The field of view available to the ventral occipito-temporal reading circuitry

Rosemary Le*

Psychology Department, Stanford University, Stanford, CA, USA

Nathan Witthoft*

Psychology Department, Stanford University, Stanford, CA, USA

The Gonda Brain Research Center and Department of English Literature and Linguistics, Gonda Multidisciplinary Brain Research Center, Bar-Ilan University, Ramat-Gan, Israel

Michal Ben-Shachar

Brian Wandell

Psychology Department, Stanford University, Stanford, CA, USA

Skilled reading requires rapidly recognizing letters and word forms; people learn this skill best for words presented in the central visual field. Measurements over the last decade have shown that when children learn to read, responses within ventral occipito-temporal cortex (VOT) become increasingly selective to word forms. We call these regions the VOT reading circuitry (VOTRC). The portion of the visual field that evokes a response in the VOTRC is called the field of view (FOV). We measured the FOV of the VOTRC and found that it is a small subset of the entire field of view available to the human visual system. For the typical subject, the FOV of the VOTRC in each hemisphere is contralaterally and foveally biased. The FOV of the left VOTRC extends ~9° into the right visual field and ~4° into the left visual field along the horizontal meridian. The FOV of the right VOTRC is roughly mirror symmetric to that of the left VOTRC. The size and shape of the FOV covers the region of the visual field that contains relevant information for reading English. It may be that the size and shape of the FOV, which varies between subjects, will prove useful in predicting behavioral aspects of reading.

Introduction

The ability to rapidly recognize letters and words varies extensively across the visual field. Letters and words are recognized most efficiently in the central visual field, even when letters are appropriately scaled (Chung, Mansfield, & Legge, 1998) and adequately spaced (Chung, 2002). With eccentric viewing, performance is not equal at equal eccentricities. Word recognition is faster for text presented in the lower visual field than in the left or right visual field (Petre, Hazel, Fine, & Rubin, 2000), and faster for words presented in the right visual field compared to the left visual field (Babkoff & Ben-Uriah, 1983; Young & Ellis, 1985).

Neuroimaging and intracranial electrode measurements provide evidence that regions in ventral occipito-temporal (VOT) cortex are part of the neural circuitry that subserves reading (Dehaene & Cohen, 2011; Kravitz, Vinson, & Baker, 2008a; Price & Devlin, 2011; Rauschecker, Bowen, Parvizi, & Wandell, 2012; Wandell, Rauschecker, & Yeatman, 2012). These regions are involved in the perception of visually presented word forms and are reproducibly activated in fMRI experiments across individuals and orthographies (Baker et al., 2007; Bolger, Perfetti, & Schneider, 2005; Glezer & Riesenhuber, 2013; Krafnick et al., 2016; Martin, Schurz, Kronbichler, & Richlan, 2015). We analyze the regions within VOT that are more responsive to visually presented words than to other categories of visual stimuli, and we refer to these regions as the VOT reading circuitry (VOTRC). The VOTRC includes but is not limited to the visual word form area (VWFA, Cohen et al., 2000).

The neural basis for differential reading performance across the visual field may be found in properties of the
VOTRC. For example, Yu, Jiang, Legge, and He (2015) demonstrate that foveally and peripherally presented words have differing response properties in the VWFA. Similarly, the neural basis of the right visual field advantage is attributed to lateralization of word processing in cortex (Cohen et al., 2002). In this paper we develop methods to quantify the portion of the visual field that reliably evokes responses in the VOTRC of individual subjects. We refer to this portion of the visual field as the field of view (FOV). We measure the FOV using population receptive field (pRF) methods (Dumoulin & Wandell, 2008; Kay, Winawer, Mezer, & Wandell, 2013) in typical adult readers.

Given that responses in the VOTRC change with reading development (Ben-Shachar, Dougherty, Deutsch, & Wandell, 2011a; Brem et al., 2006; Dehaene et al., 2010), it is likely that the visual information relayed to the VOTRC influences reading performance. The signals in this region may serve as a bottleneck, or may influence the way a subject reads (e.g., the pattern of eye movements). The FOV methods and measurements we describe can serve as a foundation for future analyses of reading in clinical and developmental populations.

### Methods

#### Subjects

Twenty right-handed subjects (eight males, 12 females; median age 24 years, range 20–35 years) participated in the study, which was approved by the Institutional Review Board at Stanford University. All subjects gave informed consent, were right-handed native English speakers, and had normal or corrected-to-normal vision with no reported history of learning disabilities. The majority (18 out of the 20) subjects participated in a ~2 hr scan session completed in a single visit. These subjects were given a break during the scanning session, when the head coil was changed. Two subjects were scanned on separate days due to scheduling conflicts.

#### Displays

Stimuli were presented with two types of displays. Retinotopic-mapping stimuli were presented with an Eiki LC-WUL100L projector. The Eiki projector has a native resolution of 1920 × 1200 pixels and 10-bit color resolution. Localizer stimuli were presented with either the Eiki projector or a 47" LCD display. The 47" LCD display has a native resolution of 1920 × 1080 and eight-bit color resolution. Both displays were viewed through a mirror placed either at ~5 cm (projector) or ~15 cm (LCD display) from the eyes. Stimuli projected using the LCD display had a vertical extent of 12°, while stimuli presented with the Eiki projector had a vertical extent of 32°.

#### MR acquisitions

Data were acquired at the Stanford Center for Cognitive and Neurobiological Imaging using a 3T General Electric MR 750 scanner with either a Nova 16- or 32-channel head coil. The Nova 16-channel is similar to the Nova 32-channel coil, but with the front half of the coil removed to allow for an unobstructed field of view. Head motion was minimized by padding around the head.

#### Anatomical

Anatomical data were acquired with the 32-channel coil using a 3D Fast SPGR scan (166 sagittal slices, resolution 0.9375 × 0.9375 × 1 mm). For each subject, one to three anatomical volumes were acquired and then averaged. Data were resampled to 1mm isotropic voxels and aligned to the anterior commissure-posterior commissure (AC-PC) plane using an affine transformation.

#### Functional

Functional data for pRF mapping and the large-field localizer (see below) were acquired with the 16-channel coil. Thirty-six slices covering occipitotemporal cortex were defined: 2.5 mm isotropic voxels, TR 2000 ms, TE 29 ms, flip angle 77°, field-of-view 200 × 200 mm. Functional data for the small-field localizer (see below) were acquired with the 32-channel coil. Forty-eight slices covering occipitotemporal cortex were defined: 2.4 mm isotropic voxels, TR 1000 ms, TE 30 ms, flip angle 62°, field-of-view 192 × 192 mm. An in-plane anatomical image that matched the functional slice prescription was acquired before each set of functional runs. These images were used to align the functional data to the anatomical volume data.

#### MR data analysis

#### Preprocessing

MR data analyses relied on the open-source code in vistalab (https://github.com/vistalab/vistasoft). The basic preprocessing steps included estimation and removal of motion artifacts, and registration of the functional data to the high resolution anatomical
images. Motion artifacts within and across runs were corrected using an affine transformation of each volume in a session to the first volume of the first run. In all subjects, head movement was less than 1 voxel (in most cases less than 0.4 voxel). The first six time frames of each functional run were discarded. Baseline drifts were removed from the time series by high-pass temporal filtering.

The inplane anatomical image was aligned to the average whole brain T1-weighted anatomical image by calculating a rigid body transformation that maximized the mutual information between the inplane anatomy and the resliced volume anatomy. These alignment parameters were then used to align the functional data to the anatomical data.

**Defining the VOT reading circuitry (VOTRC)**

We defined the VOTRC using a combination of functional and anatomical constraints. One of two localizers was used to choose voxels based on their functional selectivity: a small-field or a large-field localizer. In the small-field localizer, stimuli consisted of a single item (pseudoword, body, face, place, object, or number) overlaid on a phase-scrambled background (Figure 1A). Each phase scrambled background was generated from an image selected randomly from the entire set of images. The background image subtended a 12° × 12° portion of the visual field. For the pseudoword category, each word spanned approximately 3° × 8°. We call this the small-field localizer because the category of interest was only present in the more central part of the visual field. The small-field localizer was used to identify the VOTRC in Subjects 1–12.

In the large-field localizer, a 30° × 30° portion of the visual field was tiled with either words, faces, or a phase-scrambled object (Figure 1B). The purpose of this localizer was to present words to both central and peripheral visual field locations in order to reduce stimulus-based bias of the voxel selection. English words were presented rather than pseudowords (both words and pseudowords are well known to evoke strong responses in the VOTRC; Dehaene, Le Clec'H, Poline, Le Bihan, & Cohen, 2002). The words were one to six letters long, had a minimum frequency of 200 per million, were displayed in Helvetica font, and the
lowercase “x” character had a vertical size of 1.3°. For each of the face images, 16 faces were chosen at random from a database of 144 faces and arranged in a 4 × 4 configuration. Each face had a vertical size of approximately 6°. In this localizer, participants were only presented with word, face, and phase-scrambled object categories. The large-field localizer was used to identify the VOTRC in Subjects 1, 13–20.

In both localizers, stimuli were presented in blocks of eight images from the same category at a rate of 2 Hz. Each category was presented six times per run, and three runs were collected per subject. The categorical stimuli for both localizers were taken from the fLoc functional localizer package (Stiglani, Weiner, & Grill-Spector, 2015; http://vpnl.stanford.edu/fLoc/). Participants fixated a small red dot at the center of the screen. Zero, one, or two phase-scrambled images appeared in a block with probabilities of 0.25, 0.5, and 0.25, respectively. Participants pressed a button when a phase-scrambled image appeared.

The boundaries of the VOT are the inferior temporal sulcus (lateral), the collateral sulcus (medial), hV4 (posterior), and an imaginary horizontal line drawn from the collateral sulcus to the inferior temporal sulcus, starting at the point where the parieto-occipital sulcus would meet the collateral sulcus (anterior). The VOT is defined individually in all subjects. In most subjects, hV4 is found in the posterior transverse collateral sulcus (Witthoft et al., 2014). This region is illustrated in Figure 1C. The word-selective voxels within this region in each hemisphere are defined as the left or right VOTRC. Selective voxels are more responsive to words than to other visual categories (t test, p < 0.001, uncorrected), but not necessarily contiguous. The VOTRC falls in or near the occipital temporal sulcus (Figure 3A) in both hemispheres. The VOTRC as defined for all subjects is shown in Supplementary Figure 1.

**PRF mapping**

Stimuli for the pRF mapping experiment were presented using the Eiki LC-WUL100L projector and using the Psychophysics Toolbox (Brainard, 1997). A small dot (0.15° visual angle in diameter) at the center of the screen served as the fixation point. Participants were instructed to fixate the dot and to press a button when the dot changed color (randomly between red, green, and black every 1–5 s). Participants maintained fixation as a moving bar swept through the visual field in four orientations (0°, 45°, 90°, and 135° from vertical) and two motion directions. Eye-tracking was not used, but analysis of visual field coverage of the early visual field maps (V1–V3) shows the expected hemifields and quarterfields (Figure 2B, Figure 5CD) suggesting that stable fixation was maintained by participants (Amano, Wandell, & Dumoulin, 2009; Winawer & Witthoft, 2015).

In one set of runs, the bars contained checkerboard stimuli as contrast patterns. Code from the Vistadisp software package (https://github.com/vistalab/vistadisp/) was used to display these stimuli. There were 12 bar positions for each sweep across the visual field. Each position lasted 2 s for a total of 24 s per sweep. There were eight sweeps of the bar, for a total of 192 s per run. Four, 12 s blank periods replaced some portions of the bar sweeps, as per Dumoulin and Wandell (2008), to estimate larger pRF sizes. The checkerboards drifted in the long direction of the bar and had a contrast reversal rate of 2 Hz. Each side of the checkerboard squares was approximately 1.3°. Three runs, each 192 s (total of 576 s), of checkerboard retinotopy were collected for each subject. The stimuli were also used in previous work (Amano, Wandell, & Dumoulin, 2009; Dumoulin & Wandell, 2008).

In a second set of runs, the bars contained word stimuli as the contrast pattern. Code from the analyzePRF software package (http://kendrickkay.net/analyzePRF/) was used to display these stimuli. Each sweep of the bar moved continuously across the visual field for 31 s for each of the eight orientations. There was a 16 s blank period at the beginning, a 16 s blank period in the middle, and a 20 s blank period at the end of each run. The words were rendered in black Helvetica font on a white background. The vertical extent of the lower-case letter ‘x’ was 1.3°. One-to-six letter words with frequency greater than 200 per million were used as the lexicon; words were randomly chosen from the lexicon to create a page of text. The text within the aperture was refreshed at 4 Hz. Each run lasted 300 s, and two runs were collected (total of 600 s). The word and checkerboard runs were interleaved. For half of the subjects, word retinotopy was shown first; for the other half, checker retinotopy was shown first. The order of the remaining runs was randomized.

We model the population receptive field in each voxel using the compressive spatial summation (CSS) model (Kay et al., 2013). The CSS model estimates the pRF of a voxel by selecting parameters for a 2D Gaussian that optimally predicts the response to a translating stimulus (moving bars). The parameters of the pRF include location (x, y) and size (σ) in degrees of visual angle. The value of the 2D Gaussian at its peak is normalized to 1. The CSS model extends earlier linear models (Amano et al., 2009; Dumoulin & Wandell, 2008) by including a compressive nonlinear response exponent to account for subadditive responses which appear to increase across the visual hierarchy.
Estimating the FOV for a region of interest (ROI)

The FOV of a region of interest (ROI) defines the portion of the visual field that reliably evokes a response in any of the voxels in the ROI. The FOV of an ROI is obtained by aggregating over the voxels’ pRFs (the portion of the visual field where stimuli reliably drive responses). For each location in the visual field, the FOV value is the maximum pRF value over the voxels in the ROI.

We use a bootstrapping procedure to reduce the effect of pRFs that are anomalous in size and/or...
position (Amano et al., 2009; Winawer, Horiguchi, Sayres, Amano, & Wandell 2010; Winawer & Witthoft, 2015). Suppose an ROI has N voxels that exceed the 20% variance explained threshold for a pRF model. For each bootstrap step, N voxels are sampled with replacement and the FOV is calculated. This procedure is repeated 50 times. The average of the 50 FOV samples is the FOV of the ROI. The FOV is a map of continuous values; so to characterize its size and shape, we describe the half-max contour, a contour encompassing the part of the visual field where the FOV values exceed 0.5. Example FOV and half-max contour are shown in Figure 2B (for V1, V2v, and V3v) and Figure 3B (for left VOTRC). The group summary FOV is obtained from averaging over the individuals’ FOVs (Figure 3C). The individual and group average FOVs are distinguished with different color maps because different operations are applied: The individual FOV represents the maximum pRF value at each point in the visual field, while the group average FOV represents the mean pRF value across subjects. Group average FOVs are not normalized; thus, a value close of 1 (or a value very close to 1) indicates that every subject had a pRF center at (or very near) that location. Unless otherwise stated, only voxels with greater than 20% variance explained by the pRF model are used in measuring the FOV. For word stimuli, all subjects had at least one voxel pass this threshold in left and right VOTRC. For checker stimuli, one subject did not have any voxels pass threshold in the left VOTRC, and three subjects did not have any voxels pass threshold in the right VOTRC. In these few cases, the FOV was assigned to have all zeros. The basic results are unchanged if we exclude these subjects from the FOV analysis.

The FOV of the VOTRC

We begin by considering FOV measurements made using words as the contrast pattern. The left VOTRC and the FOV for a single subject are shown in Figure 3A and B. The FOV is biased for the central portion of the contralateral visual field. In the typical subject, more than 98% of pRF centers (median across subjects; minimum 54%, maximum 100%) are located within 5° of fixation. More than 97% of pRF centers (median across subjects; minimum 37%, maximum 100%) are located in the right visual field. The half-maximum contour extends further along the horizontal than the vertical meridian (Figure 3B). The FOV of the group (n = 20) is shown in Figure 3C. The individual and group average FOVs are distinguished with different color maps because different operations are applied: The individual FOV represents the maximum pRF value at each point in the visual field, while the group average FOV represents the mean pRF value across subjects. Group average FOVs are not normalized; thus, a value
of 1 (or close to 1) means that every subject had a pRF center at or very near that location. The group average FOV demonstrates the same foveal and contralateral bias as the individual data. See Supplementary Table 1 for a quantification of the foveal and contralateral bias in the left VOTRC and other visual areas.

The FOV is reliable within subjects and varies between subjects

To assess within-subject reliability, we visualize the half-maximum contours from pRF models fit to individual runs of data (Figure 4). For the most part, the independent half-max contours cover similar parts of the visual field (e.g., Subjects 04–07, Subjects 10–11, and Subject 16). In some cases the contours from the two runs differ substantially (e.g., Subject 14, Subject 20). In the subjects where the half-maximum contour does differ substantially, the half-maximum contour of one of the runs matches the half-maximum contour as calculated from the average of the two runs, and this run has higher variance explained compared to the other run. We suggest that accurate FOV estimates require multiple (at least two) independent runs.
The FOV of the left VOTRC varies between individuals. There are subjects in whom both runs reveal small FOVs (Subject 04) and others in which both runs reveal large FOVs (Subject 16). In other subjects the FOV seems reliably oriented in different directions (Subject 05 and Subject 06). Despite these differences, the half-maximum contour covers the contralateral and foveal portion of the visual field for all subjects.

The FOV of the left VOTRC, but not V1, is stimulus dependent

We measured pRFs using two different types of stimuli within the moving bar aperture: checkerboards and words. When measured with words, the average FOV of the left VOTRC extends 9° along the horizontal meridian in the contralateral visual field and 4° in the ipsilateral visual field (n = 20, Figure 5A). When measured with checkerboard stimuli, the half-maximum contour of these same voxels extends further into the periphery but remains biased for the horizontal meridian (Figure 5B). The FOV does not exhibit the same stimulus dependence in early visual cortex. For example, the FOV of left V1 extends to the periphery and covers the contralateral hemifield when measured with words and checkerboard stimuli (Figure 5C and D), in agreement with Winawer and Witthoft (2015) and Amano et al. (2009).

**Stimulus dependency of the FOV is not explained by a difference in voxel selection or model accuracy**

We consider two methodological explanations for the stimulus dependence of the FOV in left VOTRC (Figure 5A vs. 5B). One explanation may be due to the fact that nonidentical voxels in left VOTRC pass the 20% variance explained threshold for each stimulus type. This means that different voxels are used in the calculation of the FOV in Figures 5A and 5B. We can exclude this possibility because choosing identical voxels for the calculation of the FOV results in the same stimulus dependence (see Figure 6C for a comparison of group average FOVs within identical voxels).

The FOV of VOTRC is similar for two different localizers

In the small-field functional localizer, word images appear within the central 8° of the visual field (collected on Subjects 1–12). Restricting the stimuli to a small
portion of the visual field may bias the voxel selection process towards voxels with a foveal preference. To address this concern, we use a large-field localizer in nine additional subjects (Subjects 1, 13–20). The large-field localizer contains words that tile the central 32° and should identify word-responsive voxels in VOT that have positional preferences anywhere within this part of the visual field, eliminating any central field bias which might result strictly from the stimulus.

The VOTRC defined with the large-field localizer is similar to the definition with the small-field localizer (see Supplementary Figure 1). One subject took part in both localizers and the difference is roughly what we find for repeated measurements (Figure 7A). The mean surface area of word-responsive cortex in VOT is 31 mm² for the small-field and 39 mm² for the large-field functional localizer, both with a standard error near 8 mm² (surface areas for each subject are listed in Supplementary Figure 2B). The group average FOV estimated with the two localizers is also similar (Dice coefficient ≈ 0.91). In both cases, the FOV of the left VOTRC is biased for the horizontal meridian, extending ~9° in the contralateral visual field and ~4° in the ipsilateral visual field (Figure 7B).

**The FOV of right VOTRC is contralaterally and foveally biased**

The localizer experiments produced word-selective voxels in both hemispheres, allowing for the definition of a right VOTRC in every subject (Ben-Shachar, Dougherty, Deutsch, & Wandell, 2007; Cohen et al., 2002; Glezer, Jiang, & Riesenhuber, 2009; Rauschecker et al., 2012; Vigneau, Jobard, Mazoyer, & Tzourio-Mazoyer, 2005). Like the left VOTRC, the FOV of the right VOTRC exhibits a foveal and contralateral bias, extending 8° into the contralateral visual field and 3°...
into the ipsilateral visual field (Figure 8A). See Supplementary Figure 4 for the FOVs of individual subjects. The Dice coefficient between the FOVs of the left and right VOTRC (when flipped) is 0.86; the two FOVs are mirror symmetric, with the right VOTRC FOV being slightly smaller (compare Figure 8B with Figure 5A, or see Supplementary Figure 2A).

The distribution of pRF parameters in the left and right VOTRC is shown in Supplementary Figure 2C–F. The principle difference is that there are more word-selective voxels in the left hemisphere (Supplementary Figure 2B). Of the word-selective voxels, 49% in left VOTRC and 48% in right VOTRC have variance explained that exceeds 20% by the pRF model (word stimuli). To compute the FOV of the bilateral VOTRC, voxels are pooled across hemispheres and treated as a single ROI. The FOV of the bilateral VOTRC (combined left and right VOTRC) covers an elliptical portion of the visual field that is centered on the fovea and extends mainly along the horizontal meridian (Figure 8D).


The VOTRC’s FOV is restricted to a central region of the VOT’s FOV

Whereas there is variation in the size and shape of the FOV between subjects, the FOV of the bilateral VOTRC of the average subject has a distinctive shape. The average FOV extends about 9° along the horizontal meridian and about 6° along the vertical meridian. The FOVs of the left and right VOTRC are roughly mirror symmetric, although the FOV of the right VOTRC is slightly smaller. The FOV of the bilateral VOTRC is a fraction of the FOV of the full VOT, which extends farther into the periphery in both the horizontal and vertical directions (Figure 9).

PRF model accuracy

The FOV is fit by a model. We would like to evaluate model accuracy—defined as the extent to which the model extracts reliable information from the data. Model accuracy is derived from two calculations that use the root mean squared error (RMSE)

\[
\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (x_i - y_i)^2}{n}}
\]

where \( n = 144 \) (number of time points) and \( x_i \) and \( y_i \) correspond to the values of the time series being compared at a specific time point \( i \). The first RMSE calculation is between independent time series data collected in runs one and two. We call this \( \text{Data}_{\text{rmse}} \), and it can be interpreted as a measure of test-retest reliability. The second RMSE calculation is between the model prediction obtained from run one and the actual time series measured in run two. We call this \( \text{Model}_{\text{rmse}} \). Since run two is independent of run one, \( \text{Model}_{\text{rmse}} \) is a cross-validated error measure of the model prediction. Finally, Relative RMSE is the ratio of the two RMSE calculations: \( \text{Model}_{\text{rmse}} / \text{Data}_{\text{rmse}} \) and we use this ratio to summarize model accuracy (see Figure 10). All time series are responses obtained from pRF mapping with word stimuli.

More than 99% of the bilateral VOTRC voxels have a \( \text{Relative}_{\text{rmse}} < 1 \). This means the model predicts run two with higher accuracy (smaller error) than the test-retest reliability (\( \text{Model}_{\text{rmse}} \) is smaller than \( \text{Data}_{\text{rmse}} \)). A model is better than test-retest reliability when it captures an essential part of the response and discards measurement noise. If a model is a perfect description of the mean signal and the measurement noise at each time point is independently distributed with a Gaussian distribution, the Relative\text{rmse} values should be at \( 1 / \sqrt{2} = 0.707 \) (Rokem et al., 2015). The median value of the Relative\text{rmse} across all VOTRC voxels is 0.73, close to the level of a model that extracts all of the reliable information available in the data.

Figure 7. FOV comparison using the small- and large-field localizers. (A) The left VOTRC (black outline) in Subject 01 as defined by the two localizers: small-field (left) and large-field (right). Colored voxels have variance explained greater than 20% by the pRF model (word stimuli). (B) The left panel shows the average FOV for Subjects 1–12 whose left VOTRC was defined using the small-field localizer. The right panel shows the average FOV for Subjects 1, 13–20 whose left VOTRC was defined using the large-field localizer. The gray curves are 50 half-maximum contours of the bootstrapped group averages.
Position sensitivity in the VOT

The VOTRC is embedded within a larger cortical territory (VOT) that contains several other functionally specialized regions (e.g., Epstein & Kanwisher, 1998; Grill-Spector & Weiner, 2014). We consider position sensitivity and FOV estimates of the VOT generally and then discuss specific features of the reading circuitry.

There is consensus that the functional responses of many VOT voxels are sensitive to stimulus visual field position (e.g., Arcaro, McMains, Singer, & Kastner, 2009; Kravitz, Vinson, & Baker, 2008b; Uyar, Shomstein, Greenberg, & Behrmann, 2016). For example, more lateral VOT regions including the fusiform gyrus are more responsive to stimuli within 1.5° of the fovea while medial regions, including the collateral sulcus and lingual gyrus, are relatively more responsive to stimuli presented at greater than 7.5° eccentricity (see figure 4 in Hasson, Levy, Behrmann, Hendler, & Malach, 2002). The VOT in general is more responsive to stimuli presented in the contralateral than ipsilateral visual field (Kravitz et al., 2008a; Rauschecker et al., 2012).

Several groups have used pRF mapping and FOV measurements to characterize extrastriate visual cortex (Amano et al., 2009; Larsson & Heeger, 2006; Wandell & Winawer, 2011). A number of recent papers have specifically measured pRFs and FOV in VOT regions located near the reading circuitry (Kay, Weiner, & Grill-Spector, 2015; Silson, Chan, Reynolds, Kravitz, & Baker, 2015). Silson et al. (2015) report that pRF centers in the occipital face area (OFA) and fusiform face area (FFA) are located within the central 5° and have a contralateral bias. They report that the FOVs in these two regions represent the lower (OFA) and upper (FFA) quarterfields. Kay et al. (2015) and Witthoft et
al. (2016) agree that face-selective regions have a contralateral bias and that pRF centers are consistently near the fovea. They do not report the lower-upper field bias for the FOV in apparently corresponding face-selective regions (IOG and FFA respectively), even though the IOG may have more pRF centers in the lower visual field.

The pRF model implementations and the cortical regions of interest differ between these three studies. Kay et al. (2013) use a model that includes a compressive nonlinearity. Silson et al. (2015) use a linear model implemented within AFNI, and Witthoft et al. (2016) use a linear model implemented in Vistasoft. In addition to the model differences, the anatomical definitions differ. Kay et al. (2013) and Witthoft et al. (2016) subdivide the FFA into two distinct regions, mid fusiform, mFus, and posterior fusiform, pFus (Weiner & Grill-Spector, 2012). Silson et al. (2015) refer to the OFA which may not be precisely the same as the IOG. Hence, the differences in findings, primarily the upper-lower visual field bias, may be due to the variation of methods used between groups. The contralateral and foveal biases that remain despite the methodological differences suggest that these are robust properties.

Figure 9. Comparison of FOVs for the VOT and the VOTRC. The dashed lines indicate half-maximum contours of the group average FOV. The outer contour and the color map show the FOV of the VOT, both hemispheres (Figure 1C). The inner contour shows the FOV of the bilateral VOTRC (Figure 8D).

Figure 10. Comparing model accuracy and test-retest reliability. (A) The logical organization of the measures. Datarmse is the RMSE between run one and run two. Modelrmse is the RMSE between the model (the prediction of run one) and run two. Relativermse is Modelrmse / Datarmse, where a value < 1 means that the model predicts an independent data set better than test-retest reliability. (B) Top: Time series from run one (orange) and the predicted time series (black) from the pRF fit to run one. Middle: an example time series from run two (teal) and the predicted time series from the pRF fit to run one. Bottom: the time series from run one and run two. Run two (teal) is independent of run one and of the model prediction. (C) Distribution of Relativermse for all voxels in the combined VOTRC, for all subjects. The red line shows the median (0.73).
Lateralization of the VOTRC

Most work on word-responsive regions in visual cortex focuses on the left hemisphere (e.g., Bouhali et al., 2014) because the responses in the left are more selective and larger than those in the right (Cohen et al., 2002; others). Additionally, there is a substantial neurological literature that associates alexia with lesions to the left hemisphere and emphasizes differences in representational capacity between the two hemispheres (Damasio & Damasio, 1983; Dehaene & Cohen, 2011; Déjerine, 1892). Nonetheless, it is important to recognize that word-selective responses are reliably observed in both hemispheres (Ben-Shachar, Dougherty, Deutsch et al., 2007; Cohen et al., 2002; Glezer et al., 2009; Rauschecker et al., 2012; Vigneau et al., 2005). Dehaene, Cohen, Sigman, and Vinckier (2005) propose that word-form information from each hemifield is represented separately until hV4 (previously known as V4/V8) after which responses are sent to left OTS (VWFA). Rauschecker et al. (2012) propose that the visual information in each visual field may be represented separately up until the VOTRC before being communicated to the language areas. The asymmetry of the FOV relative to fixation in the left and right VOTRC (Figure 5A and 8C respectively) supports this proposition. Evidence in callosal patients suggests that the right hemisphere alone is sufficient for some reading functionality (Baynes, Tramo, & Gazzaniga, 1992). The left lateralization of responses emerges during development, and some authors propose that learning to read reduces word-selective responses in the right hemisphere (Shaywitz et al., 2007; Simos, Breier, Fletcher, Bergman, & Papanicolaou, 2000). The FOV of the left VOTRC is approximately mirror symmetric to the right VOTRC (Figure 8C and D). The principal difference is that there are more word-selective voxels in the left hemisphere (Supplementary Figure 2B).

Some earlier models of object recognition (Riesenhuber & Poggio, 1999; Rolls, 2000) and earlier experiments involving the visual word form area (VWFA; Cohen et al., 2002; Dehaene & Cohen, 2011) report that responses to words are invariant to position in the visual field. However, the contralateral bias of the FOV is a rich source of information for decoding word position. Examination of the data in these earlier experiments shows that responses in the VWFA are indeed elevated for words presented in the right visual field (e.g., figure 5 in Cohen et al., 2002; also see figure 5 in Ben-Shachar, Dougherty, Deutsch et al., 2007; and figure 5 in Rauschecker et al., 2012). Hence, the findings here are consistent with earlier reports of spatial sensitivity, but extend them by reporting a quantitative measurement at every location in the visual field.

Stimulus dependency within the VOT

Using checkerboards rather than words within the aperture decreases the variance explained by the pRF model and increases the FOV (Figure 5). This effect is not well explained by methodological considerations, as we showed in detail (Figure 6, Supplementary Figure 5). We discuss three possible explanations of the stimulus dependence of the FOV.

One possibility is that word stimuli and checkerboard stimuli are carried to the VOT through different pathways with slightly different terminations in early visual cortex. For example, Rauschecker et al. (2011) proposed that pattern contrast and motion contrast are communicated to the VOTRC via different routes. An alternative possibility is that all the information is transmitted, but different neuronal populations within the VOTRC preferentially respond to checkerboard and word stimuli (Harvey, Fracasso, Petridou, & Dumoulin, 2015). In this case, stimulus dependence would be explained if these populations have different FOVs. A third possibility is that the stimulus dependence is due to a task and/or attentional difference. The fixation task that the subject performs does not require all attentional resources. It is possible that for word stimuli, subjects read the letters when the bar containing word stimuli passes over fixation. The more foveal FOV for word stimuli may reflect this task difference, as well as the fact that letters 1.4° in size cannot be read in the periphery. Future experiments might systematically vary the task demand to explore attentional effects, or control readability by varying font size and spacing at different eccentricities (Chung et al., 1998; Vinckier, Qiao, Pallier, Dehaene, & Cohen, 2011). Future experiments may also analyze stimulus-dependency in other category-selective regions of the VOT. Preliminary data for face-selective cortex is included in Supplementary Figure 6.

Currently our model represents the input stimulus as a binary mask that indicates where there is some stimulus contrast (i.e., only indicating the position of the moving bar and not the image within the bar). A model that takes into account the image within the moving bar may be able to account for some of the stimulus dependency. Further, it may be possible to create circuit models that account for the contributions of multiple cortical processing regions (Kay & Yeatman, 2017); such models may also account for the stimulus dependence.

Relationship to psychophysical measures of reading

Performance is not equal at equal eccentricities. When presented at equal eccentricities, word recogni-
tion is faster for text presented in the lower visual field than in the left or right visual field (Petre, Hazel, Fine, & Rubin, 2000). There is a moderate right visual field behavioral advantage for recognizing familiar words (Young & Ellis, 1985; Babkoff & Ben-Uriah, 1983). The right visual field advantage effect size is on the order of $d' = 0.5$ (figure 5, Cohen et al., 2002; figure 1, Barca et al., 2011). The neural basis of the phenomenon has been attributed to lateralization of responses in the visual word form area (e.g., Cohen et al., 2002).

The data suggest that the relatively large differences we observe between the FOVs in different subjects are reliable (Figure 4, compare Subject 13 and Subject 16). The FOV measurements describe a characteristic of the reading circuitry, and the size and shape of the FOV may be related to certain psychophysical measures of reading. For example, the amount of contralateral bias in the left VOTRC FOV may correspond to the degree to which an individual exhibits the right visual field advantage. We suspect there may be a compliance range for the field of view in the VOTRC, such that an inadequate (too small) or an irrelevant (too large) FOV disrupts processes that are necessary for learning to rapidly recognize words. If this is true, then the FOV will be predictive of some reading-related measures. Two candidate psychophysical measures that may be predicted by the shape and size of the FOV are the visual span (Legge, Ahn, Klitz, & Luebker, 1997) and the perceptual span (McConkie & Rayner, 1975).

The visual span is the number of letters of a given size, arranged horizontally, that can be reliably recognized without moving the eyes. In typical readers of English under typical reading conditions, the expected range of visual span values is about seven to 11 letters (Legge et al., 1997). Individual differences in reading speed depend on multiple factors, and the visual span positively correlates with reading speed and accounts for a 34%–52% of this variation (Kwon, Legge, & Dubbels, 2007; Legge et al., 2007). The visual span can also be thought of as the number of letters that are not crowded (Pelli et al., 2007). At the age where children learn to read, the crowding effect decreases with age (Bondarko & Semenov, 2004) and is likely reflected in the size of the visual span increasing through the school years (Kwon et al., 2007). The size of the visual span is also dependent on stimulus characteristics such as character size, contrast, and spacing (Legge et al., 2007; Yu et al., 2007). We have preliminary data showing that the FOV is smaller when measured with letters of a smaller font size. A more systematic manipulation is needed to model the dependency of FOV measurements on text characteristics.

The perceptual span assesses reading performance while allowing for the influence of contextual information and linguistic factors. The perceptual span is defined as the region of the visual field available to the reader at a single fixation which allows the individual to maintain their typical reading speed (McConkie & Rayner, 1975). In typical English readers, the perceptual span encompasses about 15 characters to the right of fixation and three to four characters to the left of fixation (McConkie & Rayner, 1975). The perceptual span refers to what can be seen with the help of linguistic knowledge or context, while the visual span refers to what can be seen without that help (O’Regan, 1991). The visual and perceptual spans are often studied separately, but it seems likely that the size of the perceptual span is influenced by the size of the visual span.

Variations of the input from early visual cortex to the reading circuitry may account for some of the variation in reading performance. Pelli et al. (2007, figure 9) demonstrate that there is considerable individual variation in the critical spacing factor (the smallest distance between two letters that avoids crowding) at different eccentricities, and that the visual span is simply the number of characters that are not crowded. The between-subject differences in the FOV may predict individual differences in the visual span, which is likely predictive of the size of the perceptual span. Since the size of the visual and perceptual spans is related to reading speed (Legge et al., 2007; Rayner et al., 2010), the FOV may also predict reading speeds at different locations in the visual field. There is preliminary evidence ($n = 7$) that the size of the FOV is positively correlated with sight word efficiency, as measured with the TOWRE (Test of Word Reading Efficiency). However, linking the FOV to behavior will require more theoretical development and additional measurements. Precise position information may be available in the population activity of the VOTRC rather than in just the pRF and FOV estimates (Majima et al., 2016). Additionally, note that other systems (such as eye movements, language, and memory) contribute to reading so that differences in the FOV may not map directly to differences in reading performance.

**Implications for reading development for impaired and nonimpaired readers**

The VOTRC represents a restricted subset of the information available from earlier retinotopic cortical regions (Wandell et al., 2012). From a developmental perspective, one possibility is that the restricted FOV is present even prior to print exposure. Another possibility is that the foveal and horizontal bias in the FOV of the VOTRC is carved out during the early school years, as the VOT circuitry develops enhanced sensitivity to written words (Ben-Shachar, Dougherty,
Deutsch, & Wandell, 2011b). These two hypotheses could be discriminated by measuring the FOV in prereading children (4–5 years) using words and checkerboards. If the FOV develops through education, the word FOV in prereaders will resemble the checkerboard FOV in both groups. With greater exposure to print, the word-FOV may show increased bias for the central and horizontal parts of the visual field.

The reduction of the word-FOV may reduce irrelevant signals to the reading circuitry and improve word-processing efficiency (Kwon et al., 2007). On the other hand, a FOV that is too small may constrain the use of direct access, holistic strategies for visual word recognition (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). This may lead to slower, piecemeal recognition of written words, particularly for longer words. Indeed, some dyslexics exhibit letter-by-letter reading (Wimmer & Schurz, 2010), and it may be the case that such behavior is correlated with a reduced FOV in the VOTRC. Importantly, reading acquisition is clearly limited by other factors, most prominently phonological processing (e.g., Ziegler & Goswami, 2005). Reading relies on multiple brain circuits that may be constrained or impaired in multiple ways, giving rise to the highly variable individual profiles of poor readers (Ben-Shachar, Dougherty, & Wandell, 2007; Stanovitch, 1986; Wandell & Yeatman, 2013).

Limitations

These measurements and analyses have both biological and experimental limitations. The BOLD measurements are limited by the presence of large sinuses in the neighborhood of the ventral occipitotemporal region. The transverse sinus interferes with the functional responses, which may have an impact on the FOV measurements (Winawer et al., 2010). The variable position of this sinus across participants (see Supplementary Figure 7 for an estimation of its location) may account for some of the between-subject variation that we observe in the FOV of the reading circuitry.

A very large number of experimental parameter choices must be made in designing the localizer. In some cases, the choice does not alter the FOV (small-field vs. large-field localizer). In other cases, the FOV measurements may be less robust to experimental changes. Experimental parameters include stimulus categories, baseline stimuli, and task requirements. Additional experimental choices are made in the design of the pRF portion of the experiment. Here, relevant parameters include the size of the words, the size of the bar aperture, and the rate at which the words refresh within the aperture, in addition to the task require-ments. Further experiments will be needed to assess the effect of these choices on the FOV estimates.

Data analysis introduces further choices. In defining the VOTRC, we select voxels exceeding an uncorrected statistical threshold of \( p < 0.001 \). In calculating the FOV in response to word stimuli, we include voxels for which the pRF model explains greater than 20% of the variance in the data. These parameters select the voxels with the largest and best-modeled responses, but there are many voxels with smaller responses that may contribute to reading performance. The results we report are robust to a wide range of parameters at the level of the group average, but individual FOV estimates may be sensitive to these parameters in a way that is yet to be determined in future studies.

Conclusion

The FOV of the VOTRC is a small subset of the entire field of view available to the human visual system. On average, the FOV extends about 9° into the contralateral visual field and 4° into the ipsilateral visual field, and is largely confined to the horizontal meridian in each visual field. Earlier work demonstrated spatial sensitivity in the VOTRC, which is not the same as reporting a quantitative measurement at every location in the visual field. We use the pRF method to estimate the FOV of the reading circuitry, analyze the reliability of the FOV estimates, and show stimulus-dependence within fixed cortical regions.

In VOT broadly, the size and center of pRFs depend on factors such as attention, task requirements, and stimulus structure (e.g., the size of the words in the aperture; Harvey et al., 2015; Kay et al., 2013; Kay et al., 2015; Silson et al., 2015; Sprague & Serences, 2013). The restricted FOV of the VOT circuitry may be determined by the properties of the neurons within the VOT or the major projections to these neurons from other parts of cortex. It is also possible that voxels contain multiple populations of neurons with different sensitivities (Harvey et al., 2015). Current pRF models do not capture such stimulus and task sensitivities, and thus a next generation of models may be required to explain the broader properties of the VOTRC responses (Kay & Yeatman, 2017). We have archived the experimental data and metadata in a sharable database for anyone who would like to carry out further analyses (Wandell, Rokem, Perry, Schaefer, & Dougherty, 2015).

The properties of the neurons and projections in VOT may develop in response to the extensive training that many children undergo in school. This training may impact several systems, including the VOT reading circuitry, eye-movement circuitry, and connections to
language cortex. There may be multiple system solutions for normal reading performance; a limitation in one part of the circuitry (e.g., a small FOV) may be compensated by another part of the reading circuitry (e.g., a more efficient eye-movement system). The timing of the training with respect to the available plasticity in these different systems may matter for these learning processes. We hope that understanding the developmental processes in the reading circuitry will contribute to effective educational practice (Yeatman, Dougherty, Ben-Shachar, & Wandell, 2012).

Keywords: pRF, retinotopy, functional MRI, word perception, stimulus-dependence, model accuracy

Acknowledgments

This work was supported by NSF grant NSF-BCS1228397 to BW and the Binational Science Foundation (BSF) Grant 2011314 to BW and MB. We thank Michael Barnett for his help with data collection as well as construction of the display in the magnet bore. We thank Justin Gardner, Karen Larocque, Kendrick Kay, Jason Yeatman, Jon Winaer, and Lee M. Perry, and the reviewers and editor for the helpful comments and feedback.

*RL and NW contributed equally to this work.
Commercial relationships: none.
Corresponding author: Rosemary Le.
Email: rosemary.le@stanford.edu.
Address: Psychology Department, Stanford University, Stanford, CA, USA.

References


What role for the visual word form area? Neuroimage, 27(3), 694–705.


