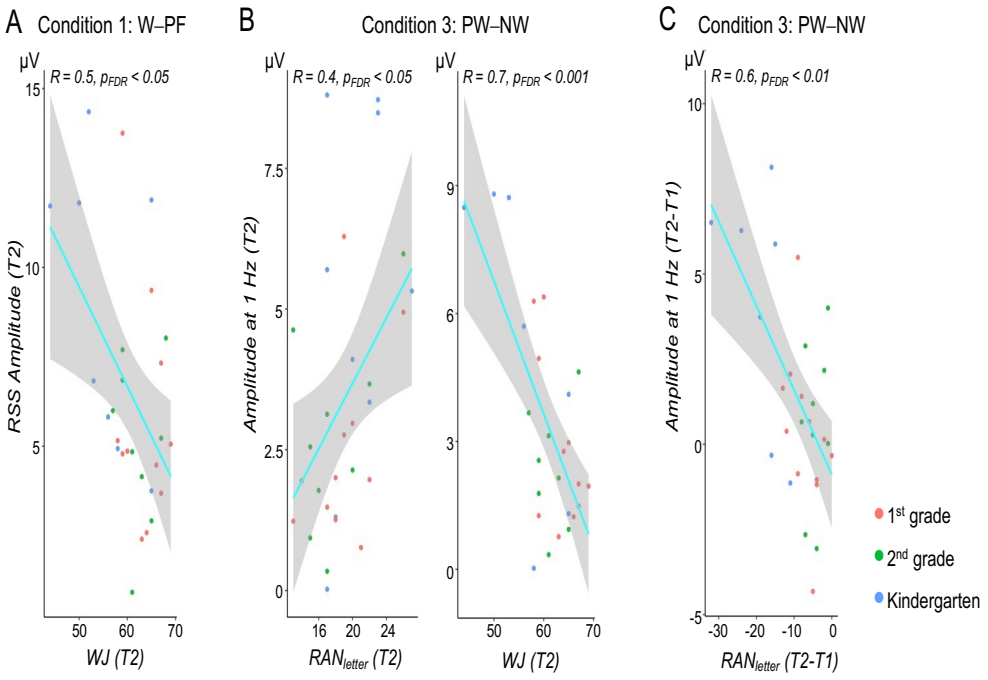

Fig. 4. Statistically significant correlations between reading and brain response amplitudes for words–pseudofonts and pseudowords–nonwords, after outlier removal. A: Regression between RSS amplitude for words–pseudofonts and word decoding abiliy (Woodcock-Johnson, WJ) at T2; B: Regressions between response amplitude at 1 Hz for pseudowords–nonwords and rapid automatized naming of letters ($\mathrm{RANletter}$) and WJ at T2; C: Regression between response amplitude improvement (T2-T1) at 1 Hz of pseudowords-nonwords and $\mathrm{RANletter}$ improvement (T2-T1). No significant relations were found for words–pseudowords.

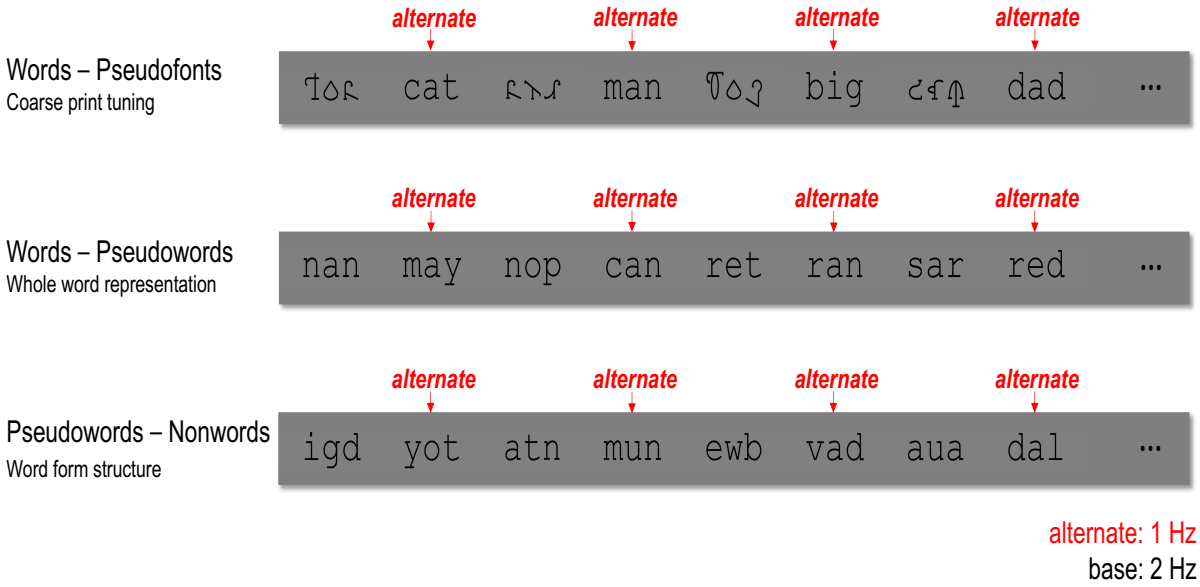

Fig. 5. Experimental Design. Examples of stimuli presented in the experiment. 1-Hz alternates were embedded within a 2-Hz base stream in all three contrasts. The first contrast assessed coarse print tuning with words–pseudofonts alternation. The second contrast assessed whole word representaion processes with words alternating with orthographically legal pseudowords (words–pseudowords alternation). The third contrast assessed visual word form structure processing with orthographically legal pseudowords alternating with orthographically illegal nonwords (pseudowords–nonwords alternation).

Supporting Information Text

SI Methods.

Participant Sample and Data Collection. EEG and behavioral data were collected during October through November, 2019 (T1), and again two years later (T2). The T1 sample included 10 kindergarteners, 12 first graders, and 9 second graders ($m = 6.78$ years, $s = 0.81$ years), and correspondingly 10 second graders, 12 third graders, and 9 fourth graders at T2 (mean age = 8.78,s age = 0.79).

Behavioral Assessments. Behavioral assessments were administered in separate sessions on average 7.97 ($s = 4.97$) days and 4.58 ($s = 7.21$) days either before or after the EEG session for T1 and T2, respectively. All children were tested on two sub-tests of the Comprehensive Test of Phonological Processing, Second Edition (CTOPP-II, (1))—Rapid Automatized Naming (RAN) of colors and letters—for their phonological awareness and rapid naming abilities. In addition, the Sight Word Efficiency (SWE) and the Phonemic Decoding Efficiency (PDE) subtests of the Test of Word Reading Efficiency, Second Edition, (TOWRE-II, (2)) were used to measure efficiency of sight word recognition and phonemic decoding ability. We used the letter-word identification subset of Woodcock-Johnson Tests of Achievement, Fourth Edition (WJ-IV; (3)) to assess word decoding ability. Finally, the handedness of each participant was defined by the Edinburgh Handedness Inventory (4).

Stimuli and Experimental Paradigm. Four types of stimuli were used: Words, pseudofonts, pseudowords, and nonwords. Words were high-frequency (average: 2110 per million, range: 629–5851 per million) three-letter CVC English words rendered in Courier New font. Pseudofonts were word stimuli presented in the Brussels Artificial Character Set font (BACS-2, (5)), which provide well-matched visual controls of Courier New letters. Pseudowords were generated on an item-by-item basis by semi-randomly rearranging letters of words stimuli while retaining a CVC structure, rendering them still pronounceable with well-matched bigram frequencies ($t(30) = 0.26, p = 0.79$) and orthographic neighborhood sizes ($t(30) = 0.79, p = 0.43$) with words. Finally, nonwords stimuli were also built on an item-by-item basis by semi-randomly shuffling letters across pseudowords to produce unpronounceable exemplars not following English orthographic and phonological rules. Bigram frequencies and orthographic neighborhood sizes of nonwords and pseudowords differed significantly (both $t(30) > 6.26$, both $p < 0.001$). All stimuli images, spanning 7.5 (horizontal) by 2 (vertical) degrees of visual angle, were presented in black font on a gray background at the center of the screen. Stimuli images and background contrast was set at 95%.

Twelve successive 1-second epochs (each containing 2 stimuli) comprised one trial. For each condition, twelve trials containing pre-randomized sequences were presented in the same order for every participant: Four "non-target" trials, four "terminal" trials, and four "catch" trials depending on whether and when repetition targets appeared. Specifically, non-target trials contained no repeated stimuli (Figure S1A); in terminal trials, repeated stimuli appeared at the end (Figure S1B); while repeated stimuli randomly appeared elsewhere during catch trials (Figure S1C). Only one target appeared in each terminal and catch trial. Feedback on task performance was provided after the end of each trial. Participants were given breaks as needed between trials.

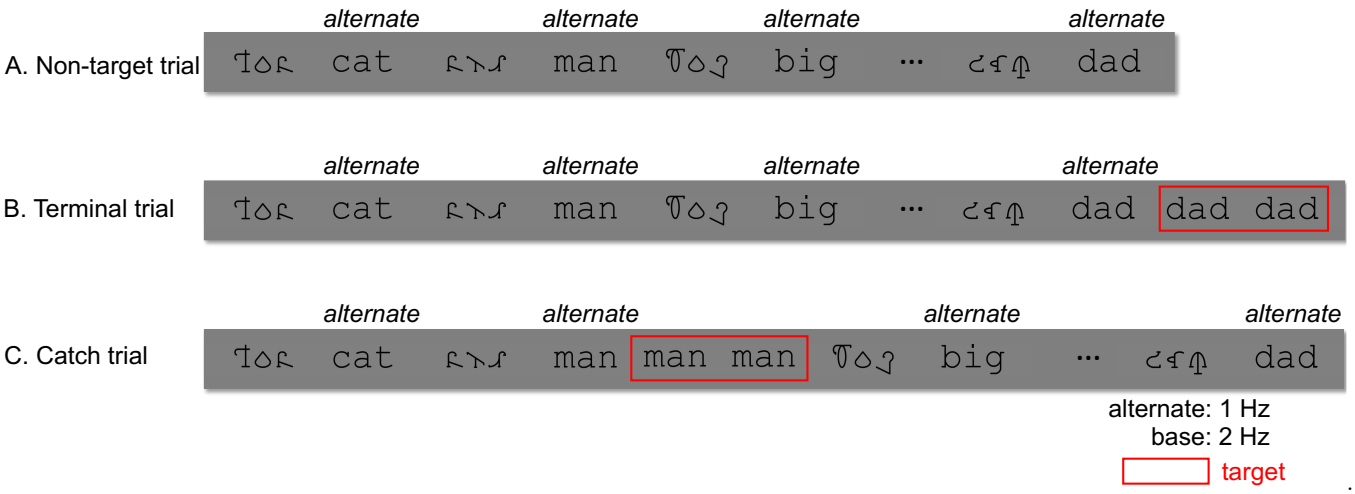

Fig. S1. Different types of trials in words–pseudofonts contrast. Twelve trials of pre-randomized sequences were presented for each condition (words–pseudofonts contrast shown here as an example): Four non-target trials (A), four terminal trials (B), and four catch trials (C). EEG data corresponding to the four catch trials were excluded from analysis due to excessive response-related movements during recording.

Because EEG data from catch trials contained movement artifacts related to the key press, EEG data corresponding only to the four non-target trials and four terminal trials were analyzed. For these 8 trials included per participant, the first and last 1-second epochs of each 12-second trial were excluded to avoid transient responses associated with ocular artifacts which occurred more at beginnings and endings of trials. In all, 10 epochs (i.e., seconds) of data from each of 8 trials per participant were analyzed.

EEG Preprocessing and Analysis. During preprocessing, EEG data were first re-sampled to 420 Hz. This was done to ensure an integer number (herein 7) of time samples per video frame given the frame rate of 60 Hz, as well as integer numbers of frames (i.e., 60 frames per 1 Hz cycle and 30 frames per 2 Hz cycle) per stimulus cycle given the current stimulation frequencies of 1 Hz and 2 Hz. Then, sensors were interpolated with the average of data from six neighboring sensors if more than 15 % of samples exceeded a $\pm$ 60 μV amplitude threshold. The continuous EEG data were then re-referenced to average reference (6) and segmented into 1-second epochs. Epochs with more than 10 % of time samples exceeding a $\pm$ 60 μV noise threshold, or with any time sample exceeding an artifact threshold of ($\pm$ 70 μV) (reflecting e.g., eye blinks, eye movements or body movements), were rejected from further analyses.

Recursive Least Squares (RLS) filters were then used to filter the epoched data in the time domain (7). The filters were tuned to each of the analysis frequencies (i.e., base frequency 2 Hz, alternate frequency 1 Hz, and their harmonics) and converted to complex amplitude values by means of the Fourier transform. Complex-valued RLS outputs were decomposed into real and imaginary Fourier coefficients as input for subsequent analyses.

Reliable Components Analysis (RCA). Instead of analyzing periodic responses from several pre-selected (literature- or SNR-based) sensors that may lead to a bias in reporting false positives (8) and fail to separately describe activities underlying multiple cortical sources (9), we spatially filtered the sensor-space data using RCA, which computes optimal weightings of sensors to maximize between-trial covariance relative to within-trial covariance (10, 11). The resulting weight vectors (spatial filters) are vectors of length N_{sensor} which represent linear weightings of sensors (i.e., linear spatial filters). Sensor-space EEG data can then be projected through (i.e., multiplied by) these weight vectors to transform the spatial dimension of the data from sensors to spatial components, where the number of components is typically much smaller than the number of sensors. The spatial filtering process of RCA is conceptually similar to process of deriving linear spatial components using Principal Components Analysis (PCA) in that both involve eigenvalue decompositions returning multiple components sorted in descending order of criteria explained; however, where PCA optimizes sensor weightings to maximize variance explained within one data matrix, RCA computes sets of weights to maximize covariance across pairs of data matrices (trials). Moreover, in contrast to PCA, the spatial filters computed by RCA are not constrained to be orthogonal (10). Technical details on the RCA technique are provided by (10, 11).

RCA produces three outputs that are of interest in our present analyses. First, the weight vectors themselves, which are the eigenvectors of the RCA calculation, can be visualized as scalp topographies using a forward-model projection (12). These visualizations are thought to reflect the propagation of synchronized activity captured by the weight vectors onto the scalp (12). Next, each weight vector (eigenvector) is accompanied by an eigenvalue, which we refer to as the coefficient of that component. These coefficients give a measure of optimized across-trials correlation for each component. Finally, the component-space data (i.e., the spatially filtered sensor-space data) undergo subsequent analysis and visualization.

RCA Calculations. Given 1-second epochs of sensor-by-feature (i.e., real and imaginary Fourier coefficients at selected frequencies of interest) EEG data matrix, RCA computes optimal weightings of sensors (i.e., linear spatial filters), through which the data transformed to 1-second epochs of component-by-feature matrices, with each component represents a linear combinations of sensor (11). For each calculation we computed five RCs.

Statistical Analyses of RC Data. Statistical significance of the RCA coefficients (eigenvalues) was assessed using permutation test. As in (13), we generated surrogate versions of the sensor-space data: For every epoch of a 10-second trial (accounting for within-trial, across-epoch correlations introduced by RLS filtering) the phases of the data were rotated by a random angle, independently for each harmonic. We computed RCA over 500 such surrogate versions of the data and treated the resulting distributions of RCA coefficients as null distributions for computing p-values of the observed (intact) coefficients. These p-values were then corrected for multiple comparisons using FDR (14). The significance of coefficients for *alternate* RCA calculations were tested separately for each condition, given RCA was computed separately for each condition in order to investigate different neural mechanisms underpinning multiple levels of word information processing. In contrast, the significance of coefficients for *base* RCA calculations were tested on RCA outputs that performed on three conditions together, which enables direct quantitative comparisons of responses to low-level visual stimulus features in a shared component space (13).

When group-level responses were statistically significant at at least three harmonics, estimated latency was calculated using $d = 1/360 \times \delta\varphi/\delta f$, where $\delta\varphi$ is the change in phase over harmonics in degrees and δf is the change in frequency in Hz (15).

Base Frequency Analysis. For the base stimulation frequency we first computed RCA over the first five harmonics (2 Hz, 4 Hz, 6 Hz, 8 Hz, and 10 Hz). As done in a previous analysis of a superset of T1 data (13), these calculations were performed on all conditions together to enable direct quantitative comparisons of responses to low-level visual stimulus features in a shared component space. However, we computed RCA for T1 and T2 separately in order to investigate potential developmental changes of such responses.

For base harmonics results (RCA trained on three conditions together), we performed a two-way repeated-measures ANOVA (within-subjects factors of condition and testing time point) on RSS projected amplitudes for each RC separately. Using the Circular Statistics Toolbox (16), we also computed circular ANOVA across three conditions at each harmonic within time point, followed by pairwise circular t tests; results were FDR-corrected across 5 comparisons (5 significant harmonics).

Visualization of RCA Data. In visualizing the EEG results, we first present scalp topographies (i.e., forward-model projection of the spatial filters) of maximally reliable component (RC1). Second, we present mean responses of projected data as vectors in

the 2D complex plane: The vector length represents the response amplitude and the angle of the vector relative to 0 degrees (counterclockwise from the 3 o'clock direction) represents the phase. Error ellipses around the vectors are the standard errors of the mean (SEM). Third, we present amplitudes (μV) at each harmonic for each component in bar plots, with statistically significant responses indicated with asterisks, as determined by adjusted p_{FDR} values of Hotelling's two-sample t^2 tests of combined real and imaginary coefficients. Finally, phase values (in degrees) were plotted across harmonics, with the line of best fit through significant harmonics and latency estimate in ms when latency could be estimated and compared between two testing time point.

Analysis of the correlations of alternate amplitudes across three conditions. We examined amplitude relationships across three conditions using RSS projected amplitudes from alternate RCA output for words–pseudofons and 1-Hz (the only significant harmonic) amplitude for words–pseudowords and pseudowords–nonwords. To test the association between amplitudes for different conditions at different developmental stages, we performed linear regressions of response amplitudes, separately for T1 and T2. Outliers determined by Cook's Distance (17) based on the regression model were removed if they exceeded the $4/n$ threshold (n total data points).

Analysis of laeralization of brain responses to three conditions. To investigate lateralization of brain responses to each stimulus contrast, first, we simulated regions of interest (ROIs) for the left and right visual word form area (VWFA), which is specialized for computation of abstract visual word forms (18) using EEG source simulation toolbox (19). Second, we projected electrode-by-frequency-by-trial sensor space data matrix at both testing time points onto left and right VWFA ROIs, respectively, resulting projected data matrix of component-by-frequency-by-trial. Finally, a three-way repeated-measures ANOVA with within-subjects factors of condition, testing time point, hemisphere was computed on RSS projected amplitude Figure S5A.

SI Results.

Behavioral Results. We computed *d'* based on the z-transformed probabilities of hits and false alarms (20). A two-way repeated-measures ANOVA of *d'* revealed a main effect of testing time point, with a higher hit rate at T2 compared to T1 ($F(1, 185) = 17.04, p < 0.001$). No significant difference was found across the three conditions ($F(2, 185) = 1.95, p = 0.15$), and no interaction between condition and testing time point was found ($F(2, 185) = 1.80, p = 0.17$).

Table S1. *d'* of the repetition detection task performance during EEG sessions for each of three conditions at T1 and T2.*d'* was computed based on the z-transformed probabilities of hits and false alarms. Values are mean(±SD).

	words–pseudofonts	words–pseudowords	pseudowords–nonwords
T1	1.00(±0.20)	1.04(±0.30)	1.10(±0.27)
T2	1.22(±0.18)	1.53(±0.24)	1.28(±0.23)

Permutation Test Results. Permutation testing was performed on alternate and base RCA coefficients. As shown in Figure S2, for alternate RCA calculations, permutation testing revealed statistically significant (permutation test $p_{FDR} < 0.001$, each corrected for 5 comparisons) RC1–5 coefficients for words–pseudofonts contrast at both T1 (Figure S2A) and T2 (Figure S2B). For words–pseudowords contrast, RC1–2 and RC1–5 were statistically significant (permutation test $p_{FDR} < 0.001$, each corrected for 5 comparisons) at T1 and T2, respectively. For pseudowords–nonwords, RC1 and RC1-3 were statistically significant (permutation test $p_{FDR} < 0.05$, each corrected for 5 comparisons) at T1 and T2, respectively. For base RCA calculations, RC1–5 coefficients were all statistically significant (permutation test $p_{FDR} < 0.001$, each corrected for 5 comparisons) at both T1 and T2.

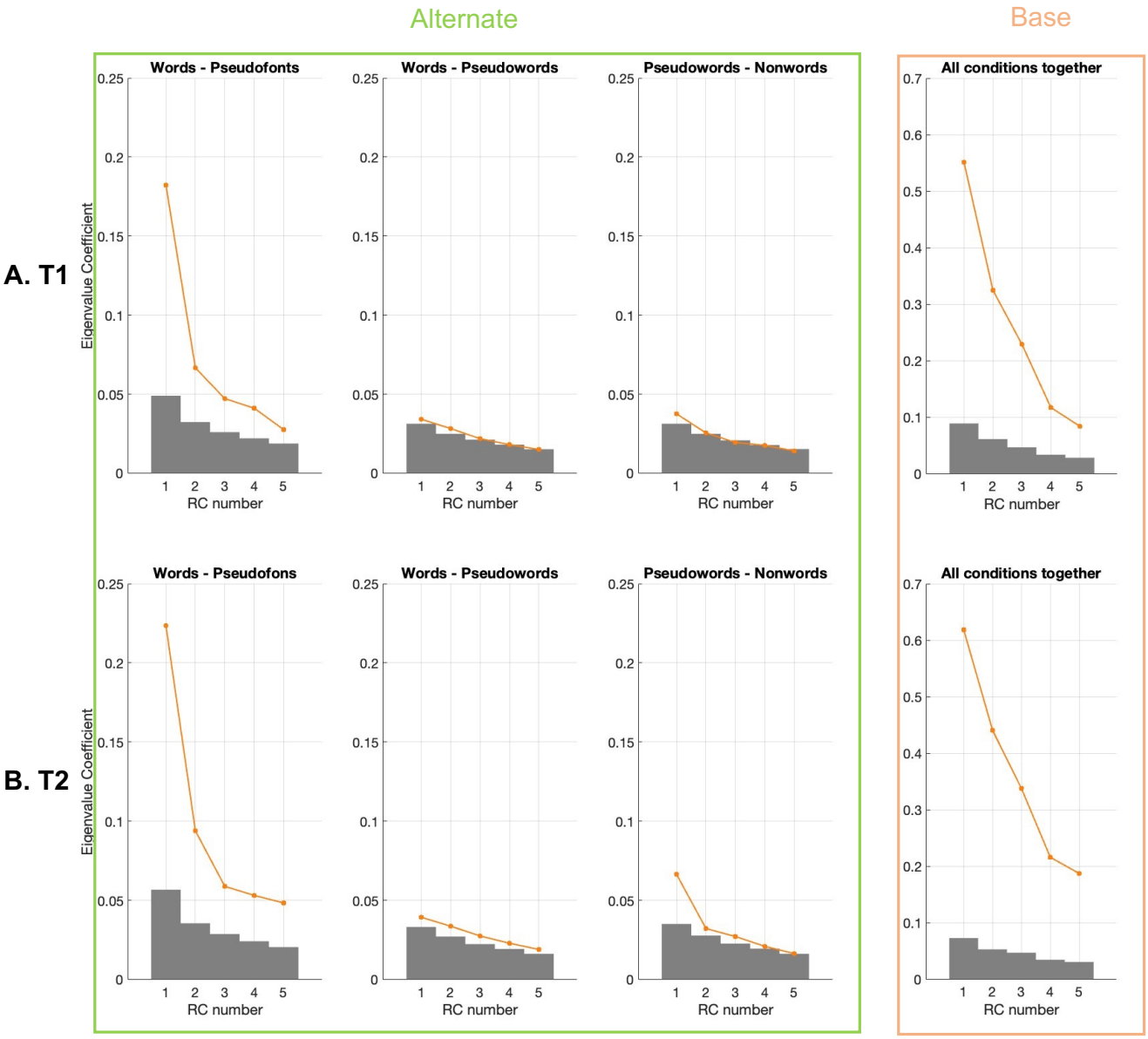

Fig. S2. Statistical analyses of eigenvalue coefficients of RC1–5 at (A) T1 and (B) T2. The line plots denote the distribution of component coefficients for each RCA calculation, and the shaded gray area denotes the 95th percentile of each component's null distribution. Alternate: RC1–5 coefficients for words–pseudofonts were all statistically significant (permutation test $p_{FDR} < 0.001$, each corrected for 5 comparisons) at both T1 and T2; for words–pseudowords, RC1–2 were statistically significant (permutation test $p_{FDR} < 0.05$, corrected for 5 comparisons) at T1, while RC1–5 were all statistically significant (permutation test $p_{FDR} < 0.01$, corrected for 5 comparisons) at T2; for pseudowords–nonwords, RC1 and RC1-3 were statistically significant (permutation test $p_{FDR} < 0.05$, each corrected for 5 comparisons) at T1 and T2, respectively. Base: RC1–5 coefficients for base RCA performed on three conditions together were all statistically significant (permutation test $p_{FDR} < 0.001$, each corrected for 5 comparisons) at both T1 and T2.

Base RCA Results. We performed RCA on EEG responses at the base frequency and its harmonics (2 Hz, 4 Hz, 6 Hz, 8 Hz, and 10 Hz) in order to investigate neural activity underlying low-level visual processing. We focused the base results on the maximally correlated component RC1. Figure S3A displays the topographies of RC1 separately for T1 (left) and T2 (right). RC1s are similar across testing time points, which are distributed over bilateral occipito-temporal region. Figure S3B presents projected data (i.e., projecting data through the spatial filter) in the complex plane and shows overlapping amplitudes (vector lengths) and phases (vector angles) across three conditions at both time points. Amplitudes (bar plots) and phase information (best fit of line) are presented in Figure S3C. Two-way repeated-measures ANOVA with within-subjects factors condition and testing time point revealed that neither the main effects nor the interaction were significant (all $F < 2.61$, all $p > 0.1$) on the projected amplitudes at each harmonic for RC1. Comparisons of phase across three conditions within a testing time point showed no significant difference (T1: all $F < 2.61$, all $p > 0.1$, T2: all $F < 2.61$, all $p > 0.1$) than RC1 (circular ANOVA; $p_{FDR} < 0.01$ for both 2 Hz and 4 Hz, corrected for 2 comparisons). Thus, we consider the responses at the base frequency to be comparable across conditions and testing time points.

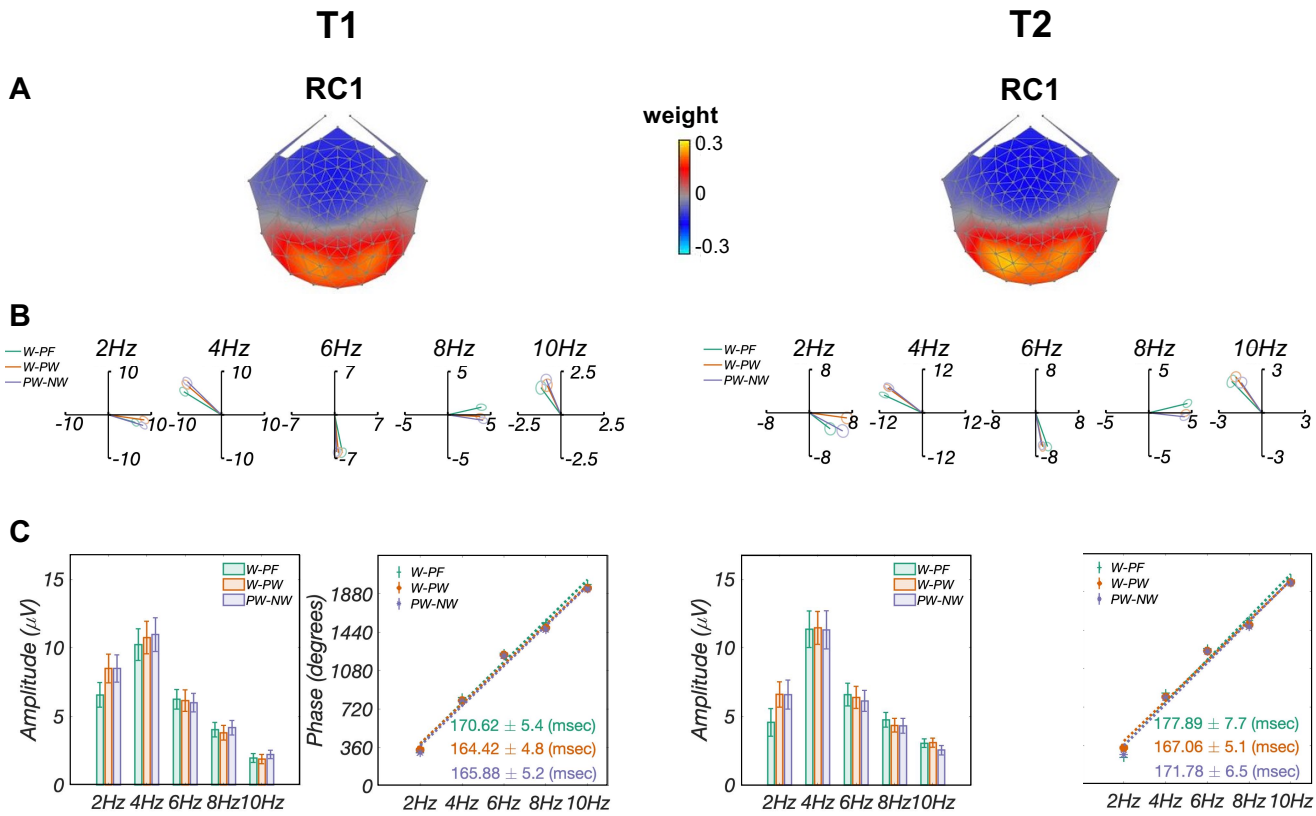

Fig. S3. Base analysis comparisons across three conditions and two testing time points. A: Topographic visualizations of the spatial filters (T1: left; T2: right) for the first component (RC1); B: Response data of three conditions presented in the 2D complex plane, where amplitude information is represented as the length of the vectors, and phase information in the angle of the vector relative to 0 degrees (counterclockwise from 3 o'clock direction), ellipse indicates standard error of the mean (SEM); C: Comparison of projected amplitude and latency estimation across three conditions (T1: left; T2: right). Response amplitudes did not differ significantly across the five harmonics (RC1: $p > 0.1$), latency estimation derived from phase slopes across harmonics were also similar across conditions and testing points (words–pseudofonts, W–PF: 170.62 ms and 177.89 ms for T1 and T2, respectively; words–pseudowords, W–PW: 164.42 ms and 167.06 ms for T1 and T2, respectively; pseudowords–nonwords, PW–NW: 165.88 ms and 171.78 ms for T1 and T2, respectively).

Correlations between amplitudes of three conditions. Linear regressions between alternate amplitudes of three conditions were performed to assess the relationship between different levels of word information processing. As shown in Figure S4, response amplitudes for words–pseudofonts contrast (i.e., coarse neural tuning) were significantly correlated with amplitudes for pseudowords–nonwords contrast (i.e., sensitivity to word form structure), instead of words-pseudowords contrast (i.e., sensitivity to whole word representation), at both T1 ($p_{FDR} < 0.01$, corrected for 3 comparisons) and T2 ($p_{FDR} < 0.001$, corrected for 3 comparisons). No significant relations were found between response amplitudes to words-pseudowords contrast and pseudowords–nonwords contrast.

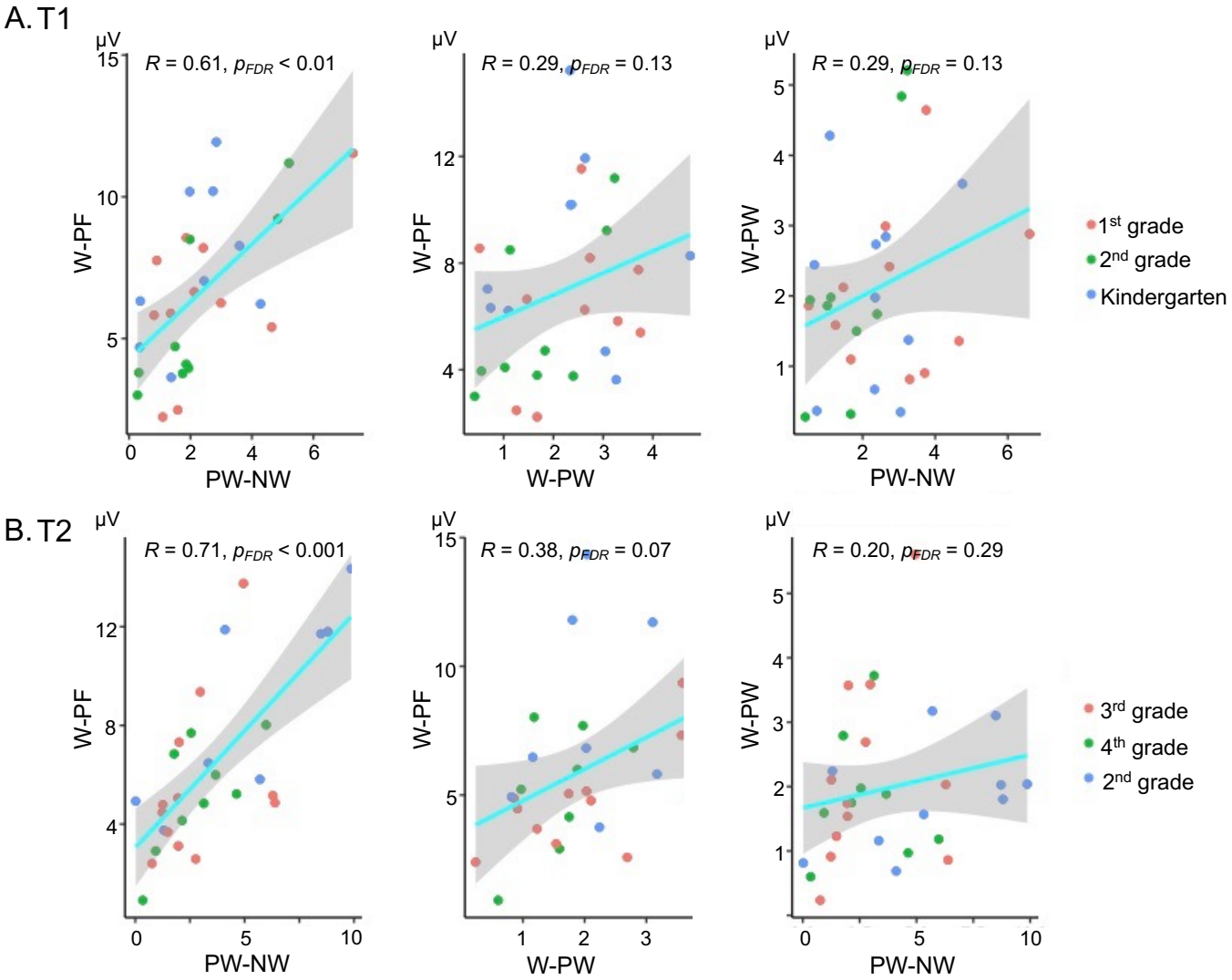

Fig. S4. Correlations between amplitudes of the three conditions at two testing time points. Linear regressions between amplitudes for words–pseudofonts (W–PF) and pseudowords–nonwords (PW–NW), words–pseudofonts (W–PF) and words-pseudowords (W–PW), and words-pseudowords (W–PW) and pseudowords–nonwords (PW–NW) at T1 (A) and T2 (B). Left panel: Statistically significant correlations between response amplitudes for words–pseudofonts (W–PF) contrast (i.e., coarse neural tuning) and pseudowords–nonwords (PW–NW) contrast (i.e., sensitivity to word form structure) at both T1 ($p_{FDR} < 0.01$, corrected for 3 comparisons) and T2 ($p_{FDR} < 0.001$, corrected for 3 comparisons); middle panel: Response amplitudes for words–pseudofonts (W–PF) contrast (i.e., coarse neural tuning) were not significantly correlated with amplitudes for words-pseudowords (W–PW) contrast (i.e., sensitivity to whole word representation). right panel: No significant relations were found between response amplitudes to words-pseudowords (W–PW) contrast and pseudowords–nonwords (PW–NW) contrast.

Lateralization of brain responses to three conditions. A three-way repeated-measures ANOVA with wihin-subjects factors condition, testing time point, and hemisphere on RSS projected amplitudeS to three stimulus contrasts was computed to test the lateralization changes of brain responses. As shown in Figure S5B, we revealed statistically significant main effects of condition ($F(2, 360) = 26.3, p < 0.001$) and hemisphere ($F(1, 360) = 41.6, p < 0.001$), as well as interaction effect between condition and hemisphere ($F(2, 360) = 10.6, p < 0.001$), on the RSS projected amplitude. Post-hoc t tests showed that brain responses to coarse print tuning (words–pseudofonts) were left lateralized at both testing time point (both $p < 0.001$, Bonferoni corrected); Brain responses to visual word form structure processing (pseudowords–nonwords) were more left lateralized at T2 ($p < 0.001$, Bonferroni corrected) than T1 ($p = 0.168$). Trend but not statistically significant left lateralization effects were found for brain responses to whole word representation processing (words–pseudowords).

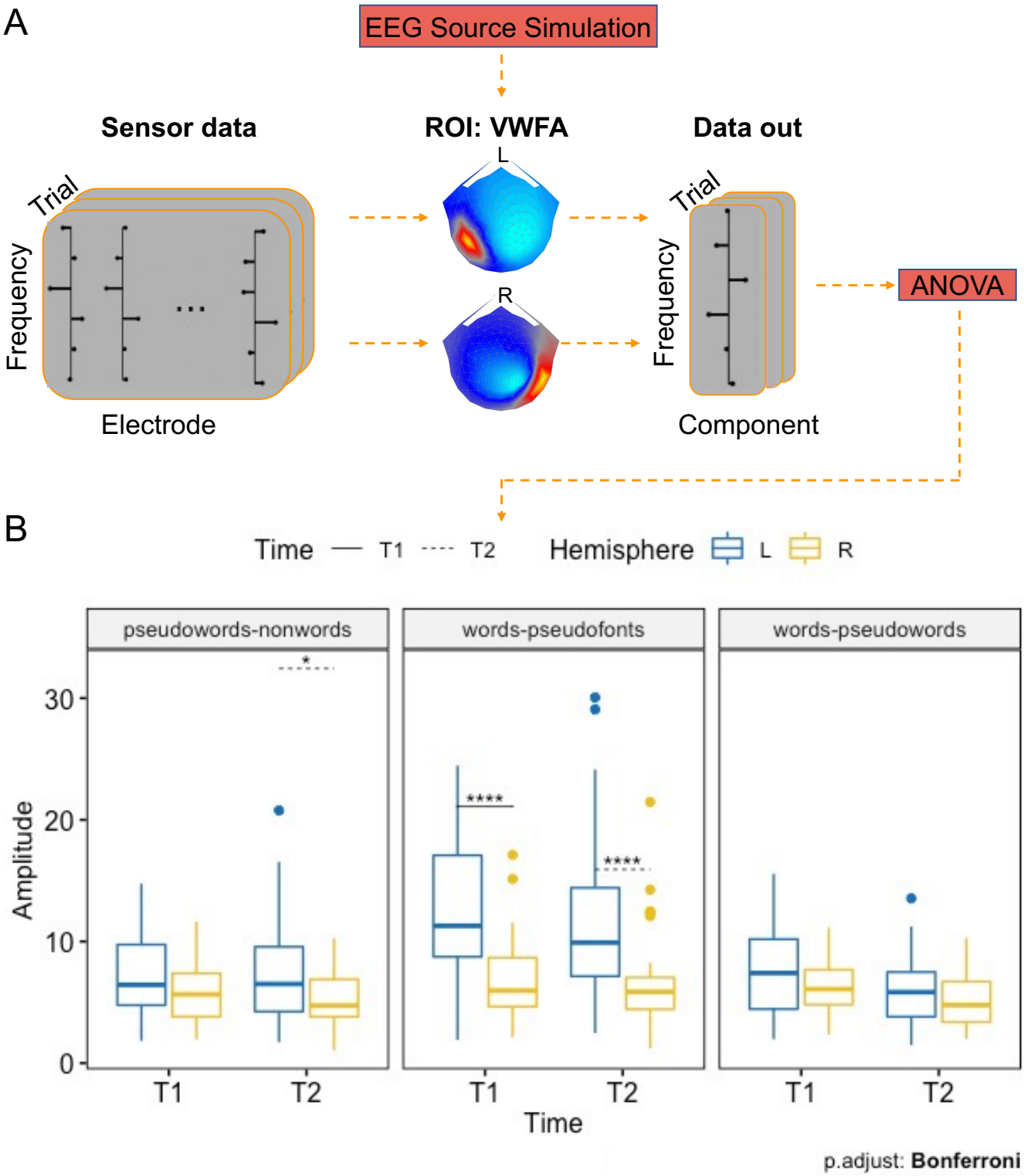

Fig. S5. Procedure and test results for lateralization of brain responses in three conditions. A: Procedure of testing brain responses lateralization. First, regions of interest (ROIs) for left and right visual word form area (VWFA) were simulated using EEG source simulation toolbox (19). Second, we projected electrode-by-frequency-by-trial sensor data matrix at both testing time points onto left and right VWFA ROIs, respectively, in which the resulting output data (i.e., projected data) turned to a matrix of component-by-frequency-by-trial. Finally, projected data were analyzed with a three-way repeated-measures ANOVA with within-subject factors condition, testing time point, and hemisphere; B: Results of lateralization test. Significant main effects of condition ($p < 0.001$) and hemisphere ($p < 0.001$), as well as interaction effect of condition by hemisphere ($p < 0.001$) were revealed. Post-hoc tests showed that brain responses were left lateralized for coarse tuning effect (words-pseudofonts) at both T1 and T2; for visual word form structure processing (pseudowords-nonwords), brain responses were more left lateralized at T2 compared with T1. Trend but not statistically significant left lateralization effects were found for whole word presentation processing. *: $p_{adj} < 0.05$; ***:$p_{adj} < 0.001$

References

1. R Wagner, J Torgesen, C Rashotte, N Pearson, Comprehensive test of phonological processing–second edition (2013).
2. JK Torgesen, RK Wagner, CA Rashotte, TOWRE-2 examiner's manual. *Austin, TX: Pro-Ed* (2012).
3. FA Schrank, KS McGrew, N Mather, BJ Wendling, EM LaForte, *Woodcock-Johnson IV tests of cognitive abilities*. (Riverside), (2014).
4. RC Oldfield, The assessment and analysis of handedness: the edinburgh inventory. *Neuropsychologia* **9**, 97–113 (1971).
5. C Vidal, F Chetail, BACS: The Brussels Artificial Character Sets for studies in cognitive psychology and neuroscience. *Behav. Res. Methods* **49**, 2093–2112 (2017).
6. D Lehmann, W Skrandies, Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalogr. clinical neurophysiology* **48**, 609–621 (1980).
7. Y Tang, AM Norcia, An adaptive filter for steady-state evoked responses. *Electroencephalogr. Clin. Neurophysiol. Potentials Sect.* **96**, 268–277 (1995).
8. J Kilner, Bias in a common eeg and meg statistical analysis and how to avoid it. *Clin. Neurophysiol.* **10**, 2062–2063 (2013).
9. F Wang, et al., Distinct neural sources underlying visual word form processing as revealed by steady state visual evoked potentials (SSVEP). *Sci. reports* **11**, 1–15 (2021).
10. JP Dmochowski, P Sajda, J Dias, LC Parra, Correlated components of ongoing EEG point to emotionally laden attention–a possible marker of engagement? *Front. human neuroscience* **6**, 112 (2012).
11. JP Dmochowski, AS Greaves, AM Norcia, Maximally reliable spatial filtering of steady state visual evoked potentials. *NeuroImage* **109**, 63–72 (2015).
12. LC Parra, CD Spence, AD Gerson, P Sajda, Recipes for the linear analysis of EEG. *NeuroImage* **28**, 326–341 (2005).
13. F Wang, et al., Early detection of sublexical and lexical processing in beginning readers: evidence from steady-state visual evoked potentials (ssveps). *bioRxiv* (2021).
14. Y Benjamini, D Yekutieli, The control of the false discovery rate in multiple testing under dependency. *The Annals Stat.* **29**, 1165–1188 (2001).
15. AM Norcia, A Yakovleva, B Hung, JL Goldberg, Dynamics of contrast decrement and increment responses in human visual cortex. *Transl. Vis. Sci. & Technol.* **9**, 6–6 (2020).
16. P Berens, , et al., Circstat: a matlab toolbox for circular statistics. *J Stat Softw* **31**, 1–21 (2009).
17. RD Cook, Detection of influential observation in linear regression. *Technometrics* **19**, 15–18 (1977).
18. 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).
19. E Barzegaran, S Bosse, PJ Kohler, AM Norcia, Eegsourcesim: A framework for realistic simulation of eeg scalp data using mri-based forward models and biologically plausible signals and noise. *J. neuroscience methods* **328**, 108377 (2019).
20. NA Macmillan, CD Creelman, *Detection theory: A user's guide*. (Psychology press), (2004).