Winter Semester Project 2005

Seeing the Forest but Not the Trees

Reference: Prof. Sabine Süsstrunk
Supervisor: Roberto Costantini
Authors: Christian Cornaz and Mounir Krichane

Laboratory for Audiovisual Communications
Swiss Federal Institute of Technology
Lausanne, Switzerland

February 4, 2005

Abstract

In an experiment on orientation identification under crowding conditions, Parkes and her colleagues propose that crowding is not the result of masking, but rather a consequence of stimulus pooling. They suggest that the visual system performs an estimate of the mean orientation of the stimuli ensemble, which corresponds to performing a textural analysis of the stimuli. Our goal is to investigate Parkes *et al.* pooling model. Our study focuses on the effect on crowding due to the variation of the distance between a central stimulus and all surrounding stimuli, as well as the effect due to the variation of the size of all the stimuli in an ensemble.

1 Intoduction

There is a well-known expression in English that says: "seeing the trees but not the forest", which means that we tend to focus more on details rather than on the global view. The title of our paper is the opposite. "Seeing the forest but not the trees" means in our case that we perceive the global view, but we loose the details. In Fig. 1, the patches surrounding the central one act like the "forest", while the central patch acts like a "tree". When this image is viewed parafoveally, the orientation of the central patch is inaccessible; nonetheless an observer is still able to perceive a global orientation (Parkes *et al.*, 2001).

Figure 1: The central patch acts like a tree, and the surrounding patches like the forest.

In Parkes, the central patch is called the *target*, while the surrounding patches are called the *distractors*. The target is tilted from the horizontal, and the distractors are all horizontal. All patches are Gabor patches. When this stimulus is viewed parafoveally, an observer is unable to report the orientation of the central patch. This phenomenon is known as *crowding*. Parkes *et al.* showed that despite their inability to report the orientation of an individual patch, observers can reliably estimate the average orientation of all the patches. This implies that the local orientation signals are combined rather than lost (*pooling*). Their results suggest that such crowding is related to texture perception. Hence it is distinct from ordinary *masking*, in which the distractors corrupt the visual system's estimate of the target's tilt.

In Fig. 2, the reader can experience crowding. Here the dense string of letters within the word FOREST are together unresolvable, even when the letter E can clearly be distinguished on its own.

Figure 2: When fixating the dot on the top left, the E to the right can be easily identified. However, when fixating the dot underneath, the E seems to meld into a jumbled texture. This is the crowding effect of the surrounding items on the availability of an individual item.

The purpose of our study is to investigate Parkes *et al.* pooling model. Our aim is to study the effect on crowding due to the variation of the distance between the central patch and all the surrounding patches. Furthermore, we would like to explore the effect on crowding due to the variation of the size of all the patches (*scaling*).

This paper is organized as follows. In section 2 we give an explanation of Parkes *et al.* pooling model. The proposed approach for investigating this model is presented in section 3. Our experimental protocol is presented in section 4. The data analysis and the underlying statistical method described by Wichmann and Hill (2001) are presented in section 5. Finally, in section 6 we present our experimental results and draw some conclusions.

2 Parkes *et al.* pooling model

In Parkes, it is supposed that each target (or distractor) is perceived by the observer with a given error. These errors are independent and identically distributed. We indicate them as E_i, and we assume they are Gaussian Random Variables with zero mean and variance σ_e^2. The perceived orientation will be the mean orientation of the stimuli, corrupted by a later noise, here indicated by a zero mean Gaussian Random Variable X with variance equal to σ_l^2. The angle that is perceived is thus a Random Variable defined as:

$$\Theta = \frac{1}{(n_t + n_d)}\left[\sum_{i=1}^{n_t+n_d}(\theta_i + E_i)\right] + X$$

Here n_t and n_d is respectively the number of targets and distractors, and θ_i their orientation. If the distractors are all horizontal (i.e. $\theta_i = 0$) and the targets have the same angle θ_t, then we have:

$$\Theta = \frac{1}{(n_t + n_d)}\left[\sum_{i=1}^{n_t+n_d} E_i\right] + \frac{\theta_t n_t}{(n_t + n_d)} + X$$

A visual configuration is said to be clockwise if $\theta > 0$, counterclockwise in the opposite case. For a given angle $\theta_t = \mu$ we have:

$$P[\Theta > 0] = P\left[\frac{1}{(n_t + n_d)}\left[\sum_{i=1}^{n_t+n_d} E_i\right] + X > -\frac{n_t \mu}{(n_t + n_d)}\right]$$

Since E_i and X are two Gaussian Random Variables, their linear combination is also Gaussian. Thus the new Random Variable on the left handside of the previous equation is Gaussian with zero mean and variance equal to $\frac{\sigma_e^2}{(n_t + n_d)} + \sigma_l^2$. Furthermore, we may write the previous equation as:

$$P[\Theta > 0] = 1 - P\left[\frac{1}{(n_t + n_d)}\left[\sum_{i=1}^{n_t+n_d} E_i\right] + X \le -\frac{n_t \mu}{(n_t + n_d)}\right]$$

If X is a normal random variable with mean μ and variance σ^2, then its cumulative density function can be related to the standard normal $\Phi(x)$ in the following way:

$$P[X \le x] = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

Therefore in our case we have:

$$P[\Theta > 0] = 1 - \Phi\left[\frac{-n_t \mu}{(n_t + n_d)\sqrt{\frac{\sigma_e^2}{(n_t + n_d)} + \sigma_l^2}}\right]$$

Figure 3: Parkes *et al.* pooling model for orientation perception.

The performance of an observer on a psychophysical task is typically summarized by reporting one or more response thresholds. In their experiment, Parkes *et al.* use a tilt threshold for 75% accuracy, which is estimated from the observers' responses. The following graph shows the cumulative density function of the standard normal distribution:

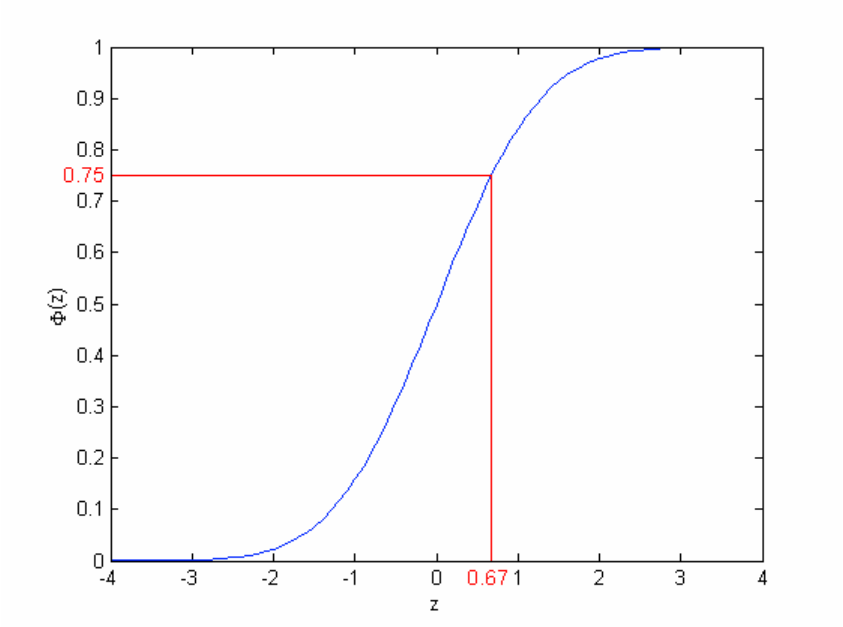

Figure 4: Cumulative density function of the standard normal distribution.

On this graph, we see that the probability that a standard normal variable has a value less than 0.75 is approximately equal to 0.67. Thus 75%-correct threshold can be calculated based on the equation given in Fig. 3 as the following:

$$P^{-1}(\mu) = 1 - \Phi(-0.67)$$

3 Proposed approach

In order to study the effect on crowding due to the variation of the distance between the central patch and all the surrounding patches, and to explore the effect due to scaling, we choose to use a configuration of seven Gabor patches: one central patch surrounded by six patches placed at each vertex of a hexagon. In addition to the parameter used by Parkes *et al.* defining the number of targets: in our case 1, 2, 3, or 4 targets, we defined two other parameters: distance, and size. We choose three distances: short, medium, and long, and two sizes: normal, and large. Note that the large scale can only be applied using the longest distance, since using a shorter distance would cause an overlap of the patches. Using this configuration we conducted a total of sixteen experiments for each of the two subjects. These experiments are summarized in the following table:

Short distance/normal scaling	1 target	2 targets	3 targets	7 targets
Medium distance/normal scaling	1 target	2 targets	3 targets	7 targets
Long distance/normal scaling	1 target	2 targets	3 targets	7 targets
Long distance/large scaling	1 target	2 targets	3 targets	7 targets

Figure 5: The sixteen experiments.

At first, this configuration allows us to recreate the experiment from Parkes *et al.*, where we study the effect on the tilt threshold due to the variation of the number of targets. In this case, the data analysis was done for each number of targets with a fixed distance/scale (e.g. the orange row in Fig. 5). Note that the experiment from Parkes *et al.* only corresponds to the first row in Fig. 5 (using what we call short distance and normal scaling). Secondly, it allows us to investigate the effect of the distance and the scaling on each configuration with a fixed number of targets (e.g. the blue column in Fig. 5).

4 Experimental protocol

At this point, we would like to present some essential information concerning the equipment used, the calibration of the Barco monitor, and after this we will dive into the core of our experiment.

In order to make this experiment in good conditions, we needed a dark room in which all veiling glare and luminance superposition due to external sources were avoided. A Barco monitor was also essential, since it is capable of displaying the different stimuli in good experimental conditions. This monitor had to be calibrated, for that purpose the relationship between the frame-buffer values (generated by the graphics card) and the light intensity emitted by the visual display was measured. The calibration was made using a Minolta colorimeter, and was validated using the Minolta CS-1000 spectroradiometer. The calibration allowed us to verify that the colors being used for the stimuli were displayed accurately. In order to calibrate the visual display we had to approximate the curve relating the relative intensity I of the light emitted from the CRT display, and the frame-buffer values. The following function is used for that purpose:

$$I = \alpha v^{\gamma} + \beta$$

Here α and β are two fitting parameters. At this point, we have a good knowledge about the display capability of our gamma-corrected monitor.

The stimuli were generated by a Cambridge Research Systems VSG graphics card with 12-bit luminance resolution and a frame rate of 84 Hz. The stimuli were displayed on a uniform grey background with a mean luminance of 10 cd/m^2. The individual patches were horizontal 2-cycle/degree Gabor patches (i.e. sine wave gratings windowed by a Gaussian) with a space constant equal to 1.4 times the wavelength of the grating (i.e. $\sigma = \lambda\sqrt{2}$ with $\lambda=1/12$) for the normal scaling. For the large scaling we multiplied σ by a factor of 1.4. The center–to-center separation of the central patch to every other patch was $d \cdot \lambda \cdot \sqrt{2}$ with d—{2.5 (short), 3 (medium), or 4 (long)}—this represents the distance parameter.

At this stage, we would like to quickly describe the general experimental procedure. The subject faces the monitor and is placed at a distance of 200 cm from the monitor. The subjects were the two authors (CC and MK). On each trial, the observer fixated a central point on the monitor and pressed a response button to see the next stimulus. An auditory

warning signal is played whenever the subject answer is erroneous. The array flashed for 100 ms randomly to the left or right of the fixation point at a horizontal eccentricity of 2.5° (parafoveal), measured from the fixation point to the central element of the array. The general pattern was a patch surrounded by 6 other patches placed on each vertex of a surrounding hexagon. In each block of trials, 1, 2, 3, or 7 of the patches were slightly tilted clockwise or counterclockwise from the horizontal (targets). The remaining patches (distractors) were all horizontal. Furthermore, we used three different distances: short, medium, and long, as well as two sizes: normal, and large. Thus each subject is submitted to a total of sixteen experiments. The observer had to press one of two buttons (forced choice) to report whether the array appeared tilted clockwise or counterclockwise. Inside each block, 40 trials at each of 5 levels of tilt from the horizontal were randomly interleaved. Hence each block comprised 200 trials, and was repeated three times resulting in a total of 600 trials for each session.

Figure 6: Four examples of hexagonal configuration of Gabor patches showing the different distances and scaling.

5 Data analysis

This section is extensively based on the original paper of Felix A. Wichmann and N. Jeremy Hill, *The psychometric function: I. Fitting, sampling, and goodness of fit*, 2001.

The performance of each of the two subjects in our psychophysical experiments are summarized by reporting a response threshold (level of tilt required to produce 75%-correct answer) and by characterization of the rate at which performance improves with increasing level of tilt. This measure is derived from a *psychometric function* (Fig. 7), which describes the dependence of the subject's performance on some physical aspect of

the stimulus. In our case it relates the probability of a correct response to the amount of tilt. In this section we explain how the data is fitted with a psychometric function in order to find the threshold. We then explain how we assess lack of fit using a Monte Carlo generation of the statistic's distribution and a set of goodness of fit tests.

5.1 Fitting psychometric functions

5.1.1 Finding the threshold

To model the process underlying experimental data, we usually assume the number of correct responses y_{in} in a given block i to be the sum of random samples from a Bernoulli process with a probability of success p_i. A model must then provide a psychometric function $\psi(x)$, which specifies the relationship between the underlying probability of a correct response p and the stimulus intensity x. A frequently used general form is:

$$\psi(x;\alpha,\beta,\gamma,\lambda) = \gamma + (1-\gamma-\lambda)F(x;\alpha,\beta)$$

Here F is the cumulative distribution function, $\theta = [\alpha, \beta, \gamma, \lambda]$ is the set of parameters defining the shape of the curve, γ is the lower bound of $\psi(x,\theta)$ (here we use 2AFC so $\gamma = 0.5$), and finally λ is the miss rate (rate at which the observer lapses, $0 \le \lambda \le 0.06$, $\lambda = 0$ denotes the perfect observer).

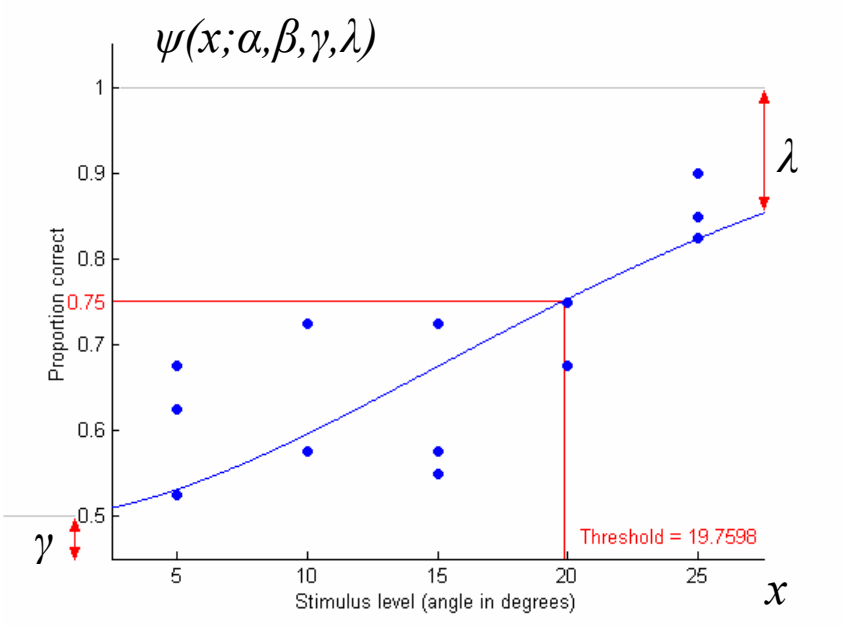

Figure 7: Fitting our data with a Weibull function. In this case 75% accuracy corresponds to a tilt threshold of 19.7598 degrees.

To determine each threshold in our case, we fitted a two-parameter Weibull function to our data (proportion correct answer for each block of trials, blue points in Fig. 7) and computed the inverse of that function for the desired performance level (75%-correct answer, see example in Fig. 7).

5.1.2 Parameters estimation

In order to perform the parameters estimation, we used the likelihood maximization. In our case, provided that the values of $\boldsymbol{y}$ are assumed to have been generated by Bernoulli processes, it is straightforward to compute a likelihood value for a particular set of parameters θ, given the observed values $\boldsymbol{y}$.

The maximum-likelihood estimator $\hat{\theta}$ of θ is simply that set of parameters for which the likelihood value is largest: $L(\hat{\theta}; y) \geq L(\theta; y)$ for all θ. Since the logarithm is a monotonic function, the log-likelihood function $l(\hat{\theta}; y) \geq l(\theta; y)$ is also maximized by the same estimator $\hat{\theta}$. In our case, it is given by:

$$l(\theta; y) = \sum_{i=1}^{K} \log \binom{n_i}{y_i n_i} + y_i n_i \log \psi(x_i; \theta) + (1 - y_i) n_i \log[1 - \psi(x_i; \theta)]$$

Bayesian priors are then used to constrain the fit according to the assumptions of our model. It is sometimes possible that the maximum-likelihood estimate $\hat{\theta}$ contains parameter values that are either nonsensical or inappropriate ($\lambda \leq 0$ or $\lambda \geq 0.06$). In both cases, it would be better for the fitting algorithm to return parameter vectors that may have a lower log-likelihood than the global maximum, but that contain more realistic values. Bayesian priors provide a mechanism for constraining parameters within realistic ranges, based on the experimenter's prior beliefs about the likelihood of particular values. A prior is simply a relative probability distribution $W(\boldsymbol{\theta})$, specified in advance, which weights the likelihood calculation during fitting. The fitting process therefore maximizes $W(\boldsymbol{\theta})L(\boldsymbol{\theta};\boldsymbol{y})$ or $\log W(\boldsymbol{\theta}) + l(\boldsymbol{\theta};\boldsymbol{y})$ instead of the unweighted metrics.

We avoid bias caused by observers' lapses in the following manner. Usually, we fix $\lambda = 0$, so that the upper bound of $\psi(x, \boldsymbol{\theta})$ is always 1.0. Doing so, we assume that observers make no stimulus-independent errors. With the following example taken from the original paper, we show the importance of letting the parameter λ free during fitting.

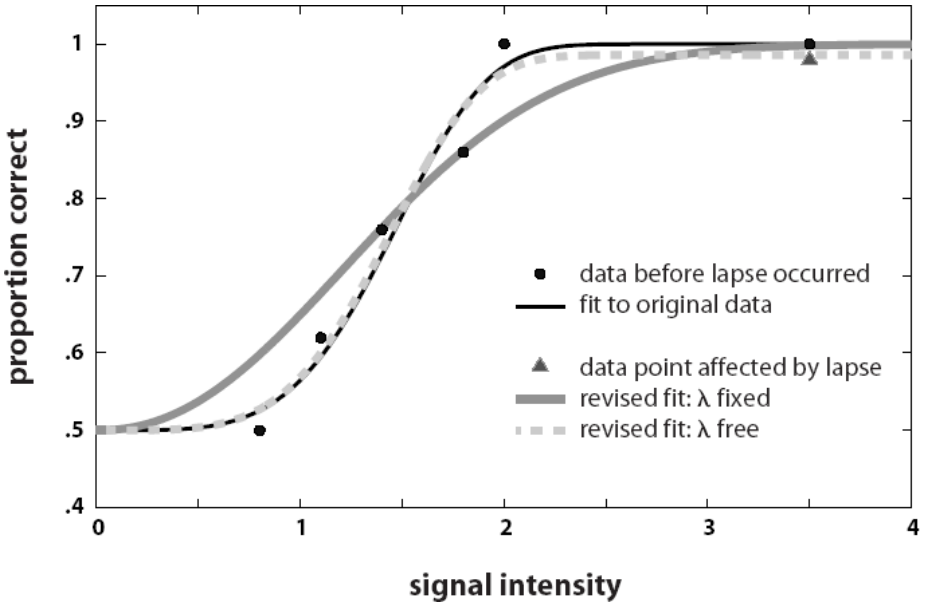

Figure 8: The importance of letting the parameter λ free during fitting.

The dark circles indicate the proportion of correct responses made by an observer in six blocks of trials in a 2AFC visual detection task. Each datapoint represents 50 trials, except for the last one, at stimulus value 3.5, which represents 49 trials: the observer still has one trial to perform to complete the block. If we were to stop here and fit a Weibull function to the data, we would obtain the curve plotted as a dark solid line. Whether or not λ is fixed at 0 during the fit, the maximum-likelihood parameter estimates are the same: $\{\hat{\alpha} = 1.573,\ \hat{\beta} = 4.360,\ \hat{\lambda} = 0\}$. Now suppose that, on the 50th trial of the last block, the observer blinks and misses the stimulus, is consequently forced to guess, and happens to guess wrongly. The new position of the datapoint at stimulus value 3.5 is shown by the light triangle: it has dropped from 1.00 to .98 proportion correct.

The solid light curve shows the results of fitting a two parameters psychometric function (i.e. allowing α and β to vary, but keeping λ fixed at 0). The new fitted parameters are $\{\hat{\alpha} = 2.604,\ \hat{\beta} = 2.191\}$. Note that the slope of the fitted function has dropped dramatically in the space of one trial—in fact, from a value of 1.045 to 0.560. If we allow λ to vary in our new fit, however, the effect on parameters is slight—$\{\hat{\alpha} = 1.543,\ \hat{\beta} = 4.347,\ \hat{\lambda} = 0.014\}$—and thus, there is little change in slope: $\mathrm{d}\hat{F}/\mathrm{d}x$ evaluated at $x = F_{0.5}^{-1}$ is 1.062.

The misestimating of parameters shown in Figure 8 is a direct consequence of the binomial log-likelihood error metric because of its sensitivity to errors at high levels of predicted performance: since $\psi(x;\boldsymbol{\theta}) \Rightarrow 1$, so, in the third term of the previous equation, $(1-y_i)n_i \log[1-\psi(x_i;\theta)] \Rightarrow -\infty$ unless the coefficient $(1-y_i)n_i$ is 0. Since y_i represents observed proportion correct, the coefficient is 0 as long as performance is perfect. However, as soon as the observer lapses, the coefficient becomes nonzero and allows the large negative log term to influence the log-likelihood sum, reflecting the fact that

observed proportions less than 1 are extremely unlikely to have been generated from an expected value that is very close to 1. Log-likelihood can be raised by lowering the predicted value at the last stimulus value, $\psi(3.5;\boldsymbol{\theta})$. Given that λ is fixed at 0, the upper asymptote is fixed at 1.0; hence, the best the fitting algorithm can do in our example to lower $\psi(3.5;\boldsymbol{\theta})$ is to make the psychometric function shallower.

Finally, one could say that fitting a tightly constrained λ is intended as a heuristic to avoid bias in cases of non-stationary observer behavior. It is, as well, to note that the estimated parameter $\hat{\lambda}$ is, in general, not a very good estimator of a subject's true lapse rate. Lapses are rare events, so there will only be a very small number of lapsed trials per data set. Furthermore, their directly measurable effect is small, so that only a small subset of the lapses that occur (those at high x values where performance is close to 100%) will affect the maximum-likelihood estimation procedure; the rest will be lost in binomial noise.

In addition, Wichmann and Hill have shown that for their simulations $N = 120$ appears too small a number of trials to be able to obtain reliable estimates of thresholds and slopes for some sampling schemes, even if the variable-λ fitting regime is employed. Other studies have shown that the total number of experimental trials should be at least as high as 200 to obtain reliable estimates of thresholds.

5.2 Goodness of fit

A problem arises with traditional goodness-of-fit methods, since psychophysical data tend to consist of small numbers of points and it is, hence, by no means certain that such tests are accurate. A promising technique that offers a possible solution is *Monte Carlo simulation*, which is potentially well suited to the analysis of psychophysical data, because its accuracy does not rely on large numbers of trials.

Lack of fit may result from failure of one or more of the assumptions of one's model. First and foremost, lack of fit between the model and the data could result from an inappropriate functional form for the model (F is inappropriate). Second, our assumption that observer responses are binomial may be false (e.g. there might be serial dependencies between trials within a single block). Third, the observer's psychometric function may be nonstationary during the course of the experiment, be it due to learning or fatigue.

Usually, inappropriate models and violations of independence result in *overdispersion*: "bad fits" in which datapoints are significantly further from the fitted curve than was expected. Experimenter bias in data selection (e.g. informal removal of outliers), on the other hand, could result in *underdispersion*: fits that are "too good to be true," in which datapoints are significantly closer to the fitted curve than one might expect.

For our experiment we only focus on the case of overdispersion. We use a set of goodness-of-fit tests for psychometric functions that rely on different analyses of the residual differences between data and fit (i.e. sum of squares) and a Monte Carlo generation of the statistic's distribution, against which to assess lack of fit.

5.2.1 Assessing overdispersion

Summary statistics measure closeness of the data set as a whole to the fitted function. In maximum-likelihood parameter estimation, the parameter vector $\hat{\theta}$ returned by the fitting routine is such that $L\left(\hat{\theta}; y\right) \geq L(\theta; y)$ for all θ. Thus, whatever error metric Z is used to assess goodness-of-fit, $Z\left(\hat{\theta}; y\right) \geq Z(\theta; y)$ should hold for all θ.

5.2.2 Deviance (global measure)

The log-likelihood ratio, or *deviance*, is a monotonic transformation of likelihood and therefore fulfills the criterion set out in the previous equation. Deviance is used to assess goodness of fit, rather than likelihood or log-likelihood directly, because, for correct models, deviance for binomial data is asymptotically distributed as χ^2_K, where K denotes the number of datapoints (blocks of trials).

The distribution of deviances is obtained using Monte Carlo simulation in the following way. We generate B data sets y_i^* using the best-fitting psychometric function $\psi\left(x, \hat{\theta}\right)$. Then for each data set y_i^* we compute the deviance D_i^* yielding the deviance distribution D^* (which should be expected from an observer whose correct responses are binomially distributed with success probability $\psi\left(x, \hat{\theta}\right)$). Let D_{emp} denote the deviance of our empirically obtained data set, then if $D_{emp} > D^{*(.975)}$ (two-sided 95% confidence interval) we are in the presence of overdispersion. Note that the number of data sets B must be large ($B \geq 10000$). The following figure shows an example of assessing goodness of fit with deviance. In this example, the test would be negative (we are not in presence of overdispersion), since $D_{emp} < D^{*(.975)}$.

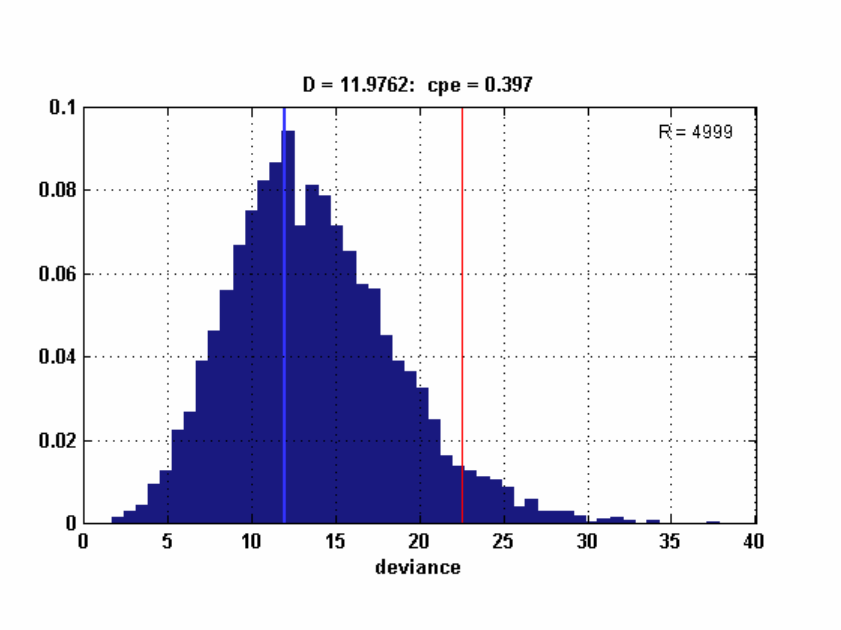

Figure 9: Assessing goodness of fit with deviance. PDF of *D* (dark blue), D_{emp} (light blue), and the 95% confidence interval (red).

5.2.3 Deviance residuals (model checking)

One of the most effective ways of identifying an incorrect model is to examine the agreement between individual datapoints and the corresponding model prediction. One simple way of looking at the residuals is to compute the correlation coefficient between the residuals and the **p** values predicted by one's model. This allows the identification of a systematic (linear) relation between deviance residuals **d** and model predictions **p**, which would suggest that the chosen functional form of the model is inappropriate (*F* is inappropriate). The following figure shows an example of model checking using deviance residuals. In this example, the test would be negative (*F* is not inappropriate), since *r* is inside the two-sided 95% confidence interval.

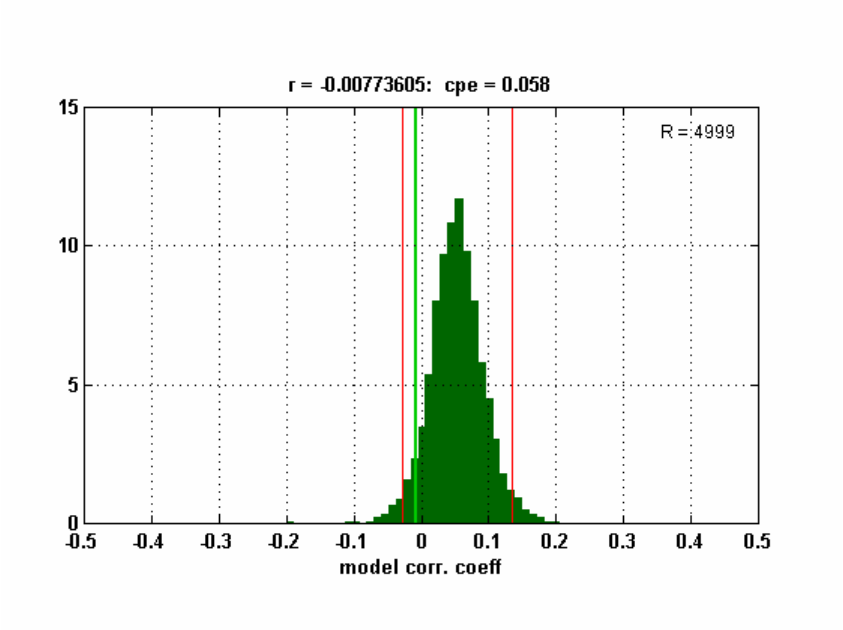

Figure 10: Checking the model with deviance residuals. The correlation coefficient *r* (light green), and the 95% confidence interval (red).

5.2.4 Learning (nonstationary observer performance)

Analysis of the deviance residuals **d** as a function of temporal order can be used to show perceptual learning. The approach is equivalent to that described for model checking, except that the correlation coefficient of deviance residuals is assessed as a function of the order in which the data were collected (indices **k**). Assuming that perceptual learning improves performance over time, one would expect the fitted psychometric function to be an average of the poor earlier performance and the better later performance. Deviance residuals should thus be negative for the first few datapoints and positive for the last ones. As a consequence, the correlation coefficient *r* of deviance residuals **d** against their indices **k** is expected to be positive if the subject's performance improved over time. The following figure shows an example of perceptual learning identification using deviance residuals. In this example, the test would be negative (there was no perceptual learning), since *r* is inside the two-sided 95% confidence interval.

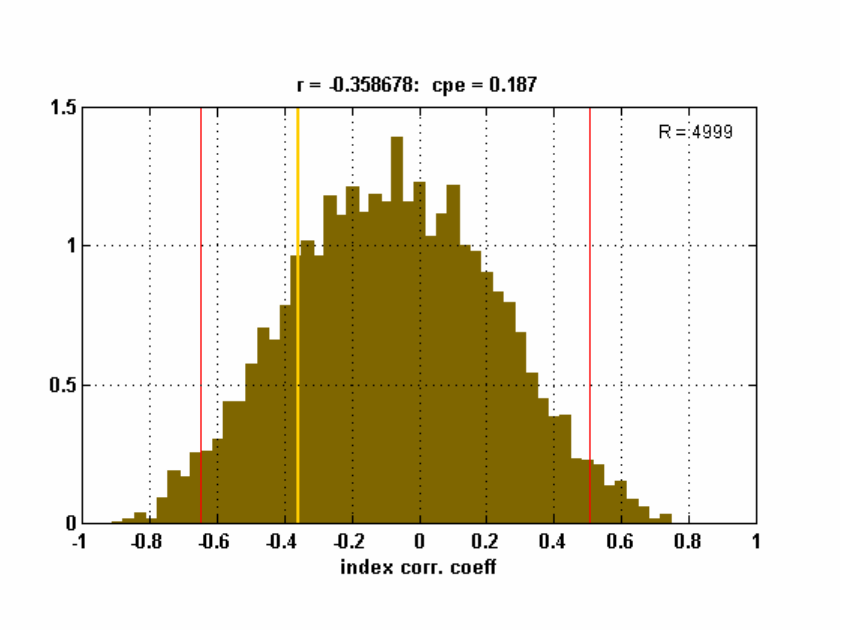

Figure 11: Identifying perceptual learning with deviance residuals. The correlation coefficient r (yellow), and the 95% confidence interval (red).

5.3 Software

In our experiment, psychometric functions were fitted using the *psignifit* toolbox version 2.5.6 for MATLAB (see http://bootstrap-software.org/psignifit/) which implements the maximum-likelihood method described by Wichmann and Hill (2001). Moreover, confidence intervals were found by the BC_a bootstrap method implemented by *psignifit*, based on 999 simulations.

6 Results

The first set of figures (Fig. 12, 13, and 14) allows us to study the effect on the tilt threshold due to the variation of the number of targets for each of the three distances and the large scaling. And the second set of figures (Fig. 15 and 16) allows us to investigate the effect of the distance and the scaling on each configuration with a fixed number of targets. Refer to the section 3 entitled "Proposed approach" for more information.

The upper-left plots (red) in Fig. 12 and 13 correspond to Parkes *et al.* experiment (short distance, normal scaling). In this case, thresholds fell as the number of targets rose. The data are well fitted on log-log axes by a line with a slope of -1, as predicted by the pooling model. The upper-right and bottom-left plots (respectively green and blue) in Fig. 12 and 13 were obtained using a set of stimuli with a longer distance between the central patch and all the surrounding patches compared to Parkes (medium and large distances). Here the thresholds still fell as the number of targets rose, and the data are also well fitted

on log-log axes by a line with a slope of -1, even with the largest distance. These results show that the distance between the central patch and all the surrounding patches does not affect Parkes *et al.* pooling model, even if it is 1.6 times longer than the original. In other words, crowding is still present when using these distances, and pooling still takes place. The bottom-right plots in Fig. 12 and 13 show the three other plots together to facilitate the comparison. As one can see, the three plots corresponding to the three distances are quite similar, especially for the short and medium distances (respectively green and blue). All these results are true for both subjects.

There are however differences between the two subjects: the plots for the two subjects have the same general shape but not the same amplitude, and variances for Subject MK are much larger in some cases. The difference in amplitude comes from the fact that the two subjects don't show the same performance when it comes to orientation identification under crowding conditions. Typically, Subject MK is able to reach smaller thresholds than Subject CC. The larger variances for Subject MK are a consequence of his performance. In order to achieve a small threshold, the proportion of correct answers must be close to one. When these proportions are all above 0.75 the fitting is harder to compute, and the resulting threshold has a large variance. This phenomenon happened to Subject MK when the stimuli contained a large number of targets. Note that this phenomenon may also be observed for Subject CC, for stimuli containing only one target. It is also a consequence of his performance: the orientation of the target was very hard to identify, thus his proportion of correct answers were all bellow 0.75, yielding a hard-to-compute fitting.

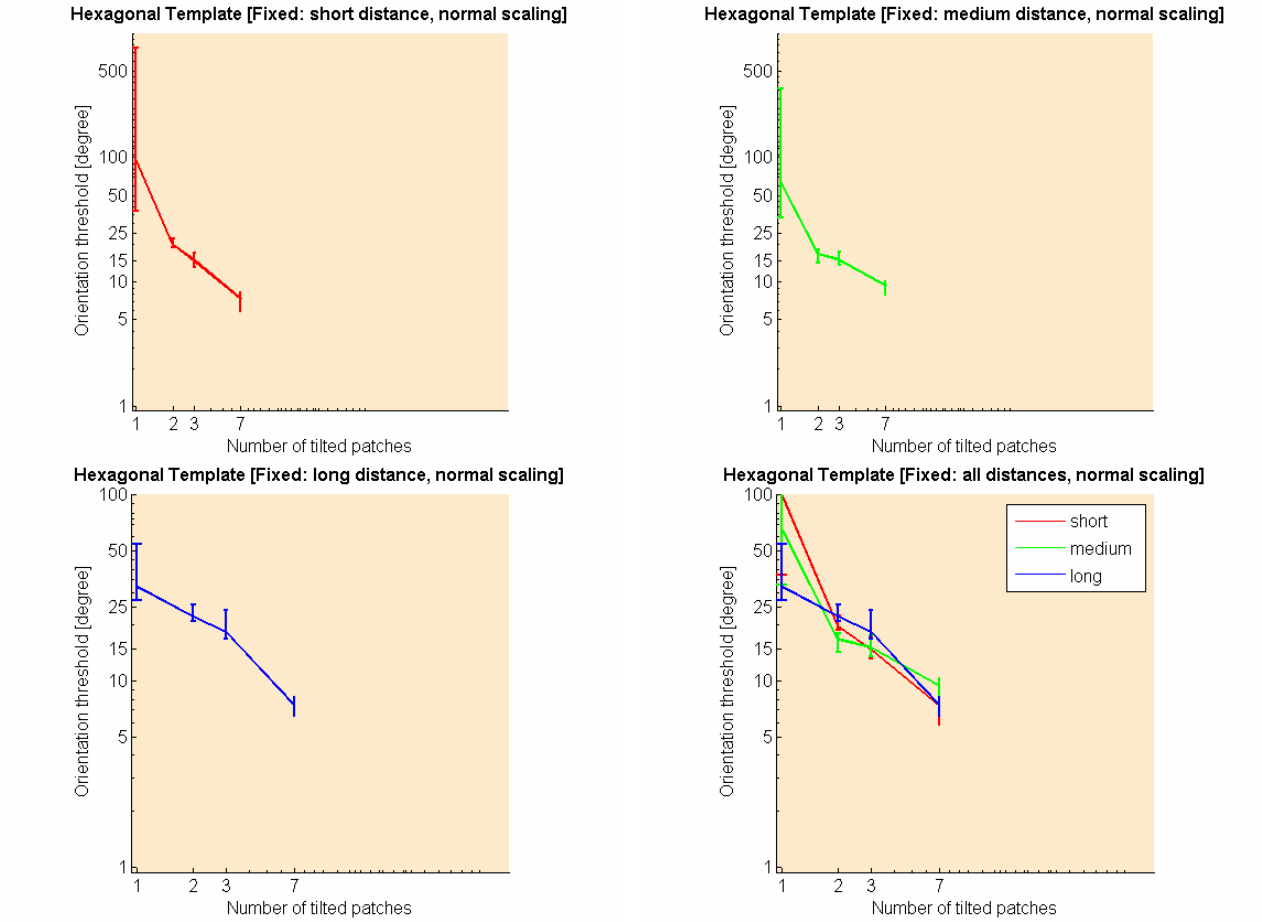

Figure 12: The effect on the tilt threshold due to the variation of the number of targets for each of the three distances [Subject CC].

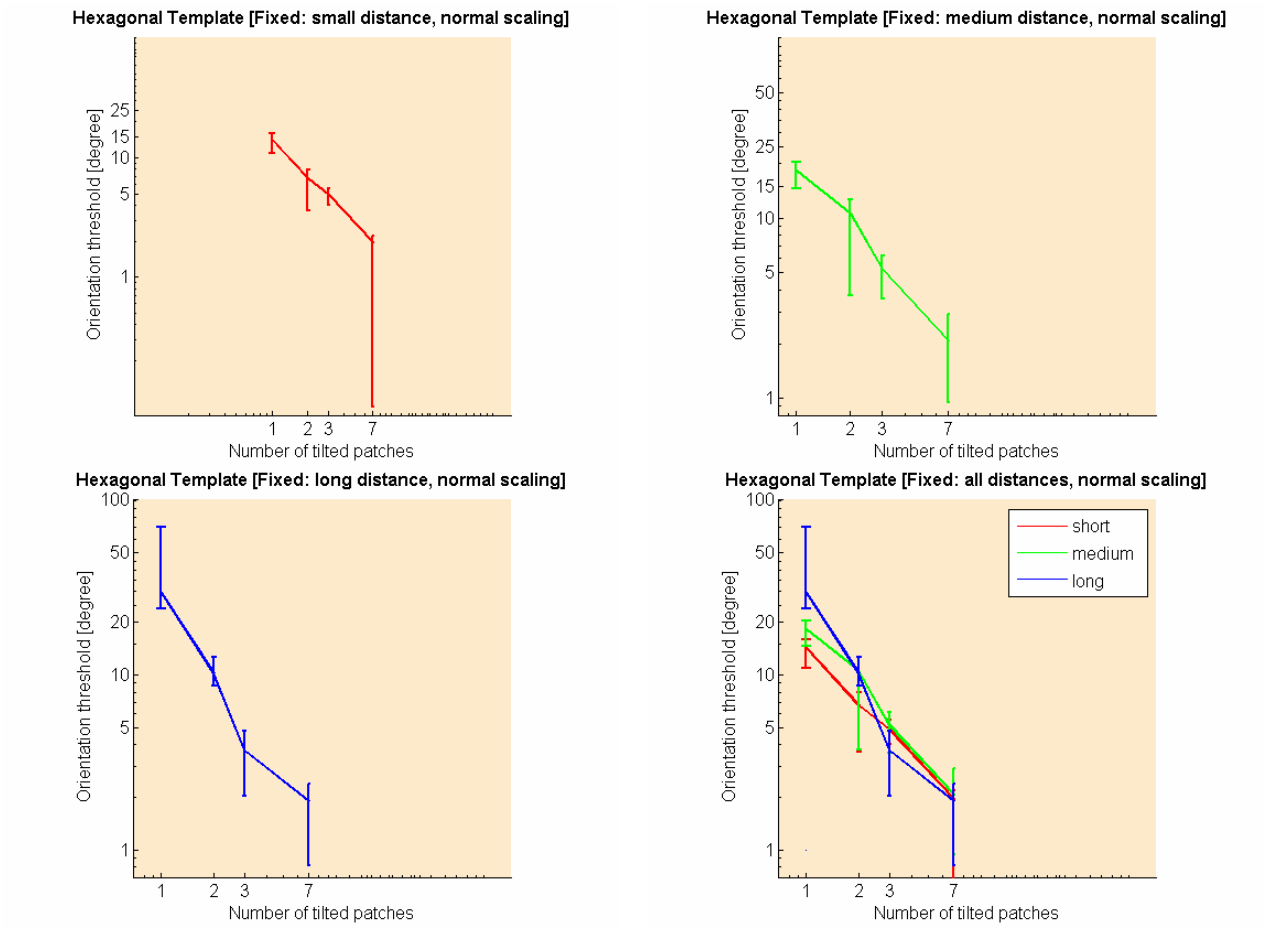

Figure 13: The effect on the tilt threshold due to the variation of the number of targets for each of the three distances [Subject MK].

The two magenta plots in Fig. 14 were obtained using a set of stimuli with the longest distance between the central patch and all the surrounding patches and the large scaling. Here the thresholds still fell as the number of targets rose, and the data are also well fitted on log-log axes by a line with a slope of -1. These results show that the size of all the patches does not affect Parkes *et al.* pooling model, even if it is 1.4 times larger than the original.

The only consequence of increasing the overall size of the stimuli is that it reduces the tilt thresholds. Indeed, Subject CC was able to reach a tilt threshold lower than 5°, and Subject MK was able to reach a tilt threshold lower than 1°. This can simply be explained by the fact that bigger stimuli are easier to perceive, hence the global orientation is simpler to identify. Here the large variances for Subject MK can be explained as stated above.

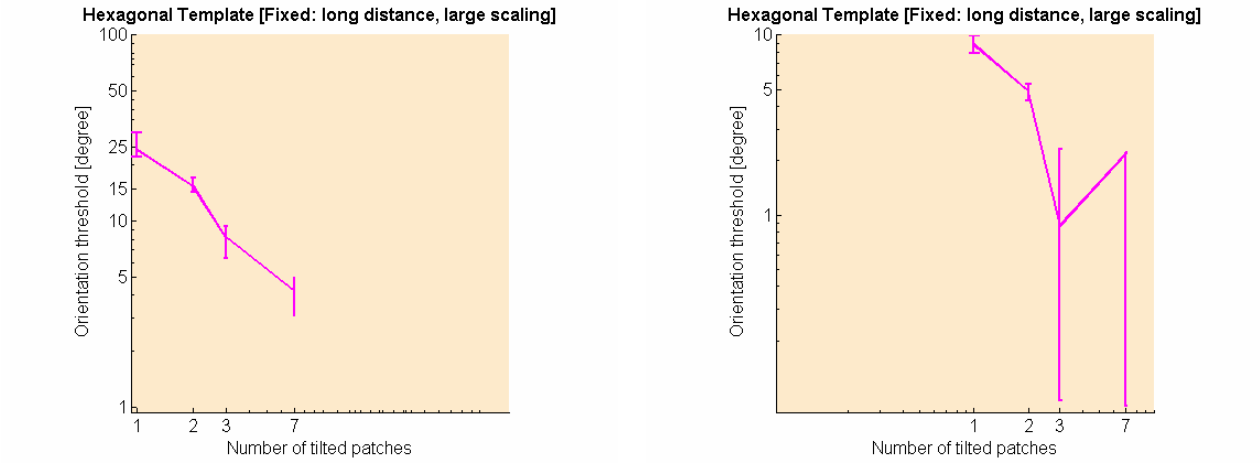

Figure 14: The effect on the tilt threshold due to the variation of the number of targets for the large scaling [Subjects CC (left) and MK (right)].

The plots in Fig. 15 and 16 show the effect on the tilt threshold due to the variation of the distance and the scaling on each configuration with a fixed number of targets. The X-coordinate of each plot show our four configurations (short, medium, and long distances, and long distance + large scaling). The Y-coordinate of each plot show the corresponding tilt threshold.

By observing the two upper plots in Fig. 15 and 16, one can see that for a small number of targets (1 or 2 targets) the thresholds tend to rise when the distance is increased. In this case, it shows that the two subjects were having more difficulty identifying a global orientation when the distance was increased. Here the density of the patches seems to help the observer identify the orientation. This is less true for a large number of targets (3 or 7 targets). The two bottom plots in Fig. 15 and 16 show that the thresholds stay more or less constant when the distance is increased. In all cases, thresholds fell when the size of the patches was increased.

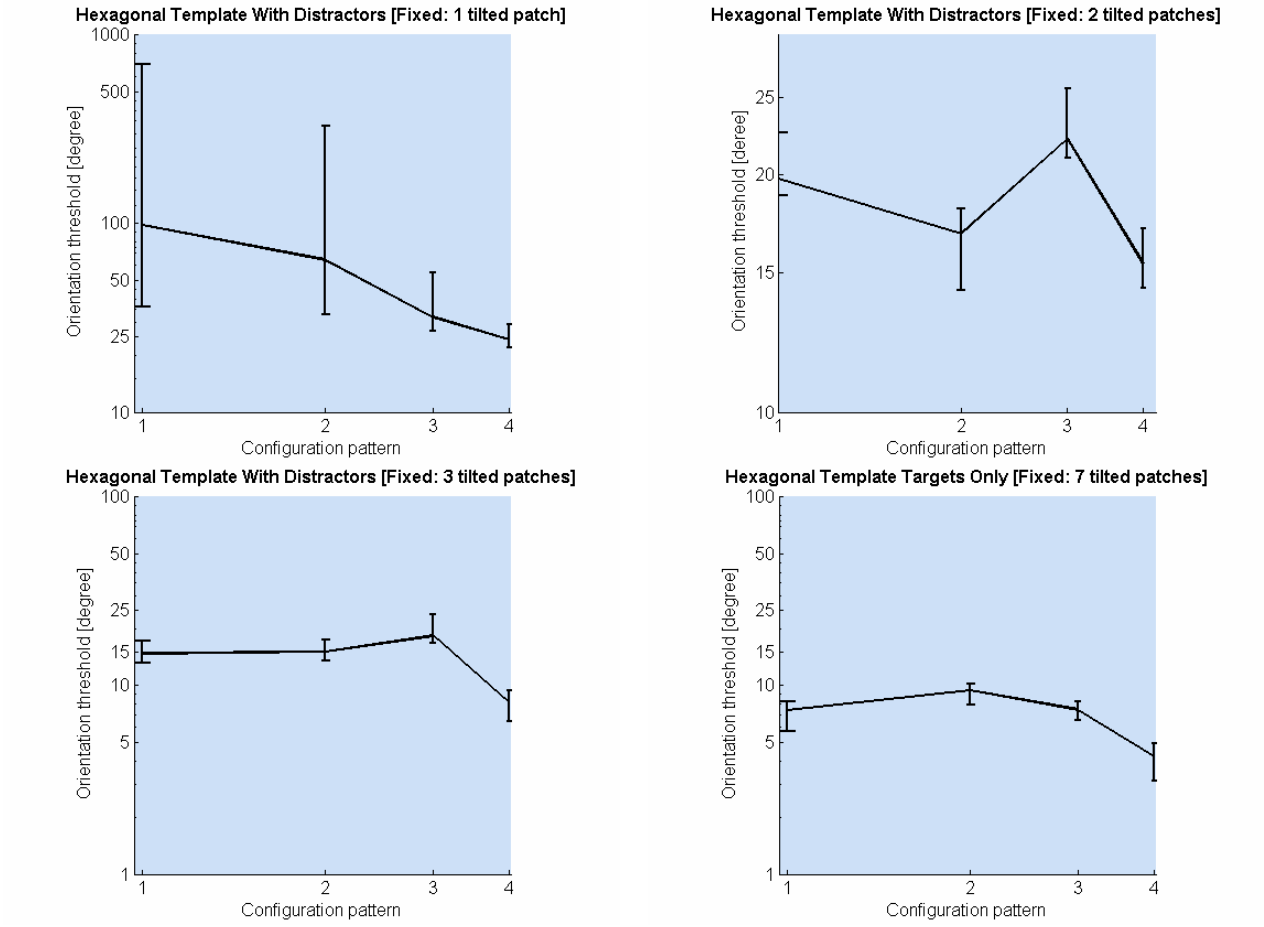

Figure 15: The effect of the distance and the scaling on each configuration with a fixed number of targets [Subject CC].

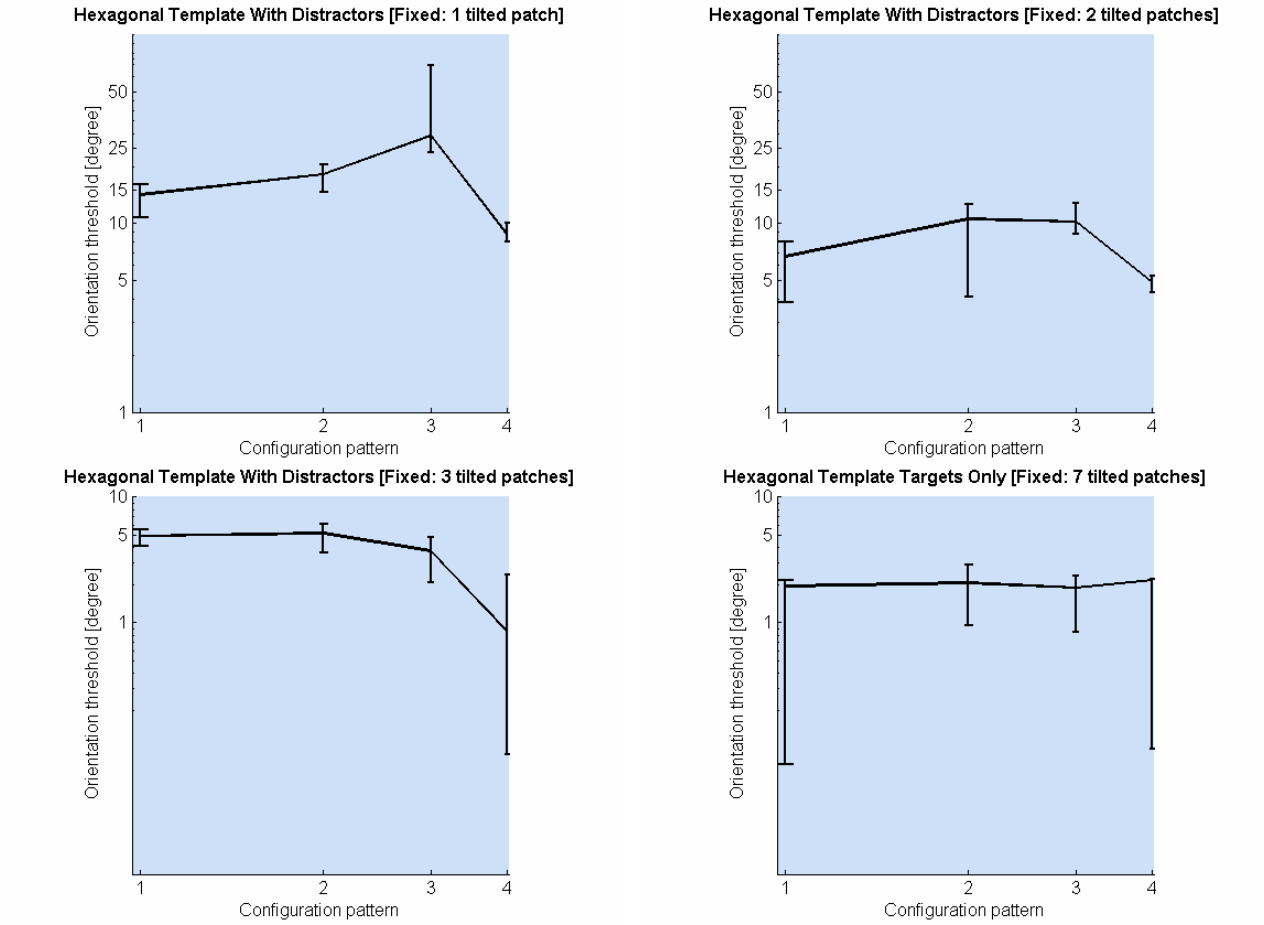

Figure 16: The effect of the distance and the scaling on each configuration with a fixed number of targets [Subject MK].

Conclusion

Our results show that increasing the distance up to 1.6 times compared to Parkes between the central patch and all the surrounding patches has no effect on the variation of the tilt threshold. Thresholds for our two subjects fell as the number of targets rose, and the data are well fitted on log-log axes by a line with a slope of -1. These results show that the distance between the central patch and all the surrounding patches does not affect Parkes *et al.* pooling model. In other words, crowding is still present when using these distances, and pooling still takes place.

Furthermore, increasing the size of all the patches yield the same result. Thresholds still fell as the number of targets rose, and the data are also well fitted on log-log axes by a line with a slope of -1. These results show that the size of all the patches does not affect Parkes *et al.* pooling model, even if it is 1.4 times larger than the original. The only consequence of increasing the overall size of the stimuli is that it reduces the tilt thresholds.

In addition, when each stimulus contains only one or two tilted patches, thresholds tend to rise when the distance is increased. Our results show that the two subjects were having more difficulty identifying a global orientation when the distance was increased. Here the density of the patches seems to help the observer identify the global orientation when only a few patches are tilted.

Finally, the discrimination capacity of the tilted patches is highly dependent of interindividual differences. The two subjects from this study don't show the same performance when it comes to orientation identification under crowding conditions.

Future studies could focus on other factors that may influence crowding, such as target shape, spatial frequency, or color. An important issue is why does crowding happens. Is it a simple unavoidable phenomenon linked to the limited capacity of the human visual system, or is it something that could help performing some visual tasks?

Acknowledgement

We would like to thank our supervisor Roberto Costantini for his notes on crowding experiments, and all his good pieces of advice.

References

- Cavanagh P. (2001). Seeing the forest but not the trees (News and Views). *Nature Neuroscience*, 4, 673-674
- L. Parkes, J. Lund, A. Angelucci, J. A. Solomon, and M. Morgan, "Compulsory averaging of crowded orientation signals in human vision", Nat. Neurosci., 4, 739744, 2001
- F. A. Wichmann, "The psychometric function I: fitting, sampling and goodness-of-fit", Percept. Psychophys., 63, 12931313, 2001
- F. A. Wichmann, "The psychometric function II: bootstrap-based confidence intervals and sampling", Percept. Psychophys., 63, 13141329, 2001
- M. P. Simunovic, R. Calver, "Crowding under scotopic conditions", *Vision Research*, 963969, 2003
- Tripathy SP, Cavanagh P. Extent of crowding in peripheral vision does not scale with target size. *Vision Research*, 42:2357-2369, 2002
- Dan Ariely, (2001), "Seeing sets: Representation by statistical properties," *Psychological Science*, 12 (2), 157- 162
- Eckstein, M.P., The lower efficiency for conjunctions is due to noise and not limited capacity visual attention, (1997) Psychological Science