Supplementary Methods

Testing took place either in a room at the child's school or in a university laboratory. To maintain a constant low level of illumination during the psychophysical tests, rooms either did not have windows, or windows were covered with opaque paper.

Reading and cognitive ability testing battery

The Word Identification subtest of the Woodcock Reading Mastery Tests, Revised (WRMT-R, Form G) was used to assess single word reading[1]. The Formal Reading Inventory (FRI) was used to assess silent reading comprehension[2]. The Word Attack test of the WRMT-R, Form G requiring children to pronounce printed nonsense words, assessed phonological decoding skill[1]. The Elision subtest from the Comprehensive Test of Phonological Processing (CTOPP) assessed phonological awareness[3]. Orthographic skill was evaluated with an Exception Word Reading test, which consisted of 70 irregularly spelled items that increased in difficulty and rarity (e.g., *two…beauty…*).

The Receptive One-Word Picture Vocabulary Test (ROWPVT) assessed children's receptive vocabulary[4]. Two subtests of the Clinical Evaluation of Language Fundamentals (CELF-3), Recalling Sentences and Concepts and Directions, assessed other aspects of language skill[5].

General cognitive ability was estimated with four subtests of the Wechsler Intelligence Scales for Children – III (WISC-III)[6]: Vocabulary and Similarities (for Verbal Estimate) and Block Design and Picture Completion (for Performance Estimate).

Participants

Twenty-eight dyslexic (18 male, 10 female) and 27 non-dyslexic children (8 male, 19 female) between 9 to 14 years old were identified. Children who scored at or below the 25th percentile on a either Word Identification or Word Attack qualified for the dyslexia group, while children who scored at or above the 40th percentile on both tests qualified for the non-dyslexia group. Both groups were further required to have Performance IQ Estimates at or above a standard score of 7. Ten of the dyslexic children (7 male, 3 female) also exhibited language impairments (LI), with scores more than a standard deviation below the national mean on two out three oral language measures (ROWPVT, Recalling Sentences, or Concepts and Directions).

There were 20 children tested who had reading scores falling between the 25th and 40th percentiles or who did not meet the Performance IQ criteria.

One dyslexic child scored within the clinical range for Hyperactivity-Impulsivity according to the Disruptive Behavior Rating Scale[7] (a questionnaire filled out by the children's parents, used to screen for ADHD), and four dyslexic children scored within the clinical range for Inattention. None of these five children exhibited LI.

Psychophysics Specifications

The visual psychophysical task was programmed using Matlab 5.2, with Psychtoolbox extensions[8]. The vertical refresh rate was 75 Hz, with a resolution of 640 x 480 pixels. Screen and stimulus luminance were determined by measuring the monitor phosphers, as well as the gray levels. Mean background luminance was 16 cd/m^2. Children sat 210 cm from the computer screen and performed the task with the lights extinguished. Everyone was given the opportunity to adapt to the darkness of the testing room. The stimuli were Gabor patterns of sinusoidal gratings, with a Gaussian contrast envelope (σ = 0.24°), in a square area subtending a ~1.09° x 1.09° visual angle. Stimuli were ramped on and off with a cosine window over 200 ms. Two versions were tested utilizing magnocellular-type (M) stimuli and parvocellular-type (P) stimuli. M stimuli had a peak spatial frequency of 2 cycles/° and flickered in counter-phase at a rate of 15 reversals/s, providing low spatial frequency and high temporal frequency stimulation to maximize magnocellular activation. P stimuli had a peak spatial frequency of 8 cycles/° and did not reverse phase, providing high spatial frequency and low temporal frequency stimulation to maximize parvocellular activation. The M and P stimuli were either presented as is in the low noise condition, or else with an accompanying noise checkerboard in the high noise condition. Noise areas were the same size as the signal areas, made of 2 x 2 pixel patches, each subtending 0.03° x 0.03°. The contrast of each pixel patch was sampled from a Gaussian distribution with a mean of 0 and standard deviation of 0 (in the zero noise condition) and 0.33 in the high noise condition. The maximum standard deviation was 33% as a guarantee that the noise fit the Gaussian distribution. In the M version, noise reversed phase with the stimuli, while in the P version noise was static.

The tasks followed a 2-alternative forced-choice design. A fixation cross appeared at the center of the screen for 250 ms, and remained on for the duration of the trial. Children were shown two simultaneous stimulus regions (intervals) on either side of a fixation cross for 200 ms. The space between the regions was the size of one stimulus square, and thus the entire display subtended a ~3.27° x 1.09° visual angle. One region contained the Gabor signal pattern, and the children had to indicate which region that was. In the low noise condition, the empty interval was a gray that matched the background. In the high noise condition, two noise patterns appeared, with only one containing the complementary signal checkerboard. Responses involved pressing the "?/" key or the "z" key and were typed into the keyboard by the experimenter. Correct responses were indicated with a simple computer "beep", while incorrect responses were greeted with an audio clip of a woman saying, "Uh oh." The luminance contrast of the sine wave grating with the background was determined by a 2-down/1-up staircase, with separate staircases for the low and high noise conditions. A step amounted either to a decrease or increase of contrast by 10% of the present contrast level (e.g., a change of 5% for a contrast level of 50% and 0.5% for a contrast level of 5%). Each task had 160 trials, a randomized mix of 80 of each of the noise conditions. Both versions were preceded by 30 practice trials (15 mixed trials of both low and high noise), in which the staircases were started at high contrast. Children re-did practice as necessary. Contrast threshold was determined by averaging the staircase endpoints, discarding the first four endpoints to account for initial learning. Version order was counter-balanced across participants.

Statistical Analyses

We analyzed the children's contrast thresholds using Welch's method for heteroscedastic means[9]. In the high noise conditions only, dyslexic children's contrast thresholds were significantly higher than non-dyslexic children's, in both M and P tasks: M: $W = 2.03$, df = 41, $P < 0.05$; P: $W = 3.01$, df = 30, $P < 0.01$. Dyslexic children had a threshold of 28.7% for M and 38.3% for P, while non-dyslexic children had 23.9% for M and 29.7% for P.

We conducted a 2 x 2 x 3 ANOVA to evaluate the interaction between noise level (no noise/high noise), task (M/P) and reading group. The reading group variable consisted of three groups, non-dyslexic children, and dyslexic children either with (LI group) or without language impairments. There was a main effect of noise level ($F_{1,32}$ = 257, $P < 0.0001$) and a main effect of task ($F_{1,32} = 23.5$, $P < 0.0001$). Across all children, magnocellular contrast thresholds were lower than parvocellular ones. There was a significant interaction between noise level and reading group ($F_{2,32} = 8.38$, $P < 0.001$). The high noise conditions yielded graded effects for both M and P tasks, with the LI dyslexic children having the highest thresholds, the non-LI dyslexic children intermediate, and the non-dyslexic children the lowest thresholds. In contrast, none of the group differences in the no-noise conditions approached significance. No other interactions involving noise level, task or reading group were significant. Given the absence of an interaction between noise level and task, we collapsed M and P high noise conditions and found a strong effect of reading group ($F_{2,32} = 10.25$, $P < 0.0001$). Dyslexic children with LI had thresholds of 33.8% for M and 42.9% for P, while dyslexic children without LI had thresholds of 25.7% and 35.9%, respectively.

1. Woodcock, R. W. *Woodcock Reading Mastery Tests – Revised* (American Guidance Services, MN, 1987).

2. Weiderholt, J. L. *Formal Reading Inventory* (Pro-Ed, Austin, TX, 1986).

3. Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. *Examiner's Manual. Comprehensive Test of Phonological Processing* (Pro-Ed, Austin, TX, 1999).

4. Brownell, R. (ed.) *Manual. Receptive One-Word Picture Vocabulary Test* (Academic Therapy Publications, Novato, CA, ed. 2, 2000).

5. Semel, E., Wiig, E. H. & Secord, W. A. *Examiner's Manual. CELF®* (The Psychological Corporation, San Antonio, TX, ed. 3, 1995).

6. Wechsler, D. *Wechsler Intelligence Scales for Children – III*™ (The Psychological Corporation, San Antonio, TX, 1991).

7. Barkley, R. A. & Murphy, K. R. *Attention-Deficit Hyperactivity Disorder. A Clinical Workbook* (Guilford Press, New York, ed. 2, 1998).

8. Brainard, D. H. *Spat. Vis.* **10**, 433-436 (1997).

9. Wilcox, R. R. *Applying Contemporary Statistical Techniques* (Academic Press, San Diego, CA, 2003).