Vision Research 194 (2022) 108012

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Spatial and temporal proximity of objects for maximal crowding

Susana T.L. Chung [a,*], Saumil S. Patel [b]

[a] School of Optometry, University of California, Berkeley, Berkeley, CA, USA
[b] Department of Neuroscience, Baylor College of Medicine, Houston, TX, USA

ARTICLE INFO

Keywords:
Crowding
Flash-lag effect
Physical proximity
Perceptual proximity

ABSTRACT

Crowding refers to the deleterious visual interaction among nearby objects. Does maximal crowding occur when objects are closest to one another in space and time? We examined how crowding depends on the spatial and temporal proximity, retinally and perceptually, between a target and flankers. Our target was a briefly flashed T-stimulus presented at 10° right of fixation (3-o'clock position). It appeared at different target-onset-to-flanker asynchronies with respect to the instant when a pair of flanking Ts, revolving around the fixation target, reached the 3-o'clock position. Observers judged the orientation of the target-T (the crowding task), or its position relative to the revolving flankers (the flash-lag task). Performance was also measured in the absence of flanker motion: target and flankers were either presented simultaneously (closest retinal temporal proximity) with different angular spatial offsets, or were presented collinearly (closest retinal spatial proximity) with different temporal onset asynchronies. We found that neither retinal nor perceptual spatial or temporal proximity could account for when maximal crowding occurred. Simulations using a model based on feed-forward interactions between sustained and transient channels in static and motion pathways, taking into account the differential response latencies, can explain the crowding functions observed under various spatio-temporal conditions between the target and flankers.

1. Introduction

Our ability to recognize an object is often impaired when the object is in close proximity to other objects. This phenomenon is referred to as *crowding*. Crowding is ubiquitous and affects almost all visual tasks including clinical visual acuity measurement (e.g. Ehlers, 1953; Stuart & Burian, 1962; Flom, Heath & Takahashi, 1963; Flom, Weymouth & Kahneman, 1963), laboratory measurements of Vernier acuity (Levi & Klein, 1985; Levi, Klein & Aitsebaomo, 1985; Westheimer & Hauske, 1975) and orientation judgement of lines and Gabors (e.g. Levi & Carney, 2009; Parkes et al., 2001; van den Berg, Roerdink & Cornelissen, 2010; Westheimer, Shimamura & McKee, 1976). Everyday tasks such as reading (Chung, 2002; He & Legge, 2017; Pelli et al., 2007), face recognition (Kalpadakis-Smith, Goffaux & Greenwood, 2018; Louie, Bressler & Whitney, 2007; Martelli, Majaj & Pelli, 2005) and even driving (Xia et al., 2020) are also affected. As such, crowding is widely believed to be the bottleneck on object recognition, especially in peripheral vision (Levi, 2008; Pelli, 2008; Whitney & Levi, 2011).

The classical feature of crowding is that its effect increases with decreased separation between the object of interest (target) and its surrounding objects (flankers) (e.g. Bouma, 1970; Chung, Legge & Levi, 2001; Pelli, Palomares & Majaj, 2004; Toet & Levi, 1992). This implies that crowding should be maximal when the target and flankers are in closest spatial proximity, at least in peripheral vision where crowding is prominent (Levi, 2008; Pelli et al., 2004).[‡] In many cases, "closest spatial proximity" means that the physical positions of the target and flankers are closest together on the retina (we will also refer to the target-flanker configuration on the retina as the *physical* configuration), especially when the target and flankers are stationary and appear simultaneously. Under situations in which the perceived position of an object differs from its physical position, a question arises as to whether crowding is maximal when the target and flankers are closest in space physically or perceptually, or neither of these. Dakin, Greenwood, Carlson and Bex (2011), and Maus, Fischer and Whitney (2011) sought to answer this

* Corresponding author.
E-mail address: s.chung@berkeley.edu (S.T.L. Chung).

[‡] At the fovea, performance for identifying a target is often worst (spatial interference is strongest) not when the target-flanker separation is the smallest; but when the target is separated from the flankers by a finite (still small) distance (e.g. Flom, Heath & Takahashi, 1963; Flom, Weymouth & Kahneman, 1963; Westheimer & Hauske, 1975). This effect is often attributed to cues such as luminance cue when the target and flankers are spatially closest to one another. Apparently this cue is not effective in peripheral vision.

https://doi.org/10.1016/j.visres.2022.108012
Received 15 June 2021; Received in revised form 14 November 2021; Accepted 6 January 2022
Available online 15 January 2022

question by exploiting the motion-induced position shift illusion in which the position of a Gabor is perceived to be shifted in the direction of motion of its carrier grating, despite the Gaussian window being physically stationary (DeValois & DeValois, 1991; Chung, Patel, Bedell & Yilmaz, 2007). Both studies showed that crowding was stronger when the carrier of the flanking Gabors drifted toward the target, such that the perceived target-flanker separation was smaller than the physical separation. These studies demonstrated that crowding depends not only on the positions of objects on the retina, but also correlates with their perceived positions. Chambers, Johnston and Roach (2018) used a different approach to induce a difference between the physical and perceived separation between a crowded target and its flankers. They adapted their observers using an annular array of dynamic dots. After adaptation, observers perceived a compression of space. However, the magnitude of the crowding effect showed no relationship with the perceived target-flanker separation, which presumably was reduced because of the adaptation-induced spatial compression. These authors explained their finding that crowding did not depend on the perceived target-flanker separation as a dissociation between the adaption effect on spatial appearance and discriminability. Alternatively, the inconsistent results of whether perceived target-flanker separation induced by motion or adaptation affects crowding could be due to the different stages of processing with respect to the different tasks. It is also possible that their finding was limited by the weak adaptation effects coupled with the large individual variability (two of their four subjects did not show any adaptation effect).

Crowding has been studied extensively as a spatial phenomenon. Considering that vision has both spatial and temporal limitations, and that visual signals experience a range of latencies as they travel from the retina to various areas in the brain (e.g. Maunsell & Gibson, 1992; Schmolesky et al., 1998), it is reasonable to expect that crowding is also affected by temporal properties of the target and flankers. Based on what we observed in spatial crowding, one might ask whether maximal crowding occurs when the target and flankers are in closest proximity in time, i.e., when the target and flankers coexist in time. A follow-up question would be whether the "closest proximity in time" refers to the temporal proximity on the retina or the temporal proximity at higher stages in the visual processing hierarchy. Several studies have examined how crowding depends on the relative timing of target and flankers. Ng and Westheimer (2002) measured visual acuity using a Landolt C target presented for 150 ms, with flankers (four bars) presented for 50 ms. By varying the onset asynchronies between the target and flankers such that the flankers could appear before the target, during various phases of the target presentation, or after the target, crowding was shown to be maximal when the flankers appeared between 50 and 100 ms after the target onset. Harrison and Bex (2014) also evaluated the effect of onset (as well as offset) asynchronies between target and flankers by presenting the target for a fixed duration of 58 ms and the flankers for durations between 75 and 508 ms so that the target was always accompanied by the flankers, but the flankers appeared for an additional amount of time before the target appeared, or after the target disappeared. Their findings showed that the critical spacing for crowding was larger when the flankers remained on the screen after the target disappeared, and smaller when the flankers preceded the target appearance. In both the studies of Ng and Westheimer (2002) and Harrison and Bex (2014), because the target and flankers always overlapped in time by approximately 50 to 58 ms, these studies could not address the question of whether maximal crowding requires that the target and flankers coexist in time. To address this question, Chung (2016) measured crowding using target and flankers that were presented for 50 ms with a range of target-flanker onset asynchronies such that the flankers could appear up to 100 ms before the target, or 150 ms after the target. The main finding was that at small target-flanker spatial separations, maximal crowding occurred when the target and flankers coexisted; but at larger target-flanker spatial separations, maximal crowding occurred when the target and flankers did not overlap in time (or with minimal overlapping). These findings provide evidence that closest temporal proximity is not a universal requirement for maximal crowding.

The studies cited above provide compelling evidence that crowding depends strongly on the relative timing of target and flankers. However, these timings refer to the presence of static target and flankers on the retina. The implication of Chung (2016) that closest temporal proximity is not a requirement for maximal crowding might therefore, refer only to the temporal proximity of static target and flankers on the retina. It is conceivable that maximal crowding occurs when the target and flankers are closest in temporal proximity at other post-retinal stages of visual processing. Further, as shown by Chung (2016), the temporal proximity of target and flankers in determining the amount of crowding depends on the spatial separation between the target and flankers. In other words, crowding represents a spatio-temporal, instead of simply a spatial, bottleneck for object recognition.

The primary goal of this study was to examine how crowding depends on the spatial and temporal properties of target and flankers. We were particularly interested in studying how the perceived timing of target and flankers relate to the crowding effect. To do so, we used a target-flanker configuration that induced the flash-lag effect. The flash-lag effect refers to the phenomenon wherein a continuously moving object is perceived to be spatially ahead of a briefly flashed stationary object when the retinal images of both objects are physically aligned (see MacKay, 1958; Nijhawan, 1994). In this study, we measured crowding and the flash-lag misalignment separately using a briefly flashed stationary target and a pair of continuously moving objects as flankers (*motion conditions*). When the target and flankers were aligned on the retina, according to the flash-lag effect, the flanker positions should be perceived as ahead of the target position. By manipulating the time of appearance of the briefly flashed target with respect to the time when the flankers reached a position that was physically collinear with the target, we could systematically examine how crowding depends on the physical and perceived spatio-temporal relationship between the target and flankers. For comparison, crowding was also measured for a conventional configuration in which the target and flankers were all static and appeared simultaneously for the same duration. They were presented either with different spatial offsets that corresponded to those between the static target and the moving flankers in the motion conditions (*static spatial condition*), or with different temporal onset asynchronies when the target and flankers were collinear (*static temporal condition*). A secondary goal was to develop a mechanistic model of crowding observed in this study. The results of the experiments are interpreted and discussed in the context of this model which is based on differential latencies in sustained and transient channels in non-motion (static) and motion pathways. Predecessors of this model have been used to explain visual processing of moving and static targets (Öğmen, Patel, Bedell & Camuz, 2004), visual masking (Öğmen, 1993; Öğmen, Breitmeyer & Melvin, 2003) and temporal crowding (Chung, 2016). According to this model, the onset of each stimulus (target or flanker) generates signals in the sustained and transient channels, and crowding arises when the sustained signals from the target, which presumably code for the identity and location of the target, are inhibited by the sustained and/or transient signals from the flanker. Because the latency in the sustained channel can be different from that in the transient channel, and because latency in static pathways can be different from that in motion pathways, the inhibition, or the crowding effect, would depend on the relative timing of the onset of target and flankers and the pathways utilized in the task.

2. Methods

To examine how crowding depends on the physical or perceived spatial and temporal separations between a target and its flankers, observers made two types of judgment (orientation or position) about the target in several conditions as described below. In the main experiment,

flankers were moving whereas in the control experiments they were stationary. Because a robust crowding effect has been observed in a target orientation identification task (e.g. Bex, Dakin, & Simmers, 2003; Coates, Chin, & Chung, 2013; Coates & Chung, 2016; Harrison & Bex, 2014; Kooi, Toet, Tripathy, & Levi, 1994; Ng & Westheimer, 2002; Toet & Levi, 1992), we used this task to measure the crowding effect in this study. Further, a robust position misalignment between retinal and perceptual representations has been observed in a flash-lag task (Nijhawan, 1994; Purushothaman et al., 1998), therefore we used this task to forge a dissociation of retinal proximity from perceived proximity in both the spatial and temporal domains.

2.1. Observers

Four observers, including one of the authors (SC), participated in the study. SC is an experienced psychophysical observer while the other three, all naïve to the purpose of this study, had participated in one or two psychophysical studies prior to the present study. All observers had normal or corrected-to-normal vision (at least 20/20 in each eye) and no known ocular anomalies. Oral and written informed consent were obtained from each observer after the procedures of the study were explained and before data collection commenced. The research was approved by the Institutional Review Board at the University of California, Berkeley, and was conducted in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

2.2. Apparatus

Experimental procedures were controlled using software written in Matlab (The MathWorks, MA) with the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). Stimuli were generated by the Matlab script and were presented on a Mitsubishi color graphics monitor (model# N0701, Diamond Plus 73) with a resolution of 640 × 480 pixels (31.3 × 23.5 cm) at a refresh rate of 90 Hz. The phosphor was P22, with a decay time under 1 ms.

2.3. Main Experiment: Motion conditions

In the main experiment, stimuli comprised a static target T that was briefly presented at 10° eccentricity to the right of a fixation target (we shall refer to this as the 3-o'clock position), and two flankers (also Ts) that revolved around the fixation target (Fig. 1). The orientation of each of these Ts was randomly chosen from four possible options: up, down, right or left, and was independent of one another. When the target T appeared on the display, it might not be spatially aligned with the two flankers; instead, the two moving flankers might be physically ahead of (leading), or behind (trailing) the target on the display. Depending on the target-flanker separation (1.4° or 2°, center-to-center),[§] the two flankers were either 8.6° and 11.4° from fixation, or 8° and 12° from fixation. Each T-stimulus was made up of two bars, with the bar-width being one-fifth the height of the T, which was 1.1°, large enough for observers to identify its orientation at approximately an accuracy of 80–90% when presented alone at 10° eccentricity. The luminance of the target and flanker Ts measured 112 cd/m^2, whereas that of the background was 0.12 cd/m^2.

Testing was performed binocularly in a dimly-lit room. On each trial, observers first fixated at the center of a small cross at a viewing distance of 40 cm. After a delay of 1 s, the pair of flankers appeared on the right of fixation and started revolving around the fixation target in a clockwise direction at an angular velocity of either 18.75 or 37.5 revolutions per minute (rpm) for 1.5 revolutions. After approximately one full revolution and when the pair of flankers was close to the 3-o'clock position, the target T appeared for two video frames (22 ms) at the 3-o'clock position. We manipulated the timing of the onset of the target T with respect to the instant at which the flankers reached the 3-o'clock position — the target-onset-to-flanker asynchrony. Negative (positive) values of target-onset-to-flanker asynchrony mean that the target T appeared at the 3-o'clock position before (after) the flankers reached there, or, the flankers were above (below) the target T on the display (see Fig. 1A). Although the target-onset-to-flanker asynchrony refers to the timing between the appearance of the target and the flankers reaching the 3-o'clock position, at the time when the target appeared, the temporal offset was also exhibited as a spatial offset between the target and flankers that was based on the velocity of the flankers.

Using this stimulus configuration, observers performed two different judgments in separate blocks of trials. In each block of trials, we used the method of constant stimuli to present eight (for velocity of 37.5 rpm) or nine (for velocity of 18.75 rpm) target-onset-to-flanker asynchronies with each asynchrony repeated ten times. For the *orientation judgment* task, which allowed us to evaluate the crowding effect, observers identified the orientation of the target T (up, down, right or left) using the four arrow keys of a keyboard. For the *position judgment* task, which allowed us to evaluate the flash-lag effect, observers judged whether the target T was perceived above (spatially lagging) or below (spatially leading) the two flankers. Each observer completed at least three blocks for each condition. Data were pooled across the different blocks of trials for the same observer for analyses.

Because the flash-lag effect depends on the detectability of the flashed and moving objects (Purushothaman et al., 1998), we determined the detectability of the target T and the moving flanking Ts by finding the neutral density filters that rendered the target T or the moving Ts to disappear. Across observers, the target T was 3.4 to 3.6 log units above detection and the moving Ts were 4.4 to 4.8 log units above detection at 18.75 rpm and 4.4 to 4.6 log units above detection at 37.5 rpm. The lower detectability of the flashed target relative to the moving flankers ensured that we would observe a robust flash-lag effect — the flashed target T perceived as lagging behind the moving flankers — when they were physically collinear.

2.4. Control Experiments: Static conditions

In an attempt to understand the contribution of flanker motion on the perceived spatial and temporal offset between the target and flankers in the orientation and position judgement tasks, we tested our observers with two additional conditions in which there was no stimulus motion, i.e., the target and flankers were presented as static stimuli. In the *static spatial* condition, the target T and the flankers were in closest temporal proximity on the retina, i.e. they appeared on the display as an ensemble for two video frames (22 ms), but were presented with a range of angular spatial offsets that resembled those between the target and the flankers in the main experiment when the target appeared (Fig. 1B). This is akin to taking a snapshot of the instant when the target appeared in the main experiment, with the flankers frozen in their motion tracks. With the range of angular spatial offsets (–20 to 20 angular degrees (adeg) in steps of 5 adeg), the flankers could appear above, aligned with, or below the target T. Like in the main experiment, in separate blocks of trials, observers either identified the orientation of the target T or judged the relative position of the target with respect to the flankers. In the *static temporal* condition, the target T and the flankers were in closest spatial proximity on the retina, but were presented at a range of temporal onset asynchronies between them. Specifically, the target T and the flankers were aligned along the horizontal meridian to the right of fixation (Fig. 1C). Both the target and the flankers appeared for two video frames (22 ms), but the target onset could occur before (negative temporal

[§] The stated target-flanker separations of 1.4° and 2° referred to the center-to-center separation between the target T and one of the flankers when the three Ts were collinear, i.e. with zero spatial offset. Because we measured performance for a range of angular spatial offsets between the target and flankers, the actual separations between the target and its flankers could be larger than the nominal values of 1.4° or 2°.

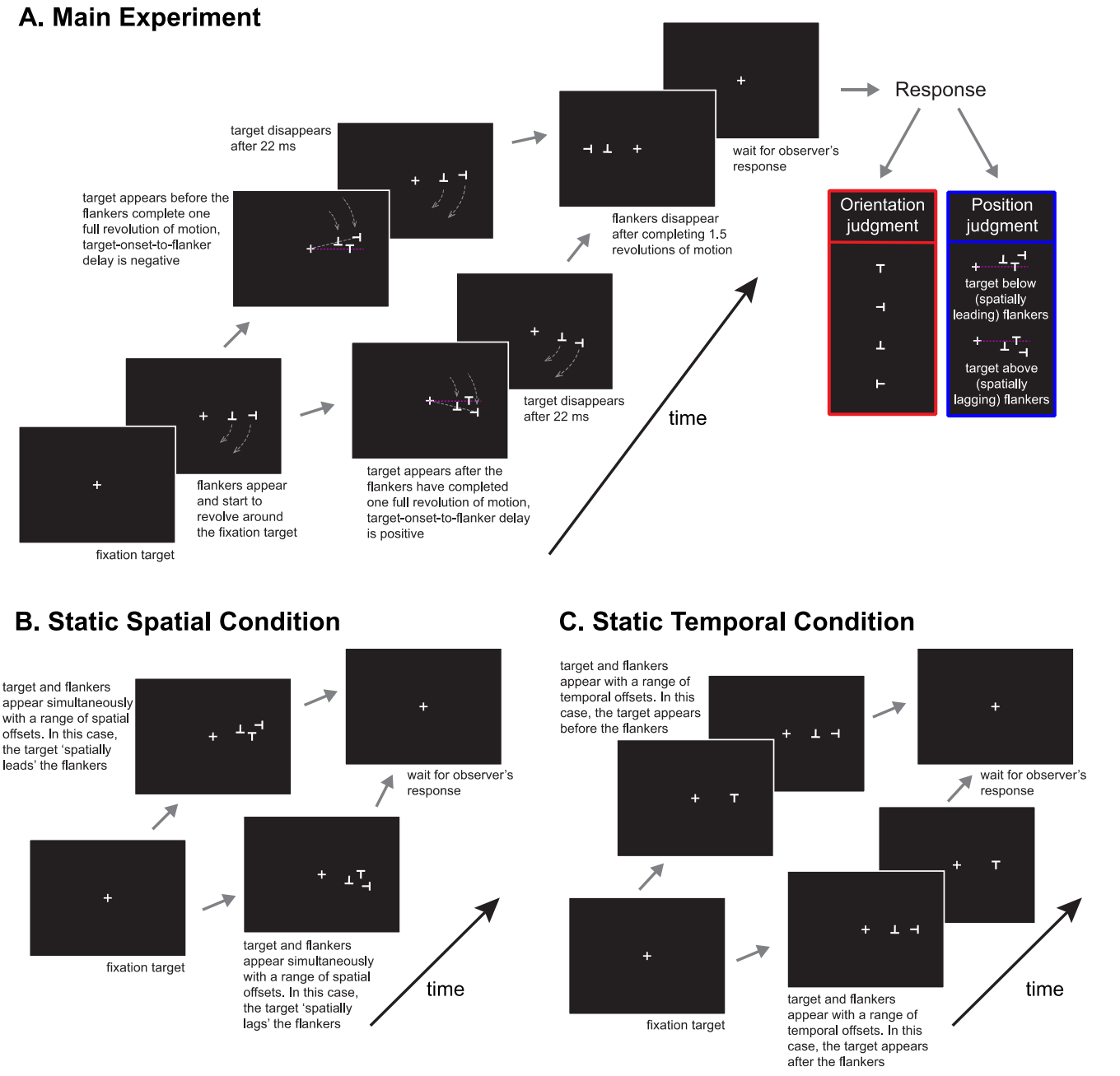


Fig. 1. A schematic figure depicting trials in (A) the main experiment, (B) the static spatial condition and (C) the static temporal condition. For each of these conditions, the spatial/temporal offset between the target and flankers is shown for both directions — target presented ahead (in space or time) of, or after, the flankers. In all these examples, the target T is drawn as an upright T and the two flankers include an upside-down T and one that is rotated to the right. In the actual experiment, the orientation of each of these stimuli was randomly rendered in one of four options: upright (top), upside-down (bottom), right or left.

onset asynchrony: up to 20 frames (222 ms) in steps of 4 frames (44 ms)), simultaneously with, or after (positive temporal onset asynchrony: up to 12 frames (133 ms) in steps of 4 frames (44 ms)) the flanker onset.[**] Observers' task was to identify the orientation of the target T. The luminance of target and flankers in both static conditions remained identical to that used in the main experiment and this allowed comparisons of modeling results between static and motion conditions.

2.5. Modeling

A model was simulated in Matlab Simulink (The MathWorks, MA) to obtain a simulated crowding effect as a function of target-onset-to-flanker asynchrony, flanker velocity and target-flanker separation. The focus of our modeling effort was to gain insights into how the data obtained in our present experiments could be explained with minimal parametric changes to a previously proposed model (Chung, 2016), in which crowding was modeled as a feed-forward inhibitory interaction from sustained and transient sub-units responding to the flanker upon a sustained sub-unit's response to the target (Fig. 2, dashed lines represent inhibition and solid lines represent excitation; neuronal dynamics defined in Appendix A1 (Grossberg, 1972)). A pair of sustained and transient sub-units formed a unit in the model. The units were arranged

[**] The notations of positive and negative temporal onset asynchronies in the static temporal condition are consistent with the definitions of positive and negative target-onset-to-flanker asynchronies in the main (motion) experiment, but note that they are in the opposite directions as defined in Chung (2016).

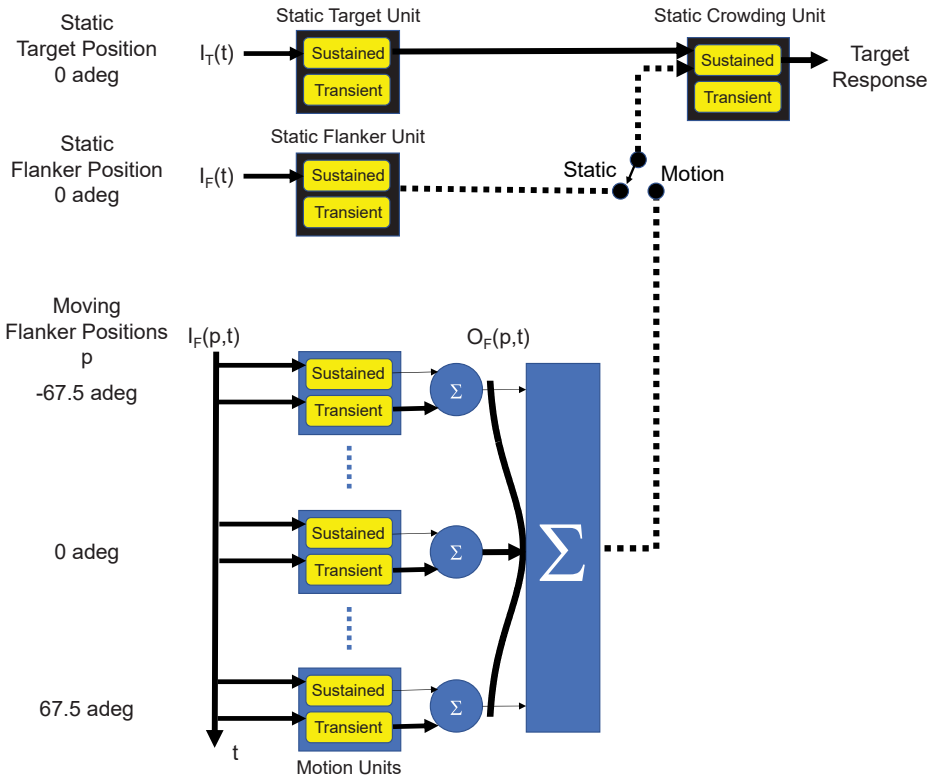


Fig. 2. A schematic figure of the model used to simulate observers' performance for identifying the orientation of the target T for the static or motion conditions. At the core of the model are units that respond to the visual targets at retinotopically specific locations, where 0 angular degree (adeg) corresponds to the 3-o'clock location. Each of these units comprises separate sub-units that generate either a sustained or a transient response. The premise of the model is that crowding (a reduction in orientation identification performance) arises as a result of the feed-forward inhibitory interaction from the output of the flanker's sustained and/or transient sub-units on the output of the target's sustained sub-unit (Chung, 2016). Static units are represented in black and motion units are represented in blue. For the motion conditions, our simulation includes 15 motion units sampling equally a spatial range of ± 67.5 adeg. Excitatory (inhibitory) connections are represented by solid (dashed) arrows. The switch is in the static (motion) position for simulations of static (motion) experimental conditions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

retinotopically in the model and the inhibitory interaction was assumed to occur in a unit upstream (the crowding unit) from where the above mentioned signals related to the target and flanker were generated. For the purpose of our model, we made no presumption about the exact neural site of the crowding unit and we did not focus strictly on the spatial aspects of crowding. We also did not intend to imply that crowding is mediated by a small number of units in the brain. The model was designed to capture the essential temporal properties of neural populations mediating crowding. Each sub-unit in the model had a pure delay (see parameters *e_delay* and *i_delay* in the model of sustained and transient sub-units in Appendices A1 and A2) to represent the transmission delay of signals from the source (retina or neighboring neuron in retinotopic map). A single unit was used to obtain the simulated target's response to a static stimulus ($I_T(t)$, where t represents time). The spatial location of the target was assumed to be at 0 adeg, i.e. 3-o'clock position (same as in the experiment). The path along which motion occurred was simulated by a set of units (presently 15 locations × 2 sub-units) sampling equally over a spatial range of ± 67.5 adeg. This spatial sampling corresponded to a temporal range of ± 0.6 and ± 0.3 s for the moving flanker stimulus ($I_F(p,t)$, where p represents the angular spatial position along the motion path) with velocities of 18.75 and 37.5 rpm, respectively. The middle location in the set of units (i.e. 0 adeg) represented the 3-o'clock position. In the simulated *motion conditions*, the net inhibitory signal sent to the crowding unit was computed as a weighted sum of outputs from the flanker's sustained and transient sub-units ($O_F(p,t)$). The weighting function qualitatively mimicked a normalized version of the crowding function in the *static spatial* condition with peak inhibition arising from the unit corresponding to 0 adeg (illustrated as a black Gaussian curve in Fig. 2). Initially, we included in the model an inhibitory signal such that the output from a stimulated flanker unit inhibited the previously stimulated neighboring units. The purpose of this backward inhibition signal was to reduce the extent of persistent excitation in units encoding previous positions in the motion trajectory. Because the simulation results showed that there were virtually no effects of adding this backward inhibition (see Discussion), this backward inhibition was subsequently removed (and not shown in Fig. 2) and all the simulation results reported in this paper are based on the model without the backward inhibition. The target's response from the crowding unit (for present experiments, between 25 and 75 ms after response onset) was integrated and used as a direct proxy for the identification performance of the target in the *orientation judgement* task. A larger integrated signal corresponds to a higher identification performance. In the simulated static conditions (spatial or temporal), in accordance with the separation of static and moving stimulus processing (Patel et al., 2000; Öğmen, Patel, Bedell & Camuz, 2004; Chung et al., 2007), inhibitory signals sent to the crowding unit were computed from the output of a static flanker unit at the 0 adeg position. The difference in the latency of response between the target's sustained sub-unit and the moving flanker's sustained sub-unit corresponding to 0 adeg was a proxy for the performance in the *position judgement* task. To model the effect of target-flanker separations used in the experiment, parameters of transient sub-units in motion units were adjusted (see parameters in Appendix A2). The equations used for the computation of signals within the sustained and transient units are given in Appendix A1. The parameters of the model were chosen manually to fit the qualitative aspects of the empirical data. The *e_delay* parameter of the sustained sub-unit in the set of motion units was fixed for each condition as the difference between the flash-lag effect and the *e_delay* parameter of the sustained sub-unit of the target unit. Hence, the flash-lag effect was trivially explained in the model.

2.6. Statistical method

Mixed model analyses were performed using SPSS 18 (IBM, NY). Several models were used based on the dependent variables and fixed factors that were tested. All models used autoregressive repeated covariance structure with observer as the random variable and task (orientation judgment, position judgment), target-flanker separation (1.4°, 2.0°) and flanker velocity (18.75 rpm, 37.5 rpm) as fixed factors. When significant interactions between factors were found, separate

models were used for each of the interacting factors. Bonferroni adjustment of confidence intervals was used to account for multiple comparisons. We have reported all significant effects but non-significant effects are only reported in certain cases.

3. Results

3.1. Motion Conditions: Orientation judgment task

Observers' performance for the orientation judgment task is presented in the top row of Figs. 3 and 4 (red symbols), for flanker velocities of 37.5 rpm and 18.75 rpm, respectively. Position judgment results are summarized in the bottom row in these figures (blue symbols), and will be discussed in the next section. In the top row of both figures, each column presents the data of one observer, in which the proportion-correct for identifying the orientation of the target T is plotted as a function of target-onset-to-flanker asynchrony, for target-flanker separations of 1.4° (unfilled circles) and 2° (filled circles). Across all observers and for both flanker velocities, performance for identifying the target orientation shows a V-shaped function with the target-onset-to-flanker asynchrony. Performance is best when the target onset was either well ahead of the flankers reaching the 3-o'clock position or well after the flankers passed the 3-o'clock position on the display, and is worst when the target was presented approximately 20 to 100 ms before the flankers reached the 3-o'clock position on the display. Because the target was presented for only 22 ms, it also means that worst performance, or maximal crowding, occurred when the target and flankers were not spatially aligned on the retina. To quantify the crowding effect (the reduction in performance when the flankers were close to the target), we fit each set of data using the following equation, which represents a split-Gaussian function, as used in Chung (2016):

$$f(t) = \begin{cases} 1 + A_p \times exp\left(-\left(\frac{t - t_{max}}{\sigma_L} \right)^2 \right) & if\ t < t_{max} \\ 1 + A_p \times exp\left(-\left(\frac{t - t_{max}}{\sigma_R} \right)^2 \right) & if\ t \geq t_{max} \end{cases} \quad (1)$$

where t represents the target-onset-to-flanker asynchrony, A_p is the peak amplitude of the Gaussian function (maximal magnitude of crowding), t_{max} is the target-onset-to-flanker asynchrony 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. Based on the fitted function, we compared three parameters across different conditions: t_{max}, A_p and the temporal extent of crowding, σ_E, defined as the sum of σ_R and σ_L.

A full factorial mixed model analysis using t_{max} as the dependent variable with target-flanker separation and flanker velocity as fixed factors found a significant effect of target-flanker separation ($F[1,12] = 16.4$, $p = 0.002$). Specifically, t_{max} shifted toward a more negative value for the larger target-flanker separation (from −34.1 ms to −68.2 ms), meaning that the target needed to be presented even earlier before the flankers reached the 3-o'clock position on the display (mean change in $t_{max} = 34.1$ ms, CI = [25.1, 43.0]). As for the effect of flanker velocity, it was only marginally significant ($F[1,9.4] = 4.5$, $p = 0.06$). When flanker velocity increased, t_{max} shifted toward a less negative value (from −56.9

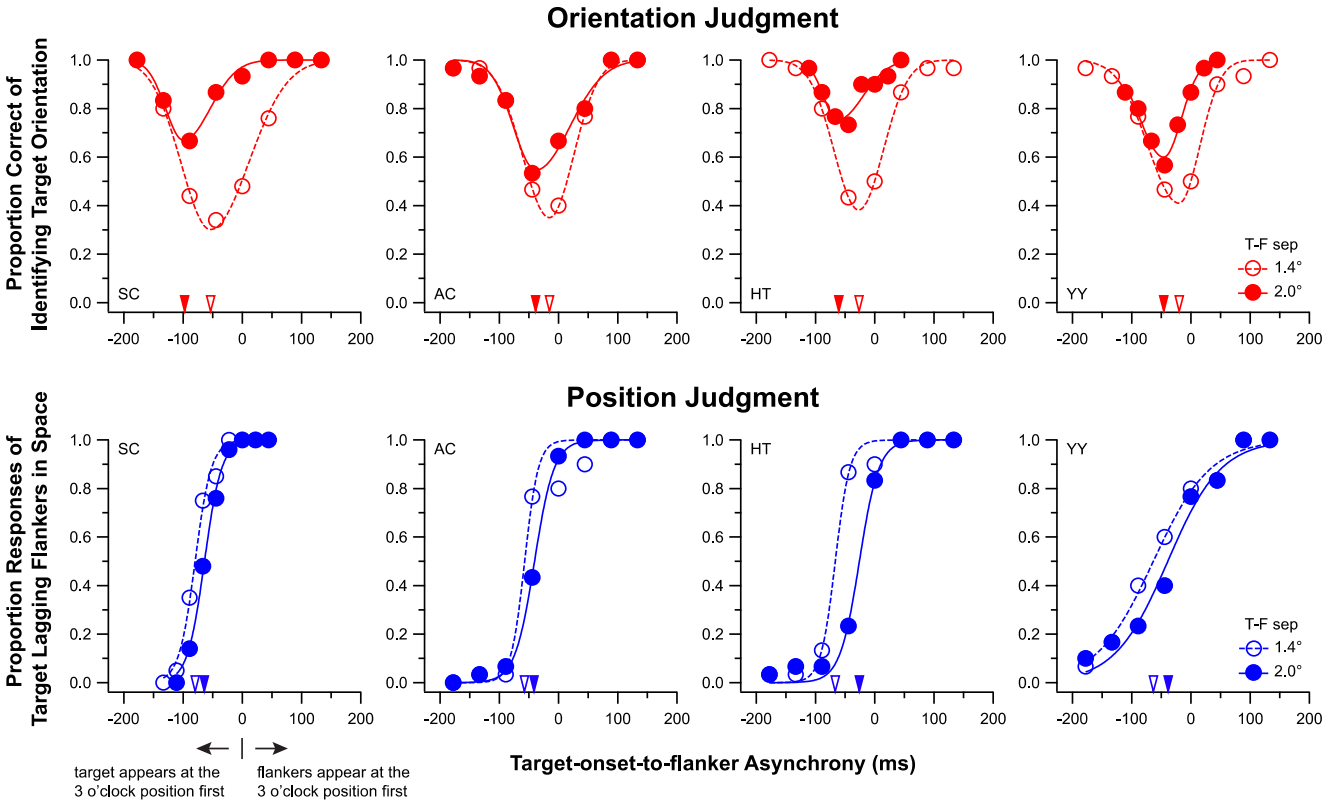


Fig. 3. Observer's performance is plotted as a function of target-onset-to-flanker asynchrony for the orientation (top) and position (bottom) judgment tasks. Data were obtained when flankers revolved around the fixation target at a velocity of 37.5 rpm. For the orientation judgment task (top row), proportion correct of identifying the orientation of the target T is represented on the ordinate. For the position judgment task (bottom row), the proportion responses of target spatially lagging the flankers (target perceived above the flankers, Fig. 1A) are plotted on the ordinate. Each column shows the performance of one observer. In each panel, data are summarized for the two target-flanker separations (T-F sep) of 1.4° (unfilled symbols) and 2° (filled symbols). Smooth curve drawn through each set of data represents the best-fit function for the task (see main text for details). Downward pointing triangles plotted on the abscissa in each panel represent the target-onset-to-flanker asynchronies corresponding to maximal crowding for the orientation task (t_{max}) or the time-point when the target was perceived to be spatially collinear with the flankers (t_{half}).

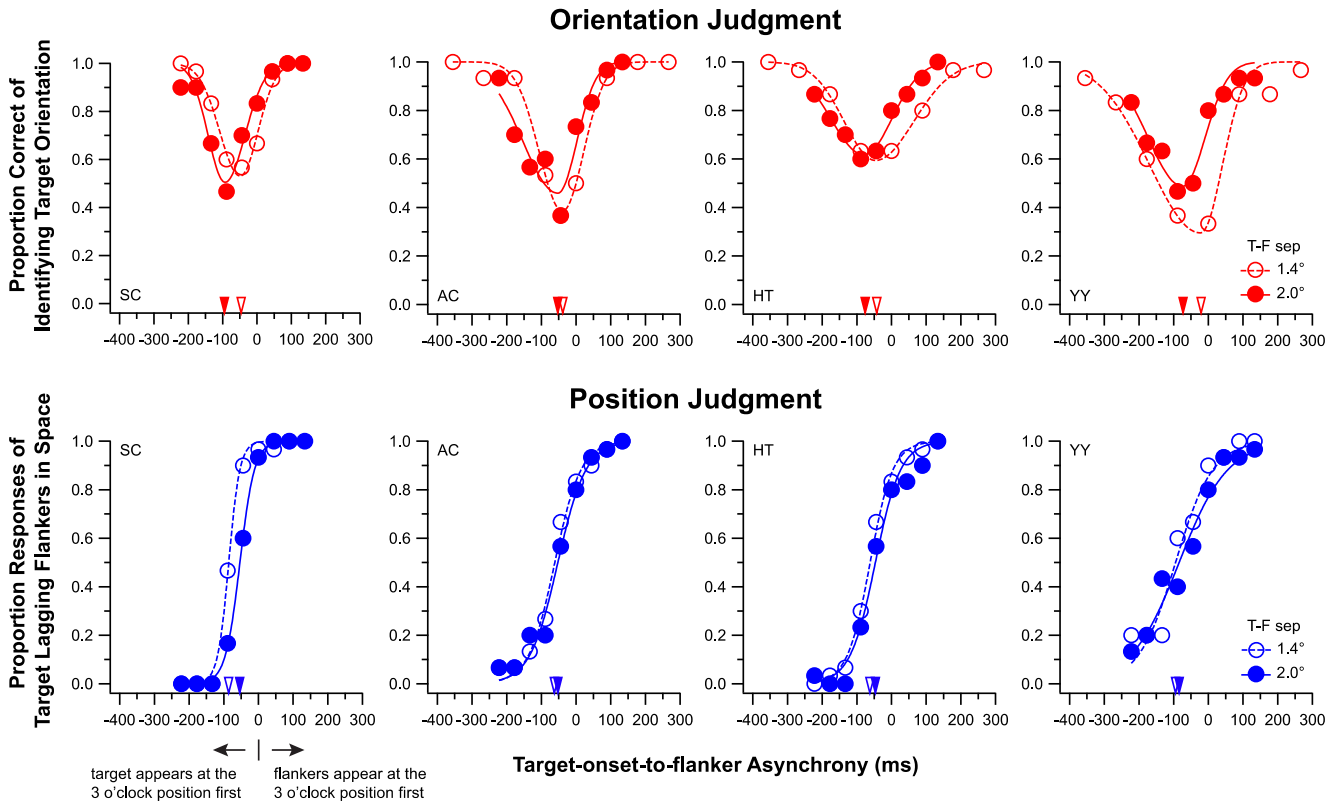


Fig. 4. Observer's performance is plotted as a function of target-onset-to-flanker asynchrony for the orientation (top row) and position (bottom row) judgment tasks. Data were obtained when flankers revolved around the fixation target at a velocity of 18.75 rpm. Details of this figure are the same as those of Fig. 3.

ms to −45.4 ms, mean change = 11.5 ms, CI = [2.5, 20.5]). Interactions between flanker velocity and target-flanker separation were not significant (F[1,8.5] = 0.06, p = 0.8), implying that there was no evidence that the amount of change in t_{max} with target-flanker separation was different between the two flanker velocities. Importantly, t_{max} being clearly different from zero for both target-flanker separations suggests that physical temporal proximity cannot account for when maximal crowding occurs.

Recall that the temporal offset between the target and flankers was also exhibited as a spatial offset between them that was based on the velocity of the flankers. Therefore, we ran the same mixed model with spatial offset at maximal crowding (t_{max}* flanker velocity in adeg/ms) as the dependent variable. There were significant effects of target-flanker separation (F[1,11.8] = 8.9, p = 0.01) and flanker velocity (F[1,7.8] = 7.24, p = 0.03). As target-flanker separation increased, the spatial offset became more negative (from −5.5 adeg to −11.1 adeg, mean change = 5.7 adeg, CI = [3.8, 7.5]). Similarly, as velocity increased, the spatial offset became more negative (from −6.4 adeg to −10.2 adeg, mean change = 3.8 adeg, CI = [1.0, 6.6]). The interactions between flanker separation and velocity were not significant (F[1,9.3] = 1.09, p = 0.32).

As expected based on previous studies (e.g. Bouma, 1970; Chung, Legge & Levi, 2001; Pelli et al., 2004), the magnitude of crowding, A_p, became smaller (from 0.60 to 0.42) when the target-flanker separation increased from 1.4° to 2° (Figs. 3 and 4). A full factorial mixed model analysis using A_p as the dependent variable with target-flanker separation and flanker velocity as fixed factors confirmed a significant effect of target-flanker separation (mean change in A_p = 0.18, CI = [0.075, 0.28]; F[1,13.6] = 11.3, p = 0.005) but no evidence for an effect of flanker velocity (F[1,6.5] = 0.66, p = 0.45). The interactions between target-flanker separation and flanker velocity were significant (F[1,9.5] = 8.4, p = 0.02). Specifically, the reduction in the magnitude of crowding with increase in target-flanker separation was significantly larger for flanker velocity of 37.5 rpm than for 18.75 rpm.

A full factorial mixed model analysis using the temporal extent of crowding (σ_E) as the dependent variable with target-flanker separation and flanker velocity as fixed factors revealed a significant effect of flanker velocity (F[1,4.5] = 14.7, p = 0.02). As flanker velocity increased from 18.75 rpm to 37.5 rpm, σ_E reduced from 221.4 ms to 119.7 ms (mean reduction = 101.8 ms (CI = [56.5, 147.0]). There was no evidence for an effect of target-flanker separation (F[1,9.6] = 0.97, p = 0.35) or any interactions between flanker velocity and target-flanker separation (F[1,11.6] = 0.03, p = 0.87). When the same mixed model was run with the extent of crowding converted into its spatial equivalent (σ_E*flanker velocity in adeg/ms) as the dependent variable, there were no significant effects found. In other words, there was no evidence that the spatial extent of crowding depended on target-flanker separation (F[1,8.9] = 2.54, p = 0.15) or flanker velocity (F[1,4.0] = 0.27, p = 0.63).

3.2. *Motion Conditions: Position judgment task*

The proportion of responses of the target spatially lagging the flankers (target perceived above the flankers) is plotted as a function of target-onset-to-flanker asynchrony in the bottom row of panels in Fig. 3 (flanker velocity = 37.5 rpm) and 4 (flanker velocity = 18.75 rpm). Data are plotted for the two target-flanker separations (1.4°: unfilled blue symbols; 2°: filled blue symbols). When the target was presented at the 3-o'clock position way ahead of the flankers reaching there (negative value on the abscissa), observers rarely reported that the target lagged the flankers. With the target-onset-to-flanker asynchrony becoming less negative, observers' responses of the target lagging the flankers increased. To quantify the target-onset-to-flanker asynchrony at which the target was perceived as collinear with the flankers, we fit each set of data with a logistic function of the form:

$$f(t) = \frac{1}{1 + \exp(-k(t - t_{half}))} \quad (2)$$

where t refers to the target-onset-to-flanker asynchrony, k represents the steepness of the function and t_{half} the target-onset-to-flanker asynchrony corresponding to the midpoint of the function, or, the point of subjective equality when the target was perceived as collinear with the flankers. We used t_{half} to represent the flash-lag effect.

A full factorial mixed model analysis using t_{half} as the dependent variable and target-flanker separation and flanker velocity as fixed factors revealed a significant effect of target-flanker separation ($F[1,11.5] = 7.05$, $p = 0.02$). When the separation increased from 1.4° to 2°, t_{half} shifted toward a less negative value, from −71.2 ms to −50.8 ms. The mean reduction in t_{half} was 20.5 ms (CI = [12.8, 28.2]). There was a non-significant trend that t_{half} became less negative as flanker velocity increased (from −67.5 ms to −54.5 ms, mean difference = 13.0 ms, CI = [0.11, 25.8]; $F[1,7] = 4.6$, $p = 0.07$). There was no evidence for interactions between target-flanker separation and flanker velocity ($F[1,9.4] = 0.34$, $p = 0.58$). When the same mixed model was run with the flash lag effect converted into its spatial equivalent (t_{half} *flanker velocity in adeg/ms) as the dependent variable, there were significant effects of target-flanker separation ($F[1,11.4] = 8.47$, $p = 0.01$) and flanker velocity ($F[1,6.9] = 20.92$, $p < 0.01$). As target-flanker separation increased, the spatial flash lag became less negative (from −11.8 adeg to −8.1 adeg, mean change = 3.7 adeg, CI = [1.8, 5.5]). As velocity increased, the spatial flash lag became more negative (from −7.6 adeg to −12.3 adeg, mean change = 4.7 adeg, CI = [2.2, 7.1]). The interactions between flanker separation and velocity were not significant ($F[1,9.7] = 2.6$, $p = 0.14$).

3.3. Motion Conditions: Orientation vs. Position judgment tasks

Considering that we used the same stimulus configuration for the orientation and position judgment tasks, are t_{max} (orientation judgment) and t_{half} (position judgment) similarly affected by our stimulus factors? Figs. 3 and 4 show that t_{max} and t_{half} shifted toward opposite directions when the target-flanker separation increased from 1.4° to 2°. This finding is more clearly illustrated in Fig. 5A. Specifically, t_{max} shifted toward a *more negative* value, implying that for maximal crowding to occur, the target needed to appear even earlier before the flankers reached the 3-o'clock position for the larger target-flanker separation. In contrast, t_{half} shifted toward a *less negative* value when the target-flanker separation was larger, implying that the flash-lag magnitude was smaller, or that the perceived position of flankers became closer to the physical one. As for the effect of flanker velocity, both t_{max} and t_{half} were affected similarly by flanker velocity with the effect being only marginally significant: both shifted toward a less negative value as flanker velocity increased (mean difference: t_{max} = 11.5 ms vs. t_{half} = 13.0 ms; Fig. 5B).

A mixed model analysis was performed to directly compare t_{max} and t_{half} in the two motion (18.75 and 37.5 rpm) conditions. The dependent variable in this analysis was t_{diff} ($t_{max} - t_{half}$). Target-flanker separation had a significant effect on t_{diff} ($F[1,8.2] = 58.2$, $p < 0.001$). Interestingly, and as described earlier (Fig. 5A), t_{diff} changed from being positive to being negative when target-flanker separation increased (mean t_{diff} = 37.1 ms, CI = [25.1, 49.1] for 1.4° separation; mean t_{diff} = −17.4 ms, CI = [−3.1, −31.7] for 2.0° separation). The results remain the same if t_{diff} is expressed in spatial units ($s_{diff} = s_{max} - s_{half}$) by taking into account the corresponding flanker velocities ($F[1,6.8] = 41.7$, $p < 0.001$; mean s_{diff} = 6.3 adeg, CI = [4.1, 8.5] for 1.4° separation; mean s_{diff} = −3.0 adeg,

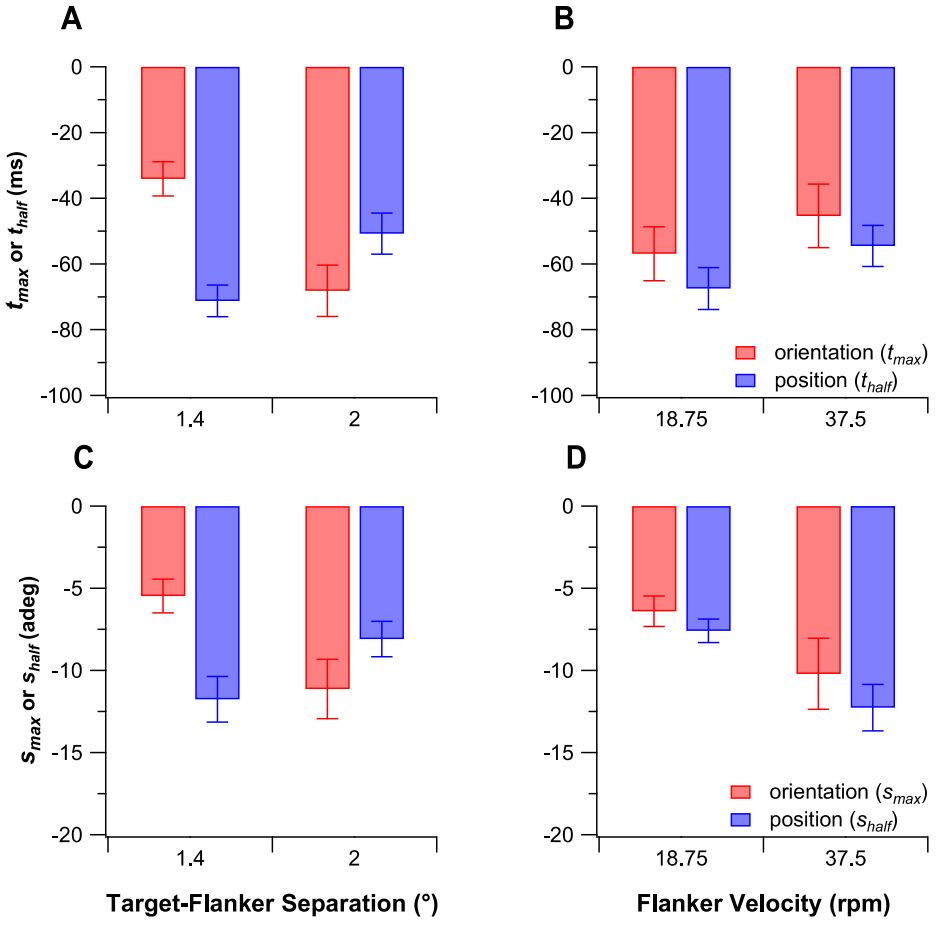


Fig. 5. (A) Group-average values of t_{max} for orientation judgment (red) and t_{half} for position judgment (blue) are plotted for the two target-flanker separations. Values plotted are averaged across the two flanker velocities. (B) Group-average values of t_{max} for orientation judgment (red) and t_{half} for position judgment (blue) are plotted for the two flanker velocities. Values plotted are averaged across the two target-flanker separations. (C & D) t_{max} and t_{half}, converted to their spatial equivalents (multiplying the temporal units by the flanker velocity), s_{max} and s_{half}, are plotted for the two target-flanker separations in C and for the two flanker velocities in D. Error bars represent the ± 1 SE. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

CI = [0.48, −5.6] for 2.0° separation). There was no evidence that flanker velocity had an effect on t_{diff} or s_{diff}. Additionally, the magnitudes of t_{diff} and s_{diff} are significantly different from zero (t_{diff}: t(15) = 6.1; mean magnitude = 28.8 ms, CI = [19.6, 38.0]; s_{diff}: t(15) = 5.6; mean magnitude = 4.9 adeg, CI = [3.2, 6.6]). The implication of the mean magnitudes of t_{diff} and s_{diff} being different from 0 and that both changed sign as a function of target-flanker separation is that *perceptual* proximity in space or time cannot account for when maximal crowding occurs.

3.4. Control experiments

Orientation and position judgment performance for the static spatial and temporal conditions are summarized together with data from the motion conditions in Fig. 6. Data plotted are average values of the four observers.

3.4.1. Crowding in motion vs. Static spatial conditions

Several mixed model analyses were performed to compare the fitted parameters of the crowding functions between the motion and the static spatial conditions. In the static spatial conditions, because the target and flankers were presented simultaneously, technically speaking, we could not compare t_{max} (the target-onset-to-flanker asynchrony at which crowding magnitude was maximal) and σ_E (the temporal extent of crowding), although we could still compare A_p (the maximal magnitude of crowding). Therefore, for these comparisons, we converted t_{max} and σ_E in the motion conditions for flanker velocity of 18.75 and 37.5 rpm to their spatial units, s_{max} (the spatial separation between the target and flankers when crowding was maximal, in adeg, Fig. 6B) and the spatial extent of crowding (Fig. 6E). With regard to s_{max}, there was a significant difference between each of the motion conditions and the spatial condition (18.75 rpm vs. spatial: t(12.9) = −4.4, p = 0.001; 37.5 rpm vs. spatial: t(17.9) = −6.43, p < 0.001). s_{max} was more negative in both motion conditions (−6.4 adeg for 18.75 rpm and −10.2 adeg for 37.5 rpm) compared with the spatial condition (0.6 adeg), yielding a mean difference of 7.0 adeg (CI = [5.2, 8.8]) for 18.75 rpm and 10.8 adeg (CI = [6.6,15.1]) for 37.5 rpm. This suggests that *physical* spatial proximity is not a sufficient necessity for maximal crowding. The spatial extent of crowding also differed significantly between each of the motion conditions and the spatial condition (18.75 rpm vs. spatial: t(10.4) = 3.3, p = 0.008; 37.5 rpm vs. spatial: t(18) = 3.7, p = 0.002). Specifically, the spatial extent of crowding was larger in both motion conditions compared with the spatial condition (mean difference: 12.6 adeg, CI = [8.3, 16.8] for 18.75 rpm; 14.6 adeg, CI = [11.2, 18.0] for 37.5 rpm). However, there was no evidence that the amplitude of crowding (A_p) differed between the motion conditions and the spatial condition for both flanker velocities (Fig. 6C). These findings clearly show that crowding in the motion conditions is different from that in the static spatial condition, consistent with the processing of moving and static flankers by separate pathways.

3.4.2. Crowding in motion vs. Static temporal conditions

Similarly, several mixed model analyses were performed to compare t_{max}, σ_E and A_p of the crowding functions obtained for the motion and the static temporal conditions. t_{max} refers to the target-onset-to-flanker asynchrony that yielded maximal crowding for the motion condition or the temporal onset asynchrony between the target and flanker that yielded maximal crowding for the static temporal condition (Fig. 6A). σ_E and A_p refer to the temporal extent of crowding (Fig. 6D) and the maximal magnitude of crowding (Fig. 6C), respectively, for both the motion and static temporal conditions. There was no evidence that t_{max} differed between the temporal and motion conditions for either flanker velocity. Consistent with the finding in Section 3.1, there was a significant main effect of target-flanker separation on t_{max} (F[1,11.4] = 17.5, p = 0.001). σ_E differed significantly between the motion and the temporal condition only for flanker velocity of 37.5 rpm (t(18) = −2.6, p = 0.02). Specifically, σ_E was smaller in the 37.5 rpm motion condition compared with the temporal condition (mean difference: 81.0 ms, CI = [52.6, 109.3]). Also consistent with the finding in Section 3.1, target-flanker separation did not have an effect on σ_E. There was no evidence that the amplitude of crowding (A_p) differed between the temporal and motion conditions for either flanker velocity. Consistent with our report in Section 3.1, A_p differed significantly between the two target-flanker separations (F[1,18.5] = 17.0, p = 0.001). The finding that σ_E is different between the static temporal and flanker velocity of 37.5 rpm conditions adds support to the notion that crowding due to moving or static flankers occur via separate pathways.

3.4.3. Crowding in motion conditions: Temporal vs spatial units

Fig. 6D shows that the temporal extent of crowding was significantly different between the two flanker velocities. However, when converted to their corresponding spatial units, the extent became largely independent of velocity (Fig. 6E). This indicates that the crowding zone for a static target with moving flankers may be a fixed spatial zone (i.e. does not vary with flanker velocity) but a more systematic study is necessary to confirm this observation. Note that the spatial crowding zone for a static target with moving flankers is clearly different from that with static flankers, once again supporting the notion that moving and static flankers are processed by separate pathways.

3.4.4. Flash misalignment in motion vs. Static spatial conditions

A mixed model analysis was performed to compare t_{half} (the target-onset-to-flanker asynchrony at which the target was perceived as collinear with the flankers) in the motion and the static spatial conditions. For the comparison, t_{half} in the motion conditions (18.75 and 37.5 rpm) were converted to their spatial units (s_{half}: Fig. 6G). For the static spatial condition, s_{half} essentially represents the bias of alignment judgment for static targets (mean = 0.11 adeg, CI = [−0.5, 0.7]), which is very close to zero. Therefore, it is not surprising that s_{half} differed significantly between each of the motion conditions and the static spatial condition (18.75 rpm vs. spatial: t(11.9) = −6.2, p < 0.001; 37.5 rpm vs. spatial: t(16.4) = −9.7, p < 0.001). Flash misalignment magnitudes were greater and in the lagging direction in both motion conditions (−7.6 adeg for 18.75 rpm and −12.3 adeg for 37.5 rpm), compared with the spatial condition (0.11 adeg), yielding a mean difference of 7.7 adeg, CI = [6.0, 9.4] for 18.75 rpm and 12.4 adeg, CI = [9.5, 15.2] for 37.5 rpm. These findings confirm that the flash misalignment observed during the motion conditions cannot be attributable to observers' bias of misalignment of static stimuli.

3.5. Modeling results

Simulated responses were obtained from the model shown in Fig. 2 under three conditions (static temporal, motion at 18.75 rpm and motion at 37.5 rpm), for target-flanker separations of 1.4° and 2°, and for target-onset-to-flanker asynchronies ranging from −600 to 600 ms, in steps of 1 ms. Fig. 7 shows examples of the time sequences of responses to the target and flankers following their onset at the 3-o'clock (0 adeg) position, and the output from the static crowding unit, for a target-flanker separation of 1.4° and three different target-onset-to-flanker asynchronies. Similar plots for a target-flanker separation of 2° are given in Appendix A3. The three target-onset-to-flanker asynchronies were chosen such that the middle one corresponded to the asynchrony that yielded maximal crowding, along with a longer and a shorter asynchrony. In each set of time sequences, the top row shows the onset of the target at the 3-o'clock position, which is set at 0 s. Timings for all subsequent events are aligned with respect to the onset of the target. The second row shows the onset of the flankers at the 3-o'clock position, which is shifted relative to that of the target by the target-onset-to-flanker asynchrony. The third row shows the sustained response to the target after a delay of *e_delay* (as shown in Appendix A2, we used 100 ms as the *e_delay* for the static target unit in these simulations). The fourth row shows the transient response to the flankers after a delay. For the

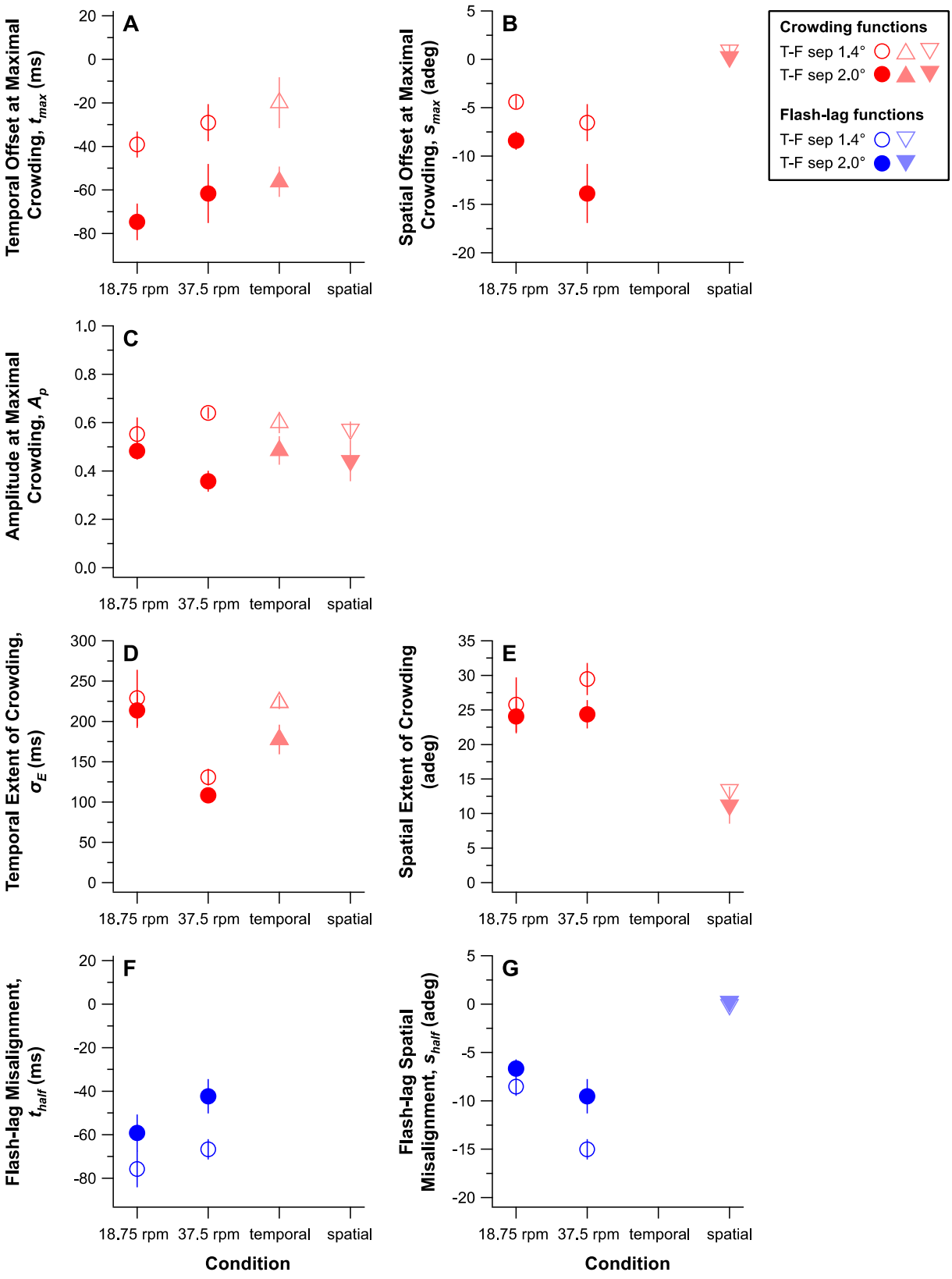
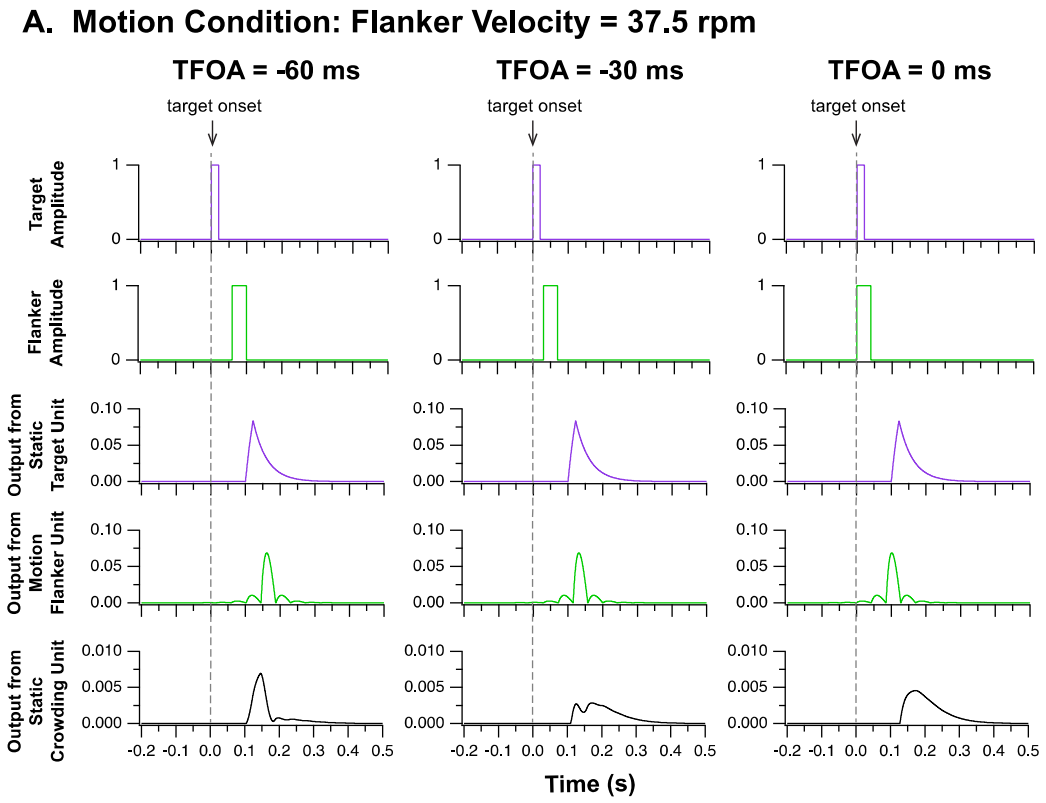


Fig. 6. Summary of the fitted parameters based on the best-fit split-Gaussian functions (crowding judgment) or psychometric functions (position judgment) for the two motion conditions, the static temporal and the static spatial conditions. Except for the amplitude that corresponds to maximal crowding (panel C), other parameters are plotted in time units (ms) in the left column for the motion and static temporal conditions, and in spatial units (adeg) in the right column for the motion and the static spatial conditions.



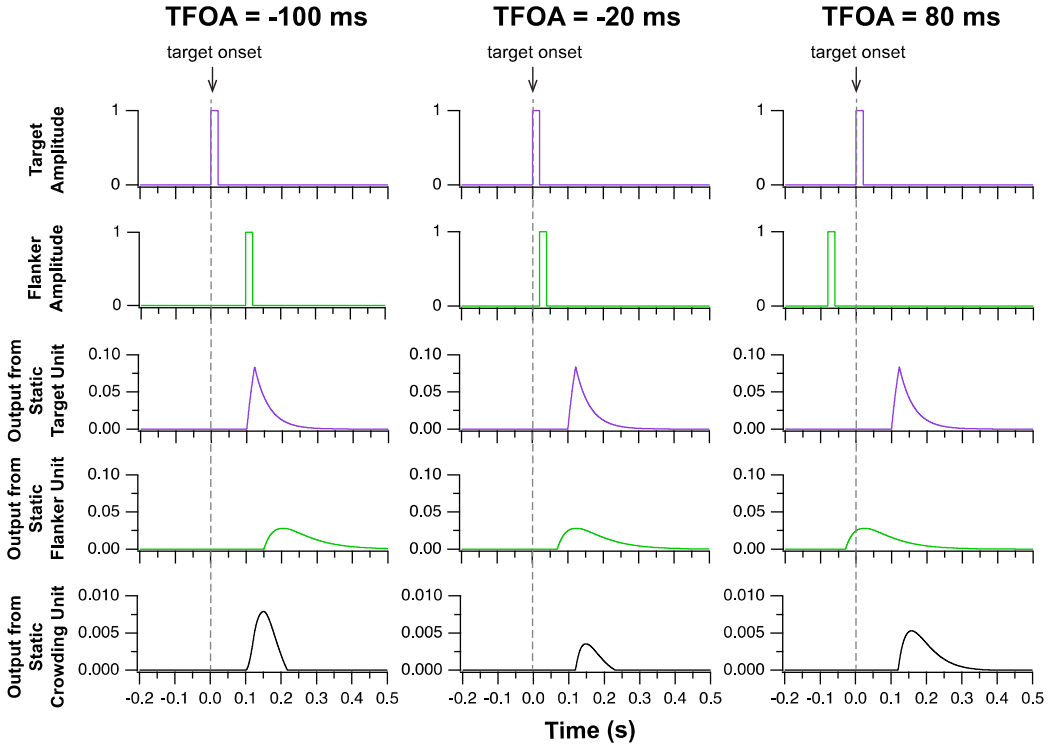


Fig. 7. Examples of the time sequence of the target and flanker unit responses to the onset of the target and flankers at the 3-o'clock position based on simulation results. (A) Results are shown for the motion condition for flanker velocity of 37.5 rpm, a target-flanker separation of 1.4°, and target-onset-to-flanker asynchronies (TOFA) of −60 ms, −30 ms and 0 ms, where maximal crowding occurred at an asynchrony around −30 ms. (B) Results are shown for the static temporal condition for a target-flanker separation of 1.4°, and target-onset-to-flanker asynchronies of −100 ms, −20 ms and 80 ms, where maximal crowding occurred at an asynchrony around −20 ms. In each set of time sequences, the top row shows the onset of the target at the 3-o'clock position and the second row shows the onset of the flankers at this position after the stated TOFA. The third and fourth rows represent the response to the target and the flankers, respectively, and the last row shows the output of the static crowding unit.

motion conditions shown in panel A (37.5 rpm), we used 70 ms for the delay (this delay was reduced to 45 ms for the 18.75 rpm motion condition, see Appendix A2). For the static temporal conditions shown in panel B, we used a delay of 50 ms. Note that we only show the transient responses to the flankers but not the sustained responses to the flankers because the latter contribute minimally to the overall effect in the present experiments (see Table 1 in Discussion). The last (fifth) row shows the output of the static crowding unit, after combining the responses to

the target and flankers. A larger output response corresponds to a higher performance in identifying the orientation of the target. Note that a larger temporal overlap between signals shown in the third (target unit's output) and the fourth (flanker unit's output) row results in a smaller output signal (lower performance).

Fig. 7 shows how the output from the static crowding unit was derived for several selected target-onset-to-flanker asynchronies. With results from the range of target-onset-to-flanker asynchronies from −600 to 600 ms, we could construct crowding functions. The simulated crowding functions obtained for the static temporal and the two motion conditions and for both separations are shown in Fig. 8. The unevenness in the simulated crowding functions is due to limited spatial resolution in the model, particularly evident in simulations of lower flanker velocity, and a much smoother function is expected if additional motion units were used. Based on the fitted function, we again used the three parameters, t_{max}, A_p and σ_E, to compare the simulated data with the experimental data. Results of these comparisons are shown in Fig. 9. All three parameters (t_{max}, A_p and σ_E) of the simulated crowding functions have close to unity relationship with the corresponding empirically measured values, clearly illustrating that with simple adjustments to the model parameters (shown in Appendix A2), a model based on feed-forward interactions of transient and sustained channels in static and motion pathways, taking into consideration the differential latencies, can explain the crowding functions observed under various static and motion conditions.

4. Discussion

The goal of this study was to examine how crowding depends on the spatial and temporal properties of target and flankers, in both physical (retinal) and perceptual domains. Previous studies have provided conflicting evidence regarding the relationship of crowding with the perceptual and physical properties of target and flankers (Chambers et al., 2018; Dakin et al., 2011; Maus et al., 2011). In this study, using a combination of a conventional crowding paradigm and a paradigm with moving flankers, we were able to dissociate the physical and perceptual properties (spatial and temporal) of target and flankers, which allowed us to investigate how crowding depends on these properties. We clearly showed that the relationship between crowding and the perceptual or physical properties of target and flankers is not a simple one.

To dissociate the physical and perceptual separation between a target and its flankers, we made use of the flash-lag effect. The premise was that if crowding were related to the perceptual temporal (or spatial) proximity between the target and flankers, when we changed the target-flanker spatial separation or the flanker velocity, the magnitude of flash-lag (t_{half}) should change similarly with the time at which maximal crowding occurred (t_{max}). Although t_{half} and t_{max} showed a trend to change similarly (in both direction and magnitude) when flanker velocity increased (Fig. 5B), they shifted toward opposite directions when the target-flanker separation increased (Fig. 5A), providing strong evidence that perceptual temporal (or spatial) proximity of the target and flankers was not the only determining factor of maximal crowding. Further, t_{max} was clearly non-zero for both flanker velocities, which implies that physical temporal proximity also cannot account for when maximal crowding occurs. Additionally, t_{max}, when converted to spatial coordinates (Fig. 6B), was lower in both motion conditions when compared with the static spatial condition and thus *physical* spatial proximity also cannot account for when maximal crowding occurs. Given all these findings, we can only conclude that maximal crowding depends on spatial and temporal proximity of the target and flankers, but the dependence is not simply in the physical or perceptual domain.

Traditionally, the flash-lag effect has been studied using simple

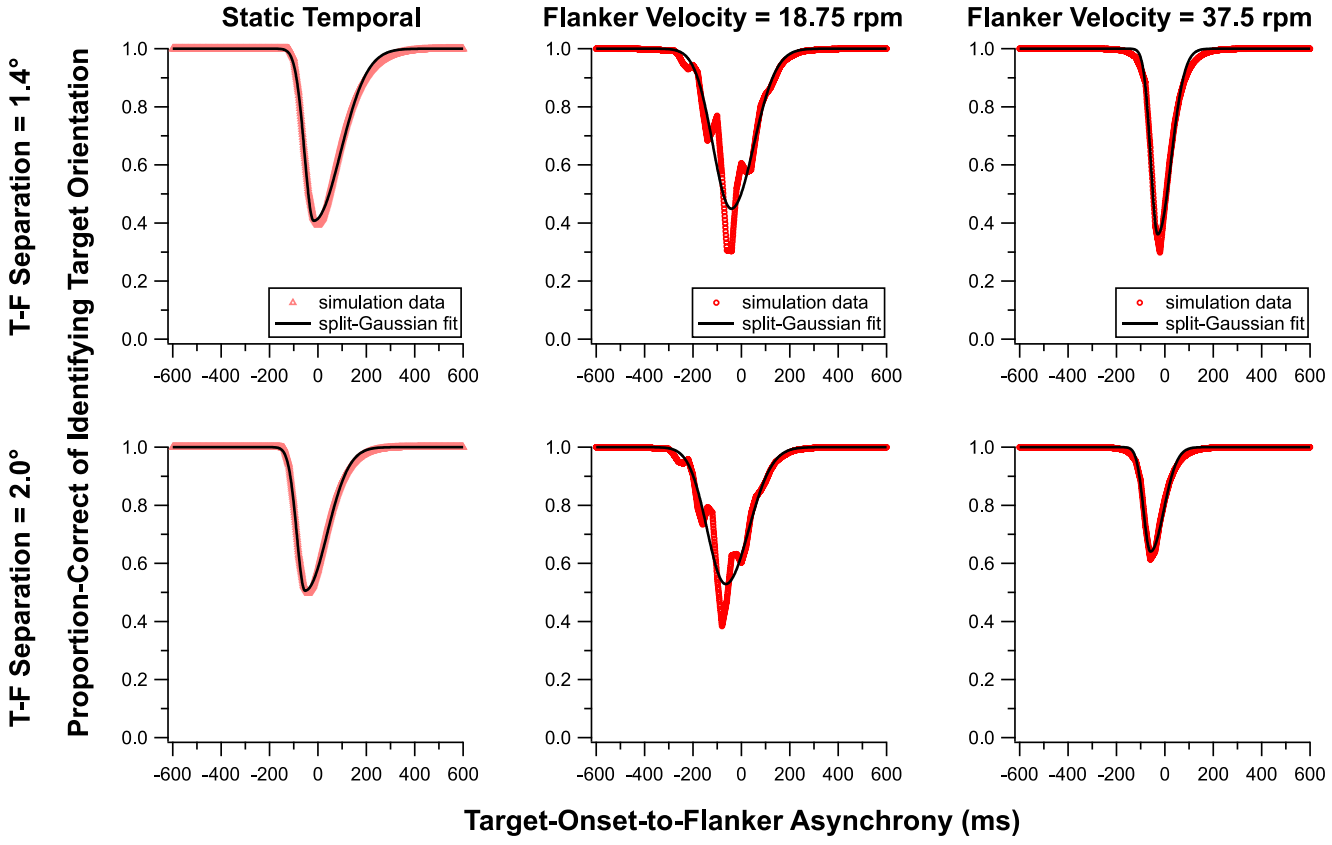


Fig. 8. Simulated crowding functions for (left to right) static temporal, flanker motion of 18.75 rpm and 37.5 rpm are shown for target-flanker separations of 1.4° (top row) and 2.0° (bottom row). A split-Gaussian function described by Eq. (1) was fit (solid black line) to each set of simulated output (pink and red symbols). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

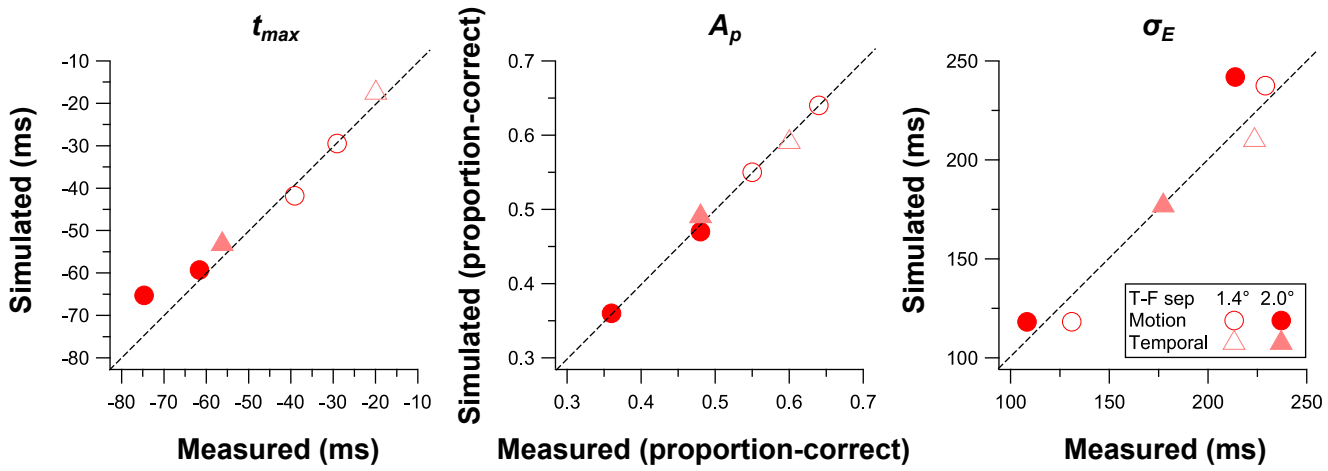


Fig. 9. Comparison of simulated crowding function parameters (from left to right: t_{max}, A_p, σ_E) with those observed in experiments (based on fitted functions in Figs. 3 and 4). In all panels, the abscissa represents measurements averaged across the four observers in the experiments and the ordinate represents corresponding values obtained from simulated crowding functions shown in Fig. 8. Each symbol represents one of the three conditions that was simulated (pink triangles for static temporal and red circles for motion conditions). Unfilled and filled symbols represent target-flanker separations of 1.4° and 2.0°, respectively. Dashed diagonal lines represent unity slope. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

stimuli, such as line segments, small dots and rings (e.g. Baldo & Klein, 1995; Krekelberg & Lappe, 1999; Nijhawan, 1994; Patel et al., 2000; Purushothaman et al., 1998). Here, we showed that even with our T-stimuli, which seem to be more complex than the traditional stimuli, we still observed a substantial flash-lag effect. This finding may not be surprising given that our ability to judge the relative position between random-shaped objects can be as good as that for dot stimuli (Patel, Bedell & Ukwade, 1999); but one logical question is how comparable with the values reported in the literature is our observed flash-lag effect. To our knowledge, there is no standard value reported for the flash-lag effect, but most studies reported a flash-lag magnitude of less than 100 ms, although it is important to note that the magnitude of the effect strongly depends on the relative detectabilities of the moving and flashed objects (Öğmen et al., 2004; Purushothaman et al., 1998). For instance, Purushothaman et al. (1998) showed that an increase in the detectability of the moving object of one log unit increased the perceptual lag of the flashed object by about 60 ms. Critically, for a low detectability moving object, increasing the detectability of the flashed object by 1–2 log units reversed the perception of the flashed objected from lagging behind the moving object by ~ 80 ms to leading the moving object by ~ 40 ms. Given the large range of reported values, the observed temporal lag of our target T (40 – 80 ms) definitely falls within the range reported in the literature. Additionally, we observed a larger flash misalignment (target-onset-to-flanker asynchrony shifted toward a more negative value) when the spatial separation between the target and flankers increased, consistent with a previous report that showed the same effect without a confounding eccentricity effect (Kanai, Sheth & Shimojo, 2004).

Another stimulus property that we manipulated was flanker velocity. Previous studies in humans and primates, using translational or rotational motion, generally reported that the spatial flash-lag effect increased with the speed of the moving objects (Krekelberg & Lappe, 1999; Nijhawan, 1994; Subramaniyan et al., 2013, 2018; Wojtach et al., 2008). Our data in spatial coordinates (Fig. 6G) are consistent with the above findings. When expressed in temporal units, we saw that the flash-lag magnitude was smaller for the higher flanker velocity (Fig. 6F), although the difference in the magnitude of the effect between the two velocities was only moderate and did not reach statistical significance. The reduction in the flash-lag effect in temporal units with the speed of the moving objects is qualitatively consistent with the report of Cantor and Schor (2007).

4.1. Crowding: Comparing the motion vs the static spatial (conventional) conditions

Static stimuli are the traditional and most conventional stimuli used for studying crowding, likely because crowding has long been considered as a spatial phenomenon. How do our results compare with the current understanding of crowding? Let us begin by comparing the results of our static spatial condition with the relevant information in the traditional crowding literature. A signature of crowding is that the spatial extent of crowding scales with the eccentricity of the target, leading to the well-known "Bouma's Law". The scaling factor, sometimes referred to as the Bouma constant, is about 0.4–0.5 in the radial dimension (Bouma, 1970; Chung, 2013; Coates et al., 2021; Pelli et al., 2004; Toet & Levi, 1992) and 0.1–0.2 in the tangential dimension (Chung et al., 2001; Chung, 2013; Coates et al., 2021; Toet & Levi, 1992). These data lead to another signature of crowding — the radial-tangential anisotropy of the crowding zone, at least in the visual periphery (Toet & Levi, 1992; Chung, 2013). In other words, the two-dimensional shape of the spatial crowding zone is elliptical, with the size of the crowding zone (the spatial extent of crowding) larger in the radial dimension than in the tangential dimension. For our static spatial condition, the spatial extent of crowding averaged approximately 12 adeg, equivalent to a spatial offset between the target and flankers of 2.1 visual degrees (calculated as the length of the arc of a 10°-radius circle subtended by an angle of 12° and thus this dimension is tangential with respect to fixation). Therefore, our measurement of the spatial offset between the target and the flankers, essentially representing the spatial extent of crowding in the tangential dimension, yields a Bouma constant of 0.21, consistent with the reported values in the literature.

In the presence of flanker motion, the spatial extent of crowding doubled that in the static spatial condition, averaging approximately 26 adeg, equivalent to 4.5 visual degrees. A few studies have examined crowding in the presence of image motion, and in these studies, both the target and flankers moved together. It has been reported that persons with nystagmus, with constant image motion on the retina, exhibit more crowding, in terms of both magnitude and extent, than persons with normal eye movements (Chung & Bedell, 1995; Pascal & Abadi, 1995; Tailor et al., 2021). In normal vision, Bex et al., 2003 examined the effect of motion on critical spacing by moving a target T around fixation, along with one or four flankers (+ symbols). They found that the size of the crowding zone did not change with image motion, at least up to 8 rotational deg/frame (corresponding to degrees of visual angle of 21°/s

to 84°/s, for 2° to 8° eccentricity). Potentially, image motion does not cause an enlargement of the crowding zone when the target moves together with the flankers. In relation to our study, it is noteworthy that when the extent of crowding was expressed in the temporal domain, it was reduced by half when the flanker velocity increased from 18.75 to 37.5 rpm, leading to the conclusion that the spatial extent of crowding was largely constant for the two flanker velocities (Fig. 6E). These findings can be combined and interpreted as follows: the size of a crowding zone (or the integration field) is mainly limited by the properties of the target and less on those of the flankers. For a static target, the crowding zone assumes a certain size that depends on the eccentricity of the target, which has been suggested to be limited by the length of the horizontal connections in V1 (Levi, 2008; Pelli et al., 2004; Coates et al., 2021). In the presence of flanker motion, the crowding zone enlarges, but the enlargement seems to be tuned to the *presence* of flanker motion, instead of the speed of the motion. This enlargement of the crowding zone could potentially be the result of feedback signals from MT or higher cortical areas. Of course it is also entirely possible that the crowding zone enlarges with motion speed for slower motion, and reaches a constant size once a threshold speed is exceeded and that both the flanker velocities used in our study exceeded this threshold speed, leading to the observation that the spatial extent of crowding was a constant. Further studies are needed to examine whether this conjecture is true by testing slower motion speed.

Another alternative hypothesis is that the difference in the size of the spatial crowding zone for static and moving flankers suggest that the two types of flankers are processed in separate pathways. In the case of static flankers, interactions that result in crowding arise within the static pathway whereas for moving flankers, interactions arise between the motion and the static pathways. This is the architecture used in our model. One may then interpret the differences in terms of within-pathway and between-pathway interactions, i.e., larger spatial zones for between-pathway interactions compared with within-pathway interactions.

4.2. Crowding: Comparing the motion vs the static temporal conditions

Next, we shall compare our results for the static temporal condition with similar crowding experiments in the literature that displayed target and flankers with different onset asynchronies. The study most similar to the static temporal condition is Chung (2016) in which the effect of crowding was evaluated at the fovea and at 10° eccentricity. In that study, target and flankers were presented for the same duration, either 50 ms or 100 ms, but with a range of onset asynchronies. For the 50-ms presentation duration condition tested at 10° eccentricity, when the target and flankers were spatially separated by 1.13° (converted based on the average letter size used and the nominal spacing of 1.25 ×), maximal crowding occurred when the target was presented ~ 30 ms before the flankers. When the target-flanker spatial separation increased to 1.8° (nominal spacing of 2 ×), the target had to be presented even earlier (~50 ms before the flanker onset) for maximal crowding to occur. Despite differences in the exact target-flanker spatial separation and the target and flanker duration, the values obtained for t_{max} in our static temporal condition were very close to those reported by Chung (2016). Besides t_{max}, we also extracted values for the temporal extent of crowding from Chung (2016), which was reduced from 188.5 ms for a target-flanker separation of 1.13° to 164.7 ms for a target-flanker separation of 1.8°, a result that was qualitatively similar to what we observed for the static temporal condition when the target-flanker separation increased. Further, in the current study, in the presence of flanker motion, the temporal extent of crowding was similar between the static condition and the condition with a flanker velocity of 18.75 rpm, and was reduced by half when the flanker velocity increased to 37.5 rpm. However, as discussed earlier, the reduction in the temporal extent of crowding for the faster motion condition might simply reflect a fixed spatial size of the crowding zone.

4.3. Interpretation of model parameters

As shown in Figs. 8 and 9, simulations using our model (Fig. 2) match our experimental results very well. Given that the parameters were selected manually, one might wonder whether these parameters are interpretable in a meaningful way, because in a neural network model like the one used here, many parameters are needed to model a neuron (Appendix A1). First, it is well accepted that static and moving objects are processed in separate pathways (e.g. Kulikowski & Tolhurst, 1973; Legge, 1978; Marrocco, 1976; Schiller, Finlay & Volman, 1976; Schiller, 1998). Even as early as V1, latencies of static and moving objects are different in non-human primates (Subramaniyan et al., 2018). Second, it is a highly plausible assumption that some neurons respond to sustained inputs and some respond to changes in input (Wilson, Babb, Halgren & Crandall, 1983). Third, it is not unreasonable to assume that different populations of neurons will respond to different velocities and with different latencies (Cheng, Hasegawa, Saleem & Tanaka, 1994; Felleman & Van Essen, 1987; Lagae et al., 1994; Lisberger & Movshon, 1999). As can be seen from Appendix A2, most of the parameters are identical for all neurons used in the model and in all tested conditions. We will discuss below the significant parameters and their role in explaining specific aspects of data.

The latency of target and flanker static sustained sub-units (*e_delay*) was 100 ms and the latency of static transient sub-unit (*e_delay*) was 50 ms for the smaller target-flanker separation and 25 ms for the larger separation (see Appendix A2 and panel A in Appendix A4). This reduction in latency partly accounts for the increase in t_{max} in the static temporal condition as target-flanker separation changed from 1.4° to 2.0° (Fig. 6A). Panel A in Appendix A4 also shows the latencies in different motion conditions. The only counter-intuitive designation we encountered was in the motion condition at a target-flanker separation of 1.4°. To account for flash-lag and crowding data for both velocities, the latency of sustained sub-units have to be shorter than that of transient sub-units. The largest difference between the transient and sustained sub-units was 40 ms. We do not know if such signal pathways exist but we do know that other than this likely anomaly, the model proposed here accounts for all aspects of our experimental data with amazing accuracy. That said, we did find evidence for longer latency visual responses in transient cells than in sustained cells in human but these cells were not in the visual cortex (Wilson, Babb, Halgren & Crandall, 1983). Note that the sustained inhibition from flanker units did not play a role in explaining crowding data in the present experiments (see Table 1), therefore another possibility is that the sustained units involved in judgements of position are distinct from those involved in orientation judgements.

To model a change in input signal, we used a low-pass filtered differentiator with a rectifier. The linear low-pass filter of this differentiator has a time constant of 40 ms. All transient sub-units in static and motion pathways have this differentiator with the same low-pass filter time constant. The excitatory input to all transient sub-units (in static and motion pathways) was scaled by a factor of 10 compared with that to all sustained units to keep the input in a reasonable range after differentiation.

The decay constant, A, of transient sub-units increased from 15 (67 ms) to 25 (40 ms) as the target-flanker separation increased from 1.4° to 2.0°. The number enclosed within each pair of parentheses represents the reciprocal of the decay constant and can be viewed as the time constant of the neuron when all inputs are zero. A larger value for A represents a faster decay time. The increase in the decay constant partly accounted for the decrease in σ_E as target-flanker separation increased (Fig. 6D). In contrast, the decay constant of all sub-units in motion pathway was 40 (25 ms), faster than all sub-units in the static pathway.

The output gain of the neuron (g) represents the strength of interaction with the post-synaptic unit because the synaptic weights of all neurons were equal and set to unity. In the static pathway, the output gain of sustained sub-units (0.4) was lower than that of transient sub-

units (~6), but because inputs to the sustained and transient sub-units were different, it is not easy to compare the output gains between them. However, they can be compared across different experimental conditions that use the motion pathway, as shown in panel B of Appendix A4. The output gains for the smaller target-flanker separation were higher compared with those for the larger target-flanker separation. Further, at the smaller separation, the output gains were equal for the two velocities and this partly accounted for the similar amplitudes of crowding in these conditions (Fig. 6C). In contrast, at the larger separation, the output gain was lower at higher velocity and this partly accounted for the reduced crowding amplitude at the higher velocity (Fig. 6C).

The interactions from the motion pathway onto the static pathway where the crowding unit resided occurred via a Gaussian-like weighting function whose peak was centered at the motion unit at 0 adeg, i.e., the 3-o'clock position (see Motion Unit parameters in Appendix A4). The weights from two neighboring units on either side of 0 adeg were slightly modified to account for the data from the two velocities. This alteration represents the involvement of a different set of velocity tuned units for the two velocities used in our experiments. There were no additional alterations for the different target-flanker separations. A similar weighting function in the static pathway (not modeled here) can be used to account for the spatial aspects of the crowding effect in static target-flanker configurations.

The last parameter worth discussing is the threshold parameter *s*. This parameter was set to 0 for all static sub-units and 0.0023 for all motion sub-units. Adding a threshold was a way to control the extent of the inhibitory interaction from motion onto static pathway and this partly accounted for the decreased crowding extent at higher velocity compared to static temporal condition (Fig. 6D).

4.4. Key components of the model

As we mentioned earlier in the Methods, initially we included in the motion pathway of the model (Fig. 2) an inhibitory signal such that the output from a stimulated flanker unit inhibited the previously stimulated neighboring units. Because the results were virtually identical with and without this backward inhibition, this inhibitory signal was subsequently removed, and we concluded that there was very little contribution from the backward inhibition. Similarly, we could perform this in-silico lesion and evaluate the importance of the components of the model by systematically removing the components one at a time. Table 1 summarizes t_{max}, A_p and σ_E of the crowding functions constructed based on the simulation output from the full model (all components included) or when specific components were removed. We also added a simulation condition in which the motion conditions shared the same parameters as those used for the static temporal condition, essentially testing the hypothesis of whether it was necessary to invoke different pathways for motion and non-motion conditions.

For both target-flanker separations, removing the flanker transient inhibition or using the same pathway for the static temporal and motion conditions yielded very different results, when compared with the results for the full model (or one without the backward inhibition), implying that the flanker transient inhibition was an important component in the model and that separate pathways should be used for the static temporal and motion conditions. Removing the flanker sustained inhibition had very minimal effects on the fitted parameters, implying that the contribution of the flanker sustained inhibition was minimal. Nevertheless, we kept this component in our model because its output defined the flash-lag effect, although we did not show it in the time sequences in Fig. 7 and Appendix A3. In short, the in-silico lesion results imply that the transient responses to flankers, but not the sustained responses to flankers, play a key role in accounting for the crowding results in the present experiments, and that separate motion and non-motion pathways are necessary to explain the results when static or moving flankers were used.

4.5. Generality of the model

Our model was developed to explain the data from the experiments in the present paper, therefore it is not surprising that it does a good job at explaining the data. If crowding really arises as a consequence of the feed-forward inhibitory interactions between sustained and transient responses of the flankers and target, then our model should also be able to account for, at least qualitatively, the results of other crowding experiments in which the spatial and temporal properties of the target or flankers were manipulated differently. Here, we shall consider several categories of studies, based on their choice of the spatial and temporal properties of the target and flankers.

4.5.1. Flankers presented before the target yield a smaller crowding effect than flankers presented after the target

Many studies reported a smaller crowding effect when flankers were presented before the target, compared with when flankers were presented after the target (e.g. Chung, 2016; Greenwood, Sayim & Cavanagh, 2014; Harrison & Bex, 2014; Huckauf & Heller, 2004; Ng & Westheimer, 2002; Scolari, Kohnen, Barton & Awh, 2007; Soo, Chakravarthi & Andersen, 2018). Several authors referred to the effect as 'flanker (distractor) previewing' (Huckauf & Heller, 2004; Scolari et al., 2007; Soo et al., 2018) and suggested that top-down attentional processes help to better 'mark' new items by deprioritizing old objects (Watson & Humphreys, 1997). However, the effect can be easily

Table 1
Simulation results from the model, with all components included (full model), or with components removed one at a time.

	Static Temporal			Motion (18.75 rpm)			Motion (37.5 rpm)		
	t_{max} (ms)	A_p	σ_E (ms)	t_{max} (ms)	A_p	σ_E (ms)	t_{max} (ms)	A_p	σ_E (ms)
A. Target-flanker separation = 1.4°									
Full model	−19	0.586	209	−41	0.551	237	−29	0.639	118
Backward inhibition removed	–	–	–	−41	0.551	237	−29	0.639	118
Flanker transient inhibition removed	5	0.01	173	−23	0.051	234	10	0	60
Flanker sustained inhibition removed	−20	0.577	209	−42	0.503	237	−29	0.639	118
Same pathway for static and motion conditions	–	–	–	12	0.856	316	−11	0.756	229
	Static Temporal			**Motion (18.75 rpm)**			**Motion (37.5 rpm)**		
	t_{max} (ms)	A_p	σ_E (ms)	t_{max} (ms)	A_p	σ_E (ms)	t_{max} (ms)	A_p	σ_E (ms)
B. Target-flanker separation = 2°									
Full model	−54	0.492	175	−65	0.471	242	−59	0.359	118
Backward inhibition removed	–	–	–	−64	0.476	243	−59	0.359	118
Flanker transient inhibition removed	−1	0.008	136	−3	0.045	234	10	0	60
Flanker sustained inhibition removed	−55	0.489	174	−67	0.444	236	−59	0.359	118
Same pathway for static and motion conditions	–	–	–	−26	0.654	289	−47	0.617	196

accounted for by our model. When flankers appear well before the target, the transient responses to the flankers occur before the sustained responses to the target. Taking into account the target-flanker onset asynchrony and the differential latencies of the transient (shorter in most cases) and sustained channels (relatively longer in most cases), the transient responses may not overlap with the sustained responses to the target, thus there would be minimal interference (crowding) between the two signals. In contrast, when flankers appear after the target, the transient responses to the flankers have a greater chance of overlapping in time with the sustained responses to the target, thus giving rise to crowding. Simply put, the observation of a smaller crowding effect when flankers precede the target, compared with when they follow the target, can be explained based on the interaction of sustained and transient responses to the target and flankers, which depend on the latencies of the respective channels and the target-flanker onset asynchrony, and that explanations based on top-down attentional mechanisms may not be necessary.

4.5.2. *Crowding is reduced when a 'blink' occurs during target presentation, but not during flanker presentation*

Greenwood et al. (2014) reported that in a standard crowding paradigm in which a target was surrounded by flankers, crowding was significantly reduced when a 'blink' (turning a stimulus off and on quickly) occurred during target presentation, but not when it occurred during flanker presentation. These authors postulated that the blink introduced a (second) transient onset signal. When the blink occurred during target presentation, the transient channel would contain only the target signal but not that of the flanker, thus if the transient channel could support target identification, then target identification would not be affected by the flankers (i.e. no crowding). Similarly, when the blink occurred during flanker presentation, the transient channel would contain only the flanker signal, but the sustained channel would contain both the target and flanker signals, allowing crowding to occur. In short, the mechanism proposed by Greenwood et al. (2014) is similar to the model proposed in Chung (2016). Like the proposed qualitative mechanism of Greenwood et al. and our earlier model, the present model also relies on the interactions of the sustained and transient signals in response to a stimulus and utilizes specific neurally plausible interactions to quantitatively explain the data presented here. Note however, Greenwood et al's model does not explicitly utilize interactions between transient and sustained channels whereas our model is very general and allows for both within and between channel interactions, e.g. the target sustained unit can interact with the flanker sustained unit, but it can also interact with the flanker transient unit. In the model presented here, because the main goal was to understand the difference between crowding caused by static and moving flankers, the target is identified using only the signals from the target's sustained unit. Applying our model to the experiments of Greenwood et al.'s, when a blink is introduced during target presentation, it will give rise to a second sustained and transient response (the first sustained and transient responses are due to the onset of the target presentation), but because there is still a relatively long target presentation following the blink, and that the second sustained response to the target does not receive any inhibition from the flanker transient unit, there would be little crowding if identification relies on the target's sustained response after the blink. However, when the blink is introduced during flanker presentation, the second transient response to the flanker would interact with the target's sustained response, thus causing crowding, and it does not matter whether the target's sustained response before or after the blink mediates identification.

4.5.3. *Identification of a target is adversely affected by stimuli that come before or after the target at the exact spatial location*

One paradigm that has been applied to study the temporal interference on object identification is to present non-target stimuli before and after the target but at the same spatial location. Bonneh, Sagi and Polat (2007) presented a rapid sequence of digits in which one of those in the middle of the sequence was the target (smaller in size) and participants were asked to identify the target digit. They found that when the stimulus onset asynchrony between successive digit presentation was longer, participants were able to identify the target digit at a smaller size, but when the asynchrony was shorter, participants required a larger digit size for identification. Yeshurun and her colleagues used a similar paradigm to study the interference of objects in time, by presenting a non-target item before and another one after the target, all at the same spatial location, instead of using a rapid sequence containing multiple non-target items (Yeshurun, Rashal & Tkacz-Domb, 2015; Tkacz-Domb & Yeshurun, 2017; 2021). These authors also examined the interplay between spatial and temporal separations on object identification by adding non-target items to the target spatially (Yeshurun et al., 2015). In other words, their target was presented with two flankers for a fixed duration, but a triplet containing different items from the target and flankers were presented before, and another triplet after the target and flanker presentation. By adjusting the inter-stimulus interval, the authors could examine the effect of spatial and temporal separations on target identification. Although these authors refer to their paradigm as 'temporal crowding', the presence of non-target items before and after the target resemble the classic forward and backward masking paradigm, and their results would have included the masking effect — an effect that the authors acknowledged. Therefore, this paradigm is not directly comparable with our current study in which we did not have pre- and post-masks on our target, and that our flankers were spatially adjacent to the target, but were never presented at the same spatial location as that of the target (not even when the target and flankers were separated temporally). Yet, we believe that our model, with some additional components to account for masking, such as those reported earlier (e.g. Öğmen, 1993; Öğmen, Breitmeyer & Melvin, 2003), could account for data that include both masking and crowding. Such a model would be very helpful in teasing apart the relative contributions of masking and crowding for the paradigm described here, as well as for explaining other studies that also included a masking component in addition to crowding.

4.6. *Summary*

By manipulating the relative timing of appearance of the target and flankers at a given location (the 3-o'clock position), we found that the performance for identifying the orientation of the target T demonstrated a V-shaped function with the target-onset-to-flanker asynchrony. The attributes of this crowding function, t_{max}, A_p and σ_E, are substantially different between the static- and the moving-flanker conditions, suggesting that interactions on the target due to static or moving flankers occur via different pathways. We also showed that neither the physical nor the perceptual (spatial or temporal) proximity between the target and flankers is necessary for maximal crowding. It is well known that in the primate visual system, visual information from the retina travels along parallel pathways and is processed by a hierarchy of specialized centers that receive this information with different latencies (Marrocco, 1976; Schiller, Finlay & Volman, 1976; Schiller, 1998; Subramaniyan et al., 2018). Further, these latencies depend on the spatio-temporal properties of the target-flanker configuration (e.g. retinal eccentricity, static vs moving, luminance of target and flankers). Thus we conclude that factors that determine the attributes of crowding are the spatio-temporal properties of target-flanker configuration and the location of processing centers where strongest interactions between target and flankers occur (i.e. the latencies of target and flanker information propagation). A parallel model with separate motion and non-motion pathways, and in which crowding was modeled as a feed-forward inhibitory interaction from sustained and transient sub-units responding to the flanker upon a sustained sub-unit's response to the target, which have different latencies, was able to account for most of the experimental data.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors are grateful to Drs. Harold Bedell and Edward Tehovnik for their helpful comments on an earlier version of this paper. This study was supported by research grants R01-EY012810 from the National Institutes of Health and BCS 0924636 from the National Science Foundation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.visres.2022.108012.

References

Baldo, M. V. C., & Klein, S. A. (1995). Extrapolation or attention shift? *Nature, 378*, 565–566.
Bex, P. J., Dakin, S. C., & Simmers, A. J. (2003). The shape and size of crowding for moving targets. *Vision Research, 43*, 2895–2904.
Bonneh, Y. S., Sagi, D., & Polat, U. (2007). Spatial and temporal crowding in amblyopia. *Vision Research, 47*, 1950–1962.
Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature, 226*, 177–178.
Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 443–446.
Cantor, C. R. L., & Schor, C. M. (2007). Stimulus dependence of the flash-lag effect. *Vision Research, 47*, 2841–2854.
Chambers, A., Johnston, A., & Roach, N. W. (2018). Visual crowding is unaffected by adaptation-induced spatial compression. *Journal of Vision, 18*(3), 12.
Cheng, K., Hasegawa, T., Saleem, K. S., & Tanaka, K. (1994). Comparison of neuronal selectivity for stimulus speed, length, and contrast in the prestriate visual cortical areas V4 and MT of the macaque monkey. *Journal of Neurophysiology, 71*, 2269–2280.
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.
Chung, S. T. L. (2013). Cortical reorganization after long-term adaptation to retinal lesions in humans. *Journal of Neuroscience, 33*, 18080–18086.
Chung, S. T. L. (2016). Spatio-temporal properties of letter crowding. *Journal of Vision, 16*(6), 8.
Chung, S. T. L., & Bedell, H. E. (1995). Effect of retinal image motion on visual acuity and contour interaction in congenital nystagmus. *Vision Research, 35*, 3071–3082.
Chung, S. T. L., Legge, G. E., & Levi, D. M. (2001). Spatial-frequency and contrast properties of crowding. *Vision Research, 41*, 1833–1850.
Chung, S. T. L., Patel, S. S., Bedell, H. E., & Yilmaz, O. (2007). Spatial and temporal properties of the illusory motion-induced position shift for drifting stimuli. *Vision Research, 47, 231–243 Vision Research, 47*, 231–243.
Coates, D. R., Chin, J. M., & Chung, S. T. L. (2013). Factors affecting crowded acuity: Eccentricity and contrast. *Optometry & Vision Science, 90*, 628–638.
Coates, D. R., & Chung, S. T. L. (2016). Crowding in the S-cone pathway. *Vision Research, 122*, 81–92.
Coates, D. R., Ludowici, C. J. H., & Chung, S. T. L. The generality of the critical spacing for crowded optotypes: From Bouma to the 21st century. *Journal of Vision, 21(11), 18*.
Dakin, S. C., Greenwood, J. A., Carlson, T. A., & Bex, P. J. (2011). Crowding is tuned for perceived (not physical) location. *Journal of Vision, 11*(9), 2.
DeValois, R. L., & DeValois, K. K. (1991). Vernier acuity with stationary moving Gabors. *Vision Research, 31*, 1619–1626.
Ehlers, H. (1953). Clinical testing of visual acuity. *AMA Archives of Ophthalmology, 49*, 431–434.
Felleman, D. J., & Van Essen, D. C. (1987). Receptive field properties of neurons in area V3 of macaque monkey extrastriate cortex. *Journal of Neurophysiology, 57*, 889–920.
Flom, M. C., Weymouth, F. W., & Kahneman, D. (1963). Visual resolution and contour interaction. *Journal of the Optical Society of America, 53*, 1026–1032.
Flom, M. C., Heath, G. G., & Takahashi, E. (1963). Contour interaction and visual resolution: Contralateral effects. *Science, 142*, 979–980.
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).
Grossberg, S. (1972). A neural theory of punishment and avoidance, II: Quantitative theory. *Mathematical Biosciences, 15*, 253–285.
Harrison, W. J., & Bex, P. J. (2014). Integrating retinotopic features in spatiotopic coordinates. *Journal of Neuroscience, 34*, 7351–7360.
He, Y., & Legge, G. E. (2017). Linking crowding, visual span, and reading. *Journal of Vision, 17*(11), 11.
Huckauf, A., & Heller, D. (2004). On the relations between crowding and visual masking. *Perception & Psychophysics, 66*, 584–595.
Kalpadakis-Smith, A. V., Goffaux, V., & Greenwood, G. A. (2018). Crowding for faces is determined by visual (not holistic) similarity: Evidence from judgements of eye position. *Scientific Reports, 8*, 12556.
Kanai, R., Sheth, B. R., & Shimojo, S. (2004). Stopping the motion and sleuthing the flash-lag effect: Spatial uncertainty is the key to perceptual mislocalization. *Vision Research, 44*, 2605–2619.
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.
Krekelberg, B., & Lappe, M. (1999). Temporal recruitment along the trajectory of moving objects and the perception of position. *Vision Research, 39*, 2669–2679.
Kulikowski, J. J., & Tolhurst, D. J. (1973). Psychophysical evidence for sustained and transient detectors in human vision. *Journal of Physiology, 232*, 149–162.
Lagae, L., Maes, H., Raiguel, S., Xiao, D. K., & Orban, G. A. (1994). Responses of macaque STS neurons to optic flow components: A comparison of areas MT and MST. *Journal of Neurophysiology, 71*, 1597–1626.
Legge, G. E. (1978). Sustained and transient mechanisms in human vision: Temporal and spatial properties. *Vision Research, 18*, 69–81.
Levi, D. M. (2008). Crowding—an essential bottleneck for object recognition: A mini-review. *Vision Research, 48*, 635–654.
Levi, D. M., & Carney, T. (2009). Crowding in peripheral vision: Why bigger is better? *Current Biology, 19*, 1988–1993.
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.
Lisberger, S. G., & Movshon, J. A. (1999). Visual motion analysis for pursuit eye movements in area MT of macaque monkeys. *Journal of Neuroscience, 19*, 2224–2246.
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.
MacKay, D. M. (1958). Perceptual stability of a stroboscopically lit visual field containing self-luminous objects. *Nature, 181*, 507–508.
Marrocco, R. T. (1976). Sustained and transient cells in monkey lateral geniculate nucleus: Conduction velocities and response properties. *Journal of Neurophysiology, 39*, 340–353.
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.
Maunsell, J. H., & Gibson, J. R. (1992). Visual response latencies in striate cortex of the macaque monkey. *Journal of Neurophysiology, 68*, 1332–1344.
Maus, G. W., Fischer, J., & Whitney, D. (2011). Perceived positions determine crowding. *PLoS ONE, 6*(5), Article e19796.
Ng, J., & Westheimer, G. (2002). Time course of masking in spatial resolution tasks. *Optometry and Vision Science, 79*, 98–102.
Nijhawan, R. (1994). Motion extrapolation in catching. *Nature, 370*, 256–257.
Öğmen, H. (1993). A neural theory of retino-cortical dynamics. *Neural Networks, 6*, 245–273.
Öğmen, H., Breitmeyer, B. G., & Melvin, R. (2003). The what and where in visual masking. *Vision Research, 43*, 1337–1350.
Öğmen, H., Patel, S. S., Bedell, H. E., & Camuz, K. (2004). Differential latencies and the dynamics of the position computation for moving targets, assessed with the flash-lag effect. *Vision Research, 44*, 2109–2128.
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.
Pascal, E., & Abadi, R. V. (1995). Contour interaction in the presence of congenital nystagmus. *Vision Research, 35*, 1785–1789.
Patel, S. S., Bedell, H. E., & Ukwade, M. T. (1999). Vernier judgments in the absence of regular shape information. *Vision Research, 39*, 2349–2360.
Patel, S. S., Öğmen, H., Bedell, H. E., & Sampath, V. (2000). Flash-lag effect: Differential latency, not postdiction. *Science, 290*, 1051a.
Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442.
Pelli, D. G. (2008). Crowding: A cortical constraint on object recognition. *Current Opinion in Neurobiology, 18*, 445–451.
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.
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.
Purushothaman, G., Patel, S. S., Bedell, H. E., & Öğmen, H. (1998). Moving ahead through differential visual latency. *Nature, 396*, 424.
Schiller, P. H. (1998). The neural control of visually-guided eye movements. In: Cognitive Neuroscience of Attention, J. Richards (ed), Lawrence Erlbaum, pp. 3–50.
Schiller, P. H., Finlay, B. L., & Volman, S. F. (1976). Quantitative studies of single cell properties in monkey striate cortex. *III. Spatial frequency. Journal of Neurophysiology, 39*, 1334–1351.
Schmolesky, M. T., Wang, Y., Hanes, D. P., Thompson, K. G., Leutgeb, S., Schall, J. D., & Leventhal, A. G. (1998). *Journal of Neurophysiology, 79*, 3272–3278.
Scolari, M., Kohnen, A., Barton, B., & Awh, E. (2007). Spatial attention, preview, and popout: Which factors influence critical spacing in crowded displays? *Journal of Vision, 7*(2).
Soo, L., Chakravarthi, R., & Andersen, S. K. (2018). Critical resolution: A superior measure of crowding. *Vision Research, 153*, 13–23.
Stuart, J. A., & Burian, H. M. (1962). A study of separation difficulty: Its relationship to visual acuity in normal and amblyopic eyes. *American Journal of Ophthalmology, 53*, 471–477.

Subramaniyan, M., Ecker, A. S., Berens, P., & Tolias, A. S. (2013). Macaque monkeys perceive the flash lag illusion. *PLoS One, 8*, Article e58788.

Subramaniyan, M., Ecker, A. S., Patel, S. S., Cotton, J., Bethge, M., Pitkow, X., … Tolias, A. S. (2018). Faster processing of moving compared with flashed bars in awake macaque V1 provides a neural correlate of the flash-lag illusion. *Journal of Neurophysiology, 120*, 2430–2452.

Tailor, V. K., Theodorou, M., Dahlmann-Noor, A. H., Dekker, T. M., & Greenwood, J. A. (2021). Eye movements elevate crowding in congenital idiopathic nystagmus. bioRxiv, doi: https://doi.org/10.1101/2021.01.16.426927.

Tkacz-Domb, S., & Yeshurun, Y. (2017). Spatial attention alleviates temporal crowding, but neither temporal nor spatial uncertainty are necessary for the emergence of temporal crowding. *Journal of Vision, 17*(3).

Tkacz-Domb, S., & Yeshurun, Y. (2021). Temporal crowding is a unique phenomenon reflecting impaired target encoding over large temporal intervals. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-021-01943-8

Toet, A., & Levi, D. M. (1992). The two-dimensional shape of spatial interaction zones in the parafovea. *Vision Research, 32*, 1349–1357.

van den Berg, R., Roerdink, J. B. T. M., & Cornelissen, F. W. (2010). A neurophysiologically plausible population code model for feature integration explains visual crowding. *PLoS Computational Biology, 6*(1), Article e1000646.

Watson, D. G., & Humphreys, G. W. (1997). Visual marking: Prioritizing selection for new objects by top-down attentional inhibition of old objects. *Psychological Review, 104*, 90–122.

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, 14*, 160–168.

Wilson, C. L., Babb, T. L., Halgren, E., & Crandall, P. H. (1983). Visual receptive fields and response properties of neurons in human temporal lobe and visual pathways. *Brain, 106*, 473–502.

Wojtach, W. T., Sung, K., Truong, S., & Purves, D. (2008). An empirical explanation of the flash-lag effect. *Proceedings of the National Academy of Sciences, 105*, 16338–16343.

Xia, Y., Manassi, M., Nakayama, K., Zipser, K., & Whitney, D. (2020). Visual crowding in driving. *Journal of Vision, 20*(6), 1.

Yeshurun, Y., Rashal, E., &n Tkacz-Domb, S. (2015). Temporal crowding and its interplay with spatial crowding. *Journal of Vision, 15(3),* 11.1–11.16.