Article:

Vilidaite, Greta and Baker, Daniel Hart orcid.org/0000-0002-0161-443X (2017) Individual differences in internal noise are consistent across two measurement techniques. Vision Research. 30 - 39. ISSN 0042-6989

https://doi.org/10.1016/j.visres.2016.10.008

Individual differences in internal noise are consistent across two measurement techniques

Greta Vilidaite & Daniel H. Baker
Department of Psychology, University of York

Abstract

Internal noise is a fundamental limiting property on visual processing. Internal noise has previously been estimated with the equivalent noise paradigm using broadband white noise masks and assuming a linear model. However, in addition to introducing noise into the detecting channel, white noise masks can suppress neural signals, and the linear model does not satisfactorily explain data from other paradigms. Here we propose estimating internal noise from a nonlinear gain control model fitted to contrast discrimination data. This method, and noise estimates from the equivalent noise paradigm, are compared to a direct psychophysical measure of noise (double-pass consistency) using a detailed dataset with seven observers. Additionally, contrast discrimination and double-pass paradigms were further examined with a refined set of conditions in 40 observers. We demonstrate that the gain control model produces more accurate double-pass consistency predictions than a linear model. We also show that the noise parameter is strongly related to consistency scores whereas the gain control parameter is not; a differentiation of which the equivalent noise paradigm is not capable. Lastly, we argue that both the contrast discrimination and the double-pass paradigms are sensitive measures of internal noise that can be used in the study of individual differences.

1 Introduction

Internal noise is intrinsic to the assumptions of signal detection theory (Green & Swets, 1974; Macmillan & Creelman, 2005) and signal degradation due to internal variability is evident in both electronic systems (e.g. amplifiers) and living organisms. Neural internal noise is inherent to sensory neurons and acts as a limiting factor in signal transduction (Faisal, Selen, & Wolpert, 2008). In psychophysics, this leads to the psychometric function taking the shape of a sigmoid rather than transitioning sharply between sub-threshold and supra-threshold stimuli (Burgess & Colborne, 1988). A substantial body of research has attempted to measure noise psychophysically for many different visual cues, including luminance (Barlow, 1956), orientation (Jones, Anderson, & Murphy, 2003), shape (Sweeny, Grabowecky, Kim, & Suzuki, 2011), motion perception (Barlow & Tripathy, 1997) and contrast (Burgess & Colborne, 1988; Lu & Dosher, 2008; Pelli, 1985).

Differences in internal noise have been reported in normal human development (Skoczenski & Norcia, 1998) and ageing (Pardhan, 2004) and in clinical conditions such as amblyopia (Levi, Klein, & Chen, 2007), macular degeneration (McAnany, Alexander, Genead, & Fishman, 2013) and autism (Dinstein et al., 2012; Milne, 2011). Furthermore, individual differences in contrast sensitivity for neurotypical adults have also been explained as being partly due to noise (Baker, 2013). In order to assess differences in internal noise levels between observers it is crucial to use a paradigm that is capable of distinguishing internal noise effects from other performance-influencing factors (such as sensitivity, suppression, uncertainty or efficiency). We now discuss several candidate psychophysical methods that might be used to achieve this aim.

1.1 *Equivalent noise*

Most commonly, the influence of internal noise on psychophysical task performance is assessed by purposefully degrading the performance of the observer by presenting external stimulus noise (such as 2D isotropic white noise; Pelli, 1985). The most widely adopted method is the equivalent noise (EN) paradigm (Legge, Kersten, & Burgess, 1987; Pelli, 1985) in which observers perform a two-alternative-forced-choice (2AFC) detection experiment with white noise masks shown in both intervals and a target stimulus added to one. Detection thresholds are obtained for several mask contrast levels, and the mask noise level at which performance begins to decline is taken as an estimate of the amount of internal noise in the system.

The EN paradigm assumes a linear amplifier model (Pelli, 1985), that defines thresholds as:

$$C_{thresh} = \frac{\sqrt{2\sigma_{ext}^2 + 2\sigma_{int}^2}}{\beta} \quad (1)$$

where C_{thresh} is the threshold target contrast level, β is a parameter reflecting efficiency (Lu & Dosher, 2008) and σ_{ext} and σ_{int} are the levels of external (stimulus) noise and internal noise respectively. The model posits a linear relationship between stimulus input and signal output, with additive internal noise. External stimulus noise introduces variability into the detecting mechanism that impairs performance at high noise contrasts (when $\sigma_{ext} > \sigma_{int}$).

However, there is abundant evidence that the relationship between stimulus contrast and visual response is not linear but rather accelerating at low contrasts and saturating at high contrasts (Baker, 2013; Boynton, Demb, Glover, & Heeger, 1999; Legge & Foley, 1980; Tsai, Wade, & Norcia, 2012). Furthermore, due to the broad frequency and orientation profile of white noise masks, non-target channels will also be activated by the mask and in turn inhibit the target channel. It has recently been demonstrated that broadband white noise has a strong suppressive effect similar to that of narrowband cross-oriented masks (Baker & Vilidaite, 2014). This suggests that impaired performance at high mask contrasts in the EN paradigm could be due to cross-channel suppression from white noise rather than within-target-channel noise (Baker & Meese, 2012).

One potential solution to this is to inject variability only to the detecting channel tuned to the target. This is possible by removing from the mask all off-channel spatial frequency and orientation information. The result is a mask that is spatially identical to the target grating, but with a randomly selected contrast – a 'zero-dimensional' (0D) noise mask (Baker & Meese, 2012). Similar approaches have been previously used in luminance (Cohn, 1976), orientation (Dakin, Bex, Cass, & Watt, 2009) and auditory tone perception (Jones, Moore, Amitay, & Shub, 2013). The contrast level of the mask is randomly sampled from a Gaussian distribution to create interval-by-interval contrast jitter. It has been shown that this type of mask produces stronger masking effects than white noise (Baker & Meese, 2012; Baker, 2013), and does not show evidence of cross-channel suppression, so it may offer a more suitable alternative to white noise masks.

However, it has been pointed out (Allard & Faubert, 2013) that zero-dimensional noise masks tend to produce near perfect efficiency, implying that estimates of internal noise using this paradigm are determined entirely by detection thresholds in the absence of a noise mask! In addition, the EN paradigm still assumes a linear model that is at odds with contemporary accounts of contrast transduction (e.g. Baldwin, Baker, & Hess, 2016). In order to take into account the nonlinearity of the human visual system, paradigms and models that have more accurate underlying assumptions must be considered.

1.2 *Pedestal masking*

One possible alternative to the equivalent noise approach is to obtain an estimate of internal noise by measuring and modelling discrimination data. This type of noise estimate has been used in auditory research where the fitted noise parameter was shown to be a good predictor of other measures of internal noise in the auditory system (Buss, Hall, & Grose, 2009; Jones et al., 2013). The same method can be implemented in visual contrast discrimination (Baker, 2013; Baldwin et al., 2016). In this paradigm, a fixed contrast pedestal stimulus is presented in both intervals of a 2AFC experiment with a target contrast increment added to one of the intervals. A staircase procedure is used to obtain discrimination thresholds at several pedestal contrast levels. The resulting function takes the shape of a dipper (Nachmias & Sansbury, 1974), with a facilitatory effect at low pedestal levels and threshold elevation from masking at higher levels of pedestal contrast. The contrast response function underlying the dipper (e.g. Boynton et al., 1999) is well described by a transducer nonlinearity (Legge & Foley, 1980; Tsai et al., 2012) adapted from the Naka-Rushton equation (Naka & Rushton, 1966):

$$resp = \frac{C^p}{Z + C^q} + \sigma_{int} \quad (2)$$

where C is the stimulus contrast, p and q are exponents that produce an accelerating response across low contrasts and a compressive response across high contrasts, Z is the saturation constant (the gain control parameter) and σ_{int} is proportional to the participant's internal noise. To simulate contrast discrimination experiments, a response (*resp*) is generated for each of the two intervals (with zero mean Gaussian noise added to each), and the interval with the larger response is selected. The influences of gain control and internal noise can be differentiated (see Figure 1): increasing the gain control parameter (Z) elevates thresholds only at low pedestal levels, whereas changing the noise parameter (σ_{int}) shifts the function vertically at all pedestal contrasts. Fitting the model to empirical contrast discrimination data will therefore provide an estimate of internal noise that is decoupled from estimates of sensitivity (or gain). However, it is currently unknown how accurate noise estimates using this method are, so it would be useful to compare it to a more direct measure.

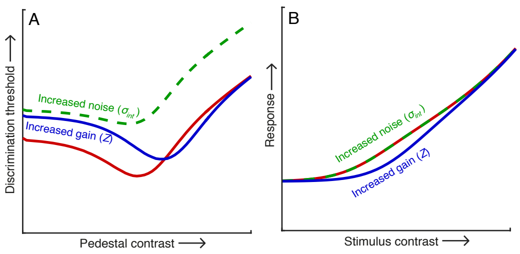

Figure 1. Panel A. Model predictions for contrast discrimination with different model parameters. The red curve shows a typical dipper function for reference (parameter values: σ_{int}=0.2, Z=8); the green curve shows the vertical shift of the whole dipper function when the noise parameter (σ_{int}) is increased by a factor of 3.5; and the blue curve shows the diagonal shift of the function when the gain control parameter (Z) is increased by a factor of 4 (at low pedestal contrasts thresholds increase, but the dipper handles converge at high contrasts). Panel B. Corresponding contrast response curves. Red and green lines here overlap showing that changes in σ_{int} do not produce a shift in the function whereas an increase in Z produces a rightward shift.

1.3 *Double-pass consistency*

When there is no variability in the stimulus, most variability in an observer's responses must be due to internal noise. One way of way estimating internal noise, therefore, is to present a sequence of noisy stimuli multiple times and look at the consistency of responses across repetitions. This method is considered to be a direct way of measuring internal noise (Burgess & Colborne, 1988; Lu & Dosher, 2008), and is typically performed with two passes (and referred to as the double pass method). Double-pass methods are well established both in auditory (Green, 1964; Jones et al., 2013) and visual modalities (Burgess & Colborne, 1988; Hasan, Joosten, & Neri, 2012; Lu & Dosher, 2008), and have also been extended to more cognitive tasks (Diependaele, Brysbaert, & Neri, 2012). To estimate double pass consistency for contrast transduction, a 2AFC detection-in-noise experiment is run twice with identical sequences of noise in the two passes. If the consistency of responses between passes is high, there is low internal noise, if the consistency is low, the internal noise is high.

1.4 *Aim*

All three of the above-mentioned paradigms are widely used in contrast perception research, with double-pass and equivalent noise specifically aimed at estimating internal noise. There has been some attempt to compare pedestal masking and EN paradigms (Baker, 2013; Baldwin et al., 2016) as well as EN and double-pass (Baker & Meese, 2012; Lu & Dosher, 2008). However, estimates of internal noise from pedestal masking and double-pass consistency experiments are yet to be compared. Given that internal noise is an important limiting factor in signal transduction and an underlying cause of individual and, in clinical research, group differences it is of importance to determine the most accurate way of measuring it. This paper compares all three methods with detailed data sets for seven observers and a further investigation of double-pass and pedestal masking paradigms with a larger sample.

2 Methods

2.1 *Observers*

Seven observers (three males) completed Study 1 and 46 observers (16 males) completed Study 2. Six of the 46 observers were excluded from the analysis as their performance was at chance for most or all of the conditions, suggesting either poor understanding of the task or an inability to follow instructions. All participants were neurotypical adults and reported normal or corrected to normal vision. Informed consent was obtained from all observers.

2.2 *Materials*

The stimuli were displayed on a gamma corrected Iiyama VisionMaster Pro 510 monitor running at 100Hz. To enable accurate rendering of low contrast stimuli, we used a ViSaGe device (Cambridge Research Systems Ltd., Kent, UK) running in 14-bit mode. Responses were made using a computer mouse.

The stimuli were patches of 0.5c/deg sine-wave grating with horizontal stripes, windowed by a circular raised cosine envelope (i.e. a circle blurred by a cosine function, with a full-width at half-height of 2.4 degrees, and blur width of 0.6 degrees, see Figure 2 for examples). The equivalent noise and double-pass experiments used zero-dimensional (0D) noise masks (see Baker & Meese, 2012). The mask was identical to the target and had a contrast level randomly drawn from a Gaussian distribution of contrasts centred around 0% (negative contrasts constitute a polarity inversion). Stimuli flickered sinusoidally between zero and their maximum contrast at a rate of 7Hz (three cycles, lasting 430ms), preserving the phase polarity of the stimulus during presentation. Contrast levels were expressed as percent Michelson contrast ($C_{\%} = 100*(\frac{L_{max}-L_{min}}{L_{max}+L_{min}})$, where L_{max} and L_{min} are the maximum and minimum luminances of the grating), or in decibels (dB), defined as $C_{dB} = 20*\log_{10}(C_{\%})$.

2.3 *Procedure*

Experiments in Study 1 were completed over several days in sessions lasting 30-60 minutes. All observers completed the experiments in the same following order: pedestal masking, noise masking and double-pass experiment. Study 2 was completed by each individual in a single 50-60 minute session. The pedestal masking took approximately 20 minutes and the double-pass experiment took 30-35 minutes to complete with short breaks in between blocks. For all experiments, the observers sat in a darkened room 105cm from the monitor with their heads supported by a chin rest. The instructions for all experiments were to 'Choose the interval in which the bar in the middle looks brighter'. The stimuli were presented foveally, along with a continuously presented central fixation cross. Each interval within a trial was presented for 430ms with an inter-stimulus interval of 400ms.

2.3.1 *Study 1 methods*

2.3.1.1 Equivalent noise experiment

Each trial contained a mask only interval and a mask + target interval (example in Figure 2b). The mask contrast was drawn from a normal distribution with a mean of 0 and standard deviations of 0, 0.5, 1, 2, 4, 8, 16 and 32% Michelson contrast. Negative contrast values reversed the polarity of the stimulus. A 3-down-1-up staircase procedure with a step size of 3dB controlled the target contrast. The staircase terminated after the lesser of 12 reversals or 70 trials and was repeated 3 times. We used Probit analysis (Finney, 1971) to fit a psychometric function to the pooled data across all repetitions, to estimate a threshold at 75% correct.

2.3.1.2 Double-pass experiment

The method of constant stimuli was used in these experiments. The stimuli had the same temporal and spatial configuration as in the equivalent noise experiments, with the mask and target + mask intervals presented in a random order on each pass (see Hasan et al., 2012). In the second pass, the samples of noise used in the first pass were repeated. Three levels of noise standard deviation were used with six target contrast levels each: i) 0% mask, target levels 0.5, 0.7, 1, 1.4, 2 and 3%; ii) 2% mask, target levels 1, 1.4, 2, 3, 4 and 5.6%; iii) 32% mask, target levels 8, 11, 16, 22, 32, 45%. Each mask standard deviation also had a target absent condition where the target contrast was set to 0% (21 conditions in total). Each condition had 200 trials (100 trials in each pass). The accuracy of responses was calculated as the proportion of correct responses out of all 200 trials in a condition; the consistency scores were calculated as the proportion of consistent responses across the two passes (Burgess & Colborne, 1988). For target absent trials nominal accuracy was calculated relative to an arbitrarily determined 'target' interval.

2.3.1.3 Pedestal masking experiment

Pairs of three-down-1-up staircases (terminating after 12 reversals or 70 trials) were used to obtain 75% correct thresholds (estimated using Probit analysis) for 9 pedestal contrasts (0.25, 0.5, 1, 2, 4, 8, 16, 32, 64%) and also in a detection condition where the pedestal contrast was set to zero. Participants completed four repetitions of each condition.

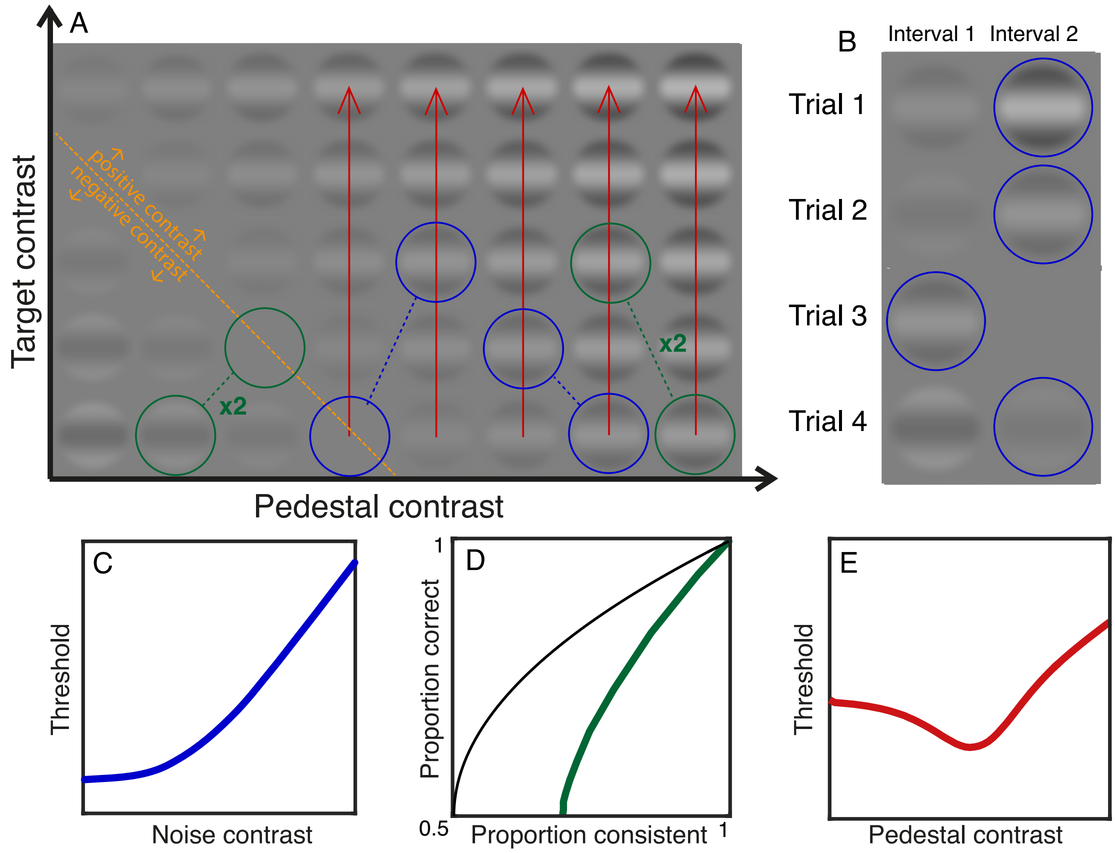

Figure 2. Illustration of methods used in the study with relation to a common contrast intensity space (panel A). For the equivalent noise paradigm (model curve shown in panel C) two independent mask contrast samples are selected for each trial with the target contrast being added to one sample. Example selections for two trials can be seen in blue circles (panel A) with blue dotted lines connecting intervals within a trial. The same procedure applies for the double-pass paradigm (model curve shown in panel D) with green circles and dotted lines showing example stimuli used. Each pair of contrasts in the double-pass experiment was presented twice. Panel B shows four more examples of trials for both EN and double-pass experiments with blue circles indicating the higher positive contrast that the observer would be expected to select. The red arrows in panel A indicate the range of possible target values for each pedestal contrast in the pedestal masking experiment. The orange dotted line indicates 0% contrast below which the sine-wave gratings reverse in phase polarity. Thresholds for contrast discrimination experiments follow a characteristic dipper shape (panel E).

2.3.1.4 Model fitting

Equivalent noise data were fitted with the linear amplifier model (equation 1) with two free parameters (β and σ_{int}) for each observer and for the average data across observers. The gain control model (equation 2) was also used to simulate and predict EN masking data (100,000 stimulated trials for each condition) with p and q parameters fixed at 2.4 and 2 (Legge & Foley, 1980) in order to keep the same number of free parameters as for the LAM. The two free parameters were the saturation constant (Z) and the internal noise (σ_{int}) parameters. Data from each observer and the average were fitted 50 times each with random starting values and the model that produced the lowest mean square error was chosen. This same procedure was used for fitting dipper data in Study 2. All models were fitted using a downhill simplex algorithm.

Pedestal masking data were fitted with the gain control model using a downhill simplex algorithm with the same two free parameters. The parameters obtained from modelling EN with LAM and pedestal masking with the gain control model were then used to simulate the double-pass experiment (100000 simulations with the gain control model and 1000000 simulations with LAM) and compare the predictions to the empirical data.

2.3.2 *Study 2 methods*

In Study 2, a smaller selection of the most informative conditions from the pedestal masking and double-pass experiments were run on a large number of observers in order to further compare the two methodologies. For pedestal masking, the same procedure and stimuli were used as in Study 1, albeit with pedestal contrast levels of 0, 2, 8 and 32%. Staircases for each condition were repeated 3 times. The double-pass procedure in Study 2 was kept the same, however, there were only two conditions: no target and 4% contrast target. In both conditions the noise standard deviation was 4% contrast. All observers completed the pedestal masking experiment first.

3 Results

The raw data are available online at:
https://dx.doi.org/10.6084/m9.figshare.3824250

3.1 Study 1

3.1.1 *Equivalent noise*

Results for the equivalent noise paradigm had the typical form, with thresholds increasing as a function of noise contrast, and the upper limb of the masking functions having a slope of 1 (Figure 3). The largest differences between participants can be seen at low noise levels (up to 0dB) where thresholds range between 0 and 5dB. At higher mask contrasts, all thresholds converge on the line of unity, x=y, consistent with previous reports that observer efficiency is near perfect for this task (Allard & Faubert, 2013). Best fits of the LAM (blue curves) described the data well (all RMS errors < 1.54dB), but estimates of the efficiency parameter (β) were similar across subjects (see values in each panel of Figure 3). This means that the only meaningful degree of freedom in this model was the internal noise parameter (σ_{int}), which determined both detection threshold and the inflection point on the noise masking function.

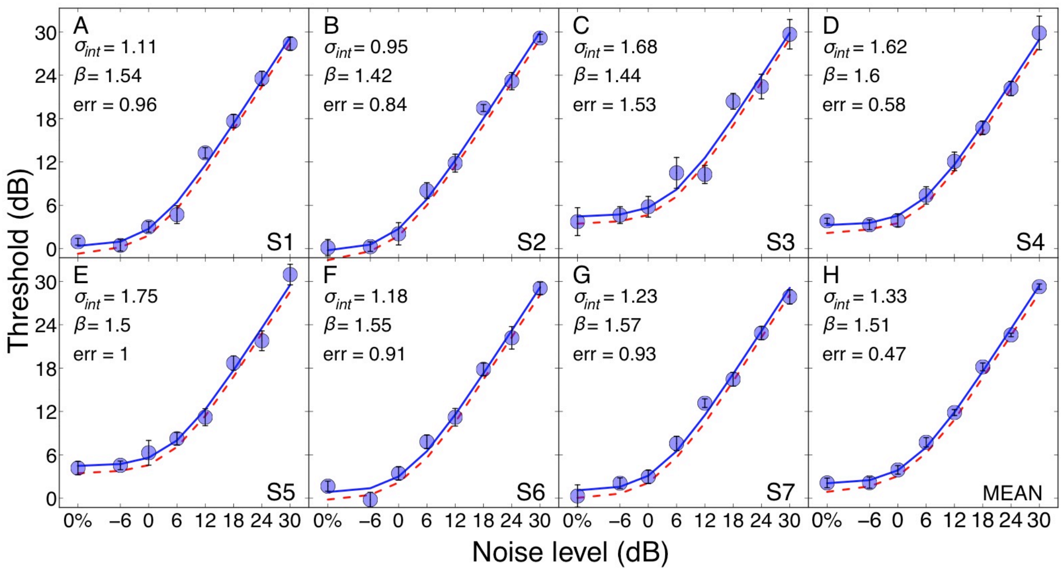

Figure 3. Noise masking thresholds from the equivalent noise experiment plotted as a function of noise contrast level. Blue dots show data points for all observers (panels S1-S7) with error bars indicating the standard error of the Probit fits. In panel H data points show the mean data averaged across observers (error bars show ±1SE across observers). Blue curves in all panels show simulated fits of the linear amplifier model and red dashed curves show simulated predictions of the gain control model. Values of parameters σ_{int} and β and the RMS error of the fit are shown in the upper left corner of each panel.

3.1.2 *Pedestal masking*

Figure 4 shows contrast discrimination data for 7 observers and their average, all of which display the characteristic 'dipper' shape first reported by Campbell and Kulikowski (1966). The gain control model with two free parameters was fitted to each observer's data individually and also to the mean data (fits to the mean data are duplicated in each panel with a red dashed curve, for comparison with the data of each observer). The model provided good fits to the data for all subjects (root mean square errors of less than 2.3dB). There is a noticeable influence of the gain control parameter (Z) on the threshold at the first four levels of pedestal contrast. For example, S4 with Z=10.57 has a much higher threshold at low pedestal conditions compared to the mean whereas S2 with a lower Z=2.78 has lower thresholds at those pedestal levels.

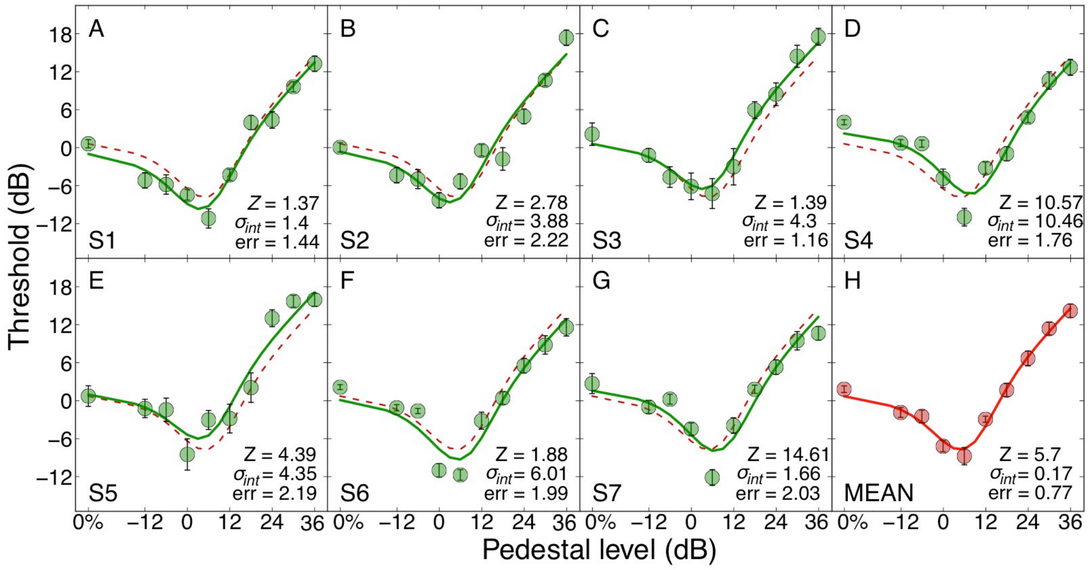

Figure 4. Thresholds at 75% correct plotted against pedestal contrast for each observer (green dots) and the mean data across observers (red dots, panel H). Error bars in panels S1-S7 show ±1SE of the Probit fit; error bars in panel H show ±1SE across observers. Green curves are the gain control model fits with two free parameters for each observer separately; red dashed curves are the model fit to the mean data and can be used as a reference for how different values of saturation constant (Z) and internal noise (σ_{int}) influence the curves. Values of both parameters used for each model are indicated in the lower right of each panel along with the RMS error in dB units.

3.1.3 *Double-pass consistency*

Accuracy and consistency scores were calculated for each noise mask and target contrast condition in the double pass experiment (Figure 5). Increasing the variance of external noise produced increasingly consistent responses, whereas increasing target contrast levels produced increasingly accurate responses. Simulated predictions for double-pass data were made using LAM fits to the EN data and gain control model fits to the pedestal masking data individually for each observer. For the majority of the observers the predictions for 0% and 2% mask contrasts were reasonably accurate from both the LAM and the gain control model. Both models produced comparatively poorer predictions for the 32% mask contrast conditions, tending to overestimate the level of consistency relative to that in the data (see also Lu & Dosher, 2008). The errors between double-pass data points and model predictions were calculated and averaged over conditions for each observer. A paired-samples t-test showed the gain control model predictions had significantly smaller errors (mean=0.11, SD=0.03) than the LAM predictions (mean=0.14, SD=0.01, t=-4.14, p=0.004).

Akaike's Information Criteria (AIC=n * log(RMS) +2p, where n is the number of data points modelled, RMS is the root mean squared error and p is the number of free parameters in the model) were calculated for these the two original models as well as for LAM with a single free parameter (β fixed at 1) and for a four free parameter gain control model (exponents p and q were also free). The gain control model with two free parameters performed best (AIC=20.17) compared to other models even when the number of free parameters is taken into account.

As it is difficult to draw population-level inferences about the consistency of noise measurements on a between observer basis with only seven observers, the conditions that seemed to show the strongest individual differences were selected for a follow up experiment with 40 observers.

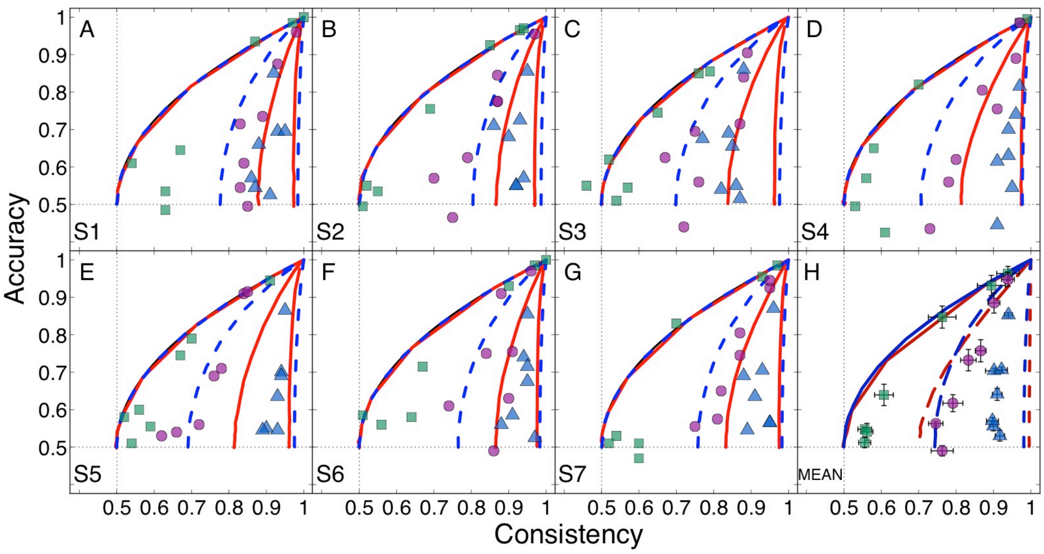

Figure 5. Double-pass consistency (x-axis) and accuracy (y-axis) for the seven observers (panels A-G) and their average (panel H) at (i) 0% (green squares), (ii) 2% (purple circles) and (iii) 32% mask standard deviation (blue triangles). Target contrast levels are not specified on the plots but generally follow an upward trend with increasing target contrast. Red curves show gain control model predictions for the three mask contrast levels for each observer and mean data; blue curves show LAM predictions. The error bars in panel H indicate ±1SE of the mean for accuracy (vertical) and consistency (horizontal) across observers.

3.2 Study 2

3.2.1 *Pedestal masking*

Using similar methods to Study 1, contrast discrimination thresholds were obtained for 40 observers (Figure 6a) and the same modelling procedure was implemented as described above. Thresholds varied between observers by 12dB (a factor of 4) or more at all pedestal levels. Pearson's correlations were carried out between the Z and σ_{int} parameters obtained from the gain control model fits and the thresholds at each pedestal level of the dipper function in order to examine the influence of these parameters at different pedestal contrasts. Scatterplots for these correlations are shown in Figure 7, however, most importantly, the Z parameter significantly correlated with individual thresholds at detection (no pedestal condition; R=0.60, p<0.0001) and at low pedestal contrast (R=0.56, p=0.0002) but did not significantly correlate at higher pedestal contrasts of 18 and 30dB (R=-0.13, p=0.426 and R=-0.17, p=0.283 respectively). This is in line with the prediction (see Figure 1) that changes in gain produce changes in threshold only at low pedestal contrasts. Conversely, the internal noise parameter σ_{int} significantly correlated with thresholds throughout the dipper function ($0.69 \leq R \geq 0.87$, p<0.0001) demonstrating that changes in the internal noise parameter shift the whole dipper function vertically in proportion to the magnitude of internal noise.

3.2.2 *Double-pass consistency*

Double-pass consistency and accuracy scores for the target and no target conditions were calculated in the same manner as in Study 1 with data from all individual observers and their mean plotted in Figure 6b. For comparison with other variables, we averaged the consistency scores across the two target contrast conditions, with high levels of consistency implying low levels of internal noise. This measure was then correlated with the four pedestal masking thresholds and Z and σ_{int} parameters from the fits shown in Figure 6a. The double-pass consistency and the fitted internal noise parameter (σ_{int}) showed a significant strong negative correlation (R=-0.68, p<0.0001) indicating consistency between these two methods of estimating internal noise. On the other hand, double-pass consistency did not significantly correlate with the gain control parameter Z (R=-0.14, p=0.378) indicating that contrast gain control estimated from pedestal masking data is not a measure of internal noise, and does not confound double pass consistency estimates.

Pearson's correlations showed that double-pass consistency was negatively correlated with dipper thresholds at all pedestal contrasts ($-0.65 \leq R \geq -0.44$, p<0.005), see Figure 9. This reiterates the point that internal noise has an influence across the entire contrast discrimination function.

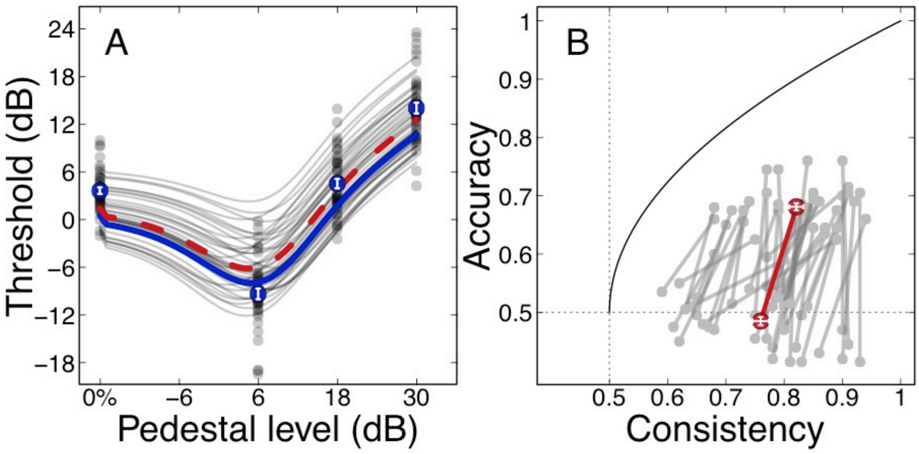

Figure 6. Panel A shows contrast discrimination thresholds as a function of pedestal contrast. Grey dots show data points for each of the pedestal levels for all 40 observers and grey lines show the gain control model fits to each observer's data. Blue dots show the mean of 40 observers with white error bars signifying inter-observer standard error of the mean. Thicker curves show the model fit for the 40 observers (blue) and model fit for 7 observers from Study 1 (red dashes). Panel B shows accuracy and consistency scores from the double-pass experiment of Study 2 for all 40 observers (grey dots and lines) and mean scores (red), with white error bars showing inter-observer standard error of the mean. Dotted lines show chance performance levels and the black curve shows the expected performance with no external noise (Klein & Levi, 2009).

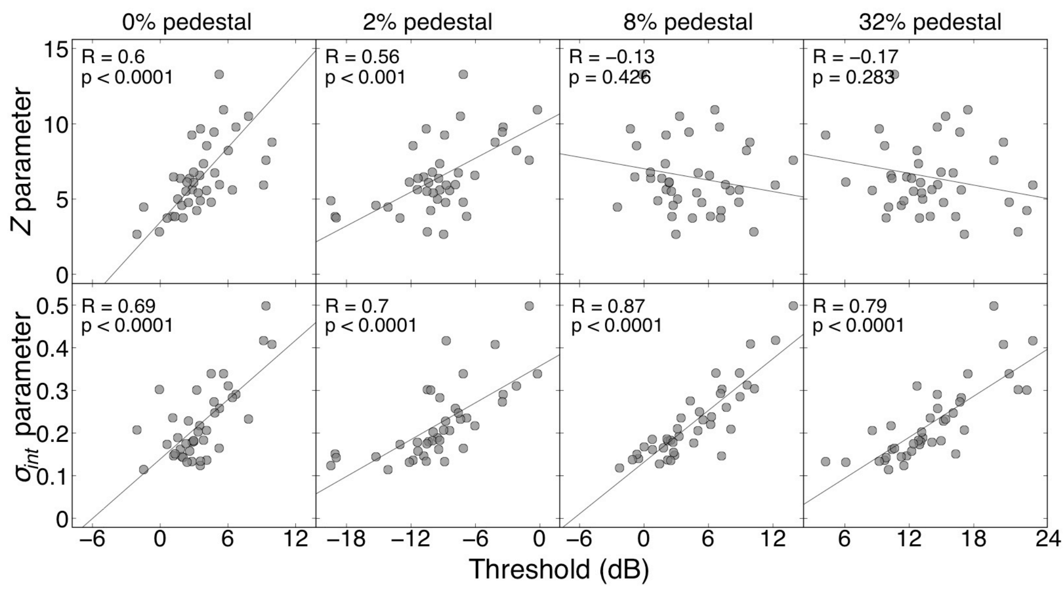

Figure 7. Correlations between the gain control parameter *Z* and pedestal masking (dipper) thresholds (top row) and correlations between the internal noise parameter σ_{int} and thresholds (bottom row). Black lines represent Deming regression lines. R and *p* values from the Pearson's correlations are shown in the upper left hand corner of each plot.

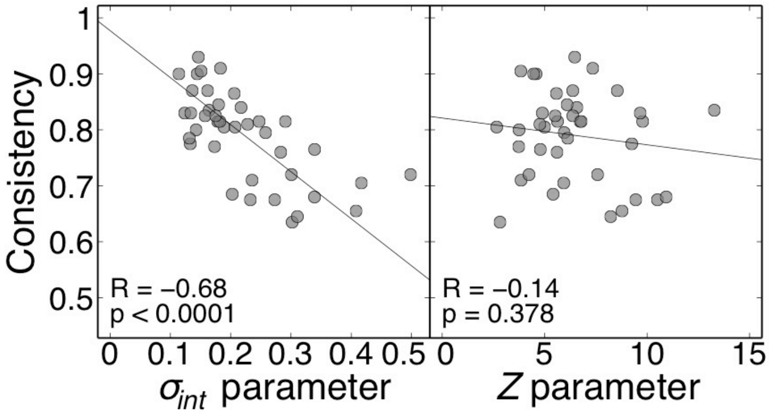

Figure 8. Scatterplots showing correlations between fitted parameters σ_{int} (left panel) and Z (right panel) and double-pass consistency scores averaged over the no target and target present conditions. Black lines represent Deming regression lines. R and p values from the Pearson's correlations are shown in the lower left hand corners of the scatterplots.

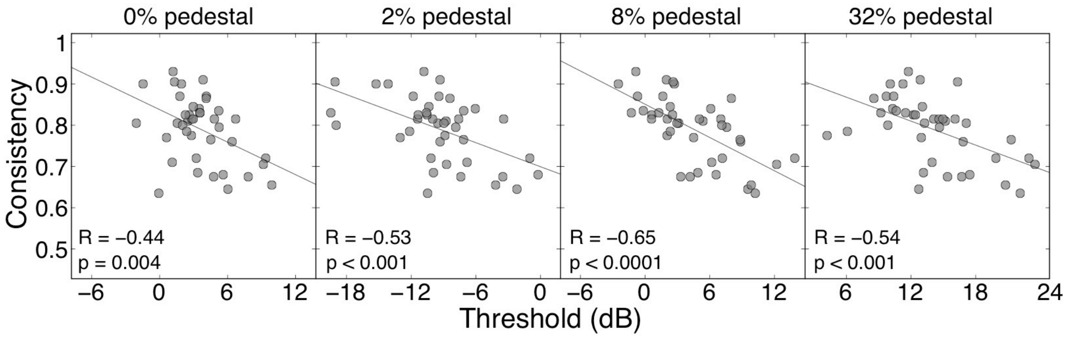

Figure 9. Scatterplots showing correlations between double-pass consistency scores averaged over the no target and target present conditions and pedestal masking thresholds at pedestal contrasts of 0, 2, 8 and 32% (from left to right). Black lines represent Deming regression lines. R and p values from the Pearson's correlations are shown in the lower left hand corners of the scatterplots.

4 Discussion

We compared three different techniques for estimating internal noise. In our first study, we showed that a nonlinear model fitted to contrast discrimination data was able to predict performance in both an equivalent noise experiment and a double pass consistency experiment. In our second study, we showed that the noise parameter from a model fitted to contrast discrimination data was strongly correlated with double pass consistency, indicating that these two paradigms measure the same internal variable. We now discuss further details of the methods, and the practicalities of running experiments to estimate internal noise.

4.1 *Comparing 2AFC discrimination with yes/no tasks*

The suggestion to use contrast discrimination paradigms as a measure of internal noise is reasonably novel (Baker, 2013; Baldwin et al., 2016), and may seem surprising to some. However, the general approach is entirely orthodox in studies that use a yes/no paradigm, where it is equivalent to measuring the slope of the yes/no psychometric function, or a just-noticeable-difference (JND). In such experiments, stimulus intensity (contrast, luminance, pitch, facial expression etc.) for a single target is typically compared to a standard (either explicit or implicit), with participants indicating whether the target appears higher ('yes') or lower ('no') in intensity than the standard. The results are plotted on a linear x-axis, with steep psychometric functions indicating low internal noise (good discriminability), and shallow functions indicating high internal noise (poor discriminability). Often a JND 'threshold' is also estimated at some criterion performance level (typically 25% and 75%). Two example simulated psychometric functions for this paradigm are shown in Figure 10a, illustrating that individuals with higher internal noise produce shallower functions with larger JNDs.

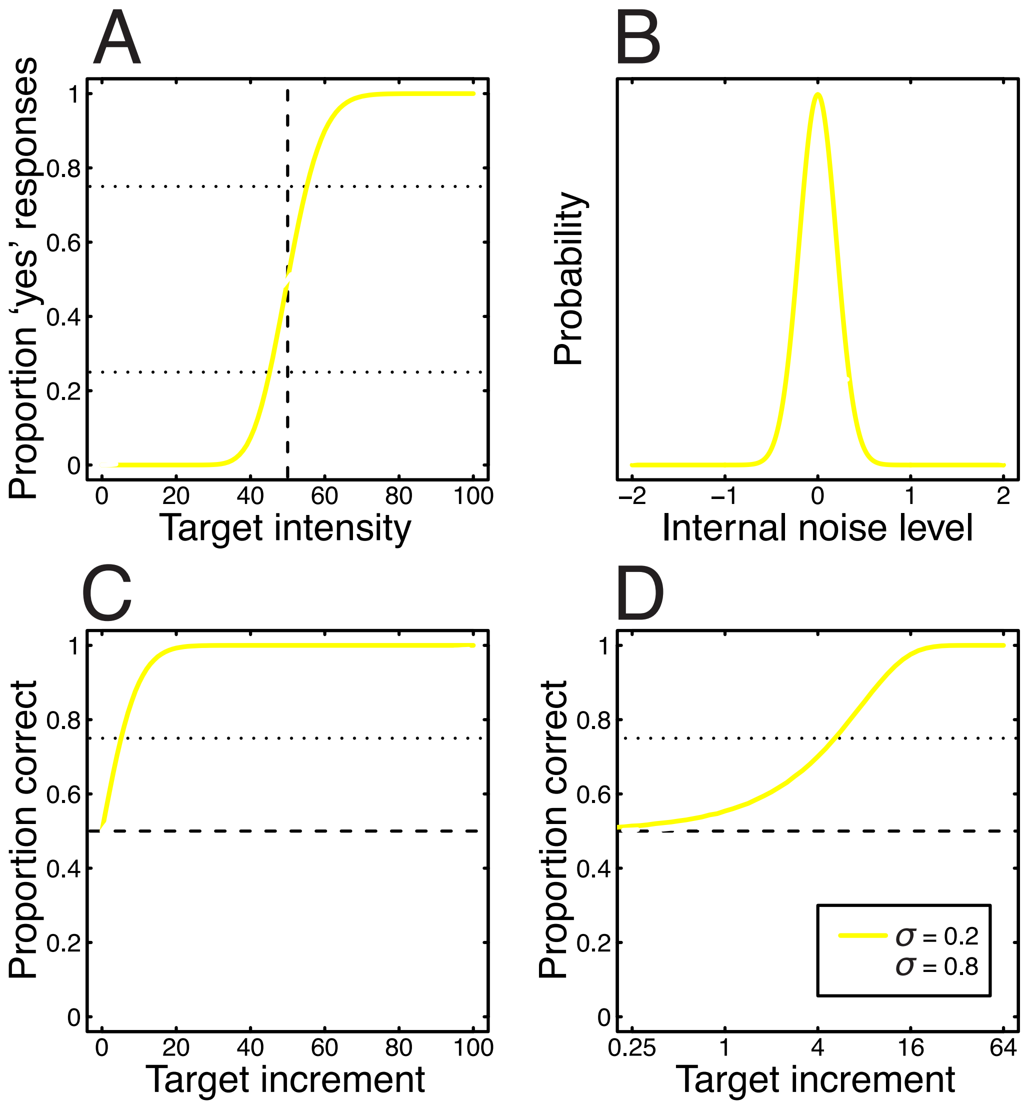

Figure 10: Illustration of the relationship between yes/no and 2AFC paradigms for intensity discrimination experiments. Panel A shows simulated yes/no psychometric functions for an intensity discrimination task in which a target was compared to a standard with an intensity of 50 units (given by the vertical dashed line). A low noise participant (blue) will have a steep psychometric function, with small just noticeable differences (JNDs) at the 25% and 75% points. A high noise participant (red) will have a shallower psychometric function and larger JNDs. Panel B shows the noise distributions for the two simulated observers. Panel C shows psychometric functions for a 2AFC discrimination task, with a pedestal level of 50 units, and a range of target increments, which are always added to the pedestal. Again, the functions are shallower for the higher noise observer when plotted on a linear x-axis, and the 75% correct threshold is higher. Panel D shows the same data replotted on a logarithmic x-axis. Now the psychometric functions are approximately parallel, and the high noise observer is differentiated only by having a higher threshold. All simulations used the gain control model given by equation 2, with parameters fixed at p=2.4, q=2, Z=1, and involved 1000000 simulated trials per target level.

In two alternative forced choice discrimination experiments, such as those described here, a pedestal is presented in one interval, and a pedestal plus target increment in the other. The pedestal level is fixed, and target stimuli constitute an increment to the pedestal contrast (though some studies have also examined decrements, i.e. (Foley & Chen, 1999). As such, effectively only the upper portion of the yes/no function is measured, as shown in Figure 10c. However, for contrast discrimination experiments the target values are conventionally plotted on a logarithmic x-axis (or alternatively converted to logarithmic units, such as the dB units used here). The log scaling of the target contrast values means that a zero point is not present, and the functions do not change in slope with changes in internal noise (see Figure 10d). Instead, only the threshold (at 75% correct) varies as noise increases. Given the close relationship between these paradigms, estimating noise levels from the dipper function is not particularly radical, and we are somewhat surprised that it has rarely been attempted.

4.2 *Why is consistency overestimated?*

We attempted to predict the double pass consistency data using models fit to either the equivalent noise thresholds or the contrast discrimination thresholds (see Figure 5). Both models overestimated the empirical double-pass consistency, especially at high mask levels. This is similar to findings from previous studies using white noise, which also found that a linear model fitted to threshold data overestimated consistency (Lu & Dosher, 2008). One possible solution is to invoke additional processes, such as induced multiplicative noise that is caused by the mask. However direct tests of this approach have not provided evidence for changes in consistency via such mechanisms (Baker & Meese, 2013). Alternatively, lower than predicted response consistency could be explained by several biases and higher level decision strategies that are relatively independent of perception, such as interval bias, finger error (lapsing), and 'superstitious' behaviours (i.e. choosing the opposite interval to the one selected on the previous trial). These low frequency events are difficult to isolate, particularly for binary decision tasks. Future work could use reports of confidence (e.g. Baker & Cass, 2013), involve an explicit mechanism for remediating trials on which an observer believes they have lapsed (Meese & Harris, 2001), or measure eye movements or other physiological variables to provide a basis for rejecting some trials. For our zero-dimensional noise, it is conceivable that observers might erroneously make judgements based on absolute contrast (ignoring phase polarity) on some trials, which would further reduce consistency estimates.

4.3 *What is the best way to measure internal noise?*

As previous studies have demonstrated, detection in white noise experiments is confounded by suppression from the mask (Baker & Meese, 2012; Baker & Vilidaite, 2014). However, using a zero-dimensional noise mask to avoid this problem results in near-perfect efficiency (Allard & Faubert, 2013), so that the inflection point of the noise masking function merely reflects detection threshold (see Figure 3). One alternative presented here and elsewhere (Baker, 2013; Baldwin et al., 2016) is to estimate internal noise using a discrimination paradigm. This is feasible for well-characterised processes such as contrast transduction, and previous findings can be reinterpreted in this context. For example, Greenaway, Davis, & Plaisted-Grant (2013) recently reported a contrast discrimination deficit in autism spectrum disorders, that could well be a consequence of increased internal noise in this population (Dinstein et al., 2012; Milne, 2011).

However, discrimination paradigms may not be suitable for more complex stimulus domains, in which the mapping between stimulus and internal representation is unknown, and perhaps nonmonotonic. In such cases, the double pass method can still be applied, as it is relatively invariant to differences in the underlying transfer function, since the addition of external noise causes 'Birdsall linearization' (Smith & Swift, 1985) that neutralises nonlinearities. For example, Baker & Meese (2013) recently showed that double pass consistency is unaffected by strong gain control suppression from a narrowband mask. The method has been successfully adapted to lexical decision tasks (Diependaele et al., 2012) and pitch discrimination (Jones et al., 2013), and we have recently run experiments using faces that vary in emotional expression, as well as value judgement tasks (Vilidaite, Yu, & Baker, 2016). Data can be obtained without prohibitively large numbers of trials (here we used 200 trials per target level), and the interpretation of results is reasonably straightforward. Additionally, double-pass shows good internal reliability with split-half analysis showing a very high correlation (R=0.88, $p<0.0001$).

4.4 *Implications for understanding individual differences*

A previous analysis of 18 studies concluded that individual differences in gain control could account for more of the variance in contrast sensitivity than could internal noise (Baker, 2013). To see if this was also the case here, we conducted a further analysis of study 2. Individual observer data from the dipper experiment (Figure 6a) were fitted as before but allowing only one free parameter, either Z or σ_{int}, fixing the other to the value obtained from modelling the average data. This procedure should reveal which of the free parameters can explain the largest proportion of the population variance. A paired samples t-test was used to compare mean RMS errors between these two fits, and revealed that RMS errors were significantly lower when σ_{int} was a free parameter (mean=2.89dB, SD=1.34dB) than when Z was a free parameter (mean=3.86dB, SD=1.75dB, $t(39)=5.52$, $p<0.001$). This suggests that internal noise had a larger influence on individual differences in contrast discrimination in this study than did gain control.

The discrepancy between studies could be due to the fixed, low spatial frequency (0.5c/deg) used here, and the variety of spatial frequencies included in the analysis by Baker (2013). This seems a plausible explanation given that differences in sensitivity caused by changes in spatial frequency are largely due to differences in gain control, and not internal noise (Baldwin, Baker & Hess, 2016). This could imply that noise accounts for a greater proportion of inter-individual variation at some spatial frequencies than others, perhaps because optical and neural factors limit sensitivity more at higher spatial frequencies. Indeed, previous work that has addressed individual differences in contrast sensitivity has revealed independently varying factors that likely relate to channels tuned to different spatiotemporal scales (Peterzell, Werner & Kaplan, 1995; Peterzell & Teller, 1996). Although it may be tempting to relate these channels to magnocellular and parvocellular systems, we note that disambiguating these psychophysically is fraught with problems (e.g. Goodbourn et al., 2012; Skottun, 2000).

In general, we take the theoretical position that internal noise is a stable and measureable property of the visual system that could, in principle, vary across individuals and clinical groups. Our aim here was to determine which experimental techniques might best be used to measure internal noise, with the intention of applying them in specific contexts (i.e. with different clinical groups). Because they are highly correlated with each other, double pass consistency and contrast discrimination appear to be suitable measures. Future work might use these tools to focus on how internal noise changes as a function of both genetic and environmental factors (e.g. ageing, diet, visual experience etc.), and how noise in one system (i.e. vision) relates to noise in other senses and tasks, or measured using different methodologies.

4.5 *Conclusions*

We compared three methods for estimating internal noise in contrast processing. Estimates from contrast discrimination and double pass consistency paradigms were highly correlated, and so are likely to be measuring the same underlying phenomenon. Depending on the dimension of interest, one or both of these methods appear to provide a good measure of internal variability, and could be used in individual differences research, or with different clinical groups.

5 References

Allard, R. & Faubert, J. (2013). Zero-dimensional noise is not suitable for characterizing processing properties of detection mechanisms. *Journal of Vision*, *13*(10):25, 1–3. http://doi.org/10.1167/13.10.25.doi

Baker, D. H. (2013). What Is the Primary Cause of Individual Differences in Contrast Sensitivity? *PLoS ONE*, *8*(7). http://doi.org/10.1371/journal.pone.0069536

Baker, D. H. & Cass, J. (2013). A dissociation of performance and awareness during binocular rivalry. *Psychological Science*, *24*(12), 2563–2568.

Baker, D. H. & Meese, T. S. (2012). Zero-dimensional noise : The best mask you never saw. *Journal of Vision*, *12*(10):20, 1–12. http://doi.org/10.1167/12.10.20

Baker, D. H. & Meese, T. S. (2013). Regarding the benefit of zero-dimensional noise. *Journal of Vision*, *13*(10):26, 1–6. http://doi.org/10.1167/13.10.26

Baker, D. H. & Vilidaite, G. (2014). Broadband noise masks suppress neural responses to narrowband stimuli. *Frontiers in Psychology*, *5*, 1–9. http://doi.org/10.3389/fpsyg.2014.00763

Baldwin, A. S., Baker, D. H. & Hess, R. F. (2016). What Do Contrast Threshold Equivalent Noise Studies Actually Measure ? Noise vs . Nonlinearity in Different Masking Paradigms. *PLoS ONE, 11*(3): e0150942

Barlow, H. B. (1956). Retinal noise and absolute threshold. *Journal of the Optical Society of America*, *46*(8), 634–639.

Barlow, H., & Tripathy, S. P. (1997). Correspondence noise and signal pooling in the detection of coherent visual motion. *The Journal of Neuroscience*, *17*(20), 7954–66. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9315913

Boynton, G. M., Demb, J. B., Glover, G. H. & Heeger, D. J. (1999). Neuronal basis of contrast discrimination. *Vision Research*, *39*(2), 257–269. http://doi.org/10.1016/S0042-6989(98)00113-8

Burgess, A. E., & Colborne, B. (1988). Visual signal detection. IV. Observer inconsistency. *Journal of the Optical Society of America. 5*(4), 617–627. http://doi.org/10.1364/JOSAA.5.000617

Buss, E., Hall, J. W. & Grose, J. H. (2009). Psychometric functions for pure tone intensity discrimination: slope differences in school-aged children and adults. *The Journal of the Acoustical Society of America*, *125*(2), 1050–1058. http://doi.org/10.1121/1.3050273

Campbell, B. F. W. & Kulikowski, J. J. (1966). Orientational selectivity of the human visual system. *J Physiol*, *187*, 437–445.

Cohn, T. E. (1976). Detectability of a luminance increment : Effect of superimposed random luminance fluctuation. *Journal of Optical Society of America*, *66*(12), 1426–1428.

Dakin, S. C., Bex, P. J., Cass, J. R. & Watt, R. J. (2009). Dissociable effects of attention and crowding on orientation averaging. *Journal of Vision*, *9*(11):28, 1–16. http://doi.org/10.1167/9.11.28

Diependaele, K., Brysbaert, M. & Neri, P. (2012). How noisy is lexical decision? *Frontiers in Psychology*, *3*, 1–9. http://doi.org/10.3389/fpsyg.2012.00348

Dinstein, I., Heeger, D. J., Lorenzi, L., Minshew, N. J., Malach, R. & Behrmann, M. (2012). Unreliable evoked responses in autism. *Neuron*, *75*, 981–991, http://dx.doi.org/10.1016/j.neuron.2012.07.026

Faisal, A. A., Selen, L. P. J. & Wolpert, D. M. (2008). Noise in the nervous system. *Nature Reviews Neuroscience*, *9*, 292–303. http://doi.org/10.1038/nrn2258

Finney, D. J. (1971). *Probit Analysis* (3rd ed.). London: Cambridge University Press.

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

Goodbourn, P.T., Bosten, J.M., Hogg, R.E., Bargary, G., Lawrance-Owen, A.J. & Mollon, J.D. (2012). Do different 'magnocellular tasks' probe the same neural substrate? *Proc R Soc B*, *279*, 4263-4271.

Green, D. M. (1964). Consistency of auditory detection judgments. *Psychological Review*, *71*(5), 392–407. http://doi.org/10.1037/h0044520

Green, D. M., & Swets, J. A. (1974). *Signal detection theory and psychophysics (A reprint, with corrections of the original 1966 ed.)*. Huntington, NY: Robert E. Krieger Publishing Co.

Greenaway, R., Davis, G., & Plaisted-Grant, K. (2013). Marked selective impairment in autism on an index of magnocellular function. *Neuropsychologia*, *51*(4), 592–600. http://doi.org/10.1016/j.neuropsychologia.2013.01.005

Hasan, B. A. S., Joosten, E., & Neri, P. (2012). Estimation of internal noise using double passes : Does it matter how the second pass is delivered ? *Vision Research*, *69*, 1–9. http://doi.org/10.1016/j.visres.2012.06.014

Jones, D. G., Anderson, N. D., & Murphy, K. M. (2003). Orientation discrimination in visual noise using global and local stimuli. *Vision Research*, *43*(11), 1223–1233. http://doi.org/10.1016/S0042-6989(03)00095-6

Jones, P. R., Moore, D. R., Amitay, S., & Shub, D. E. (2013). Reduction of internal noise in auditory perceptual learning. *The Journal of the Acoustical Society of America*, *133*(2), 970–81. http://doi.org/10.1121/1.4773864

Klein, S. A., & Levi, D. M. (2009). Stochastic model for detection of signals in noise, *J Opt Soc Am A*, *26*(11), 110–126.

Legge, G. E., & Foley, J. M. (1980). Contrast masking in human vision. *Journal of the Optical Society of America*, *70*(12), 1458–1471. http://doi.org/10.1364/JOSA.70.001458

Legge, G. E., Kersten, D., & Burgess, A. E. (1987). Contrast discrimination in noise. *Journal of the Optical Society of America*, *4*(2), 391–404. http://doi.org/10.1364/JOSAA.4.000391

Levi, D. M., Klein, S. A., & Chen, I. (2007). The response of the amblyopic visual system to noise. *Vision Research*, *47*(19), 2531–2542. http://doi.org/10.1016/j.visres.2007.06.014

Lu, Z.-L., & Dosher, B. A. (2008). Characterizing observers using external noise and observer models: Assessing internal representations with external noise. *Psychological Review*, *115*(1), 44–82. http://doi.org/10.1037/0033-295X.115.1.44

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A users guide* (2nd ed.). New York: Cambridge University Press.

McAnany, J., Alexander, K. R., Genead, M. A., & Fishman, G. A. (2013). Equivalent intrinsic noise, sampling efficiency, and contrast sensitivity in patients with retinitis pigmentosa. *Investigative Ophthalmology and Visual Science*, *54*(6), 3857–3862. http://doi.org/10.1167/iovs.13-11789

Meese, T. S., & Harris, M. G. (2001). Broad direction bandwidths for complex motion mechanisms. *Vision Research*, *41*, 1901–1914.

Milne, E. (2011). Increased intra-participant variability in children with autistic spectrum disorders: Evidence from single-trial analysis of evoked EEG. *Frontiers in Psychology*, *2*, 1–12. http://doi.org/10.3389/fpsyg.2011.00051

Nachmias, J., & Sansbury, R. V. (1974). Grating contrast: discrimination may be better than detection. *Vision Research*, *14*, 1039–1042. http://doi.org/10.1111/j.1528-1167.2007.01060_5.x

Naka, K. I., & Rushton, W. A. H. (1966). S-potentials from colour units in the retina of fish (Cyprinidae). *Journal of Physiology*, *185*, 536–555.

Pardhan, S. (2004). Contrast sensitivity loss with aging: sampling efficiency and equivalent noise at different spatial frequencies. *Journal of the Optical Society of America*, *21*(2), 169–175. http://doi.org/10.1364/JOSAA.21.000169

Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America* *2*(9), 1508–1532. http://doi.org/10.1364/JOSAA.2.001508

Peterzell, D.H., Werner, J.S. & Kaplan, P.S. (1995). Individual differences in contrast sensitivity functions: longitudinal study of 4-, 6- and 8-month-old human infants. *Vision Res*, *35*, 961-979.

Peterzell, D.H. & Teller, D.Y. (1996). Individual differences in contrast sensitivity functions: the lowest spatial frequency channels. *Vision Res*, *36*, 3077-3085.

Skoczenski, A. M., & Norcia, A. M. (1998). Neural noise limitations on infant visual sensitivity. *Nature*, *391*, 697–700. http://doi.org/10.1038/35630

Skottun, B.C. (2000). The magnocellular deficit theory of dyslexia: the evidence from contrast sensitivity. *Vision Res*, *40*, 111-127.

Smith, R. A., & Swift, D. J. (1985). Spatial-frequency masking and Birdsall's theorem, *J Opt Soc Am A*, *2*(9), 1593–1599.

Sweeny, T. D., Grabowecky, M., Kim, Y. J., & Suzuki, S. (2011). Internal curvature signal and noise in low- and high-level vision. *Journal of Neurophysiology*, *105*(3), 1236–1257. http://doi.org/10.1152/jn.00061.2010

Tsai, J. J., Wade, A. R., & Norcia, A. M. (2012). Dynamics of normalization underlying masking in human visual cortex. *The Journal of Neuroscience*, *32*(8), 2783–9. http://doi.org/10.1523/JNEUROSCI.4485-11.2012

Vilidaite, G., Yu, M., & Baker, D. H. (2016). Highly correlated internal noise across three perceptual and cognitive modalities. *Journal of Vision*, 16(12): 809.