Vision Research 221 (2024) 108422

Contents lists available at ScienceDirect

Vision Research

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

Spatial summation for motion detection

Joshua A. Solomon [a,*], Fintan Nagle [a], Christopher W. Tyler [a,b]

[a] Centre for Applied Vision Research, City, University of London, UK
[b] Smith-Kettlewell Eye Research Institute, San Francisco, CA, USA

ARTICLE INFO

Keyword:
Motion

ABSTRACT

We used the psychophysical summation paradigm to reveal some spatial characteristics of the mechanism responsible for detecting a motion-defined visual target in central vision. There has been much previous work on spatial summation for motion detection and direction discrimination, but none has assessed it in terms of the velocity threshold or used velocity noise to provide a measure of the efficiency of the velocity processing mechanism. Motion-defined targets were centered within square fields of randomly selected gray levels. The motion was produced within the disk-shaped target region by shifting the pixels rightwards for 0.2 s. The uniform target motion was perturbed by Gaussian motion noise in horizontal strips of 16 pixels. Independent variables were field size, the diameter of the disk target, and the variance of an independent perturbation added to the (signed) velocity of each 16-pixel strip. The dependent variable was the threshold velocity for target detection. Velocity thresholds formed swoosh-shaped (descending, then ascending) functions of target diameter. Minimum values were obtained when targets subtended approximately 2 degrees of visual angle. The data were fit with a continuum of models, extending from the theoretically ideal observer through various inefficient and noisy refinements thereof. In particular, we introduce the concept of sparse sampling to account for the relative inefficiency of the velocity thresholds. The best fits were obtained from a model observer whose responses were determined by comparing the velocity profile of each stimulus with a limited set of sparsely sampled "DoG" templates, each of which is the product of a random binary array and the difference between two 2-D Gaussian density functions.

1. Introduction

Summation is one of three general paradigms (adaptation and masking being the other two; Graham, 1989) available to psychophysicists for answering questions about the spatial characteristics of mechanisms responsible for detecting visual targets. In the summation paradigm, spatial characteristics may be inferred from the relationship between target size and the amplitude required for detection. While several studies have measured direction thresholds as a function of the extent of moving targets (Lappin et al., 2009; Tadin et al., 2003; van de Grind et al., 1983), none have measured motion detection thresholds, or analyzed them in terms of the ideal observer theory for the adaptive receptive field structure underlying the motion detection performance or its efficiency assessed with masking noise, as we do in the present study.

For spots of light on an otherwise-uniform visual field, the retinal illumination required for detection decreases as the area of the spot increases to a size now known as "Riccò's area," in honor of Riccò (1877). In central vision, that area can be anywhere between 0.025 and 0.4 deg^2 (corresponding to radial extents between 0.09 and 0.35 deg; Barlow, 1958), depending on background illumination and exposure duration. Beyond this limit, thresholds decrease more slowly or asymptote to a constant level (see Fig. 1). By convention, summation within a receptive field is assumed to be linear. Consequently, the measurements of Riccò and others have allowed inferences regarding the area (i.e., 0.4 deg^2) of the smallest receptive fields responsible for detecting spots of light in central vision.

Just as Riccò (1877) established the limits of spatial summation for detecting luminance-defined targets, we set out to discover whether there are analogous limits for detecting motion-defined targets (see Fig. 2). Whereas detection of a brief, monochromatic target within Riccò's area depends primarily on the number of quanta it contains (i.e., target intensity), detection of motion in random-dot stimuli depends primarily on target velocity (Nakayama & Tyler, 1981). Thus, whereas Riccò and others measured threshold intensity as a function of target area, we measured threshold velocity as a function of target area.

* Corresponding author.
E-mail address: j.a.solomon@city.ac.uk (J.A. Solomon).

https://doi.org/10.1016/j.visres.2024.108422
Received 5 August 2023; Received in revised form 20 April 2024; Accepted 22 April 2024

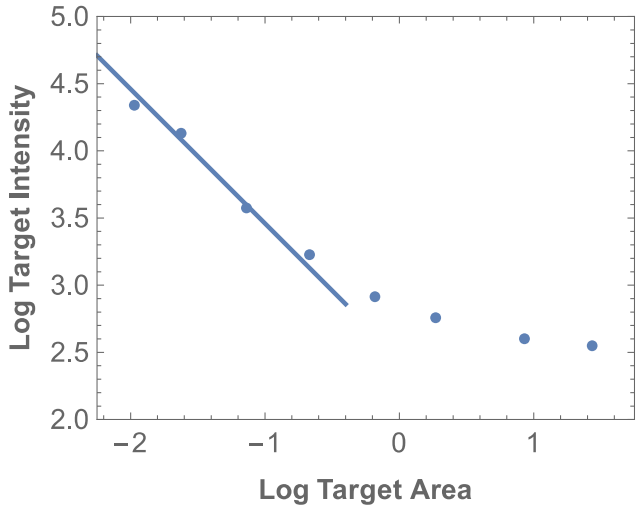

Fig. 1. Detection thresholds for a 0.93-s flash of light against an unilluminated background. The line has a gradient of −1. Intensities in 507 mμ quanta/(s•deg^2), areas in deg^2. Data replotted from Barlow (1958).

2. Methods

These methods were granted formal approval from the Optometry Proportionate Review Committee at City, University of London, and written informed consent was provided by all observers.

2.1. Stimuli

All stimuli were centered on a mean-gray background. The typical size of the stimulus field (as in Fig. 2b) was 256 × 256 pixels. However, in some cases, we used larger fields (see Table 1). In all cases, the stimulus duration was 0.5 s (30 video frames at 60 Hz). Target motion was confined to the middle 12 video frames. In fact, there was technically no motion on the first of these 12 frames ($f = 1$), because it was identical to each of the preceding frames, wherein each pixel was independently selected from the full gamut of gray levels.

Each (horizontal) raster of pixels was notionally divided into 16 or 17 strips of pixels. In the latter case, the central 15 of these strips had 16 pixels each, leaving n and 16 −n pixels for the outer two strips. For each raster and each stimulus field, the number n was independently selected from the integers between 0 and 15. When $n = 0$, the raster was equally divided into 16 notional strips.

In successive frames ($f = 2,3,\ldots,12$), the gray levels within each 16-pixel strip shifted rightwards (with respect to the first frame) by the number of positions corresponding to the integer closest to $(f-1)v(r;a,r_{\text{targ}},\sigma_{\text{ext}})$. The latter factor in this expression is the disk-shaped profile of velocities:

$$v(r;a,r_{\text{targ}},\sigma_{\text{ext}}) = \begin{cases} a+\sigma_{\text{ext}}\,Z, r < r_{\text{targ}} \\ \sigma_{\text{ext}}\,Z, r \geq r_{\text{targ}} \end{cases}, \quad (1)$$

where r represents the distance (measured in degrees of visual angle) between the center of the 16-pixel strip and the center of the stimulus, a is the target's amplitude (corresponding to its expected velocity), Z is an independent sample (one for each strip) from a standard normal

Table 1
Methodological details. Total number of trials: 34,800.

Apparatus	Observer	Size (pixels)	σ_{ext}	Target sizes	Trials/Target
USA	CWT	256 × 256	0	7	200
USA	CWT	256 × 256	0.2	7	200
USA	CWT	256 × 256	0.4	7	200
USA	CWT	512 × 512	0	8	600
USA	CWT	512 × 512	0.4	8	200
UK	JAS	256 × 256	0	7	200–300
UK	JAS	256 × 256	0.2	7	200
UK	JAS	256 × 256	0.4	7	200–300
UK	JAS	512 × 512	0	5	200
UK	JAS	512 × 512	0.2	5	200
UK	JAS	512 × 512	0.4	5	200–300
UK	JAS	736 × 736	0	5	200–300
UK	JAS	736 × 736	0.2	3	100
UK	JAS	736 × 736	0.4	5	300
UK	KZC	256 × 256	0	7	200
UK	KZC	256 × 256	0.2	7	200
UK	KZC	256 × 256	0.4	7	200
UK	KZC	512 × 512	0	3	200
UK	KZC	736 × 736	0	5	200
UK	KZC	736 × 736	0.4	5	200
UK	FSN	256 × 256	0	5	200
UK	FSN	256 × 256	0.2	5	200
UK	FSN	256 × 256	0.4	7	200
UK	FSN	736 × 736	0	5	200
UK	FSN	736 × 736	0.2	5	200
UK	FSN	736 × 736	0.4	5	300

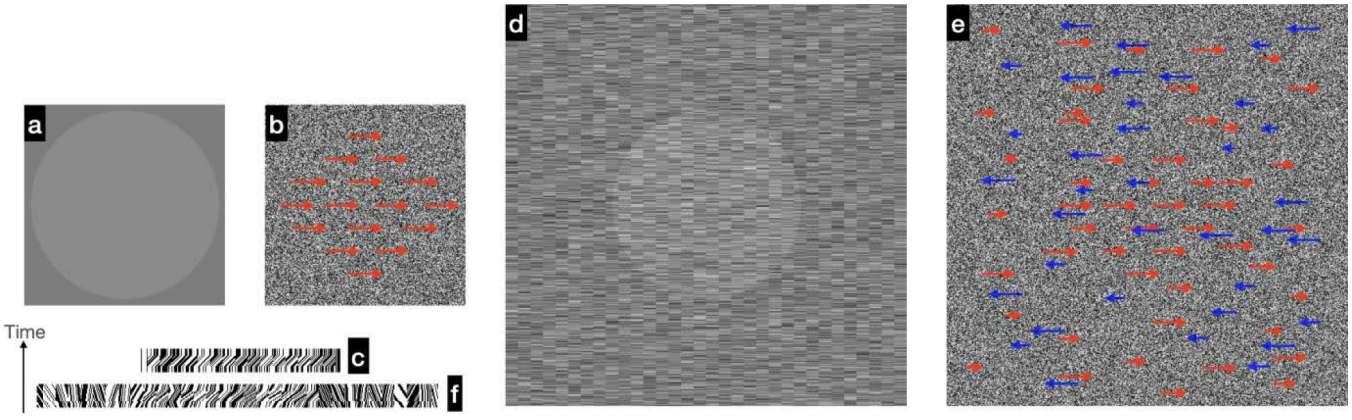

Fig. 2. Static representation of motion-defined targets (panels **b**, **c**, **e**, and **f**) and analogous targets in the luminance domain (panels **a**, **d**). Panel **a** contains a disk that is brighter than its small (256 × 256) inscribing square. Thin, red arrows in panel **b** are meant to represent a disk of rightward motion. Panel **c** contains a space–time representation of the central raster in panel **b**, with one raster per frame of a 30-frame stimulus. Note that the target's motion is confined to the middle 12 frames. Panels **d**, **e**, and **f** show "noisy," medium-sized (512 × 512) stimulus fields containing the same targets. In these stimuli, each row of pixels was divided into 16-pixel strips. These 16 pixels moved with uniform velocity. Correspondingly, each row of pixels in the luminance-domain analogue (panel **d**) has been divided into 16-pixel strips of uniform luminance. For a movie of representative targets, with and without motion noise, see http://www.staff.city.ac.uk/~solomon/MotionSummation.zip.

distribution, and target size r_{targ} and noise level σ_{ext} are independent variables. Target diameters were constrained to contain pixels that approximated integer powers of $\sqrt{2}$: 32, 45, 64, …, 724. Gray levels moving beyond one edge of a 16-pixel strip were replaced by independently selected gray levels at the opposite edge.

2.2. Apparatus

Travel restrictions resulting from the COVID-19 pandemic required parallel data collection in geographically separated laboratories. In the UK (see Table 1) stimuli were presented on a gamma-linearized Dell CRT with a spatial resolution of 1600 × 1200 pixels. Video signals were generated by a mid-2012 MacBook Pro, running OS X 10.12.6, MATLAB R2016b, and version 3.0.12 of the Psychtoolbox (Brainard, 1997). Maximum and minimum luminances were 69.09 and 0.027 cd/m^2, respectively. Viewing distance was fixed at 1.67 m, producing pixels having a retinal subtense of 0.45 arcmin and velocities that were 0.45 deg/s times the velocity profile described in Eqn. (1). This value ($0.45\ \text{deg/s} = 1\text{pixel/frame}$) corresponds to the maximum (expected) target velocity. When velocity noise is present, some parts of the stimulus will have greater velocity, others will have less.

In the USA stimuli were presented on a gamma-linearized Asus monitor with a spatial resolution of 2560 × 1440 pixels. Video signals were generated by a 2020 quad core Pentium i7, running MATLAB 2019b and version 3.0.15 of the Psychtoolbox. Maximum and minimum luminances were 250 and 1 cd/m^2, respectively. Viewing distance was fixed at 1.45 m, producing pixels having a retinal subtense of 0.67 arcmin and velocities that were 0.67 deg/s times the velocity profile described in Eqn. (1).

2.3. Procedure

On each trial of this two-alternative, forced-choice (2AFC) task, two stimulus fields were shown. One contained no target ($a = 0$); the other contained a target whose amplitude ($a > 0$) was adjusted using a QUEST adaptive staircase (Watson & Pelli, 1983), configured to converge on the "threshold" amplitude required for the observer to indicate which of the two fields contained the target with an accuracy of 81% correct.[1] Immediate auditory feedback followed each incorrect response. The inter-stimulus interval was fixed at 0.5 s, during which the screen was uniformly illuminated at the background level.

In the UK setup, when the field size was 256 × 256 pixels, the inter-trial interval (between the previous trial's feedback and exposure of the first stimulus) was 1.0 s. When the field size was 512 × 512 pixels, stimulus-computation time caused the inter-trial interval to increase to 2.7 s. When the field size was 736 × 736 pixels, it increased to 4.7 s. In the USA setup, it was <1 s for all field sizes.

2.4. Observers and conditions

Authors JAS, FSN, and CWT served as the primary observers. Additional observations were carried out by KZC. Target size (r_{targ}) and motion noise level (σ_{ext}) were fixed within each block of 100 trials. Further details are given in Table 1 and Fig. 3. Viewing was binocular, through natural pupils. Fixation at the center of the display was encouraged.

[1] Even with the most sophisticated adaptive psychometric methods, the number of trials required for precise, simultaneous estimates of threshold and psychometric slope was prohibitive. Consequently, we fixed the latter parameter (β in Watson & Pelli's Eqn. (13) at the value of 1.3, which was both the median value obtained in pilot experiments (not reported here) and all linear models limited by Gaussian noise (May & Solomon, 2013).

3. Results

Raw data are available at https://www.staff.city.ac.uk/~solomon/MotionSummation.zip, where all the modelling code is included. In general, faster-moving stimuli were easier to detect. Maximum-likelihood estimates of (81 % correct) thresholds are shown in Fig. 3.[2] For all targets, displacement = velocity × 0.2 s. Thresholds fell as the area of the disk increased to a value of approximately 3 deg^2, before rising again in the form of a "swoosh." This area corresponds to a radius of about 1 deg. Motion noise was effective in elevating detection thresholds for all sizes of target, though to different extents across the observers.

3.1. Models[3]

In 2AFC detection tasks, Signal-Detection Theory (Green & Swets, 1966) ascribes all incorrect responses to visual signals in the target's absence exceeding those in the target's presence. Probability correct is given by

$$\Psi = \delta + \psi - 2\delta\psi, \tag{2}$$

where δ represents the lapse rate (i.e., the incorrect proportion of responses to massively suprathreshold targets, which was assumed to be 0.01 in our implementation of QUEST) and

$$\psi = \int_{-\infty}^{\infty} F_N(x) F_S'(x) dx. \tag{3}$$

In Eq. (3), $F_N(x)$ is the cumulative distribution function (CDF) for the visual signal N elicited in the absence of a target and $F_S'(x)$ is the derivative of the CDF (i.e., it is the density) for the signal S elicited in the presence of the target.

3.1.1. The ideal observer

Behavior of the ideal observer is based on signals described by the dot product between each stimulus field and the expected target:

$$X = v\left(r; a_X, r_{\text{targ}}, \sigma_{\text{ext}}\right) \bullet \left[1 - H\left(r - r_{\text{targ}}\right)\right], \tag{4}$$

where H is the Heaviside step function. Consequently, the CDFs for the target (where $X = S$ and $a_S > 0$) and nontarget (where $X = N$ and $a_N = 0$) can be written as

$$F_X(x) = \frac{1}{2}\text{erfc}\left[\frac{-x + a_X \pi r_{\text{targ}}^2}{\sqrt{2\pi\sigma_{\text{ext}}^2 r_{\text{targ}}^2}}\right], \tag{5}$$

[2] These estimates were obtained with $\beta = 1.3$ (see Footnote 1). When simultaneously fitting all the data described in this paper, the single maximum-likelihood value for this parameter was 1.0. We must stress that, although our estimates of psychometric slope should not be considered precise, when fitting just the conditions without motion noise (i.e., $\sigma_{\text{ext}} = 0$), psychometric functions were – if anything – even shallower: $\beta = 0.7$. This result seems incompatible with intrinsic uncertainty models, as described below. Psychometric functions for 2AFC detection of contrast-defined targets are almost never this shallow (see Mayer & Tyler, 1986, and May & Solomon, 2013 for a review). Finally, although a full investigation of these psychometric slopes is beyond this scope of this study, there are various potential explanations for why they may be shallower than those predicted by linear models with constant, Gaussian noise: i) the limiting noise is platykurtic (e.g., Neri, 2013), ii) the variance of limiting noise increases with signal strength (e.g., Solomon, 2007), iii) stimulus velocity undergoes compressive, nonlinear transduction (e.g., according to a power law with exponent < 1), and iv) pooling data across sessions in which sensitivity changes, due to the effects of practice and/or fatigue.

[3] This section (Models) and the next section (Model Fits) may be skipped without loss of continuity.

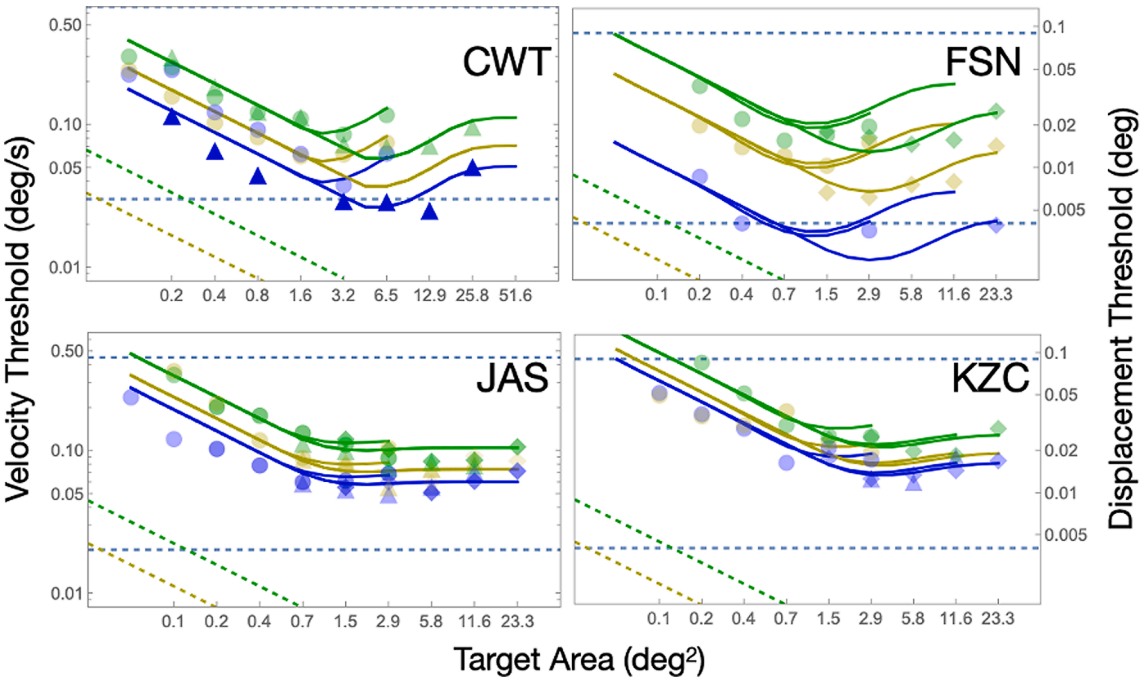

Fig. 3. Velocity/displacement thresholds and best-fitting (swoosh-like) curves from the "DoGS" model for four observers. Dashed horizontal lines indicate maximum and minimum non-zero target velocities available to each observer. Green, amber, and blue symbols illustrate 81%-correct thresholds in the presence of noise having RMS amplitudes of 40 %, 20 %, or 0 % of the maximum target velocity, respectively. Symbol opacity is proportional to the number of trials (see Table 1). Circles, triangles, and diamonds correspond to field sizes of 256 × 256, 512 × 512, and 736 × 736 pixels, respectively. Solid curves illustrate best-fits of the DoGS model (curves corresponding to large-field fits extend to the right of the largest targets contained within smaller fields), while dashed amber and green lines indicate the ideal observer's performance with noise amplitudes of 20% and 40%, respectively. Field size, target radius, and noise amplitude were held constant within each block of 100 trials. Note that CWT's horizontal axis differs from those of the other three observers because the viewing distance on his apparatus (see Methods) was smaller.

where erfc is the complementary error function. Note that there are no free parameters in the ideal observer model. The expected signal $\mathrm{E}(S)$ grows in proportion to the square-root of target area, $\sqrt{\pi r_{\mathrm{targ}}^2}$. Consequently, the ideal observer's thresholds fall in proportion to this quantity. On log–log axes such as the panels of Fig. 2, this proportionality manifests as a line with gradient $-1/2$.

Our human observers' thresholds were always higher than the ideal observer's. Indeed, in the absence of external motion noise (i.e., when $\sigma_{\mathrm{ext}} = 0$), the ideal observer will never make an incorrect response; its threshold is effectively 0.[4] However, it is noteworthy that the initial response gradients for the smaller targets appear to follow the same trend as the ideal observer, but at an elevated level.

3.1.2. The noisy observers

To elevate the (otherwise) ideal observer's thresholds, we assume that its visual signals are noisy, even in the absence of any motion noise in the stimulus. By definition, "late" Gaussian noise is injected after the velocity profile has been matched with the template (i.e., the expected target). In this case we have

$$F_X(x) = \frac{1}{2}\mathrm{erfc}\left[\frac{-x + a_X \pi r_{\mathrm{targ}}^2}{\sqrt{2\pi\left(\sigma_{\mathrm{ext}}^2 r_{\mathrm{targ}}^2 + \sigma_{\mathrm{L}}^2\right)}}\right]. \tag{6}$$

Late noise can be considered implicit whenever human performances correspond to threshold differences in a model mechanism's deterministic response (e.g., Betts et al., 2012; Solomon, 2022; Schallmo et al., 2018; Tadin & Lappin, 2005). In such cases, threshold is determined by the standard deviation of the implicit noise. Consequently, in the absence of external noise, the expected visual signal grows in proportion to the target area, not its square-root; and on log–log axes such as the panels of Fig. 2, this proportionality would manifest as a line with gradient -1, which is not well-supported by any of the datasets.

As an alternative to late noise, "early" Gaussian noise can corrupt visual signals before their velocity profiles are matched with the template. In this case we have

$$F_X(x) = \frac{1}{2}\mathrm{erfc}\left[\frac{-x + a_X \pi r_{\mathrm{targ}}^2}{\sqrt{2\pi\left(\sigma_{\mathrm{ext}}^2 + \sigma_{\mathrm{E}}^2\right) r_{\mathrm{targ}}^2}}\right]. \tag{7}$$

The variance of this early noise (σ_{E}^2) is functionally equivalent to an increase in the variance of motion noise in the stimulus. (It can be considered "equivalent input noise," e.g., Pelli, 1990.) Consequently, thresholds for an otherwise-ideal observer with early noise would fall with a gradient of $-1/2$, just like those for an ideal observer with motion noise in the stimulus.

Now consider what happens when external motion noise (i.e., random velocity fluctuations) is added to the stimulus field (i.e., $\sigma_{\mathrm{ext}} > 0$). When motion noise is large compared to the internal noise (early or late), the variance of visual signals will increase, and thresholds should rise. Sure enough, thresholds did rise when motion-defined targets were presented with random velocity fluctuations (see Fig. 3). However, so much internal noise (early or late) would be required to make the thresholds for this noisy-but-otherwise-ideal observer commensurate with those of our human observers that external noise would always remain negligible. Consequently, we infer that internal noise cannot be the only reason that human thresholds are so high.

[4] In the absence of velocity noise, the ideal observer's performance is limited only by quantal fluctuations. Compared to all the other noise sources discussed in this study, these fluctuations are literally negligible, which is synonymous with "effectively 0.".

3.1.3. The noisy, inefficient observer

Aside from internal noise, another potential explanation for the high thresholds of our human observers is that their detection processes have a less-than-ideal sampling efficiency because their templates never match the shape of the external signal (Burgess et al., 1981).

Here we consider a further modification to the noisy (but otherwise ideal) observer model with early noise, in which imperfectly matched templates nonetheless grow in proportion to the target, thereby achieving a constant sampling efficiency. Constant sampling efficiency could happen if each template were the product of a disc identical to the target and sparse array of local receptive fields (see Fig. 4). The coverage of that array (η) would be equal to the sampling efficiency (constrained to the interval [0,1]), and the distribution of visual signals would be

$$F_X(x) = \frac{1}{2}\mathrm{erfc}\left[\frac{-x + a_X \pi \eta r_{\mathrm{targ}}^2}{\sqrt{2\pi\eta\left(\sigma_{\mathrm{ext}}^2 + \sigma_{\mathrm{E}}^2\right) r_{\mathrm{targ}}^2}}\right]. \tag{8}$$

This noisy, inefficient, but-otherwise-ideal observer can produce both realistically high thresholds and realistically large effects of motion noise in the stimulus. Where this model fails is in its prediction of a constant log–log gradient of $-1/2$. Although thresholds for our human observers did fall at this rate when target area was less than approximately 3 deg^2, they reversed direction when target area exceeded this value.

3.1.4. The noisy, inefficient, size-limited observers

We consider two further modifications of the ideal observer model, in which its templates have a minimum and/or maximum size. These modifications can be achieved by adding free parameters r_{min} and r_{max} (in addition to σ_{E} and η) to the signal distributions:

$$F_X(x) = \frac{1}{2}\mathrm{erfc}\left[\frac{-x + a_X \pi \eta \min\left(r_{\mathrm{max}}^2, r_{\mathrm{targ}}^2\right)}{\sqrt{2\pi\eta\left(\sigma_{\mathrm{ext}}^2 + \sigma_{\mathrm{E}}^2\right) \min\left[r_{\mathrm{max}}^2, \max\left(r_{\mathrm{min}}^2, r_{\mathrm{targ}}^2\right)\right]}}\right]. \tag{9}$$

When $r_{\mathrm{min}} = 0$, $r_{\mathrm{max}} < \infty$, this noisy, inefficient, size-limited observer can produce bi-linear threshold-vs-size functions. The left side of each function displays the ideal observer's summation gradient of $-1/2$, but the right side is flat. Below, we consider the possibility that maximum size (r_{max}) varies with field size (i.e., the number of pixels in the image) and/or external noise (σ_{ext}^2). When $0 < r_{\mathrm{min}} < r_{\mathrm{max}} < \infty$, this noisy, inefficient, size-limited observer can produce tri-linear threshold-vs-size functions, where the gradient doubles (to –1) on the far-left side (as in Chen et al., 2019, for contrast summation).

3.1.5. Intrinsic uncertainty

Intrinsic-uncertainty (IU) models of detection (Pelli, 1985) specify CDFs for the target (where $X = S$) and nontarget (where $X = N$) as the product of M CDFs, such that

$$F_X(x) = \left[F_{\mathrm{IU}}(x;0)\right]^{M-K}\left[F_{\mathrm{IU}}(x;a_X)\right]^{K}. \tag{10}$$

In Eq. (10), $M-K$ represents the number of "irrelevant" visual signals, which are nonetheless considered in the decision process. K represents the number of "relevant" signals, so-called because their expected values increase with the target's amplitude a_S. Summation is modelled using the additional assumption that K is proportional to target area. To complete the specification of IU models, we adopt the conventional assumption that all K relevant visual signals can be represented by independent Gaussian-based random variables with equal mean and variance, i.e.,

$$F_{\mathrm{IU}}(x;a) = \frac{1}{2}\mathrm{erfc}\left(\frac{a-x}{\sqrt{2}\,\sigma}\right). \tag{11}$$

Note that this model has three free parameters that could potentially vary with field size and/or external noise: the ratio between the number of relevant, independent signals K and target area, the total number of independent signals M, and the variance of each signal σ^2. If the number of signals were independent of external noise, then we could expect the variance of those signals to increase (linearly) with the variance of external noise. However, signals will fail to remain independent when they arise from neurons (or pools of neurons) whose receptive fields overlap the same strip of 16 pixels to which a single sample of external noise has been added. Consequently, the *effective* values of K and M cannot plausibly increase with external noise, but they could decrease.

This implication is important because a decrease in the effective values of K and M would manifest as more precipitous decreases in threshold when plotted on log–log axes, such as those in Fig. 3.[5] No such increase in steepness can be seen in our thresholds. Indeed, the range of target areas over which the IU model can produce (81%-correct) threshold curves with a log–log gradient of –1/2 is only about one log unit ($K < M$ and $M \approx 10$).

For comparison with ideal observer models, we fit a simple, 4-parameter version of the IU model to each observer's data (see below). We have not evaluated the IU model further, since it cannot capture the elevation in threshold for targets having radial extents greater than 1°.

3.1.6. The DoG observers

To produce swoosh-like summation curves, a model must have some inhibitory component. The simplest such model we could imagine has a template formed by a difference of Gaussians (DoG). Watson and Eckert (1994) previously proposed DoG-shaped templates for motion detection, but they did not examine their implications for spatial summation. Each of our DoG templates is centered on the target. On average, visual signals could be expected to increase with the size of the moving target until its edge reached the zero-crossing of the DoG, after which the expected signal would fall. Formally, the behavior of this observer is based on signals described by the dot product between each stimulus and a DoG:

$$X = v\left(r; a_X, r_{\mathrm{targ}}, \sigma_{\mathrm{ext}}\right) \bullet \left(\exp\left[-r^2/\left(2r_{\mathrm{a}}^2\right)\right] - b\,\exp\left[-r^2/\left(2r_{\mathrm{b}}^2\right)\right]\right), \tag{12}$$

where r_{b} and r_{a} respectively denote the space constants of the DoG's inhibitory and excitatory component blobs and b denotes the ratio between their (unsigned) amplitudes. Unfortunately, there is no closed-form expression for this model's distribution of visual signals:

$$F_X(x) = \frac{1}{2}\mathrm{erfc}\left[\frac{-x + a_X \eta \sqrt{2\pi}\int_{-r_{\mathrm{targ}}}^{r_{\mathrm{targ}}} g(u)du}{2\left(\sigma_{\mathrm{ext}}^2 + \sigma_{\mathrm{E}}^2\right)\sqrt{\pi\eta\left(r_{\mathrm{a}}^2 - b r_{\mathrm{b}}^2\right)}}\right], \tag{13}$$

where

$$g(u) = r_{\mathrm{a}}\exp\left[-u^2/\left(2r_{\mathrm{a}}^2\right)\right]\mathrm{erf}\left[\sqrt{r_{\mathrm{targ}}^2 - u^2}\,/\left(r_{\mathrm{a}}\sqrt{2}\right)\right] - b r_{\mathrm{b}}\exp\left[-u^2/\left(2r_{\mathrm{b}}^2\right)\right]\mathrm{erf}\left[\sqrt{r_{\mathrm{targ}}^2 - u^2}\,/\left(r_{\mathrm{b}}\sqrt{2}\right)\right]$$

and erf is the error function. Numerical methods were required to evaluate the integral.

Initially, we considered only "balanced" DoGs centered on each target (such that their dot product with large, uniform stimuli would be zero), in which $b = r_{\mathrm{a}}^{-2}$. This option produced uniformly poor fits (see Table 2). Better fits were obtained when b and the model's other 4 parameters (σ_{E}, η, r_{a}, and r_{b}) were free to vary, but these fits remained inferior to those of the noisy, inefficient observer. Finally, we examined size-limited variants of the DoG observer, whose templates have a minimum and/or maximum size, the latter of which was allowed to vary with the size of the stimulus field (i.e., the number of pixels in the image) and/or external noise. For all templates, the ratio between inhibitory

[5] Concomitantly, they would manifest as shallower psychometric functions (see Pelli, 1985 Eqns. 5.4 and 6.1).

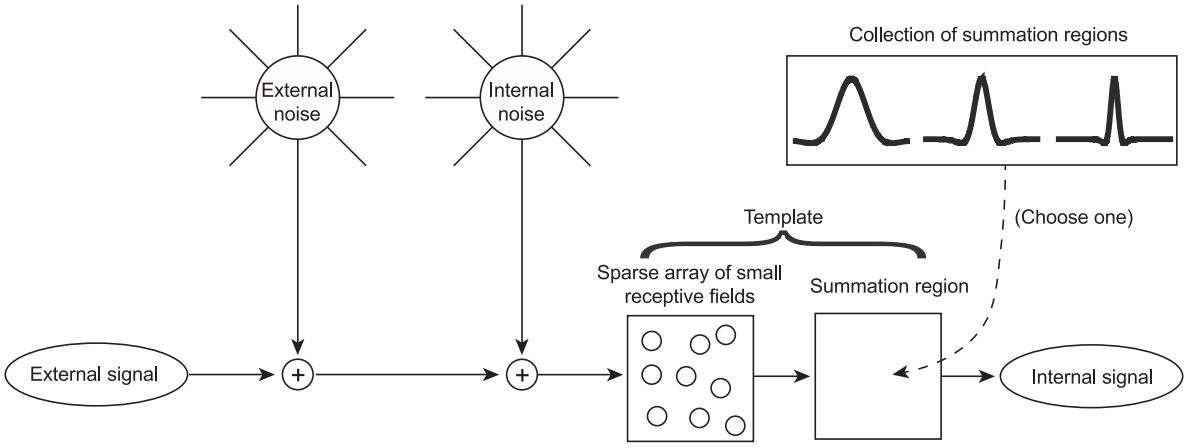

Fig. 4. Inefficient DoG-observer model with early noise. Model decisions are based on a comparison between two signals internal to the observer. Each of these signals has an expected value that is equal to the dot product between the external signal and something called a template. When the stimulus contains a motion-defined disk, the external signal has a disk shape. The amplitude of the external signal is proportional to the target's velocity. The template has a two-dimensional shape, just like the external signal. The presence of early noise causes the variance of internal signals to increase with template size, even when no velocity fluctuations are present in the stimulus.

Table 2

Pairwise comparison of maximum-likelihood fits to the data. Differences in AIC are shown for the four observers across each pair of models. For the model at the head of each column, values in that column represent substantial support (red), essentially no support (blue) or intermediate values (black). Conversely, blue values support the model at the head of their corresponding row, and red values represent essentially no support for it.

	01.LNO	02.NO	03.NIO	04.NISLO	05.NISLOM	06.NISLOS	07.NISLON	08.NISLOSN	09.NISLOSNM	10.IU	11.BDoGO	12.DoGO	13.BDoGS	14.DoGS	15.DOGN
02.NO	1516.7 911.1 772.3 559.														
03.NIO	1665.2 1196.4 853. 790.5	148.5 285.3 80.7 231.5													
04.NISLO	2712.8 1651.4 1216.9 1297.9	1196. 740.4 444.6 738.9	1047.5 455.1 363.9 507.4												
05.NISLOM	2710.7 1649.4 1212.8 1295.9	1194. 738.4 440.5 736.9	1045.5 453.1 359.8 505.4	-2. -2. -4.1 -2.											
06.NISLOS	2717.5 1693.6 1226.8 1304.4	1200.8 782.6 454.4 745.4	1052.2 497.3 373.8 513.9	4.7 42.2 9.8 6.5	6.7 44.2 13.9 8.5										
07.NISLON	2749.1 1650. 1223.5 1297.7	1232.4 738.9 451.2 738.8	1083.9 453.6 370.5 507.3	36.3 -1.5 6.6 -0.1	38.4 0.5 10.7 1.9	31.6 -43.7 -3.3 -6.6									
08.NISLOSN	2735.9 1704.9 1232.6 1308.6	1219.2 793.8 460.3 749.6	1070.7 508.5 379.6 518.1	23.1 53.4 15.7 10.7	25.2 55.4 19.7 12.7	18.4 11.3 5.8 4.2	-13.2 54.9 9.1 10.8								
09.NISLOSNM	2735.2 1704.6 1230.6 1312.7	1218.5 793.6 458.3 753.7	1070. 508.3 377.6 522.2	22.5 53.2 13.7 14.8	24.5 55.2 17.7 16.8	17.8 11. 3.8 8.3	-13.8 54.7 7.1 14.9	-0.7 -0.2 -2. 4.1							
10.IU	2634.1 1549.3 1202.7 1250.2	1117.4 638.2 430.4 691.3	968.9 352.9 349.7 459.8	-78.7 -102.1 -14.3 -47.6	-76.6 -100.1 -10.2 -45.6	-83.4 -144.3 -24.1 -54.2	-115. -100.7 -20.8 -47.5	-101.8 -155.6 -29.9 -58.3	-101.1 -155.4 -27.9 -62.4						
11.BDoGO	2071.2 1403.3 1217.2 1051.1	554.5 492.2 444.9 492.1	406. 206.9 364.2 260.6	-641.6 -248.2 0.3 -246.8	-639.5 -246.2 4.4 -244.8	-646.3 -290.4 -9.5 -253.3	-677.9 -246.7 -6.3 -246.7	-664.7 -301.6 -15.4 -257.5	-664.1 -301.4 -13.4 -261.6	-562.9 -146. 14.5 -199.2					
12.DoGO	2706.8 1648.3 1215.3 1295.2	1190.1 737.2 442.9 736.2	1041.6 451.9 362.3 504.7	-6. -3.2 -1.7 -2.6	-3.9 -1.2 2.4 -0.7	-10.7 -45.4 -11.5 -9.2	-42.3 -1.7 -8.2 -2.5	-29.1 -56.6 -17.3 -13.4	-28.4 -56.4 -15.3 -17.4	72.7 99. 12.6 45.	635.6 245. -2. 244.1				
13.BDoGS	2353.7 1675.7 1139.4 1267.5	837. 764.6 367.1 708.5	688.5 479.4 286.4 477.	-359. 24.3 -77.5 -30.4	-357. 26.3 -73.4 -28.4	-363.7 -17.9 -87.4 -36.9	-395.4 25.7 -84.1 -30.3	-382.2 -29.2 -93.2 -41.1	-381.5 -28.9 -91.2 -45.2	-280.4 126.4 -63.3 17.2	282.5 272.4 -77.8 216.4	-353.1 27.4 -75.9 -27.7			
14.DoGS	2712.2 1729.2 1226.6 1343.7	1195.5 818.2 454.2 784.7	1046.9 532.9 373.6 553.2	-0.6 77.8 9.6 45.8	1.5 79.8 13.7 47.8	-5.3 35.6 -0.2 39.3	-36.9 79.3 3.1 46.	-23.7 24.3 -6. 35.1	-23.1 24.6 -4. 31.	78.1 179.9 23.9 93.5	641. 325.9 9.3 292.6	5.4 81. 11.3 48.5	358.5 53.5 87.2 76.2		
15.DOGN	2743. 1648. 1219.7 1297.6	1226.3 736.9 447.4 738.6	1077.8 451.6 366.7 507.1	30.2 -3.4 2.8 -0.3	32.3 -1.4 6.9 1.7	25.5 -45.6 -7.1 -6.8	-6.1 -2. -3.8 -0.2	7.1 -56.9 -12.9 -11.	7.7 -56.7 -10.9 -15.1	108.9 98.7 17. 47.3	671.8 244.7 2.5 246.5	36.2 -0.3 4.5 2.3	389.3 -27.7 80.3 30.1	30.8 -81.2 -6.9 -46.2	
16.DOGSN	2736.3 1725.4 1224.5 1347.8	1219.6 814.3 452.2 788.9	1071.1 529. 371.5 557.4	23.5 74. 7.6 50.	25.6 76. 11.7 52.	18.8 31.8 -2.2 43.5	-12.8 75.4 1. 50.1	0.4 20.5 -8.1 39.3	1. 20.8 -6.1 35.2	102.2 176.1 21.8 97.6	665.1 322.1 7.3 296.8	29.5 77.1 9.3 52.6	382.6 49.7 85.1 80.4	24.1 -3.8 -2. 4.1	-6.7 77.4 4.8 50.3

and excitatory space constants remained fixed at $\rho_{ba} = r_b/r_a$. A diagram of the DoG observer is provided in Fig. 4.

3.2. Model fits[6]

Details, including parameter values, of maximum-likelihood fits to the data from each observer are available at http://www.staff.city.ac.uk/~solomon/MotionSummation.zip. Pairwise comparisons between these fits to the data appear in Table 2. The four numbers in each cell represent differences in Akaike Information Criteria (AIC)[7]; one difference for each observer (JAS, CWT, KZC, and FSN, respectively). Values less than 2 represent "substantial support" for the model of lower rank (i.e., the model named at the head of that column) over the model in the specified row (Burnham & Anderson, 2003, p. 70). These values appear in red. Values greater than 10 represent "essentially no support" for the model of lower rank. These values appear in blue. All intermediate values appear in black. The rankings, which are essentially nominal, appear in the leftmost column.

Models 1–9 are all based on the ideal observer, with disk-shaped summation regions matching each target profile. The acronyms for these 9 cases should thus be understood to include the implicit modifier that they are "otherwise ideal" in addition to the specified characteristics. The ideal observer's maximisation of signal-to-noise ratio requires those summation regions to change for each motion-defined target. Analogous changes in the templates of our models for human observers can be ascribed to an attentional strategy (Chen et al., 2019).

1. LNO, the (late) noisy observer. An otherwise-ideal observer with late noise (a.k.a. "decision noise"). One free parameter.
2. NO, the (early) noisy observer. An otherwise-ideal observer with early noise (a.k.a. "equivalent input noise"). One free parameter.
3. NIO, the noisy, inefficient observer. A noisy observer whose otherwise-ideal template covers a fixed proportion of the target's 16-pixel strips. Two free parameters.
4. NISLO, the noisy, inefficient, size-limited observer. A noisy, inefficient observer whose templates are limited to an arbitrary maximum size regardless of the size of the stimulus field. Three free parameters.
5. NISLOM, the noisy, inefficient, size-limited observer with a minimum template size. A noisy, inefficient observer, whose templates have both a maximum and a minimum size. Four free parameters.
6. NISLOS, the noisy, inefficient observer with size-specific template maxima. A noisy, inefficient observer whose largest templates are free to increase with size of the stimulus field (i.e., the number of pixels in the image). Two free parameters, plus one for each field size.
7. NISLON, the noisy, inefficient observer with noise-specific template maxima. A noisy, inefficient observer whose largest templates are free to increase with stimulus noise level. Two free parameters, plus one for each level of stimulus noise.
8. NISLOSN, the noisy, inefficient observer with size- and noise-specific template maxima. A noisy, inefficient observer whose largest templates are free to increase with field size and/or stimulus noise. Two free parameters, plus one for each combination of field size and stimulus noise.
9. NISLOSNM, the noisy, inefficient observer with noise- and size-specific template maxima and a minimum template size. A noisy, inefficient observer whose largest templates are free to increase with field size and/or stimulus noise, and whose smallest template is also limited to an arbitrary size (but is invariant with field size or stimulus noise). Three free parameters, plus one for each combination of field size and stimulus noise.

Intrinsic Uncertainty Model

10. IU, a simple form of the intrinsic-uncertainty model. Four free parameters remain invariant with field size and external noise: the ratio between the number of relevant independent signals and target area, the total number of independent signals, and two parameters describing the linear relationship between signal variance and external noise.

Models 11–16 employ Difference of Gaussian (DoG)-shaped summation regions that increase in size with a range of target sizes.

11. BDoGO, the balanced DoG one-max-fits-all model. Identical to the noisy, inefficient observer, except it uses DoG-shaped templates that are perfectly balanced. Five free parameters (early noise σ_E, coverage η, ratio of inhibitory:excitatory space-constants ρ_{ba}, ratio between excitatory space-constant and target radius ρ_{at}, and maximum excitatory space-constant r_{all}).
12. DoGO, the DoG one-max-fits-all model. Identical to the balanced DoG one-max-fits-all model except that its DoGs need not be balanced. Six free parameters. (In addition to BDoGO's five, there is another free parameter, b, for the ratio between inhibitory gain and excitatory gain.)
13. BDoGS, the balanced DoG model with size-specific template maxima. Identical to the balanced DoG one-max-fits-all model, except there is a maximum space-constant that is allowed to increase with field size. Four free parameters, plus one for each field size.
14. DoGS, the DoG model with size-specific template maxima. Identical to the DoG one-max-fits-all model, except there is a maximum space-constant that is allowed to increase with field size. Five free parameters, plus one for each field size (r_{256}, r_{512}, and/or r_{736}).
15. DoGN, the DoG model with noise-specific template maxima. Identical to the DoG one-max-fits-all model, except there is a maximum space-constant that is allowed to increase with the level of stimulus noise. Five free parameters, plus one for each level of stimulus noise.
16. DoGSN, the DoG model with size- and noise-specific template maxima. Identical to the DoG one-max-fits-all model, except there is a maximum space-constant that is allowed to increase with field size and/or stimulus noise. Five free parameters, plus one for each combination of field size and stimulus noise.

Note that every number in the first three columns of Table 2 is blue. Thus, our data provide essentially no support for LNO, NO, or NIO over any of the other models we consider. Those first three models are too restrictive. Similarly restrictive are all models with a minimum template size, and all balanced DoG models. Also take note of the comparison between DoGSN and DoGO. Here again, almost every number is blue (at 9.3, the comparison for KZC can be considered almost blue), indicating that DoGO is too restrictive. However, whereas it is primarily the restriction against noise-specific template maxima that hampers DoGO's fit to JAS's data, it is primarily the restriction against size-specific maxima that hampers its fit to data from CWT, KZC, and FSN.

Although the shape of summation curves can vary between models, when any given model's parameter values are fixed, all the curves predicted by that model are constrained to have the same shape. This shape can be thought of as a compromise between the different noise levels and field sizes. Fig. 3 shows that the maximum-likelihood fits of DoGS to JAS's and KZC's data produce "compromises" with relatively shallow swooshes. Best-fitting parameter values $[\sigma_E, \eta, \rho_{ba}, b, \rho_{at}, r_{256}, r_{512}, r_{736}]$ were $[0.28, 0.018, 1.6, 0.10, 1.1, 53, 60, 60]$ for JAS; $[0.20, 0.20,$

[6] This section may be skipped without loss of continuity.

[7] That is, $2(k_1 - k_2) + 2\ln(l_2/l_1)$, where k_1 and l_1 are the number of free parameters and maximum likelihood of the lower-ranked model (respectively) and k_2 and l_2 are the number of free parameters and maximum likelihood of the higher-ranked model.

1.2, 0.66, 0.75, 51, 80, *] for CWT; [0.32, 0.0096, 2.2, 0.077, 1.5, 83, 110, 120] for KZC; and [0.069, 0.022, 2.0, 0.18, 1.7, 75, *, 120] for FSN, where asterisks replace maximum space constants for unused field sizes.

4. Discussion

The primary result of this study is the finding that the maximum summation region for motion detection is about 3 deg^2 in the center of the visual field. This value is between 1 and 2 orders of magnitude larger than Riccò's area for luminance summation (possibly better understood as unsigned contrast summation, see Thibos et al., 2019). It corresponds to a retinal subtense of 2 deg, which is a good match for one-half the wavelength of the sinusoidally modulated flow field that was most easily detected by CWT (Nakayama & Tyler, 1981). It is also a fairly good match to the average receptive field size of foveal neurons in the middle temporal area of macaque, where even a single action potential can convey a substantial amount of directional information about slowly moving stimuli (Bair & Movshon, 2004). (Albright & Desimone's, 1987 regression of receptive field sizes against eccentricity indicates that this size should be approximately 1.2 degrees of visual angle.)

Much larger receptive fields have been inferred from psychophysical experiments with large-scale optic flow. Burr et al. (1998) recorded a roughly constant (~10%) sampling efficiency with which the direction of a target flow-field's motion could be identified, regardless of its retinal subtense. This result may seem inconsistent with the swoosh-like summation curves reported here and elsewhere (e.g., Tadin et al., 2003), but note that Burr et al. manipulated target subtense by changing the observer's viewing distance. Consequently, motion speed varied from approximately 12 deg/s for their large targets to 190 deg/s for their small ones. It seems likely that direction-discrimination of this relatively high-velocity optic flow engages different visual mechanisms than the detection of our low-velocity, motion-defined targets.

When our motion-defined targets were smaller than 3 deg^2, thresholds fell with with an increase in the square root of target area. This relationship implies a constant sampling efficiency and stands in sharp contrast to the thresholds for luminance-defined targets, which fall with target area (not its square-root) when that area is smaller than Riccò's area (Barlow, 1958). Constant sampling efficiency implies a capability of accessing templates whose sizes match the target sizes, even if those matches aren't perfect. Efficiency can be defined as the square of the ratio between ideal and human thresholds in high noise (Pelli, 1990). Perfect ("ideal") matches produce efficiencies of 100%. The largest we found (2.5%) was achieved by CWT with the smallest target in 40% noise. KZC never achieved better than 0.3% efficiency. (These numbers are well-matched by the best-fitting values for the coverage parameter in DoGS.) This coverage, which we attribute to a sparse array of small receptive fields, may correspond to the relatively sparse sampling area of a neuron's dendritic tree, compared to the size of visual field over which it extends.

4.1. Comparison with previous investigations of spatial summation with moving stimuli

Whereas our measurements of velocity threshold can be considered directly analogous to Riccò's measurements of luminance threshold, others have characterized spatial summation for motion detection using alternative independent variables. Perhaps most notable is a study by Tadin, Lappin, Gilroy, and Blake (2003), who reported swoosh-like summation curves of threshold duration for targets defined by random dots and drifting gratings. In the absence of external noise, our empirical results mirror theirs. However, in some cases, the detection of their drifting gratings was *facilitated* by the addition of external noise (randomized pixel contrasts). Unlike Tadin et al., we did not find any evidence for facilitation from external noise.

Several other laboratories have investigated spatial summation with moving targets. van de Grind et al. (1986) measured the widths of just-detectable, vertical strips of translating texture having various velocities against a static background. They also measured threshold signal-to-noise ratios for motion detection (1983). Like Tadin et al. (2003), Lappin et al. (2009) measured duration thresholds for direction discrimination with targets having various sizes. Both of these groups used target velocity as an independent variable. Our work can be considered complementary, as we report threshold velocities for targets of various size and signal-to-noise ratio.

As an index of the relationship between grating velocity and the effect of target size on the duration threshold for direction discrimination, Lappin et al. (2009) used log ratio between the thresholds for large (110 deg^2) and small (4.91 deg^2) targets. Consistent with the swoosh-like summation curves reported by Tadin et al. (2003), this index was positive for high speeds ($\geq$ 0.7 deg/s). However, it was close to zero for speeds in the range of the threshold velocities reported here. Prima facie, this behavior may suggest that our studies probe distinct visual mechanisms. However, there is no genuine inconsistency because our summation curves are defined by different dependent variables. To evaluate the suggestion of distinct mechanisms, we would need a model for how duration and velocity affect both the mean and the variance of the visual signals used for direction discrimination and motion detection.

Qualitatively similar to the monotonically decreasing duration thresholds for direction-discrimination with low-contrast gratings (Tadin & Lappin, 2005; Tadin et al., 2003), Anderson and Burr (1987, 1991), reported that contrast thresholds decrease monotonically, regardless of whether the task is detection or direction discrimination with a drifting Gabor target. Tadin and colleagues attributed the very different, swoosh-like shape of summation curves (mapping target size to duration threshold for direction discrimination) for high-contrast targets to their greater detectability.

Assuming that some form of surround suppression is responsible for the non-monotonicity of our summation curves (and, indeed, we offer no alternative explanation), our results indicate that highly visible targets are not required for its activation. Similar non-monotonicity can be seen in all our detection thresholds, including those obtained at relatively low signal-to-noise ratios. However, it must be noted that the "signal" and "noise" in our stimuli were not proportional to contrast or grey level; they were proportional to velocity. Therefore, there really is no conflict between our results and earlier literature.

Although Bell and Lappin (1973) argued that direction discrimination was a better test of motion processing than displacement detection, Nakayama and Tyler (1981) demonstrated that detection of oscillatory displacements in random-dot stimuli was a constant function of velocity for various temporal frequencies. In other words, motion-defined form could not be detected until the random dots that comprised it were moving with sufficient velocity. Identifiable displacements in random-dot stimuli were dubbed 'short-range apparent motion' by Baker and Braddick (1985). They suggested that the minimum-displacement thresholds were, "likely determined by receptive field properties in area 17."

4.2. Comparison with previous modeling efforts

Tadin and Lappin (2005) considered several models for the swoosh-like summation functions in the absence of external noise. All their models employ a sigmoidal transducer, without input from neighboring mechanisms. The relationship between parameter values and their model's behavior is somewhat opaque (as noted by Betts et al., 2012) because sigmoidal transduction (with its inherent divisive inhibition) is only the first stage and does not produce swoosh-like summation curves. To obtain swooshes, their models rely on a second stage with subtractive inhibition, a feature analogous to the inhibitory component of our DoG models. However, unlike our DoG models, it is not clear how to modify the formulas of Tadin and Lappin (or those of Betts et al.) for compatibility with external noise. Of course, it would be possible to calculate the distribution of signals in their excitatory channel as well as the

distribution of signals in their inhibitory channel, but unless both channels were linear (and Tadin and Lappin's formulas collapsed to a simple DoG model), those signals would be correlated with a coefficient whose calculation would be intractable.

Our DoGN model contains most of the spirit of Tadin and Lappin (2005) models. Specifically, it has an inhibitory surround and a template size that increases with decreased visibility (increased external noise). However, it only showed an advantage over the DoGS model for one of our four observers, so it was not well-supported in our implementation. We therefore regard the DoGS model with the fixed maximum template size, as plotted in Fig. 2, as the preferred model for human motion summation.

We concentrated on a class of linear models wherein the expected visual signal is given by the dot product between the spatial layout of stimulus velocities and an internal template mediating motion detection. This template might correspond to a single physiological receptive field or the superposition of several, but within the context of our psychophysical model it is merely a function mapping spatial position to a unitless quantity that can be considered the detector's gain. A summary of our modeling can be described as follows. If template size were a constant function of target size (as in the models described by Tadin & Lappin, 2005), then the model's thresholds would either fall too rapidly (over small sizes) or not at all (over large sizes), as the target size increased. Thus, we reasoned that template sizes must adaptively increase with (small-to-medium) target sizes, such that the perceptual system selected the best-matching available template for these target sizes. Also, if the variance of internal signals did not similarly increase with template (and target) size, then the model's thresholds would fall too rapidly to match the data. Thus, we inferred that internal noise affected signal variance in just the same way as external noise; it could be considered "early."

If target size never exceeded the template's central, excitatory region, then the model's thresholds would not rise to form the ascending parts of the swoosh-like curves seen in Fig. 3, implying that large targets must encroach upon the largest template's inhibitory surrounds. However, any simple difference-of-Gaussians template would produce either implausibly low thresholds or no effect of external noise. Therefore, to account for the inefficiency of the performance, we conjectured that each DoG-shaped template must linearly combine inputs from similarly sparse arrays of small receptive fields. Best fits (smaller AIC values) were obtained when larger DoG-shaped templates were allowed with larger field sizes.

CRediT authorship contribution statement

Joshua A. Solomon: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Fintan Nagle:** Writing – review & editing, Software, Investigation. **Christopher W. Tyler:** Writing – review & editing, Methodology, Formal analysis, Conceptualization, Investigation.

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.

Data availability

The data are available online. http://www.staff.city.ac.uk/~solomon/MotionSummation.zip.

Acknowledgment

The authors would like to acknowledge Kristina Zeljic for help with data collection.

References

Albright, T. D., & Desimone, R. (1987). Local precision of visuotopic organization in the middle temporal (MT) area of the macaque. *Experimental Brain Research, 65*, 582–592.
Anderson, S. J., & Burr, D. C. (1987). Receptive field size of human motion detection units. *Vision Research, 27*, 621–635.
Anderson, S. J., & Burr, D. C. (1991). Spatial summation properties of directionally selective mechanisms in human vision. *Journal of the Optical Society of America A, 8*, 1330–1339.
Bair, W., & Movshon, J. A. (2004). Adaptive temporal integration of motion in direction-selective neurons in macaque visual cortex. *Journal of Neuroscience, 24*, 7305–7323.
Baker, C. L., & Braddick, O. J. (1985). Eccentricity-dependent scaling of the limits for short-range apparent motion perception. *Vision Research, 25*, 803–812.
Barlow, H. B. (1958). Temporal and spatial summation in human vision at different background intensities. *Journal of Physiology, 141*, 337–350.
Bell, H. H., & Lappin, J. S. (1973). Sufficient conditions for the discrimination of motion. *Perception & Psychophysics, 14*, 45–50.
Betts, L. R., Sekuler, A. B., & Bennett, P. J. (2012). Spatial characteristics of motion-sensitive mechanisms change with age and stimulus spatial frequency. *Vision Research, 53*, 1–14.
Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433–436.
Burgess, A. E., Wagner, R. F., Jennings, R. J., & Barlow, H. B. (1981). Efficiency of human visual signal discrimination. *Science, 214*, 93–94.
Burnham, K. P., & Anderson, D. R. (2003). *Model selection and multimodel inference: A practical information-theoretic approach*. Springer Science & Business Media.
Burr, D. C., Morrone, M. C., & Vania, L. M. (1998). Large receptive fields for optic flow detection in humans. *Vision Research, 38*, 1731–1743.
Chen, C.-C., Yeh, Y.-H., & Tyler, C. W. (2019). Length summation in noise. *Journal of Vision, 19*(9), 1–13.
Graham, N. V. S. (1989). *Visual pattern analyzers*. New York: Oxford University Press.
Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
Lappin, J. S., Tadin, D., Nyquist, J. B., & Corn, A. L. (2009). Spatial and temporal limits of motion perception across variations in speed, eccentricity, and low vision. *Journal of Vision, 9*(1), 1–14.
May, K. A., & Solomon, J. A. (2013). Four theorems on the psychometric function. *PLoS One1, 8*, Article e74815.
Mayer, M. J., & Tyler, C. W. (1986). Invariance of the slope of the psychometric function with spatial summation. *Journal of the Optical Society of America A, 3*, 1166–1172.
Nakayama, K., & Tyler, C. W. (1981). Psychophysical isolation of movement sensitivity by removal of familiar position cues. *Vision Research, 21*, 4267–4433.
Neri, P. (2013). The statistical distribution of noisy transmission in human sensors. *Journal of Neural Engineering, 10*, Article 016014.
Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A, 2*, 1508–1532.
Pelli, D. G. (1990). The quantum efficiency of vision. In C. Blakemore (Ed.), *Vision: Coding and efficiency* (pp. 3–24). Cambridge: Cambridge University Press.
Riccò, A. (1877). Relazione fra il minimo angolo visuale e l'intensità luminosa. *Annali di Ottalmologia, 6*, 373–479.
Schallmo, M.-P., Kale, A. M., Millin, R., Flevaris, A. V., Brkanac, Z., Edden, R. A. E., Bernier, R. A., & Murray, S. O. (2018). Suppression and facilitation of human neural responses. *eLife, 7*, Article e30334.
Solomon, J. A. (2007). Intrinsic uncertainty explains second responses. *Spatial Vision, 20*, 45–60.
Solomon, J. A. (2022). An image-driven model for pattern detection, resistant to Birdsall linearisation. *Vision Research, 201*, Article 108121.
Tadin, D., & Lappin, J. S. (2005). Optimal size for perceiving motion decreases with contrast. *Vision Research, 45*, 2059–2064.
Tadin, D., Lappin, J. S., Gilroy, L. A., & Blake, R. (2003). Perceptual consequences of centre–surround antagonism in visual motion processing. *Journal of the Optical Society of America A, 36*, B35–B43.
Thibos, L. N., Bradley, A., Xu, R., & Lopez-Gil, N. (2019). Ricco's law and absolute threshold for foveal detection of black holes. *Journal of the Optical Society of America A, 36*, B35–B43.
van de Grind, W. A., Konderink, J. J., & van Doorn, A. J. (1986). The distribution of human motion detector properties in the monocular visual field. *Vision Research, 26*, 797–810.
van de Grind, W. A., van Doorn, A. J., & Konderink, J. J. (1983). Detection of coherent movement in peripherally viewed random-dot patterns. *Journal of the Optical Society of America A, 73*, 1674–1683.
Watson, A. B., & Eckert, M. P. (1994). Motion-contrast sensitivity: Visibility of motion gradients of various spatial frequencies. *Journal of the Optical Society of America A, 11*, 496–505.
Watson, A. B., & Pelli, D. G. (1983). Quest: A Bayesian adaptive psychometric method. *Perception and Psychophysics, 33*, 113–120.