

Bayesian adaptive estimation of threshold versus contrast external noise functions: The quick TvC method [☆]

Luis Andres Lesmes ^{a,*}, Seong-Taek Jeon ^b, Zhong-Lin Lu ^{b,*}, Barbara Anne Doshier ^c

^a Vision Center Laboratory (VCL), Salk Institute for Biological Studies, La Jolla, CA 92037, USA

^b Laboratory of Brain Processes (LOBES), Departments of Psychology and Biomedical Engineering, and Neuroscience Graduate Program, University of Southern California, Los Angeles, CA 90089-1061, USA

^c Memory, Attention, and Perception (MAP) Laboratory, Department of Cognitive Sciences and Institute of Mathematical Behavioral Sciences, University of California, Irvine, CA 92697-5100, USA

Received 12 July 2005; received in revised form 31 March 2006

Abstract

External noise paradigms, measuring contrast threshold as a function of external noise contrast (the “ TvC ” function), provide a valuable tool for studying perceptual mechanisms. However, measuring TvC functions at the multiple performance criteria needed to constrain observer models has previously involved demanding data collection (often > 2000 trials). To ease this task, we developed a novel Bayesian adaptive procedure, the “quick TvC ” or “ $qTvC$ ” method, to rapidly estimate multiple TvC functions, by adapting a strategy originally developed to estimate psychometric threshold and slope [Cobo-Lewis, A. B. (1996). An adaptive method for estimating multiple parameters of a psychometric function. *Journal of Mathematical Psychology*, 40, 353–354; Kontsevich, L. L., & Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research*, 39(16), 2729–2737]. Exploiting the regularities observed in empirical TvC functions, the $qTvC$ method estimates three parameters: the optimal threshold c_0 , the critical noise level N_c , and the common slope, η , of log-parallel psychometric functions across external noise conditions. Before each trial, the $qTvC$ uses a one-step-ahead search to select the stimulus (jointly defined by signal and noise contrast) that minimizes the expected entropy of the three-dimensional posterior probability distribution, $p(N_c, c_0, \eta)$. The method’s accuracy and precision, for estimating TvC functions at three performance criteria (65%, 79%, and 92% correct), were evaluated using Monte-Carlo simulations and a psychophysical task. Simulations showed that less than 300 trials were needed to estimate TvC functions at three widely separated criteria with good accuracy (bias < 5%) and precision (mean root mean square error < 1.5 dB). Using an orientation identification task, we found excellent agreement (weighted $r^2 > .95$) between TvC estimates obtained with the $qTvC$ and the method of constant stimuli, although the $qTvC$ used only 12% of the data collection (240 vs 1920 trials). The $qTvC$ may hold considerable practical value for applying the external noise method to study mechanisms of observer state changes and special populations. We suggest that the same adaptive strategy can be applied to directly estimate other classical functions, such as the contrast sensitivity function, elliptical equi-discrimination contours, and sensory memory decay functions.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: External noise; Adaptive; Psychophysics; Bayesian estimation; Noise

1. Introduction

External noise methods measure sensory thresholds for stimuli embedded in external noise (Pelli & Farell, 1999). The classical external noise function, the threshold versus noise contrast (TvC) function (Fig. 1), describes how signal contrast thresholds change with external noise strength. Initially used to characterize fundamental properties of the

[☆] MATLAB code for the method can be downloaded from <http://lobes.usc.edu/qMethods>.

* Corresponding authors.

E-mail addresses: lu@salk.edu (L.A. Lesmes), zhonglin@usc.edu (Z.-L. Lu).

URL: <http://lobes.usc.edu> (Z.-L. Lu).

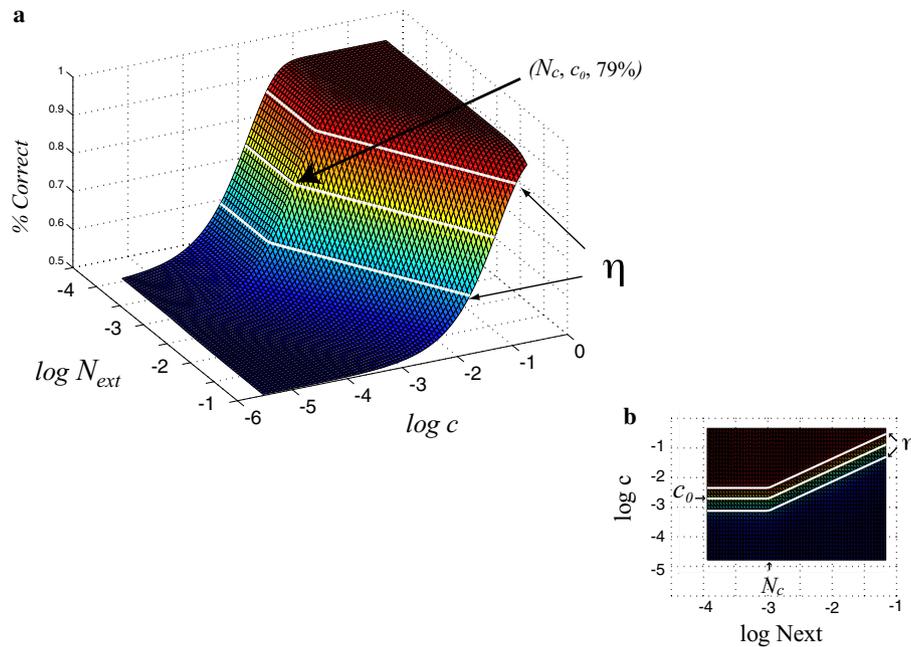


Fig. 1. In external noise studies, the family of psychometric functions measured across different external noise levels can be characterized using three parameters: (1) c_0 defines the contrast level corresponding to the 79% threshold, over the low external noise region (2) N_c defines the critical noise level at which thresholds start increasing from c_0 , and (3) η defines the psychometric function's slope, which determines the constant threshold ratio observed across external noise levels. (a) The psychometric surface (percent correct as a function of noise and signal contrast) defined by these three parameters; (b) the typical 2-D presentation of multiple TvC functions (Lu & Doshier, 1998).

perceptual system (Ahumada, 1987; Ahumada & Watson, 1985; Barlow, 1956; Burgess, Wagner, Jennings, & Barlow, 1981; Legge, Kersten, & Burgess, 1987; Legge, Rubin, Pelli, & Schleske, 1988; Nagaraja, 1964; Pelli & Farell, 1999; Pelli, 1990), external noise methods have also been used to assay mechanisms of attention (Eckstein, Pham, & Shimozaki, 2004; Lu & Doshier, 1998; Lu & Doshier, 2004b), perceptual learning (Chung, Levi, & Tjan, 2005; Doshier & Lu, 1998; Gold, Bennett, & Sekuler, 1999), object recognition (Tjan, Braje, Legge, & Kersten, 1995), and adaptation (Dao, Lu, & Doshier, 2006). Recently, the external noise method has been applied to special populations, to characterize perceptual deficits accompanying amblyopia (Huang, Tao, Zhou, & Lu, submitted; Levi & Klein, 2003; Pelli, Levi, & Chung, 2004; Xu, Lu, Qiu, & Zhou, under review), dyslexia (Sperling, Lu, Manis, & Seidenberg, 2005), prosopagnosia (Mangini & Biederman, 2002), development (Kiorpes & Movshon, 1998; Skoczenski & Norcia, 1998), and senescence (Bennett, Sekuler, & Ozin, 1999; Pardhan, 2004).

Several observer models have been developed to interpret TvC functions, including a linear amplifier model (Pelli, 1981), a multiplicative noise model (Burgess & Colborne, 1988), a multiplicative noise plus uncertainty model (Eckstein, Ahumada, & Watson, 1997), and a perceptual template model (Lu & Doshier, 1999). With only two parameters describing equivalent internal noise and efficiency, the simplest linear amplifier model is fully constrained by measuring a TvC function at a single performance criterion (e.g., 75% correct). However, because a particular parameterization of the linear amplifier model only characterizes the observer for that given performance level, addi-

tional parameters are needed to characterize the observer for other performance levels (Chung et al., 2005; Lu & Doshier, 2004a; Tjan, Chung, & Levi, 2002). To model TvC functions across the full performance range with a single set of parameters, the linear amplifier model needs elaboration, to include multiplicative noise (Burgess & Colborne, 1988) and either nonlinear transducer functions (Lu & Doshier, 1999) or channel uncertainty (Eckstein et al., 1997; Pelli, 1985). We have shown that the perceptual template model (PTM) accommodates the known standard properties of TvC functions, and provides the best qualitative and quantitative account of data across a range of paradigms (Lu & Doshier, 2002; Lu & Doshier, in preparation). To fully constrain the PTM, and thereby comprehensively account for TvC functions across the full performance range, it is necessary and sufficient to measure TvC functions at three performance criteria (Lu & Doshier, 1999).¹ For other noisy observer models, measurements of TvC functions are also needed to characterize the perceptual system over a wide performance range. Lastly, measuring TvC functions at multiple performance criteria is critical for distinguishing mechanism mixtures in the study of attention and perceptual learning (Doshier & Lu, 1999; Doshier & Lu, 2000; Lu & Doshier, 2000).

Unfortunately, measuring multiple TvC functions has so far required a large investment in data collection. For example, adequate threshold estimates at three perfor-

¹ TvC functions measured at three widely separated performance levels provide excellent proxies for full psychometric functions over all external noise conditions.

mance criteria, over eight external noise conditions—using the method of constant stimuli to measure psychometric functions at each noise level—typically demand more than 2000 trials. Because such data collection is cumbersome for all applications of the external noise paradigm—and untenable for applications in special populations—we have developed a new adaptive procedure to estimate TvC functions with good accuracy and precision, using relatively few experimental trials.

2. The $qTvC$ procedure

Our procedure, named the “quick TvC ” or “ $qTvC$ ” method, builds on recent developments in Bayesian adaptive procedures for estimating the threshold and slope of psychometric functions (Cobo-Lewis, 1996; Kontsevich & Tyler, 1999). Exploiting the regularities observed in empirical TvC functions, the $qTvC$ method uses only three parameters to characterize TvC functions over the full performance range: the optimal threshold c_0 , the critical noise level N_c , and the slope, η , of the psychometric function. The method uses responses to stimuli, jointly specified by the contrast of signal and external noise, $\vec{x} = (s, N_{\text{ext}})$, to estimate a three-dimensional probability distribution, $p(\vec{v})$, defined over a space of possible TvC functions; the posterior probability distribution $p(\vec{v})$ describes the probability that the vector $\vec{v} = (N_c, c_0, \eta)$ represents the observer’s TvC functions.

The adaptive psychophysical strategy, namely, using the subject’s previous responses to present the most “informative” stimulus on each trial (Leek, 2001; Treutwein, 1995) is formulated using the entropy of $p(\vec{v})$. By simulating the next trial’s outcome for all possible stimulus conditions, the $qTvC$ method uses a one-step-ahead search (Cobo-Lewis, 1996, 1997; Kontsevich & Tyler, 1999) to choose the stimulus \vec{x} providing the minimum expected entropy of $p(\vec{v})$. Because entropy reflects the level of uncertainty in a probability distribution, minimizing the entropy of $p(\vec{v})$ corresponds to gaining information about the underlying TvC functions.

2.1. Parameterizing TvC functions

Several properties of TvC functions (for reviews, see Lu & Doshier, in preparation; Pelli, 1990) have been reported in the classical external noise literature: (1) over low external noise levels, thresholds are nearly constant² (Nagaraja, 1964; Pelli, 1981); (2) over high external noise levels, contrast thresholds increase with noise contrast, with slope ≈ 1.0 in log–log space³ (Griffiths & Nagaraja, 1963;

Nagaraja, 1964; Stromeyer & Julesz, 1972; van Meeteren & Boogaard, 1973; as cited by Pelli, 1981), and (3) across all noise levels, the ratio between thresholds at two performance levels (e.g., 79% vs 65% correct), is constant (Chung et al., 2005; Lu & Doshier, 1999; Lu & Doshier, 2002). Collectively, these empirical regularities allow us to approximate TvC functions across the full performance range with three parameters: (1) the optimal threshold c_0 —the threshold contrast measured in low external noise levels (2) the critical noise level N_c —the external noise contrast at which thresholds start increasing from c_0 , and (3) the slope of the psychometric function, η .

The third property reported above—invariance of threshold ratios across noise contrasts—is consistent with the observation that psychometric functions measured in different noise conditions are log-parallel (Burgess, Humphrey, & Wagner, 1979; Cohn, 1976; Gold, 2001; Griffiths & Nagaraja, 1963; Lu & Doshier, 1999; Pelli, 1981; Pelli, 1985). Therefore, a psychometric function that translates on log-axes (e.g., log-Weibull or log-Gaussian; see Klein, 2001; Strasburger, 2001a, 2001b; Treutwein, 1995; Watson & Pelli, 1983), can be fit to data obtained at different noise levels with a common slope η , although thresholds, $\alpha_{N_{\text{ext}}}$, depend on external noise: $\psi_{N_{\text{ext}}} = f(\alpha_{N_{\text{ext}}}, \eta)$.

Properties 1 and 2 describe how log contrast thresholds, $\alpha_{N_{\text{ext}}}$, approximately change with external noise contrast. Over low noise levels, (below the critical noise level N_c), log thresholds can be approximated by the optimal threshold:

$$\alpha_{N_{\text{ext}}} \approx \alpha_0 = \log(c_0), \quad \text{if } N_{\text{ext}} \leq N_c. \quad (1a)$$

Property 2 reflects that, when external noise exceeds the critical noise level, N_c , the corresponding increase in psychometric threshold (in log units), $\Delta\alpha = \alpha_{N_{\text{ext}}} - \alpha_0$, is approximately equal to the log difference between the external noise contrast, N_{ext} , and the critical noise contrast, N_c :

$$\begin{aligned} \alpha_{N_{\text{ext}}} &= \alpha_0 + \Delta\alpha \\ &\approx \log(c_0) + \log(N_{\text{ext}}) - \log(N_c), \quad \text{if } N_{\text{ext}} > N_c. \end{aligned} \quad (1b)$$

Therefore, $\psi_{N_{\text{ext}}}(s)$, describing the expected percent correct as a function of signal and external noise contrast, s and N_{ext} , can be defined using the log-Weibull psychometric function⁴ as:

$$\psi_{N_{\text{ext}}}(s) = \begin{cases} \gamma + (1 - \gamma - \lambda/2) \\ \quad \times \{1 - \exp(-\exp(\eta(\log(s) - \alpha_0)))\}, & \text{if } N_{\text{ext}} \leq N_c \\ \gamma + (1 - \gamma - \lambda/2) \\ \quad \times \{1 - \exp(-\exp(\eta(\log(s) - \alpha_0) \\ \quad - \log(N_{\text{ext}}) + \log(N_c)))\}, & \text{otherwise} \end{cases} \quad (2)$$

where γ and λ are nuisance parameters determining the lower and upper asymptotes of the psychometric function

² The dip in TvC functions observed in contrast-pedestal paradigms (Legge & Foley, 1980) results from a phase-interaction between the target stimulus and the pedestal. Because such interaction does not occur consistently between the target stimulus and broadband noise, the dip is usually not observed in external noise paradigms.

³ This property, observed in nearly all paradigms using white noise (Pelli, 1990), is also predicted by all the existing noisy observer models (Lu & Doshier, in preparation).

⁴ Although the theoretical basis and implication of fitting Weibull functions to psychometric data has been recently debated (Mortensen, 2002; Tyler & Chen, 2000), the Weibull is used here as a good description of psychometric data, not for its inference of underlying noise distributions or nonlinearities.

(Treutwein & Strasburger, 1999; Wichmann & Hill, 2001a). These parameters, respectively, describing the chance performance level in the absence of a stimulus (for example, $1/m$ in *mAFC*), and the proportion of lapsed trials at the highest stimulus intensities, are assumed to be independent of external noise level.

Fig. 1 presents an example of the psychometric surface, percent correct as a function of signal and noise contrast, defined by Eq. (2). This surface, illustrating the family of psychometric functions typically measured in external noise paradigms (see Fig. 1, inset), is fully specified by the three-dimensional vector $\vec{v} = (N_c, c_0, \eta) = (10\%, 5\%, 2.5)$. Therefore, given an estimate of $\vec{v} = (N_c, c_0, \eta)$, Eq. (2) can be reversed to estimate *TvC* functions at any performance criterion. The *qTvC* method is designed to estimate these three parameters using a minimum expected entropy criterion (Cobo-Lewis, 1997; Kontsevich & Tyler, 1999) to adaptively place stimuli at signal and noise levels providing the most information about defining features of multiple *TvC* functions.

2.2. Estimating *TvC* functions using the *qTvC*

We define a parameter space $T_{\vec{v}}$, where $\vec{v} = (N_c, c_0, \eta)$ specifies the *TvC* functions defined by the particular values of N_c , c_0 , and η . The entire parameter space therefore represents all the possible *TvC* functions specified by all the possible \vec{v} 's in $T_{\vec{v}}$. A stimulus space, $X_{\vec{x}}$, is defined, containing the combinations of signal and noise contrast, $\vec{x} = (s, N_{\text{ext}})$, which will possibly be presented in the experiment. Following the observer's response to stimulus $\vec{x}_t = (s, N_{\text{ext}})$ on trial t , the *qTvC* method uses Bayes' rule to evaluate the posterior probability, $p_{t+1}(\vec{v})$, that a vector \vec{v} in parameter space $T_{\vec{v}}$ represents the *TvC* functions underlying the observer. Before running trial $t + 1$, the *qTvC* calculates $p_{t+1}(\vec{v})$ for all possible stimulus conditions (i.e., combinations of signal and external noise contrast in the stimulus space $X_{\vec{x}}$), and calculates the expected entropy of $p_{t+1}(\vec{v})$, representing the degree of uncertainty about the particular \vec{v} best describing the observer. On trial $t + 1$, the method presents to the observer the stimulus condition providing the minimum expected entropy for $p_{t+1}(\vec{v})$ and therefore the most information about the observer's *TvC* functions.

To implement the *qTvC* method, it is useful to initialize a conditional probability lookup table,⁵ $p(\text{correct}|\vec{x}, \vec{v})$, specifying the probability of a correct response for all the possible stimulus conditions and all the possible *TvC* functions defined in the parameter space $T_{\vec{v}}$. Before the experiment, we define an a priori probability distribution $p_o(\vec{v})$, representing our knowledge of $p(\vec{v})$. Following the general sequence prescribed by Kontsevich and Tyler (1999), before trial $t + 1$ ($t = 0, \dots$), the *qTvC* method estimates:

- (1) The probability of a correct response in a given stimulus condition $\vec{x} = (s, N_{\text{ext}})$:

$$p_{t+1}(\text{correct}|\vec{x}) = \sum_{\vec{v}} p(\text{correct}|\vec{v}, \vec{x}) p_t(\vec{v}).$$

- (2) The posterior probability distributions $p_{t+1}(\vec{v})$ following a correct and an incorrect response to each possible stimulus \vec{x} in trial $t + 1$ (the Bayes' rule):

$$p_{t+1}(\vec{v}|\vec{x}, \text{correct}) = \frac{p_t(\vec{v})p(\text{correct}|\vec{x}, \vec{v})}{\sum_{\vec{v}} p_t(\vec{v})p(\text{correct}|\vec{x}, \vec{v})},$$

$$p_{t+1}(\vec{v}|\vec{x}, \text{incorrect}) = \frac{p_t(\vec{v})[1 - p(\text{correct}|\vec{x}, \vec{v})]}{\sum_{\vec{v}} p_t(\vec{v})[1 - p(\text{correct}|\vec{x}, \vec{v})]}.$$

- (3) The entropies of the estimated posterior probability $p_{t+1}(\vec{v})$ following a correct and an incorrect response to \vec{x} :

$$H_{t+1}(\vec{x}, \text{correct}) = - \sum_{\vec{v}} p_{t+1}(\vec{v}|\vec{x}, \text{correct})$$

$$\times \log(p_{t+1}(\vec{v}|\vec{x}, \text{correct})),$$

$$H_{t+1}(\vec{x}, \text{incorrect}) = - \sum_{\vec{v}} p_{t+1}(\vec{v}|\vec{x}, \text{incorrect})$$

$$\times \log(p_{t+1}(\vec{v}|\vec{x}, \text{incorrect})).$$

- (4) The expected entropy after a trial with stimulus \vec{x} :

$$E[H_{t+1}(\vec{x})] = H_{t+1}(\vec{x}, \text{correct})p_{t+1}(\text{correct}|\vec{x})$$

$$+ H_{t+1}(\vec{x}, \text{incorrect})p_{t+1}(\text{incorrect}|\vec{x}).$$

- (5) The stimulus condition providing the lowest expected entropy:

$$\vec{x}_{t+1} = \arg \min_x E[H_{t+1}(\vec{x})].$$

After the observer finishes trial $t + 1$, the posterior probability function corresponding to \vec{x}_{t+1} and the observed response (correct or incorrect) is saved and used as the prior probability function for the subsequent trial.

In the next two sections, we describe Monte-Carlo simulations and a psychophysical validation of the *qTvC* method.

3. Testing the *qTvC* Method: Simulations

3.1. Method

Monte-Carlo simulations (1000 iterations) were used to estimate the precision and accuracy of the *qTvC* method. Each simulated experiment consisted of 1000 trials. Possible signal contrasts (s) ranged from 1.5% to 100% in 1 dB steps. Possible noise contrasts (N_{ext}) ranged from 2% to 33%, sampled in 3 dB steps. We selected the range of possible \vec{v} 's (defining the parameter space $T_{\vec{v}}$) based on the results of previous external noise studies (Chung et al., 2005; Dao et al., 2006; Gold, 2001; Lu & Doshier, 1998, 1999, 2002, 2004a, 2004b, in preparation; Lu, Jeon, & Doshier, 2004;

⁵ A lookup table is computed over a discretely sampled space of possible stimuli $X_{\vec{x}}$, and a sampled space of possible *TvC* functions, $T_{\vec{v}}$ (see Section 3.2.5 for discussion of issues related to parameter and stimulus space sampling).

Pelli, 1990): N_c ranged from 2.5% to 25%; c_0 ranged from 2.5% to 33%; and η ranged from .5 to 7, with 1 dB sampling over each parameter's range. Like other Bayesian methods, the efficiency of the $qTvC$ method can be improved by using a better informed *a priori* probability distribution $p_o(\vec{v})$. The present simulations assumed that the *a priori* probability distribution $p_o(\vec{v})$ was uniform over $T_{\vec{v}}$. In certain applications, better informed priors can be used.

The simulated observer, specified by $\vec{v}_0 = (10\%, 12\%, 2.5)$, matched a typical observer in a 2AFC orientation identification experiment (Lu & Dosher, 1999). With a lapse rate of 4%, the observer's psychometric function exhibited a lower asymptote of 50% and an upper asymptote of $100 - \frac{4}{2} = 98\%$. Based on these parameters and Eqs. (1) and (2), \vec{v}_0 completely specified the observer's performance (percent correct) in all possible stimulus conditions. For each simulated trial, \vec{v}_0 and Eqs. (1) and (2) were used to compute the expected percent correct p , based on signal and noise contrast. After drawing a random number r from a uniform distribution over $[0, 1]$, the observer's response was marked correct if $r < p$, and marked incorrect otherwise.

Every 10 trials, \hat{v} was estimated from $p(\vec{v})$ using the mean of the posterior's marginal probability distributions (King-Smith & Rose, 1997; Kontsevich & Tyler, 1999). TvC functions at three performance criteria were estimated from \hat{v} . Reversing Eq. (2), the threshold estimate $\hat{\tau}_{ij}$ was calculated for each external noise condition, $N_{\text{ext}}^{(j)}$ ($j = 1, \dots, 9$), and each performance criterion, $P_i = 65\%$, 79%, and 92% ($i = 1, 2, 3$):

$$\hat{\tau}_{ij} = \exp \left(\alpha_{N_{\text{ext}}^{(j)}} + \frac{1}{\eta} \left(\log \left(-\log \left(1 - \frac{P_i}{1 - \gamma - \lambda/2} \right) \right) \right) \right), \quad (3)$$

where $\gamma = .5$, $\lambda = .04$, and $\alpha_{N_{\text{ext}}}$, describing psychometric threshold as a function of external noise contrast, is defined by Eq. (1). The influence of such factors as the assumed lapse rate, sampling range and grain of the stimulus space, $X_{\vec{x}}$, and parameter space, $T_{\vec{v}}$, and the functional form of the psychometric function is examined later, (Sections 3.2.5 and 3.2.6), with more conceptual detail and comprehensive simulations.

3.2. Results

We present the $qTvC$ method's stimulus placement strategy and its accuracy and precision for estimating (a) the parameter vector \vec{v}_0 , and (b) TvC functions ($\hat{\tau}_{ij}$ generated by \hat{v}). Estimates of TvC functions obtained with the $qTvC$ are compared with estimates obtained with the method of constant stimuli (MCS).

3.2.1. Stimulus placement

Fig. 2 illustrates the $qTvC$'s sampling of the stimulus space, $X_{\vec{x}}$. Presentation frequency, as a function of noise and signal contrast, is presented for different experimental epochs. To aid evaluation, the simulated observer's "true"

TvC functions (at 65%, 79%, and 92%) are overlaid. Over the first 50 trials, stimuli are exclusively presented at two noise levels (one low and one high), with a unimodal distribution of signal contrasts across a wide range of performance levels on both psychometric functions. For subsequent epochs of the experiment (for example, trials 51–240), the distribution of signal contrasts is bimodal: presentation alternates between two signal contrasts, roughly corresponding to performance at 70% and 90% (Kontsevich & Tyler, 1999). As the session progresses, stimuli are placed at intermediate noise levels, over which signal contrasts alternate between 90% performance (for noise levels on the elbow's left side) and 70% performance (for levels on the elbow's right side).

Two more notable properties of the $qTvC$'s stimulus placement strategy are (1) relatively infrequent sampling of the stimulus space region corresponding to the elbow of the TvC function, (N_c , c_0); and (2) relatively frequent sampling of high signal contrasts: observer performance averages 85%. In fact, over 50% of trials occur in stimulus conditions corresponding to >90% correct performance. This is consistent with the observation that large numbers of trials near the upper asymptote of the psychometric function are needed to accurately estimate its threshold and slope (Green, 1990; Kontsevich & Tyler, 1999; Wichmann & Hill, 2001a, 2001b). This feature of the $qTvC$ method bodes well for applications to special populations: such a high rate of correct response is comfortable for non-experienced psychophysical observers.

3.2.2. Accuracy

The $qTvC$'s accuracy for estimating the parameters N_c , c_0 , and η , and TvC functions, is evaluated. The bias (in dB) between the actual parameter, $\vec{v}_0^{(i)}$, and its estimate, $\hat{v}^{(i)}$, for $i = 1, 2, 3$, is calculated from $N = 1000$, ($k = 1, \dots, N$), simulation repetitions:

$$\text{bias}(\hat{v}_0^{(i)}) = \frac{\sum [20 \log_{10}(\hat{v}_k^{(i)}) - 20 \log_{10}(v_0^{(i)})]}{N}. \quad (4)$$

Fig. 3a presents bias estimates for each parameter, (N_c , c_0 , and η), as a function of completed trial number. For each $qTvC$ parameter, estimation bias decreases to less than ± 2 dB (<2.5%) after 300 trials, and thereafter continues decreasing. Using Eq. (3), the $qTvC$ estimate, \hat{v} , obtained in each of N simulations provided a threshold estimate, $\hat{\tau}_{ij}$, for each external noise condition, $N_{\text{ext}}^{(j)}$ ($j = 1, \dots, 9$), and each performance criterion, $P_i = 65\%$, 79%, and 92% ($i = 1, 2, 3$). The mean bias of TvC estimates (for $J = 9$ noise conditions), averaged over $N = 1000$ simulations, is calculated for each performance criterion:

$$\text{bias}(\hat{\tau}_i) = \frac{\sum_j [20 \log_{10}(\hat{\tau}_{ij}) - 20 \log_{10}(\tau_{ij}^t)]}{J}. \quad (5)$$

Fig. 4a illustrates the mean bias for TvC estimates as a function of the number of simulated trials. Bias of thresh-

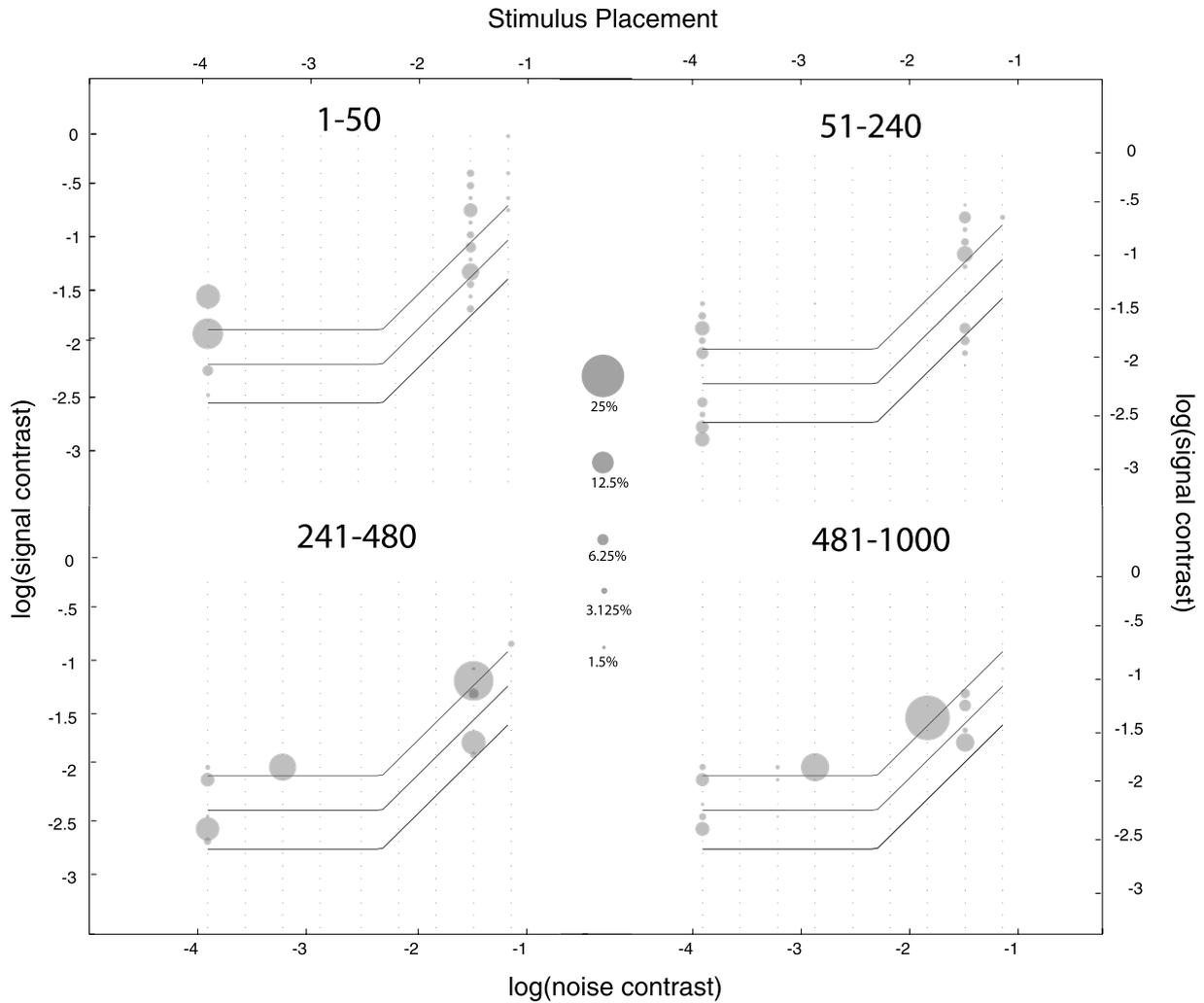


Fig. 2. The minimum entropy criterion used by the *qTvC* presents stimuli (defined by signal and external noise contrast) that maximize the information gained about the observer's *TvC* functions on each trial. The frequency of stimulus presentation, as a function of signal and noise contrast, is presented for different epochs (trials 1–50, 51–240, 241–480, and 481–1000) of a *qTvC* experiment, overlaid on a simulated observer's *TvC* functions (65%, 79%, and 92%). The inset circles provide a standard for estimating stimulus presentation frequency as a function of signal and noise contrast: the diameter of the circle presented at each point in the discretely sampled stimulus space, $X_{\vec{x}}$, is proportional to stimulus presentation frequency.

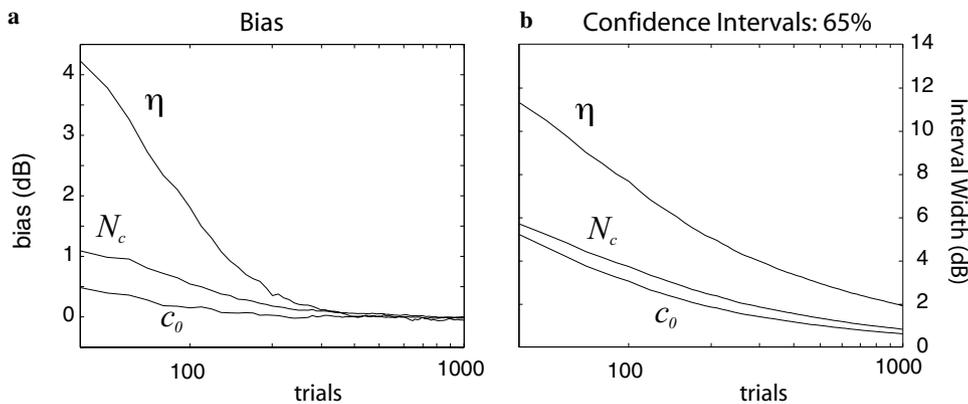


Fig. 3. Accuracy and precision for estimating the three *qTvC* parameters: N_c , c_0 , and η . The true *qTvC* parameter vector generating the simulated data was $\vec{v} = (10\%, 12\%, 2.5)$. (a) Bias for N_c , c_0 , and η estimates, as a function of completed trials. (b) The mean width of 65% confidence intervals (in dB) for N_c , c_0 , and η estimates, as a function of completed trial number. These simulations suggest that *qTvC* parameter estimates reach reasonable accuracy (bias < 1 dB) and precision (65% confidence interval width < 4 dB) in the region of 200–300 trials.

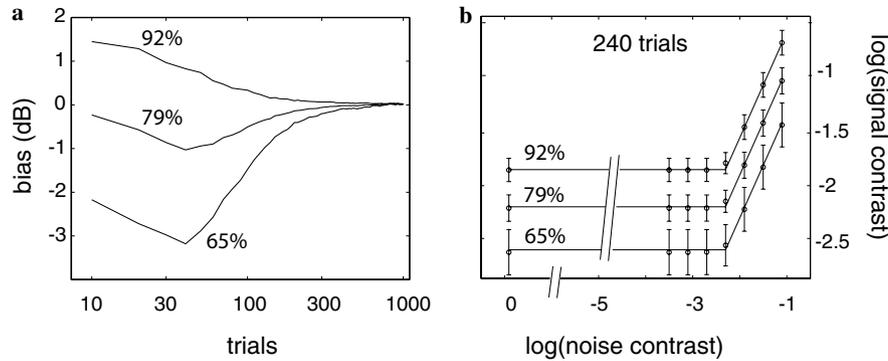


Fig. 4. Accuracy and precision of TvC estimates, at three performance criteria (65%, 79%, and 92% correct), obtained with the $qTvC$. (a) Bias estimates for TvC functions at 65%, 79%, and 92% correct are presented as a function of completed trials. Bias decreases to less than ± 0.5 ($<5\%$) dB after 200–300 trials. (b) The solid lines represent the “true” TvC functions defined by the stimulated observer, $\vec{v} = (10\%, 12\%, 2.5)$. Circles mark the mean threshold estimates obtained from the $qTvC$ with 240 trials, and error bars represent ± 1.0 SD. For the three criteria, these error bars reflect threshold variability (1 SD) between 1 and 1.5 dB.

old estimates for each performance criteria falls below ± 0.5 dB, ($<5\%$), after 240 trials. It’s apparent that the small bias in the psychometric slope estimate ($\text{bias}(\eta) \sim 2.5\%$) at 240 trials does not greatly bias estimates of TvC functions.

3.2.3. Precision

Fig. 3b presents the mean width (in dB) of the 65% confidence intervals, for estimated parameters \hat{N}_c , \hat{c}_0 , and $\hat{\eta}$ as a function of the number of completed trials. By 300 trials, the mean width of the 65% confidence intervals decreases to 2 dB for \hat{N}_c and \hat{c}_0 , and to 4 dB for $\hat{\eta}$. These interval widths correspond to standard errors of $\pm 11.5\%$ and $\pm 23\%$, which are typically acceptable stop criteria for psychophysical experiments (Kontsevich & Tyler, 1999).

The standard deviation (in dB) of each threshold estimate, $\varepsilon(\hat{\tau}_{ij})$, was calculated from $N = 1000$ ($k = 1, \dots, N$) simulations, for each external noise condition, $N_{\text{ext}}^{(j)}$ ($j = 1, \dots, 9$), and each performance criterion, $P_i = 65\%$, 79%, and 92% ($i = 1, 2, 3$):

$$\varepsilon(\hat{\tau}_{ij}) = \sqrt{\frac{\sum (20 \log_{10}(\bar{\tau}_{ij}) - 20 \log_{10}(\hat{\tau}_{ij}^k))^2}{N - 1}}. \quad (6)$$

Fig. 4b presents the average TvC estimates (over $N = 1000$ simulations) provided by 240 $qTvC$ trials overlaid on the “true” TvC functions. The error bars represent standard deviation estimates calculated with Eq. (6).

The precision for estimating multiple TvC functions is summarized by the mean error of threshold estimates, over $J = 9$ noise levels and $I = 3$ performance criteria:

$$\bar{\varepsilon}_{TvC} = \frac{\sum_{i,j} \varepsilon(\hat{\tau}_{ij})}{I \times J}. \quad (7)$$

Fig. 5a, presenting the mean error, $\bar{\varepsilon}_{TvC}$, of TvC estimates as a function of completed trial number, demonstrates that standard error falls below 1.5 dB after 200 trials and decreases to 1 dB after 300 trials. Simulations showed that this precision for estimating TvC functions agrees with that predicted by the precision of $qTvC$ parameter estimation.

Together with the previous bias analysis, these results suggest that about 300 trials are sufficient for accurate (bias $< 5\%$) and precise estimates (error < 1.5 dB) of both $qTvC$ parameters and TvC functions at three performance criteria (65%, 79%, and 92%), although fewer trials may be needed when estimating TvC functions.

3.2.4. Simulated comparison with the method of constant stimuli

We compared the precision of TvC estimates, at three performance levels, obtained with the $qTvC$ and the method of constant stimuli (MCS). The same simulated observer was tested using the method of constant stimuli, using up to 3600 trials per simulation. Constant stimuli were placed at five contrast levels: 58%, 78%, 100%, 129%, and 150% of threshold for each external noise level, based on the specification of the simulated observer, $\vec{v}_0 = (10\%, 12\%, 2.5)$; these stimulus levels were near optimal for estimating threshold and slope of psychometric functions at each noise contrast (Green, 1990; Wichmann & Hill, 2001b). The simulated results were fit with the psychometric functions defined in Eq. (2), although $\alpha_{N_{\text{ext}}}$, describing the relationship between thresholds at different noise contrasts, was defined two ways: $\alpha_{N_{\text{ext}}}$ was assumed to (1) follow the bilinear approximation, defined by N_c and c_0 , described in Eqs. (1), or (2) vary independently across external noise conditions. Following Eq. (2), both models assumed a common psychometric slope across noise conditions. Hereafter, the estimates of TvC functions provided by these models will be referred to as $TvC_{MCS_constrained}$ and TvC_{MCS} estimates.

For each MCS method, we calculate the root mean square error (RMSE) of threshold estimates, Eq. (6), as a function of completed trial number. The results are shown in Fig. 5a. Examining the relative shape and position of the error curves presented in Fig. 5a suggests that reaching a given precision level (ranging from .5 to 2 dB) with the $qTvC$ requires 25% of the trials needed by TvC_{MCS} , and 80% of the trials needed by $TvC_{MCS_constrained}$ estimates. Not surprisingly, combining the $qTvC$ parameterization

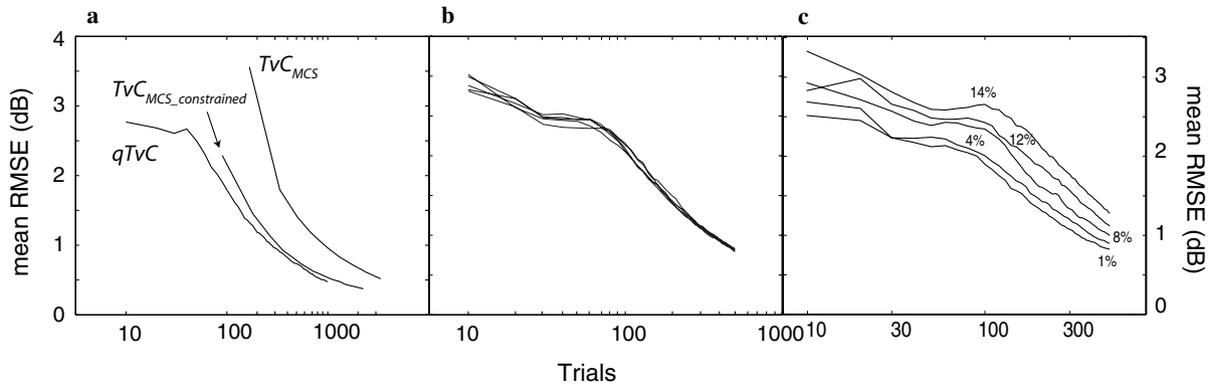


Fig. 5. Precision of TvC function estimates, when (a) compared to TvC estimates obtained with the method of constant stimuli (MCS); (b) the sampling grain of the parameter space T_v is varied, and (c) the observer's lapse rate varies (from 1 to 14%), from the $qTvC$'s assumed lapse rate (8%). For the simulations presented in (a), psychometric data was simulated using constant stimulus levels providing near optimal sampling of signal contrast at each of nine noise levels. To generate TvC estimates, the MCS psychometric data obtained at each noise level were fit with two alternative models, using maximum-likelihood: (1) independent thresholds (9 thresholds) and single slope parameter (TvC_{MCS}), or (2) the three parameter description of TvC functions (Eq. (2)) used by the $qTvC$ ($TvC_{MCS_constrained}$). For each simulated data set, threshold estimates at three performance criteria, for nine noise levels, were estimated from the best-fitting models. We calculate the root mean squared error (RMSE) for each threshold over all simulations, and present the mean RMSE over all the threshold estimates in (a). The relative position of the curves for the three procedures suggests that the $qTvC$ parameterization offers its own advantage, independent of adaptive stimulus placement. However, it should be noted that the precision of MCS estimates is probably overestimated, as they depend on optimal stimulus placement, which is difficult when attempting to measure nine psychometric functions. (b) The precision of TvC estimates is presented, for simulations varying the sampling grain (in dB) of the parameter space, T_v . For sampling schemes (in dB) ranging from fine, (.5, .5, 1.25), to coarse (1.5, 1.5, 2), the $qTvC$'s precision, as a function of completed trial number, was unchanged. (c) The precision of TvC estimates as a function of the observer's lapse rate. With the lapse rate assumed by the $qTvC$ fixed at 8%, the precision of TvC estimates depended not on the lapse rate mismatch, but on the lapse rate magnitude (lower lapse rates provided better precision).

with the near optimal stimulus placement assumed by our MCS simulations results in very efficient $TvC_{MCS_constrained}$ estimates. However, these simulations overestimate the efficiency of TvC estimation using constant stimuli; because optimal sampling of the psychometric function, dependent on a priori specification of thresholds over all external noise conditions, is usually impossible in an experimental setting, the practical efficiency of the method of constant stimuli is lower. In contrast, the only a priori knowledge the experimenter needs to apply the $qTvC$ method is readily available from the literature (or minor preliminary data collection).

3.2.5. The $qTvC$'s stimulus and parameter space: Sampling grain and range

The basic experimenter input needed by the $qTvC$ method is the specification of sampling range and grain for the stimulus space X_s and the parameter space T_v . Signal contrasts used in external noise paradigms, selected to span psychometric functions in low and high noise conditions, can vary widely across tasks (e.g., from .004 to 18% in motion discrimination and 3–99% for orientation discrimination); however, within a task, the range of signal contrasts is typically 30–35 dB. For experiments estimating thresholds that may change over multiple measurements (e.g., perceptual learning), a signal contrast range of 40 dB, with 1 dB sampling, is appropriate.⁶ Capping the

sampling range of external noise contrast at 33% (for displays with a $\pm 100\%$ contrast range) maintains a good approximation of a Gaussian distribution in the maximum external noise condition. If the minimum noise contrast used is 2%, then 3 dB sampling over this noise range (2–33%) corresponds to nine noise levels.

Similar issues apply to sampling the parameter space, T_v . Inter-observer variability suggests a 20 dB sampling range is appropriate for optimal thresholds, c_0 . There is less inter-observer and inter-task variability for critical noise, N_c : dozens of experimental data sets show critical noise levels typically fall between 5 and 10% (Chung et al., 2005; Dao et al., 2006; Gold, 2001; Lu & Doshier, 1998, 1999, 2002, 2004a, 2004b, in preparation; Lu et al., 2004; Pelli, 1990). Therefore, if 2.5–25% is used as the parameter range for critical noise N_c , the range includes many critical noise values that have not been previously observed. In Fig. 7c, the lightly shaded region marks the recommended sampling range for optimal threshold, c_0 (2.5–33%), and critical noise (2.5–25%); the darkly shaded region marks the range of normally observed critical noise values (5–10%). The range of psychometric slope (.5–7) was suggested by previous investigations (King-Smith & Rose, 1997; Kontsevich & Tyler, 1999).

Additional simulations examined how the $qTvC$'s precision is affected by the sampling grain of the parameter space T_v . The same simulated observer was tested with the $qTvC$, using five different sampling schemes s , where $s = (s_1, s_2, s_3)$ defines the respective sampling grain (in dB) for N_c , c_0 , and η : (.5, .5, 1.25), (1, .5, 1.5), (1, 1, 1.5), (1.5, 1, 2), and (1.5, 1.5, 2). Fig. 5b demonstrates the negli-

⁶ This sampling grain for stimulus intensity is standard for adaptive procedures. A 40 dB sampling range is wide enough that, given a particular task, an experimenter should have a clear sense of the appropriate range.

ble effects of sampling scheme on the precision of the estimated TvC functions: the number of trials (250–350) needed to reach TvC estimates of good precision (RMSE ~ 1 – 1.25 dB) does not depend on sampling grain. The independence of parameter sampling grain (over the range studied) makes it possible to implement the $qTvC$ method on even relatively old equipment: with a PowerMac G4 (running OS 9 and *MATLAB* with 512 MB RAM), the trial-to-trial computing needed by the $qTvC$ method takes less than 500 ms.

3.2.6. Observer-method mismatches: Lapse rate and parameter range

When applying adaptive Bayesian methods, it is important to consider the possible bias and imprecision introduced by mismatches between the true properties of the observer and those assumed by the method (Alcala-Quintana & Garcia-Perez, 2004). In the current study thus far, simulations have assumed complete agreement between the simulated observer’s veridical lapse rate (λ_{true}) and that assumed by the $qTvC$: $\lambda_{qTvC} = 4\%$. Because such consistency is rare in experimental applications, an additional set of

simulations examined how $qTvC$ estimates were affected by lapse mismatch (Alcala-Quintana & Garcia-Perez, 2004; Green, 1995; Kontsevich & Tyler, 1999; Wichmann & Hill, 2001a). The same simulated observer, $\vec{v}_0 = (10\%, 12\%, 2.5)$, was used, but her veridical lapse rate, λ_{true} , could take values of 1%, 2%, 8%, 12%, or 14%. To increase the possible range of lapse mismatch (both over- and under-estimation), the $qTvC$ ’s assumed lapse rate was increased: $\lambda_{qTvC} = 8\%$. Fig. 6c (top row) presents bias estimates for each $qTvC$ parameter as a function of trial number for each lapse rate tested. Fig. 6a summarizes the parameter bias results, presenting the asymptotic bias for each $qTvC$ parameter as a function of the observer’s veridical lapse rate, with dotted line signifying the method’s assumed lapse rate. Similarly, Fig. 6b summarizes the lapse mismatch effects for estimating TvC functions: asymptotic bias for each performance criterion (65%, 79%, and 92%) as a function of the observer’s veridical lapse rate.

As suggested by previous simulations, there is no systematic bias for $qTvC$ parameters when the observer’s lapse rate matches the method’s assumed lapse rate: $\lambda_{true} = \lambda_{qTvC}$. Under conditions of lapse mismatch, $\lambda_{true} \neq \lambda_{qTvC}$, the bias

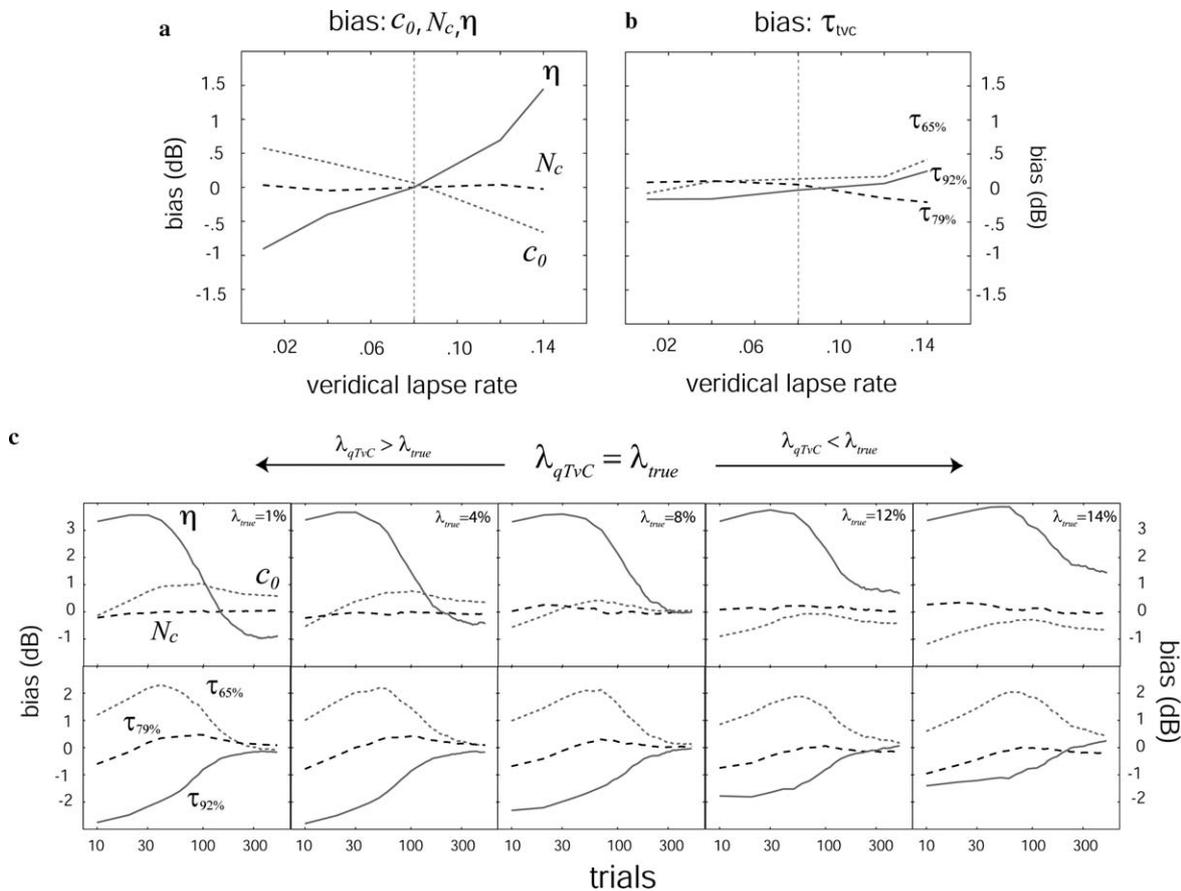


Fig. 6. Simulations examined the bias introduced by the mismatch between the observer’s lapse rate, λ_{true} , and the method’s assumed lapse rate: $\lambda_{qTvC} = 8\%$. (c) Presents the bias for $qTvC$ parameters (top row) and TvC estimates at 65%, 79%, and 92% (bottom row) as a function of completed trials, for five “veridical” lapse rates: $\lambda_{true} = 1\%$, 4%, 8%, 12%, and 14%. The asymptotic bias estimates for $qTvC$ parameters and TvC estimates are summarized in (a and b). The lapse mismatch introduces biases estimates of optimal threshold, c_0 , and psychometric slope, η , but not for critical noise, N_c . Interestingly, biased estimates of c_0 and η apparently tradeoff, so that bias is low, ($< \pm 5$ dB), for TvC estimates over a broad range of lapse rates (1–14%).

pattern for psychometric slope, η , and optimal threshold, c_0 , is clearly illustrated in Fig. 6a: when the observer's lapse rate is underestimated ($\lambda_{qTvC} < \lambda_{true}$), c_0 is likewise underestimated (bias < 0), and psychometric slope, η , is overestimated (bias > 0). The opposite pattern emerges when lapse rate is overestimated ($\lambda_{qTvC} > \lambda_{true}$): that is, c_0 is overestimated and η is underestimated. Furthermore, consistent with the conclusions of Alcalá-Quintana and García-Pérez (2004), the magnitude of bias was greater when the assumed lapse rate underestimated the observer's true lapse rate. Therefore, a conservative approach to achieving accurate estimates of $qTvC$ parameters, especially in applications with lapse-prone observers, (e.g., special populations) would be to slightly overestimate lapse rate.

When the lapse rate of a given observer \bar{v}_0 increases from 1 to 14%, the corresponding change in the psychometric function's upper asymptote can change the observer's TvC function at 92% correct by up to 2.5 dB.⁷ Therefore, for each lapse rate, the bias of threshold estimates at each performance criterion is calculated relative to the empirical (non-guessing-corrected) TvC functions generated by the lapsing observer (see Fig. 6c, bottom row). Unlike the patterns observed for $qTvC$ parameters (top row), the bias pattern in the bottom row of Fig. 6c is remarkably consistent over conditions of lapse rate mismatch ($\lambda_{true} = 1\text{--}14\%$ vs $\lambda_{qTvC} = 8\%$). Fig. 6b demonstrates that the asymptotic bias in TvC estimates (over the three criteria: 65%, 79%, and 92%) does not exceed ± 0.5 dB across the range of lapse rates studied.

Although the accuracy of TvC estimates is not affected by high lapse rates, there are effects on the precision of TvC threshold estimates (see Fig. 5c). The laminarity exhibited by precision curves in Fig. 5c clearly demonstrates that the $qTvC$'s precision improves as the observer's lapse rate decreases, even when the observer's low lapse rate mismatches the assumed lapse rate (8%).

Bias and imprecision for adaptive Bayesian estimates can also result when the assumed parameter space, $T_{\bar{v}}$, does not sufficiently encompass the observer's psychometric parameters (King-Smith & Rose, 1997). The parameter space, $T_{\bar{v}}$, specified for previous simulations, was based on empirical results reported in the literature (see Section 3.2.5). Though the range of the recommended parameter space was conservative, there remains the possibility of an atypical observer $v = (4\%, 3\%, 1.0)$, corresponding to the edges of the parameter space, (see Fig. 7c). The observer's odd TvC functions, and their relation to the parameter range, are apparent by inspection of Fig. 7c's inset, which displays the same parameter range relative to typical TvC functions. Simulations examined the accuracy and precision of $qTvC$ estimates when the assumed parameter range did not adequately match the TvC functions of an atypical,

aberrant observer. Although the estimates of the observer's $qTvC$ parameters and TvC functions are biased (see Figs. 7a and b), the TvC functions estimated with only 100 $qTvC$ trials (see Fig. 7c) clearly reflect the aberrance of the observer's functions. At 100 trials, the bias of $qTvC$ parameters and TvC estimates are less than 2 dB. One possible way to avoid the effects of parameter range mismatch⁸ is a stopping criterion: if the marginal probabilities at the borders of $p(N_c, c_0, \eta)$ exceed 15%, $qTvC$ sessions can be interrupted and rerun (with a properly specified parameter range). This problem should recede as computing power increases: it should become possible to specify more expansive parameter spaces that easily encompasses nearly all TvC functions likely to be empirically observed.

A last set of simulations examined the effects of another parametric mismatch: the specific form of the psychometric function. The simulated observer's performance (TvC functions) was generated by a log-Gaussian psychometric function, which mismatched the log-Weibull psychometric function assumed by the $qTvC$. As expected from the small practical differences between these psychometric functions (Pelli, 1987), psychometric function mismatch only introduced a small systematic bias ($\sim 5\%$) for TvC estimates at the 65% percent correct criterion.

Taken together, these simulation results suggest that the most precise and accurate estimates of $qTvC$ parameters and TvC functions occur when the observer's lapse rate is both (1) low and (2) equal to the method's assumed lapse rate. Surprisingly, when the observer's lapse rate mismatches the assumed rate, the resulting biases in $qTvC$ parameters trade-off, and the accuracy of TvC functions is relatively maintained. Furthermore, even under conditions when the parameter range or functional form of the psychometric function is misinformed, the $qTvC$ delivers reasonably accurate (bias < 1 dB) and precise (error < 2 dB) estimates of TvC functions with about 300 trials. These simulations demonstrate the robustness of the $qTvC$ method under non-optimal conditions likely to be encountered in experimental and clinical settings.

4. Testing the $qTvC$ method: Psychophysical validation

We conducted a psychophysical experiment to directly compare TvC estimates obtained with the $qTvC$ procedure and the method of constant stimuli (MCS).

4.1. Method

4.1.1. Apparatus

The experiment was conducted on a Macintosh Power G4 computer running Psychtoolbox extensions (Brainard, 1997; Pelli, 1997). The stimuli were presented on a Hewlett Packard hp91 color monitor at a 120 Hz refresh rate.

⁷ The parameter vector \bar{v}_0 specifies the observer's guessing-corrected psychometric function. However, an increasing lapse rate truncates the psychometric function at the upper asymptote, thereby shifting TvC functions defined by the same \bar{v}_0 downwards.

⁸ Another possibility is to perform post hoc fitting of psychometric data with maximum-likelihood or Bayesian fitting procedures (Jeon, Lu, Lesmes, & Doshier, in preparation; Treutwein & Strasburger, 1999).

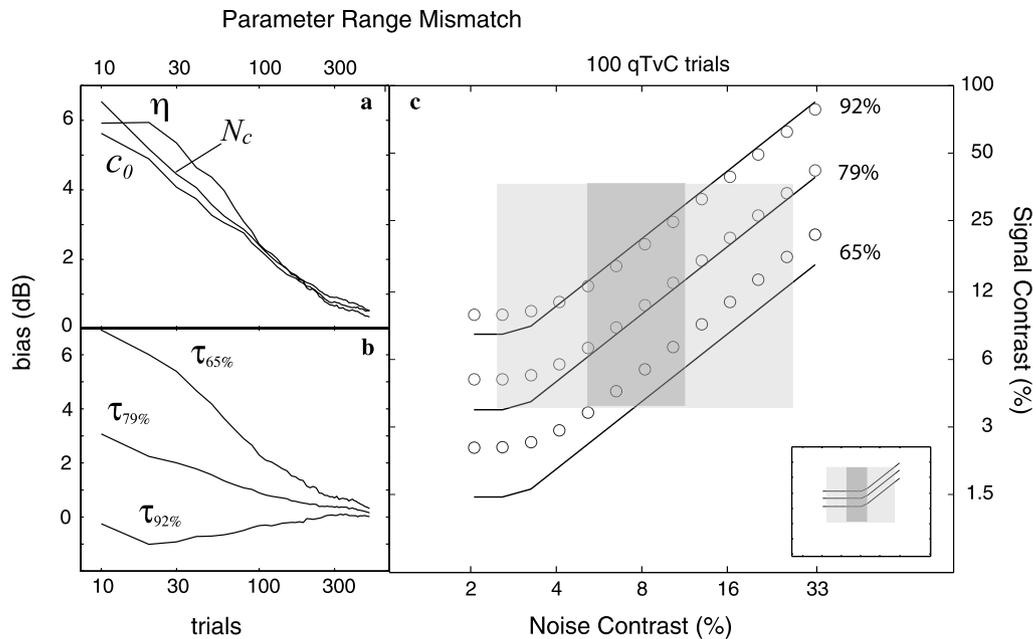


Fig. 7. The results of simulations testing the $qTvC$'s accuracy and precision when the method's assumed parameter range does not adequately contain the observer's TvC functions. The above figure presents the bias of (a) $qTvC$ parameters and (b) TvC threshold estimates as a function of trial number for an aberrant observer. (c) The observer's true TvC functions are overlaid and the inset demonstrates the coverage of normal TvC functions provided by the assumed parameter range for optimal threshold, c_0 , and critical noise, N_c . TvC estimates, obtained with 100 $qTvC$ trials, are represented by circles. Even though 100 trials provide $qTvC$ parameter and TvC threshold estimates with bias ~ 2 dB, the method does provide a good sense of the observer's aberrant functions. One way to avoid such mismatch may be to stop the experiment if some probability level is exceeded by the parameters at the edge of the parameter space. If the marginal probabilities observed at the borders of the parameter range are too large (e.g., $>15\%$) the $qTvC$ session should be re-run.

A special circuit, (Li, Lu, Xu, Jin, & Zhou, 2003), changed the display to a monochromatic mode, with high grayscale resolution (>12.5 bits). A lookup table was used to linearize the luminance levels. Stimuli were viewed binocularly with natural pupil at a viewing distance of approximately 72 cm in dim light. Observers used a chinrest to maintain head position and aid fixation throughout the experiment.

4.1.2. Participants

Two naïve observers (C.C. and W.C.) and the second author (S.J.) participated in the experiment. All observers had corrected-to-normal vision and were experienced in psychophysical studies.

4.1.3. Stimuli

The signal stimuli were Gaussian-windowed sinusoidal gratings, oriented $\theta = \pm 45$ degrees from vertical. The luminance profile of the Gabor stimulus is described by:

$$L(x, y) = L_0 \left\{ 1.0 + c \sin[2\pi f(x \cos \theta + y \sin \theta)] \exp \left[-\frac{(x^2 + y^2)}{2\sigma^2} \right] \right\} \quad (8)$$

where c is the signal contrast, $\sigma = 0.57^\circ$ is the standard deviation of the Gaussian window, and the background luminance L_0 was set in the middle of the dynamic range of the display ($L_{\min} = 1 \text{ cd/m}^2$; $L_{\max} = 55 \text{ cd/m}^2$).

The signal stimuli were rendered on a 64×64 pixel grid, extending 2.78×2.78 deg of visual angle. External noise images were constructed using 2 by 2 pixel elements

(0.087×0.087 deg). Each noise element's contrast level was drawn independently from a Gaussian distribution with mean of 0 and standard deviation ranging from .01 to .33. Because the maximum achievable contrast is $\pm 100\%$, a noise sample with standard deviation of 0.33 conforms reasonably well to a Gaussian distribution. In a given trial, external noise images were made of elements with jointly independent, identically distributed contrasts. Eight external noise levels (0%, 3%, 4.5%, 6.7%, 10%, 15%, 22%, and 33%) were used in the experiment.

4.1.4. Design

TvC functions were obtained with the $qTvC$ method and the method of constant stimuli (MCS). The MCS psychometric functions (for each of eight external noise levels) were sampled at five signal contrast levels, specified for each observer based on preliminary data. For the $qTvC$ method, signal contrast was sampled from a pool of 40 possible contrast levels, ranging from 2 to 100% with 1 dB step size. For the parameter space $T_{\bar{v}} : N_c$ ranged from 2.5 to 25%; c_0 ranged from 2.5 to 33%; and η ranged from .5 to 7, with 1.5 dB sampling over each parameter's range. For two observers (W.C. and C.C.), the lapse rate was assumed to be 1%. For the other observer (S.J.), the lapse rate was assumed to be higher: 8%.

Each observer completed four sessions of 960 trials. In each session, observers ran 480 $qTvC$ trials and 480 constant stimulus trials, randomly interleaved. An experimental session lasted about 1 h.

4.1.5. Procedure

In the beginning of each trial, a fixation-cross was presented in the center of the screen for 500 ms. The subsequent stimulus sequence consisted of three 8.3 ms frames: a noise frame, a signal frame, and another (independent) noise frame. Observers were instructed to identify the orientation of the *Gabor* stimulus by pressing different keys on the computer keyboard. A beep immediately followed each incorrect response. The intertrial interval (3 s), held constant across the *qTvC* and constant stimulus conditions, included the time (~ 1 s) needed for the *qTvC* to compute the stimulus prescribed for the next trial.

4.2. Results

4.2.1. Re-checking the regularities of the *TvC* functions

To verify the assumptions underlying the *qTvC* parameterization, the data collected with the method of constant stimuli was analyzed for the *TvC* function regularities described in Eqs. (1) and (2). Data collected at each noise level were fit with log-Weibull psychometric functions:

$$\psi_{N_{\text{ext}}}(s) = .5 + (1 - .5 - \lambda/2) \times \{1 - \exp(-\exp(\eta_{N_{\text{ext}}}(\log(s) - \alpha_{N_{\text{ext}}}))\}, \quad (9)$$

with threshold and slope parameters depending on noise level ($\eta_{N_{\text{ext}}}$, $\alpha_{N_{\text{ext}}}$) and a lapse rate parameter, λ , independent of noise level.

Psychometric thresholds were classified into low noise thresholds ($\alpha_{N_{\text{ext}}}^{\text{low}}$, when $N_{\text{ext}} < 8\%$) and high noise thresholds ($\alpha_{N_{\text{ext}}}^{\text{high}}$, when $N_{\text{ext}} > 8\%$). Fig. 8a presents low noise thresholds, $\alpha_{N_{\text{ext}}}^{\text{low}}$, measured for each observer, with error bars signifying standard error estimates (1 *SD*) obtained with a resampling procedure, and each observer's mean threshold (across the four noise levels) signified by a dotted line. A nested χ^2 test, based on the likelihood ratio, compared the fit of two models. The full model posited independent psychometric threshold and slopes for each low external noise level. The alternative (reduced) model posited that a single psychometric function (with a single set of threshold and slope parameters) fits psychometric data over the low noise levels. The appropriate χ^2 statistic is defined as:

$$\chi^2(df) = 2 \log \left(\frac{\max \text{likelihood}_{\text{full}}}{\max \text{likelihood}_{\text{reduced}}} \right), \quad (10)$$

where $df = k_{\text{full}} - k_{\text{reduced}}$, the difference of parameter number between the two models. A nested χ^2 test suggested that for all three observers, the model positing a single psycho-

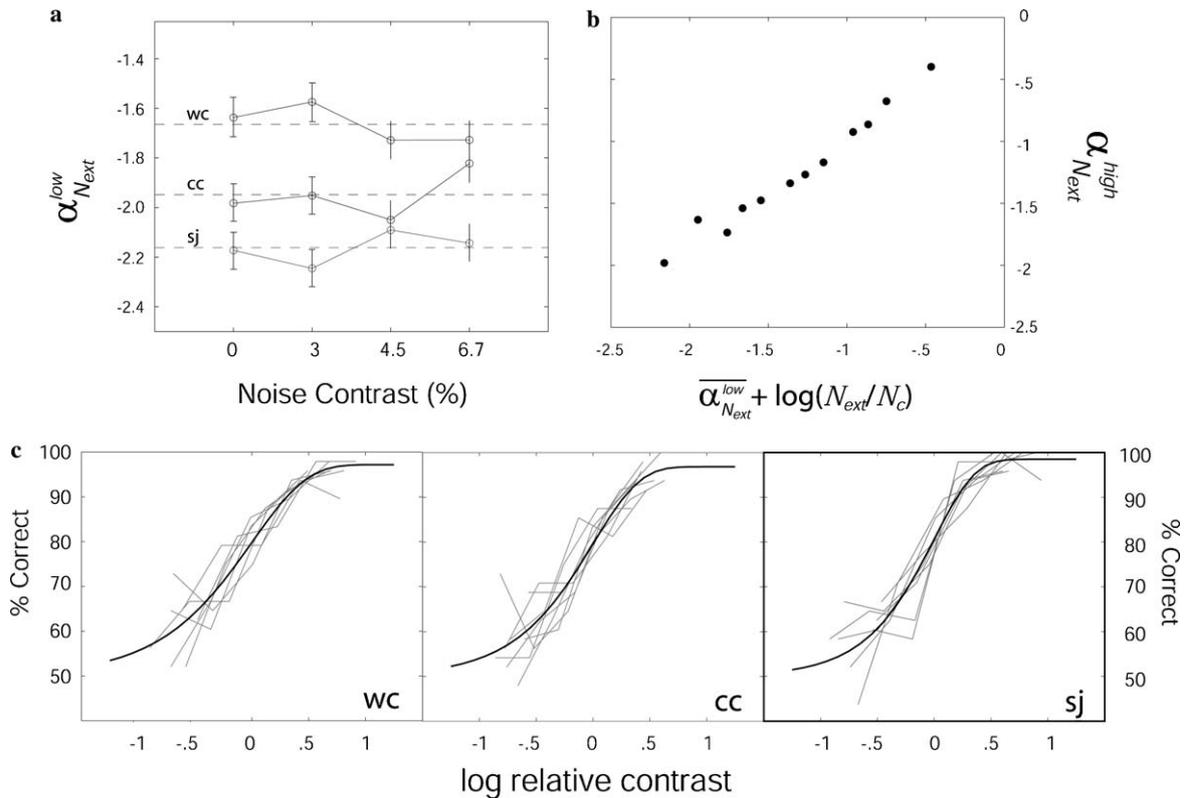


Fig. 8. To justify the *qTvC*'s assumptions concerning external noise functions, the psychophysical data collected with the *MCS* is analyzed for the *TvC* regularities. (a) The thresholds measured over low noise levels, $\alpha_{N_{\text{ext}}}^{\text{low}}$, are presented for the three subjects. A χ^2 test supports what is apparent by eye, namely that psychometric thresholds over low noise levels are constant (Property 1). (b) The thresholds measured over high noise levels, $\alpha_{N_{\text{ext}}}^{\text{high}}$, are presented as a function of low noise threshold, $\alpha_{N_{\text{ext}}}^{\text{low}}$, plus the log ratio of external noise to the critical noise level: $\alpha_{N_{\text{ext}}}^{\text{high}}$ vs $\alpha_{N_{\text{ext}}}^{\text{low}} + \log(N_{\text{ext}}) - \log(N_c)$, where $N_c = 10\%$. The excellent correspondence ($r^2 = .97$) suggests that, at high noise levels exceeding the critical noise levels, psychometric thresholds increase proportionally (Property 2). (c) Signal contrasts (s) are normalized by each noise level's psychometric threshold. The resulting overlap of the raw psychometric functions provides evidence for slope invariance of psychometric functions measured over external noise levels (Property 3).

metric function over low noise levels provided the optimal fit ($p > .25$ for each observer). Thus, this analysis verifies the proposition that psychometric thresholds over low noise contrasts are constant (Property 1).

To verify that thresholds increase proportionally at external noise levels exceeding the critical noise level (Property 2), high noise thresholds, $\alpha_{N_{\text{ext}}}^{\text{high}}$, were analyzed. Because preliminary examination suggested that the critical noise, N_c , differed little across the three subjects ($N_c \sim 9\text{--}11\%$), the analysis was simplified by fixing the N_c estimate (10%) across the three observers. In Fig. 8b, high noise thresholds, $\alpha_{N_{\text{ext}}}^{\text{high}}$, are plotted against the mean threshold measured over low noise levels, $\alpha_{N_{\text{ext}}}^{\text{low}}$ (calculated in the previous analysis), plus the log difference between the corresponding noise contrast and N_c ; that is, $\alpha_{N_{\text{ext}}}^{\text{high}}$ vs $\alpha_{N_{\text{ext}}}^{\text{low}} + \log(N_{\text{ext}}) - \log(N_c)$ (see Eq. (1b)). The excellent correspondence evident in Fig. 8b ($r^2 = .97$) provides evidence that thresholds increase proportionally above the critical noise level; or equivalently, that the rising slope of the TvC function over high noise levels is 1.0 (Property 2).

To verify psychometric slope invariance across noise levels (Property 3), Fig. 8c presents psychometric data, re-calculated and presented as raw psychometric functions on log-relative-contrast (Strasburger, 2001a, 2001b). That is, for each noise level, signal contrasts, s , were transformed to log-threshold units, $s' = \log(s) - \alpha_{N_{\text{ext}}}$, by normalizing signal contrast by the noise level's psychometric threshold. Fig. 8c demonstrates excellent correspondence between the transformed psychometric data measured across different noise levels. The psychometric function defined on log-relative-contrast, s' , can be fit with only a slope parameter, η , and a lapse rate parameter, λ :

$$\psi(s') = .5 + (1 - .5 - \lambda/2)\{1 - \exp(-\exp(\eta \times s'))\}. \quad (11)$$

The best-fitting psychometric functions (overlaid in the figure) account well for the raw psychometric data, (mean $r^2 = .89$; $SD = .025$), across noise levels. Furthermore, a nested χ^2 test, based on the likelihood ratio between full and reduced models, in which the full model proposes distinct slopes across all noise levels and the former uses a single slope parameter across all levels, suggests that additional slope parameters do not appreciably improve the fit to psychometric data ($p > .15$ for each observer). Taken together, these analyses of the MCS data verify the regularities of TvC functions and justify the parameterization forming the basis of the $qTvC$ method.

4.2.2. Comparing the $qTvC$ and MCS

A previously fit psychometric model (independent thresholds and a single slope parameter), provides the empirical TvC functions presented in Fig. 9. Thresholds obtained with the MCS , for three performance criteria (65%, 79%, and 92% correct), across eight noise conditions, are plotted with circles, with error bars reflecting bootstrapping estimates of threshold variability (1 SD) (Wichmann &

Hill, 2001b). The constant stimulus TvC functions (τ_{MCS}) are repeated in the top and bottom rows of Fig. 9.

Threshold variability of $qTvC$ threshold estimates was estimated using a bootstrapping procedure. The mean and standard deviation of $qTvC$ parameters, (measured across four $qTvC$ runs of 240 or 480 trials), provided estimates of the mean and standard error for sampling distributions of each subject's $qTvC$ parameters. Independently sampling these distributions 5000 times provided 5000 sets of simulated threshold estimates. The variability of each threshold estimate (in dB) was calculated from the simulation results:

$$\sigma_{\hat{\tau}} = \sqrt{\frac{\sum(20\log_{10}(\bar{\tau}) - 20\log_{10}(\hat{\tau}^k))^2}{N - 1}}. \quad (12)$$

The shaded region provides the ± 1 SD region of $qTvC$ threshold estimates for three performance levels (65%, 79%, and 92% correct). The $qTvC$ threshold estimates obtained with 240 and 480 trials are, respectively, presented in the top and bottom rows of Fig. 9. The mean threshold variability at 65%, 79%, and 92% correct was 1.49, 1.13, and 1.35 dB across the external noise conditions for $qTvC$ estimates obtained with 240 trials and 1.24, .91, and 1.11 dB for those obtained with 480 trials. The mean threshold variability, for the method of constant stimuli (with 1920 trials), was .81, .72, and .76 dB for the three performance levels. Although threshold variability was lower for the method of constant stimuli, the $qTvC$'s threshold variability compares favorably, considering the adaptive procedure used only 12–25% of number of experimental trials.

To illustrate the consistency of individual $qTvC$ runs, and their correspondence with MCS results, Fig. 10 presents the aggregate TvC estimates (for 4 $qTvC$ runs for each of 3 observers). TvC estimates defined at three performance criteria (65, 79, and 92%) obtained with the $qTvC$ (4 runs of 240 or 480 trials), are plotted against MCS estimates obtained with 1920 trials. A regression analysis was performed to examine correspondence between mean TvC estimates obtained with the $qTvC$ and MCS . Due to differences in variability between threshold estimates, both within and between estimation methods, a Deming (type II) weighted regression analysis (Linnet, 1998; Martin, 2000) was used to analyze the agreement between thresholds obtained with the $qTvC$ procedure and the MCS . This analysis, accounting for measurement errors in both methods, finds the linear relation:

$$\log(\tau_{qTvC}) = a + b \log(\tau_{MCS}). \quad (13)$$

A b close to 1.0 would imply excellent agreement between the two different procedures and provide an experimental validation of the $qTvC$ procedure.

The analysis was applied to the mean aggregate data, to relate mean $qTvC$ estimates obtained with 240 and 480 trials with MCS estimates obtained with 1920 trials (see Fig. 9). There was excellent agreement between TvC estimates: for 240 trials, slope $b = 0.974$ ($\sigma_b = 0.024$), intercept $a = -0.139$ ($\sigma_a = 0.062$), and weighted correlation coefficient $r = .984$; for 480 trials, $b = 0.937$ ($\sigma_b = 0.021$),

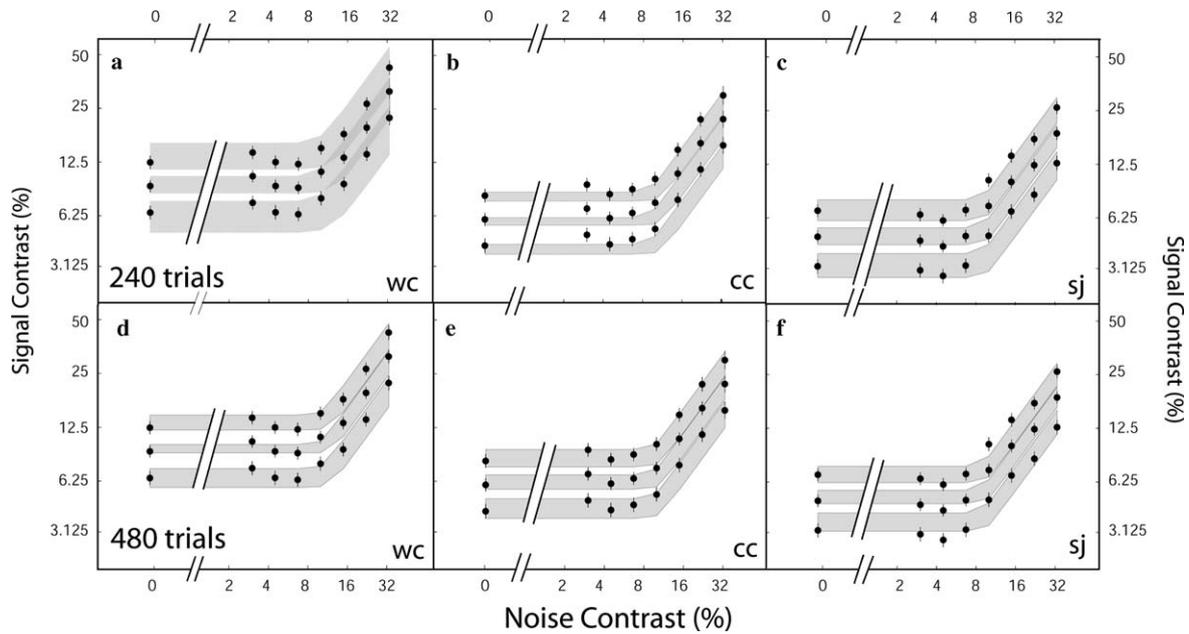


Fig. 9. Empirical TvC functions (obtained with $qTvC$ and MCS) are measured for three observers. The TvC functions (at 65%, 79%, and 92%) collected with the MCS , presented as circles, are repeated in the top and bottom rows, with error bars reflecting variability estimates (1 SD). Presented as shaded regions ($\pm 1 SD$) are TvC functions estimated with the $qTvC$, with 240 trials (a–c) or 480 trials (d–f). There is excellent agreement between the TvC estimates obtained with both methods. The variability of MCS estimates is less than that exhibited by $qTvC$ estimates, but $qTvC$ estimates used only 12–25% of the trials.

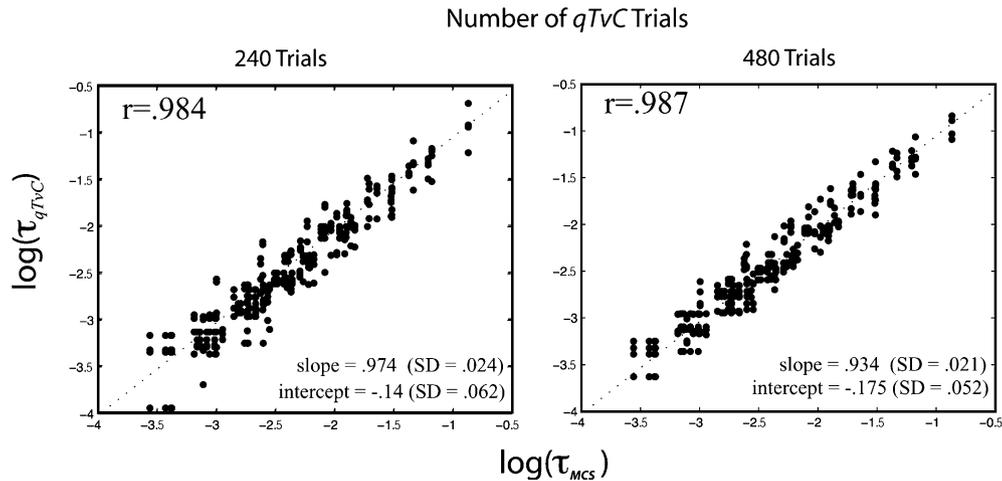


Fig. 10. Comparing TvC estimates obtained with $qTvC$ and the method of constant stimuli (MCS). Presented are TvC estimates provided by four $qTvC$ runs, plotted against an estimate obtained with constant stimuli. A weighted Deming regression, used for method comparison (Martin, 2000), provides the least squares line fit and weighted correlation coefficient, between the mean $qTvC$ and MCS threshold estimates.

$a = -0.175$ ($\sigma_a = 0.052$), and $r = .987$. Furthermore, following Martin (2000), we computed the systematic bias, $\hat{B}(c)$, between threshold estimates at a given contrast threshold c , obtained with the $qTvC$ and the MCS , as:

$$\hat{B}(c) = a + c(b - 1) \quad (14)$$

and the standard deviation of estimated bias as:

$$\sigma_{\hat{B}(c)} = \sqrt{\sigma_a^2 + c \times \sigma_b^2 \times (c - 2 \times \log(\tau_{MCS}))}, \quad (15)$$

where σ_a is the standard deviation of the intercept estimate a , σ_b is the standard deviation of the slope estimate b , and $\log(\tau_{MCS})$ is the weighted mean of threshold estimates

obtained with the method of constant stimuli. We computed systematic bias, $\hat{B}(c)$, for three contrast levels, $c = 3.6\%$, 16.4% , and 72.7% , representing the lowest, middle, and highest thresholds measured in the present experiment. Respectively, systematic bias estimates at these contrast levels were, for 240 trials: $\hat{B}(c) = 0.134, -0.002$, and -0.037 , and $\sigma_{\hat{B}(c)} = .084, .061$, and $.054$; and for 480 trials, $\hat{B}(c) = 0.129, 0.003$, and -0.030 , and $\sigma_{\hat{B}(c)} = 0.074, 0.053$, and 0.046 . These results suggest that over the range of threshold measurements observed in the present experiment, there was no significant systematic bias between $qTvC$ and MCS threshold estimates.

4.3. Summary

The results of simulations and psychophysical studies show that the $qTvC$ method needs less than 300 trials to successfully estimate TvC functions at three performance criteria, with acceptable precision (< 1.5 dB). Further, both studies showed that the $qTvC$ method provides accurate estimates of TvC functions at three performance criteria (e.g., 65%, 79%, and 92% correct) with only 240 trials of data collection.

With this number of trials, the precision of these estimates is well within that expected for clinical or practical applications. As the psychophysical results demonstrate, more trials (~ 480) may be necessary for the precision desired in experimental applications. With 480 trials, the precision of the $qTvC$ is comparable with the method of constant stimuli with 1920 trials. We believe that the small precision loss is compensated by: (1) reducing data collection by 75%; and (2) not worrying about stimulus placement. Without ideal stimulus placement, the estimates provided by constant stimulus methods rapidly lose precision. The $qTvC$ possesses the advantage of placing stimuli without any input from the experimenter.

The results of both the Monte-Carlo simulations and the psychophysical experiment strongly suggest that the $qTvC$ method can measure external noise functions rapidly with reasonable accuracy and precision.

5. Discussion

In addition to using the external noise method to investigate mechanisms underlying observer state changes (e.g., perceptual learning, memory decay, attention), there has been great interest in applying the method to characterize visual deficits in human observers. The goal of the current research is to alleviate the burden of data collection without compromising the rich data provided by external noise methods. Built on an adaptive method developed to estimate psychometric functions (Cobo-Lewis, 1996; Kontsevich & Tyler, 1999), we developed a Bayesian adaptive procedure that exploits the regularities of TvC functions to efficiently sample the stimulus space of signal and noise contrast. Recently, Jeon et al. (in preparation) have developed a maximum-likelihood procedure to estimate the parameters of the perceptual template model using the data obtained from the $qTvC$ method. The procedure provides a good avenue for theoretical modeling of the outcomes of the $qTvC$ method.

By exploiting the regularities of the TvC functions, the $qTvC$ method is much more efficient than the method of constant stimuli and other adaptive procedures that independently estimate thresholds in each external noise condition. The efficiency of the $qTvC$ emerges from its parameterization of external noise functions (suggested by the empirical TvC 's regularities) and its stimulus placement algorithm. Therefore, the $qTvC$ method holds great advantages in domains in which these regularities hold. On the other hand, it would be prudent to use classical

methods to validate the assumed regularities in each new domain of application. The experiment in this study provides an example procedure for validating the $qTvC$ method. We believe the $qTvC$ method will greatly facilitate the application of external noise methods to examine perceptual deficits accompanying dyslexia, amblyopia, or senescence.

There are several avenues for improving the application of the $qTvC$ method. Successful applications of adaptive methods depend on the assumption that subjects do not themselves adapt their behavior to the stimulus placement algorithm. Although the agreement between TvC estimates obtained with the MCS and the $qTvC$ suggests that this assumption was not violated in our study, future development could constrain stimulus placement. For example, the stimulus placement algorithm could be computed over partitions of the stimulus space, which can be alternated between trials (as with interleaving staircases). Though such constraints might make stimulus placement slightly less than optimal, it may improve the practical application of the $qTvC$ method.

Two other aspects of the $qTvC$ application can also be greatly improved: (1) In the current method, entropy calculations for the posterior probability density (defined over the three-dimensional parameter space) weigh all points in the parameter space equally. Future development could apply differential weights on regions of the parameter space, as Tanner and colleagues have suggested for estimating psychometric functions (Tanner, Hill, Rasmussen, & Wichmann, 2005). (2) The current method determines the most informative stimulus for the next trial using a one-step ahead exhaustive search over a discretely sampled stimulus space. Although adequate in the current study, more efficient optimization of stimulus placement may result from using smooth functions to approximate the expected entropy function (Cobo-Lewis, 1997) or the marginal probability distributions in the parameter space. Searches for optimal stimuli could then be conducted over continuous stimulus dimensions. Additionally, in addition to entropy, other information-theoretic measures, such as Fisher information (Cobo-Lewis, 1999), or mutual information (Cobo-Lewis, 1996; Paninski, 2005), can be used to optimize stimulus sampling. Currently, such measures can be computationally intensive, but this approach could be desirable for future development.

The adaptive Bayesian framework, originally developed to estimate psychometric threshold (Watson & Pelli, 1983), has been since refined by others to estimate the slope, lower, and upper asymptotes of psychometric functions (Cobo-Lewis, 1996, 1999; King-Smith & Rose, 1997; Kontsevich & Tyler, 1999; Tanner et al., 2005). In this paper, we elaborated the framework to estimate TvC functions at multiple performance criteria. The framework can be applied to measure any behavioral function with a relatively simple and well-established functional form.

Building on the example of the $qTvC$ method, we have developed other adaptive Bayesian procedures

(“quick Methods” or “qMethods”) that search one-step ahead of the current trial (over a multi-dimensional stimulus space), and use the minimum expected entropy criterion to choose the stimulus maximizing the information (defined over a multi-dimensional parameter space) on the next trial. For example, the *quick CSF* (*qCSF*) method maximizes the information gained about four parameters describing the contrast sensitivity function (Campbell & Robson, 1968; Watson & Ahumada, 2005), by choosing stimuli of the appropriate spatial frequency and contrast on each trial. Other procedures we have developed measure psychometric functions in Yes/No tasks (Green & Swets, 1966), temporal contrast sensitivity functions (de Lange, 1958; Kelly, 1984), elliptical equi-discrimination contours (MacAdam, 1942; MacLeod & Boynton, 1979), visual perimeteric fields (Bengtsson, Heijl, & Rootzen, 1997), and sensory memory decay functions (Sperling, 1960). These classical functions, and their corresponding psychophysical models, have previously provided great insight into perceptual mechanisms, but required much data collection. Developing rapid and accurate methods for measuring these behavioral functions will allow the application of rich psychophysical methods without the burden of data collection previously needed in the laboratory setting.

Acknowledgments

This research was supported by US Air Force Office of Scientific Research, Visual Information Processing Program, NSF grants BCS-9911801 and BCS-9910678, and NIMH grant 1 R01 MH61834-01. We would like to thank Alan Cobo-Lewis and Mark Steyvers for useful conversations.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.visres.2006.04.022](https://doi.org/10.1016/j.visres.2006.04.022).

References

- Ahumada, A. J. (1987). Putting the visual system noise back in the picture. *Journal of the Optical Society of America A*, 4, 2372–2378.
- Ahumada, A. J., & Watson, A. B. (1985). Equivalent-noise model for contrast detection and discrimination. *Journal of the Optical Society of America A*, 2(7), 1133–1139.
- Alcala-Quintana, R., & Garcia-Perez, M. A. (2004). The role of parametric assumptions in adaptive bayesian estimation. *Psychological Methods*, 9(2), 250.
- Barlow, H. B. (1956). Retinal noise and absolute threshold. *Journal of the Optical Society of America*, 46, 634–639.
- Bengtsson, B. O. J., Heijl, A., & Rootzen, H. (1997). A new generation of algorithms for computerized threshold perimetry, SITA. *Acta Ophthalmologica Scandinavica*, 75, 368–375.
- Bennett, P. J., Sekuler, A. B., & Ozin, L. (1999). Effects of ageing on calculation efficiency and equivalent noise. *Journal of the Optical Society of America A*, 16(3), 654.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10(4), 433–436.
- Burgess, A. E., & Colborne, B. (1988). Visual signal detection: IV. Observer inconsistency. *Journal of the Optical Society of America A*, 2, 617–627.
- Burgess, A. E., Humphrey, K., & Wagner, R. F. (1979). Detection of bars and discs in quantum noise. *Proceedings of the Society of Photo-optical Instrumentation Engineers*, 173, 34–40.
- Burgess, A. E., Wagner, R. F., Jennings, R. J., & Barlow, H. B. (1981). Efficiency of human visual signal discrimination. *Science*, 214(4516), 93–94.
- Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology (London)*, 197, 551–566.
- Chung, S. T. L., Levi, D. M., & Tjan, B. S. (2005). Learning letter identification in peripheral vision. *Vision Research*, 45(11), 1399–1412.
- Cobo-Lewis, A. B. (1996). An adaptive method for estimating multiple parameters of a psychometric function. *Journal of Mathematical Psychology*, 40, 353–354.
- Cobo-Lewis, A. B. (1997). An adaptive psychophysical method for subject classification. *Perception & Psychophysics*, 59, 989–1003.
- Cobo-Lewis, A. B. (1999). Parameterization-invariant adaptive methods for psychophysics. *Journal of Mathematical Psychology*, 43, 580–588.
- Cohn, T. E. (1976). Detectability of a luminance increment—Effect of superimposed random luminance fluctuation. *Journal of the Optical Society of America*, 66(12), 1426–1428.
- Dao, D. Y., Lu, Z. -L., & Doshier, B. A. (2006). Adaptation to sine-wave gratings selectively reduces the sensory gain of the adapted stimuli. *Journal of Vision*, in press.
- de Lange, H. (1958). Research into the dynamic nature of the human fovea-cortex systems with intermittent and modulated light. II. Phase shift in brightness and delay in color perception. *Journal of the Optical Society of America*, 48, 784–789.
- Doshier, B. A., & Lu, Z.-L. (1998). Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. *Proceedings of the National Academy of Sciences of the United States of America*, 95(23), 13988–13993.
- Doshier, B., & Lu, Z.-L. (1999). Mechanisms of perceptual learning. *Vision Research*, 39, 3197–3221.
- Doshier, B. A., & Lu, Z.-L. (2000). Noise exclusion in spatial attention. *Psychological Science*, 11(2), 139–146.
- Eckstein, M. P., Ahumada, A. J., Jr., & Watson, A. B. (1997). Visual signal detection in structured backgrounds: II. Effects of contrast gain control, background variations, and white noise. *Journal of the Optical Society of America A*, 14(9), 2406–2419.
- Eckstein, M. P., Pham, B. T., & Shimozaki, S. S. (2004). The footprints of visual attention during search with 100% valid and 100% invalid cues. *Vision Research*, 44(12), 1193–1207.
- Gold, J. (2001). Signal and noise in perceptual learning. Ph.D. dissertation. Toronto, Canada: University of Toronto.
- Gold, J., Bennett, P. J., & Sekuler, A. B. (1999). Signal but not noise changes with perceptual learning. *Nature*, 402(6758), 176.
- Green, D. M. (1990). Stimulus selection in adaptive psychophysical procedures. *The Journal of the Acoustical Society of America*, 87(6), 2662–2674.
- Green, D. M. (1995). Maximum-likelihood procedures and the inattentive observer. *Journal of the Acoustical Society of America*, 97(6), 3749–3760.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: John Wiley.
- Griffiths, J. W. R., & Nagaraja, N. S. (1963). Visual detection in intensity-modulated displays. *Radio and Electronic Engineer*, 25, 225–240.
- Huang, C. B., Tao, L. M., Zhou, Y. F., & Lu, Z.-L. (submitted for publication). Treated amblyopes remain deficient in spatial vision: A contrast sensitivity and external noise study. *Investigative Ophthalmology & Visual Science*.
- Jeon, S. T., Lu, Z. L., Lesmes, L. A., & Doshier, B. A. (in preparation). Fitting the perceptual template model to adaptive external noise functions.
- Kelly, D. H. (1984). Retinal inhomogeneity: I. Spatiotemporal contrast sensitivity. *Journal of the Optical Society of America A*, 1(1), 107–113.

- King-Smith, P. E., & Rose, D. (1997). Principles of an adaptive method for measuring the slope of the psychometric function. *Vision Research*, 37(12), 1595.
- Kiorpes, L., & Movshon, J. A. (1998). Peripheral and central factors limiting the development of contrast sensitivity in macaque monkeys. *Vision Research*, 38(1), 61–70.
- Klein, S. A. (2001). Measuring, estimating, and understanding the psychometric function: A commentary. *Perception & Psychophysics*, 63(8), 1421–1455.
- Kontsevich, L. L., & Tyler, C. W. (1999). Bayesian adaptive estimation of psychometric slope and threshold. *Vision Research*, 39(16), 2729–2737.
- Leek, M. R. (2001). Adaptive procedures in psychophysical research. *Perception & Psychophysics*, 63(8), 1279–1292.
- Legge, G. E., & Foley, J. M. (1980). Contrast masking in human vision. *Journal of the Optical Society of America*, 70(12), 1458–1471.
- Legge, G. E., Kersten, D., & Burgess, A. E. (1987). Contrast discrimination in noise. *Journal of the Optical Society of America A*, 4(2), 391–404.
- Legge, G. E., Rubin, G. S., Pelli, D. G., & Schleske, M. M. (1988). Understanding low vision reading. *Journal of Visual Impairment & Blindness*, 82(2), 54–59.
- Levi, D. M., & Klein, S. A. (2003). Noise provides some new signals about the spatial vision of amblyopes. *Journal of Neuroscience*, 23(7), 2522–2526.
- Li, X., Lu, Z.-L., Xu, P., Jin, J., & Zhou, Y. (2003). Generating high gray-level resolution monochrome displays with conventional computer graphics cards and color monitors. *Journal of Neuroscience Methods*, 130, 9–18.
- Linnet, K. (1998). Performance of Deming regression analysis in case of misspecified analytical error ratio in method comparison studies. *Clinical Chemistry*, 44(5), 1024–1031.
- Lu, Z.-L., & Doshier, B. A. (1998). External noise distinguishes attention mechanisms. *Vision Research*, 38(9), 1183–1198.
- Lu, Z.-L., & Doshier, B. A. (1999). Characterizing human perceptual inefficiencies with equivalent internal noise. *Journal of the Optical Society of America A*, 16(3), 764–778.
- Lu, Z.-L., & Doshier, B. A. (2000). Spatial attention: Different mechanisms for central and peripheral cues? *Journal of Experimental Psychology: Human Perception and Performance*, 26, 1534–1548.
- Lu, Z.-L., & Doshier, B. A. (2002). External noise methods and observer models. *Investigative Ophthalmology & Visual Science*, 43, U821.
- Lu, Z.-L., & Doshier, B. A. (2004a). Perceptual learning retunes the perceptual template in foveal orientation identification. *Journal of Vision*, 4, 44–56.
- Lu, Z.-L., & Doshier, B. A. (2004b). Spatial attention excludes external noise without changing the spatial frequency tuning of the perceptual template. *Journal of Vision*, 4(10), 955–966.
- Lu, Z.-L., & Doshier, B. A. (in preparation). Using external noise methods and observer models to characterize observer states.
- Lu, Z.-L., Jeon, S. T., & Doshier, B. A. (2004). Temporal tuning characteristics of the perceptual template and endogenous cuing of spatial attention. *Vision Research*, 44(12), 1333–1350.
- MacAdam, D. L. (1942). Visual sensitivities to color differences in daylight. *Journal of the Optical Society of America*, 32, 247–273.
- MacLeod, D., & Boynton, R. M. (1979). Chromaticity diagram showing cone excitation by stimuli of equal luminance. *Journal of the Optical Society of America*, 69, 1183–1186.
- Mangini, M. C., & Biederman, I. (2002). Prosopagnosics have low internal noise? *Journal of Vision*, 2(7), 609a.
- Martin, R. F. (2000). General deming regression for estimating systematic bias and its confidence interval in method-comparison studies. *Clinical Chemistry*, 46(1), 100–104.
- Mortensen, U. (2002). Additive noise, Weibull functions and the approximation of psychometric functions. *Vision Research*, 42(20), 2371–2393.
- Nagaraja, N. S. (1964). Effect of luminance noise on contrast thresholds. *Journal of the Optical Society of America*, 54(7), 950–955.
- Paninski, L. (2005). Asymptotic theory of information—Theoretic experimental design. *Neural Computation*, 17, 1480–1507.
- Pardhan, S. (2004). Contrast sensitivity loss with aging: Sampling efficiency and equivalent noise at different spatial frequencies. *Journal of the Optical Society of America A*, 21(2), 169–175.
- Pelli, D. G. (1981). Effects of visual noise. Ph.D. dissertation, Cambridge, England: University of Cambridge.
- Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A*, 2, 1508–1532.
- Pelli, D. G. (1987). On the relation between summation and facilitation. *Vision Research*, 27(1), 119–123.
- Pelli, D. G. (1990). The quantum efficiency of vision. In C. Blakemore (Ed.), *Vision: Coding and efficiency* (pp. 3–24). Cambridge, UK: Cambridge University Press.
- Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442.
- Pelli, D., & Farell, B. (1999). Why use noise? *Journal of the Optical Society of America A*, 16(3), 647–653.
- Pelli, D. G., Levi, D. M., & Chung, S. T. L. (2004). Using visual noise to characterize amblyopic letter identification. *Journal of Vision*, 4(10), 904–920.
- Skoczenski, A. M., & Norcia, A. M. (1998). Neural noise limitations on infant visual sensitivity. *Nature*, 391, 697–700.
- Sperling, G. (1960). The information available in brief visual presentation. *Psychological Monographs*, 74(11, Whole No. 498), 29.
- Sperling, A. J., Lu, Z.-L., Manis, F. R., & Seidenberg, M. S. (2005). Deficits in perceptual noise exclusion in developmental dyslexia. *Nature Neuroscience*, 8(7), 862–863.
- Strasburger, H. (2001a). Converting between measures of slope of the psychometric function. *Perception & Psychophysics*, 63(8), 1348–1355.
- Strasburger, H. (2001b). Invariance of the psychometric function for character recognition across the visual field. *Perception & Psychophysics*, 63(8), 1356–1376.
- Stromeyer, C. F., & Julesz, B. (1972). Spatial frequency masking in vision: Critical bands and spread of masking. *Journal of the Optical Society of America*, 62, 1221–1232.
- Tanner, T. G., Hill, N. J., Rasmussen, C. E., & Wichmann, F. A. (2005). Efficient adaptive sampling of the psychometric function by maximizing information gain. In H. H. Bülthoff, A. Mallot, R. Ulrich, & F. A. Wichmann (Eds.), *Proceedings of the 8th Tübingen Perception Conference* (Vol. 106). Kirchentellisfurt: Knirsch Verlag.
- Tjan, B. S., Braje, W. L., Legge, G. E., & Kersten, D. (1995). Human efficiency for recognizing 3-D objects in luminance noise. *Vision Research*, 35(21), 3053–3069.
- Tjan, B. S., Chung, S., & Levi, D. (2002). Limitation of ideal-observer analysis in understanding perceptual learning. In *Twenty-seventh annual interdisciplinary conference*, Jackson Hole, Wyoming.
- Treutwein, B. (1995). Adaptive psychophysical procedures. *Vision Research*, 35(17), 2503–2522.
- Treutwein, B., & Strasburger, H. (1999). Fitting the psychometric function. *Perception & Psychophysics*, 61(1), 87–106.
- Tyler, C. W., & Chen, C.-C. (2000). Signal detection theory in the 2AFC paradigm: Attention, channel uncertainty and probability summation. *Vision Research*, 40(22), 3121–3144.
- van Meeteren, A., & Boogaard, J. (1973). Visual contrast sensitivity with ideal image intensifiers. *Optik*, 37, 179–191.
- Watson, A. B., & Ahumada, A. J. (2005). A standard model for foveal detection of spatial contrast. *Journal of Vision*, 5(9), 717–740, <http://journalofvision.org/5/9/6/>.
- Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, 33(2), 113–120.
- Wichmann, F. A., & Hill, N. J. (2001a). The psychometric function I: Fitting, sampling and goodness-of-fit. *Perception & Psychophysics*, 63(8), 1293–1313.
- Wichmann, F. A., & Hill, N. J. (2001b). The psychometric function II: Bootstrap based confidence intervals and sampling. *Perception & Psychophysics*, 63(8), 1314–1329.
- Xu, P. J., Lu, Z.-L., Qiu, Z. P., & Zhou, Y. -F. (under review). Identifying mechanisms of amblyopia in gabor orientation identification with external noise. *Vision Research*.