Hierarchical Bayesian modeling of contrast sensitivity functions in a within-subject design

Yukai Zhao
Center for Neural Science, New York University, New York, NY, USA

Luis Andres Lesmes
Adaptive Sensory Technology Inc., San Diego, CA, USA

Fang Hou
School of Ophthalmology & Optometry and Eye Hospital, Wenzhou Medical University, Wenzhou, Zhejiang, China

Division of Arts and Sciences, NYU Shanghai, Shanghai, China

Center for Neural Science and Department of Psychology, New York University, New York, NY, USA

NYU-ECNU Institute of Brain and Cognitive Science, Shanghai, China

Zhong-Lin Lu

Recent development of the quick contrast sensitivity function (qCSF) method has made it possible to obtain accurate, precise, and efficient contrast sensitivity function (CSF) assessment. To improve statistical inference on CSF changes in a within-subject design, we developed a hierarchical Bayesian model (HBM) to compute the joint distribution of CSF parameters and hyperparameters at test, subject, and population levels, utilizing information within- and between-subjects and experimental conditions. We evaluated the performance of the HBM relative to a non-hierarchical Bayesian inference procedure (BIP) on an existing CSF dataset of 112 subjects obtained with the qCSF method in three luminance conditions (Hou, Lesmes, Kim, Gu, Pitt, Myung, & Lu, 2016). We found that the average $d'$s of the area under log CSF (AULCSF) and CSF parameters between pairs of luminance conditions at the test-level from the HBM were 33.5% and 103.3% greater than those from the BIP analysis of AULCSF. The increased $d'$ resulted in greater statistical differences between experimental conditions across subjects. In addition, simulations showed that the HBM generated accurate and precise CSF parameter estimates. These results have strong implications for the application of HBM in clinical trials and patient care.

Introduction

The contrast sensitivity function (CSF), which quantifies the visibility (1/threshold) of narrow-band filtered stimuli over a wide range of spatial frequencies, provides a comprehensive measure of spatial vision (Ginsburg, 1981; Ginsburg, 2003; Hess, 1981). It is closely related to daily visual functions (Ginsburg, 2003), and can better quantify deficits in spatial vision than visual acuity (Jindra & Zemon, 1989; Marmor, 1986). It has long been recognized that the CSF provides important information for monitoring progression of vision change and evaluating treatment efficacy in eye diseases (Bellucci, Scialdone, Buratto, Morselli, Chierego, Criscuolo, Criscuoli, Moretti, & Piers, 2005; Ginsburg, 2006; Loshin & White, 1984; Levi & Li, 2009; Tan & Fong, 2008; Zhou, Huang, Xu, Tao, Qiu, Li, & Lu, 2006).

Despite its clinical promise, precise and efficient CSF assessment has presented a challenge. The CSF charts provide a fast but imprecise assessment of contrast sensitivity due to coarse sampling of both spatial frequency and stimulus contrast (Bradley, Hook, & Haeseker, 1991; Buhren, Terzi, Bach, Wesemann, & Kohnen, 2006; Hohberger, Laemmer, Adler, Juenemann, & Horn, 2007; Pesudovs, Hazel, Doran, & Elliott, 2004; van Gaalen, Jansonius, Koopmans, Terwee, & Kooijman, 2009). On the other hand, the long testing time (30–60 minutes) required for measuring the CSF with conventional psychophysical methods has prevented their clinical applications (Kelly & Savoie, 1973; Treutwein, 1995). The quick contrast sensitivity function (qCSF) was developed to address the challenges (Lesmes, Lu, Baek, & Albright, 2010). Based on active learning principles, it estimates

A Bayesian Inference Procedure (BIP) has been developed to make statistical inference on CSF changes in a within-subject design based on CSF metrics extracted from each subject in each experimental condition (Hou, et al., 2016; Kuss, Jäkel, & Wichmann, 2005; Prins, 2013; Schütt, Harmeling, Macke, & Wichmann, 2016). Because it scores each test independently with an uninformative prior without considering potential relationships of CSF parameters across subjects and experimental conditions, the BIP may have overestimated the variance of each test and resulted in reduced statistical power (Borm, Fransen, & Lemmens, 2007; Egbewale, Lewis, & Sim, 2014; Wilcox, 2012). In addition, a single summary metric, the area under log CSF (AULCSF), is usually used to compare CSFs in different experimental conditions, potentially leaving out information in the multidimensional joint distribution of CSF parameters.

In this study, we developed a Hierarchical Bayesian Model (HBM) to reduce the variability of estimated CSF parameters for each test and to further improve the ability to detect between-condition CSF changes in a within-subject design. The HBM is a generative model framework that uses Bayes’ rule to quantify the joint distribution of test-, subject-, and population-level parameters and hyperparameters (Kruschke, 2015; Lee, 2006; Lee, 2011; Rouder & Lu, 2005; Wilson, Cranmer, & Lu, 2020). It explicitly quantifies the covariance of the hyperparameters and parameters (Daniels & Kass, 1999; Klotzke & Fox, 2019; Thall, Wathen, Bekele, Champlin, Baker, & Benjamin, 2003; Wang, Lin, & Nelson, 2020; Yang, Zhu, Choi, & Cox, 2016). By sharing information within and across levels via conditional dependencies, it reduces the variance of the test-level estimates through (1) decomposition of variabilities from different sources (test, subject, and population) with parameters and hyperparameters (Song, Behmanesh, Moaveni, & Papadimitriou, 2020), and (2) shrinkage of the estimated parameters at the lower levels toward the mean of the higher levels when there is not sufficient data at the lower level (Kruschke, 2015; Rouder & Lu, 2005; Rouder, Sun, Speckman, Lu, & Zhou, 2003).

Although it has been used in many different disciplines, such as astronomy (Thrane & Talbot, 2019), ecology (Reum, Hovel, & Greene, 2015; Wikle, 2003), genetics (Storz & Beaumont, 2002), machine learning (Li & Perona, 2005), cognitive science (Ahn, Krawitz, Kim, Busmeyer, & Brown, 2011; Lee, 2006; Lee & Mumford, 2003; Merkle, Smithson, & Verkuilen, 2011; Molloy, Bahg, Li, Steyvers, Lu, & Turner, 2018; Molloy, Bahg, Lu, & Turner, 2019; Rouder & Lu, 2005; Rouder et al., 2003; Wilson et al., 2020) and visual acuity (Zhao, Lesmes, Dorr, & Lu, 2021), HBM has not been applied to analyze the CSF. Here, we develop a three-level HBM to model the entire CSF dataset in a single-factor (luminance), multi-condition (3 luminance conditions), and within-subject experiment design. We modeled the data with CSF parameters at the test level and hyperparameters at the individual and population levels, with conditional dependencies across levels. We evaluated the performance of the HBM relative to the BIP using an existing dataset of 112 subjects tested with qCSF in three luminance conditions (Hou et al., 2016), which was collected to mimic mild, medium, and large CSF changes observed in clinical settings (Bellmann, Unnebrink, Rubin, Miller, & Holz, 2003; Haymes, Roberts, Cruess, Nicolela, LeBlanc, Ramsey, Chauhan, & Artes, 2006; Kalia, Lesmes, Dorr, Gandhi, Chatterjee, Ganesh, Bex, & Sinha, 2014; Kleiner, Enger, Alexander, & Fine, 1988; Midena, Degli Angeli, Blarzino, Valenti, & Segato, 1997; Owsley, Sekuler, & Siemsen, 1983). In addition, a simulation study was conducted to evaluate and compare the accuracy and precision of the estimates from the HBM and BIP. We hypothesized that, relative to the BIP, the HBM would reduce the variability of the estimated CSF parameters from each test, increase the d's of CSF changes between luminance conditions for each subject, and improve statistical inference across subjects.
Bayesian modeling of the CSF

Overview

In a typical within-subject design CSF experiment with multiple conditions, the trial-by-trial data can be organized as $y_{ijkm} = (f_{ijkm}, c_{ijkm}, r_{ijkm})$, where $r_{ijkm}$, either correct or incorrect, is individual $i$’s response in trial $m$ of test $k$ in experimental condition $j$ tested with a stimulus of spatial frequency $f_{ijkm}$ and contrast $c_{ijkm}$. The BIP consists of four components (Hou et al., 2016): (1) a log parabola model of the contrast sensitivity function with several parameters, (2) a likelihood function that specifies the probability of making a correct or incorrect response in each stimulus condition, (3) a Bayesian procedure to infer the posterior distribution of the CSF parameters for each subject in each test, and (4) inference based on statistics computed from posterior distributions either at the subject level or aggregated across subjects. In this section, we first provide a brief review of the BIP, and then introduce the HBM.

The Bayesian inference procedure

In the BIP (Figures 1, 2a), the contrast sensitivity $S(f_{ijkm}, \theta_{ijk})$ at spatial frequency $f_{ijkm}$ is modeled with a log parabola function with three parameters, $\theta_{ijk} = (\gamma_{ijk}^{max}, f_{ijk}^{max}, \beta_{ijk})$ (Lesmes et al., 2010; Rohaly & Owsley, 1993; Watson, & Ahumada Jr, 2005):$^1$

$$\log_{10}\left(S(f_{ijkm}, \theta_{ijk})\right) = \log_{10}\left(\gamma_{ijk}^{max}\right) - 4 \log_{10}\left(2\right) \left(\frac{\log_{10}\left(f_{ijkm}\right) - \log_{10}\left(f_{ijk}^{max}\right)}{\beta_{ijk}}\right)^2,$$  \hspace{1cm} (1)

where $\gamma_{ijk}^{max}$ is the peak sensitivity, $f_{ijk}^{max}$ is the peak spatial frequency (cycles/degree), and $\beta_{ijk}$ is the bandwidth (octaves) at half of the peak sensitivity. The probability of making a correct response is described with a psychometric function (Hou et al., 2015):
Figure 2. (a) The Bayesian Inference Procedure (BIP) computes the posterior distribution of CSF parameters for each test independently. (b) A three-level hierarchical Bayesian model (HBM) of CSFs across multiple individuals, conditions and tests. At the population level, \( \mu \) and \( \Sigma \) are the mean and covariance hyperparameters of the population. At the individual level \( \rho_{ij} \) and \( \phi_j \) are the mean and covariance hyperparameters of individual \( i \) in experimental condition \( j \). At the test level, \( \theta_{ijk} \) is the CSF parameter of individual \( i \) in test \( k \) of condition \( j \).

\[
p(r_{ijkm} = 1|\theta_{ijk}, f_{ijkm}, c_{ijkm}) = g + \left( 1 - g - \frac{\lambda}{2} \right) \Phi \times \left( \frac{\log_{10}(c_{ijkm}) + \log_{10}(S(f_{ijkm}, \theta_{ijk}))}{\sigma} \right), \tag{2}
\]

where \( g \) is the guessing rate, \( \lambda \), usually set to 0.04 (Lesmes et al., 2010; Wichmann & Hill, 2001), is the lapse rate, \( \Phi \) is the standard cumulative Gaussian function, and \( \sigma \) determines the steepness of the psychometric function. The probability of making an incorrect response is:

\[
p(r_{ijkm} = 0|\theta_{ijk}, f_{ijkm}, c_{ijkm}) = 1 - p(r_{ijkm} = 1|\theta_{ijk}, f_{ijkm}, c_{ijkm}). \tag{3}
\]

Equations 2 and 3 define the likelihood function, that is, the probability of making a correct or incorrect response given the stimulus and CSF parameters in a trial. The goal in most experiments is to infer the CSF parameters from the experimental data, that is, estimate the posterior distribution \( p(\theta_{ijk}|Y_{ijk}) \) — the distribution of the CSF parameters \( \theta_{ijk} \) given the experimental data \( Y_{ijk} = \{y_{ijkm}\} \), for \( m = 1, \ldots, M \), where \( M \) is the total number of trials in a test. This can be accomplished using Bayes’ rule:

\[
p(\theta_{ijk}|Y_{ijk}) = \frac{\prod_{m=1}^{M} p(r_{ijkm}|\theta_{ijk}, f_{ijkm}, c_{ijkm}) p_0(\theta_{ijk})}{\int \prod_{m=1}^{M} p(r_{ijkm}|\theta_{ijk}, f_{ijkm}, c_{ijkm}) p_0(\theta_{ijk}) \, d\theta_{ijk}}, \tag{4}
\]

where \( p_0(\theta_{ijk}) \) is the prior probability distribution of the CSF parameters for individual \( i \) in test \( k \) of experimental condition \( j \), which is usually uninformative and the same for all subjects and experimental conditions, and the denominator is the integral across all possible values of \( \theta_{ijk} \), and is a constant for a given dataset and BIP.

The hierarchical Bayesian model

We developed a three-level HBM to account for the entire dataset, incorporating conditional dependencies across test, individual, and population levels to improve estimates for each test (see Figure 2b). The HBM is based on three properties: (1) CSF parameters at the test level are conditionally dependent on hyperparameters at the individual level, (2) CSF hyperparameters at the individual level are conditionally dependent on those at the population level (“conditional dependency”), and (3) the probability \( p(r_{ijkm}|\theta_{ijk}, f_{ijkm}, c_{ijkm}) \) of response \( r_{ijkm} \) is determined only by the CSF parameters \( \theta_{ijk} \) in that test (Equations 2, 3).

In the HBM, the joint distribution of CSF hyperparameter \( \eta \) across all the \( J \) experimental conditions at the population level, \( p(\eta) \), is modeled as a mixture of \( 3 \times J \)-dimensional Gaussian distributions \( \mathcal{N} \) with mean \( \mu \) and covariance \( \Sigma \), which have distributions \( p(\mu) \) and \( p(\Sigma) \):

\[
p(\eta) = \mathcal{N}(\eta, \mu, \Sigma) \, p(\mu) \, p(\Sigma). \tag{5}
\]
The joint distribution of CSF hyperparameter $\tau_{i,j}$ of individual $i$ across all experimental conditions $1:J$ at the individual level, $p(\tau_{i,j} | \eta)$, is modeled as mixtures of three-dimensional Gaussian distributions with mean $\rho_{ij}$ and covariance $\phi_{ij}$, which have distributions $p(\rho_{i,j} | \eta)$ and $p(\phi_j)$:

$$p(\tau_{i,j} | \eta) = p(\rho_{i,j} | \eta) \prod_{j=1}^J \mathcal{N}(\tau_{ij}, \rho_{ij}, \phi_j) p(\phi_j),$$

where $p(\rho_{i,j} | \eta)$ denotes that $\rho_{i,j}$ is conditioned on $\eta$, and $\phi_j$ is a $3 \times 3$ covariance matrix in experimental condition $j$. Finally, at the test level, $p(\theta_{ijk} | \tau_{ij})$, the joint distribution of the CSF parameters, $\theta_{ijk}$, is conditioned on $\tau_{ij}$.

The probability of obtaining the entire dataset is computed by probability multiplication:

$$p(Y_{1:J,1:K,1:M} | X) = \prod_{i=1}^I \prod_{j=1}^J \prod_{k=1}^K \prod_{m=1}^M p(r_{ijk,m} | \theta_{ijk}, f_{ijk,m}, c_{ijk,m})$$

$$\times p(\theta_{ijk} | \tau_{ij}) p(\tau_{i,j} | \eta) p(\eta)$$

$$= \prod_{i=1}^I \prod_{j=1}^J \prod_{k=1}^K \prod_{m=1}^M p(r_{ijk,m} | \theta_{ijk}, f_{ijk,m}, c_{ijk,m})$$

$$\times p(\theta_{ijk} | \tau_{ij}) \mathcal{N}(\tau_{ij}, \rho_{ij}, \phi_j) p(\phi_j) p(\rho_{i,j} | \eta)$$

$$\times \mathcal{N}(\eta, \mu, \Sigma) p(\mu) p(\Sigma),$$

where $X = \{\theta_{i,j,k,m} \mid \theta_{i,j,k} \in \{\phi_j, \mu, \Sigma\} \}$ are all the parameters and hyperparameters in the HBM.

We use Bayes' rule to compute the joint posterior distribution of $X$ (Kruschke, 2015; Lee, 2006; Lee, 2011; Rouder & Lu, 2005; Wilson et al., 2020):

$$p(X | Y_{1:J,1:K,1:M})$$

$$= \prod_{i=1}^I \prod_{j=1}^J \prod_{k=1}^K \prod_{m=1}^M p(r_{ijk,m} | \theta_{ijk}, f_{ijk,m}, c_{ijk,m})$$

$$\times p(\theta_{ijk} | \tau_{ij}) \mathcal{N}(\tau_{ij}, \rho_{ij}, \phi_j) p(\phi_j)$$

$$\times \mathcal{N}(\eta, \mu, \Sigma) p(\mu) p(\Sigma)$$

$$= \frac{1}{\int \prod_{i=1}^I \prod_{j=1}^J \prod_{k=1}^K \prod_{m=1}^M p(r_{ijk,m} | \theta_{ijk}, f_{ijk,m}, c_{ijk,m})$$

$$\times p(\theta_{ijk} | \tau_{ij}) \mathcal{N}(\tau_{ij}, \rho_{ij}, \phi_j) p(\phi_j)$$

$$\times \mathcal{N}(\eta, \mu, \Sigma) p(\mu) p(\Sigma)$$

$$\times \text{sech}(\theta_{a, \text{confidence}} \times (\log_{10}(\theta_a) - \log_{10}(\theta_{a, \text{mode}}))) \times (\theta_{a, \text{mode}} = (100, 1, 3)).$$

where $\text{sech}(x) = \frac{2}{e^x + e^{-x}}$, $\theta_a = \gamma_{ijk}^{\text{max}}$, $f_{ijk}^{\text{max}}$, and $\beta_{ijk}$ for $a = 1, 2$, and 3, respectively, $\theta_a, \text{confidence} = (0.5, 0.5, 0.5)$, and $\theta_{a, \text{mode}} = (100, 1, 3)$. 

**Methods**

**Data**

The dataset used in this study included 112 college-aged subjects, each tested once ($K = 1$) in three luminance conditions (low = 2.62 cd/m$^2$, medium = 20.4 cd/m$^2$, and high = 95.4 cd/m$^2$) with the qCSF method (Hou et al., 2016). Each test consisted of 150 trials. Three test trials were presented in each display consisting of three filtered letters of the same size, randomly sampled with replacement from 10 SLOAN letters (C, D, H, K, N, O, R, S, V, and Z), with the center spatial frequency and contrasts of the letters determined by qCSF. Subjects were asked to verbally report the identity of the letters on the screen.

**Apparatus**

All analysis was conducted on a Dell computer with Intel Xeon W-2145 @ 3.70 GHz CPU (8 cores and 16 threads) and 64 GB installed memory (RAM). The BIP was implemented in Matlab R201Xa (MathWorks Corp., Natick, MA, USA) and the HBM was implemented in JAGS (Plummer, 2003) in R (R Core Team, 2020).

**Implementation of the BIP**

Because a 10-alternative forced-choice identification task was used in the experiment, we set $g$ to 0.1, and $\sigma$ to 0.1485 in Equation 2 (Foley & Legge, 1981; Hou et al., 2015; Legge, Kersten, & Burgess, 1987; Lesmes et al., 2010; Lu & Dosher, 1999). Following the qCSF procedure (Hou et al., 2015; Lesmes et al., 2010), we defined a three-dimensional CSF parameter space with 60 log-linearly spaced $\gamma_{ijk}^{\text{max}}$ values between 1.05 and 1050, 40 log-linearly spaced $f_{ijk}^{\text{max}}$ values between 0.1 and 20 cycles/degree, and 27 log-linearly spaced $\beta_{ijk}$ values between 1 and 9 octaves. The weakly informative prior, $p_0(\theta_{ij})$, identical across all the tests, subjects, and experimental conditions, was defined by a hyperbolic secant function (Lesmes et al., 2010):

$$p_0(\theta_{ij1})$$

$$= \sum_{a=1}^3 \text{sech}(\theta_{a, \text{confidence}} \times (\log_{10}(\theta_a) - \log_{10}(\theta_{a, \text{mode}}))) \times (\theta_{a, \text{mode}} = (100, 1, 3)).$$

where $\text{sech}(x) = \frac{2}{e^x + e^{-x}}$, $\theta_a = \gamma_{ijk}^{\text{max}}$, $f_{ijk}^{\text{max}}$, and $\beta_{ijk}$ for $a = 1, 2$, and 3, respectively, $\theta_a, \text{confidence} = (0.5, 0.5, 0.5)$, and $\theta_{a, \text{mode}} = (100, 1, 3)$. 

**References**


The posterior distributions of the CSF parameters $p(\theta_{j|i}|Y_{ij})$ was computed using Equation 4. Convergence of the BIP solutions was quantified by the half-width of 68.2% credible interval (HWCI: Clayton & Hills, 1993; Edwards, Lindman, & Savage, 1963), equivalent to the standard deviation of the distribution if it is normal. With sufficient number of trials in the qCSF, the HWCI can reach its asymptotic minimum (Hou et al., 2015; Lesmes et al., 2010).

### Implementation of the HBM

In the current implementation of the HBM, the prior of $\mu$, $p_0(\mu)$, was a nine-dimensional uniform distribution:

$$p_0(\mu) = U(\log(\mu_{0,\text{min}}), \log(\mu_{0,\text{max}})),$$

with $\mu_{0,\text{min}}$ and $\mu_{0,\text{max}}$ of the three parameters in the three luminance conditions specified in Table 1.

The weakly informative prior distribution of $\Sigma$, $p_0(\Sigma)$, was specified by a $9 \times 9$ precision matrix $\Omega$ with a Wishart distribution:

$$p_0(\Omega) = W\left(\Sigma_{\text{BIP}}^{-1}, \nu, \nu\right),$$

where the degrees of freedom $\nu = 9$, and the expected mean, $\Sigma_{\text{BIP}}^{-1}$, was based on the covariance matrix of the estimated CSF parameters $\Sigma_{\text{BIP}}$ across all the subjects and luminance conditions from the BIP procedure.

The weakly informative prior distribution of $\phi_j$, $p_0(\phi_j)$, was specified with a $3 \times 3$ precision matrix $\Lambda_j$ with a Wishart distribution:

$$p_0(\Lambda_j) = W\left(\phi_{\text{BIP},j}^{-1}, \nu_j, \nu_j\right),$$

where the degrees of freedom $\nu_j = 3$, and the expected mean, $\phi_{\text{BIP},j}^{-1}$, was based on the average covariance matrix $\phi_{\text{BIP},j}$ computed from the estimated CSF parameters across all the subjects in luminance condition $j$ from the BIP procedure.

The R (R Core Team, 2020) function `autorun.jags` in JAGS (Plummer, 2003) was used to compute representative samples of the posterior distributions of $\theta_{j|i}$ (3 parameters/condition $\times$ 3 conditions $\times$ 112 subjects = 1008 parameters), $\rho_{i,1:j}$ (9 parameters $\times$ 112 subjects = 1008 parameters), $\phi_j$ (6 parameters/condition $\times$ 3 conditions = 18 parameters), $\mu$ (9 parameters), and $\Sigma$ (45 parameters) in three Markov Chain Monte Carlo (MCMC) chains. The MCMC is an algorithm used to efficiently sample the joint posterior distribution (Kruschke, 2015). It started at a randomly selected position in the 2088-dimensional parameter space. In each step, one of the 2088 parameters was selected randomly. The one-dimensional conditional posterior probability distribution of the selected parameter was evaluated by fixing the values of all the other 2087 parameters at the current position. A new value of the selected parameter was chosen based on the one-dimensional conditional probability distribution (Equation 8). By reiterating this process, the probability of visiting a location in the random walk approximated the joint posterior distribution of all the 2088 parameters in Equation 8. These steps were re-iterated until the convergence criterion was reached.

### Statistical analysis

#### Goodness of fit

Bayesian predictive information criterion (BPIC; Ando 2007; Ando 2011) was used to quantify the goodness of fit to the trial-by-trial data. The BPIC quantifies the likelihood of the data based on the joint posterior distribution of the parameters of the model and penalizes model complexity.

#### Posterior distributions of the area under log CSF

The posterior distributions of AULCSF were constructed by computing the AULCSFs from samples of the corresponding posterior distributions of $\theta_{j|i}$ from the HBM and BIP.

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>M</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma_{\text{max}}$</td>
<td>1.05</td>
<td>1.05</td>
<td>1.05</td>
</tr>
<tr>
<td>$f_{\text{max}}$</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>$\beta$</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 1. $\mu_{0,\text{min}}$ and $\mu_{0,\text{max}}$ of the uniform prior of $\mu$. H, high; L, low; M, medium.
**d : Between-condition discriminability of distributions**

Discriminability $d'$ quantifies the signal (mean separation) to noise (variability) ratio of two probability distributions. We used the difference distribution between conditions to compute $d'$. Each sample in the difference distribution represented the difference between two randomly drawn samples from the corresponding distributions.

For a one-dimension difference distribution, $d'$ is defined as (Green & Swets, 1966):

$$d' = \sqrt{2 \frac{\Delta}{\sigma}}, \quad (11)$$

where $\Delta$ is the mean separation, and $\sigma$ is the standard deviation of the difference distribution.

For a multidimensional difference distribution, $d'$ is defined as (Ashby & Townsend, 1986):

$$d' = \sqrt{2 \Delta \cdot \text{cov}((\Delta)^{-1} \cdot \Delta^T)}, \quad (12)$$

where $\Delta$ and $\text{cov}(\Delta)$ are the mean separation and covariance matrix of the difference distribution, $\text{cov}((\Delta)^{-1})$ is the inverse of $\text{cov}(\Delta)$, $\Delta^T$ is the transpose of $\Delta$, and $\ast$ represents matrix multiplication.

**Statistical tests on CSF parameters and AULCSF**

We compared the mean (expected value) and variance of the posterior distributions of $\theta_{ij1}$ from the HBM and BIP using Hotelling’s T-squared test (Anderson, 2003) in R (R Core Team, 2020; Nordhausen, Sirkiä, Oja, & Tyler, 2018). We also compared the correlation coefficients of pairs of CSF parameters from the two methods with paired $t$-test.

To quantify the between-condition discriminability across subjects, we compared the means of the posterior distributions of $\theta_{ij1}$ and AULCSF between pairs of experimental conditions from each method with Hotelling’s T-squared test and paired $t$-test, respectively.

**Simulation**

To compare the accuracy and precision of the BIP and HBM estimates, we conducted a simulation study to investigate the bias and variability of the estimated CSF parameters for each test. The dataset consisted of 336 qCSF tests (112 subjects $\times$ 3 conditions). $\theta_{1:1:1:1:1}$ of the simulated tests were a random sample from the posterior distribution of $\tau_{1:1:1:1:1}$ obtained from the HBM fit to the real data. Each qCSF test consisted of 150 trials, identical to the real experiment (Hou et al., 2016), with the trial-by-trial responses determined by the CSF parameters of the simulated subject (Equations 1 to 3). Both the HBM and BIP were fit to the simulated dataset. The mean of the posterior distribution of $\theta_{ij1}$ was used as the best estimate for each test. The bias, root mean square error (RMSE), variance, $d'$, and $t$ statistics were computed based on the posterior distributions of $\theta_{ij1}$ from both methods.

**Results**

**Goodness of fit**

The BPIC for the BIP and HBM were 34886 and 34225, respectively, indicating that the HBM fit the data better than the BIP. Figure 3 shows the estimated CSFs of one subject from the BIP and HBM in three luminance conditions.

Many “image-computable” models have used the CSF as the front-end filter on actual images to predict human performance in image processing and object recognition (Chung, Legge, & Tjan, 2002; Malo, Pons, Felipe, & Artigas, 1997; Schütt, & Wichmann, 2017; Watson, 2000; Watson & Ahumada Jr, 2005; Watson, & Malo, 2002). Although recent studies have suggested that a more comprehensive model may require additional parameters related to non-linearities in the visual system (Chen, Hou, Yan, Zhang, Xi, Zhou, Lu, & Huang, 2014; Hou, Lu, & Huang, 2014), the CSF filtered images nevertheless provide an excellent demonstration of human visual processing. To illustrate the differences between the CSF estimates from the BIP and HBM and their implications for image-computable models (Figure 4), we applied the mean - SD, mean, and mean + SD CSFs from the BIP and HBM in the high luminance condition to filter a letter K (Lu & Dosher, 2013). Although the mean CSFs from the two methods are very similar and generated very similar filtered K’s, the CSFs from the BIP exhibited much larger uncertainty.

**Posterior distributions from the HBM**

Figure 5 shows the three-dimensional posterior distributions of hyperparameters $\eta$ (marginalized), $\tau_{i:1:1:1}$ for one individual, and $\theta_{i:1:1:1:1}$ for one individual in one test from the HBM.

**Population level**

Table 2 shows the mean and covariance matrix of $\eta$. The correlation coefficients were positive and significant between all three pairs of experimental conditions. Table 3 shows the $d$’s of $\eta$ between the three pairs of experimental conditions. In the HBM, the posterior distributions of $\eta$ constrained $\tau_{1:1:1}$. The large $d$’s of the posterior distributions of $\eta$ between
different experimental conditions indicated that the posterior distributions of $\eta$ provided strong constraints on $\tau_{1,1:J}$.

**Individual level**

Table 4 shows the average covariance matrix of $\tau_{1,1:J}$ across all 112 individuals and experimental conditions. Figure 5(b) illustrates the three-dimensional posterior distributions of $\tau_{1,1:J}$ for one individual in all three luminance conditions. Table 3 shows the average $d'$s of $\tau_{ij}$ between the three pairs of experimental conditions. In the HBM, the posterior distributions of $\tau_{1,1:J}$ constrained $\theta_{1,1:J}$. The large $d'$s of the posterior distributions of $\tau_{1,1:J}$ between different experimental conditions indicated that the posterior distributions of $\tau_{1,1:J}$ provided strong constraints on $\theta_{1,1:J}$.

**Test level**

We computed the mean, covariance, and correlation coefficient based on the estimated test-level CSF parameters $\theta_{1,1:J}$ in the three luminance conditions from the HBM and compared them with the results from the BIP.

The means of the posterior distributions of $\theta_{ij}$ from the HBM and BIP were significantly different ($t^2(9,103) = 5.34, p < 0.001$), and the average variance of the estimated CSF parameters from the HBM (mean = 0.00139 log10 units; range = 0.00030 to 0.00739 log10 units) was 65.8% less than that from the BIP (mean = 0.00407 log10 units; range = 0.00035 to 0.11893 log10 units) ($t^2(9,103) = 109, p < 0.001$), consistent with the well-known variance shrinkage effect of the HBM (Kruschke, 2015).

Figures 6 and 7 show histograms of the difference between the expected values of $\theta_{ij}$ from the BIP and HBM. Whereas most of the differences between the expected values of $\theta_{ij}$ from the two methods were small (mean absolute difference = 0.027 log10 units), there were thirteen instances (out of a total of 3 parameters $\times$ 3 conditions $\times$ 112 subjects = 1008) in which the absolute difference was greater than 0.2 log10 units (range = 0.200 to 0.676 log10 units). The discrepancies were associated with large variances of the BIP estimates in those instances: their average variance of 0.065 log10 units (range = 0.027 to 0.119 log10 units).
Figure 5. Three-dimensional posterior distributions of $\eta$ (marginalized) (a), $\tau_{i,j}$ for one individual (b), and $\theta_{i,j}$ for one individual in one test in the HBM. The colors represent log10 probability density.

<table>
<thead>
<tr>
<th>Covariance</th>
<th>Parameter</th>
<th>L</th>
<th>M</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\eta^1$</td>
<td>6.2E-03</td>
<td>1.4E-03</td>
<td>3.4E-05</td>
<td>4.7E-03</td>
</tr>
<tr>
<td>$\eta^2$</td>
<td>1.1E-03</td>
<td>9.5E-03</td>
<td>-1.9E-03</td>
<td>3.2E-03</td>
</tr>
<tr>
<td>$\eta^3$</td>
<td>4.4E-05</td>
<td>-1.9E-03</td>
<td>1.1E-03</td>
<td>3.0E-05</td>
</tr>
<tr>
<td>$\eta^4$</td>
<td>4.7E-03</td>
<td>3.2E-03</td>
<td>3.0E-05</td>
<td>6.9E-03</td>
</tr>
<tr>
<td>$\eta^5$</td>
<td>1.9E-03</td>
<td>4.5E-03</td>
<td>-6.9E-04</td>
<td>1.7E-03</td>
</tr>
<tr>
<td>$\eta^6$</td>
<td>-4.7E-04</td>
<td>-5.3E-04</td>
<td>5.6E-04</td>
<td>-1.1E-04</td>
</tr>
<tr>
<td>$\eta^7$</td>
<td>4.2E-03</td>
<td>2.6E-03</td>
<td>-1.2E-04</td>
<td>5.4E-03</td>
</tr>
<tr>
<td>$\eta^8$</td>
<td>-2.3E-04</td>
<td>1.8E-03</td>
<td>-5.4E-04</td>
<td>-1.5E-04</td>
</tr>
<tr>
<td>$\eta^9$</td>
<td>8.4E-04</td>
<td>1.0E-03</td>
<td>5.3E-04</td>
<td>1.2E-03</td>
</tr>
</tbody>
</table>

Table 2. Mean and covariance of $\eta$. H, high; L, low; M, medium.

was 16 times the mean variance (0.00407 log10 units) of $\theta_{j1}$ in the BIP procedure, suggesting that BIP did not converge well in those cases. On the other hand, the HBM generated more precise estimates with on average a 93.7% reduction of variance (mean = 0.00407 log10 units; range = 0.00139 to approximately 0.00681 log10 units) compared to the BIP in the 13 cases by incorporating data from all the subjects and conditions in a single model.

Table 5 lists the average correlation coefficients between $\theta_{j1}$ in pairs of luminance conditions across subjects from the HBM and BIP procedures. All correlations were negative, with the strongest between $f^\text{max}$ and $\beta$. Across all the subjects, 97.4% and 97.9% of the correlation coefficients from the BIP and HBM were statistically significant, respectively. Although the paired t-test showed that the correlation coefficients between $\gamma^\text{max}$ and $\beta$ in the high luminance condition ($p = 0.003$) and between $f^\text{max}$ and $\beta$ in all three luminance conditions ($p < 0.001$) from the two procedures were significantly different, the magnitudes of the differences were very small and probably not of practical importance.

Table 6 shows the average $d'$s of $\theta_{j1}$ and AULCSF between pairs of luminance conditions across all the subjects. Averaged across the three pairs, the AULCSF $d'$ from the HBM was 33.5% greater than that from the BIP. Compared to AULCSF, incorporating information from the three-dimensional joint distributions of $\theta_{j1}$ led to an average $d'$ increase of 66.6% for the BIP and 51.7% for the HBM. Compared to the AULCSF $d'$ from the BIP, using $\theta_{j1}$ in the HBM increased $d'$ by 103.3% across the three pairs of luminance conditions.

Statistics on $\theta_{j1}$ and AULCSF across individuals

Table 7 shows $\sqrt{\chi^2(3, 109)}$ and $t(111)$ of the means of $\theta_{j1}$ and AULCSF among the three pairs of experimental conditions. The HBM generated larger $t$ values than the BIP for both $\theta_{j1}$ and AULCSF in all pairs of experimental conditions. Averaged across the
Table 4. Average covariance matrix of $\tau_{i,j}$, H, high; L, low; M, medium.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\tau_{1,i,j}$</th>
<th>$\tau_{2,i,j}$</th>
<th>$\tau_{3,i,j}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau_{1,i,j}$</td>
<td>2.6E-03</td>
<td>-9.8E-04</td>
<td>-1.0E-06</td>
</tr>
<tr>
<td>$\tau_{2,i,j}$</td>
<td>-9.8E-04</td>
<td>4.6E-03</td>
<td>-1.8E-03</td>
</tr>
<tr>
<td>$\tau_{3,i,j}$</td>
<td>-1.0E-06</td>
<td>-1.8E-03</td>
<td>8.7E-04</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\tau_{1,i,j}$</th>
<th>$\tau_{2,i,j}$</th>
<th>$\tau_{3,i,j}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau_{1,i,j}$</td>
<td>2.2E-03</td>
<td>-8.0E-04</td>
<td>-3.2E-05</td>
</tr>
<tr>
<td>$\tau_{2,i,j}$</td>
<td>-8.0E-04</td>
<td>3.6E-03</td>
<td>-1.4E-03</td>
</tr>
<tr>
<td>$\tau_{3,i,j}$</td>
<td>-3.2E-05</td>
<td>-1.4E-03</td>
<td>7.2E-04</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\tau_{1,i,j}$</th>
<th>$\tau_{2,i,j}$</th>
<th>$\tau_{3,i,j}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau_{1,i,j}$</td>
<td>3.4E-03</td>
<td>-1.9E-03</td>
<td>5.5E-05</td>
</tr>
<tr>
<td>$\tau_{2,i,j}$</td>
<td>-1.9E-03</td>
<td>8.0E-03</td>
<td>-2.9E-03</td>
</tr>
<tr>
<td>$\tau_{3,i,j}$</td>
<td>5.5E-05</td>
<td>-2.9E-03</td>
<td>1.4E-03</td>
</tr>
</tbody>
</table>

Discussion

The HBM provides a general framework that can be adapted to different experiment designs. In this paper, we developed a three-level HBM to account for CSF data of 112 subjects in a single-factor (luminance), multi-condition (3 luminance conditions), and within-subject experimental design. We applied the HBM to quantify the joint distribution of CSF parameters and hyperparameters at the population, individual, and test levels and compared the performance of the model with that of the BIP. The HBM generated more precise estimates for each test than the BIP by incorporating information across subjects and conditions to constrain the estimates. The increased precision led to increased
Relative to the BIP, the HBM increased the average subject, and bigger statistical differences across subjects.

Experimental conditions at the test level for each

Table 7. \(d\)'s of AULCSF and CSF parameters between different experimental conditions at the test level for each subject, and bigger statistical differences across subjects. Relative to the BIP, the HBM increased the average \(d\)'s of AULCSF and \(\theta_{ij}\) between conditions at the test level by 24.5% and 20.5%, and the corresponding \(t\)(111) and \(\sqrt{1/2(3,109)}\) by 51.2% and 49.6%, respectively. Simulations also showed that the HBM generated accurate and precise CSF parameter estimates.

The HBM generated larger \(d\) and \(t\) statistics at the test level because it reduced the variance of \(\theta_{ij}\) by 65.8% relative to the BIP (0.00139 vs. 0.00407 log10 units). In addition, the 13 instances in which the absolute difference of \(\theta_{ij}\) from the HBM and BIP were greater than 0.2 log10 units further demonstrated the benefit of incorporating information across tests, subjects, and conditions in the HBM (Kruschke, 2015; Rouder & Lu, 2005; Rouder, Sun, Speckman, Lu, & Zhou, 2003). In those cases, the variances of the BIP estimates were very large (16 times the mean variance), suggesting that the BIP did not converge well. On the other hand, the HBM generated much more precise CSF estimates for each test by incorporating data across subjects and conditions in a single model. The ability of the HBM to generate more precise estimates from insufficient or poor-quality data can be quite valuable in clinical trials.

The HBM can be used to conduct two types of power analyses. First, a replication power analysis computes the power of different sample sizes in replicated experiments with the exact same experimental design (Kruschke, 2015). Simulated data for new subjects can be generated from the posterior distributions of the hyperparameters based on the HBM fit to the existing dataset, just as we did in the simulation. In this case, the original data of 112 subjects shall be combined with the simulated data to compute the power for each new sample size. A more interesting application of the HBM is in prospective power analysis (Kruschke, 2015). In that case, no data have been collected for a new experiment; simulated data of the new experiment must be based on the generative model constructed based on results from a different experiment. An HBM for the new experimental design is then constructed and fit to the simulated data. Therefore, data from existing studies can only be used as prior knowledge and cannot be combined with the simulated data.

A certain sample size is required for the joint posterior distribution of the CSF hyperparameters and parameters in the HBM to become stable. This can be done by evaluating the stability of the estimates with different sample sizes. Relative to the noninformative diffuse prior used in the BIP; Gu, Kim, Hou, Lesmes, Pitt, Lu, and Myung (2016) showed that (1) an informative prior from the HBM fit to as few as five subjects could provide significant improvement in qCSF measurements of new subjects in the hierarchical adaptive design optimization (HADO) procedure, (2) prior constructed from larger samples further improved the accuracy and precision of the estimation, and (3) the improvement stabilized when the sample size was about 30. Simulation studies are necessary to determine the minimum sample size required for the HBM to converge for a given experiment.

Although the MCMC algorithm automatically selected by the JAGS (Plummer, 2003) provided an efficient sampling method, the weakly informative priors of the covariance matrices at both the subject and population levels were very helpful. With these priors, it took about 54 hours for each MCMC chain to generate at least 10,000 effective samples (Kruschke, 2015) for all parameters on a computer with eight cores. Fifty-four hours are practical given that the actual physical time decreases with increasing number of CPUs used in parallel computation (Kruschke, 2015). On the other hand, the HBM using diagonal covariance...
matrices as the priors took 18% longer (63.5 hours) to generate 10,000 effective samples for all parameters, and did not converge, as indicated by the larger variances of the estimated CSF parameters at both the subject (0.00298 log10 units and 114% increase) and population levels (0.212 log10 units and 2873% increase), although the Gelman and Rubin’s diagnostic was below 1.05 for all parameters, which is based on the ratio of between-chain and within-chain variances but not the magnitudes of the variances. The effects of priors on covariance estimation were consistent with previous studies (Hobert & Casella, 1996; Rouder et al., 2003).

Although the HBM in the current study was developed to account for group differences in a within-subject design, HBM-based approaches can be developed to detect deviations in individual patients belonging to different subpopulations (e.g. healthy versus different stages of an eye disease, or different eye diseases) using different tests (null hypothesis significance testing versus estimation of magnitude/effect size) in both frequentist and Bayesian approaches (Kruschke & Liddell, 2018). The HADO method (Gu, Kim, Hou, Lesmes, Pitt, Lu, & Myung, 2016; Kim, Pitt, Lu, Steyvers, & Myung, 2014) provides a potential framework. HADO uses an informative prior obtained from all previously tested subjects with the same CSF characteristics (e.g. testing in the same luminance condition). It took 20 to 30 trials for the qCSF method with an uninformative diffuse prior to achieve the same initial precision level of the HADO procedure with the informative prior. Moreover, HADO with a mixture prior that represent a wide range of CSF properties (e.g. across different luminance conditions) still achieved higher precision than qCSF with an uninformative prior, and could mitigate the problem of mis-specified prior and improve the qCSF method on testing individual subjects. The joint posterior distributions of the hyperparameters at the population and individual levels from HBM can provide informative priors within the HADO framework for new individuals and repeated tests of the same individual, respectively. Furthermore, the HBM can be extended to model additional covariance between parameters of different measurements (e.g. CSF and visual acuity) in a joint modeling approach (Palestro, Bahg, Sederberg, Lu, Steyvers, & Turner, 2018; Turner, Forstmann, Wagenmakers, Brown, Sederberg, & Steyvers, 2013) to account for multiple test results of multiple subjects and conditions, and potentially further increase statistical power in detecting changes of functional vision in normal and clinical populations. Therefore, the HBM framework can be used to take advantage of all available information at different levels to enable sensitive detection of CSF changes, and thereby improve patient care and clinical trials with increased statistical power.

Conclusions

In this paper, we developed a three-level HBM to account for CSF data of 112 subjects in a within-subject, single-factor (luminance), multi-condition (3 luminance conditions) experimental design. The HBM was used to compute the joint distribution of CSF parameters and hyperparameters at population, individual, and test levels to fully utilize information across levels to accurately estimate the CSF in each test. Relative to the BIP, the HBM increased the average $d'$s of AULCSF and $\theta_{ij}$ between conditions at the test level by 24.5% and 20.5 %, and the corresponding $t$ statistics by 51.2% and 49.6%, respectively. Future research will further evaluate the potential value of the HBM for analyzing clinical changes in contrast sensitivity, whether in individual patients or groups in clinical trials.

Keywords: contrast sensitivity function, quick contrast sensitivity function (qCSF), hierarchical Bayesian model (HBM), $d'$

Acknowledgments

Supported by the National Eye Institute (EY021553 and EY017491 to Z.L.), and by the Qianjiang Talent Project of Zhejiang Province (QJD1803028 to H.F.).

Commercial relationships: L.A.L. and Z.L.L. have intellectual property interests in methods for measuring and applying contrast sensitivity functions (US 7938558, WO2013170091, and PCT/US2015/028657), and equity interest in Adaptive Sensory Technology, Inc. (San Diego, CA). In addition, L.A.L. has an intellectual property interest in methods for quantitative visual acuity testing (US 10758120B2) and holds employment in AST.

Corresponding author: Zhong-Lin Lu.
Email: zhonglin@nyu.edu.
Address: Center for Neural Science and Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, USA.

Footnote

1 Hou et al., (2016) used a four-parameter CSF model with a low frequency truncation level $\delta$ as the fourth parameter (Lesmes et al., 2010; Watson & Ahumada, 2005). Because of the range of low-frequencies that were tested in this dataset, we used the three-parameter CSF model without $\delta$ because it was not very well constrained. The three-parameter log-parabola CSF also reduced the total number of parameters and the complexity of the HBM.
References


