City Research Online

City, University of London Institutional Repository

This is the published version of the paper.

This version of the publication may differ from the final published version.

Length summation in noise

Chien-Chung Chen

Department of Psychology, National Taiwan University, Taipei, Taiwan
Neurobiology and Cognitive Science Center, National Taiwan University, Taipei, Taiwan

Yu-Hsin Cynthia Yeh

Department of Psychology, National Taiwan University, Taipei, Taiwan

Christopher W. Tyler

Smith-Kettlewell Eye Research Institute, San Francisco, CA, USA
Department of Ophthalmology, City University of London, London, UK

To investigate the effect of background noise on visual summation, we measured the contrast detection thresholds for targets with or without a white noise mask in luminance contrast. The targets were Gabor patterns placed at 3° eccentricity to either the left or right of the fixation and elongated along an arc of the same radius to ensure equidistance from fixation for every point along the long axis. The task was a spatial two-alternative forced-choice (2AFC) paradigm in which the observer had to indicate whether the target was on the left or the right of the fixation. The threshold was measured at 75% accuracy with a staircase procedure. The detection threshold decreased with target length with slope −1/2 on log-log coordinates for target lengths between 30′ and 300′ half-height full-width (HHFW), defining a range of ideal matched-filter summation extending up to about 200′ (or about 16× the center width of the Gabor targets). The summation curves for different noise contrasts were shifted copies of each other. For the threshold versus mask contrast (TvN) functions, the target threshold was constant for noise levels up to about −22 dB, then increased with noise contrast to a linear asymptote on log-log coordinates. Since the "elbow" of the target threshold versus noise function is an index of the level of the equivalent noise experienced by the visual system during target detection, our results suggest that the signal-to-noise ratio was invariant with target length. We further show that a linear-nonlinear-linear gain-control model can fully account for these results with far fewer parameters than a matched-filter model.

Introduction

The visibility of a sufficiently small visual target increases with its size (Barlow, 1958; Baumgardt, 1959). This effect, known in the literature as spatial summation, takes several forms. In the first, called Ricco's law (Barlow, 1958; Baumgardt, 1959) or complete summation, target detection is inversely proportional to target size. This law applies when the target size is small enough to fit into the spatial extent of the smallest receptive field of the target detector mechanisms (assuming that the receptive field is linear and that the noise within the target detector is independent of the visual stimulation). Enlarging target size would increase the overlap between the target and the receptive field and in turn the response of the target detector in proportion. However, since the noise level within the target detector is independent of the target size and thus constant throughout the experiment, the signal-to-noise ratio experienced by the visual system should increase linearly with target size. Thus, assuming a linear transducer function, the threshold should decrease with target size with a slope of −1 on log-log coordinates as long as the stimulus is smaller than the receptive field. Thus, the upper limit of such summation has been used as a psychophysical index of the limiting receptive field size for contrast processing (Watson, Barlow, & Robson, 1983; Polat & Tyler, 1999; Tyler & Chen, 2006; Kao & Chen, 2012).

In the second type of spatial summation, called Piper's law the target threshold decreases with the square root of target size. To explain this law, one can assume that the observer has complete knowledge of

the stimulus and uses a single matched filter to detect its presence (Wiener, 1949). Thus, when the stimulus size increases, the optimal observer would just employ a filter with a larger receptive field to detect it. This optimal strategy can be implemented as a filter receiving inputs from an array of discrete sensors with independent and identically distributed noise sources (Tyler & Chen, 2000). A larger filter would allow proportionately more sampling on the target within its receptive field (Green & Swets, 1966). The response of the matched filter to the target is proportional to the sum of the responses to the individual samples. Thus, assuming that there are summation mechanisms matching each stimulus size, and that the noises in the local filters are independent and identically distributed, the mean and the variance of the matched filter response should increase proportionally with target size.

Now, in signal detection theory (Green & Swets, 1966), the detectability of a target depends on the ratio of the mean to the standard deviation of the response. The threshold should thus decrease with the square root of target size, giving a −1/2 log-log slope summation function over the range of available matching filters. It is suggested that such matched filters can be implemented as second-order filters[1] combining numbers of local filters in proportion to target size (Tyler & Chen, 2000). Some formulations thus simply use the square root of the number of local channels as the dominator of d' calculation (see Kingdom, Baldwin, & Schmidtmann, 2015). Such matched-filter behavior may also be called "ideal summation" behavior (Tanner & Jones, 1960; Tyler & Chen, 2000).

Another type of spatial summation is often termed "probability summation" (Pelli, 1985; Meese & Summers, 2012; Kingdom et al., 2015), which is governed by a selection or attention mechanism that identifies the channels with the maximum signal-to-noise ratio across the array of stimulated local channels (rather than summing their outputs linearly). It is thus better termed "max-rule summation." This process was first analyzed in relation to the two-alternative forced choice (2AFC) paradigm by Tyler & Chen (2000), who showed that under a variety of attention models the summation curve has a maximum slope of −1/4 on log-log coordinates, and decreases as the number of local units being summed increases. Here, we note that the same mechanism could operate over the outputs of multiple parallel second-order filters.

One can assume that the decision mechanism has access to a fixed number of independent detectors and is able to detect the target when the response of any of the monitored detectors is larger than a criterion. Thus, the 2AFC detectability of the target depends on the monitored detector with the greatest signal-to-noise

ratio. As target size increases, there are more monitored detectors with an increased response and thus the chance that at least one detector has a large enough response also increases. Since the detectability is determined by the detector with the largest response, the probability of target detection is then governed by the maximum order statistics, whose mean increases and the standard deviation decreases with the number of channels (Pelli, 1985; Chen & Tyler, 1999b). Such behavior, as shown by computer simulation (Tyler & Chen, 2000; Meese & Summers, 2012) displays the property that detection threshold decreases with target size with a log-log slope of at most −1/4 (see Tyler & Chen, 2000, for detailed treatment).

Note that different types of spatial summation are based on how the noise experienced by the visual system varies with target size. Thus, an estimation of noise across the range of target sizes would be a good test for the theories of spatial summation. In the literature, it is suggested that the equivalent noise method (Nagaraja, 1964; Kersten, 1984, 1987; Legge, Kersten, & Burgess, 1987; Pelli, 1990; Lu & Dosher, 1998, 2008; Pelli & Farell, 1999) can be a useful tool for internal noise estimation.

In a typical equivalent noise paradigm, the experiment measures the 2AFC threshold for a target under various amount of external noise, typically implemented as random-dot patterns at various levels of contrast. Thus, the detection is limited by both specified external and prevailing internal noise sources. When the external noise level is low, visual performance would be limited by the intrinsic noise of the visual mechanisms. Progressive increases in external noise from this low level would have little effect on the detection threshold until it approximates that of the internal noise. On the other hand, when the external noise level is high, it would swamp the effect of the intrinsic noise and become the limiting factor for detection. In this high-noise regime, the detection threshold would increase in proportion to the external noise level. As a result, the target threshold versus noise intensity (TvN) function is expected to be flat at low external noise contrast and to rise linearly with noise contrast at high contrasts. The "elbow," or the transition between the two regimes, demarcates the level where the external noise equals that of the intrinsic noise. Thus, if the threshold change with size, as discussed already, reflects the change of the signal-to-noise ratio experienced by the detection mechanism, one should observe a change of elbow position with target size.

Past studies on spatial summation have not fully addressed this issue. Kersten (1984) measured the detection threshold of Gabor patterns of different spatial frequency on high contrast noise. However, he used only one noise level and thus the data were unable
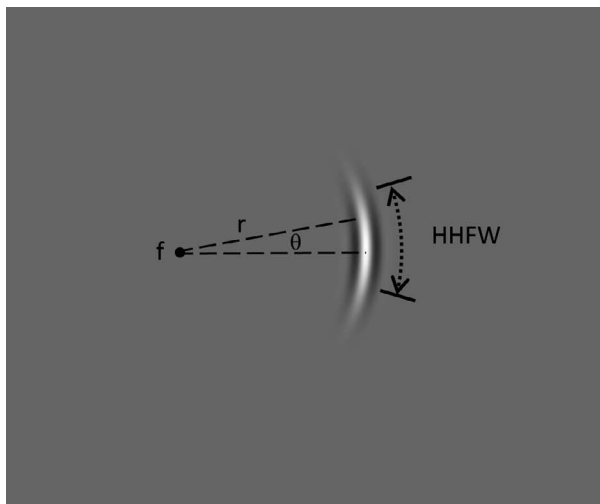
(A)
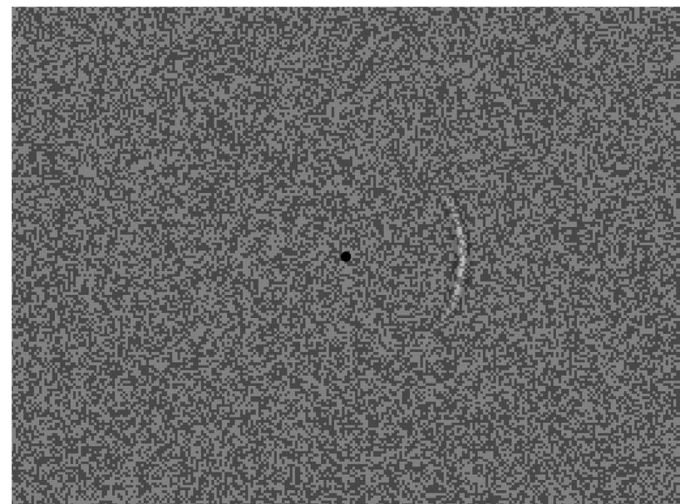
(B)

Figure 1. (A). The targets were Gabor arcs, centered at the fixation point, f, of different lengths defined in polar coordinates with radius r and angle θ. For a better comparison with the literature, we reported the target size in HHFW, or half-height full width. (B) The target embedded in a binary noise mask.

to provide an estimation of equivalent noise. Nagaraja (1964) measured the threshold of disks on various levels of noise. However, he used only three target sizes spread over a wide range. Thus, it is unclear which spatial summation regime they were in. Kersten (1987) varied the background noise level for the detection of noise targets, concluding that the human visual system had an unexpectedly high efficiency that could be adapted to the scale of the stimuli, but only used two stimulus sizes. We thus overcame the limitations of previous studies by using a sufficient number of levels of both target size and noise level for a quantitative analysis of the equivalent noise behavior for contrast processing.

Method

Apparatus

All stimuli were presented on two Viewsonic pf75+ 15-in. monitors controlled by a Radeon 7200 video card on an Apple MacPro computer. The Radeon video card (Radeon Technologies Group, Sunnyvale, CA) provided a 10-bit digital-to-analog converter depth that allowed an accurate representation of visual stimuli at low contrasts. We presented the mask on one monitor and the target on the other. Lights from the two monitors were combined by a beam splitter so that the observer would experience the stimuli from the two monitors superposed. This arrangement provided the

advantage of independent control of the contrast of the target and the noise mask. The viewing field was 13.9° (horizontal) by 10.4° (vertical) with a resolution of 1,024 horizontal by 768 vertical pixels, giving 48 pixels/° at the 135 cm viewing distance. The refresh rate of the monitors was 85 Hz, with a mean luminance of 24.25 cd/m^2. The luminance of the monitor at each output setting were measured with a PhotoResearch PR655 radiometer, which was used to construct a linearized lookup table.

Stimuli

The target (Figure 1) was an elongated Gabor patch arced along the circumference of an invisible circle of 3° radius centered at the fixation (f in Figure 1A). That is, the target was defined by

$$G(r, \theta; \sigma_\theta, c_t) = L + L * c_t * \cos(2\pi f r) \\ * \exp\left(-\frac{(r-u)^2}{2\sigma_r^2}\right) * \exp\left(-\frac{\theta^2}{2\sigma_\theta^2}\right),$$

where r and θ were the radius and the angle of a pixel in the polar coordinates (as noted in Figure 1A); L was the mean luminance of the display; c_t was the contrast of the target; f was the spatial frequency of 2.5 c/°; σ_r and σ_θ were the scale parameters (standard deviation) of the Gaussian envelope along the radius and circumference respectively; and the center of the Gabor arc, u, was at +3° and −3° visual eccentricity for patterns presented on the left or right of screen respectively. The parameters σ_r and σ_θ controlled the

size of the stimuli. The value of σ_r was fixed at 0.14° of visual angle while σ_θ varied from 0.5 to 40° along the circumference. Thus, the targets were Gabor arcs with a range of lengths from 1° to 80° circumferential angle. This arrangement allowed the stimuli to remain at the same cortical magnification factor regardless their length (see Robson & Graham, 1981). For better comparison with the results of the previous studies, we subsequently specify the length of the targets by their half-height at full-width (denoted HHFW in Figure 1A) of the Gaussian envelope, which is $(-\ln(0.5){*}2)^{0.5}$ $*u * \sigma_\theta = 7.06 * \sigma_\theta$, where σ_θ is in arc radians and HHFW is in degree of visual angle. Figure 1B illustrates what an observer might see in a high-noise trial.

The masks were binary noise patterns covering the whole display. The luminance of each pixel was drawn from a binomial distribution with equal probability outcomes and scaled by a contrast factor, c_m. That is, the noise mask

$$N(x, y) = L + L * c_m * B(0.5, x, y),$$

where c_m denotes the mask contrast and B denotes sampling from a binomial distribution. The noise was updated for each trial.

Procedure

We used a spatial 2AFC to measure the target threshold. On each trial, the target was presented either to the left or to the right of the fixation. On each trial, a 200 ms long auditory tone signaled the start of the trial, followed by a 295 ms stimulus presentation, and then by the response interval lasting until a valid response was recorded. The next trial started 800 ms after the response. The task of the observer was to respond whether the Gabor arc target was to the right or the left side of the display.

In each run, the noise contrast and the target size were the same, although the noise pattern updated for each trial. There were six possible noise contrasts ($20{*}\log_{10}(c_m) = -\infty, -26, -22, -18, -14, -10$ dB) and 10 target sizes, making a total of 60 test conditions. The target contrast in each trial was determined by the Ψ threshold-seeking algorithm (Kontsevich & Tyler, 1999), which was to measure the threshold at 75% correct response level. There were 40 trials following two practice trials in each run. Each reported datum point was an average of four repeated measures. The sequence of test conditions and repetition were all randomized.

Two observers participated in this study: YYH was one of the authors and DTJ was a paid observer naïve to the purpose of the experiment. Both participants had normal or corrected-to-normal visual acuity (20/20).

Results

Length summation

Figure 2 shows the length summation functions, or the variation of threshold as a function of Gabor arc length, at a range of noise levels. Each row shows the data for one observer, while each column shows the data split into two subsets for clarity. The smooth curves are fits of our model discussed as follows. The following qualitative characterization of the data is supported by the statistical assessment through the model fitting described in the Discussion.

Above about −20 dB, increases of mask contrast increased detection threshold. Conversely, the detection threshold decreased with target size with a log-log slope of approximately −1/2 (thick dashed line) at all noise levels. Such −1/2 slope summation is commonly observed in spatial summation with periodic patterns (Polat & Tyler, 1999; Tyler & Chen, 2006; Kingdom et al., 2015), faces (Tyler & Chen, 2006) and texts (Kao & Chen, 2012). There were telltale signs of the function asymptoting to a −1 slope (dotted line in Figure 2) for small target sizes and a trend of flattening more than the slope of −1/2 for the largest sizes. Such flattening may be consistent with the −1/4 slope summation observed with spatial summation with random dots (Tyler & Chen, 2006) or periodic pattern extended along the axis orthogonal to its orientation (Robson & Graham, 1981; Polat & Tyler, 1999). To quantify these trends, we fit a generic three-component model (Tyler & Chen, 2006), which considers all three types of spatial summation, to the data. That is,

$$\begin{aligned}
\text{threshold} \\
= \Big((a_1 - \log(\text{size}))^4 + (a_2 - 0.5 \log(\text{size}))^4 \\
+ (a_3 - 0.25 \log(\text{size}))^4 \Big)^{1/4}
\end{aligned}$$

where threshold is in dB units, and a_1, a_2, and a_3 are intercepts for each component. This model, with 18 free parameters for each observer, explains 97% of the variation in the averaged data with root mean squared error (RMSE) 0.97 dB overall. Removing the −1/2 slope component in the model degrades the best fit dramatically (RMSE 1.48, $F(12, 82) = 9.03$, $p < 0.0001$). Removing the −1 slope component but keeping others in the model reduces goodness-of-fit (RMSE 1.19, $F(12, 82) = 3.38$, $p = 0.0005$). Removing the −1/4 component, on the other hand, shows a barely significant effect on the fit (RMSE 1.13, $F(12, 82) = 2.4$, $p = 0.01 = \alpha$). Thus, there was strong support for full summation (for small Gabor arc lengths) and weak support for attentional, or probability, summation (for large Gabor arc lengths).
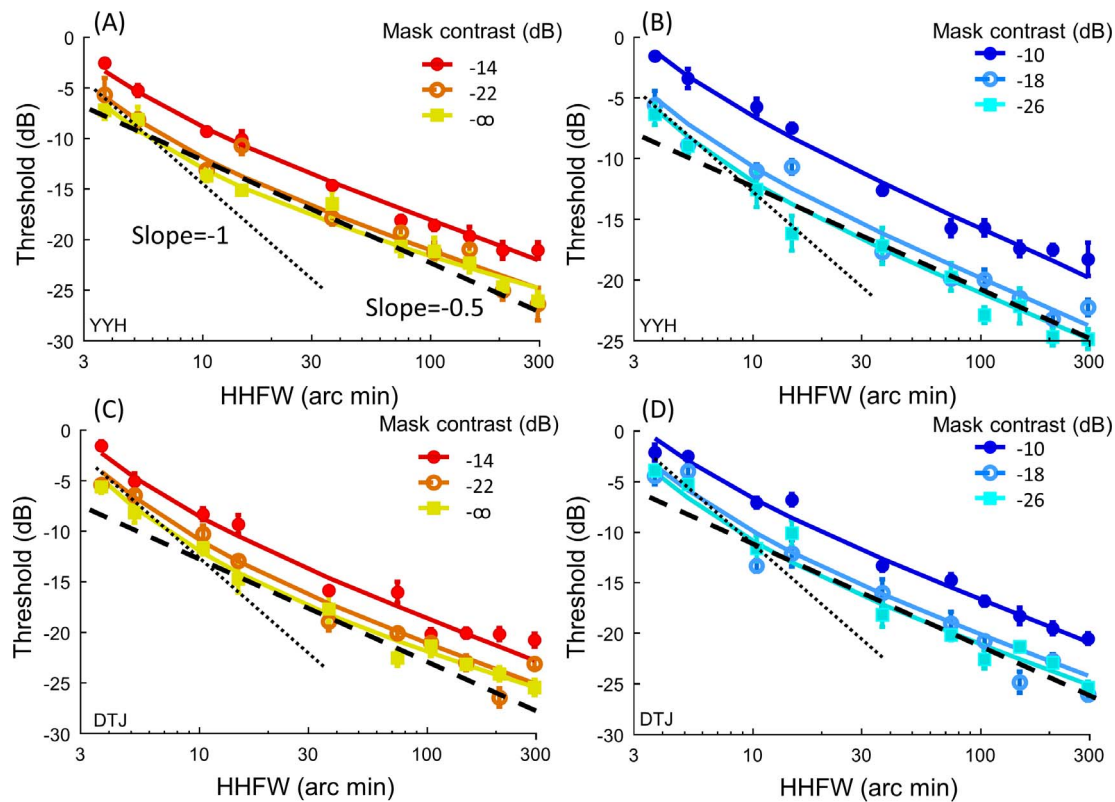
Figure 2. Gabor arc length summation functions at different noise level. The smooth curves are fits of our model. The dashed line denotes the −1/2 slope threshold reduction while the dotted line has a −1 slope. The error bars are one standard error of measurement.

For illustrative purposes, Figure 3 shows the fit of the three-component model to the summation curve averaged across mask contrast and observers. As expected, the full summation (−1 slope) and attentional summation (−1/4 slope) are at the extremes of the curve. Without the ideal summation component (−1/2 slope), it cannot capture the nature of the data. One can ask whether our data may be accounted for by a model with a single channel with limited receptive field size. In such a case, the summation curve would have a slope transitioning directly from −1 to 0. However, as shown in the blue curve in Figure 3, this model deviates from our data dramatically further than other generic models tested here. Finally, fitting the data with just the ideal summation model of a channel matching every stimulus size, giving a −1/2 slope, can explain as much as 96.4% of the variability, with RMSE 1.25. Thus, the summation data are mostly in the ideal-summation range, though with some deviation at the extremes.

Target threshold versus noise mask contrast (TvN) functions

Figure 4 shows the TvN functions for the two observers as a function of target size. Different colors and symbols represent TvN functions measured for the various target sizes. The smooth curves are the fits of

our model described below. At all target sizes, the TvN functions were flat at low mask contrast and then increased with a slope of about 0.5 at high mask contrast (Figure 4A). This functional form is consistent with previous noise masking studies (Legge et al., 1987; Pelli, 1990; Lu & Dosher, 1998).

The mask contrast at which the "elbow," or transition from the flat to the rising segment, occurs is usually considered to be a measure of the level of intrinsic noise. The formulation, according to Legge et al. (1987), is $E = k (N_{eq} + N_{ext})$, where E, or threshold energy, is the squared threshold; N_{ext}, or the external noise, is the squared masking noise contrast; and k, a scale constant, and N_{eq}, the equivalent noise estimate, are the two free parameters for each curve. We estimated the elbow for each TvN function by fitting the equivalent noise model to each TvN function. Figure 5A shows the equivalent noise fits for the same data as in Figure 4A. The R^2 of the fits was 0.97. We thus obtained a reliable estimation of the equivalent noise according to this standard model.

Figure 5B shows the estimated level of the equivalent noise, expressed in mask contrast s (i.e., the "elbow"), as a function of target size for the two observers (YYH: blue circles and DJT: red crosses). The equivalent noise estimate was practically constant for all target size: The slopes of the regression lines in Figure 5B were not significantly greater than zero (two-tailed $t(1) = -0.18$,

$p = 0.44$ for YYH and $t(1) = -1.26$, $p = 0.21$ for DTJ). As a result, the TvN functions for different target sizes appeared as vertically shifted copies of each other. We thus conclude that there was no evidence of any dependence of the level of equivalent noise on target size. Nagaraja (1964), as re-analyzed by Pelli (1990), also showed that equivalent noise did not change with the area of a disk.

Supplementary material

Supplementary File S1 is a Microsoft Excel file that tables all our data. Different sheets are data for different participants respectively. In each sheet, the first column is target size, the second, pedestal contrast, the third, measured target threshold, and the fourth, the psychometric function slope estimated by PSI. Each condition has four measurements, one for each row.

Discussion

In this study, we measured the contrast detection threshold for Gabor arc targets of various lengths embedded in different levels of noise. We found that, at all noise mask contrasts, the contrast detection thresholds decreased with target size with a slope approximating $-1/2$ on log-log coordinates. Thus, most of our summation data were consistent with Piper's law and hence were in the ideal summation range. As discussed earlier, a key implication of these spatial summation functions, given the conventional theory (Tanner & Birdsall, 1958; Tyler & Chen, 2000), is that the noise limiting the visual processing for target detection increases with target size at a rate approximating the square root of the signal strength.

The shape of the TvN functions is typical of the noise masking literature (Kersten, 1987; Pelli, 1990; Lu & Dosher, 2008): the target threshold is flat at low mask contrast and rises up when the mask contrast is greater than a critical value. The noticeable feature of our data is that all TvN functions started to rise at similar mask contrasts. As a result, the TvN functions for different target sizes are essentially vertically shifted copies of each other on log-log coordinates. Since the "elbow" of the TvN function indicates the level of equivalent noise estimated in conventional equivalent noise theory (Legge et al., 1987; Pelli, 1990), our TvN function result would imply that the internal noise does not change with target size.

The usual interpretation of the $-1/2$ slope summation derived from the ideal observer analysis (Tanner & Jones, 1960; Green & Swets, 1966) is that the observer has complete knowledge of the stimulus and uses a
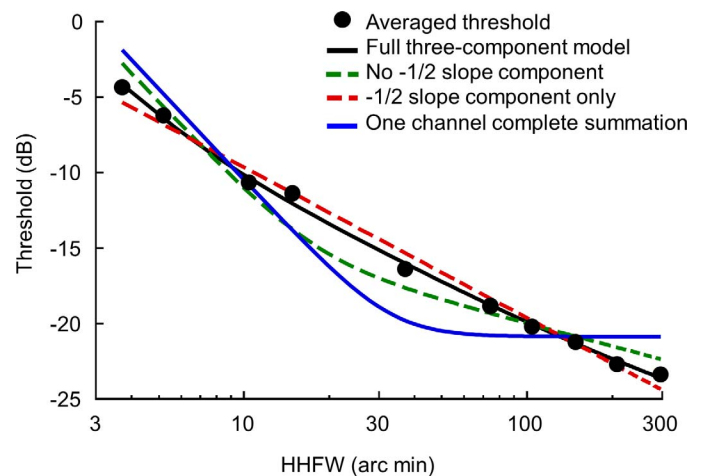

Figure 3. Example fits of four generic models to thresholds as a function of Gabor arc length, averaged across mask contrasts and observers (black dots). The ideal observer model of a $-1/2$ slope (red dashed line) deviates mildly at the two ends of the curve. A model of a single channel size with attentional summation across space (green dashed curve), implying the transition from a slope of -1 to $-1/4$, gives a strongly different function shape from the data. A similar one single channel model without attentional summation (blue solid curve), implying a slope transition from -1 to -0, gives an even worse fit. Only a three-component model with slopes of -1, $-1/2$, and $-1/4$ (black solid line) gives an accurate account of the summation behavior over the full range of sizes.

matched filter to detect its presence (Wiener, 1949). As discussed in Introduction, both signal and noise level (expressed as variance) in the system increase proportionally with size (Tyler & Chen, 2000). In the common formulation of equivalent noise analysis, $E = k (N_{eq} + N_{ext})$, where E, or threshold energy, is the squared threshold; N_{ext}, or the external noise, is the squared masking noise contrast; k, a scale constant, and N_{eq}, is the equivalent noise estimate. Here, the equivalent noise should contain all the noise experienced by the system during the detection task except that from the external noise mask. At first glance, since the noise in the system increases with Gabor arc length, one would expect the "elbow" position to increase with target size (shown as the black dashed line in Figure 5B). However, this prediction requires an independent sampling of noise from the target and the mask, in which the noise from the target increases with size while that from the mask depends only on its contrast. This is possible only if the dominant noise in the system during target detection is intrinsic to the matched filter *after* the sampling of the stimuli, or late noise. In this scenario, the increase of noise with size is solely due to the increase of the matched filter size. However, this late noise prediction is inconsistent with our result that, as shown in Figure 5, the equivalent noise estimate
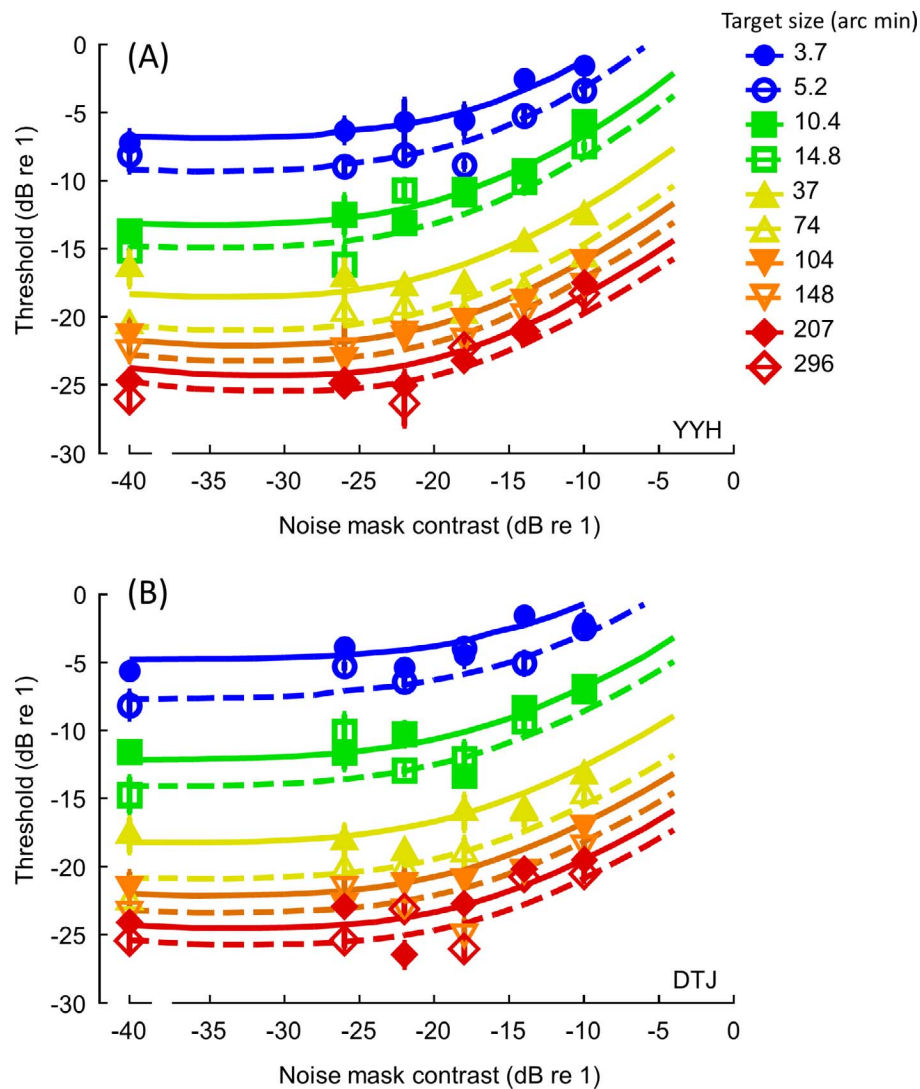
Figure 4. The target threshold versus mask contrast (TvN) functions for two observers (Panels (A) and (B) respectively) and at various Gabor arc lengths (see legend). The error bars represent 1 standard error of measurement. The smooth curves are our full model fits.

derived from the "elbow" position was the same for all target sizes we tested.

One way to resolve this conflict is to assume that the dominant noise occurs at the sampling of the stimuli. That is, as the target size increases, the system is able to take more samples on the target. However, since the target is embedded in the noise mask, the mask would affect every sample of the target. Thus, the effect of the noise mask should also increase with target size. Since the effects of noise produced by both target and mask increase with the square root of target size, the noise from the mask would surpass that from the target at the same mask level regardless target size. As a result, the "elbows" of the TvN functions would occur at a constant size regardless target size. However, this solution would imply that the effect of external noise is target dependent. Accepting this extension would imply that the system is operating on the signal-to-noise ratio

of the optimum filter size rather than a fixed noise threshold, which is natural behavior in a 2AFC task.

On the other hand, Lu and Dosher (1998) proposed a gain-control model for noise masking, in which the target pattern was first processed by a linear template, with the d' being controlled by the template output divided by the square root of the sum of all stimulus-related noise sources. This model offers more flexibility in interpreting data than the conventional equivalent noise model. For instance, our result of vertical shift of the TvN functions could be explained as due to different sensitivities of the visual system to targets of different size without a change in the internal noise. Such sensitivity changes are also consistent with the decrease of threshold with target size. However, the Lu and Dosher model offers no explanation as to why the threshold reduction should have a slope of −1/2. To account for this, it would be necessary to implement a
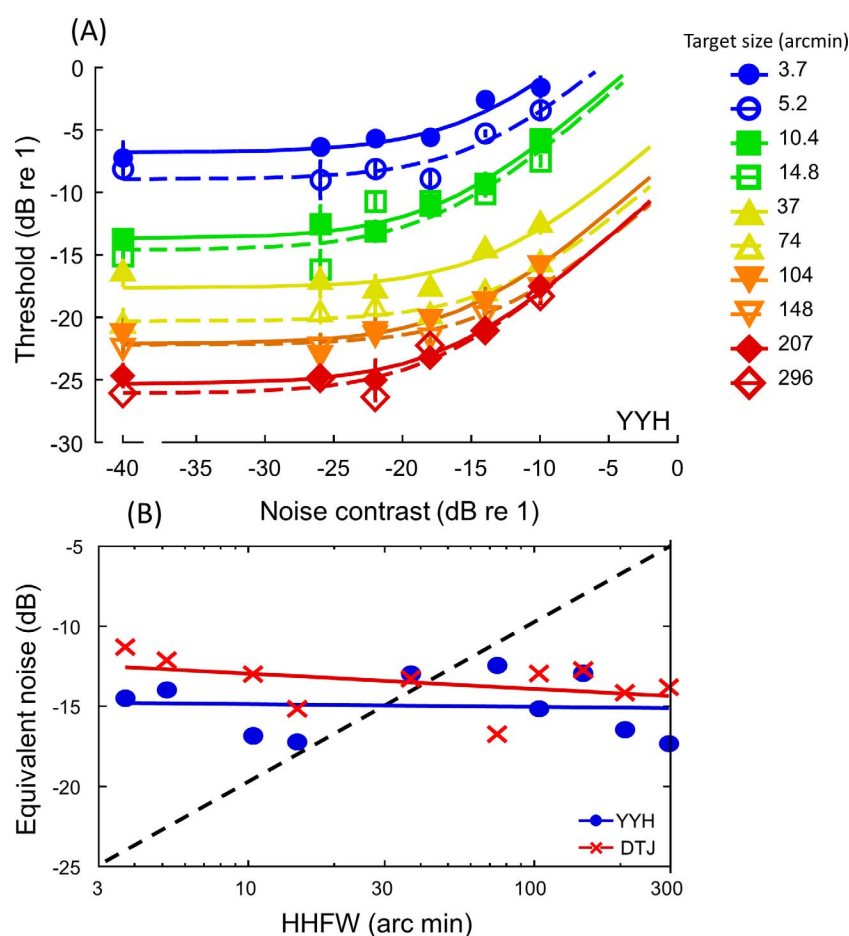
Figure 5. (A) The equivalent noise analysis fits to the same data as Figure 4A. (B) The resulting equivalent noise estimates are expressed in mask contrast as the function of target size. Blue circles: YYH; red crosses: DTJ. Blue and red solid lines are regression lines for YYH and DTJ, respectively. Dashed line shows the prediction from the late noise ideal summation, in which the noise experienced by the target detector increases with the square root of size. See Discussion for detail.

deus ex machina with prior knowledge of the stimulus size to control internal noise sources to match to stimulus extent, which is an arbitrary construct in the absence of further assumptions.

Recently, multiple stage models have been proposed to explain various spatial summation phenomena (Baker & Meese, 2011; Meese & Summers, 2012; Kingdom et al., 2015). Although there are variations among the different models, they generally start with a band of local filters followed by a nonlinear transform, and an array of summation mechanisms pooling the nonlinear responses of a number of local filters. In particular, Meese and Summers (2007) incorporated local contrast gain control into their model, to account for the masking effects of the pedestal with various target-pedestal size combinations (although their later 2012 model eliminated this gain control stage). This kind of mechanism is an implementation of the divisive transducer function formalism of Legge and Foley (1980) or Foley (1994), although the numerator of their response functions (Meese & Summers, 2007, Equation

B1) would imply the target and the pedestal being processed by different local mechanisms. However, those models were designed to account for spatial summation for the target alone (Kingdom et al., 2015) or the target on a patterned pedestal (Meese & Summers, 2012). It is unclear how these models can explain the effect of a noise mask.

Model

To fit our new results, we propose a straightforward model that can account for both for noise masking and spatial summation. This model contains three elements: (1) the contrast normalization that accounts for pattern detection; (2) "ideal" summation; and (3) decision making under noise. The combination of elements (1) and (3) would be similar to the model used by Chen and Tyler (2010) or Chou, Yeh, and Chen (2014) to account for the TvN functions for the noise masking paradigm, while element 2 accounts for the areal summation
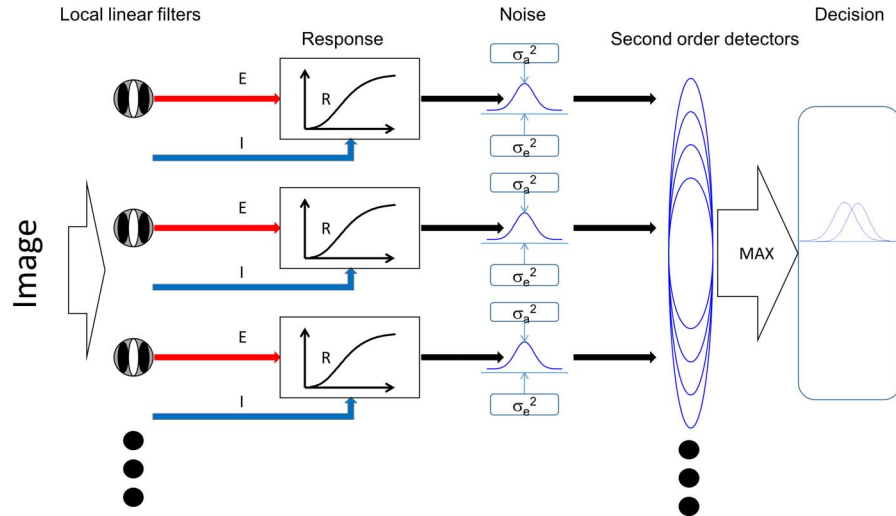
Figure 6. Diagram of the gain control model. See text for details.

behavior. Figure 6 shows a diagram of the stages of this model, which characterizes the local processing for each point in the visual field. Thus, all the terms in the model have implicit (x, y) subscripts for their visual field location.

The first stage is a set of linear filters each with its own location and form selectivity. The excitation of a linear filter is then half-wave rectified, raised to a power and scaled by a divisive inhibition input to form the normalized contrast energy response of each local filter. (The divisive inhibition may also be conceptualized as Bayesian estimator of the recent past information reaching this detector.) A set of second-order detectors each sums inputs from the responses of local filters. Due to this summation process, the second-order detectors also combine the noise from all their local filters, whether it is external or intrinsic to the local filter. Thus, the signal-to-noise ratio in the second-order detectors is then the combined with the local filter responses and divided by the combined noise. The decision variable is determined a max rule to identify the particular second-order detector with the greatest signal-to-noise ratio.

In detail, when image information arrives at the input, it is first processed by local filters. Each local filter j has a spatial sensitivity profile $f_j(x,y)$. The excitation of this linear filter to the i-th image component $C_i g_i(x,y)$, where $g_i(x,y)$ defines the contrast independent spatial variation and i can be either the target or the mask in our experiment, is given as

$$E_{ij}' = \Sigma_x \Sigma_y C_i g_i(x, y) f_j(x, y) \quad (1)$$

Summing over x and y, eq. (1) can be simplified to

$$E_{ji}' = Se_{ji} C_i \quad (1')$$

where Se_{ji} is a constant defining the excitatory

sensitivity of the mechanism to the stimulus, and $i = t$ or m, for target and mask, respectively.

The output of the linear filters is half-wave rectified (Foley, 1994; Foley & Chen, 1999; Chen & Tyler, 1999a) to produce the rectified excitation E_{ji}

$$E_{ji} = \max\left(E_{ji}', 0\right) \quad (2)$$

where max denotes the operation of choosing the greater of the two terms.

The response of the j-th local filter is the excitation of the j-th filter, E_j, raised by a power p, in which $E_j = \Sigma_i E_{i,j}$ is the sum of excitations produced by all image components, and is then divided by a divisive inhibition term I_j plus an additive constant z. That is,

$$R_j = E_j^p \Big/ \left(I_j + z\right) \quad (3)$$

where I_j is the summation of a non-linear combination of the excitations of all relevant filters to filter j. This divisive inhibition term I_j can be represented as

$$I_j = \Sigma_i \left(S_{j,i} C_i\right)^q \quad (4)$$

where $S_{j,i}$ is the weight of the contributions from each image component to the inhibition term.

At a given location in the visual field, a set of the second-order detectors pools the responses of the local filters. Each second-order detector has a different size of summation field and thus sums different number of local filters. Suppose that the k-th second-order detector receives inputs from as many as n_k local filters. The overall response of k-th second-order detectors, T_k, is simply

$$Tk = \sum_{j=1}^{n_k} R_j \quad (5)$$

The decision is based on the outputs of the second-order detectors limited by the noise experienced by these detectors, which, as considered in this paper, has two sources: (1) the internal noise inherited in the local filters $[\sigma_a^2]$, and (2) the external noise provide by the noise patterns $[\sigma_e^2]$. The variance of the internal noise is assumed to be constant for all local filters in the model. Thus, the noise experienced by each second-order detector from this source is $n_k \sigma_a^2$, according to its pooling extent, n_k.

The variance of the external noise, σ_e^2 is proportional to the square of the contrast noise mask; that is, $\sigma_e = w_m C_m^2$, where w_m is a scalar constant that determines the amount of contribution of the noise mask to the variance of the response. Pooling these two noise sources, the variance of the response distribution in the k-th second-order detector is

$$\sigma_k^2 = n_k\left(\sigma_a^2 + \sigma_e^2\right) \quad (6)$$

In the context of our experiment, the observer compares the response to the stimuli at the two possible target locations. The observer can detect the target if the difference between the response to the target+mask, $T_{k,t+m}$, and that to the mask alone, $T_{k,m}$, in at least one second-order detector is greater than the limitation imposed by the noise. In practice, we assume that the noise mask produces little excitation in the local filters and in turn negligible second-order responses. This second-order mechanism should be the one whose receptive field covers the whole target extent and nothing else. If a mechanism does not cover the whole target, its response will be smaller than the one that does. If a mechanism covers a larger area, then it would suffer the noise from the extra local filters but receive no extra responses produced by the target. Thus, one implication of our model is that *the matched filter for the ideal summation is actually the second-order mechanism with the maximum signal-to-noise ratio*. We thus only need to consider the second-order mechanism that has the greatest response to the target. Thus, we can drop the subscript k for this study and focus on the decision variable given by,

$$d' = (T_{t+m} - T_m)\big/\left(2\sigma^2\right)^{1/2} \quad (7)$$

The threshold is defined when d' reaches unity.

For the matched filter, the number of channels it monitors is simply the target size multiplied by a factor. That is,

$$n = r\,S \quad (8)$$

where r is a scaling factor and S is the target size, an independent variable of our experiment, for the matched filters.

In practice, to avoid overdetermination in the fits, we set the parameter Se to the target, Se_t, to 100 and the

contribution the external noise mask, w_m, to 1. We also found that we could set the inhibition exponent q to 2 and the excitatory and inhibitory sensitivities of the noise mask, Se_m and Si_m, respectively, to zero without affecting the goodness of fit. The latter implies that the external noise mask had little effect on the mean response of summation mechanisms, but only increased the variability. In total, there were thus only five free parameters in the model for the whole data set for one observer: inhibitory sensitivity of the local detector to the target (Si_t, Equation 5), excitatory exponent and additive constant of the nonlinear response function (p and z, respectively, in Equation 4), the level of interval noise (σ_a in Equation 7) and scaling factor r (Equation 8).

We used the Powell method (Press, Flannery, Teukolsky, & Vetterling, 1988) to find the least-square fits to the data. As shown by the smooth curves in Figures 2 and 4, this model fits the data well. The root of mean squared error (RMSE) across the curves was between 0.97 and 1.30 dB, compatible with the mean standard error of measurement 0.8 dB, and explained 96.7–97.8% of all variance in the averaged data in the two observers.

Since the external noise mask had little effect on the mean response of local filters, with the constraints discussed already and combining Equations 5 to 8 within the matched filter, the decision variable

$$d' \sim = (S * R)\big/\left(S * \left(\sigma_a^2 + C_m^2\right)\right)^{1/2}$$
$$= \left(S^{1/2} * R\right)\big/\left(\sigma_a^2 + C_m^2\right)^{1/2} \quad (9)$$

If there is no external noise ($C_m = 0$), the model can be simplified as

$$d' \sim = \left(S^{1/2} * R\right)\big/\sigma_a \quad (10)$$

and thus becomes a typical ideal summation model that accounts for the approximately $-1/2$ slope of the spatial summation curve. Adding external noise does not change this slope for the size effect: as shown in the denominator of Equation 9, its effect does not depend on noise level. However, it is noteworthy that the spatial summation curve for our gain-control model deviates from the $-1/2$ slope for small target sizes. This deviation is due to fact that, when the target contrast is high, the divisive inhibition in Equation 3 outweighs the additive constant, reducing d' further than that imposed by the noise. Thus, as illustrated in Figure 7A for averaged data, our multiple stage model (red solid curve), incorporating the contrast gain control that is ubiquitous in models of contrast masking, captures not just the ideal summation tendency (green dashed line) but also the deviation at the extremes. Thus, the prediction of our model is almost indistinguishable from that of the three-component model (blue curve)
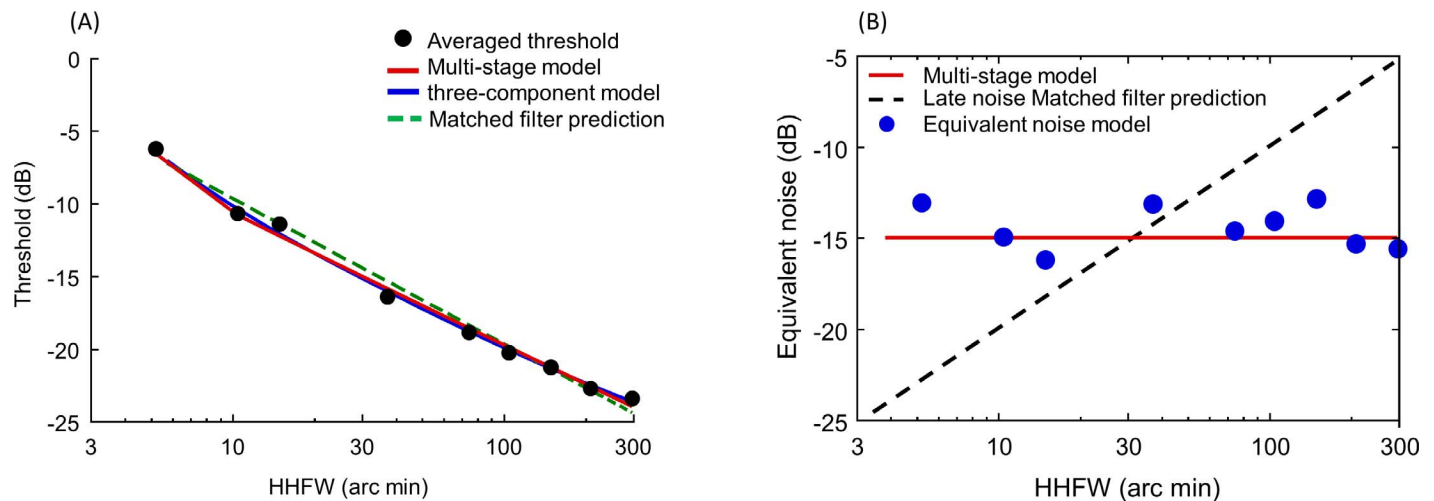
Figure 7. Illustration showing the comparison of selected model behavior of our multi-stage model with the convention models. (A) Comparison of our model (red solid curve) with the three-component summation model (blue solid curve) and the ideal summation model (green dashed line) on the averaged length-summation data from Figure 3. (B) Comparison of estimated equivalent noise from our multi-stage model, averaged across observers (red solid line) with that estimated by the equivalent noise model (Pelli, 1990; Legge et al., 1987) and with the prediction from the late noise matched filter model (diagonal dashed line) as shown in Figure 5B. The equivalent noise fits are averaged across observers for the estimations shown in Figure 5B.

that incorporates mechanisms for complete, ideal and probability summation. Indeed, our model, with only five free parameters for each observer, performs much better than the conventional ideal summation model with six free parameters for each observer. Quantitatively, the differential Bayesian information criterion (ΔBIC, Wagenmaker, 2007) of the ideal summation model, relative to our model, was -14.32 (probability that this model being more likely than ours, $p = 0.0008$) for YYH and -11.82 ($p = 0.0027$) for DTJ; and the three component model, with 18 free parameters for each observer, had ΔBIC -22.5 ($p < 0.0001$) for YYH and -86.3 ($p < 0.0001$) for TJ. Thus, taking into account the number of free parameters, the two generic models performed much worse than our model.

Another important feature of the model is that its d' is based on the second-order matched-filter summation mechanism. This output cannot distinguish whether the variability in the signal is from the internal noise in the local detector or the external noise in the stimulus. That is, the integration of the noise sources occurs early in the local detectors. As a result the denominator of d' is independent of target size, as shown in Equation 9. Thus, the constant equivalent noise (Figure 7B) found in our TvN functions is evidence for the operation of such second-order summation.

The size-dependent change in the numerator of Equation 9 also provides an explanation for the sensitivity change in the Lu and Dosher (1998) model. This fact also allows us to use fewer parameters to account for our results than the conventional equivalent noise model, which needs one free parameter for

each target size (see TvN Function section in Results and Figure 5A), Thus, with ΔBIC at -29.75 ($p < 0.0001$) for YYH and -40.6 ($p < 0.0001$) for DTJ, the equivalent noise model was dramatically outperformed by our model.

As one further comparison, we note that our model assumes that there is a full range of second-order detectors available, each with a different range of spatial summation, the decision-making being based on the one with greatest signal-to-noise ratio. A simpler model shown in Figure 3 is that the summation may be achieved by just one second-order detector for all target sizes. In this case the noise, which comes from all the local filters being monitored, should be a constant, but the signal in such mechanism would increase in proportion the target size. This model can be implemented by removing n_k from Equation 6 for noise level to form essentially a complete summation model (blue curve in Figure 3, See Spatial Summation in Results). Thus, it is not surprising that these models are much less likely than our model (ΔBIC $= -18.42$, $p = 0.0001$ for YYH and -10.22 for DTJ, $p = 0.006$).

Conclusion

In this study, we measured the contrast detection threshold of Gabor arc targets of various lengths embedded in a range of levels of external noise. We found that, at all noise mask contrasts, the contrast

detection threshold generally decreased with target length with a slope of approximately −1/2 on log-log coordinates, and thus was consistent the ideal summation behavior of Piper's law up to about 200 arcmin, or a Gabor arc length of 16 times its center width. Such ideal summation, in conventional psychophysical theory (Peterson, Birdsall, and Fox, 1954; Tanner & Jones, 1960; Green & Swets, 1966; Tyler & Chen, 2000) implies that the noise level experienced by the visual system increases with target size (since if the noise were constant, the threshold should decrease with a slope of −1). The TvN functions were all flat at low mask contrast and rose up when the mask contrast was greater than a critical value. The critical value, and in turn the equivalent noise estimates, did not change with target size. This is consistent with the early noise version of ideal spatial summation functions (Peterson et al., 1954; Tanner & Jones, 1960; Green & Swets, 1966; Tyler & Chen, 2000) but not the late noise one.

We proposed a multistage model with three elements: (1) contrast normalization in local detectors that accounts for pattern detection; (2) ideal summation behavior, in which the second order detector with the greatest signal-to-noise ratio dominates visual response; and (3) decision making based on both combined signal and noise in the best second order detector. This model accounts for the data better than the conventional matched-filter summation models (Green & Swets, 1966; Tyler & Chen, 2000, 2006; see Figure 7) or equivalent noise formulations (Legge et al., 1987; Pelli & Farell, 1999; Lu & Dosher, 1998, 2008).

Keywords: length summation, noise masking, gain control, equivalent noise, divisive inhibition

Acknowledgments

Footnote

[1] Note that we are using the term "second-order" in the sense of a second layer of summing mechanisms with more extended summing regions than the initial layer, with no implication as to whether they incorporate any form of contrast nonlinearity.

References

Baker, D. H. & Meese, T. S. (2011). Contrast integration over area is extensive: A three-stage model of spatial summation. *Journal of Vision*, *11*(14):14, 1–16, https://doi.org/10.1167/11.14.14. [PubMed] [Article]

Barlow, H. B. (1958). Temporal and spatial summation in human vision at different background intensities. *Journal of Physiology*, *141*(2), 337–350.

Baumgardt, E. (1959, December 19). Visual spatial and temporal summation. *Nature*, *184*, 1951–1952.

Chen, C. C., & Tyler, C. W. (1999a). Spatial pattern summation is phase-insensitive in the fovea but not in the periphery. *Spatial Vision*, *12*(3), 267–286.

Chen, C. C., & Tyler, C. W. (1999b). Accurate approximation to the extreme order statistics of Gaussian samples. *Communications in Statistics-Simulation and Computation*, *28*(1), 177–188.

Chen, C. C., & Tyler, C. W. (2010). Symmetry: Modeling the effects of masking noise, axial cueing and salience. *PLoS One*, *5*(4): e9840.

Chou, Y. L., Yeh, S. L., & Chen, C. C. (2014). Distinct mechanisms subserve location- and object-based visual attention. *Frontiers in Psychology*, *5*, 456.

Foley, J. M. (1994). Human luminance pattern-vision mechanisms: Masking experiments require a new model. *Journal of the Optical Society of America A, Optics, Image Science, and Vision*, *11*, 1710–1719.

Foley, J. M., & Chen, C. C. (1999). Pattern detection in the presence of maskers that differ in spatial phase and temporal offset: Threshold measurements and a model. *Vision Research*, *39*, 3855–3872.

Green, D. M., & Swets, J. A. (1966) *Signal detection theory and psychophysics*. New York, NY: Wiley.

Kao, C. H., & Chen, C. C. (2012). Seeing visual word forms: Special summation, eccentricity and spatial configuration. *Vision Research*, *62*, 57–65.

Kersten, D. (1984). Spatial summation in visual noise. *Vision Research*, *24*(12), 1977–1990.

Kersten, D. (1987). Statistical efficiency for the detection of visual noise. *Vision Research*, *27*(6), 1029–1040.

Kingdom, F. A. A., Baldwin, A. S., & Schmidtmann, G. (2015). Modeling probability and additive summation for detection across multiple mechanisms under the assumptions of signal detection theory. *Journal of Vision*, *15*(5):1, 1–16, https://doi.org/10.1167/15.5.1. [PubMed] [Article]

Kontsevich, L. L., & Tyler, C. W. (1999) Bayesian

adaptive estimation of psychometric slope and threshold. *Vision Research*, *39*, 2729–2737.

Legge, G. E., & Foley, J. M. (1980). Contrast masking in human vision. *Journal of the Optical Society of America*, *70*(12): 1458–1471.

Legge, G. E., Kersten, D., & Burgess, A. E. (1987). Contrast discrimination in noise. *Journal of the Optical Society of America A*, *4*, 391–404.

Lu, Z. L., & Dosher, B. A. (1998). External noise distinguishes attention mechanisms. *Vision Research*, *38*, 1183–1198.

Lu, Z. L., & Dosher, B. A. (2008). Characterizing observers using external noise and observer models: Assessing internal representations with external noise. *Psychological Review*, *115*, 44–82.

Meese, T. S., & Summers, R. J. (2007). Area summation in human vision at and above detection threshold. *Proceedings. Biological Sciences/The Royal Society*, *274*(1627): 2891–2900.

Meese, T. S., & Summers, R. J. (2012). Theory and data for area summation of contrast with and without uncertainty: Evidence for a noisy energy model. *Journal of Vision*, *12*(11):9, 1–28, https://doi.org/10.1167/12.11.9. [PubMed] [Article]

Nagaraja, N. S. (1964). Effect of luminance noise on contrast thresholds. *Journal of the Optical Society of America*, *54*, 950–955.

Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A*, *2*(9), 1508–1532.

Pelli, D. G. (1990). The quantum efficiency of vision. In C. Blakemore (Ed.), *Vision: Coding and efficiency* (pp. 3–24). Cambridge, UK: Cambridge University Press.

Pelli, D. G., & Farell, B. (1999). Why use noise? *Journal of the Optical Society of America A*, *16*(3), 647–653.

Peterson, W.W., Birdsall, T. G., & Fox, W. C. (1954). Theory of signal detectability. *IRE Transactions in Information Theory PGIT*, *4*, 171–212.

Polat, U., & Tyler, C. W. (1999). What pattern the eye sees best. *Vision Research*, *39*(5): 887–895.

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1988). *Numerical recipes in C.* Cambridge, UK: Cambridge University Press.

Robson, J. G., & Graham, N. (1981). Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Research*, *21*(3), 409–418.

Tanner, W. P., Jr., & Birdsall, T. G. (1958). Definitions of d' and η as psychophysical measures. *Journal of the Acoustical Society of America*, *30*(10), 922–928.

Tanner, W. P., Jr., & Jones, R. L. (1960). The ideal sensor system as approached through statistical decision theory and the theory of signal detectability. In A. Morris & E. P. Horne (Eds.), *Visual search techniques* (pp. 59–68). Washington, DC: National Academy of Sciences, National Research Council, Armed Forces NDC Committee on Vision.

Tyler, C.W., & Chen, C. C. (2000). Signal detection theory in the 2AFC paradigm: Attention, channel uncertainty and probability summation. *Vision Research*, *40*, 3121–3144.

Tyler, C. W., & Chen, C. C. (2006). Spatial summation of face information. *Journal of Vision*, *6*(10):11, 1117–1125, https://doi.org/10.1167/6.10.11. [PubMed] [Article]

Wagenmaker, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, *14*(5), 779–804.

Watson, A. B., Barlow, H. B., & Robson, J. G. (1983, March 31). What does the eye see best? *Nature*, *302*(5907), 419–422.

Wiener, N. (1949). *Extrapolation, interpolation and smoothing of stationary time series, with engineering applications.* Cambridge, MA: MIT Press.