Available online at www.sciencedirect.com
ScienceDirect
Vision Research xxx (2007) xxx–xxx
Vision Research
www.elsevier.com/locate/visres

Inverse modeling of human contrast response

Mikhail Katkov, Misha Tsodyks, Dov Sagi *

Department of Neurobiology/Brain Research, The Weizmann Institute of Science, Rehovot 76100, Israel

Received 3 November 2005; received in revised form 11 June 2007

Abstract

Mathematical singularities found in the Signal Detection Theory (SDT) based analysis of the 2-Alternative-Forced-Choice (2AFC) method [Katkov, M., Tsodyks, M., & Sagi, D. (2006a). Analysis of two-alternative force-choice Signal Detection Theory model. *Journal of Mathematical Psychology*, 50, 411–420; Katkov, M., Tsodyks, M., & Sagi, D. (2006b). Singularities in the inverse modeling of 2AFC contrast discrimination data. *Vision Research*, 46, 256–266; Katkov, M., Tsodyks, M., & Sagi, D. (2007). Singularities explained: Response to Klein. *Vision Research*, doi:10.1016/j.visres.2006.10.030] imply that contrast discrimination data obtained with the 2AFC method cannot always be used to reliably estimate the parameters of the underlying model (internal response and noise functions) with a reasonable number of trials. Here we bypass this problem with the Identification Task (IT) where observers identify one of N contrasts. We have found that identification data varies significantly between experimental sessions. Stable estimates using individual session data showed Contrast Response Functions (CRF) with high gain in the low contrast regime and low gain in the high contrast regime. Noise Amplitudes (NA) followed a decreasing function of contrast at low contrast levels, and were practically constant above some contrast level. The transition between these two regimes corresponded approximately to the position of the dipper in the Threshold versus Contrast (TvC) curves that were computed using the estimated parameters and independently measured using 2AFC.

Keywords: Psychophysics; Signal Detection Theory; Identification; Categorization; Noise; Contrast response function; Criteria variability

1. Introduction

Signal Detection Theory (SDT) is a common framework used to model performances in various psychophysical tasks. It assumes that a task-related attribute of a stimulus is encoded in the sensory system, and that the encoded information is used by a decision mechanism to choose an action to perform (Green & Swets, 1966). The encoding of a sensory stimulus is not perfect: it varies between trials due to noise, leading to confusions between stimuli with similar task-related attributes. SDT provides the analytical tools required to quantify these confusions and to analyze human performance. A long-established goal in visual sciences is to model sensory processes that can account for human performance; however, a commonly accepted model of contrast transduction in humans has not yet been established. Several groups suggested different (sometimes contradictory) models that describe the data reasonably well (Chirimuuta & Tolhurst, 2005; Foley & Legge, 1981; García-Pérez & Alcalá-Quintana, 2007; Georgeson & Meese, 2006; Gorea & Sagi, 2001; Klein, 2006; Kontsevich, Chen, & Tyler, 2002; Lu & Dosher, 1999).

In a previous work we used SDT to analyze the 2AFC contrast discrimination task, and we found that a large set of different models can account for our data (Katkov, Tsodyks, & Sagi, 2006b). In simulating discrimination performances for a constant noise model, we found that, even for unrealistically long experiments, simulated performances can be explained by very different models with monotonic noise functions that span a large (almost arbitrary) range of noise amplitudes (Katkov, Tsodyks, & Sagi, 2007). Alternative models were of a specific form related to fundamental singularities in the discrimination models (Katkov, Tsodyks, & Sagi, 2006a). Therefore, there is no

* Corresponding author. Fax: +972 8 934 4131.
E-mail address: Dov.Sagi@Weizmann.ac.il (D. Sagi).

0042-6989/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2007.06.024

practical way to "prove" constant noise model in 2AFC experiments—there are always alternative models that explain the data as well (Katkov et al., 2007). On the other hand, if simulated models were different than singular models specified in Katkov et al. (2006a) the parameters of the model could be reliably recovered. Nevertheless, psychophysical data from experiments carried out by several groups were found to be compatible with the constant noise model (Foley & Legge, 1981; Georgeson & Meese, 2006; Gorea & Sagi, 2001; Katkov et al., 2006b), and therefore, a different technique is required to obtain a unique specification of the SDT model parameters (the "true" model that describes the transducer function and noise is possibly close to a singularity; see more in Section 4 and in Katkov et al., 2007). In Katkov et al. (2006a) we suggested using experimental methods where only one stimulus is presented during a trial. Therefore, in this work we employed an Identification Task (IT, McNicol, 2005) to estimate the parameters of the sensory model, using SDT. This method is often used to estimate contrast detection thresholds and is considered to be advantageous over forced-choice methods (Klein, 2001). We assumed that the sensory model is common for all tasks and that the outcome for different tasks, such as rating or 2AFC, depends only on the decision stage. To test this hypothesis we compared experimental results measured in 2AFC experiments with the results obtained from the IT experiments.

The aim of this work was to obtain a data driven "theory-free" model for the sensory stage that underlies contrast perception, assuming only SDT with no assumptions made on the contrast response function or on the noise dependency. This approach limits the set of possible models/mechanisms for contrast transduction, and being data driven, permits models that can be speculated to be "unrealistic". More specifically, we are trying to establish how the parameters R_i, σ_i in Eqs. (1) and (3) (internal response and noise, see below) vary with stimulus contrast for the best fitting model. This problem is especially important when psychophysical data are related to physiological studies. For example, in Boynton, Demb, Glover, and Heeger (1999), an assumption of constant (additive) noise was used to compare fMRI data with psychophysical performance in contrast discrimination. Nevertheless, it is well-known that in single-cell recordings the variance of the response is proportional to its mean value (Softky & Koch, 1993). Moreover, in a previous study we found that there are models with additive Gaussian noise that result in discrimination performances which are similar to those predicted by models assuming Poisson noise (Katkov et al., 2006b).

1.1. Identification task

In an IT trial, one of N possible stimuli (contrasts) is presented and the observer responds with one of M possible responses (category, 1 to M). According to SDT, the contrast level of the stimulus is transformed into a decision variable (internal response). Here we assume that for each stimulus contrast, s_i, a decision variable, r_i, has a Gaussian distribution across trials and thus, can be described by two numbers: the mean internal response, R_i, and the trial-by-trial standard deviation, σ_i (referred to as noise). At the decision stage, the observer establishes a criterion for each category boundary. The observer reports that the presented stimulus belongs to some category when the internal response falls between the corresponding category boundaries. In the model we can compute the probability, $P_{i,n}$, of the internal response to the stimulus contrast, s_i, to be greater than a category boundary, k_n, which corresponds to the probability that an observer reports a category higher than n. Assuming criteria that are stable across trials, we can write:

$$P_{i,n} = \Phi\left(\frac{R_i - k_n}{\sigma_i}\right), \tag{1}$$

where

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-t^2/2}\,\mathrm{d}t. \tag{2}$$

If the criteria are not stable but have variability that is the same for every category boundary, the criteria variability can be added to the internal response variabilities without affecting the form of the final mathematical expression. Although there is no formal way to separate sensory noise from decision noise (criteria variability) within IT, a comparison between IT-derived models and 2AFC-derived models can support such a separation since according to SDT, no absolute criteria are involved in the 2AFC task.

By increasing the number of stimuli and the number of categories, it is possible to define an over-complete system of equations of type (1), and thus to estimate the parameters of the model (R_i, σ_i, and k_n) from the measured P. For example, in the case where the number of categories equals the number of stimuli, the number of independent measurements grows as the square of the number of stimuli. The number of parameters grows linearly—each additional stimulus adds two parameters (R_i, σ_i) and a category boundary (k_n). This model is equivalent to the Thurstonian scaling method of successive intervals (McNicol, 2005). Moreover, when two stimuli and several response categories are used, this method is reduced to the well-known rating procedure used for rapid estimation of Receiver Operating Characteristic (ROC) curves. In the present study the number of response categories was equal to the number of stimuli (5–10).

1.2. Two-alternative force-choice task

Two stimuli of different strengths are presented during a single 2AFC trial. The observer reports the stronger stimulus out of the two (the target stimulus). According to SDT, each stimulus evokes an internal response and the observer reports as a target the stimulus with the higher internal response. Using the same notation as in IT, the following

Please cite this article in press as: Katkov, M. et al., Inverse modeling of human contrast response, *Vision Research* (2007), doi:10.1016/j.visres.2007.06.024

expression describes the probability of the observer reporting that stimulus "j" has greater intensity than stimulus "i" (Thurstone, 1927)

$$P_{i,j} = \Phi\left(\frac{R_j - R_i}{\sqrt{\sigma_i^2 + \sigma_j^2}}\right). \quad (3)$$

1.3. Finger errors

Usually, there is a finite probability that the observer mistakenly presses a wrong key when reporting, or she/he did not attend the task at a particular trial. These response errors, as all responses that do not depend on the stimulus, can be added to the stimulus-related errors, yielding a performance measure that obeys the following mathematical expressions:

$$\mathrm{d}\tilde{P}_{i,n} = p_{\mathrm{fe}}\frac{1}{N} + (1 - p_{\mathrm{fe}})\mathrm{d}P_{i,n}, \quad (4)$$

for the IT, and

$$\tilde{P}_{i,j} = p_{\mathrm{fe}}\frac{1}{2} + (1 - p_{\mathrm{fe}})P_{i,j}, \quad (5)$$

for 2AFC. Here, p_{fe} represents the probability of a stimulus-unrelated response (finger error, unattended response, etc.), $\mathrm{d}P_{i,n} = P_{i,n+1} - P_{i,n}$ is the probability that i's stimulus is classified as n, and N is the number of categories in IT.

2. Methods

2.1. Experiment

Six observers with normal or corrected-to-normal vision participated in the experiments. Several experimental paradigms described next were used in the work. In all the paradigms the visual stimuli were presented on a Mitsubishi Diamond Pro 2060u computer monitor. They consisted of a single Gabor patch (GP, a product of a cosine gratings and Gaussian envelope) in the middle of the screen. In the 2AFC experiments four crosses (a cue) were presented at approximately 6° eccentricity during stimulus presentation to reduce the temporal uncertainty. The spatial frequency of the GP carrier grating was 4 cpd and the width of the envelope was the same as the period of the grating ($\sigma = \lambda = 0.25°$). The contrast is defined here as the Michelson contrast of the carrier grating. The rest of the screen was uniformly gray with a mean luminance of 41 cd/m^2. Several half-hour sessions were performed for each experimental condition. No more than two sessions were performed in a single day with a break of at least 30 min between sessions. First sessions in the IT experiments were discarded from the analysis. In the following sections the details of the experimental procedures for each paradigm will be described.

2.1.1. Identification task

Fixation point (200 ms), empty screen (300 ms), and stimulus (80 ms) were presented sequentially to the observer. This sequence was followed by an empty screen. Numeric keys on the computer keyboard were used to record the responses. Several observers participated in more than one experiment. The number of stimuli with different contrasts was different in different experiments and was in the range of 5–10 contrast levels. The number of response categories was always equal to the number of different contrast levels. The contrasts were selected individually for each observer to maximize observer response variability. This was done by first obtaining a TvC curve, which was used to select contrast levels c_i and c_{i+2},

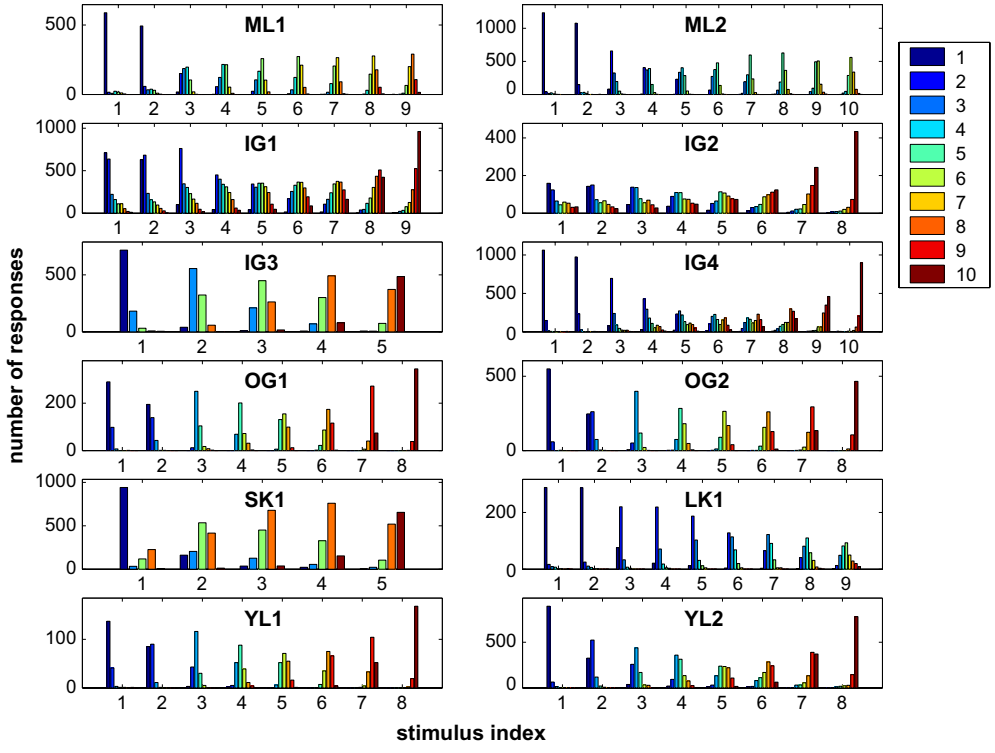

Fig. 1. Confusion matrix. Each plot depicts histograms of observer responses (which represent the confusion matrix) for a single experiment marked by the observer's initials and a number indicating the experimental condition. The stimulus index is presented below the histogram. The response category is represented by the color code. Experiments are encoded by two letters and a number.

Please cite this article in press as: Katkov, M. et al., Inverse modeling of human contrast response, *Vision Research* (2007), doi:10.1016/j.visres.2007.06.024

separated by a predefined threshold performance (80% correct discriminations). In several experiments auditory feedback for incorrect identifications was provided—lower tones for responses below the correct one and higher tones for responses above the correct one. No substantial differences in terms of transducer function and noise amplitudes were found in the results.

2.1.2. Two-alternative force-choice

Fixation point (200 ms), empty screen (300 ms), first stimulus (80 ms), empty screen (800 ms), and second stimulus (80 ms) were sequentially presented to the observer. Then, an empty gray screen was presented until the response from the observer was obtained. The observer was asked to report in which frame, the first or the second, the GP had a higher contrast. Pairs of contrasts were chosen from the same set of contrasts used in the IT (c_i, c_j, presented in random order). Pairs with large differences in contrast levels were not used in this experiment, since they are perfectly discriminated and do not provide any additional information. Auditory feedback was provided to indicate incorrect responses.

2.1.3. Threshold versus contrast

A 2AFC adaptive staircase method was used to construct the TvC curves. The contrast level for one of the stimuli was fixed in a given condition (base contrast) and the contrast level for the second stimulus was varied. After three consecutive correct discriminations for a given base contrast, the difference in contrasts between the two stimuli (contrast increment) was decreased by 0.1 log units, and after a single wrong discrimination, it was increased by 0.1 log units. The initial increment was chosen such that it is easy to discriminate between two stimuli. A geometric mean of six last reversals (first two reversals were ignored) of the staircase procedure averaged across 4–6 measurements was used to estimate the discrimination threshold. This procedure converges to approximately 80% ($\sqrt[3]{1/2} = 0.7937\ldots$) correct responses. Staircases corresponding to different base contrasts were randomly interleaved in a block of trials.

2.2. Estimation of parameters

We searched for model parameters (R_i, σ_i, k_n, p_{fe}) that minimize χ^2. This minimization process, using the model described by Eq. (1) and the confusion matrices we measured, seems to produce many local minima. Since there is no numerical method that guarantees a global minima, there is a possibility that the models we present are not optimal. To decrease this possibility, we performed χ^2 fitting (Press, Flannery, Teukolsky, & Vetterling, 2005), starting from different initial conditions, including one found using extensive search by genetic algorithms (Goldberg, 1989). In the actual implementation we were using a MATLAB (trademark of The MathWorks, Inc.) implementation of an optimization and a genetic algorithm search (functions "lsqnonlin" and "ga", respectively).

The χ^2 value for IT was computed as (Howell, 2002)

$$\chi^2 = \sum_{i,n} \frac{\left(\mathrm{d}\tilde{P}_{i,n} - \mathrm{d}\hat{P}_{i,n}\right)^2}{\mathrm{d}\tilde{P}_{i,n}} N_i, \tag{6}$$

where $\mathrm{d}\tilde{P}_{i,n} = \tilde{P}_{i,n} - \tilde{P}_{i,n-1}$; $\tilde{P}_{i,n}$ is predicted by a model's probability of a response greater than category "n" for stimulus "i" (Eq. (4)), $\mathrm{d}\hat{P}_{i,n}$ is the measured frequency of responses "n" for stimulus "i", and N_i is the number of trials for the stimulus "i". The symbol $\mathrm{d}\tilde{P}_{i,n}$ represents the predicted entry into the corresponding confusion matrix cell, and it is a counterpart of the measured $\mathrm{d}\hat{P}_{i,n}$; $k_0 = -\infty$. The number of degrees of freedom ($df = N_s N_r - N_s - (2N_s - 2 + N_r - 1 + 1)$) is the number of cells in the confusion matrix ($N_s N_r$) minus the number of stimuli (N_s) minus the number of parameters − two for each stimulus (except normalization, $2N_s - 2$) plus the number of categories minus one (the number of category

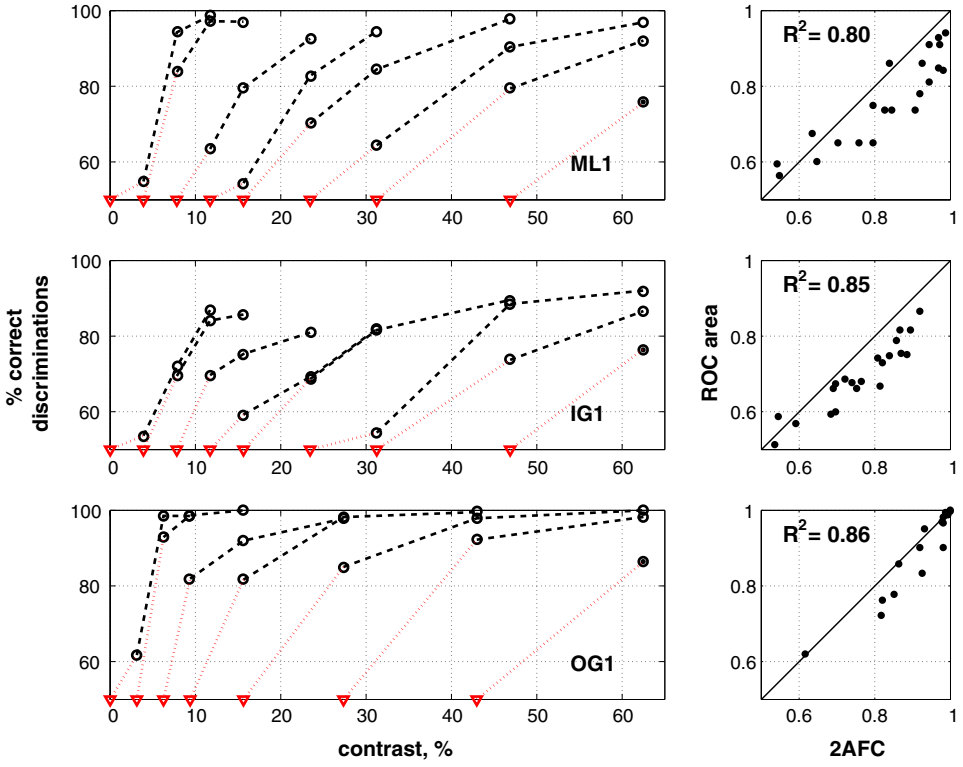

Fig. 2. 2AFC data. Each row corresponds to a singe observer. The left column represents psychometric curves measured in 2AFC experiments. Triangles represent base contrasts, and circles denote the percentage of correct discriminations between stimuli with contrast levels represented by a circle and linked triangle. The right column shows a comparison between performances measured in 2AFC (left column) and an estimation of the area under the ROC curves (ROC area) defined by pairs of lines in the confusion matrix for the corresponding stimuli (see details in the text).

Please cite this article in press as: Katkov, M. et al., Inverse modeling of human contrast response, *Vision Research* (2007), doi:10.1016/j.visres.2007.06.024

boundaries, $N_r - 1$), plus one parameter for finger errors. In total, it is the number of cells in the confusion matrix minus four times the number of stimuli plus two ($df = N_s^2 - 4N_s + 2$, $N_s = N_r$); thus, the method requires at least four stimuli to work.

The χ^2 value for the 2AFC is reduced to

$$\chi^2 = \sum_{(i,j)} \frac{(\tilde{P}_{i,j} - \hat{P}_{i,j})^2}{\tilde{P}_{i,j}(1 - \tilde{P}_{i,j})} N_{i,j}, \tag{7}$$

where $\tilde{P}_{i,j}$ is the probability of correct responses as defined by Eq. (5), $\hat{P}_{i,j}$ is the measured frequencies of correct responses, and $N_{i,j}$ is the number of trials measured for pair (i,j). The summation is performed over all pairs used in the experiment. The number of degrees of freedom is the number of pairs minus the number of parameters – twice the number of stimuli minus two.

TvC estimation was based on models obtained in the IT task. A transducer function $R(c)$ and noise amplitudes $\sigma(c)$ were linearly interpolated from the estimated model parameters. Here c denotes the contrast level. Then, the contrast increment that gives 80% correct discriminations was found numerically. More specifically, for all contrast levels, and all contrast increments, performance $P(c, c + \Delta c)$ was computed using the following equation (Kontsevich et al. (2002), Eq. (3)):

$$P(c, c + \Delta c) = \Phi\left(\frac{R(c + \Delta c) - R(c)}{\sqrt{\sigma(c)^2 + \sigma(c + \Delta c)^2}}\right). \tag{8}$$

The value of Δc for which $P(c, c + \Delta c) = 80\%$ was taken as a threshold.

3. Results

We performed a number of IT experiments. Twelve experiments are presented here with a contrast range of more than 30%. Three experiments out of 12 were complemented by 2AFC experiments. Five observers out of six performed TvC experiments. In the following text, and in the figures each experiment is encoded by two letters referring to observer and a number representing experimental condition (a set of presented contrast and the presence or absence of feedback, see Section 2.1.1).

3.1. IT and 2AFC data

In each IT experiment the observer was presented with a fixed number of stimuli. The histograms of observer responses to each presented stimulus are shown in Fig. 1 (the set of histograms represents the confusion matrix). There was considerable confusion between stimuli since neighboring stimuli were chosen to have close contrast levels. In some cases the response distributions for the stimulus with the highest and lowest contrast levels were narrow

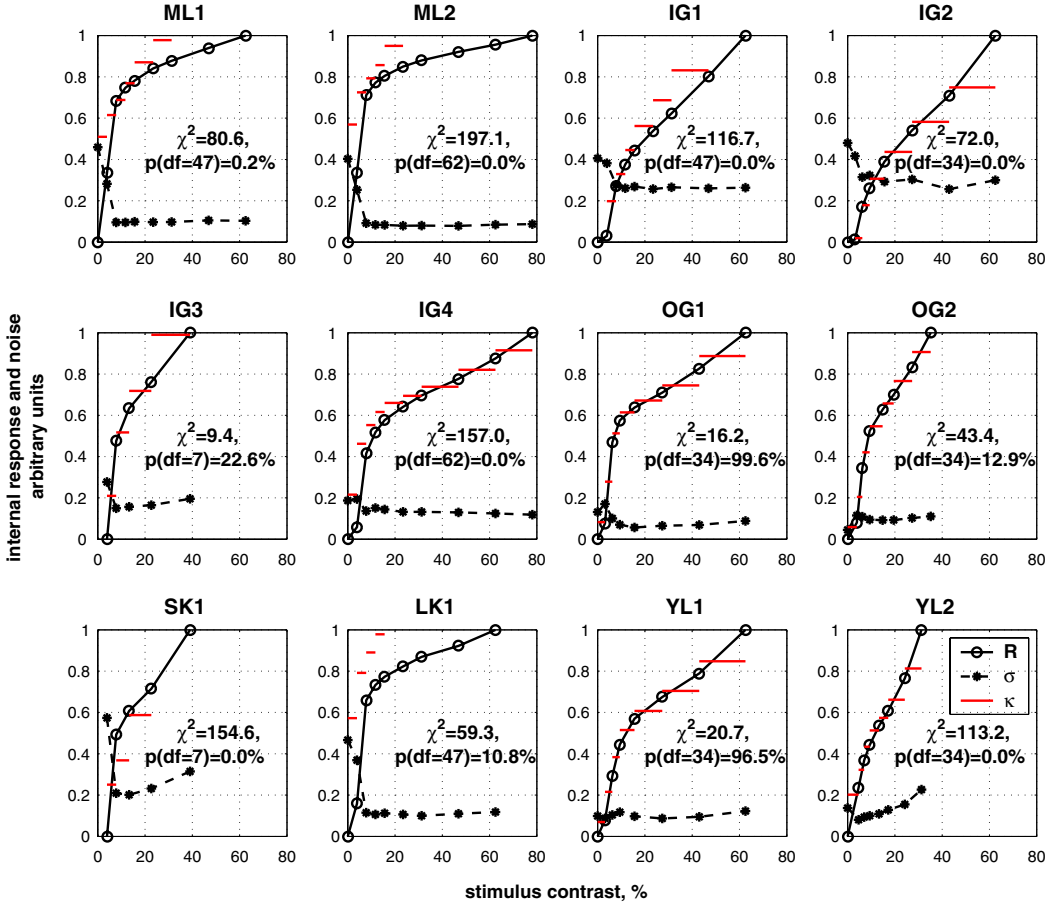

Fig. 3. Estimated model parameters. Contrast response function and noise amplitudes estimated from the data presented in Fig. 1. Each plot corresponds to an independent experiment. Most plots show relatively flat noise amplitudes above some threshold. On the other hand, many models have very large χ^2 values.

Please cite this article in press as: Katkov, M. et al., Inverse modeling of human contrast response, *Vision Research* (2007), doi:10.1016/j.visres.2007.06.024

(~2 bins). The particular shape of the response distribution depends on the distribution of the internal responses to the corresponding stimulus and on the subjective criteria that were established by the observer for category boundaries, which we assumed are stimulus-independent in a particular experiment. The placement of criteria may be affected in turn by stimulus spacing.

Three observers performed 2AFC discrimination experiments (Fig. 2, left column). It is well-known that the performance in the 2AFC experiments corresponds to the area under the ROC curve in the rating task (Green & Swets, 1966). Assuming that the criteria in IT are fixed during an experiment, the ROC curve of a stimulus pair (s_i, s_j) can be constructed from the corresponding pair of lines of confusion matrix (histograms in Fig. 1) and can be used to predict the 2AFC discrimination performance for that pair. The right column in Fig. 2 shows a comparison between the measured 2AFC performances and the estimated area under the ROC curves (ROC area) for the corresponding stimuli. The ROC area predicts more than 80% of the variance in discrimination performance. The difference between the predicted and measured performances may originate from ROC area estimations, small differences between the stimuli used in the two experiments, and from finger errors in both experiments. Another option is instability of parameters (contrast response function, noise amplitudes or category boundaries) in time during the course of the experiment either between sessions or within a single session (see more in Section 3.2).

3.2. Model parameters estimated from IT data

Model parameters with the smallest found χ^2 values are presented in Fig. 3 (alternative representation of the same figure is available in Appendix A).

Most of the noise amplitudes are relatively flat above some contrast level, but χ^2 values are very high in many cases. Such a result possibly means that some of the assumptions underlying the model are incorrect, for example, the model parameters may change with time. In the latter case, the assumptions underlying χ^2 goodness of fit are not met—samples measured in different sessions do not have the same expected frequencies, and deviation of session data from their mean across sessions are not normally distributed any more. Taking session data as samples, we can compute the distribution of χ^2 values as computed by Eq. (6) across sessions. We expect these values to be distributed according to the χ^2 distribution if the samples' data

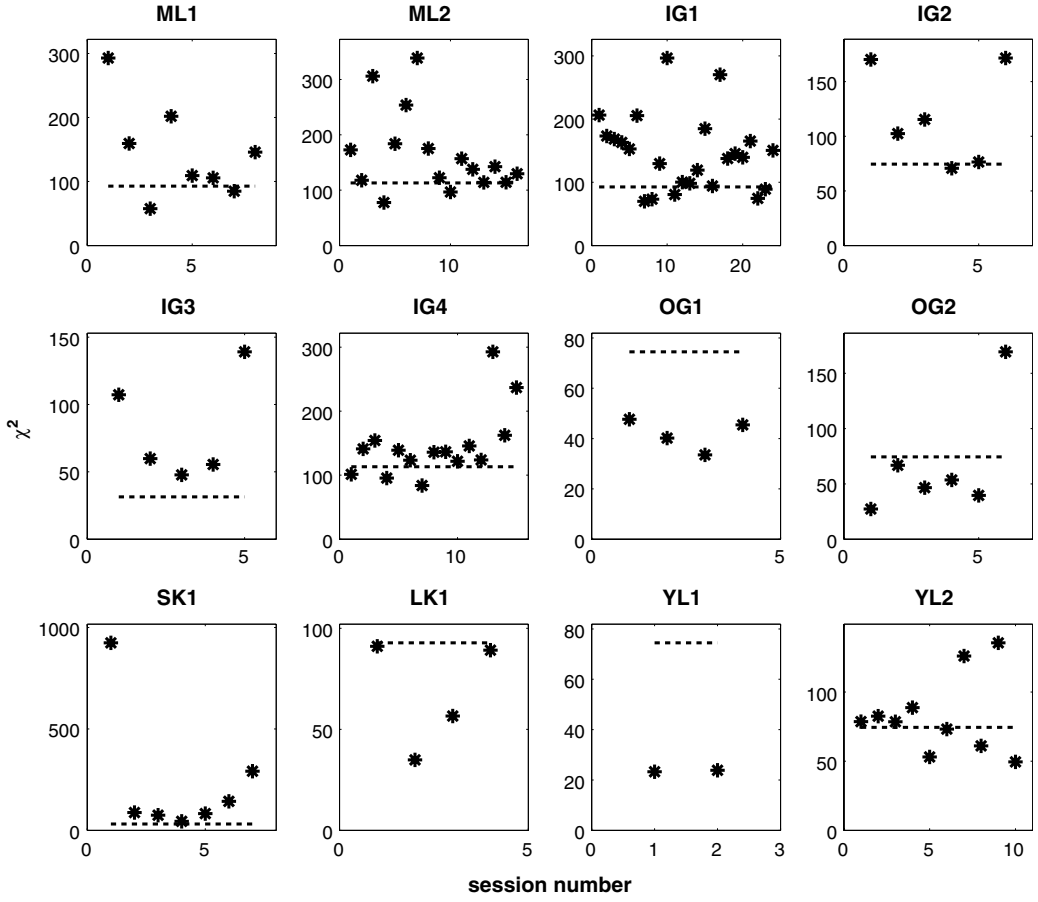

Fig. 4. χ^2 values computed for each session, taking the data averaged across sessions as expected frequencies. The dashed horizontal lines represent a 5% rejection boundary, assuming zero free-parameters. In the experiments where most of the session χ^2 values are in the acceptance range, the χ^2 values corresponding to the estimated model presented in Fig. 3 are also low.

Please cite this article in press as: Katkov, M. et al., Inverse modeling of human contrast response, *Vision Research* (2007), doi:10.1016/j.visres.2007.06.024

are drawn from the same model and only finite sampling is an issue. In particular, we do not expect that more than 5% of the computed values are above the 95 percentile line. In Fig. 4 we present χ^2 values computed for each session and a 95 percentile line for the corresponding χ^2 distribution. Often there are too many sessions with χ^2 values above the 95 percentile line, suggesting that mean data are indeed not representative of the session data. For example, it is possible that observers do not remember criteria across days, and that the session-by-session variability of criteria is not normal.

Fig. 5 shows model parameters estimated from data measured in each session separately. For the purpose of presentation, we changed the scale of the internal response axis, and the position of the origin. This transformation does not change the predicted performances (see Eq. (1)). Median values of noise amplitudes in all models were set to one (by changing the scale of internal response axis), and median values of contrast response functions (CRF) were forced to have the same value for all sessions by shifting the position of the origin. Only sessions with a significance level above 5% (by χ^2 goodness of fit test with degrees of freedom computed as specified in Section 2.2) are shown. It can be seen that the estimated parameters (category boundaries k, noise amplitudes σ in high and low contrast regions, and contrast response function R in low and high contrast regions) of the model vary between sessions. Both CRF and noise amplitudes were more stable for high contrast levels than for low contrast levels. Noise amplitude was relatively flat in the high contrast range. Nevertheless, there is inter-session variability of noise amplitudes at high contrast in some experiments. Many noise amplitude curves decrease with contrast. It is possible that some type of uncertainty is in effect here since low contrast stimuli are barely seen, whereas for high contrast, where noise is quite flat, stimuli are clearly seen. Finally, there is a session-by-session variability in the position of criteria. In experiments where response feedback was provided, criteria were positioned between the corresponding stimuli, suggesting that observers are able to learn the task, and set criteria at nearly the optimal position for the identification task.

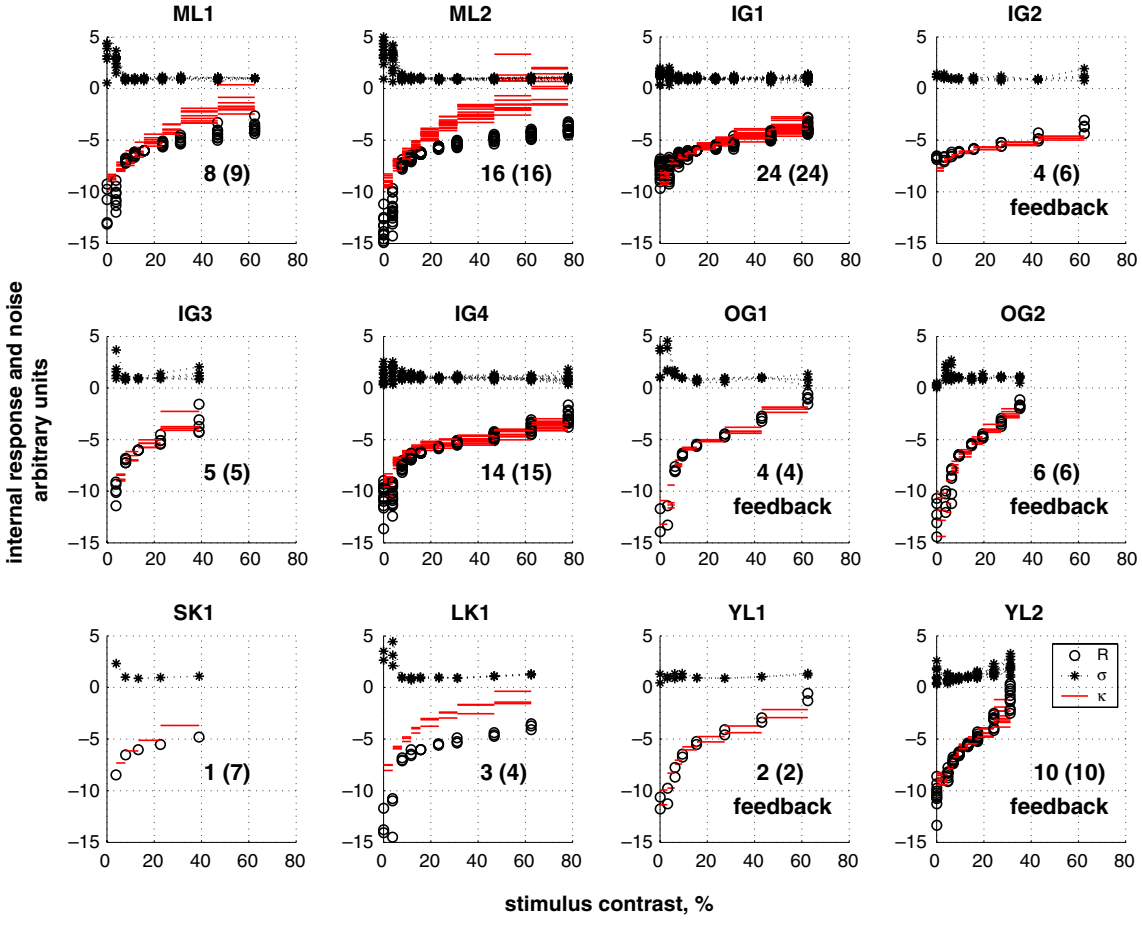

Fig. 5. Model parameters (R, σ, k) estimated for each session separately. Each box corresponds to a single session. Only sessions with a significance level above 5% (by χ^2 goodness of fit) are shown. Each plot corresponds to an independent experiment. Most plots show relatively flat noise amplitudes (filled symbols) above some threshold (compare with Fig. 3). There is a larger variability in CRF (empty symbols) and noise amplitudes for low contrast regions than for high contrast regions. In the experiments with feedback (marked with word "feedback") the positions of criteria (horizontal line segments, red) are set nearly optimally for the identification task. In many cases the contrast response function can be described as having two regimes: one with low and one with high gains. The number of accepted sessions are shown on each plot, with the total number of sessions presented in parentheses. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article in press as: Katkov, M. et al., Inverse modeling of human contrast response, *Vision Research* (2007), doi:10.1016/j.visres.2007.06.024

Criteria variability in the experiments without feedback was sometimes greater than the variability in the parameters corresponding to the transducer function, and usually were not optimal. However, some observers managed to place criteria near optimal positions without feedback.

Taking into account intra-session variability of model parameters, we can check whether the data can reject a constant noise model. It can be seen from Fig. 5 that the low contrast part of the noise function in many cases is not constant. We formally checked the constant noise assumption using the χ^2 goodness of fit test, and surprisingly found that five experiments out of 12 can be considered compatible with the constant noise model. If a few (2–4) of the smallest contrast levels are removed from the analysis, the number of compatible experiments increases to seven (Fig. 6, experiments where not more than one p-value is below the 5% line). Moreover, it can be seen (Figs. 5 and 6) that in experiments with many small p-values, there is a large variability in noise amplitudes at high contrast levels (Fig. 5, experiments IG2, IG3, YL2).

3.3. Comparison of measured and predicted TvC curves

Fig. 7 shows a comparison between discrimination thresholds measured using an adaptive staircase procedure (see details in Section 2.1.3) and the TvC curve predicted by the models estimated from the IT data (see Fig. 5). As can be expected, the predicted TvC curves exhibit variability due to variability in the estimated parameters. Nevertheless, some features are surprisingly stable, e.g., the position of the dipper, the shape of the curves on the left and on the right side of the dipper. These features are also stable across observers. There is general agreement between the predicted and measured curves. On the other hand, a more detailed study is required to explain the discrepancies between curves. For example, it is possible that within-session instability of parameters (transducer function, noise amplitudes, and criteria positions) leads to effectively larger noise amplitudes and affects threshold predictions that, in this case, can be larger than the measured ones, e.g., for observers SK1, IG1, and IG2.

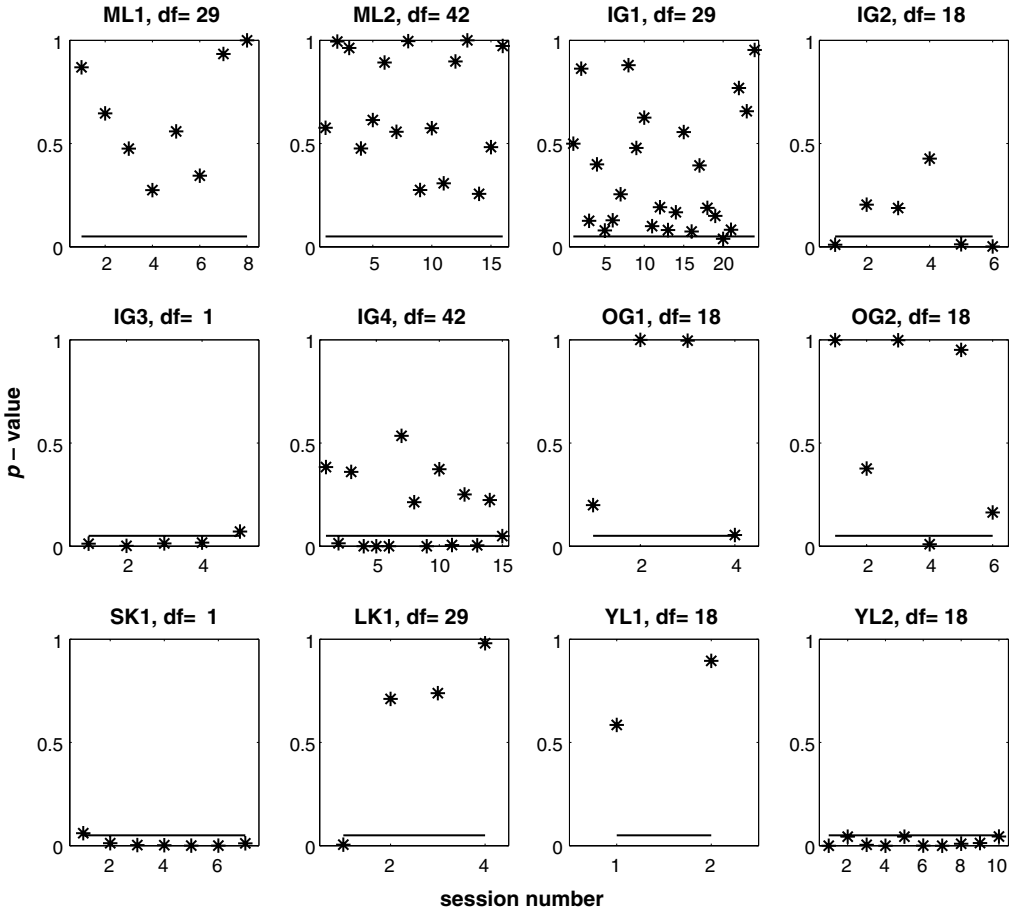

Fig. 6. p-values for the best fitting model computed for each session separately, assuming a constant noise model, and after removing a few (2–4) stimuli with the lowest contrast levels. The horizontal lines represent the 5% rejection criterion (when p-values fall below this line, a null-hypothesis about constant noise should be formally rejected). Note that for observers with many sessions below the 5% line, the model parameters have larger variability at high contrast levels (Fig. 5).

Please cite this article in press as: Katkov, M. et al., Inverse modeling of human contrast response, *Vision Research* (2007), doi:10.1016/j.visres.2007.06.024

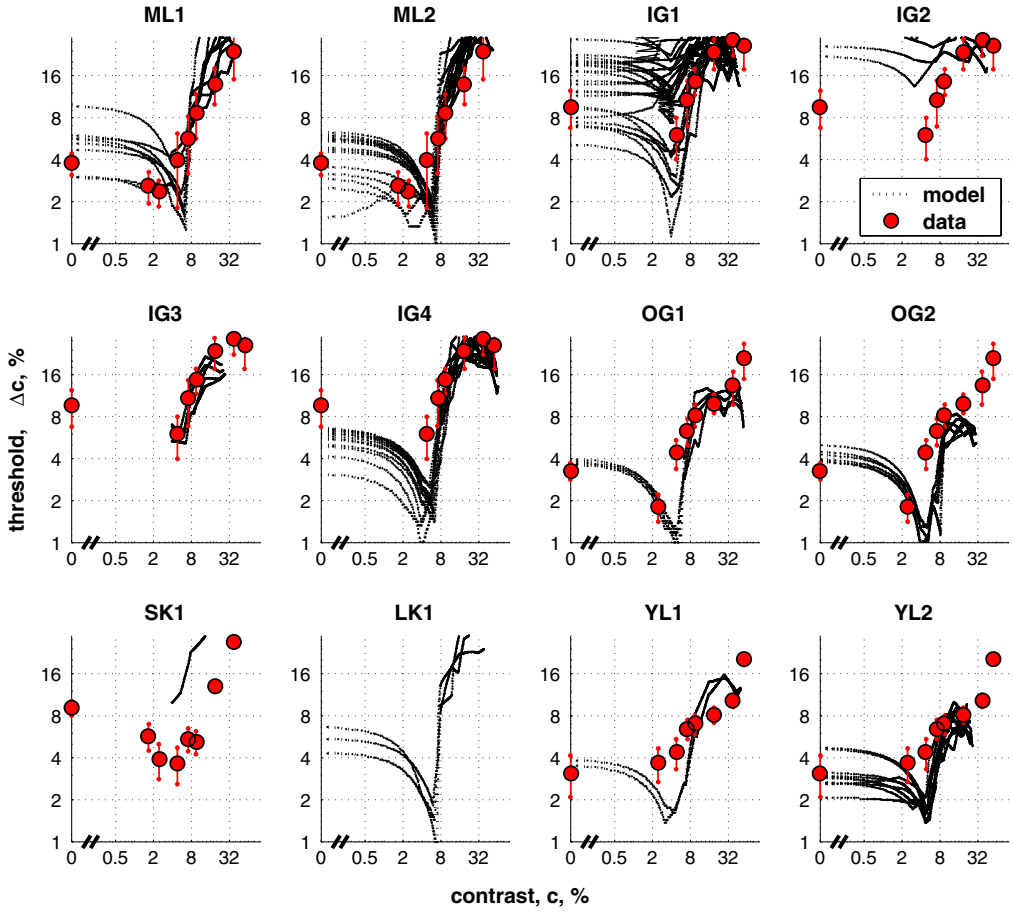

Fig. 7. Comparison between TvC curve measured in 2AFC experiment and the ones predicted from the estimated IT models. TvC curves measured by adaptive 2AFC procedure are depicted by circles. TvC curves estimated from the models presented in Fig. 5 are shown by dashed lines. Each line corresponds to the TvC curve computed from a single session data. Surprisingly, despite the session-by-session variability of the parameters, the TvC curves are often quite similar even across observers. Many predicted TvC curves are in agreement with the mean measured thresholds, even though there is no adjustment for criteria variability. Error bars represent one standard error. Observer LK did not participate in experiments related to threshold measurements.

4. Discussion

Using the Identification Task (Thurstonian scaling method of successive intervals, McNicol, 2005; Torgerson, 1958), we estimated the transducer function (contrast response function, CRF, R) and noise amplitudes (σ) for contrast perception in humans. In seven experiments out of 12, the data were compatible with the constant noise assumption for contrasts above the detection threshold (Fig. 6). In the other five experiments we found a large session-by-session variability of noise amplitudes (σ) at high contrast (Fig. 5). The estimated noise amplitudes for low contrast stimuli, below detection threshold, were not stable and varied between sessions in the same experiment. Such a result can be expected from uncertainty regarding the location or time of stimulus presentation, which selectively affects the detection of low contrast (barely visible or invisible) stimuli. In addition, we found substantial variability between sessions in the estimated decision criteria (k), which may lead to variability in the estimation of CRF (R) at low contrasts. In the full range of measured contrasts, five experiments out of 12 were consistent with a constant noise assumption.

To incorporate inter-session variability into the data analysis, we considered each daily experimental session separately. We found that not all sessions had SDT models compatible with the data (according to the χ^2 goodness of fit test). It is possible that the SDT model, despite its generality, is incapable of capturing the mechanisms underlying human contrast perception. Alternatively, it is possible that model parameters vary within a single experimental session, an assumption that calls for a more detailed study of intra-session variability. In the singular case, for example in 2AFC, very different model parameters lead to statistically indistinguishable data (Katkov et al., 2007). On the other hand, Fig. 4 shows that model parameters found in different sessions cannot explain the same data. Therefore, we suspect that the instability of model parameters originates from the biological system and not from the mathematical construction, in contrast to results we obtained in contrast discrimination experiments (Katkov et al., 2006b).

There were not enough contrast levels in the low contrast region to allow for any claim concerning the CRF behavior near zero contrast. On the other hand, the CRF

Please cite this article in press as: Katkov, M. et al., Inverse modeling of human contrast response, *Vision Research* (2007), doi:10.1016/j.visres.2007.06.024

can be described as having two regimes of different behaviors with a transition at or around the dipper contrast—a high response gain at low contrasts, which rapidly decreases at around the dipper and remains low and constant at high contrast levels. In some cases, CRF shows accelerating behavior at high contrast.

TvC curves that are computed from IT data are very similar to the measured TvC curves. Surprisingly, the predicted position of the dipper and the shape of the predicted curves were very similar even across observers. The predicted flattening of the TvC curve at high contrast can be attributed to the fact that the estimated CRF at high contrast levels becomes linear or even accelerates (Fig. 5). The "dipper" effect in TvC curves can be explained by a decreasing noise level below the dipper contrast and by changing the gain of CRF above the dipper contrast.

Previous attempts to separate signal and noise in contrast transduction, using 2AFC data, included explicit assumptions about the functional relationship between internal response and noise. Such assumptions can reduce the ambiguity found in singular cases due to the added constraints and may lead to unique models which, to a large extent, depend on the assumed functional form. Kontsevich et al. (2002) assumed a special parametric form of the transducer function and noise amplitudes—power functions. Within this class of models they found a relatively good fit with positive exponents for both the CRF and noise amplitudes. Therefore, they concluded that human observers have accelerating CRF and multiplicative noise. Georgeson and Meese (2006) pointed out that Kontsevich et al. (2002) analyzed only a particular subset of the possible models, and that other models can also describe the same data. To support this point, Georgeson and Meese (2006) found an alternative constant noise model that statistically describes the data as well as the Kontsevich et al. (2002) models for three out of four experiments. The last experiment (observer AK-T) had high χ^2 values for both types of models; thus both models were rejected. Nevertheless, Klein (2006) argued that observer AK-T showed preferences for a multiplicative model. The analysis of both Kontsevich et al. (2002) and Georgeson and Meese (2006) were limited to specific parametric forms of the transducer function, and therefore failed to describe observer AK-T. Klein (2006), following Katkov et al. (2006b), allowed CRF values to be free parameters during fitting; however, noise amplitudes were assumed to be a power function of CRF values because the data measured in Kontsevich et al. (2002) do not have enough constraints to allow all model parameters to be free. Surprisingly, he obtained the same noise exponent as Kontsevich et al. (2002), rejecting the constant noise assumption for observer AK-T. In Fig. 8 we present an alternative fit for the same observer, assuming a functional form of noise dependence on the internal response, which is consistent with the one found in the present work. The p-value of this fit is greatest among the known fits. Unlike in Klein (2006) the stimulus parameters were not modified, leading to effectively smaller degrees of freedom. Although we constrained only the relationship between CRF (R) and noise amplitudes (σ), and not the shape of CRF itself, the later shows the same tendency as the CRF fitted from our IT data—there are two regions of CRF with different slope (Figs. 3 and 5). Therefore, the data measured in Kontsevich et al. (2002) do not contradict our conclusions. Interestingly, recently García-Pérez and Alcalá-Quintana (2007) also reported that their data can be better described by decreasing with contrast noise.

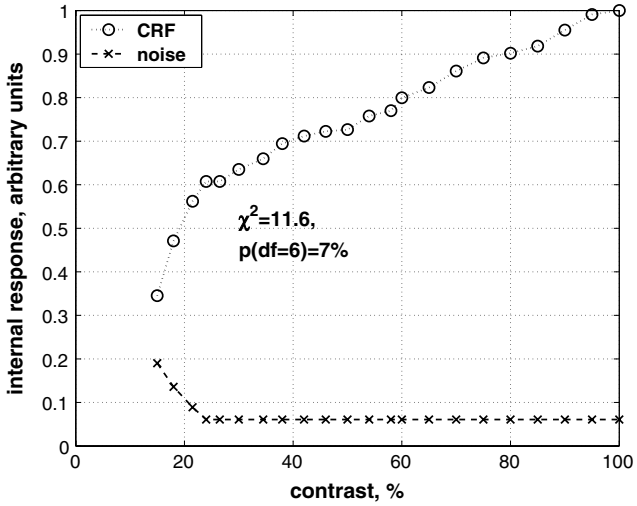

Fig. 8. Fit of Kontsevich et al. (2002) data for the observer AK-T. The lowest four contrast levels were removed from the analysis since the data there do not provide enough constraints and allow arbitrary shapes of noise amplitudes for those contrast levels (see details in the Appendix B). The models' parameters: unconstrained CRF $R(c)$, and noise amplitude $\sigma = \eta\,(C_1 - R(c) + 1)^{0.68}$ when $(R(c) < C_1)$, and η otherwise (C_1 and η are some constants). Features of the estimated model are similar to those obtained in our experiments (Fig. 5)—the transducer function changes its gain and noise changes its behavior at the same contrast level.

Nachmias and Kocher (1970) applied a closely related technique (rating scale task with only two stimuli presented during a block of trials) to detection and discrimination of luminance increments. Their results showed increasing noise amplitudes with luminance in the low luminance conditions. However, when the stimuli to be discriminated were both of high luminance (which corresponds to our viewing conditions at low contrast) the ratio between neighboring noise amplitudes were often "substantially less than one" (Nachmias & Kocher, 1970, p. 384)—similar to our findings.

Finally, many of the best fitted models had noise amplitudes (Fig. 5, σ) decreasing with contrast in the low contrast range. This contradicts an uncertainty model based on using the maximum response as the decision variable (Pelli, 1985). According to this model, at zero pedestal decision is based on the maximum response value across all detection channels (stimulus-dependent and -independent) while at some higher contrast level only (or mostly) the stimulus-dependent channels contribute to the decision. Since the distribution of the maximum of several normal random

Please cite this article in press as: Katkov, M. et al., Inverse modeling of human contrast response, *Vision Research* (2007), doi:10.1016/j.visres.2007.06.024

variables has a smaller variance than the variance of any of the original variable, this model predicts a noise function which is increasing with contrast. Nevertheless, it may be possible to describe this type of noise dependency with an uncertainty model that is using a different decision rule. The rule should gradually reduce the weight of the signal from stimulus-independent channels, or gradually reduce the number of channels effectively considered at the decision stage, when the stimulus strength grows. This process leads to a gradual reduction of noise with contrast. Probably, above dipper contrast, the process saturates, and a small fixed set of channels is monitored by the decision circuit.

5. Conclusion

Using the Identification Task, we estimated the Contrast Response Function (CRF), noise amplitudes, and positions of category boundaries (criteria) in a simple Signal Detection Theory model underlying human contrast perception. Estimated model parameters varied across sessions for the same observer. Noise amplitudes were relatively flat above some contrast. Five out of 12 experiments were compatible with the constant noise assumption in the whole range of measured contrasts. Experiments, where data were not compatible with the constant noise model, show large session-by-session variability of parameters. For low contrast levels, the noise amplitudes were mostly either decreasing functions of contrast or constant. In rare cases, they were slightly increasing functions. The CRF can be described as having two modes with a smooth transition between them. For high contrast levels, CRF has a relatively low gain, and for low contrast levels the gain is much larger. TvC curves computed from estimated model parameters had a similar shape across sessions for the same observer and across observers. The position of the dipper was in good agreement between sessions, experiments, and observers. Moreover, the position of the dipper seems to coincide with the region where CRF changes its gain and the noise amplitude changes its behavior.

Acknowledgments

The authors thank Stan Klein and an anonymous reviewer for helpful comments that substantially improved the article. This research was supported by the Basic Research Foundation, administrated by the Israel Academy of Science. We thank Yoram Bonneh for supplying the software used in the experiments.

Appendix A. Alternative representation of data in Fig. 3

As per reviewer suggestion we present here Fig. 3 in log–log (Fig. A.1) and lin–log (Fig. A.2) axes.

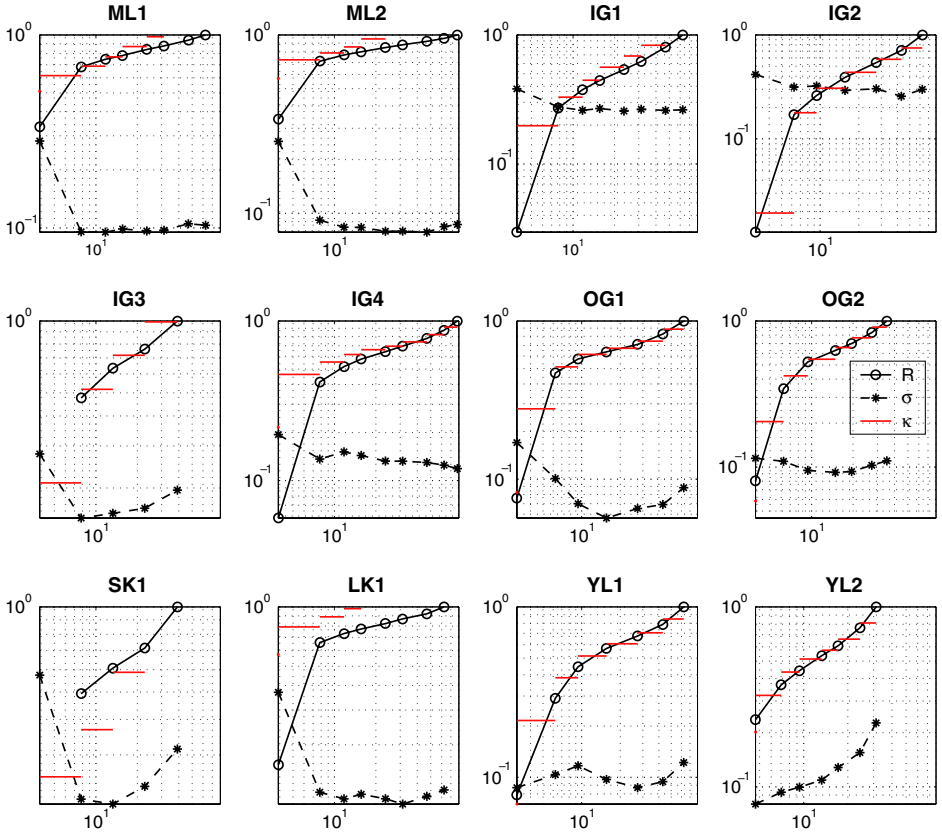

Fig. A.1. Model parameters (Fig. 3) are shown in log–log coordinates.

Please cite this article in press as: Katkov, M. et al., Inverse modeling of human contrast response, *Vision Research* (2007), doi:10.1016/j.visres.2007.06.024

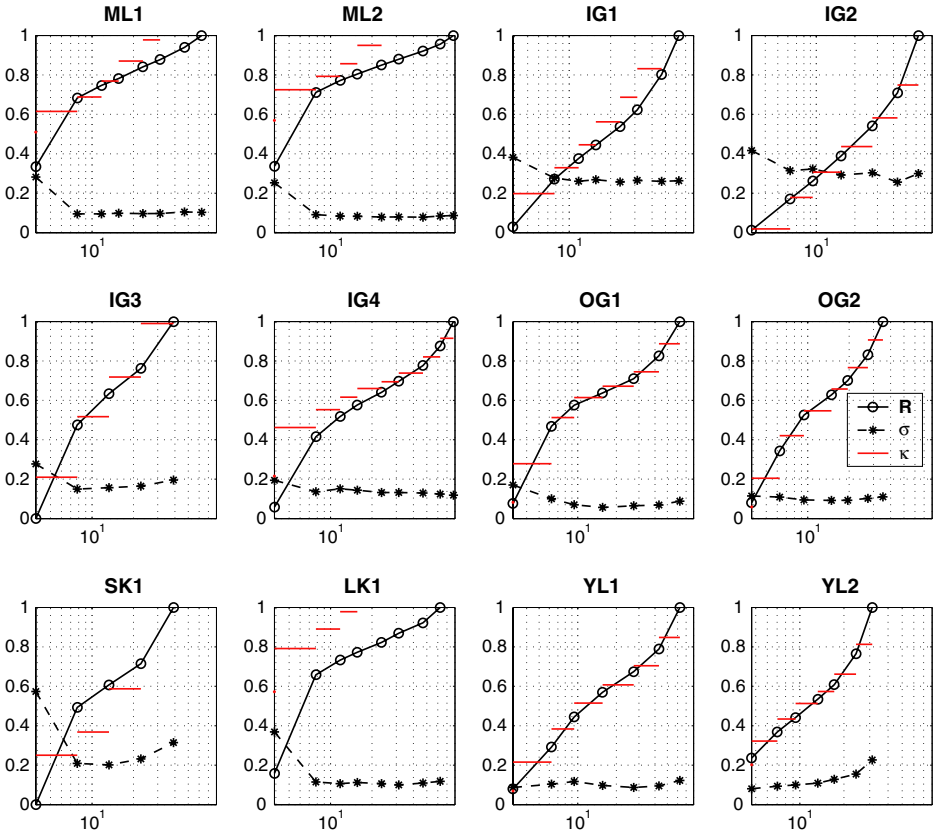

Fig. A.2. Model parameters (Fig. 3) are shown in log–lin coordinates.

Appendix B. AK-T observer

The data presented in Kontsevich et al. (2002) consisted of three psychometric curves with base contrasts 15%, 30%, and 60%. Test contrasts spanned 18–95%, 2–100%, 65–100%, respectively. Below 15% contrast there is only one psychometric curve. Therefore, only one constraint is available for each contrast level c below 15%. Nevertheless, there are two free parameters for each contrasts—R_c and σ_c. Therefore, we can always choose arbitrary σ_c values for contrast $c < 15\%$, and choose corresponding R_c such that

$$P_{c_{\text{base}},c} = \Phi\left(\frac{R_c - R_{\text{base}}}{\sqrt{\sigma_c^2 + \sigma_{\text{base}}^2}}\right).$$

Since $\Phi(x)$ is a monotonic function of x and the denominator is defined by chosen σ_c the solution for R_c exists and unique. Fig. B.1 shows an example of possible values in the low contrast range. Note, σ does not have to be monotonic at low contrasts.

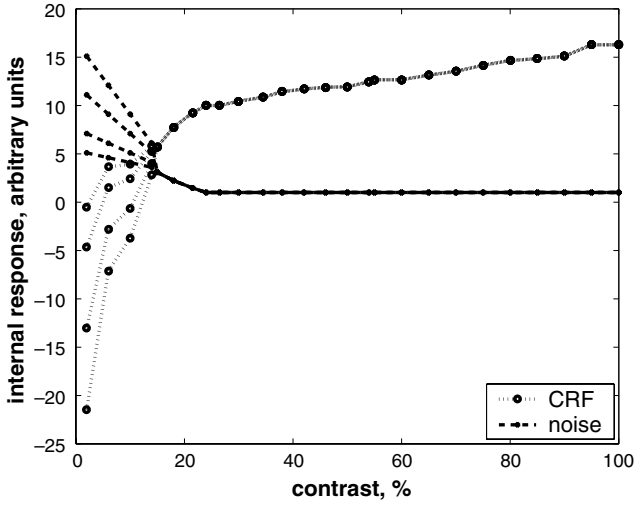

Fig. B.1. Fit of Kontsevich et al. (2002) data for the observer AK-T. Example of fits to observer AK-T with different low contrast values leading to the same χ^2.

References

Boynton, G. M., Demb, J. B., Glover, G. H., & Heeger, D. J. (1999). Neuronal basis of contrast discrimination. *Vision Research, 39*, 257–269.

Chirimuuta, M., & Tolhurst, D. J. (2005). Does a Bayesian model of V1 contrast coding offer a neurophysiological account of human contrast discrimination? *Vision Research, 45*, 2943–2959.

Foley, J. M., & Legge, G. E. (1981). Contrast detection and near-threshold discrimination in human vision. *Vision Research, 21*, 1041–1053.

García-Pérez, M. A., & Alcalá-Quintana, R. (2007). The transducer model for contrast detection and discrimination: Formal relations, implications, and an empirical test. *Spatial Vision, 20*, 5–43.

Please cite this article in press as: Katkov, M. et al., Inverse modeling of human contrast response, *Vision Research* (2007), doi:10.1016/j.visres.2007.06.024

Georgeson, M. A., & Meese, T. S. (2006). Fixed or variable noise in contrast discrimination? The jury's still out.... *Vision Research, 46*, 4294–4303.

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization & machine learning*. Addison-Wesley.

Gorea, A., & Sagi, D. (2001). Disentangling signal from noise in visual contrast discrimination. *Nature Neuroscience, 4*, 1146–1150.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Howell, D. C. (2002). *Statistical methods for psychology*. Pacific Grove: Duxbury.

Katkov, M., Tsodyks, M., & Sagi, D. (2006a). Analysis of two-alternative force-choice Signal Detection Theory model. *Journal of Mathematical Psychology, 50*, 411–420.

Katkov, M., Tsodyks, M., & Sagi, D. (2006b). Singularities in the inverse modeling of 2AFC contrast discrimination data. *Vision Research, 46*, 256–266.

Katkov, M., Tsodyks, M., & Sagi, D. (2007). Singularities explained: Response to Klein. *Vision Research*. doi:10.1016/j.visres.2006.10.030.

Klein, S. A. (2001). Measuring, estimating and understanding the psychometric function: A commentary. *Perception & Psychophysics, 63*, 1421–1455.

Klein, S. A. (2006). Separating transducer nonlinearities and multiplicative noise in contrast discrimination. *Vision Research, 46*, 4279–4293.

Kontsevich, L. L., Chen, C. C., & Tyler, C. W. (2002). Separating the effects of response nonlinearity and internal noise psychophysically. *Vision Research, 42*, 1771–1784.

Lu, Z. L., & Dosher, B. A. (1999). Characterizing human perceptual inefficiencies with equivalent internal noise. *Journal of the Optical Society of America A, 16*, 764–778.

McNicol, D. (2005). *A primer of signal detection theory*. London: Lawrence Erlbaum Associates.

Nachmias, J., & Kocher, E. C. (1970). Visual detection and discrimination of luminance increments. *Journal of the Optical Society of America, 60*, 382–389.

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.

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (2005). *Numerical recipes in C: The art of scientific computing*. New York: Cambridge University Press.

Softky, W. R., & Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *The Journal of Neuroscience, 13*, 334–350.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review, 34*, 273–287.

Torgerson, W. S. (1958). *Theory and method of scaling*. New York: Wiley.

Please cite this article in press as: Katkov, M. et al., Inverse modeling of human contrast response, *Vision Research* (2007), doi:10.1016/j.visres.2007.06.024