

UC Berkeley

UC Berkeley Previously Published Works

Title

Spatio-temporal properties of letter crowding

Permalink

<https://escholarship.org/uc/item/1fj008cz>

Journal

Journal of Vision, 16(6)

ISSN

1534-7362

Author

Chung, Susana TL

Publication Date

2016-04-18

DOI

10.1167/16.6.8

Peer reviewed

Spatio-temporal properties of letter crowding

Susana T. L. Chung

School of Optometry and Vision Science Graduate Group,
University of California, Berkeley, CA, USA



Crowding between adjacent letters has been investigated primarily as a spatial effect. The purpose of this study was to investigate the spatio-temporal properties of letter crowding. Specifically, we examined the systematic changes in the degradation effects in letter identification performance when adjacent letters were presented with a temporal asynchrony, as a function of letter separation and between the fovea and the periphery. We measured proportion-correct performance for identifying the middle target letter in strings of three lowercase letters at the fovea and 10° in the inferior visual field, for a range of center-to-center letter separations and a range of stimulus onset asynchronies (SOA) between the target and flanking letters (positive SOAs: target preceded flankers). As expected, the accuracy for identifying the target letters reduces with decreases in letter separation. This crowding effect shows a strong dependency on SOAs, such that crowding is maximal between 0 and ~100 ms (depending on conditions) and diminishes for larger SOAs (positive or negative). Maximal crowding does not require the target and flanking letters to physically coexist for the entire presentation duration. Most importantly, crowding can be minimized even for closely spaced letters if there is a large temporal asynchrony between the target and flankers. The reliance of letter identification performance on SOAs and how it changes with letter separations imply that the crowding effect can be traded between space and time. Our findings are consistent with the notion that crowding should be considered as a spatio-temporal, and not simply a spatial, effect.

Introduction

Our ability to perceive the fine details of a visual object is usually better when it is presented alone than when it is surrounded by other objects in close proximity (e.g., Bouma, 1970; Townsend, Taylor, & Brown, 1971). This phenomenon is known as crowding. Crowding has been suggested as the bottleneck for object recognition (Levi, 2008; Pelli & Tillman, 2008; Whitney & Levi, 2011). It is ubiquitous in spatial

vision and has been demonstrated to affect many spatial tasks, including Vernier discrimination (Levi & Klein, 1985; Levi, Klein, & Aitsebaomo, 1985; Westheimer & Hauske, 1975), stereopsis (Butler & Westheimer, 1978), orientation sensitivity (e.g., He, Cavanagh, & Intriligator, 1996; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001; Westheimer, Shimamura, & McKee, 1976), alphanumeric recognition (e.g., Bouma, 1970; Chung, Levi, & Legge, 2001; Pelli, Palomares, & Majaj, 2004; Strasburger, 2005), face recognition (Louie, Bressler, & Whitney, 2007; Martelli, Majaj, & Pelli, 2005) and object recognition (Wallace & Tjan, 2011). A classical property of crowding is that the degrading effect of the flanking objects in close proximity to the target object diminishes with increased distance between the target and the flankers. In addition, many characteristics of crowding such as the effect of target and or flanker contrast (e.g., Chung et al., 2001; Pelli et al., 2004), number of flankers (Pelli et al., 2004; Pöder & Wagemans, 2007), target-flanker similarity (e.g., Bernard & Chung, 2011; Chung et al., 2001; Kooi, Toet, Tripathy, & Levi, 1994), radial-tangential anisotropy in the shape of the crowding zone (Toet & Levi, 1992), and the inward-outward asymmetry effect of flankers (Banks, Bachrach, & Larson, 1977; Bouma, 1970; Petrov, Popple, & McKee, 2007) are based on studies that manipulate certain spatial characteristics of the target and or flankers. As such, crowding has been primarily regarded as a spatial phenomenon.

How does crowding occur? Currently the most popular theories of crowding, including the lower-level ones such as inappropriate feature integration or the higher-level ones such as attention, postulate that the objects of interest (could be parts of, or the whole target) in some way interact or combine with the flankers within a spatial integration region (see Levi, 2008). This spatial region within which interactions between the target and flankers occur is referred to as the crowding zone or the combination field.¹ As an attempt to locate the neural origin of crowding, many investigators used the size of the crowding zone, the *critical spacing*, as a proxy of the receptive field size at

Citation: Chung, S. T. L. (2016). Spatio-temporal properties of letter crowding. *Journal of Vision*, 16(6):8, 1–20, doi:10.1167/16.6.8.

doi: 10.1167/16.6.8

Received June 18, 2015; published April 18, 2016

ISSN 1534-7362

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.



Downloaded From: <http://jov.arvojournals.org/pdfaccess.ashx?url=/data/Journals/JOV/935165/> on 04/19/2016

the neural site at which crowding occurs, and compare how the size of the critical spacing and receptive field change as a function of eccentricity at specific cortical areas. Similar changes in critical spacing and receptive field size with eccentricity are taken as evidence for crowding to occur at a given cortical region. Using this approach, the first cortical site at which crowding occurs has been placed at V1 (Millin, Arman, Chung, & Tjan, 2014), V2 (Bi, Cai, Zhou, & Fang, 2009; Freeman & Simoncelli, 2011), and V4 (Chung, Li, & Levi, 2007; Motter, 2006), depending on the stimulus manipulation, experimental setup and observers' task, etc. Regardless of which conclusion was correct, or whether the aggregate results of these studies simply imply that crowding occurs independently at multiple stages of visual processing (Whitney & Levi, 2011), the receptive field properties of neurons have been linked to crowding. However, the properties of all receptive fields are dynamic and are subject to the spatio-temporal interactions of stimulus characteristics or manipulations. Although the spatial properties of the crowding zone are quite well characterized, the temporal properties of crowding, and in particular, the spatio-temporal limitations of the crowding zones, are less well understood.

One of the better-known temporal properties of crowding is the effect of stimulus duration. In general, the critical spacing is larger for stimulus presented for a shorter duration than for a longer one (Chung & Mansfield, 2009; Kooi et al., 1994; Tripathy & Cavanagh, 2002; Wallace et al., 2013). Chung and Mansfield (2009) reported a reduction in the critical spacing by approximately half when the stimulus duration increased from 53 to 1000 ms (for targets and flankers with the same contrast polarity). Wallace et al. (2013) compiled data from several studies (including their own), and showed that in general, the critical spacing is reduced by half when the stimulus duration increases by a factor of 13. Tripathy, Cavanagh, and Bedell (2014) also independently reported a similar magnitude of reduction in the critical spacing for a similar range of change in stimulus duration.

Fewer studies have examined the spatio-temporal limitations on crowding. Huckauf and Heller (2004) measured the accuracy of recognizing a target letter flanked by two other letters that were presented at different stimulus onset asynchronies (SOAs) with respect to the target letter. The target and flanking letters were presented for 50 ms. The authors found that the accuracy for recognizing the target letter was lower when the flankers appeared after the target (positive SOAs) than when the flankers preceded the target (negative SOAs). Most importantly, the authors observed a strong interaction between this SOA dependency of recognition accuracy and the spatial

separations between letters, or retinal eccentricity. However, because these authors used a fixed physical letter size and letter spacing at different eccentricities, when expressed in angular subtense, the letter size and spacing were quite large with respect to the resolution limit and critical spacing at their smallest (1°) eccentricity, but were small at their largest (7°) eccentricity. Therefore, the observed effects might have been confounded with the letter size and spacing effects, and might not have reflected the genuine effect of eccentricity. The target and flankers were presented for the same duration in Huckauf and Heller (2004); other studies had presented target and flankers of different durations but synchronizing their onset or offset. For example, Ng and Westheimer (2002) measured visual acuity using Landolt C targets closely flanked by four bars. The Landolt C targets were presented for 150 ms whereas the flanking bars, only briefly shown for 50 ms, were presented with different SOAs relative to the Landolt Cs. These authors found that Landolt C acuity was most degraded by flanking bars appearing 50 to 100 ms after the onset of the C, and was not affected by flanking bars appearing 50 ms after the offset of the Landolt Cs. More recently, Harrison and Bex (2014) examined how the critical spacing is affected by the timing of the target relative to the flankers. Their target and flankers always coexisted for 58 ms, but the flanker duration could be longer than that of the target, such that the onset of the flankers could occur up to 450 ms before the target onset (with synchronized target and flanker offset), or the offset of the flankers could occur up to 450 ms after the target offset (with synchronized target and flanker onset). They found that the critical spacing was smaller when flanker-onset occurred before target-onset, and larger when flanker-offset occurred after target-offset. From their data, they deduced a 45-ms window over which target and flanking letters interacted to produce a reduction in performance. Although this study clearly shows that there is a temporal window over which target and flankers presumably interact with one another, thus producing crowding, the authors only reported one single value for the temporal window. Is this 45-ms temporal window a universal temporal window that applies to all conditions? Does it change with the size of the crowding zone (hence, a spatio-temporal effect) or eccentricity?

These cited studies clearly demonstrated a substantial spatio-temporal interaction on the magnitude or the extent of crowding, but to our knowledge, we do not know of a systematic investigation on the interplay between spatial and temporal separations, and their interaction with retinal eccentricity on the magnitude of crowding. The understanding of the spatio-temporal interaction on crowding would provide useful information as to how to minimize crowding, which could remove the bottleneck on object recognition, and in

turn, improve visual performance. Therefore, the primary goals of this study were to (1) systematically examine how spatial crowding depends on the temporal characteristics of the target and flankers, in particular, the temporal asynchrony between the target and flankers; and (2) derive from the data the critical size (in both the spatial and temporal domains) of the crowding region or window.

To preview our findings, the dependence of crowding on SOA shows a clear and systematic modulation by the spatial separation between the target and its flankers, confirming previous observations that there exists a strong spatio-temporal interaction on crowding. Crowding is maximal for SOA between 0 and 100 ms, and does not require the target and flankers to coexist for the entire presentation duration. In fact, under some conditions, maximal crowding occurs when the target and flankers never coexist physically. We also observed a trade-off between the spatial and temporal separations between the target and its flankers such that crowding is absent when the spatial *or* the temporal separation is large (but not necessary for both spatial and temporal separations to be large). This result implies that we could eliminate crowding, and thus improve object recognition, by either a large spatial separation or a long temporal asynchrony between the target and its flankers.

Methods

In this paper, we used letters as stimuli, for both the target and flankers. Observers' task was to identify the middle target letters in strings of three lowercase letters (trigrams). Letters making up the trigrams were chosen randomly from the 26 lowercase letters of the Times-Roman alphabet, and were rendered as black letters (2.2 cd/m^2) on a white background (147 cd/m^2). For comparison, we also measured the identification accuracy for single letters. The difference in performance accuracy for identifying the target letters in trigrams, compared with that for single unflanked letters, represents the magnitude of crowding. To examine the temporal properties of letter crowding, we introduced an SOA between the middle target letter of a trigram and its two flanking letters, defined as the difference in the onset time (ms) of the flankers relative to the target. Positive SOAs mean that the onset of the target letter occurred before that of the flankers whereas negative SOAs mean that the onset of the flankers occurred before that of the target letter (see Figure 1). Considering that crowding is most substantial when the spatial separation between a target and its

flankers is small, and decreases with larger separations, we sought to determine if there exists a systematic interaction between SOAs and spatial separations between the target and flankers, and how this interaction changes between the fovea and periphery. Thus, we measured how identification accuracy changed with SOA for a range of letter (spatial) separations, at the fovea and at 10° eccentricity in the inferior visual field. Note that our basic experimental paradigm is similar to that of Huckauf and Heller (2004), with the following improvements. First, we tested SOAs in steps of 25 ms so that we could get a much finer resolution in the change in performance versus SOA (Huckauf & Heller used 50-ms steps). This is important for deriving the critical spatio-temporal window for crowding (Figure 6). Second, we tested a range of smaller letter separations ($0.8\text{--}2 \times$ the letter-size; compared with their range that corresponded to $1.1\text{--}5.3 \times$ the letter-width) that should yield more crowding. Third, instead of testing negative or positive SOAs in separate blocks of trials and with different groups of observers as in Huckauf and Heller (2004), we tested all the SOAs in a random order within the same block of trials to avoid observers using different strategies when responding to trials with negative or positive SOAs. Fourth, we included the baseline performance for identifying single, unflanked letters for comparison with the performance for the flanked conditions. Without the baseline condition, it is difficult to quantify the magnitude and extent of crowding, and to compare the magnitude and extent of crowding across conditions or observers. Fifth, letter size was scaled in the periphery to avoid observing effects that were limited by resolution, instead of crowding. Sixth, our analyses focused on the within-subject comparison of the different conditions, instead of a between-subject approach.

Observers

Three experienced psychophysical observers (including the author) participated in this study. All had normal or corrected-to-normal vision (20/20 or better in each eye) and had prior experience in other psychophysical studies that involved the use of peripheral vision. Except for the author, the other two observers were not aware of the purpose of the study. Testing was performed binocularly in a dimly-lit room. Written informed consent was obtained from each observer after the procedures of the experiment were explained and before the commencement of data collection. The research followed the tenets of the Declaration of Helsinki.

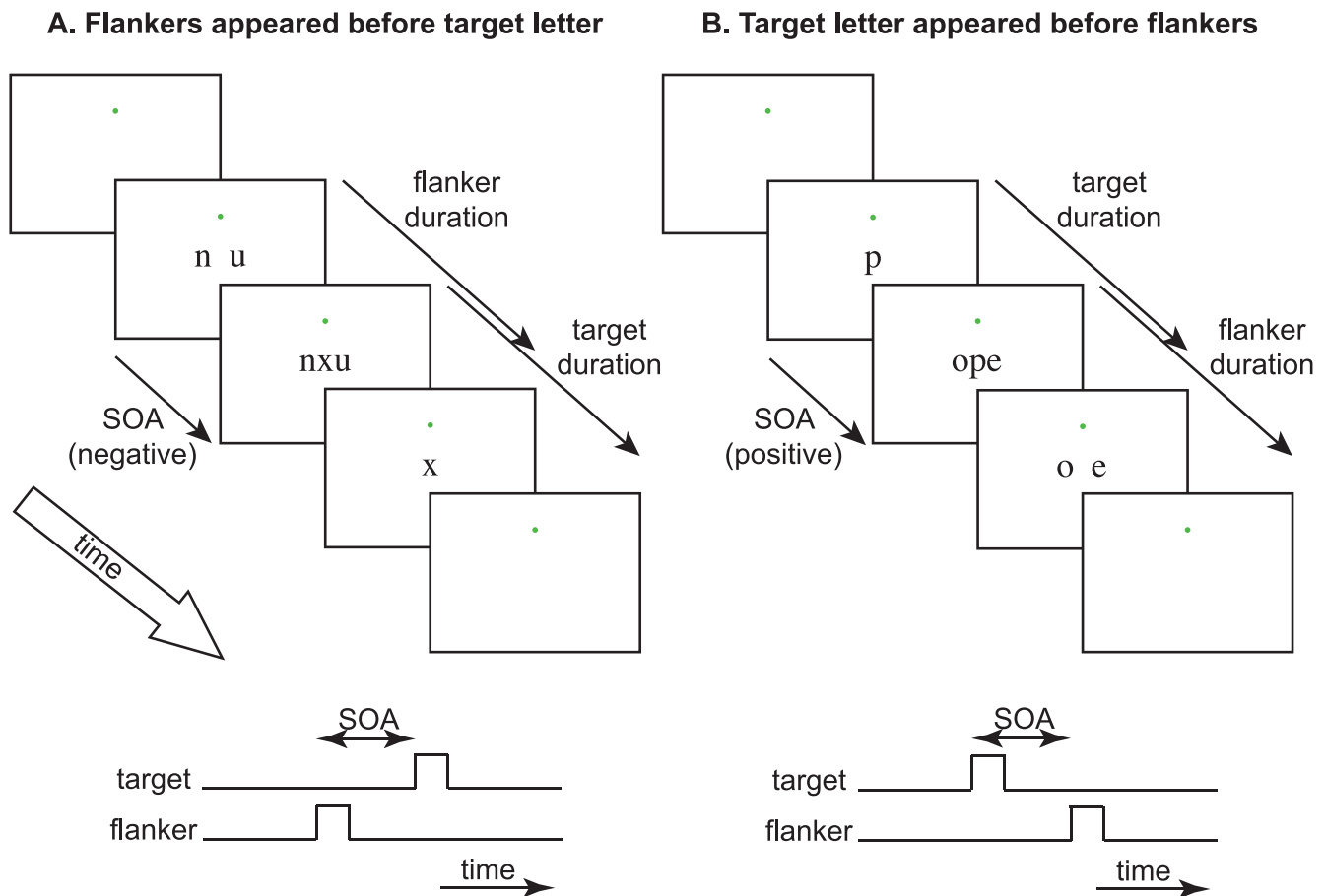


Figure 1. A schematic cartoon depicting two sample trials with a negative (A) and a positive (B) target-flanker SOA, respectively. A negative SOA means that the two flanking letters (in this example, letters *n* and *u*) appear before the target letter (*x* in this example); whereas a positive SOA means that the target letter (*p* in this example) appears before the two flanking letters (*o* and *e* in this example).

Apparatus

Stimuli were generated using a Visual Stimulus Generator (VSG) 2/5 graphics board (Cambridge Research Systems, Rochester, UK) controlled by a Dell Precision 650 workstation using custom software written in MATLAB 7.3.0 (MathWorks, Natick, MA) and presented on a Sony 24-in. color graphics display monitor (Model No. GDM-FW900, Japan). The resolution of the display was 1280×960 pixels, with a refresh rate of 80 Hz. The temporal dynamics of the display was verified using a photodetector and an oscilloscope, DSO1024A, Agilent Technologies, Santa Clara, CA). The luminance of the display was linearized and calibrated using the VSG OptiCAL software, together with a Minolta CS-100 photometer (Ramsey, NJ). A forehead and chin rest was used to minimize observers' head movements and to maintain a constant viewing distance of 300 cm for foveal testing and 75 cm for peripheral testing. At these distances, each pixel subtended a visual angle of 0.34 and 1.38 arc min, respectively.

Psychophysical procedures and stimulus parameters

Crowding can be limited by stimulus size or spacing (Chung, 2014; Coates, Chin, & Chung, 2013; Song, Levi, & Pelli, 2014) but the spacing limitation still requires that the stimulus size exceeds the resolution limit. To ensure that our letter size exceeds the resolution limit at the testing eccentricity, we first measured how identification accuracy changes with letter size, and chose a letter size for subsequent testing accordingly. For each observer and at each testing eccentricity (fovea or 10° inferior visual field), we used the method of constant stimuli to present single letters at six letter sizes (defined as the *x*-height) and measured the proportion-correct identification performance at each letter size. Letters were exposed for 50 or 100 ms (tested in separate blocks). From the psychometric function, we chose a letter size that corresponded to approximately 80%–90% correct as the target size used in subsequent experiments examining the temporal properties of crowding. This letter size exceeded the resolution threshold (usually

defined as 50% correct on the psychometric function, after correction for guessing) while ensuring that our results would not be limited by a ceiling effect (100% correct). Averaged across the three observers, these letter sizes were 0.097° (range: 0.09° – 0.1°) and 0.9° (0.7° – 1°) at the fovea and 10° inferior visual field for letter exposure duration of 50 ms; and 0.059° (0.05° – 0.07°) and 0.68° (0.56° – 0.9°) at the fovea and 10° inferior visual field for letter exposure duration of 100 ms.

Accuracy for identifying the middle target letters in trigrams was then measured for a range of letter (spatial) separations, each for a range of target-flanker SOAs. Letter separation was defined as the distance between the center of the target letter and the center of either flanking letter, and was normalized to the height of a lowercase letter “x.” Four letter separations were tested: $0.8\times$, $1\times$, $1.25\times$, and $2\times$ the *x*-height. At the smallest letter separation ($0.8\times$), adjacent letters frequently touched but did not significantly overlap with one another, except for the two wider letters “m” and “w.” Considering that all the letters were randomly drawn, with equal probability, from the set of 26 letters of the alphabet, the chance that a trigram was made up of only “m” or “w” was low. The exposure duration of the target and flanking letters were identical and were either 50 or 100 ms (tested in separate blocks). Only one combination of letter separation, eccentricity (fovea or 10° eccentricity) and letter exposure duration was tested in a block of trials. In each block of trials, we used the method of constant stimuli to present trials at 11 SOAs (ranging from -100 to $+150$ ms, in steps of 25 ms [two video frames]) when the letter-exposure duration was 50 ms, and 13 SOAs (ranging from -150 to $+150$ ms, also in steps of 25 ms) for letter-exposure duration of 100 ms, with 20 trials for each SOA. The presentation order of these trials was randomized within a block. Each condition was tested three times (blocks) for each observer, so that there were a total of 60 presentations for each combination of eccentricity, letter separation, letter exposure duration (50 or 100 ms) and SOA for each observer. The order of testing conditions was randomized for each observer, so that any potential effect of improvement in performance due to familiarity with the task² over the course of the experiment would not be limited to conditions tested later.

For foveal testing, a pair of small green dots separated vertically by 0.42° straddled the center of the display. Observers were asked to fixate the center of the two green dots, where the target letter would be presented. For peripheral testing, the fixation target was a single green dot positioned 10° above the center of the display (see Figure 1), where the target letter would be presented. Observers initiated each trial by pressing the spacebar on the keyboard. Following a short delay, the target and flanking letters appeared

with the SOA for that trial. After all the letters disappeared from the display, the observer indicated his or her response of the identity of the target letter using a computer keyboard. Feedback was not provided so as to discourage observers from learning to associate a particular spatial pattern of letters with the correct response. Eye movements were not monitored, although casual observations of the observers’ eyes from the side during testing suggested that observers fixated well at the fixation green dot. As can be seen later (see Results), identification performance as a function of SOA was very systematic across all conditions, a finding that is difficult to obtain had the observers moved their eyes to look at the target letters from time to time.

All observers practiced the task (more for peripheral viewing) for 2–3 sessions (1–1.5 hr per session) until they were comfortable with the task before actual data collection. Data collected during the practice sessions were excluded for analyses and are not reported in this paper.

Data analyses

We used Igor Pro 6.37 (WaveMetrics Inc., Lake Oswego, OR) to perform the curve-fitting shown in Figures 3 and 5. For panels showing individual observers’ data (columns 1–3), each data point represents the performance pooled across the 60 trials of the same condition. In the last column showing group data, each data point represents the performance pooled across all the trials for the three observers. For a given set of data, the best-fitted curve was one that minimized the Chi-square error between the experimental and the model fit, based on a Levenberg–Marquardt iterative algorithm (a form of nonlinear least-squares fitting). Statistical analyses were performed using the R software (R Development Core Team, 2014).

Results

50-ms letter exposure duration

Proportion-correct for identifying the target letters is plotted as a function of target-flanker SOA in Figure 2, with letter separation as a parameter. The first three columns of panels present data for the three individual observers, and the last column presents the aggregate data pooled across the three observers. Data obtained at the fovea and 10° inferior visual field are shown in the upper and lower panels, respectively. For comparison, proportion-correct performance for identifying single letters is given in each panel as the dashed line.

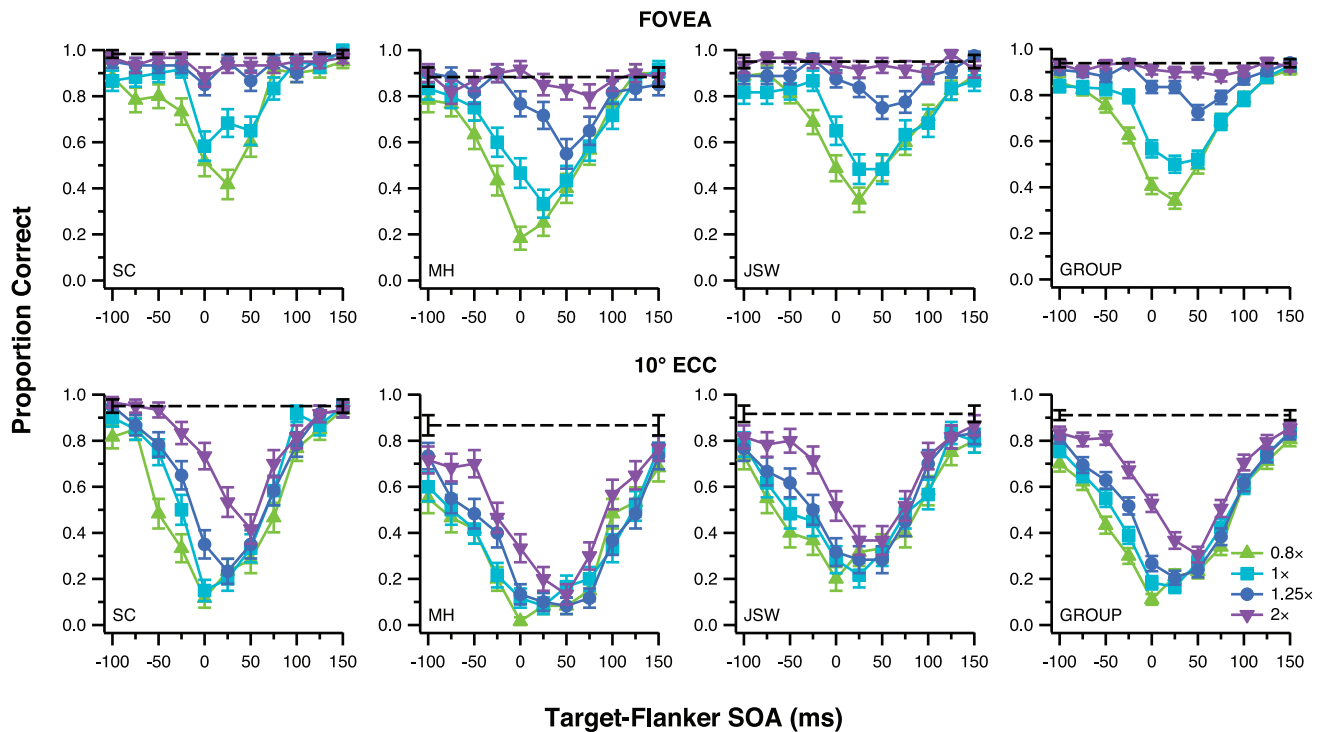


Figure 2. Proportion-correct for identifying the target letter is plotted as a function of the target-flanker SOA (in ms) for the three observers (columns 1–3), for letter exposure duration of 50 ms. The rightmost column shows the group data pooled across the three observers. Data obtained at the fovea are presented in the upper panels while data obtained at 10° in the inferior visual field are presented in the bottom panels. In each panel, results are plotted separately for the four nominal letter separations (coded by different colored symbols). The black dashed line in each panel represents the accuracy of identifying single letters. Error bars represent the standard errors of proportion.

The difference in performance between the single-letter and the flanked-letter conditions is taken to represent the magnitude of the crowding effect. In general, plots relating proportion-correct and SOA show a V-shaped function, such that the magnitude of crowding is maximal at some SOA (close to but not necessarily zero) and is reduced when the absolute value of the difference between an SOA and the SOA corresponding to maximal crowding increases, regardless of whether the flankers or the target letter appeared first. However, the reduction in the magnitude of crowding is asymmetrical on the two sides of the V-shaped function. In particular, crowding seems to be stronger on the right-hand limb than the left, meaning that positive SOAs induce more crowding than negative SOAs. This result is consistent with observations for temporal interference of Vernier discrimination (Westheimer & Hauske, 1975), stereopsis (Butler & Westheimer, 1978), and judgment of Gabor orientations (Song & Levi, 2010). For all observers and at both testing eccentricities, the four curves (coded by different colors), representing data for the four letter separations, show the following systematic changes as the letter (spatial) separation increases. First, the SOA at which maximal magnitude of crowding occurs shifts

toward larger and more positive SOAs. Second, the maximal magnitude of crowding (largest dip in the V-shaped function) becomes smaller. In addition, the right-hand limbs of the four curves appear to collapse into one single function, whereas the left-hand limbs do not.

Because the proportion-correct accuracy for identifying *single* letters differed among observers and testing eccentricities, and because proportion-correct is not a linear measurement (a reduction in proportion-correct from 0.9 to 0.6 is different from a reduction in proportion-correct from 0.8 to 0.5), to facilitate a quantitative comparison of the magnitude of crowding, we converted letter identification performance from proportion-correct to *z*-score unit. The difference in *z*-score units between the single- and flanked-letter conditions can then be used as a quantitative measurement of the magnitude of crowding. With this transformation, plots of identification performance versus SOA become an inverted-V shape, but the important characteristics of each plot such as the SOA at which the magnitude of crowding is maximal and the asymmetry between the left and right limb of each plot remain unchanged (Figure 3). To quantify the SOA tuning characteris-

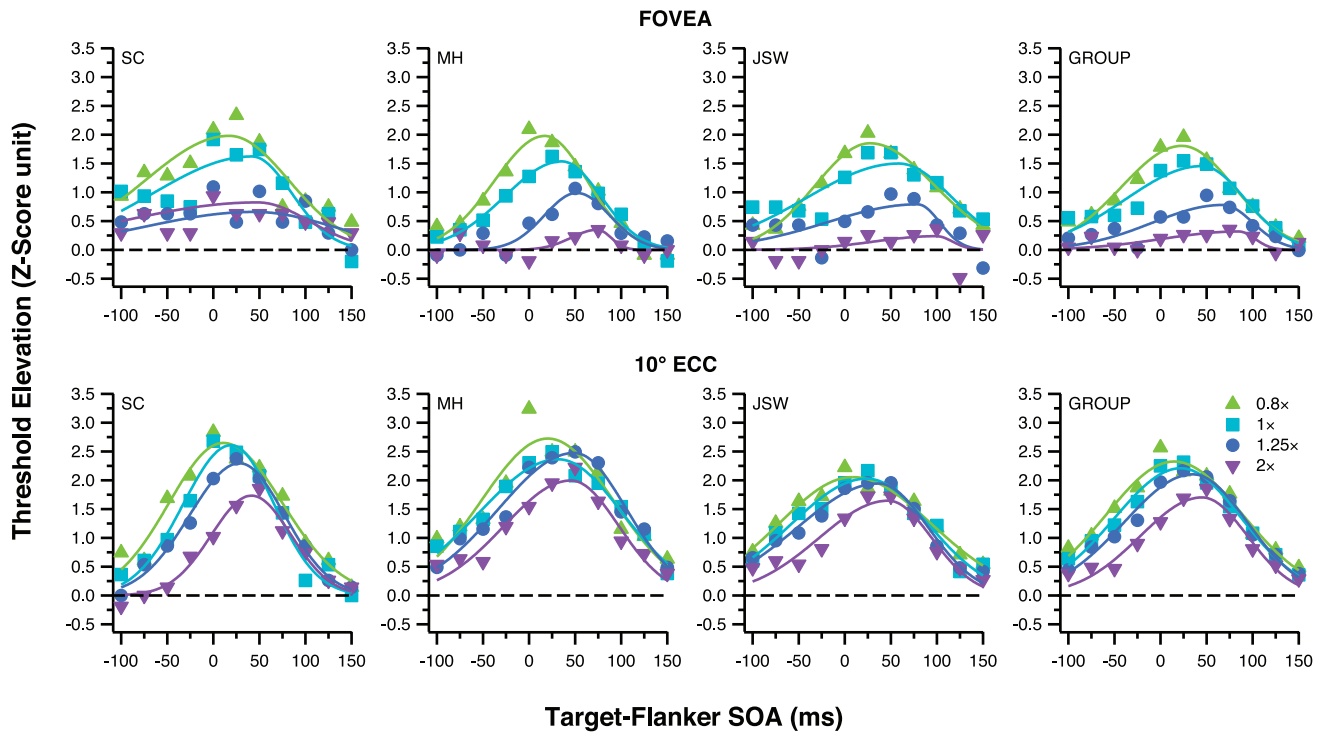


Figure 3. Proportion-correct data as shown in Figure 2 are transformed into differences in z-score units (see text for details), as a quantitative measurement of the crowding magnitude. A z-score unit of 0 implies that there is no performance difference in identifying flanked target letters and single letters, in other words, there is no crowding. Each panel shows data from one observer (the last panel in each row shows the group data pooled across the three observers) tested at the fovea (upper panels) or 10° in the inferior visual field (bottom panels). Results for the four nominal letter separations are plotted in different colored symbols, as in Figure 2. The smooth curve through each set of color symbols represents the best-fit asymmetric Gaussian function (see text for details).

tics of the data, we fit each data set using an asymmetric Gaussian function, as given by the following:

$$f(x) = \begin{cases} A \times \exp\left(-\left(\frac{x - x_p}{\sigma_L}\right)^2\right) & \text{if } x < x_p \\ A \times \exp\left(-\left(\frac{x - x_p}{\sigma_R}\right)^2\right) & \text{if } x \geq x_p \end{cases}$$

where $f(x)$ is the difference in z-score units between the flanked condition at a given SOA x and the single-letter condition, A is the peak amplitude of the Gaussian function, representing the maximal magnitude of crowding, x_p is the SOA at which crowding magnitude is maximal, σ_L and σ_R are the standard deviations of the left- and right limb of the Gaussian function, respectively. The fitted values for the different observers and conditions are given in Table 1.

In general and as expected, crowding is more substantial in the periphery than at the fovea, as illustrated by the larger differences in the z-score for the proportion-correct versus SOA plot, repeated-measures

ANOVA with Satterthwaite approximation for degrees of freedom: $F_{df=(1, 4)} = 15.48, p = 0.017$, for all four nominal letter separations.³ Crowding is also more substantial for small letter separations and decreases with larger letter separations, repeated-measures ANOVA: $F_{df=(3, 15.3)} = 74.01, p < 0.0001$. Both of these findings are consistent with what we understand based on spatial crowding when the target and flanking letters are present simultaneously (SOA = 0). The more interesting finding is that the crowding magnitude shows a strong dependency on the target-flanker SOA, repeated-measures ANOVA: $F_{df=(10, 45.6)} = 55.59, p < 0.0001$, contributing to the tuning characteristics. In particular, crowding is maximal when the target and flankers were present close in time and is reduced when the target and flankers were physically well separated in time.

Besides the main effects of eccentricity, letter separation and SOA, the magnitude of crowding also depends on the interactions of these three main factors. Specifically, the interaction of eccentricity and letter separation on the crowding magnitude is illustrated by how the four curves for the four letter separations do not stack up in the same manner (and

the same height) between the fovea and the periphery, repeated-measures ANOVA: $F_{df=(3, 10.6)} = 7.57$, $p = 0.0055$. In addition, the dependence of the crowding magnitude on SOA is different between the fovea and the periphery, repeated-measures ANOVA: $F_{df=(10, 33.5)} = 7.15$, $p < 0.0001$. This interaction effect is manifested as differences in the specific shape of the SOA-tuning functions (e.g., the location of the peak; the skewness; or the asymmetrical differences between the left- and the right-limbs of the functions) between the fovea and the periphery. The dependence of the crowding magnitude on SOA is also different across letter separations, resulting in differences in the shape and or the positions of the SOA-tuning functions for the four letter separations, repeated-measures ANOVA: $F_{df=(30, 65.8)} = 4.93$, $p < 0.0001$. Furthermore, how the crowding magnitude changes with SOA for a given nominal letter separation also changes with eccentricity, implying a significant three-way interaction effect, repeated-measures ANOVA: $F_{df=(30, 58.7)} = 3.54$, $p < 0.0001$.

These interaction effects are clearly illustrated as a shift of the peak of the SOA-tuning function along the SOA-axis, as a function of letter separation, or eccentricity, or both. To quantify these shifts, we compared across different conditions the SOA at which the magnitude of crowding is maximal. A separate repeated-measures two-way ANOVA was performed on these data, with eccentricity (two levels: fovea and 10° eccentricity) and letter separation (four levels: $0.8\times$, $1\times$, $1.25\times$, and $2\times$) as main factors. In general, SOA corresponding to maximal magnitude of crowding is not affected by eccentricity, $F_{df=(1, 2)} = 4.64$, $p = 0.16$, but is affected by letter separation, $F_{df=(3, 6)} = 28.25$, $p = 0.0006$. At the fovea, the magnitude of crowding was maximal at an SOA (averaged across the three observers) of 20.6 ms (range: 17.1–27.3 ms) for the smallest letter separation ($0.8\times$ the x -height), and shifted toward an SOA of 72.6 ms (44.3–99.2 ms) for the largest letter separation ($2\times$ the x -height). At 10° eccentricity, the SOA corresponding to maximal magnitude of crowding also shifted toward a more positive SOA as letter separation increased—from 13.2 ms (8.5–20.4 ms) for the smallest letter separation to 44.7 ms (41.5–47.2 ms) for the largest letter separation. For all conditions, the SOA corresponding to maximal crowding occurred at a nonzero value, implying that maximal magnitude of crowding does not require the target and flankers to be physically present simultaneously (SOA = 0). Another interesting observation from Figures 2 and 3 is that the SOA corresponding to maximal crowding shifts toward more positive SOAs when letter separation increases, indirectly causing the right limbs of the SOA tuning functions to collapse into a single function.

100-ms letter exposure duration

Figure 4 plots the proportion-correct for identifying the target letters as a function of target-flanker SOA, when letters (both target and flankers) were presented for 100 ms. The general characteristics of these curves are very similar to what we observed for a letter exposure duration of 50 ms (Figure 2).

Just as in Figure 3, we quantified the magnitude of crowding at each SOA by calculating the difference in z -score units between the single- and flanked-letter conditions. These differences in z -score units are plotted as a function of SOA in Figure 5. We then fit each dataset using the asymmetric Gaussian function to derive the key parameters of the crowding magnitude versus SOA tuning function, as for the 50-ms letter exposure duration data. The fitted values of the key parameters are summarized in Table 1.

In general, results for the 100-ms letter exposure duration replicate the observations for the 50-ms letter exposure duration. Crowding is more substantial in the periphery than at the fovea, repeated-measures ANOVA: $F_{df=(1, 2)} = 18.45$, $p = 0.05$; more substantial for small letter separations than for larger ones, repeated-measures ANOVA: $F_{df=(3, 6)} = 76.45$, $p < 0.0001$, and also shows a strong dependence on SOA, repeated-measures ANOVA: $F_{df=(12, 24)} = 31.70$, $p < 0.0001$. The interactions among the three main factors of eccentricity, letter separation and SOA on the crowding magnitude are also all significant (eccentricity \times letter separation: $p = 0.039$; SOA \times eccentricity: $p = 0.023$; SOA \times letter separation: $p < 0.0001$; eccentricity \times letter separation \times SOA: $p < 0.0001$).

We also performed a separate repeated-measures two-way ANOVA on the SOA that corresponds to maximal magnitude of crowding, with eccentricity and letter separation as main factors. Consistent with the results for the 50-ms letter exposure duration, SOA at which the magnitude of crowding was maximal is not affected by eccentricity, $F_{df=(1, 2)} = 3.47$, $p = 0.20$, but is affected by letter separation, $F_{df=(3, 6)} = 7.83$, $p = 0.017$. At the fovea, the magnitude of crowding was maximal at an SOA of 20.6 ms (–9.9 to 43.0 ms) for the smallest letter separation ($0.8\times$ the x -height), and shifted toward an SOA of 88.3 ms (78.7–102.4 ms) for the largest letter separation ($2\times$ the x -height). At 10° eccentricity, the SOA at which maximal crowding occurred also shifted toward more positive SOA as letter separation increased—from 27.8 ms (5.9–43.3 ms) for the smallest letter separation to 49.9 ms (39.9–64.5 ms) for the largest letter separation. Again, all of these values are greater than zero, confirming that maximal crowding does not require the target and flankers to physically coexist for the entire presentation duration.

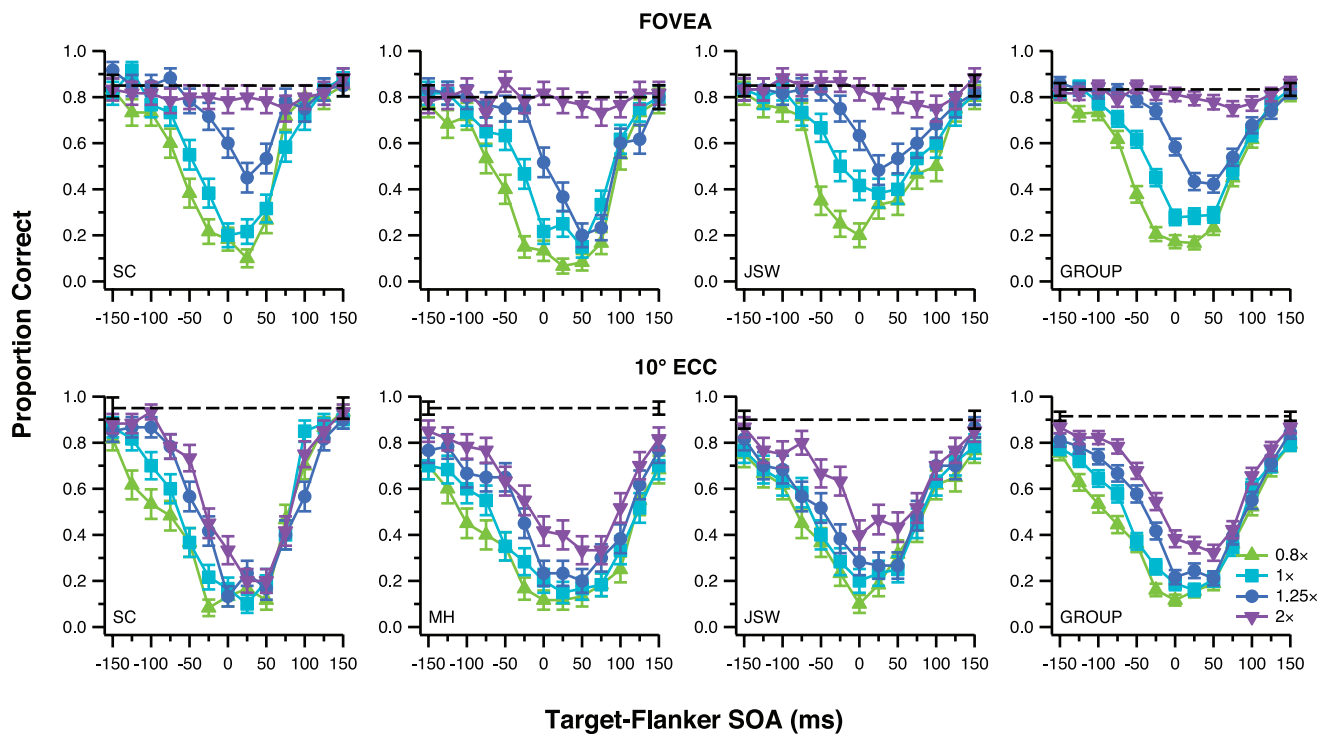


Figure 4. Proportion-correct for identifying the target letter is plotted as a function of the target-flanker SOA, for letter exposure duration of 100 ms. Details of the figure are as in Figure 2.

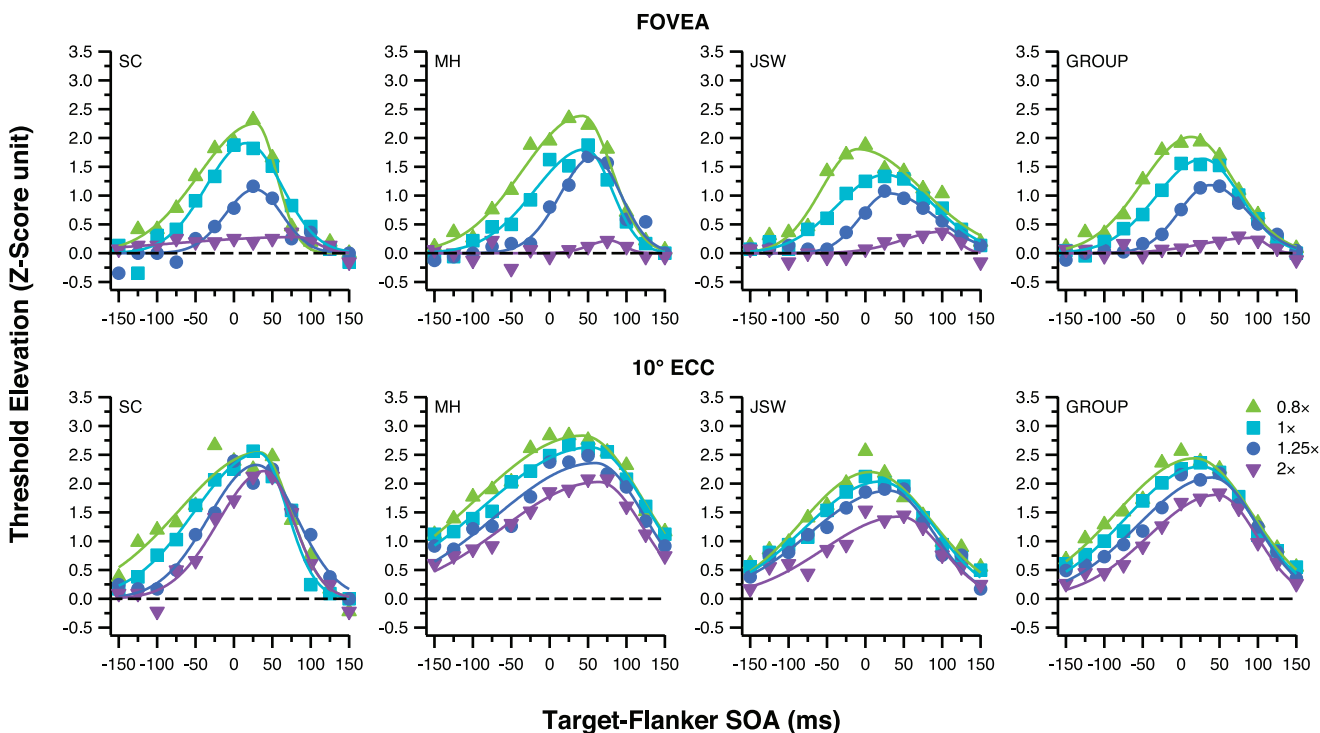


Figure 5. Data shown in Figure 4 are replotted with proportion-correct transformed into differences in z-score units (see text for details). Details of the figure are as in Figure 3.

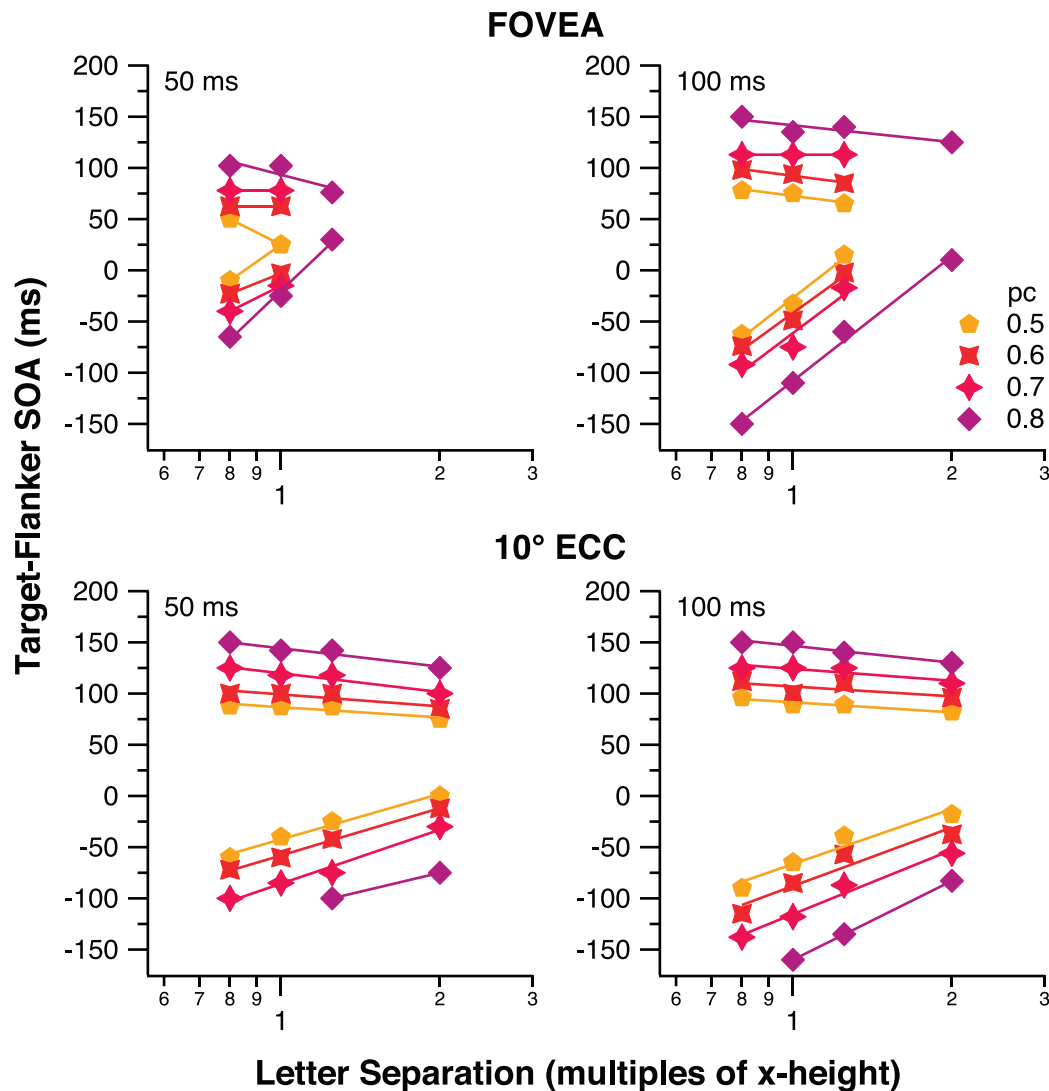


Figure 6. Criterion target-flanker SOA (in ms) is plotted as a function of nominal letter separation, for the two letter exposure durations (left: 50 ms; right: 100 ms), for data obtained at the fovea (upper panels) and 10° inferior visual field (lower panels). Each datum is derived from the fitted curve shown in Figures 3 or 5, based on the group data, and represents the combination of SOA and letter separation that yields a given criterion performance, which is color-coded for proportion correct (pc) of 0.5, 0.6, 0.7, or 0.8. Straight line through each set of colored symbols in each panel represents the best-fit line (on semilog axes). Slopes of these lines (only for data-sets with more than two data points) are given in Table 2.

At each testing eccentricity, the right limbs of the functions for the four letter separations also collapse into a single function, similar to what we observed for the data for 50-ms letter exposure duration.

Considering the qualitative similarities of the results for the two letter exposure durations, a logical question to ask is whether there are any differences in results between the two letter exposure durations. A separate ANOVA show that there is no main effect of stimulus exposure duration, $F_{df=(1, 2.5)} = 4.84$, $p = 0.13$, on the SOA corresponding to maximal magnitude of crowding, implying that the letter exposure duration is not an important factor limiting the SOA for maximal crowding to occur.

Spatio-temporal window for crowding

The second goal of this study was to derive the spatio-temporal window that needs to be exceeded to minimize crowding. To do so, we computed from the fitted asymmetric Gaussian curves for the group-data the SOAs that corresponded to a criterion level of identification performance of the target letter, for different letter separations and at the two retinal eccentricities tested. These SOAs are plotted in Figure 6 as a function of letter separation for four performance criteria—proportion of correct identification of 0.5, 0.6, 0.7, and 0.8. Essentially, these plots represent the minimum temporal (ordinate) and spatial (abscissa)

Observer	Eccentricity	Nominal letter separation	Letter exposure duration = 50 ms				Letter exposure duration = 100 ms			
			A	x_p (ms)	σ_L	σ_R	A	x_p (ms)	σ_L	σ_R
SC	0° (fovea)	0.8X	1.98 ± 0.19	17.4 ± 24.1	129.7 ± 43.8	91.8 ± 29.0	2.25 ± 0.08	28.6 ± 4.6	104.9 ± 6.7	38.3 ± 5.7
		1X	1.62 ± 0.22	41.9 ± 27.7	151.7 ± 56.1	61.1 ± 31.9	1.91 ± 0.11	16.4 ± 8.5	74.3 ± 11.0	65.7 ± 10.7
		1.25X	0.96 ± 0.15	38.3 ± 33.9	157.8 ± 71.7	81.7 ± 39.5	1.11 ± 0.13	27.4 ± 13.3	54.9 ± 16.9	50.4 ± 16.9
		2X	0.66 ± 0.10	44.3 ± 61.8	168.9 ± 108	122.6 ± 98.9	0.28 ± 0.04	83.7 ± 26.5	231.2 ± 49.9	34.9 ± 11.7
		0.8X	2.65 ± 0.13	10.7 ± 9.6	83.3 ± 12.7	90.8 ± 12.4	2.54 ± 0.14	34.0 ± 10.4	148.4 ± 17.5	55.9 ± 12.4
	10°	1X	2.61 ± 0.16	19.3 ± 8.9	73.6 ± 11.6	66.1 ± 11.1	2.53 ± 0.08	33.0 ± 5.0	118.0 ± 7.5	51.3 ± 6.1
		1.25X	2.29 ± 0.07	28.5 ± 5.0	76.9 ± 6.5	69.2 ± 6.4	2.32 ± 0.12	28.9 ± 8.8	87.9 ± 11.4	74.9 ± 11.2
		2X	1.73 ± 0.09	41.4 ± 7.0	61.6 ± 9.0	60.6 ± 9.0	2.22 ± 0.11	39.9 ± 7.3	84.9 ± 9.9	53.7 ± 8.9
		0.8X	1.98 ± 0.11	17.0 ± 8.5	77.3 ± 11.3	65.3 ± 10.6	2.38 ± 0.08	43.0 ± 5.6	110.4 ± 7.9	54.1 ± 6.8
		1X	1.54 ± 0.08	35.4 ± 8.1	87.1 ± 11.0	60.1 ± 9.9	1.80 ± 0.09	47.0 ± 7.0	93.7 ± 9.5	48.5 ± 8.5
MH	0° (fovea)	1.25X	0.99 ± 0.13	51.9 ± 14.1	49.7 ± 17.8	50.3 ± 18.0	1.67 ± 0.11	56.5 ± 7.6	60.8 ± 9.9	51.7 ± 9.7
		2X	0.35 ± 0.14	74.1 ± 44.4	40.2 ± 53.0	21.0 ± 38.7	0.22 ± 0.12	78.7 ± 17.2	38.4 ± 35.7	25.0 ± 24.9
		0.8X	2.72 ± 0.20	20.4 ± 17.0	101.8 ± 23.9	100.5 ± 22.6	2.83 ± 0.06	43.3 ± 8.5	197.9 ± 14.1	110.8 ± 12.4
		1X	2.36 ± 0.08	31.6 ± 9.3	121.7 ± 14.0	98.4 ± 12.4	2.63 ± 0.05	53.9 ± 7.5	199.3 ± 12.1	101.9 ± 11.2
		1.25X	2.47 ± 0.09	49.1 ± 8.8	116.5 ± 12.3	80.1 ± 11.4	2.36 ± 0.10	59.9 ± 13.9	184.5 ± 21.3	90.2 ± 19.9
	10°	2X	2.00 ± 0.12	45.4 ± 11.6	102.3 ± 15.8	73.3 ± 14.7	2.03 ± 0.05	64.5 ± 7.4	178.4 ± 11.1	81.4 ± 10.1
		0.8X	1.85 ± 0.08	27.3 ± 8.4	84.0 ± 10.5	98.9 ± 11.8	1.81 ± 0.10	-9.9 ± 10.9	67.8 ± 13.1	120.1 ± 16.1
		1X	1.50 ± 0.15	58.6 ± 26.1	140.6 ± 40.0	81.9 ± 35.2	1.36 ± 0.05	26.2 ± 6.4	89.2 ± 8.1	90.0 ± 8.6
		1.25X	0.79 ± 0.19	79.0 ± 35.0	138.5 ± 59.1	33.9 ± 43.4	1.04 ± 0.04	30.7 ± 6.0	52.4 ± 7.3	85.4 ± 8.5
		2X	0.24 ± 0.16	99.2 ± 22.6	95.7 ± 84.8	25.0 ± 23.1	0.38 ± 0.07	102.4 ± 7.5	76.5 ± 20.2	25.0 ± 27.8
JSW	0° (fovea)	0.8X	2.05 ± 0.07	8.5 ± 9.9	113.2 ± 15.7	120.6 ± 13.5	2.20 ± 0.09	5.9 ± 11.7	126.0 ± 16.2	114.8 ± 15.7
		1X	2.03 ± 0.09	23.8 ± 11.3	116.8 ± 17.5	95.1 ± 14.4	2.04 ± 0.06	19.8 ± 8.5	139.1 ± 12.2	99.6 ± 10.9
		1.25X	1.94 ± 0.07	36.8 ± 7.9	124.9 ± 12.8	76.1 ± 9.7	1.88 ± 0.07	31.2 ± 9.3	146.5 ± 14.1	82.0 ± 11.5
		2X	1.65 ± 0.11	47.2 ± 12.0	106.3 ± 16.8	66.8 ± 14.8	1.43 ± 0.09	45.3 ± 14.1	140.2 ± 20.0	76.9 ± 17.7
		0.8X	1.81 ± 0.08	23.2 ± 8.3	92.6 ± 11.3	79.8 ± 10.4	2.02 ± 0.05	13.3 ± 4.7	90.6 ± 6.1	80.4 ± 6.0
	10°	1X	1.45 ± 0.11	45.0 ± 14.0	116.9 ± 21.0	66.2 ± 16.9	1.64 ± 0.04	28.3 ± 4.3	85.6 ± 5.6	68.9 ± 5.4
		1.25X	0.78 ± 0.11	67.5 ± 22.1	107.8 ± 32.3	46.2 ± 26.2	1.18 ± 0.04	37.6 ± 4.4	57.1 ± 5.5	70.3 ± 5.8
		2X	0.32 ± 0.06	84.7 ± 9.1	116.3 ± 32.6	25.0 ± 11.4	0.27 ± 0.05	93.3 ± 10.5	92.4 ± 28.3	25.0 ± 26.3
		0.8X	2.33 ± 0.09	14.8 ± 9.4	103.8 ± 13.6	104.6 ± 12.4	2.44 ± 0.06	17.9 ± 7.4	145.8 ± 11.1	105.2 ± 9.7
		1X	2.21 ± 0.07	22.1 ± 6.7	102.1 ± 9.5	93.3 ± 8.7	2.30 ± 0.06	26.3 ± 6.4	139.9 ± 9.2	95.7 ± 8.4
GROUP	0° (fovea)	1.25X	2.10 ± 0.07	35.1 ± 6.9	105.1 ± 9.6	83.4 ± 8.8	2.11 ± 0.07	40.5 ± 8.5	137.1 ± 11.9	83.0 ± 10.8
		2X	1.70 ± 0.09	45.1 ± 9.9	94.8 ± 13.2	69.9 ± 12.4	1.81 ± 0.05	47.1 ± 6.3	125.9 ± 8.5	72.5 ± 7.9

Table 1. Summary of fitted parameters derived from the asymmetric Gaussian functions as shown in Figures 3 and 5. Parameters listed for “GROUP” were derived from the curves fitted to the data pooled across the three observers, instead of averages of the fitted values of the three observers.

SOA, proportion-correct	Letter exposure duration = 50 ms		Letter exposure duration = 100 ms	
	Fovea	10° eccentricity	Fovea	10° eccentricity
negative, 0.5	Insufficient data points	147.5 ± 13.2	399.9 ± 55.1	178.4 ± 31.5
negative, 0.6	Insufficient data points	153.2 ± 5.9	371.5 ± 65.5	190.2 ± 40.6
negative, 0.7	Insufficient data points	175.4 ± 17.9	387.0 ± 122.0	208.8 ± 22.9
negative, 0.8	490.2 ± 44.7	Insufficient data points	404.2 ± 24.4	255.6 ± 0.8
positive, 0.5	Insufficient data points	-33.2 ± 10.5	-67.1 ± 20.9	-32.1 ± 7.0
positive, 0.6	Insufficient data points	-39.0 ± 14.5	-67.1 ± 14.9	-31.7 ± 23.0
positive, 0.7	Insufficient data points	-60.7 ± 11.4	0 ± 0	-39.0 ± 14.5
positive, 0.8	-134.2 ± 77.4	-60.1 ± 10.5	-54.5 ± 19.6	-54.5 ± 9.9

Table 2. Summary of the slopes of the lines shown in Figure 6 (only for datasets with at least three datapoints). The slope of the lines refers to the variable m in the equation: $SOA = m(\log \text{ letter separation}) + \text{constant}$.

separation between a target and its flanking letters to yield a given level of identification performance. Clearly, the spatio-temporal window is not a constant. When the criterion is lower (e.g., proportion-correct of 0.5), observers could withstand a smaller spatial and/or temporal separation between the target and its flankers. On the contrary, when the performance criterion is higher (e.g., proportion-correct of 0.8), the target and its flankers need to be separated more in space and or time. Most importantly, for a given criterion level of performance, there is a trade-off between the spatial and temporal separations of the target and its flankers. When the spatial separation is small, the target and the flankers need to be separated more in time; but when the spatial separation is large, then the target and the flankers can tolerate a closer proximity in time. As an example, at the fovea, if the target and flanking letters are presented for 50 ms and are separated by a distance

equivalent to 0.8× or 1× the letter size, in order for an observer to recognize the target letters at an accuracy of 60% (red curves), the target and flanking letters need to be separated by 62.5 ms if the flankers appear after the target (positive SOA), or by ~2 ms if the flankers appear before the target (negative SOA). What if the flankers appear 30 ms after the target? In this case, the observer would not be able to recognize the target letters at 60% accuracy. At 10° eccentricity, for the same stimulus conditions and letter-recognition accuracy, the temporal separation between the target and flankers need to increase, such that the flankers need to appear 100 ms after, or between 62.5 and 75 ms before the target onset.

To capture the interaction of the spatial and temporal size of the crowding window, and as a function of criterion level of performance, we fit each set of data with at least three data points with a straight line (on semilog axes) in Figure 6. Table 2 summarizes the slopes of these lines. At the fovea and for the 50-ms letter presentation duration, because there is less crowding, letter identification performance exceeded 80% correct for the larger letter separations (see Figure 2). Thus in many cases, there are only two data points for each set of data and line-fit was not attempted. Because there are fewer slopes reported for the foveal conditions, we cannot meaningfully compare the slopes of these SOA versus letter separation lines between the fovea and 10° eccentricity. However, there are still several interesting observations. The slopes of the lines for the negative SOAs are generally steeper than those for the positive SOAs, implying that the change in the size of the temporal window with letter separations is faster when flankers preceded the target. The shallow slopes of the lines for the positive SOAs mean that the critical SOA is very similar for the four letter separations when flankers followed the target, for any given performance criterion. In addition, for a given combination of conditions (letter presentation duration × eccentricity), the four lines for the positive or the negative SOAs, corresponding to the four criterion levels, are essentially

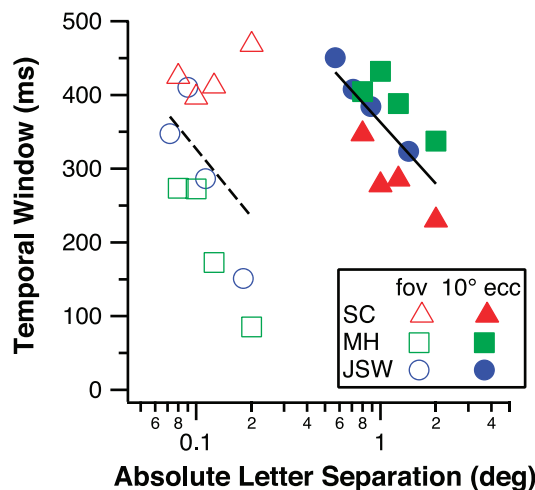


Figure 7. The size of the temporal window of crowding (ms) is plotted as a function of the absolute letter separation. Data shown represent the spatiotemporal limit beyond which crowding is not observed, for a letter exposure duration of 50 ms. Dashed and solid lines represent the best-fit line (on linear-log axes) to the foveal and 10° eccentricity data, respectively.

parallel to one another, meaning that the shape of the SOA-tuning functions for the four criterion levels shown in Figures 3 and 5 are essentially the same.

An interesting question is whether there is a law relating the spatial and the temporal size of the window of crowding. As an attempt to answer this question, we derived from each asymmetric Gaussian function shown in Figure 4 the positive and negative SOA that corresponded to a threshold elevation of 0.05 z -score unit. The sum of the absolute value of these SOAs represents the temporal window, for a given condition. Figure 7 plots the temporal window (ms) as a function of the absolute letter separation (nominal separation \times letter size, in degrees) at the fovea (unfilled symbols) and 10° eccentricity (filled symbols), for the three observers. Data for the fovea and 10° eccentricity are segregated into two clusters, suggesting that we could not derive one single law to relate the spatial and temporal size of the window of crowding based simply on the size of the spatial window. Nevertheless, the trade-off between the size of the temporal and spatial window is clear. Note that these data were obtained for a 50-ms letter exposure duration; a similar analysis for the 100-ms letter exposure duration showed a similar trend of the change in the size of the temporal window as a function of the absolute letter separation, with the temporal window shifted toward slightly larger values.

Discussion

Crowding has been traditionally studied as a spatial vision phenomenon. By measuring the recognition accuracy of a flanked letter when there is an onset asynchrony between the letter and its flankers, we showed, together with evidence from previous studies (Greenwood, Sayim, & Cavanaugh, 2014; Harrison & Bex, 2014; Huckauf & Heller, 2004; Ng & Westheimer, 2002), that crowding also shows a strong dependence on the temporal properties of the stimulus.

As summarized in the Introduction, previous studies have demonstrated an inverse relationship between the spatial critical spacing and stimulus duration such that the critical spacing is smaller for longer-duration targets and larger for shorter-duration targets (Chung & Mansfield, 2009; Kooi et al., 1994; Tripathy & Cavanagh, 2002; Wallace et al., 2013). In this study, when other stimulus conditions (eccentricity, SOA) were kept the same, we also observed a larger critical spacing for the 100-ms than the 50-ms stimulus durations (Figure 6). Additionally, our findings in relation to the spatio-temporal interactions on crowding are qualitatively consistent with previous findings that the magnitude of crowding depends strongly not only on the spatial separation, but also on the relative

timing between a target and its flankers. Most importantly, our findings provide several new insights into the spatio-temporal limitations on crowding. For example, although Ng and Westheimer (2002) already alluded to a temporal window within which flankers degrade the recognition of a target (flankers should be presented about 50 ms after the onset of the target for maximal crowding), and that Harrison and Bex (2014) quantified this temporal window as 45 ms; in these studies, the target and flankers physically coexisted for the entire presentation duration of either the target or the flankers, leading to an impression that maximal crowding requires the target and flankers to overlap for a substantial period of time. Our finding shows that this is not necessary. Maximal crowding could occur when the target and flankers never co-exist in time. For example, Figure 3 shows that for letter duration of 50 ms, maximal crowding occurred when flankers appeared after the target *offset* (SOA > 50 ms), for the larger letter separations (1.25 \times and 2 \times separations at the fovea and 2 \times separation at 10° eccentricity). A similar conclusion can be drawn from *some* of the figures in Huckauf and Heller (2004). However, because these authors sampled SOA in steps of 50 ms, and that they did not provide any error bars in their data points, it is difficult to derive with certainty whether the plotted SOA corresponding to the worst performance was indeed the SOA yielding maximal crowding. This observation was complicated by the fact that the authors tested positive and negative SOA trials with different groups of observers. In the present study, we improved the experimental paradigm of Huckauf and Heller (2004) in several ways (see Methods for details). As a result, we are able to conclude with certainty that there was a systematic shift in the SOA corresponding to maximal crowding as a function of spatial separation and or eccentricity. Also, Huckauf and Heller (2004) reported that the SOA corresponding to maximal crowding shifted from 50 ms at 1° eccentricity to 0 ms (target and flanker presented simultaneously) at 7° eccentricity, implying that the target and flankers need to completely overlap in time for maximal crowding to occur in the periphery. This finding was likely an artifact because the same letter size (0.51° in height) and letter spacing (1°) were used for all eccentricities (1° – 7°), thus the letter size with respect to the resolution limit, and the letter spacing with respect to the critical spacing, are not comparable across eccentricities. With letter size and spacing scaled in the periphery, in the current study we observed a shift in SOA corresponding to maximal crowding from small to large letter separations that was comparable between the fovea and 10° eccentricity.

The most interesting questions based on our results are what accounts for the maximal crowding at a given SOA, and why does the SOA that corresponds to

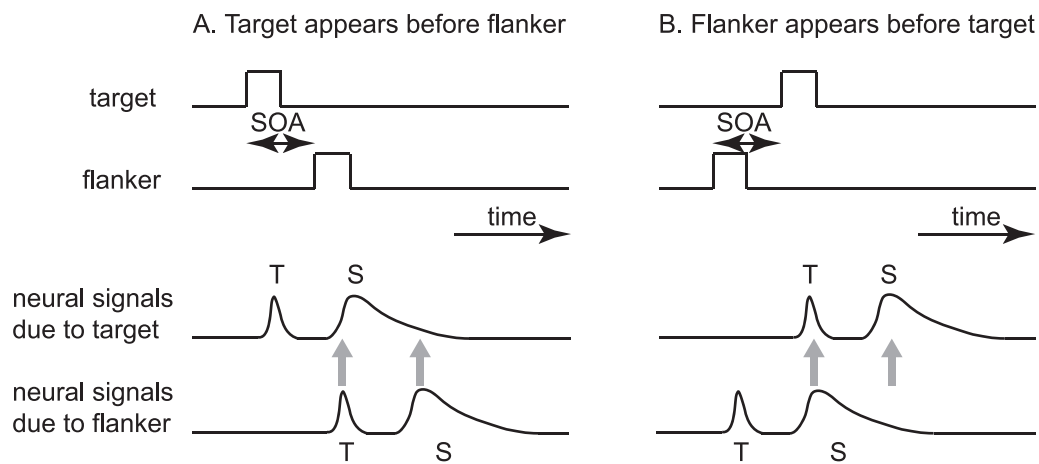


Figure 8. A schematic of how the dual-channel inhibition model can explain our data, for the scenarios when the target appears before a flanker (left: positive SOA) and when a flanker appears before the target (right: negative SOA). In each scenario, the top two traces represent the time-courses of the target and a flanking letter. The bottom two traces represent the time-courses of the neural signals generated by the target and the flanker. Each letter generates two signals—a shorter-latency transient one (T) and a longer-latency sustained one (S). (A) When the target appears before a flanker, if the flanker is offset from the target letter by an SOA such that the transient signal generated by the flanker coexists with the sustained signal arising from the target letter, the transient signal from the flanker can interfere with the sustained signal from the target letter. This is referred to as the interchannel inhibition. Depending on the SOA, in some cases, the sustained signal from the flanker may also coexist with the sustained signal from the target (intrachannel inhibition), thus causing interaction. The interference of the sustained signal of the target would primarily affect the identity information of the target. (B) When the flanker appears before the target, the transient signal from the flanker cannot interfere with the signals generated by the target, but the sustained signal from the flanker may interfere with the transient signal of the target (inter channel inhibition), affecting primarily the position information of the target. The sustained signal from the flanker may also interfere with the sustained signal of the target (intrachannel inhibition), affecting the identity information of the target.

maximal crowding shifts to a more positive value when the spatial separation of the target and flankers increases. Here, we offer several candidate explanations. The first explanation is based on a dual-channel inhibition model that involves the interactions of the sustained and transient channels (Breitmeyer & Ganz, 1976; Breitmeyer & Ögmen, 2000).⁴ According to this model, the onset of each stimulus component (the target or the flankers) gives rise to a transient and a sustained signal. The transient signal, responsible for signaling the *where* information of the target, has a shorter latency and the signal itself is short-lived (Ögmen, Breitmeyer, & Melvin, 2003). The sustained signal, responsible for providing the *what* information of the target, has a longer latency and the signal persists for a longer period of time. Thus, an inhibition of the transient or sustained signal could lead to an error in localizing or identifying an object, respectively. Specifically, in the case where flankers appear after the target, if the onset of the target and its flankers differs by a duration such that the transient signals of the flankers coexist with the sustained signal of the target, as shown in Figure 8A, then the transient signals of the flankers may inhibit the sustained signal of the target (interchannel inhibition). Further, because the sustained signal persists for a longer time, the sustained signals of the flankers may also interact with the

sustained signal of the target, causing intrachannel inhibition. The inhibition of the sustained channel (carrying the identity information) of the target makes it difficult for observers to correctly identify the target letters. This explanation is consistent with our observation that the SOA at which maximal crowding occurs shifts toward more positive SOAs when letter separation increases. A larger physical letter separation requires that the neural signals (sustained or transient) from the flankers travel over a slightly longer distance, thus, taking a longer time to reach the neural site where the inhibition of the sustained signal of the target occurs.

Conversely, when flankers precede the target, as in Figure 8B, depending on the SOA between flankers and the target, the sustained signals of the flankers may inhibit the transient signal of the target (interchannel inhibition), affecting primarily the position information of the target. However, it is also possible that the sustained signals of the flankers and the target coexist in time, thus causing intrachannel inhibition.

This dual-channel inhibition model is very popular in the visual masking literature (e.g., Breitmeyer & Ganz, 1976; Ögmen, 1993; Breitmeyer & Ögmen, 2000). In applying this model to explain our data, we are not assuming that the observed findings are completely due to masking; yet, we cannot rule out masking as a

contributing factor to our finding, nor can we separate out the relative contributions of visual masking and crowding on accounting for our data. However, recall that masking is a reduction of the *visibility* of the target by the mask (Breitmeyer & Ögmen, 2000) whereas crowding only affects the identification, but not the detection (Pelli et al., 2004), of a target. Because our letter stimuli were of high contrast, a reduction in the visibility of the target letters due to masking may not necessarily cause an identification error. In other words, although temporal masking may contribute to letter identification errors observed in this study, it is unlikely to be the sole factor accounting for the errors.

An implication of the previously mentioned explanation is that the transient and sustained signals arise in response to the onset, not the offset, of the target or flankers. At least for the transient signal, it can also arise in response to the offset of a stimulus. Therefore a logical question to ask is whether the crowding effect observed was due to the onset or offset of the target and or flankers. Because the target and flanker durations were the same in this study (either 50 or 100 ms), replotting our data with stimulus *offset* asynchrony instead of stimulus onset asynchrony on the abscissa would not change the pattern of results shown in Figures 2–5. To tease apart the contributions of the onset versus the offset of a stimulus (target or flankers) in causing errors in letter identification would require decoupling the target and flanker durations, a paradigm that was used in Harrison and Bex (2014) and Greenwood et al. (2014). Although the exact experimental paradigms differed between these two studies, both studies arrived at the conclusion that crowding is alleviated with an onset, not offset, transient of the target, implying that onset events are more important than offset events in limiting crowding.

The second explanation is based on the coarse-to-fine progression in visual processing. Substantial psychophysical and computational evidence suggests that the processing of a visual stimulus relies more on the coarse, or the low spatial-frequency information initially, and shifts its reliance toward the fine, or high spatial-frequency information with time (e.g., Marr & Poggio, 1979; Parker, Lishman, & Hughes, 1992; Watt, 1987). This coarse-to-fine phenomenon is supported by neurophysiological evidence showing that the size of receptive fields in the early visual cortex decreases during the time-course of processing (e.g., Malone, Kumar, & Ringach, 2007; Menz & Freeman, 2003). In relation to our findings, if the target and flankers were separated by a small SOA, then when the target is being analyzed by low spatial-frequency mechanisms during the early-phase of processing, there is a good chance that the flankers are also being analyzed by low spatial-frequency mechanisms. In this case, the target and flankers need to be separated by a spatial distance that

exceeds the extent of the receptive fields of these low spatial-frequency mechanisms to avoid crowding. If the target and flankers were separated by a large SOA, then the visual system should have shifted its reliance to higher spatial-frequency mechanisms to analyze the stimulus that was presented first, while relying on low spatial-frequency mechanisms to analyze the stimulus that appeared later. In this case, the target and flankers could tolerate a smaller spatial separation.

Data from the present study exhibit properties that are consistent with the predictions based on both the dual-channel inhibition model and the coarse-to-fine progression in visual processing hypothesis. To determine the exact mechanism that underlies spatio-temporal letter crowding would require additional data for modeling, for example, by decoupling the onset and offset of the target and flankers, or by examining the time course of crowding using stimuli that must be processed by either low or high spatial-frequency mechanisms. We are currently using both of these approaches to continue our quest of understanding the mechanism underlying spatio-temporal crowding.

To date, there are many proposed theories to account for conventional spatial crowding, i.e., when the target and flankers are presented simultaneously (for a review, refer to Levi, 2008, or Whitney & Levi, 2011). One popular theory is based on feature integration in which it is assumed that features of a target and its spatially proximal flankers fall within the combination field over which features are drawn and then combine to represent the final percept of the target. The theory further postulates that crowding is feature integration gone awry. This could be a consequence of, for instance, features of the target and flankers being displaced from their veridical spatial locations due to position uncertainty, especially in peripheral vision (Pelli, 1985), with some features from the flankers encroaching upon the location of the target (or vice versa), with the result that these displaced features of the flankers combine incorrectly with those of the target to form the percept of the target. Alternatively, excessive feature integration (Pelli et al., 2004) or the underutilization of some of the valid features of the target and flankers (Nandy & Tjan, 2007) might also lead to erroneous feature integration. Regardless of the exact mechanism underlying the process, erroneous feature integration may also explain our results. When features from the target and flankers are drawn to form the percept of the target, the process is unlikely to be an instantaneous one but instead, should last for some time. Therefore, as long as the signals for the target and flanker features coexist within the spatial and temporal limits for integration, features from the flankers can be combined with those of the target, causing errors in perception. This finding is consistent with the crowding effect observed with both

positive and negative values of SOA because this explanation only requires that the signals for the target and flanker features to coexist for a sufficient amount of time, but it does not matter whether the flanker features reach the combination field before the target features do, or vice versa. Also, it is the *signals* of the features that need to coexist within the integration window, but the target and flanker features do not need to physically coexist in space and time to produce the largest crowding effect.

Mislocation errors

Crowding may occur at multiple stages of visual processing (Whitney & Levi, 2011). Although our results are consistent with the erroneous feature integration theory of crowding, they are also consistent with an explanation of crowding at the symbol level, without invoking features of the letters. For instance, it is possible that the combination field may consist of a central excitatory region, surrounded by an inhibitory zone. If the two flankers fall within the inhibitory zone, they may affect the signal coming from the excitatory region encoding the target. Another potential explanation, based on the symbol level, is the mislocation of the flankers as the target. This explanation predicts that a high proportion of errors made in identifying the target letter should match one of the two flankers' identities. To determine if the mislocation of a flanker as the target can account for our results, we rescored our letter identification data using a different criterion—a letter was scored as being identified correctly if it matched any of the letters (target or flankers) presented in the trigram. The difference in the proportion-correct scored by matching only the target letter (the original method) versus matching any letters of the trigram represents the proportion of responses in which observers identified the flankers as the target. This difference, relative to the proportion of error trials (scored by the original method), represents the rate of mislocation errors (Chung & Legge, 2009). For example, if the proportion correct of letter identification was 0.56 when responses were only compared with the target letters and 0.8 when responses were compared with any letters of a trigram, then the mislocation error rate is $(0.8 - 0.56) / (1 - 0.56) = 0.55$. This means that 55% of the identification errors (for that given condition) was due to a mislocation. Figure 9 summarizes the rate of mislocation errors (data pooled across the three observers) as a function of SOA, for the four letter separations, the two testing eccentricities and the two letter exposure durations. Recall that the mislocation error rates plotted are the relative rates (relative to the proportion of error trials), two conditions yielding very different identification accu-

racies could yield the same mislocation error rate. In general, the pattern of how mislocation error rate changes with SOA is not too different across the four letter separations. The mislocation error rate is also not too different between positive and negative SOAs. An explanation of our data based solely on the interactions of sustained and transient signals generated by the target and flanking letters would predict a higher mislocation error rate for negative SOAs than for positive SOAs because in the case of negative SOAs (target following flankers), the transient signal arising from the target is likely to be inhibited by the transient or sustained signals of the flankers (see Figure 8). In contrast, when the SOAs are positive (flankers following target), the transient signal arising from the target is not inhibited by the neural signals arising from the flankers. Therefore, our finding is inconsistent with the notion that the identification errors (crowding effect) are due to the interactions of sustained and transient signals at the symbol level.

Temporal crowding versus temporal masking

A conventional way to study the temporal properties of interaction effects between two objects is to present the objects with an intervening temporal interval. This is the standard paradigm to study temporal masking. Here, we adopted the same paradigm to study crowding in the temporal domain. Considering that many of our observed findings are qualitatively similar with those reported in the masking literature, are the effects we observed simply a conventional (temporal) masking effect? Temporal masking has been traditionally referred to as a reduction in the *visibility* or the *detectability* of an object when it is closely followed, or preceded by the mask (Breitmeyer & Ögmen, 2000). Crowding does not affect detection, but only the identification or discrimination of a target (Pelli et al., 2004; Chung, 2010). The reduction in letter identification accuracy in this study certainly is consistent with a crowding effect, but we cannot rule out the contribution of masking (see the previous section). To directly evaluate the contribution of masking (or how much a reduction in stimulus visibility accounts for our results) or to compare whether or not temporal masking is the same as temporal crowding would require experimental paradigms specifically designed to test these ideas, for example, by including a detection task as in Pelli et al. (2004). This is outside the scope of the present study. Pelli et al. (2004) directly compared spatial crowding with lateral masking, by examining how the detection and identification of targets depend on different stimulus parameters such as target size and eccentricity, and arrived at several diagnostic tests to differentiate between spatial crowding and lateral masking. Future

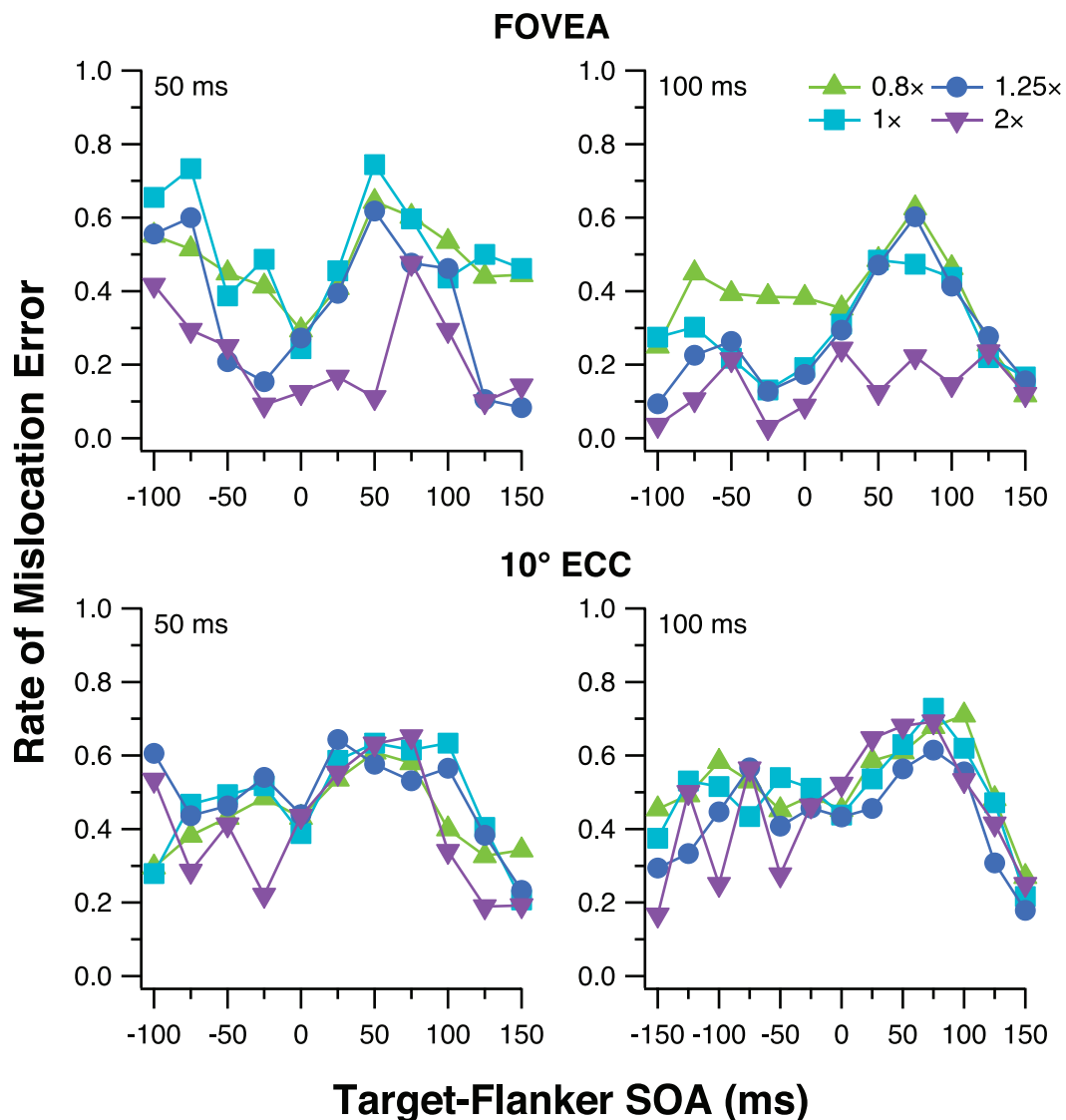


Figure 9. Rate of mislocation error (see text for definition) is plotted as a function of target-flanker SOA for the two letter exposure durations (left: 50 ms; right: 100 ms), for data obtained at the fovea (upper panels) and 10° inferior visual field (lower panels). In each panel, data are shown for the four letter separations. Data plotted are pooled across the three observers.

studies may consider a similar comparison to test if temporal crowding is simply a temporal masking effect.

Spatio-temporal window for crowding

The second goal of the study was to derive the critical size of the crowding region or window. Although the size or extent of the spatial crowding region has been studied extensively over the past 50 years or so (e.g., Bouma, 1970; Flom, Weymouth, & Kahneman, 1963; Toet & Levi, 1992), little is known about the size of the temporal crowding region, and more importantly, the interaction between the spatial and temporal extent. Here we provided empirical data showing the spatio-temporal extent of the crowding

window for a range of conditions. Not surprisingly, there is a trade-off between the spatial and temporal extent of the crowding window such that similar level of performance can be obtained for letters that are close in space but are well separated in time, or for letters that are presented close in time but are spatially well separated. This finding confirms our understanding of the spatio-temporal limitations on vision, but more importantly, it provides another means to minimize crowding. Crowding has been suggested as the bottleneck on object recognition and reading (Levi, 2008; Pelli & Tillman, 2008; Levi, Song, & Pelli, 2007), implying that our ability to recognize objects and to read faster should improve if crowding can be minimized. Previous attempts at minimizing crowding have targeted at reducing spatial crowding by increas-

ing the spatial distance between letters but with no significant improvement in reading speed (Chung, 2002; Pelli et al., 2007). The current finding presents a unique opportunity to reduce crowding without affecting the spatial relationship among letters. We are currently investigating whether reading can be improved by reducing crowding between letters with the introduction of an appropriate SOA between letters.

Keywords: crowding, spatial interaction, temporal interaction, stimulus onset asynchrony, mislocation errors, letter identification

Acknowledgments

This study was supported by research grant R01-EY012810 from NIH. The author thanks Dennis Levi, Daniel Coates, Girish Kumar, and Mehmet Agaoglu for helpful comments and suggestions on an earlier version of the manuscript.

Commercial relationships: none.

Corresponding author: Susana T. L. Chung.

Email: s.chung@berkeley.edu.

Address: School of Optometry and Vision Science Graduate Group, University of California, Berkeley, California.

Footnotes

¹ Also known by different terminologies such as the *isolation field* (Pelli et al., 2004) and *perceptive hypercolumn* (Levi et al., 1985).

² Although each observer responded to over 10,000 trials over the course of the experiment, the improvement in performance due to perceptual learning is likely to be small because perceptual learning is specific to testing conditions. In our experiment, the specific combination of testing condition (letter separation, letter exposure duration, and testing eccentricity) changed from block to block. In addition, a range of 11 or 13 target-flanker SOAs was tested within each block, with trials presented randomly. All of these are not conducive to perceptual learning.

³ We used a linear mixed-effects model for this analysis, using the *lmer* function of the *lme4* package in R (Bates, Maechler, Bolker, & Walker, 2015). The three main factors that entered into the model were: eccentricity (two levels: fovea, 10° eccentricity), nominal letter separation (four levels: 0.8×, 1×, 1.25×, and 2×) and target-flanker SOA (11 levels: –100 to 150 ms, in steps of 25 ms). The effective degrees of freedom were calculated using the Satterthwaite approximation.

The correction is necessary because the variances were pooled across several independent sample variances. A similar analysis was performed separately for the 100 ms data, with the only exception that there were 13 levels of SOAs (–150 to 150 ms, in steps of 25 ms).

⁴ Recently, Greenwood et al. (2014) used the sustained and transient channels to explain their data in which they found that a brief blink (turning off a stimulus for 20 ms) applied to the target, but not the flankers, alleviated crowding. However, there are several major differences between their model and the dual-channel inhibition model proposed here. In the present study, we assumed that the identity information is carried only by the sustained channel, whereas Greenwood et al. (2014) stated that “stimulus identification can occur somewhat independently in each (sustained or transient) channel” (page 17 in their paper). Second, according to the dual-channel inhibition model, the degradation in psychophysical performance is attributed to the inhibition (hence, the name of the model) effects between the signal from a target and the flanker either within the same sustained or transient channels, or between the sustained and transient channels. Greenwood et al.’s model attributes crowding to the coexistence of the target and flanker signals in the sustained and transient channels. A relief of crowding requires an isolation of the target, in their case, via a blink. However, an introduction of a blink should have introduced additional inhibitory signals to the transient channel according to the dual-channel inhibition model. Third, Greenwood et al.’s model assumes that the isolation of the target was via an attentional process while higher-level processes are not necessary in the dual-channel inhibition model.

References

- Banks, W. P., Bachrach, K. M., & Larson, D. W. (1977). The asymmetry of lateral inference in visual letter identification. *Perception & Psychophysics*, 22, 232–240.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48.
- Bernard, J.-B., & Chung, S. T. L. (2011). The dependence of crowding on flanker complexity and target-flanker similarity. *Journal of Vision*, 11(8):1, 1–16, doi:10.1167/11.8.1. [PubMed] [Article]
- Bi, T., Cai, P., Zhou, T., & Fang, F. (2009). The effect of crowding on orientation-selective adaptation in human early visual cortex. *Journal of Vision*, 9(11):13, 1–10, doi:10.1167/9.11.13. [PubMed] [Article]

- Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, 226, 177–178.
- Breitmeyer, B. G., & Ganz, L. (1976). Implications of sustained and transient channels for theories of visual pattern masking, saccadic suppression and information processing. *Psychological Review*, 83, 1–36.
- Breitmeyer, B. G., & Ögmen, H. (2000). Recent models and findings in visual backward masking: A comparison, review, and update. *Perception & Psychophysics*, 62, 1572–1595.
- Butler, T. W., & Westheimer, G. (1978). Interference with stereoscopic acuity: Spatial, temporal, and disparity tuning. *Vision Research*, 18, 1387–1392.
- Chung, S. T. L. (2002). The effect of letter spacing on reading speed in central and peripheral vision. *Investigative Ophthalmology & Visual Science*, 43, 1270–1276. [PubMed] [Article]
- Chung, S. T. L. (2010). Detection and identification of crowded mirror-image letters in normal peripheral vision. *Vision Research*, 50, 337–345.
- Chung, S. T. L. (2014). Size or spacing: which limits letter recognition in people with age-related macular degeneration? *Vision Research*, 101, 167–176.
- Chung, S. T. L., & Legge, G. E. (2009). Precision of position signals for letters. *Vision Research*, 49, 1948–1960.
- Chung, S. T. L., Levi, D. M., & Legge, G. E. (2001). Spatial-frequency and contrast properties of crowding. *Vision Research*, 41, 1833–1850.
- Chung, S. T. L., Li, R. W., & Levi, D. M. (2007). Crowding between first- and second-order letter stimuli in normal foveal and peripheral vision. *Journal of Vision*, 7(2):10, 1–13, doi:10.1167/7.2.10. [PubMed] [Article]
- Chung, S. T. L., & Mansfield, J. S. (2009). Contrast polarity differences reduce crowding but do not benefit reading performance in peripheral vision. *Vision Research*, 49, 2782–2789.
- Coates, D. R., Chin, J. M., & Chung, S. T. L. (2013). Factors affecting crowded acuity: Eccentricity and contrast. *Optometry & Vision Science*, 90, 628–638.
- Flom, M. C., Weymouth, F. W., & Kahneman, D. (1963). Visual resolution and contour interaction. *Journal of the Optical Society of America*, 53, 1026–1032.
- Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience*, 14, 1195–1201.
- Greenwood, J. A., Sayim, B., & Cavanagh, P. (2014). Crowding is reduced by onset transients in the target object (but not in the flankers). *Journal of Vision*, 14(6):2, 1–21, doi:10.1167/14.6.2. [PubMed] [Article]
- Harrison, W. J., & Bex, P. J. (2014). Integrating retinotopic features in spatiotopic coordinates. *Journal of Neuroscience*, 34, 7351–7360.
- He, S., Cavanagh, P., & Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature*, 383, 334–337.
- Huckauf, A., & Heller, D. (2004). On the relations between crowding and visual masking. *Perception & Psychophysics*, 66, 584–595.
- Kooi, F. L., Toet, A., Tripathy, S. P., & Levi, D. M. (1994). The effect of similarity and duration on spatial interaction in peripheral vision. *Spatial Vision*, 8, 255–279.
- Levi, D. M. (2008). Crowding—An essential bottleneck for object recognition: A mini-review. *Vision Research*, 48, 635–654.
- Levi, D. M., & Klein, S. A. (1985). Vernier acuity, crowding and amblyopia. *Vision Research*, 25, 979–991.
- Levi, D. M., Klein, S. A., & Aitsebaomo, A. P. (1985). Vernier acuity, crowding and cortical magnification. *Vision Research*, 25, 963–977.
- Levi, D. M., Song, S., & Pelli, D. G. (2007). Amblyopic reading is crowded. *Journal of Vision*, 7(2):21, 1–17, doi:10.1167/7.2.21. [PubMed] [Article]
- Louie, E. G., Bressler, D. W., & Whitney, D. (2007). Holistic crowding: Selective interference between configural representations of faces in crowded scenes. *Journal of Vision*, 7(2):24, 1–11, doi:10.1167/7.2.24. [PubMed] [Article]
- Malone, B. J., Kumar, V. R., & Ringach, D. L. (2007). Dynamics of receptive field size in primary visual cortex. *Journal of Neurophysiology*, 97, 407–414.
- Marr, D., & Poggio, T. (1979). A computational theory of human stereo vision. *Proceedings of the Royal Society of London B: Biological Sciences*, 204, 301–328.
- Martelli, M., Majaj, N. J., & Pelli, D. G. (2005). Are faces processed like words? A diagnostic test for recognition by parts. *Journal of Vision*, 5(1):6, 58–70, doi:10.1167/5.1.6. [PubMed] [Article]
- Menz, M. D., & Freeman, R. D. (2003). Stereoscopic depth processing in the visual cortex: A coarse-to-fine mechanism. *Nature Neuroscience*, 6, 59–65.
- Millin, R., Arman, A. C., Chung, S. T. L., & Tjan, B. S. (2014). Visual crowding in V1. *Cerebral Cortex*, 24, 3107–3115.
- Motter, B. C. (2006). Modulation of transient and

- sustained response components of V4 neurons by temporal crowding in flashed stimulus sequences. *Journal of Neuroscience*, 26, 9683–9694.
- Nandy, A. S., & Tjan, B. S. (2007). The nature of letter crowding as revealed by first- and second-order classification images. *Journal of Vision*, 7(2):5, 1–26, doi:10.1167/7.2.5. [PubMed] [Article]
- Ng, J., & Westheimer, G. (2002). Time course of masking in spatial resolution tasks. *Optometry and Vision Science*, 79, 98–102.
- Ögmen, H. (1993). A neural theory of retino-cortical dynamics. *Neural Networks*, 6, 245–273.
- Ögmen, H., Breitmeyer, B. G., & Melvin, R. (2003). The what and where in visual masking. *Vision Research*, 43, 1337–1350.
- Parker, D. M., Lishman, J. R., & Hughes, J. (1992). Temporal integration of spatially filtered visual images. *Perception*, 21, 147–160.
- Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, 4, 739–744.
- Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America, A*, 2, 1508–1532.
- Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision*, 4(12):12, 1136–1169, doi:10.1167/4.12.12. [PubMed] [Article]
- Pelli, D. G., & Tillman, K. A. (2008). The uncrowded window of object recognition. *Nature Neuroscience*, 11, 1129–1135.
- Pelli, D. G., Tillman, K. A., Freeman, J., Su, M., Berger, T. D., & Majaj, N. J. (2007). Crowding and eccentricity determine reading rate. *Journal of Vision*, 7(2):20, 1–36, doi:10.1167/7.2.20. [PubMed] [Article]
- Petrov, Y., Popple, A. V., & McKee, S. P. (2007). Crowding and surround suppression: Not to be confused. *Journal of Vision*, 7(2):12, 1–9, doi:10.1167/7.2.12. [PubMed] [Article]
- Pöder, E., & Wagemans, J. (2007). Crowding with conjunctions of simple features. *Journal of Vision*, 7(2):23, 1–12, doi:10.1167/7.2.23. [PubMed] [Article]
- R Development Core Team. (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from <http://www.R-project.org>
- Song, S., & Levi, D. M. (2010). Spatiotemporal mechanisms for simple image feature perception in normal and amblyopic vision. *Journal of Vision*, 10(13):21, 1–22, doi:10.1167/10.13.21. [PubMed] [Article]
- Song, S., Levi, D. M., & Pelli, D. G. (2014). A double dissociation of the acuity and crowding limits to letter identification, and the promise of improved visual screening. *Journal of Vision*, 14(5):3, 1–37, doi:10.1167/14.5.3. [PubMed] [Article]
- Strasburger, H. (2005). Unfocussed spatial attention underlies the crowding effect in indirect form vision. *Journal of Vision*, 5(11):8, 1024–1037, doi:10.1167/5.11.8. [PubMed] [Article]
- Toet, A., & Levi, D. M. (1992). The two-dimensional shape of spatial interaction zones in the parafovea. *Vision Research*, 32, 1349–1357.
- Townsend, J. T., Taylor, S. G., & Brown, D. R. (1971). Lateral masking for letters with unlimited viewing time. *Perception and Psychophysics*, 10, 375–378.
- Tripathy, S. P., & Cavanagh, P. (2002). The extent of crowding in peripheral vision does not scale with target size. *Vision Research*, 42, 2357–2369.
- Tripathy, S. P., Cavanagh, P., & Bedell, H. E. (2014). Large crowding zones in peripheral vision for briefly presented stimuli. *Journal of Vision*, 14(6):11, 1–11, doi:10.1167/14.6.11. [PubMed] [Article]
- Wallace, J. M., Chiu, M. K., Nandy, A. S., & Tjan, B. S. (2013). Crowding during restricted and free viewing. *Vision Research*, 84, 50–59.
- Wallace, J. M., & Tjan, B. S. (2011). Object crowding. *Journal of Vision*, 11(6):19, 1–17, doi:10.1167/11.6.19. [PubMed] [Article]
- Watt, R. J. (1987). Scanning from coarse to fine spatial scales in the human visual system. *Journal of the Optical Society of America A*, 4, 2006–2021.
- Westheimer, G., & Hauske, G. (1975). Temporal and spatial interference with vernier acuity. *Vision Research*, 15, 1137–1141.
- Westheimer, G., Shimamura, K., & McKee, S. P. (1976). Interference with line-orientation sensitivity. *Journal of the Optical Society of America*, 66, 332–338.
- Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*, 15, 160–168.