Vision Research 122 (2016) 124–134

Contents lists available at ScienceDirect

Vision Research

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

# Rejecting probability summation for radial frequency patterns, not so Quick!

Alex S. Baldwin *, Gunnar Schmidtmann, Frederick A. A. Kingdom, Robert F. Hess

*McGill Vision Research, Department of Ophthalmology, McGill University, Montreal, Canada*

ARTICLE INFO

*Article history:*
Received 15 October 2015
Received in revised form 17 February 2016
Accepted 8 March 2016

*Keywords:*
Form
Integration
Radial frequency
RF
SDT
Shape
Signal Detection Theory
Summation

ABSTRACT

Radial frequency (RF) patterns are used to assess how the visual system processes shape. They are thought to be detected globally. This is supported by studies that have found summation for RF patterns to be greater than what is possible if the parts were being independently detected and performance only then improved with an increasing number of cycles by probability summation between them. However, the model of probability summation employed in these previous studies was based on High Threshold Theory (HTT), rather than Signal Detection Theory (SDT). We conducted rating scale experiments to investigate the receiver operating characteristics. We find these are of the curved form predicted by SDT, rather than the straight lines predicted by HTT. This means that to test probability summation we must use a model based on SDT. We conducted a set of summation experiments finding that thresholds decrease as the number of modulated cycles increases at approximately the same rate as previously found. As this could be consistent with either additive or probability summation, we performed maximum-likelihood fitting of a set of summation models (Matlab code provided in our Supplementary material) and assessed the fits using cross validation. We find we are not able to distinguish whether the responses to the parts of an RF pattern are combined by additive or probability summation, because the predictions are too similar. We present similar results for summation between separate RF patterns, suggesting that the summation process there may be the same as that within a single RF.



## 1. Introduction

To a first approximation the visual system can be considered a series of feedforward stages, where the neurones at each stage exhibit tuning to progressively more complex stimulus features. In primary visual cortex (V1) for example, we find cells tuned to orientation and spatial frequency (Hubel & Wiesel, 1962, 1968). Beyond V1 the system diverges into the dorsal stream, handling motion information, and the ventral stream where shape information is processed (Goodale & Milner, 1992; Ungerleider & Mishkin, 1982). As we move along the ventral stream the neurones exhibit tuning to more complex shape information (see Kravitz, Saleem, Baker, Ungerleider, & Mishkin, 2013 for recent review); these properties have inspired many models of shape and object processing (Cadieu et al., 2007; DiCarlo, Zoccolan, & Rust, 2012; Riesenhuber & Poggio, 2000; Serre, Oliva, & Poggio, 2007; Van Essen, Anderson, & Felleman, 1992). Neurones in primate V2 and V4 selectively respond to stimuli that combine multiple orientations such as angles, arcs, circles, hyperbolic gratings, and polar gratings (Anzai, Peng, & Van Essen, 2007; Gallant, Connor, Rakshit, Lewis, & Van Essen, 1996; Hegdé & Van Essen, 2007). Shape representation is believed to be mediated by a population code of cells in primate V4, which have been shown to exhibit tuning to specific contour features, e.g. convex and concave curvature maxima relative to the centre of a shape (Carlson, Rasquinha, Zhang, & Connor, 2011; Pasupathy & Connor, 1999, 2002; Yau, Pasupathy, Brincat, & Connor, 2013). Further along the ventral stream in inferotemporal cortex we find neurones selective for complex shapes and objects such as faces (Albright, Desimone, & Gross, 1984; Tanaka, 1996; Tsao & Livingstone, 2008).

As evidence continues to grow for this hierarchy, where progressively more complex stimulus features are represented along the ventral stream, the question of how this is achieved arises (Loffler, 2008; Wilson & Wilkinson, 2015). Radial frequency (RF) patterns were introduced by Wilkinson, Wilson, and Habak (1998) to address this question. An RF pattern is defined as a circular contour with a sinusoidally-modulated radius. Each cycle of the sinusoid gives a bulge at its peak and an indent at its trough. The frequency of the modulation determines the number of cycles in

* Corresponding author.
*E-mail address:* alexsbaldwin@googlemail.com (A.S. Baldwin).

http://dx.doi.org/10.1016/j.visres.2016.03.003
0042-6989/

the pattern (e.g. an RF4 has four peak-and-trough cycles) and the amplitude describes the magnitude of the distortion from a circle. Discriminating RF patterns from circles could be accomplished either by comparing the outputs of local filters matched to parts of the pattern, or by a global mechanism operating at the scale of the entire shape (taking those local filters as its input). Wilkinson et al. (1998) argued that the high sensitivity for the detection of RF modulations could not be achieved simply by local orientation or curvature analysis, but requires pooling of local contour information into a global representation of the RF shape. Further support for the global integration of RF shapes at threshold amplitude comes from a range of subsequent psychophysical studies (e.g. Bell & Badcock, 2008, 2009; Bell, Badcock, Wilson, & Wilkinson, 2007; Bell, Gheorghiu, Hess, & Kingdom, 2011; Hess, Achtman, & Wang, 2001; Hess, Wang, & Dakin, 1999; Jeffrey, Wang, & Birch, 2002; Loffler, Wilson, & Wilkinson, 2003; Schmidtmann, Kennedy, Orbach, & Loffler, 2012; Wang & Hess, 2005). For amplitudes above threshold, global integration receives support from studies of RF shape aftereffects (Bell, Hancock, Kingdom, & Peirce, 2010; Bell et al., 2011).

A subset of these studies used a summation paradigm in which the number of modulated cycles $n$ in the pattern was varied and the effect on threshold measured (Bell & Badcock, 2008; Dickinson, Han, Bell, & Badcock, 2010; Dickinson, McGinty, Webster, & Badcock, 2012; Hess et al., 1999; Loffler et al., 2003; Schmidtmann et al., 2012; Tan, Dickinson, & Badcock, 2013). In a linear system that performs global pooling one expects to see an inversely proportional relationship between the threshold and the number of modulated cycles (i.e. doubling the number of modulated cycles should halve the threshold). This gives a summation slope of $-1$ when threshold is plotted against $n$ on log–log axes. This prediction can be contrasted against that from a system where there is no global pooling and each cycle of the RF pattern is detected independently. In that case the improvement in performance due to the increasing number of modulated cycles would be due to probability summation between the mechanisms responsible for detecting each individual cycle (Sachs, Nachmias, & Robson, 1971). Probability summation is typically modelled under the assumptions of High Threshold Theory (HTT; see Green & Swets, 1966). Under HTT the predicted summation slope is $-1/\beta$, where $\beta$ is the parameter controlling the slope of the psychometric function obtained from a Weibull fit to the data (Quick, 1974). The summation slopes and HTT probability summation predictions from several previous experiments are shown in Table A1. As summation slopes are typically steeper than that predicted by probability summation under HTT, the authors of these studies have rejected this model. Although the empirical summation slopes do not reach the $-1$ predicted by the linear summation model (which under HTT means that the fixed high threshold occurs *after* the global pooling, as opposed to *before* the global pooling for the probability summation model), an additive global pooling model can still account for their data if there is a nonlinearity in the response to the individual cycles before the global pooling occurs. For example, a nonlinear transducer where the local response $r_{\text{local}} = A^{\tau}$, would give a predicted summation slope of $-1/\tau$.

Although these previous studies have focused on rejecting the HTT probability summation model, it is now widely accepted that Signal Detection Theory (SDT) provides the more appropriate framework to characterise decision processes in psychophysical experiments (Green & Swets, 1966; Meese & Summers, 2012; Nachmias, 1981; Tyler & Chen, 2000). This raises the question of whether probability summation modelled under the assumptions of SDT can so easily be rejected (Kingdom, Baldwin, & Schmidtmann, 2015; Tyler & Chen, 2000). Note that under SDT detection is also affected by uncertainty (Pelli, 1985), which introduces other model forms such as those featuring template-matching. If the noise affecting the inputs is uncorrelated and the observer is able to ignore noise from irrelevant inputs (i.e. those not being stimulated) this will also reduce the measured summation. In the ideal case where each input is weighted by the expected magnitude of its stimulation the slope will be $-1/2$ (Tanner, 1956). In the case where there is both a nonlinear transducer and an adjustable template their effects on the summation slope will multiply together to give even shallower summation slopes, on par with those predicted by HTT probability summation (Wilson, 1980). It is important to note that the derivation provided in Wilson (1980) does not describe a probability summation model; this detail is sometimes overlooked (e.g. Dickinson, Cribb, Riddell, & Badcock, 2015). Recent studies in the summation of contrast over area have rejected previous probability summation accounts and concluded that a "noisy energy" model of this form (where $\tau = 2$) provides the best explanation of the results (Baldwin & Meese, 2015; Meese, 2010; Meese & Summers, 2012).

In this study we first collect receiver operating characteristic (ROC) data to demonstrate that SDT, rather than HTT, is the correct theory in the context of RF pattern discrimination. This is simple to demonstrate as under HTT the ROC should be straight whereas under SDT it should be curved (Green & Swets, 1966). This finding makes the predictions of the HTT-based probability summation model irrelevant to the study of the detection of RF pattern modulation. In the second part we perform additional experiments and modelling (Matlab code is provided as a Supplementary material) to investigate whether this rejection of HTT changes our conclusions about how summation occurs within RF patterns. We also compare summation within an RF pattern against summation between RF patterns in order to see whether summation within an RF pattern has any special properties. We find that we are unable to reject a probability summation model formulated under SDT. When comparing summation within an RF pattern to summation between RF patterns we find little difference.

## 2. Methods

### 2.1. Equipment

The stimuli were generated in Matlab (Matlab R2013a, MathWorks) and presented on a gamma-corrected Iiyama Vision Master Pro 513 CRT monitor with a resolution of $1024 \times 768$ pixels and a frame rate of 85 Hz (mean luminance 38 cd/m$^2$) using an Apple Mac Pro (3.33 GHz). Observers viewed the stimuli at a distance of 1.2 m. At this distance one pixel on the monitor subtended 0.018 degrees of visual angle (deg). Experiments were carried out under dim room illumination. Routines from the PsychToolbox were used to present the stimuli (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007).

### 2.2. Observers

Three observers participated in the complete set of experiments. Two were authors (ASB and GS), and the third was a psychophysically-experienced observer who was naive to the purposes of the experiment (AR). Two more naive observers were brought in to collect additional data (YG and TT). All observers wore their appropriate optical correction for the viewing distance. Experiments were carried out with the participants' informed consent in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) and were approved by the Biomedical B Research Ethics Board of the McGill University Health Centre.

### 2.3. Stimuli

The RF patterns were created by sinusoidally modulating the radius ($r$) of a circle

$$r_\theta = r_{\text{mean}} \times [1 + m_\theta \times A \times \sin(f\theta + \varphi)], \quad (1)$$

where $r$ and $\theta$ are polar coordinates (with the origin at the centre of the pattern), $r_{\text{mean}}$ is the mean radius, $A$ is the amplitude of the modulation expressed as a Weber fraction (absolute modulation amplitude divided by the mean radius), $f$ is the modulation frequency (number of cycles in the pattern) and $\varphi$ is the phase (equivalent to rotation of the pattern). In order to allow us to measure the detectability of individual cycles in the RF pattern we define a masking vector $m$, the values of which ($m_\theta$) within a particular cycle can either be zero (meaning the cycle will not be modulated) or one (meaning it will be modulated). In all of our experiments $f$ was fixed at 4. The popular nomenclature would refer to these as "RF4" patterns. We also fixed $r_{\text{mean}}$ at 0.5 deg, and $\varphi$ at 90°.

In order to create stimuli which were bandpass in terms of their spatial frequency content, we generated them by using the radial fourth-derivative of a Gaussian to define the luminance profile of a cross-section across the edge of the pattern, after Wilkinson et al. (1998). We first transform the polar coordinates ($r$, in degrees of visual angle, and $\theta$) of each location to give the pattern our desired peak spatial frequency ($\omega$) and modulating radius ($r_\theta$, as defined in Eq. (1))

$$F(r,\theta) = \frac{\pi\omega(r - r_\theta)}{\sqrt{2}} \quad (2)$$

and then use that to set the luminance $l$ at each location in the image

$$l(r,\theta) = c \times \left[1 - 4 \times F(r,\theta)^2 + \frac{4}{3} \times F(r,\theta)^4\right] \times \exp[-F(r,\theta)^2], \quad (3)$$

where $c$ is the luminance contrast. The spatial frequency used in our experiments was 8 c/deg, and the bandwidth (full width at half magnitude) was 1.24 octaves (Wilkinson et al., 1998).

For the two main conditions observers were presented with either a single RF pattern at fixation (Fig. 1a) or a quad of RF patterns surrounding fixation (Fig. 1b). In the Single RF pattern condition we could modulate any combination of these four cycles (positioned left, right, superior and inferior relative to fixation) in order to test sensitivity to just one cycle being modulated and to see how this changed as the number of modulated cycles increased. For the Quad RF condition the individual patterns were placed to the left, right, superior and inferior to fixation (indicated in this condition by a white circle with a diameter of 0.126 deg), and modulations were applied to the single innermost cycle of each RF pattern. The eccentricity of the centres of the Quad RF patterns was 1.8 deg. This meant that the modulations our observers were detecting in the Quad RF condition were slightly more eccentric than in the Single RF condition (1.3 deg vs. 0.5 deg). In all our experiments the stimulus duration was 300 ms and the interstimulus interval was also 300 ms.

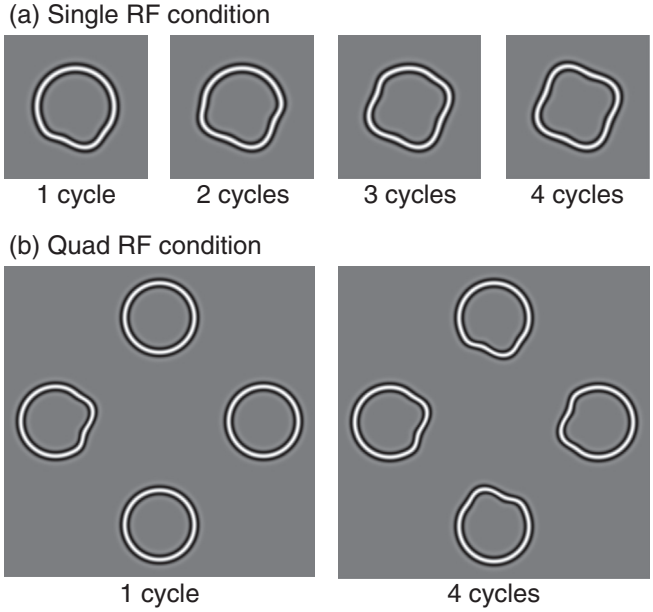


**Fig. 1.** Examples of the RF stimuli used in our experiments.

### 2.4. Receiver operating characteristics

We collected receiver operating characteristics (ROCs) using a rating scale experiment. For each observer we first conducted a two-interval forced-choice pilot experiment with stimuli that had a single cycle modulated. This was to determine the modulation level that could be discriminated from a circle 75% of the time. We tested four blocks using the method of constant stimuli (6 levels, 20 trials per level), with the modulated cycle being in a different location in each block (left, right, superior and inferior). The data were combined across location and then fit with a Quick psychometric function to give a 75% threshold value. For our rating scale tasks we randomly interleaved 100 trials where there was no modulation and 100 trials where the modulation was set to this threshold value. We later ran an additional experiment where all four cycles were modulated. For this condition we used the results from our summation experiment (below) to find modulation amplitudes that would be slightly above threshold. The stimulus condition (Quad vs. Single RF, and the location of the modulation) was kept constant within a testing block. On each trial the observer responded using a four-level scale. They were asked to rate how confident they were that the modulation was present, where 1 was "confident present", 2 was "probably present", 3 was "probably absent", and 4 was "confident absent".

The ratings were converted into points on the ROC plot by cumulatively binning the responses (1, 1–2, 1–3) and calculating $P$(hit), the probability of responding that the modulation is present when it *is* shown, against $P$(false alarm), the probability of responding that it is present when it is *not* shown, for each bin. The rating experiments were repeated twice by each observer for each condition, and nonparametric bootstrapping (1000) was performed to generate six bivariate distributions of $P$(hit) and $P$(false alarm) values for each plot. Each bootstrap sample was generated by randomly sampling (with replacement) 100 responses from the set of test trials and 100 responses from the set of null trials in each repetition, to create a simulated dataset that could be processed like our empirical data.

Note that no training was conducted on this task before data collection began. Discriminating the circles from the RF patterns in a single interval task proved to be challenging at first, as one has to maintain a template of the expected modulated stimulus. We provided the observers with an example stimulus indicating the modulated location at the beginning of each block, but they were required to keep track of the stimulus they were detecting after this point (one observer found the task impossible for one of the modulated locations in each condition, see Supplementary material Figs. S1a and S2d). As a result of this, each observer produced some inconsistent results at the beginning of testing as they were learning the task. Where observers were very inconsistent between testing blocks (determined by non-overlapping 95% confidence intervals on the $d'$ values obtained by the SDT ROC fit) a

further repetition was conducted, and the inconsistent data were not used when producing the plots that illustrate our ROC curves. This happened a few times for each observer in our initial experiment (4 repetitions for AR, 4 for ASB, and 2 for GS). These data were however still used in our statistical tests to avoid any bias. They are the ringed data points in Fig. 5, where it can be seen that the rejected points *do not* favour the HTT model. For the subsequent experiments where all four cycles were modulated the observers performed consistently and no data were rejected.

The ROC plots $P$(hit) against $P$(false alarm). The possible behaviour in response to a single stimulus intensity is described by a line on the ROC plot. Under both HTT and SDT the lines pass through the point (1, 1), representing the situation where the observer responds that the modulation was present on every trial. The major distinction between HTT and SDT is what happens as $P$(false alarm) reaches zero on the left hand side of the ROC plot. Under HTT false alarms are due to the observer guessing incorrectly, and because guess rates do not depend on the intensity of the stimulus they can be factored out in yes/no tasks (this was shown not to be the case empirically for contrast detection by Nachmias, 1981). Between $P(\text{falsealarm}) = 0$ and $P(\text{falsealarm}) = 1$ the prediction is that the two points are joined by a straight line as the HTT observer varies their guess rate from 0 to 1. The slope of the line is determined by the stimulus intensity; as it increases the line becomes shallower and the hit rate that the observer can achieve in the absence of guessing increases. Under HTT the responses in rating scale experiments are determined by the observer adopting different guess rates for each point on the scale. Differences between the two repetitions we conducted would be due to variations in those guess rates.

Under SDT on the other hand it is impossible to eliminate false alarms without also reducing the hit rate to zero. The ROC is described by a line between (0, 0) and (1, 1) which curves inward toward the top-left corner of the graph, since $P$(hit) is generally greater than $P$(falsealarm). The degree of curvature is determined by the stimulus intensity which results in a particular signal-to-noise ratio ($d'$). Each position along the line between (0, 0) and (1, 1) is the combination of values expected for a particular response criterion. The criterion is the magnitude of the internal response that the observer requires before classifying an interval as "signal present". Significantly this criterion can be exceeded on trials where the modulation was not present, as a consequence of the internal noise in the observer's visual system. In a rating scale experiment the observer assigns different criterion values to the different points on the scale, and differences between the two repetitions can be attributed to changes in those criteria.

### 2.5. Summation within and between RF patterns

A two-interval forced-choice "method of constant stimuli" design was used to obtain psychometric functions for the summation experiment. Observers indicated which of the two intervals had the stimulus with the RF modulation. We tested 36 trials at 9 modulation amplitudes. For the blocked condition the modulated cycle or cycles were the same for all trials within a block. For the interleaved condition we randomly interleaved all 15 possible stimuli within a block, testing each condition 3 times. By testing 12 of these blocks for the interleaved condition we reach the same number of trials per stimulus as was tested in the blocked condition. Psychometric functions for each condition were then fitted with a cumulative normal function using Palamedes (Prins & Kingdom, 2009). Parametric bootstrapping was performed (500 samples) to generate populations of threshold ($\alpha$) and psychometric slope (equivalent Weibull $\beta$) parameters for each observer. We fixed the psychometric slopes to be constant for particular number of modulated cycles for each observer, but allowed the thresholds to vary when those modulated cycles were in different locations. We combined data across observers by taking the mean of these bootstrap populations and then reported the median and confidence intervals of the resulting distributions.

## 3. Results and modelling

### 3.1. ROC curves

The ROCs from one observer (ASB) for the Single and Quad RF conditions are shown in Figs. 2 and 3 respectively (similar figures for the other observers are provided in the Supplementary materials Figs. S1–S4). The empirical data are shown by the black markers, with the coloured regions showing the distribution of non-parametric bootstrap samples (indicating the error associated with those black marker points). In all cases the data form a curved line between the points (0, 0) and (1, 1) that is inflected toward the top-left of the plot. This is the expected behaviour under the SDT framework, with the appropriate model shown by the solid black line. The dashed black line shows the best-fitting prediction from the HTT model. It is immediately clear that this provides a poorer account of the empirical data. The points from the two repetitions of the experiment (squares and diamonds) do not overlap, falling on distinct points along the ROC curve. This indicates that the criteria associated with the response categories varied between repetitions, while the sensitivity remained roughly constant. At the request of a reviewer, we also obtained ROC data for each observer where all four cycles were modulated (Fig. 4). These also demonstrate the curved form that we expect under SDT.

We performed two-tailed sign tests in Matlab (Matlab R2013a, MathWorks) to compare the deviances of the fits from the two models (shown in Fig. 5) on a block-by-block basis (fitting the data

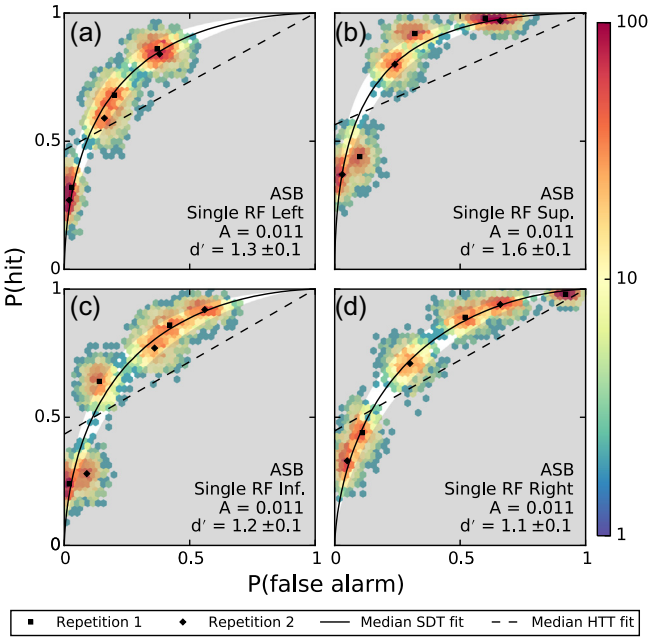


**Fig. 2.** ROC plots showing the relationship between hit and false alarm rates for the four different modulation locations in the Single RF condition for observer ASB. Data from the two repetitions are shown by the two marker symbols, with the distribution of nonparametric bootstrap samples shown by the colour-map. The median fit from the HTT model is shown by the dashed line. The median SDT fit is shown by the solid line. The area outside the 95% confidence interval on the SDT fit is shaded. Each graph also gives both the modulation amplitude $A$, as a Weber fraction and the resulting $d'$ calculated from the SDT fit to the data.

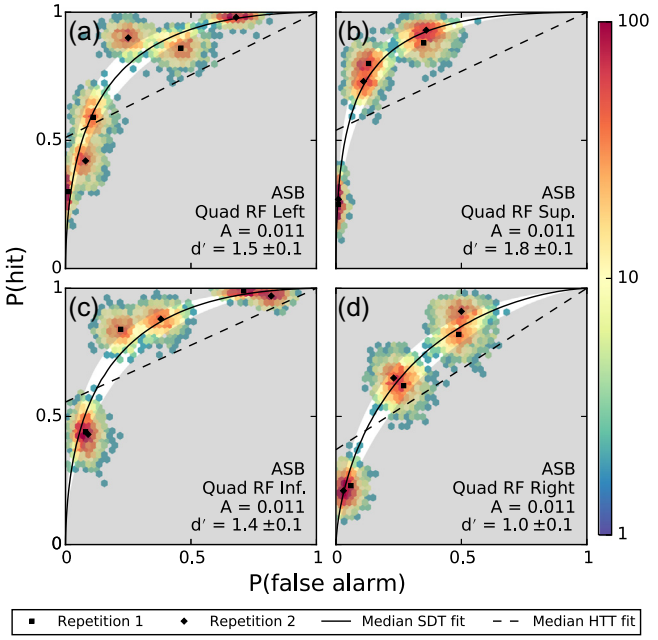


**Fig. 3.** ROCs plot of the Quad RF condition for observer ASB. See Fig. 2 caption for further details.

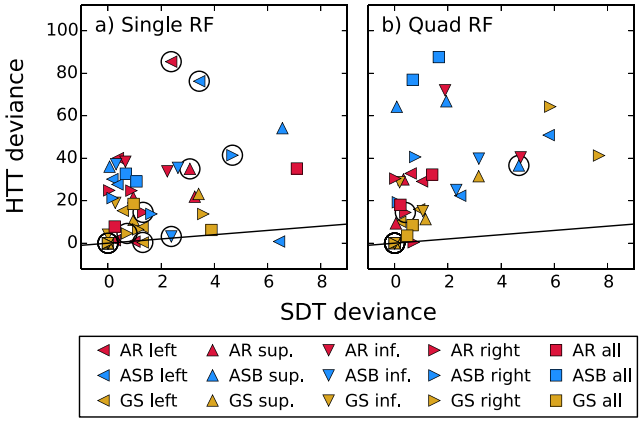


**Fig. 5.** Scatter plot showing the deviances for the fits to each repetition of the ROC data from the HTT and SDT models. All points above the $y = x$ line (the scales on our axes had to be unequal in order to clearly present the data) were better fit by the SDT model. Data that were not included in the ROC plots (see Section 2) are indicated by the black circles. These were still included in the statistical analysis.

from each repetition separately). Deviances for the SDT and HTT model fits were each calculated as twice the difference between the log-likelihood from that model and the log-likelihood from the saturated model. In the Single RF condition the SDT model provided a significantly better fit for ASB ($P = 0.002$), AR ($P = 0.002$), and GS ($P = 0.006$). This was also true in the Quad RF condition where the SDT model provided a better fit for ASB ($P = 0.001$), AR ($P = 0.012$), and GS ($P = 0.002$). We can therefore assert that the detection of RF patterns does not comply with the HTT framework. This means that the Quick pooling function (Quick, 1974) typically used to model probability summation under HTT does not apply to RF patterns. Instead one must use an SDT probability summation model such as the one we provide below.

One final point of interest from the ROC curves is the considerable variation in sensitivity between the different modulation locations. In Fig. 2 we show that with the same modulation amplitude in the Single RF condition we get a 40% difference between the lowest (modulation right) and highest (modulation superior) $d'$ values measured. For the Quad RF condition (Fig. 3) this variation was even greater (a 73% difference in $d'$ between the right and superior modulation locations). This may reflect variations in sensitivity across the visual field for shape discrimination (Schmidtmann, Logan, Kennedy, Gordon, & Loffler, 2015). For ASB they would not be consistent with different sensitivities to the different rotations of our RF4 stimulus because when the modulation is shown to the right of fixation in the Quad RF stimulus it is equivalent to a translation of the *left*-modulated single RF stimulus (Fig. 1). This can be contrasted against the results for AR in the

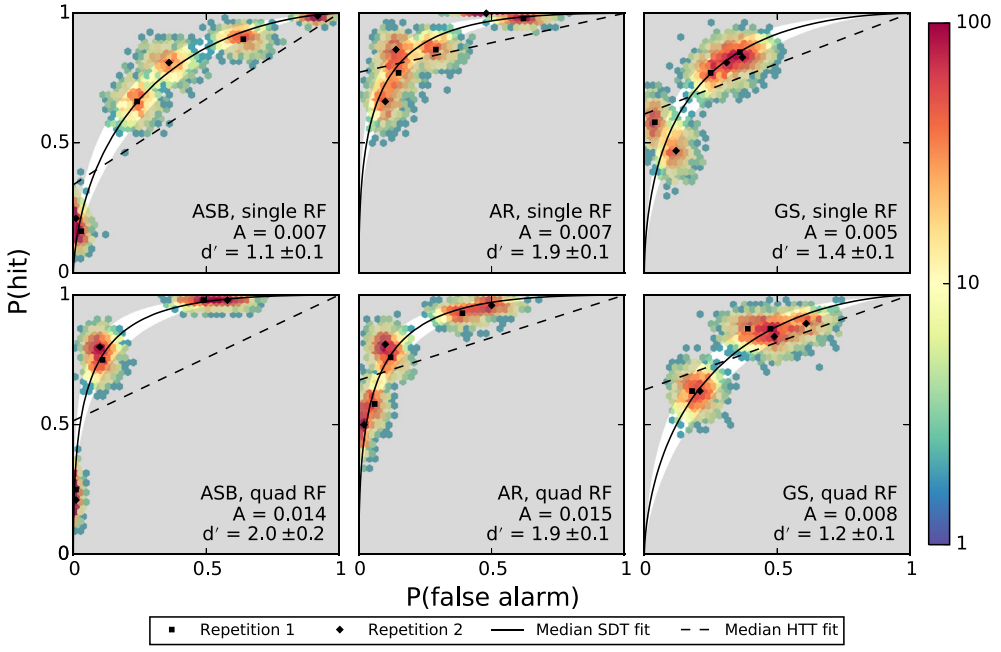


**Fig. 4.** ROC plots for detecting the RF stimuli when all four cycles were modulated. The three columns show results from different observers, with the top row containing the experiments for a single RF pattern and the bottom row a quad of RF patterns (with the innermost cycles modulated). See Fig. 2 caption for other details.

single RF left (Fig. S1a) and the Quad RF right conditions (Fig. S2d) where it seems that it is a particular rotation of the stimulus that they have difficulty detecting.

### 3.2. Summation results

The summation results averaged across our three principal observers are shown in Fig. 6. Panels (a, b) shows how thresholds change as the number of modulated cycles in the stimulus increases, (c, d) show how the psychometric slopes are affected. As expected (Kingdom et al., 2015; Tyler & Chen, 2000) the summation slopes for the interleaved condition are steeper than for the blocked condition. What is surprising though is that there seems to be very little summation at all in the blocked condition. This can be seen in the slopes of the lines fit to the data. For the Single RF condition there is a small decline in threshold as the number of modulated cycles increases (the median value of the log–log slope of the fit to 1000 bootstrapped thresholds was −0.18), and for the Quad RF condition there is no improvement at all (slope of 0.06). In our experiments it seems that the observers in the blocked condition simply chose a single cycle of the RF stimulus to monitor for the purpose of making the 2AFC decision, regardless of how many cycles were in fact being modulated. Because the observers did not exhibit summation behaviour in this condition, we will exclude it from our modelling.

For the interleaved data we find threshold vs. $n$ slopes of approximately −0.53 for the Single RF condition and −0.60 for the Quad RF condition. These are in the same range as those found in previous experiments that tested summation within a single RF (Table A1). The similarity of the slope between our Single and Quad RF conditions indicates that any special quality attributed to summation within an RF pattern (e.g. that the component cycles are additively combined in a higher level mechanism) could also apply to summation between RF patterns. In fact, the summation slopes we find may be consistent with either additive or probability summation under SDT. We shall address this in our modelling below. The only clear difference we do see between the Single and Quad RF conditions is that in the Single RF condition the thresholds for the blocked condition are higher than for the interleaved condition, whereas for the Quad RF condition they are generally lower. The Single RF results are consistent with the explanation we offer above for the lack of summation in the blocked condition, however those for the Quad RF condition are more difficult to interpret. Our best explanation is that there is some "cost" associated with integrating the local outputs over a larger area in the interleaved condition, and that this elevates the thresholds to be higher than those for the blocked condition (where we suggest that the observer is only making use of a single local output).

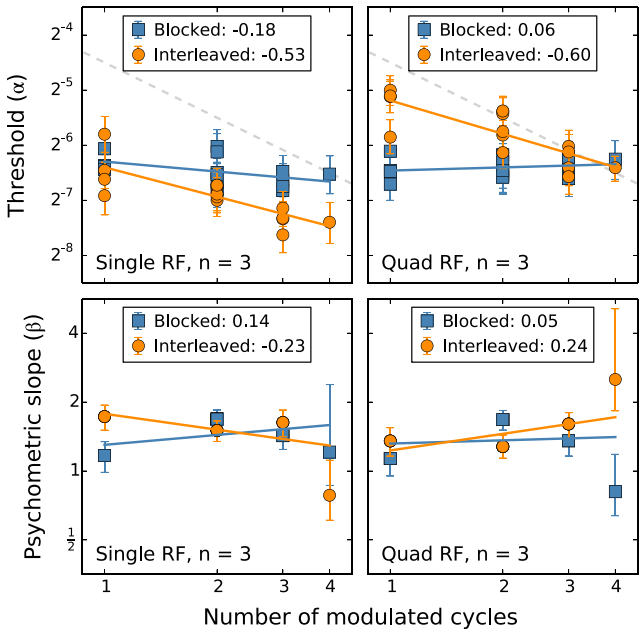


**Fig. 6.** Summation results averaged across our three principal observers. The top row shows thresholds (as Weber fractions) as a function of the number of modulated cycles in the pattern for the Single and Quad RF conditions. The bottom row shows the psychometric slopes. The data from each condition are fitted with straight lines, the slopes of which are reported in the legends. The dashed grey lines in panels a, b have a slope of −1.

One key prediction from the SDT probability summation model is that – under conditions where the threshold vs. $n$ summation slopes are steep – the psychometric function slopes should become more shallow as the number of modulated cycles increases due to a reduction in extrinsic uncertainty (Pelli, 1985). Such an effect is not immediately apparent from the data shown in Fig. 6c, d. There is a slight negative slope however in the Single RF interleaved condition (−0.23), which is comparable to the value we would predict from our SDT probability summation model (Kingdom et al., 2015). Overall though, our analysis of the data from the interleaved conditions does not provide particularly strong evidence in favour of either the additive or the probability summation models. We shall address this by developing additive and probability summation models and fitting those to our data.

### 3.3. Summation modelling

The summation models that we used are outlined briefly here. Further information and Matlab code can be found in our Supplementary material. For additive summation we calculate $d'$ by summing over each $i$ of $n$ mechanisms

$$d' = \frac{\sum_{i=1}^{n} (g_i A_i)^{\tau}}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}} \tag{4}$$

where $g_i$ is the input gain at each location, $A_i$ is the amplitude of the modulation, and $\sigma_i$ is the internal noise. In the additive summation model having different values for the internal noise at different locations does not affect performance, so we fix these to be constant across location. The resulting parameter has a directly inverse effect to the gain, so we fix all of our internal noise $\sigma = 1$ and implicitly fit any internal noise effects with our gain parameter. We test four versions of the additive summation model which are distinguished by whether they feature linear ($\tau = 1$) or nonlinear ($\tau$ is free) transduction (T0 vs. T1) and by whether the gain is fixed across locations or whether it is different for each location (G1 vs. G4). This nomenclature is summarised in the first four rows of Table 1. The $d'$ value can be converted to predicted percent-correct using the standard method

$$P(\mathrm{c}) = 0.5 + \left[\Phi_s\left(\frac{d'}{\sqrt{2}}\right) - 0.5\right] \times (1 - 2\lambda) \tag{5}$$

where $\Phi_s$ is the standard cumulative normal distribution function and $\lambda$ is the lapse rate.

For the probability summation model we used a modified version of the equation derived in Kingdom et al. (2015). For this we must define a general version of the cumulative normal distribution function $\Phi(t, \mu, \sigma)$ where $t$ is the value for which the probability is calculated (used as a dummy variable in the integration in our equations), $\mu$ is the mean of the normal distribution, and $\sigma$ is its standard deviation. We also define the probability density function $\phi(t, \mu, \sigma)$ in the same way. For detection within a single mechanism (no summation occurring) we calculate

**Table 1**
Explanation of the acronyms used to refer to the different forms of our models.

| Model | Transducer | Gain | Noise |
|---|---|---|---|
| T0G1S0 | Linear ($\tau = 1$) | Controlled by single $g_1$ | Constant ($\sigma = 1$) |
| T1G1S0 | Nonlinear | Controlled by single $g_1$ | Constant ($\sigma = 1$) |
| T0G4S0 | Linear ($\tau = 1$) | Per location $g_1 - g_4$ | Constant ($\sigma = 1$) |
| T1G4S0 | Nonlinear | Per location $g_1 - g_4$ | Constant ($\sigma = 1$) |
| T0G1S3 | Linear ($\tau = 1$) | Controlled by single $g_1$ | Per location $\sigma_1 = 1$, $\sigma_2 - \sigma_4$ vary |
| T1G1S3 | Nonlinear | Controlled by single $g_1$ | Per location $\sigma_1 = 1$, $\sigma_2 - \sigma_4$ vary |

$$P(\mathrm{c}) = \int_{-\infty}^{\infty} [\phi(t, (gA)^{\tau}, \sigma) \times \Phi(t, 0, \sigma)] dt \tag{6}$$

For every possible response $t$ in the target interval we are calculating the probability of that response occurring $\phi[t, (gA)^{\tau}, \sigma]$ and multiplying that by the probability that the response in the null interval was lower $\Phi(t, 0, \sigma)$. If the observer responds simply on the basis of which interval elicited the larger response then this should be the probability with which the observer chooses the correct interval.

The probability summation model involves these same components in a more complex arrangement

$$P(\mathrm{c}) = \sum_{i=1}^{n} \int_{-\infty}^{\infty} \phi[t, (g_i A_i)^{\tau}, \sigma_i] \times \prod_{\substack{j=1 \\ j \neq i}}^{n} \Phi[t, (g_j A_j)^{\tau}, \sigma_j] \times \prod_{j=1}^{n} \Phi[t, 0, \sigma_j] dt. \tag{7}$$

Here we are performing the calculation from Eq. (6) for each $i$ of $n$ mechanisms in order to find the probability of responding correctly on the basis of each mechanism's response. These are then combined by summing them together. The calculation for each mechanism is also modified by including a term for the probability that no other mechanism in the target interval had a greater response $\Phi[t, (g_j A_j)^{\tau}, \sigma_j]$. For both this term and the term for the responses in the other interval we must now also multiply the responses over each $j$ of our $n$ mechanisms, as we need to weight the probability for a particular response from our target mechanism by the probability that *all* other mechanisms in either interval gave a smaller response. Crucially, in this model we *do* get distinct effects for having different internal noise in each mechanism vs. different input gains. We must fix one of the noise standard deviations however, as otherwise they will interact with the gain. Models with constant internal noise are coded S0 and those with varying internal noise are coded S3, as shown in the last two rows of Table 1.

We used a maximum-likelihood procedure to fit these models to our data. Log-likelihoods for the whole experiment were calculated over the sum of our $k$ conditions as

$$\log L = \sum_{k} C_k \log[P_k(\mathrm{c}|\boldsymbol{A_k}, \boldsymbol{g}, \boldsymbol{\sigma}, \tau)] + I_k \log[1 - P_k(\mathrm{c}|\boldsymbol{A_k}, \boldsymbol{g}, \boldsymbol{\sigma}, \tau)], \tag{8}$$

where $C_k$ is the number of correct responses for that condition, and $I_k$ is the number of incorrect responses, and $P_k(\mathrm{c}|\boldsymbol{A_k}, \boldsymbol{g}, \boldsymbol{\sigma}, \tau)$ is the probability of a correct response for each condition according to the model. This probability is calculated based on the vector of amplitudes $\boldsymbol{A_k}$ of the modulation applied to each mechanism in that condition, and the current model parameters $\boldsymbol{g}$, $\boldsymbol{\sigma}$ (both vectors as there may be different values for each mechanism), and $\tau$. We used fminsearch in Matlab to minimise the negative log-likelihood, which is equivalent to finding the maximum likelihood solution.

We approached the question of which model best accounts for our data as a *model selection* problem. It is not possible to use a simple likelihood ratio test because our additive and probability summation models are two entirely distinct families (i.e. not nested one within another). We also must compensate in some way for the number of parameters in each of our models, as it is expected that within a particular family of nested models (e.g. our sets of additive and probability summation models) we should always get a superior fit from the model with the greater number of parameters, with some of this improvement being due to overfitting. One tool to compensate for this is a criterion that incorporates the number of fitted parameters as a penalty (e.g. the various forms of AIC, see Akaike, 1974; Burnham & Anderson, 2002). A more intuitive method (the outcome of which is related to that found using AIC; see Stone, 1977) is cross-validation, where the data are partitioned and the predictions made based on the fit to one part are evaluated by how well they predict another part.

In this study we used a ninefold cross-validation method. For each observer we randomly partitioned their interleaved data into nine sets. We then performed the fitting nine times. Each time a different set was chosen as the holdout set for validation, with the fitting being performed to the combined data from the other eight sets. We used the model with the parameters obtained by fitting to the rest of the data to make predictions for the holdout data set to obtain a log-likelihood value. We averaged these values, normalised by the maximum for each observer, and plot them with their standard errors in Figs. 6 and 7. The fitted parameters for each model are provided as a Supplementary material, and their values averaged across observers are provided in Tables B1–B4 of Appendix B.

The results from the cross-validation for the Single RF condition are shown for our three main observers and two additional observers in Fig. 7. The different models are represented on the abscissa according to the parameters they use (Table 1). Each model also has one final parameter, the lapse rate $\lambda$. For the models on the left side of the graph the signal combination occurs according to additive summation, on the right side is probability summation. The details of the fits for each observer are provided in our Supplementary materials. A summary giving parameters averaged across observers and combined log-likelihoods and deviances is given in Appendix B. The deviance is a goodness-of-fit measure that factors out the entropy of the dataset that the model likelihood is calculated from. It is calculated as

$$D = -2 \times (\log L_{\mathrm{test}} - \log L_{\mathrm{sat}}) \tag{9}$$

where $L_{\mathrm{test}}$ is the likelihood of the model being evaluated, and $L_{\mathrm{sat}}$ is the likelihood of the saturated model (Kingdom & Prins, 2016). In the case of this experiment the "prediction" used to calculate the likelihood for the saturated model is that in each condition (Single/Quad RF, location of modulated cycles, and signal level) the probability of the observer giving a correct response is the empirical probability found in the experiment. In this way the likelihood of the saturated model represents the inherent unpredictability of the dataset. To obtain our deviances in Eq. (9) we are using the candidate model's log-likelihood when tested against

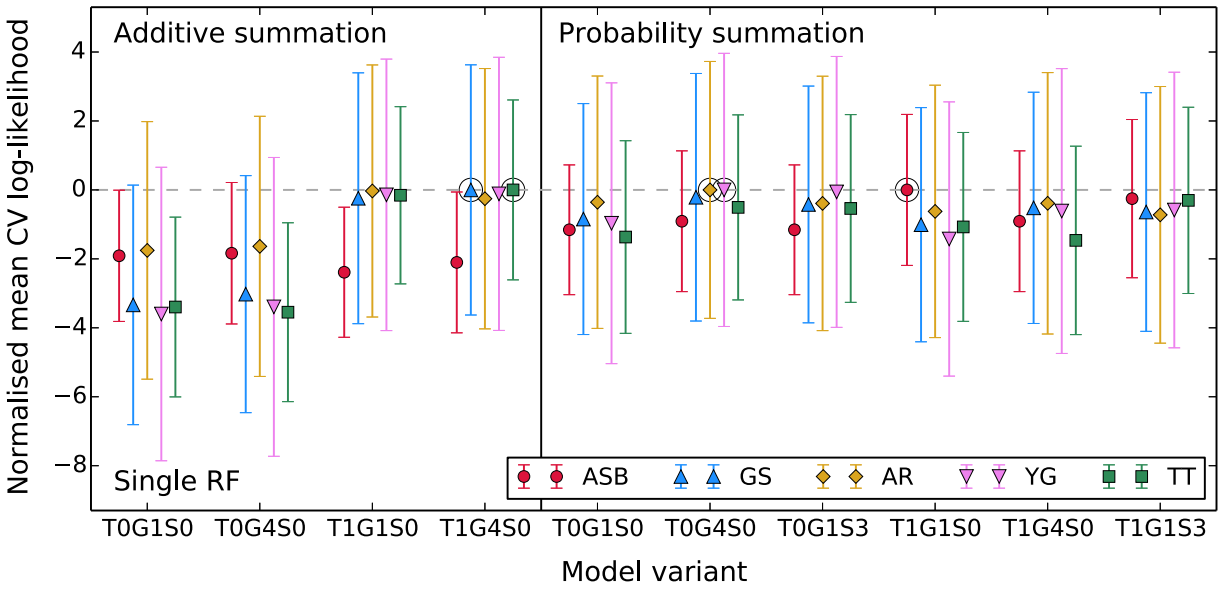


**Fig. 7.** Output of the cross-validation procedure for the Single RF condition. For each version of the model we plot the mean log-likelihood across the cross-validation folds normalised by subtracting the maximum likelihood for each observer, with its standard error. To the left of the black vertical line we plot the additive summation models, to the right probability summation models. We mark the single best-fitting model for each observer with a ring. Results from each observer are offset slightly on the *x*-axis to make the figure easier to read.

**Table B1**
Table showing the mean parameters across observers and combined log likelihoods and mean deviances obtained by fitting the additive summation model to the Single RF data. The codes defining each model are explained in Table 1.

| Model | $\tau$ | $g_1$ | $g_2$ | $g_3$ | $g_4$ | $\lambda$ (%) | $\Sigma(\log L)$ | $\bar{D}$ |
|---|---|---|---|---|---|---|---|---|
| T0G1S0 | | 103.2 | | | | 0.93 | −953.4 | 110.5 |
| T1G1S0 | 1.63 | 93.7 | | | | 1.17 | −942.4 | 106.1 |
| T0G4S0 | | 108.4 | 106.2 | 105.2 | 98.3 | 0.90 | −952.9 | 110.3 |
| T1G4S0 | 1.70 | 98.0 | 96.8 | 93.9 | 88.1 | 1.18 | −941.9 | 105.9 |

**Table B3**
Table showing the mean parameters across observers and combined log likelihoods and mean deviances obtained from a maximum-likelihood fit of the additive summation model to the Quad RF data.

| Model | $\tau$ | $g_1$ | $g_2$ | $g_3$ | $g_4$ | $\lambda$ (%) | $\Sigma(\log L)$ | $\bar{D}$ |
|---|---|---|---|---|---|---|---|---|
| T0G1S0 | | 53.0 | | | | 0.88 | −708.3 | 124.7 |
| T1G1S0 | 1.31 | 49.8 | | | | 1.08 | −706.3 | 123.4 |
| T0G4S0 | | 47.9 | 74.6 | 44.7 | 52.8 | 0.92 | −706.3 | 123.4 |
| T1G4S0 | 1.38 | 45.1 | 71.2 | 40.8 | 47.0 | 1.17 | −704.0 | 121.9 |

the holdout dataset and subtracting the likelihood of the saturated model calculated for that dataset.

Of the five observers, three of them (ASB, AR, and YG) behave in a way that is most-likely explained by probability summation (see the ringed "best fit" markers). The other two (GS and TT) behave in a way that is more compatible with the additive summation model. The worst performing models tend to be the additive summation models with a linear transducer, but otherwise there is not a great difference between them. From Tables B1 and B2 we can see that the best-fitting transducer values are higher for the additive than the probability summation models (1.7 vs. 1.4). This is to be expected, as an additive summation model with a steeper transducer behaves similar to a probability summation model with a shallower transducer. From our results we are not able to conclude whether the parts of an RF pattern are combined to improve performance by additive or by probability summation. Importantly however, this means that we are not able to *reject* the probability summation account.

For the Quad RF condition (Fig. 8) all three of the tested observers come out in favour of the probability summation model. Similar to the Single RF condition, the transducers for the best-fitting additive summation model are more steep than those for the best-fitting probability summation model (1.4 vs. linear). Once again however there is only a small difference between the quality of the fits from the different models. Even in this case where the modulations are being combined between separate RF patterns, we are not able to confidently state that detection occurs under probability summation rather than additive summation.

**Table B2**
Table showing the mean parameters across observers and combined log likelihoods and mean deviances obtained from a maximum-likelihood fit of the probability summation model to the Single RF data.

| Model | $\tau$ | $g_1$ | $g_2$ | $g_3$ | $g_4$ | $\sigma_1$ | $\sigma_2$ | $\sigma_3$ | $\lambda$ (%) | $\Sigma(\log L)$ | $\bar{D}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| T0G1S0 | | 106.2 | | | | | | | 1.01 | −944.1 | 106.8 |
| T1G1S0 | 1.23 | 99.1 | | | | | | | 1.09 | −942.0 | 106.0 |
| T0G4S0 | | 113.4 | 110.5 | 106.1 | 98.2 | | | | 1.02 | −943.3 | 106.5 |
| T1G4S0 | 1.38 | 106.9 | 99.1 | 102.3 | 85.5 | | | | 1.18 | −941.0 | 105.6 |
| T0G1S3 | | 180.7 | | | | 1.4 | 1.4 | 1.8 | 0.88 | −943.5 | 106.6 |
| T1G1S3 | 1.28 | 783.5 | | | | 4.9 | 5.0 | 14.6 | 0.95 | −941.9 | 105.9 |

**Table B4**
Table showing the mean parameters across observers and combined log likelihoods and mean deviances obtained from a maximum-likelihood fit of the probability summation model to the Quad RF data.

| Model | $\tau$ | $g_1$ | $g_2$ | $g_3$ | $g_4$ | $\sigma_1$ | $\sigma_2$ | $\sigma_3$ | $\lambda$ (%) | $\Sigma(\log L)$ | $\bar{D}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| T0G1S0 | | 53.1 | | | | | | | 0.91 | −705.6 | 122.9 |
| T1G1S0 | 1.01 | 52.9 | | | | | | | 0.92 | −705.7 | 123.0 |
| T0G4S0 | | 49.1 | 76.8 | 43.9 | 50.2 | | | | 1.01 | −702.7 | 121.0 |
| T1G4S0 | 1.10 | 47.8 | 74.9 | 42.6 | 48.1 | | | | 1.07 | −702.8 | 121.0 |
| T0G1S3 | | 73.3 | | | | 0.8 | 1.2 | 1.4 | 0.80 | −704.5 | 122.2 |
| T1G1S3 | 1.04 | 72.7 | | | | 0.8 | 1.1 | 1.4 | 0.81 | −704.6 | 122.3 |

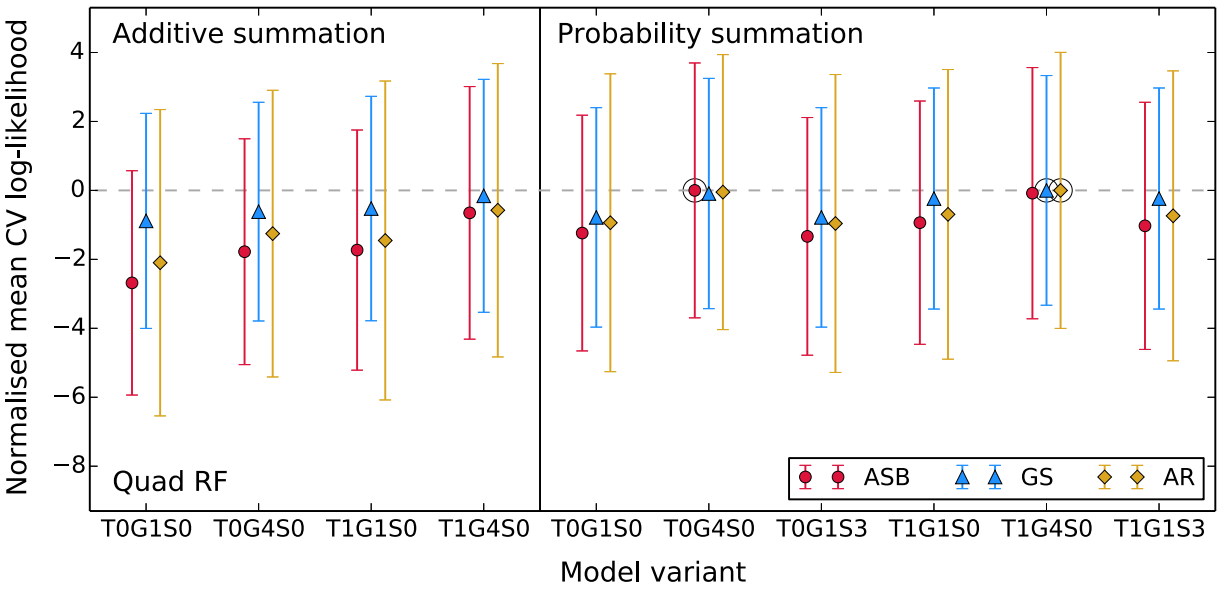


**Fig. 8.** Output of the cross-validation procedure for the Quad RF condition. For more details see Fig. 7 caption.

## 4. Discussion

By measuring the receiver operating characteristics for detecting RF pattern modulations we have demonstrated that SDT is the correct framework to use when modelling the results from experiments that use RF pattern stimuli. This means that comparisons made against previous probability summation models based on HTT must be re-examined. In our summation experiments we find that the summation slopes found using an interleaved design (with extrinsic uncertainty) are much steeper than those from a blocked design (without extrinsic uncertainty). In fact, we find almost no summation in our blocked experiments. We explain this by suggesting that observers focus on locally detecting just one cycle of the RF pattern *even when multiple cycles are presented*, rather than exploiting their ability to combine information over multiple cycles. Some previous studies have reported steeper slopes with a blocked design (Bell & Badcock, 2008; Dickinson et al., 2010, 2012; Loffler et al., 2003; Schmidtmann et al., 2012; Tan et al., 2013). One explanation for this discrepancy is that these studies typically randomised the phase of their RF stimuli within a block. This introduces uncertainty effects that will steepen the threshold vs. $n$ slope, as the observer cannot reliably monitor only the appropriate mechanisms on each trial (making the task more like the interleaved condition).

In the interleaved experiments the strategy that we suggest our observers adopted for the blocked experiments will not work (as the observer cannot rely on the modulation being applied to any one cycle) and so the observer monitors the entire stimulus. The results from this interleaved condition were closer to those found in previous studies (Table A1). These slopes can be accounted for by either an additive *or* a probability summation model under SDT (Kingdom et al., 2015). Where these previous studies were able to reject probability summation this is because they were using the version of the model derived under HTT (typically the equation provided by Quick, 1974). Even after performing extensive modelling, the only conclusion we can draw is that once an SDT probability summation model is applied it is able to behave very much like the additive summation model. For the data we collected and the analysis we conducted we are not able to reject the probability summation model. If previous studies that found similar summation slopes had performed their modelling under SDT they would have likely been unable able to reject the probability summation model either. We also find little difference in summation between our Single and Quad RF stimulus designs, suggesting that summation *between* RF patterns may occur in the same way as summation within an RF. Previous research has also suggested that the detection of RF patterns may not differ from that of modulated line stimuli (Mullen, Beaudot, & Ivanov, 2011). Poirier and Wilson (2006) developed a model that specifically accounted for the detection of RF pattern modulation. In brief, the model detects an RF pattern by generating a map of local curvatures and then matching that map to an RF template. Our results suggest that the process underlying detection may not be specific to an RF stimulus. It may be the case that the curvature map alone is sufficient for detection using only local cues, or that global templates can be generated for arbitrary shapes (in our case perhaps the shape of

the region in the centre of our Quad RF display that is bounded by the inner edges of the four RF patterns).

The findings we report are for RF4 patterns. Steeper summation slopes that may be incompatible with probability summation (even under SDT) have been found previously for RF3 patterns (e.g. Loffler et al., 2003). In general summation slopes become shallower for increasing radial frequencies (Loffler et al., 2003; Schmidtmann et al., 2012). We would therefore expect that for higher radial frequencies than the one tested here the probability summation model would become an increasingly viable candidate. Although our results suggest that the detection of RF patterns may not be a global process, it is important to note that these results indicate what is happening at threshold. Another source of evidence used to argue that RF patterns are processed globally comes from adaptation studies that investigate suprathreshold perception (Anderson, Habak, Wilkinson, & Wilson, 2007; Bell & Badcock, 2008; Bell, Gheorghiu, & Kingdom, 2009; Bell et al., 2010). It may be the case that global RF mechanisms exist but are not more sensitive than local detection at threshold. If it is the case that RF patterns are detected locally at threshold by neurotypical observers then this has implications for studies that use the detection of RF patterns to distinguish between local and global strategies in clinical populations such as in autism (Perreault, Habak, Lepore, Mottron, & Bertone, 2015).

Our result also has implications for other areas in which data are tested against a HTT probability summation model. In the case where the use of a HTT framework is not supported (e.g. by ROC curves) then one should also consider applying a SDT probability summation model. We are not the first to suggest this (Tyler & Chen, 2000), and several studies in contrast detection have already demonstrated the differences this can make (e.g. Meese & Summers, 2012). More recent work on the combination of local orientations into a coherent texture has also shown that probability summation under SDT may be involved (Schmidtmann, Jennings, Bell, & Kingdom, 2015). We hope that the further explanation and Matlab code that we provide in our Supplementary materials will encourage other researchers to consider applying these models to their results. The similarities between the predictions possible under the additive and probability summation models do present a challenge however. To solve this, and as a way of answering the outstanding question left by our study, we suggest that it is possible to design a future experiment that selectively tests conditions to distinguish between the additive and probability summation models in the shortest number of trials. This could be done with entropy-minimisation, using the same principles on which the Psi method is based (Kontsevich & Tyler, 1999).

## Acknowledgments

We would like to thank our two anonymous reviewers for their helpful comments and criticism. This work was supported by Canadian Institutes of Health Research (CIHR) grants awarded to R.F.H. (MT108-18 and MOP53346) and a Natural Sciences and Engineering Research Council of Canada (NSERC) grant awarded to F.K. (121713-11).

## Appendix A. Summation slopes from previous studies

Table A1 shows a summary of the summation slopes (log threshold vs. log number of modulated cycles) reported in previous studies along with the HTT probability summation model predictions.

**Table A1**
Summary of summation slopes measured in RF summation studies compared to the $-1/\beta$ slopes predicted by the HTT probability summation model. The slopes marked with asterisks (*) are mean values.

| Reference | $f$ | Summation slope | |
|---|---|---|---|
| | | HTT PS | Empirical |
| Hess et al. (1999) | 4 | −0.33* | −0.52* |
| | 8 | −0.33* | −0.44* |
| Loffler et al. (2003) | 3 | −0.33 | −0.86 |
| | 5 | −0.33 | −0.69 |
| | 10 | −0.33 | −0.64 |
| | 24 | −0.33 | −0.31 |
| Bell and Badcock (2008) | 5 | −0.41 | −0.76* |
| | 6 | −0.41 | −0.69* |
| Dickinson et al. (2010) | 3 | −0.43* | −0.75* |
| Dickinson et al. (2012) | 4 | −0.52* | −0.80* |
| Schmidtmann et al. (2012) | 3 | −0.33 | −0.65 |
| | 4 | −0.33 | −0.53 |
| | 5 | −0.33 | −0.61 |
| | 20 | −0.33 | −0.40 |
| Tan et al. (2013) | 3 | −0.53* | −0.92* |

## Appendix B. Mean parameters from model-fitting

The parameters obtained for each observer from the maximum-likelihood fitting of our summation models are provided as a Supplementary material. Presented in this appendix are the mean parameters averaged across observers, along with the combined log likelihoods and deviances ($D$, calculated relative to the likelihood of the saturated model for the holdout data) for each model.

## Appendix C. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.visres.2016.03.003.

## References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716–723.
Albright, T. D., Desimone, R., & Gross, C. G. (1984). Columnar organization of directionally selective cells in visual area MT of the macaque. *Journal of Neurophysiology, 51*(1), 16–31.
Anderson, N. D., Habak, C., Wilkinson, F., & Wilson, H. R. (2007). Evaluating shape after-effects with radial frequency patterns. *Vision Research, 47*, 298–308. http://dx.doi.org/10.1016/j.visres.2006.02.013.
Anzai, A., Peng, X., & Van Essen, D. C. (2007). Neurons in monkey visual area V2 encode combinations of orientations. *Nature Neuroscience, 10*(10), 1313–1321. http://dx.doi.org/10.1038/nn1975.
Baldwin, A. S., & Meese, T. S. (2015). Fourth-root summation of contrast over area: No end in sight when spatially inhomogeneous sensitivity is compensated by a witch's hat. *Journal of Vision, 15*(15), 1–12 [4].
Bell, J., & Badcock, D. R. (2008). Luminance and contrast cues are integrated in global shape detection with contours. *Vision Research, 48*, 2336–2344. http://dx.doi.org/10.1016/j.visres.2008.07.015.
Bell, J., & Badcock, D. R. (2009). Narrow-band radial frequency shape channels revealed by sub-threshold summation. *Vision Research, 49*, 843–850. http://dx.doi.org/10.1016/j.visres.2009.03.001.
Bell, J., Badcock, D. R., Wilson, H., & Wilkinson, F. (2007). Detection of shape in radial frequency contours: Independence of local and global form information. *Vision Research, 47*, 1518–1522. http://dx.doi.org/10.1016/j.visres.2007.01.006.
Bell, J., Gheorghiu, E., Hess, R. F., & Kingdom, F. A. A. (2011). Global shape processing involves a hierarchy of integration stages. *Vision Research, 51*, 1760–1766. http://dx.doi.org/10.1016/j.visres.2011.06.003.
Bell, J., Gheorghiu, E., & Kingdom, F. A. A. (2009). Orientation tuning of curvature adaptation reveals both curvature-polarity-selective and non-selective mechanisms. *Journal of Vision, 9*(12), 1–11. http://dx.doi.org/10.1167/9.12.3 (3, Introduction).
Bell, J., Hancock, S., Kingdom, F. A. A., & Peirce, J. W. (2010). Global shape processing: Which parts form the whole? *Journal of Vision, 10*(6), 1–13. http://dx.doi.org/10.1167/10.6.16 [16, Introduction].
Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436.
Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference*. New York: Springer-Verlag.
Cadieu, C., Kouh, M., Pasupathy, A., Connor, C. E., Riesenhuber, M., & Poggio, T. (2007). A model of V4 shape selectivity and invariance. *Journal of Neurophysiology, 98*, 1733–1750. http://dx.doi.org/10.1152/jn.01265.2006.

Carlson, E. T., Rasquinha, R. J., Zhang, K., & Connor, C. E. (2011). A sparse object coding scheme in area V4. *Current Biology, 21*, 288–293. http://dx.doi.org/10.1016/j.cub.2011.01.013.

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron, 73*, 415–434. http://dx.doi.org/10.1016/j.neuron.2012.01.010.

Dickinson, J. E., Cribb, S. J., Riddell, H., & Badcock, D. R. (2015). Tolerance for local and global differences in the integration of shape information. *Journal of Vision, 15*(3), 1–24.

Dickinson, J. E., Han, L., Bell, J., & Badcock, D. R. (2010). Local motion effects on form in radial frequency patterns. *Journal of Vision, 10*(3), 1–15. http://dx.doi.org/10.1167/10.3.20 [20, Introduction].

Dickinson, J. E., McGinty, J., Webster, K. E., & Badcock, D. R. (2012). Further evidence that local cues to shape in RF patterns are integrated globally. *Journal of Vision, 12*(12), 1–17. http://dx.doi.org/10.1167/12.12.16 [16, Introduction].

Gallant, J. L., Connor, C. E., Rakshit, S., Lewis, J. W., & Van Essen, D. C. (1996). Neural responses to polar, hyperbolic, and cartesian gratings in area V4 of the macaque monkey. *Journal of Neurophysiology, 76*(4), 2718–2739.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences, 15*(1), 20–25. http://dx.doi.org/10.1016/0166-2236(92)90344-8.

Green, D. M., & Swets, J. A. (1966). The sensory threshold and psychophysical method. In *Signal detection theory and psychophysics* (pp. 117–148). New York: John Wiley & Sons.

Hegdé, J., & Van Essen, D. C. (2007). A comparative study of shape representation in macaque visual areas V2 and V4. *Cerebral Cortex, 17*, 1100–1116. http://dx.doi.org/10.1093/cercor/bhl020.

Hess, R. F., Achtman, R. L., & Wang, Y.-Z. (2001). Detection of contrast-defined shape. *Journal of the Optical Society of America A, 18*(9), 2220–2227. http://dx.doi.org/10.1364/JOSAA.18.002220.

Hess, R. F., Wang, Y. Z., & Dakin, S. C. (1999). Are judgements of circularity local or global? *Vision Research, 39*, 4354–4360. http://dx.doi.org/10.1016/S0042-6989(99)00153-4.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology, 160*, 106–154.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology, 195*, 215–243.

Jeffrey, B. G., Wang, Y.-Z., & Birch, E. E. (2002). Circular contour frequency in shape discrimination. *Vision Research, 42*, 2773–2779. http://dx.doi.org/10.1016/S0042-6989(02)00332-2.

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), 1–16. http://dx.doi.org/10.1167/15.5.1 [1].

Kingdom, F. A. A., & Prins, N. (2016). Model comparisons. In *Psychophysics: A practical introduction* (pp. 247–307). London: Academic Press: an imprint of Elsevier.

Kleiner, M., Brainard, D. H., & Pelli, D. G. (2007). What's new in Psychtoolbox-3? *Perception, 36*(ECVP Abstract Supplement).

Kontsevich, L. L., & Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research, 39*, 2729–2737.

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: An expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences, 17*(1), 26–49.

Loffler, G. (2008). Perception of contours and shapes: Low and intermediate stage mechanisms. *Vision Research, 48*, 2106–2127. http://dx.doi.org/10.1016/j.visres.2008.03.006.

Loffler, G., Wilson, H. R., & Wilkinson, F. (2003). Local and global contributions to shape discrimination. *Vision Research, 43*, 519–530. http://dx.doi.org/10.1016/S0042-6989(02)00686-7.

Meese, T. S. (2010). Spatially extensive summation of contrast energy is revealed by contrast detection of micro-pattern textures. *Journal of Vision, 10*(8), 1–21. http://dx.doi.org/10.1167/10.8.14 [14, Introduction].

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), 1–28. http://dx.doi.org/10.1167/12.11.9 [9, Introduction].

Mullen, K. T., Beaudot, W. H. A., & Ivanov, I. V. (2011). Evidence that global processing does not limit thresholds for RF shape discrimination. *Journal of Vision, 11*(3), 1–21. http://dx.doi.org/10.1167/11.3.6 [6, Introduction].

Nachmias, J. (1981). On the psychometric function for contrast detection. *Vision Research, 21*, 215–223.

Pasupathy, A., & Connor, C. E. (1999). Responses to contour features in macaque area V4. *Journal of Neurophysiology, 82*, 2490–2502.

Pasupathy, A., & Connor, C. E. (2002). Population coding of shape in area V4. *Nature Neuroscience, 5*(12), 1332–1338. http://dx.doi.org/10.1038/nn972.

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.

Perreault, A., Habak, C., Lepore, F., Mottron, L., & Bertone, A. (2015). Behavioral evidence for a functional link between low- and mid-level visual perception in the autism spectrum. *Neuropsychologia, 77*, 380–386. http://dx.doi.org/10.1016/j.neuropsychologia.2015.09.022.

Poirier, F. J. A. M., & Wilson, H. R. (2006). A biologically plausible model of human radial frequency perception. *Vision Research, 46*, 2443–2455.

Prins, N., & Kingdom, F. A. A. (2009). Palamedes: Matlab routines for analyzing psychophysical data. <www.palamedestoolbox.org> Accessed 07.01.13.

Quick, R. F. (1974). A vector-magnitude model of contrast detection. *Kybernetik, 16*, 65–67.

Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. *Nature Neuroscience, 3*, 1199–1204.

Sachs, M. B., Nachmias, J., & Robson, J. G. (1971). Spatial-frequency channels in human vision. *Journal of the Optical Society of America, 61*(9), 1176–1186.

Schmidtmann, G., Jennings, B. J., Bell, J., & Kingdom, F. A. A. (2015). Probability, not linear summation mediates the detection of concentric orientation-defined textures. *Journal of Vision, 15*(16), 1–19 [6].

Schmidtmann, G., Kennedy, G. J., Orbach, H. S., & Loffler, G. (2012). Non-linear global pooling in the discrimination of circular and non-circular shapes. *Vision Research, 62*, 44–56. http://dx.doi.org/10.1016/j.visres.2012.03.001.

Schmidtmann, G., Logan, A. J., Kennedy, G. J., Gordon, G. E., & Loffler, G. (2015). Distinct lower visual field preference for object shape. *Journal of Vision, 15*(5), 1–15. http://dx.doi.org/10.1167/15.5.18 [18].

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences, 104*(15), 6424–6429. http://dx.doi.org/10.1073/pnas.0700622104.

Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *Journal of the Royal Statistical Society Series B (Methodological), 39*(1), 44–47.

Tan, K. W. S., Dickinson, J. E., & Badcock, D. R. (2013). Detecting shape change: Characterizing the interaction between texture-defined and contour-defined borders. *Journal of Vision, 13*(14), 1–16. http://dx.doi.org/10.1167/13.14.12 [12].

Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience, 19*, 109–139.

Tanner, W. P. (1956). Theory of recognition. *The Journal of the Acoustical Society of America, 28*(5), 882–888.

Tsao, D. Y., & Livingstone, M. S. (2008). Mechanisms of face perception. *Annual Review of Neuroscience, 31*, 411–437. http://dx.doi.org/10.1146/annurev.neuro.30.051606.094238 [Mechanisms].

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.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behaviour* (pp. 549–586). Cambridge, MA: The MIT Press.

Van Essen, D. C., Anderson, C. H., & Felleman, D. J. (1992). Information processing in the primate visual system: An integrated systems perspective. *Science, 255*, 419–423.

Wang, Y.-Z., & Hess, R. F. (2005). Contributions of local orientation and position features to shape integration. *Vision Research, 45*(11), 1375–1383. http://dx.doi.org/10.1016/j.visres.2004.12.003.

Wilkinson, F., Wilson, H. R., & Habak, C. (1998). Detection and recognition of radial frequency patterns. *Vision Research, 38*, 3555–3568. http://dx.doi.org/10.1016/S0042-6989(98)00039-X.

Wilson, H. R. (1980). A transducer function for threshold and suprathreshold human vision. *Biological Cybernetics, 38*, 171–178.

Wilson, H. R., & Wilkinson, F. (2015). From orientations to objects: Configural processing in the ventral stream. *Journal of Vision, 15*(7), 1–10. http://dx.doi.org/10.1167/15.7.4 [4].

Yau, J. M., Pasupathy, A., Brincat, S. L., & Connor, C. E. (2013). Curvature processing dynamics in macaque area V4. *Cerebral Cortex, 23*, 198–209. http://dx.doi.org/10.1093/cercor/bhs004.