ELSEVIER

Cognitive Science 28 (2004) 167–207

COGNITIVE SCIENCE

http://www.elsevier.com/locate/cogsci

Characterizing perceptual learning with external noise

Jason M. Gold [a,*], Allison B. Sekuler [b], Partrick J. Bennett [b]

[a] *Department of Psychology, Indiana University, Bloomington, IN 47405, USA*
[b] *Department of Psychology, McMaster University, Hamilton, Ont., Canada*

Received 9 August 2002; received in revised form 26 September 2003; accepted 1 October 2003

Abstract

Performance in perceptual tasks often improves with practice. This effect is known as 'perceptual learning,' and it has been the source of a great deal of interest and debate over the course of the last century. Here, we consider the effects of perceptual learning within the context of signal detection theory. According to signal detection theory, the improvements that take place with perceptual learning can be due to increases in *internal signal strength* or decreases in *internal noise*. We used a combination of psychophysical techniques (external noise masking and double-pass response consistency) that involve corrupting stimuli with externally added noise to discriminate between the effects of changes in signal and noise as observers learned to identify sets of unfamiliar visual patterns. Although practice reduced thresholds by as much as a factor of 14, internal noise remained virtually fixed throughout training, indicating learning served to predominantly increase the strength of the internal signal. We further examined the specific nature of the changes that took place in signal strength by correlating the externally added noise with observer's decisions across trials (response classification). This technique allowed us to visualize some of the changes that took place in the linear templates used by the observers as learning occurred, as well as test the predictions of a linear template-matching model. Taken together, the results of our experiments offer important new theoretical constraints on models of perceptual learning.

Keywords: Perceptual learning; Internal noise; Internal signal; Ideal observer; Classification image

1. Introduction

The observation that performance in perceptual tasks can improve with practice has been documented for over a century (Dresslar, 1894; Gibson, 1969; Gilbert, 1994). This improvement in perceptual discriminations with training is referred to as *perceptual learning*, and it

* Corresponding author. Tel.: +1-812-855-4635; fax: +1-812-855-4691.
E-mail address: jgold@indiana.edu (J.M. Gold).

0364-0213/$ – see front matter © 2004 Cognitive Science Society, Inc. All rights reserved.
doi:10.1016/j.cogsci.2003.10.005

has been found to occur in a wide variety of perceptual tasks, including very simple sensory discriminations such as visual and tactile acuity tasks (Fahle, Edelman, & Poggio, 1995; Fahle & Morgan, 1996; Poggio, Fahle, & Edelman, 1992; Sathian & Zangaladze, 1998), orientation discrimination (Matthews, Liu, Geesaman, & Qian, 1999; Schiltz et al., 1999; Schoups, Vogels, & Orban, 1995), motion discrimination (Ball & Sekuler, 1987; Matthews et al., 1999), texture discrimination (Fine & Jacobs, 2000; Karni & Sagi, 1991) and auditory pitch discrimination (Demany, 1985; Recanzone, Schreiner, & Merzenich, 1993). Learning also has been found to operate over a wide range of time scales, from within as little as 100 trials (Fahle et al., 1995; Poggio et al., 1992) to as long as several weeks (Fiorentini & Berardi, 1997; Karni et al., 1998; Karni & Sagi, 1993; Sathian & Zangaladze, 1998; Schoups et al., 1995). Typically, the learning is restricted to the exact specifications of the stimuli and task where training has occurred (Ahissar & Hochstein, 1997; Ball & Sekuler, 1987; Crist, Kapadia, Westheimer, & Gilbert, 1997; Fahle & Morgan, 1996; Fiorentini & Berardi, 1980), and observers often do not require feedback in order to exhibit learning (Ball & Sekuler, 1987; Fahle et al., 1995; Herzog & Fahle, 1997, 1999). The combination of these findings (learning for simple stimuli, stimulus specificity, and implicit learning) has been taken as evidence that perceptual learning occurs at relatively early stages of sensory processing (Gilbert, 1994). Consequently, much of the recent psychophysical and physiological work on this topic has been directed toward localizing the neural substrates changed by perceptual learning in different tasks and modalities. Evidence from these experiments suggests that perceptual learning may modify neural mechanisms at or before the level of primary sensory cortex. For example, several studies have found only partial or no inter-ocular transfer of learning for simple visual discrimination tasks (Ball & Sekuler, 1987; Fahle et al., 1995), suggesting some of the effects of learning for these tasks occur in monocular mechanisms before the site of binocular integration. Similarly, physiological studies have found that practice changes the response properties of neurons in primary cortical areas for simple discrimination tasks, such as visual orientation discrimination (Schiltz et al., 1999; Schoups et al., 2001) and auditory frequency discrimination (Recanzone et al., 1993). Other physiological studies have investigated the topographic changes that take place in sensory cortical maps with practice (Buonomano & Merzenich, 1998; Recanzone et al., 1993). These studies have found sensory cortex to be highly plastic, with striking amounts of cortical reorganization and reallocation taking place as a result of practice. However, there is also evidence that suggests higher-order mechanisms, such as those found in the prefrontal cortex, can change with perceptual learning (Asaad, Rainer, & Miller, 2000; Rainer & Miller, 2000) (see Goldstone, 1998 for a concise review and Fahle & Poggio, 2002 for a more detailed treatment).

1.1. Signal and noise

But what aspects of perceptual mechanisms change with learning? One way to approach this problem is to consider the effects of learning within the context of signal detection theory (Green & Swets, 1966). Signal detection theory is a general framework designed to characterize and quantify an observer's decision processes and sensitivity in a task. One of the central tenets of signal detection theory is the assumption that internal responses are probabilistic, so that a particular stimulus has only some probability of eliciting a particular internal response. The theory also assumes an observer makes decisions by comparing the internal response to a

criterion. One advantage of this theory is it provides an estimate of sensitivity (d') that is independent of the observer's response criterion. Another advantage is that it can be used to quantify optimal (ideal) decision processes, which can be used to estimate the amount of information used by non-optimal (e.g., human) observers in a given task (Geisler, 1989; Green & Swets, 1966; Tanner, 1961).

To understand how these concepts relate to perceptual learning, consider the task of learning to discriminate between two unfamiliar patterns (e.g., two faces, etc.). In such a task, the observer is briefly shown one of two patterns (chosen randomly) and must decide which pattern appeared in the interval. Psychologically, each trial will produce an internal response within the observer which must be used to make a decision about which signal was shown. If we ignore for the moment any trial-by-trial variability introduced by the stimulus (e.g., photon noise), an ideal observer that is shown the same exact stimulus at several different times will make the same decision on every presentation. However, such complete response consistency would not be expected from a human observer. Unlike an ideal observer, human observers have internal variability or 'noise' (Barlow, 1956, 1957; Green, 1964). This internal trial-by-trial variability is thought to originate from a variety of sources, ranging from the stochastic properties of sensory neurons (Croner, Purpura, & Kaplan, 1993; Tolhurst, Movshon, & Thompson, 1981; Tolhurst, Movshon, & Dean, 1983; Vogels, Spileers, & Orban, 1989) to random fluctuations in strategy or response criterion (Burgess, 1990; Raghavan, 1989). As a result, the same stimulus will not produce the same internal response on every presentation. Instead, it will produce a *distribution* of responses across identical presentations, and the variance of this distribution will be determined by the magnitude of the internal noise.

A second way that human and ideal observers differ is in terms of the goodness or 'efficiency' of the non-stochastic (i.e., non-random) aspects of any computations performed between stimulus encoding and making a decision. For many tasks (including those described in the experiment reported here), an ideal strategy is to use a linear filter or *template* that is matched to the spatial and temporal characteristics of the stimulus. Any deviation from this strategy reduces performance relative to the ideal. Adding noise to the template is one way to degrade performance, but another way is to alter the template in a non-random fashion. For example, an observer that uses just the bottom half of the stimulus will be sub-optimal (assuming there is information present in the top half of the stimulus). This would be a *deterministic* inefficiency rather than a *stochastic* inefficiency. An ideal observer uses a deterministic computation that is guaranteed to make optimal use of all of the information available in a given task. Unlike an ideal observer, human observers perform sub-optimal deterministic computations. These inefficiencies can arise from many sources, ranging from sub-optimal encoding by sensory organs (Banks, Geisler, & Bennett, 1987; Banks, Sekuler, & Anderson, 1991; Geisler, 1989) to the comparison of the sensory representation to a sub-optimal receptive field or template (Legge, Kersten, & Burgess, 1987). In terms of signal detection theory, the distance between the underlying internal response distributions (rather than their variance) is determined by the relative efficiency of the deterministic aspects of the observer's computations. Ultimately, an observer's sensitivity (d') is determined by the ratio of signal (distance between the distributions) to noise (standard deviation of the distributions) within the system.

If we now consider the problem of perceptual learning within the context of signal detection theory, we see that the effects of perceptual learning (e.g., improved performance in a task) can

result from either a decrease in internal noise or an increase in the efficiency of the deterministic aspects of observer computations (or some combination of the two).

1.2. Measuring the strength of signal and noise

It is not possible to discriminate between changes in internal noise and deterministic efficiency by simply measuring sensitivity at different points during learning. However, psychophysical techniques have been developed in recent years that, when used in combination with a simple pattern discrimination model, allow the effects of these changes to be disambiguated. The techniques are called *external noise masking* and *response consistency*.

1.2.1. External noise masking

A standard technique used by electrical engineers to estimate the intrinsic noise in an electronic device (e.g., an amplifier) is to refer the intrinsic noise to an externally added source of noise (Mumford & Schelbe, 1968). Pelli (1981) was among the first to apply a variant of this technique to human information processing. To understand fully Pelli's approach, it is useful to first consider his abstraction of the internal transformations performed by an observer in a pattern discrimination task. Pelli's 'black-box' model of an observer is illustrated schematically in Fig. 1. In this model, the observer receives a physical stimulus (in this case, a signal corrupted by an externally added noise). The stimulus is converted into an internal representation, where an internal noise of fixed variance is introduced and a calculation is performed on the representation. A decision is then made based on the resulting internal response. Notice that the model assumes that the internal noise is *added* to the representation and that both the variance of the internal noise and the calculation are invariant with respect to the magnitude of the stimulus. When the stimuli are achromatic visual images varying in luminance across space and/or time

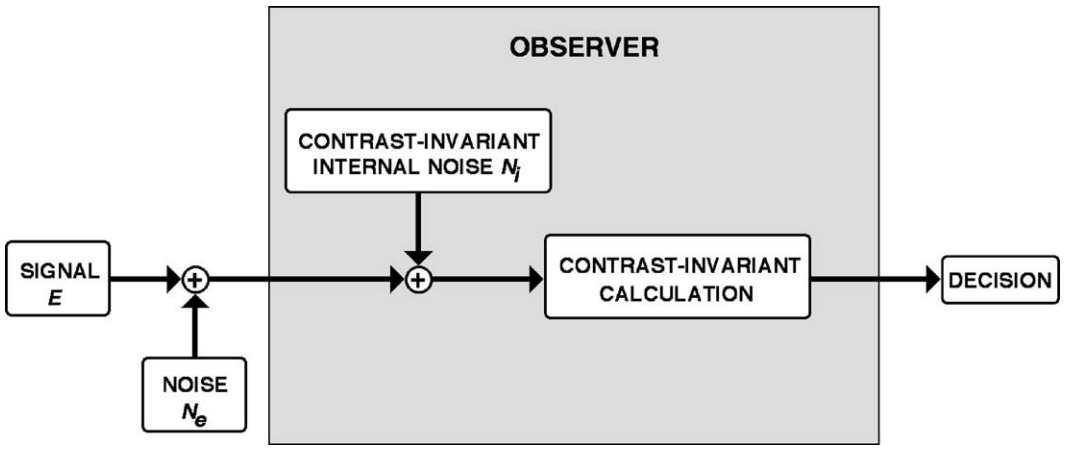

Fig. 1. A black-box model of a human observer in a perceptual discrimination task (adapted from Pelli, 1981, 1990). The observer is treated like a black-box that receives a noisy external stimulus ($E + N_e$), introduces a fixed amount of variability to the stimulus (N_i), performs a calculation that is reduced to an internal response, and makes a decision based on the magnitude of the internal response.

(as they are in the experiments reported here), the internal noise and calculation are referred to as *contrast–invariant internal noise* and a *contrast–invariant calculation*, respectively.

Given this framework, the observer's contrast threshold will be linearly related to the contrast of the external noise (Legge et al., 1987) according to the equation

$$E = k(N_e + N_i) \tag{1}$$

where E is the energy (a measure of signal contrast) of the signal at threshold, N_e is the external noise power spectral density (a measure of noise contrast), and k and N_i are free parameters (see Section 2.1 below for formal definitions of energy and noise power spectral density). The parameter N_i is often referred to as the observer's *equivalent input noise* because it is equal to the amount of external noise that must be added to the display to double the observer's noise-free threshold. The parameter k is a measure of how rapidly the observer's threshold increases with increasing external noise, and is inversely proportional to the goodness or 'efficiency' of the observer's calculation. An observer's *calculation efficiency* is computed by comparing k for a human observer to that an ideal observer, and can be interpreted as the proportion of the available information used by the human observer in the task.

Given this model of the observer, it is possible to estimate the magnitude of an observer's contrast–invariant internal noise and the efficiency of the observer's calculations by measuring thresholds in various amounts of externally added noise. The differential effects of contrast–invariant internal noise and calculation efficiency on thresholds in noise are illustrated in Fig. 2. The solid line in Fig. 2 depicts a hypothetical *noise masking function* for a human observer, where log threshold energy (E) is plotted as a function of log external noise power spectral density (N_e). Notice that the noise masking function (i.e., Eq. (1)) is curved when plotted in log–log coordinates. The solid arrow corresponds to the estimate of contrast–invariant internal noise (N_i). The dotted line in Fig. 2 shows the effects of reducing N_i by a constant c. Changing N_i in this fashion reduces thresholds at only low external noise levels, shifting the knee of the noise-masking function to a lower value (the dotted arrow in Fig. 2). The dashed line in Fig. 2 shows the effects of reducing the index of calculation efficiency k by the constant c. Changing k in this fashion will have the same effect across external noise levels, producing a uniform shift in the overall 'height' of the noise masking function when plotted in log–log coordinates.

Equivalent input noise and calculation efficiency have been measured for a wide variety of tasks, including grating detection (Pelli, 1981; Bennett, Sekuler, & Ozin, 1999), contrast discrimination (Legge et al., 1987), letter discrimination (Pelli & Farell, 1999; Raghavan, 1989; Tjan, Braje, Legge, & Kersten, 1995), object recognition (Tjan et al., 1995), divided attention (Dosher & Lu, 2000; Lu & Dosher, 1998), and motion discrimination (Lu, Liu, & Dosher, 2000). With little exception, the form of the noise masking functions have conformed well to the model described by Eq. (1).

However, there may be other sources of noise in the sensory systems that are not invariant with respect to stimulus magnitude. The effects of such a *contrast-dependent* internal noise in Pelli's black-box model can be seen by including a second independent noise source in Eq. (1):

$$E = k[N_e + N_i + m(N_e + N_i + E)^P] \tag{2}$$

where the proportionality constant m and the exponent P determine the magnitude of the contrast-dependent internal noise. There is both physiological (Tolhurst et al., 1983) and psy-

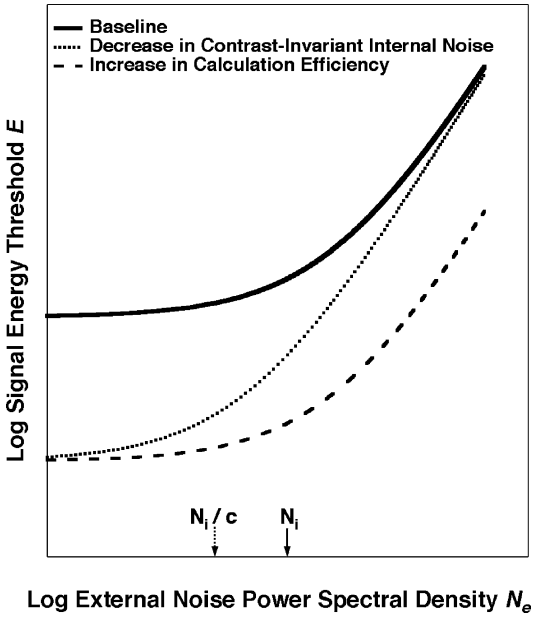

Fig. 2. Hypothetical noise-masking functions for a human observer. Log of signal energy threshold (E) is plotted as a function of external noise power spectral density (N_e). The finely dashed line depicts a reduction in equivalent input noise N_i by a constant factor c relative to the solid line. The coarsely dashed line depicts an increase in calculation efficiency (indexed by k) by a constant factor c relative to the solid line.

chophysical (Burgess & Colborne, 1988) evidence that the magnitude of the contrast-dependent internal noise varies in direct proportion to the magnitude of the stimulus (i.e., the exponent P in Eq. (2) is equal to unity). The effects of a proportional noise in Pelli's black-box model can be seen by setting P to unity in Eq. (2):

$$E = k[N_e + N_i + m(N_e + N_i + E)] = \left[\frac{k(1+m)}{1-km}\right](N_e + N_i) = k'(N_e + N_i) \qquad (3)$$

where k' is a constant equal to $k(1+m)/(1-km)$. A comparison of Eqs. (1) and (3) shows that estimates of calculation efficiency will be affected by both the magnitude of an observer's proportional internal noise (m) and the efficiency of an observer's deterministic computations (k), confounding these two factors in the context of Pelli's black-box model. Pelli (1990) is explicit about this aspect of the model, and assumes that any proportional noise stems from the stochastic properties of the contrast–invariant calculation (i.e., random changes in the calculation across trials). However, proportional noise may arise from sources other than a noisy calculation (Lillywhite, 1981). Thus, we consider contrast-dependent noise as a separate source of internal noise in the model.

1.2.2. Response consistency

Green (1964) and others (Burgess & Colborne, 1988; Spiegel & Green, 1981) devised a method of measuring internal noise that is independent of the deterministic operations of the observer. The technique is called *response consistency*, and it takes advantage of the fact that

internal noise will cause trial-by-trial variability in an observer's internal responses to identical stimuli. Consider again a task where observers must identify a signal presented in external noise. If the signal and noise shown on every trial of the experiment were recorded, and then the exact same trial-by-trial sequence was shown a second time, the task would be physically identical in both passes through the experiment. The responses of a noiseless observer would also be identical on each corresponding trial in the sequence, regardless of whether a response was correct or incorrect. However, for an observer with internal noise, there would be response inconsistency between corresponding identical trials in the two passes, and the degree of inconsistency depends on the ratio of internal to external noise at the level of the decision variable (Burgess & Colborne, 1988). Although both contrast–invariant and contrast-dependent noise will produce response inconsistency, the contrast-dependent component of an observer's internal noise can be estimated by measuring response consistency under conditions of high external noise (where the contribution of the contrast–invariant internal noise will be negligible; see Fig. 2). In the context of perceptual learning, response consistency offers a way of measuring changes in contrast-dependent internal noise as a function of learning independently of changes in the deterministic aspects of an observer's calculations.

1.3. Measuring the calculation

One of the limitations of the techniques described above is that they only provide a gross index of the goodness of observers' calculations, leaving the particular nature of their computations unspecified. For example, two observers that exhibit the same degree of improvement in calculation efficiency with practice may base their decisions on very different aspects of the same stimulus. The *response classification* technique developed by Ahumada and his colleagues (Ahumada & Lovell, 1971; Beard & Ahumada, 1998; Watson & Rosenholtz, 1997) offers a way of addressing this problem. Consider once again an identification task where an observer must identify a noisy stimulus as one of two possible patterns, S_1 or S_2. On some trials, an observer will incorrectly classify the stimulus. For example, the observer may respond that the signal was S_1 when, in fact, S_2 was shown. If the signal was embedded in a large amount of external noise, there are two possible reasons for this mistake. One possibility is that internal contrast-dependent noise was high, causing the observer to misclassify the stimulus. A second possibility is that the external noise was distributed in such a way to make the stimulus look more like S_1 than S_2. As long as the internal contrast-dependent noise is not excessively high, the external noise will lead to misclassification on many trials. The noise fields shown on each trial can be recorded and sorted into a 2×2 stimulus–response matrix. After many trials, these noise fields can be averaged in each signal–response category and summed across categories in such a fashion as to produce a *classification image*. The classification image is a map that shows the locations in the stimulus that have affected an observer's responses during the experiment. More specifically, it shows the correlation between the noise magnitude at each location in the stimulus and an observer's responses throughout the experiment. It has also been shown (Ahumada, 2002; Murray, Bennett, & Sekuler, 2002) that the classification image will be proportional to the linear template (i.e., linear calculation) used by an observer in many tasks (including the identification task described above). It is not uncommon to model the deterministic operations performed by an observer in a task

as a linear filter or template that is compared to the stimulus (Abbey, Eckstein, & Bochud, 1999; Ahumada, 2002; Bochud, Abbey, & Eckstein, 2000; Eckstein, Ahumada, & Watson, 1997; Legge et al., 1987; Levi, Klein, & Carney, 2000; Lu & Dosher, 1998, 1999, 2001a; Murray et al., 2002; Watson, 1998). Such *template matching* models have been successful in characterizing human performance in a wide variety of visual pattern discrimination tasks.

The response classification technique has been applied successfully to a wide range of auditory and visual discrimination tasks (e.g., Abbey et al., 1999; Abbey & Eckstein, 2002; Ahumada, 1996; Ahumada & Lovell, 1971; Eckstein, Shimozaki, & Abbey, 2002; Gold, Murray, Bennett, & Sekuler, 2000; Levi & Klein, 2002; Neri, Parker, & Blakemore, 1999; Neri & Heeger, 2002; Sekuler, Gold, Gaspar, & Bennett, 2001; Watson & Rosenholtz, 1997). In the context of perceptual learning, the response classification technique offers a way of specifying the nature of the changes that occur in an observer's calculation over the course of training.

1.4. Overview

The goal of the work presented here was to apply the signal detection model and external noise methods described above to illuminate the effects of perceptual learning. Experiments 1 and 2 involved using external noise masking (Experiment 1) and response consistency (Experiment 2) to partial out the relative contributions of stochastic and non-stochastic factors to perceptual learning in two visual pattern discrimination tasks: human face and abstract texture identification. Human faces were used as stimuli because they reflect a complex, real-world perceptual learning problem that the visual system must solve throughout the lifespan. The rationale for using abstract texture patterns in addition to faces stems from recent work regarding the mechanisms that mediate human face perception. Some evidence suggests that there are cortical mechanisms specifically devoted to face perception (Kanwisher, McDermott, & Chun, 1997; Perrett, Hietanen, Oram, & Benson, 1992), whereas other evidence suggests the apparent special status of faces is rooted in expertise (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999; Gauthier, Skudlarski, Gore, & Anderson, 2000; Gauthier & Tarr, 1997). If learning to recognize novel faces is mediated by face-specific mechanisms, the effects of learning found with faces may not apply to other kinds of patterns. The texture patterns were chosen to address this issue because they are dissimilar to faces.

The results of Experiment 1 showed that calculation efficiency increased by as much as a factor of 4 across learning sessions, with no corresponding decreases in equivalent input noise. The results of Experiment 2 showed little or no significant changes in response consistency in high levels of external noise. Taken together, these experiments indicated that learning of both faces and textures was mediated by purely deterministic changes in the efficiency of observers' calculations. Experiment 3 used the response classification technique to explore the specific kinds of changes that took place in observers' calculations with learning, and showed that observers' strategies became more similar to the ideal strategy as learning took place.

2. Experiment 1

The main goal of Experiment 1 was to discriminate between increased calculation efficiency and decreased contrast–invariant internal noise as possible sources of the improvements in

performance that take place with learning in a pattern discrimination task. We measured signal identification energy thresholds in a range of external noise power spectral densities, across a series of learning sessions. Measures of contrast–invariant internal noise and calculation efficiency were then derived from these data, which allowed us to trace the changes in these quantities as learning took place.

2.1. Method

2.1.1. Apparatus

Stimuli were displayed using a Macintosh G3 computer on a 13 in. Apple high resolution RGB color monitor. The monitor displayed 640 × 480 pixels, which subtended a visual angle of 12.9° × 9.6° from the viewing distance of 100 cm, at a frame rate of 67 Hz (non-interlaced). Luminance calibrations were performed with a Hagner Optikon universal spot photometer, and the calibration data were used to build a 1779-element look-up table (Tyler, Chan, Liu, McBride, & Kontsevich, 1992). The experiment was conducted in the MATLAB programming environment (version 5.1), using in-house software and the extensions provided by the Psychophysics Toolbox (Brainard, 1997) and the Video Toolbox (Pelli, 1997). Luminance on the display ranged between 0.3 and 80.2 cd/m^2. Average luminance was 28.8 cd/m^2.

2.1.2. Signals

There were two classes of signals used in the experiment. The first class of signals was digital images (256 × 256 pixels in size) of human faces that were constructed using Adobe Photoshop (version 3.0) and MATLAB. The second class of signals was randomly generated band-pass filtered Gaussian noise fields (also 256 × 256 pixels in size) generated using MATLAB (see below for details about the face and texture images). The values in each image represented the contrast (c_i) at pixel location i, defined by Eq. (4):

$$c_i = \frac{l_i - L}{L} \tag{4}$$

where L is average luminance and l_i is the luminance of the ith pixel. Each image file was normalized so that root-mean-square (RMS) contrast of the image equaled 1. RMS contrast is defined as

$$c_{\mathrm{RMS}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} c_i^2} \tag{5}$$

where n is the number of pixels in the image.

2.1.2.1. Faces. The face stimuli consisted of the images of five male and five female Caucasian faces that we used in a previous set of experiments (see Gold, Bennett, & Sekuler, 1999a for a more detailed discussion of how the face images were constructed). The images were cropped to show only the inner portion of each face, eliminating non-facial cues such as hair and ears. The shape of the visible region of each face was elliptical, and the size and height:width ratio were constant across all stimuli (198:140 pixels; 4.0° × 2.9°). The faces were centered within a 256 × 256 pixel (5.25° × 5.25°) background of average luminance (see left side of Fig. 3).

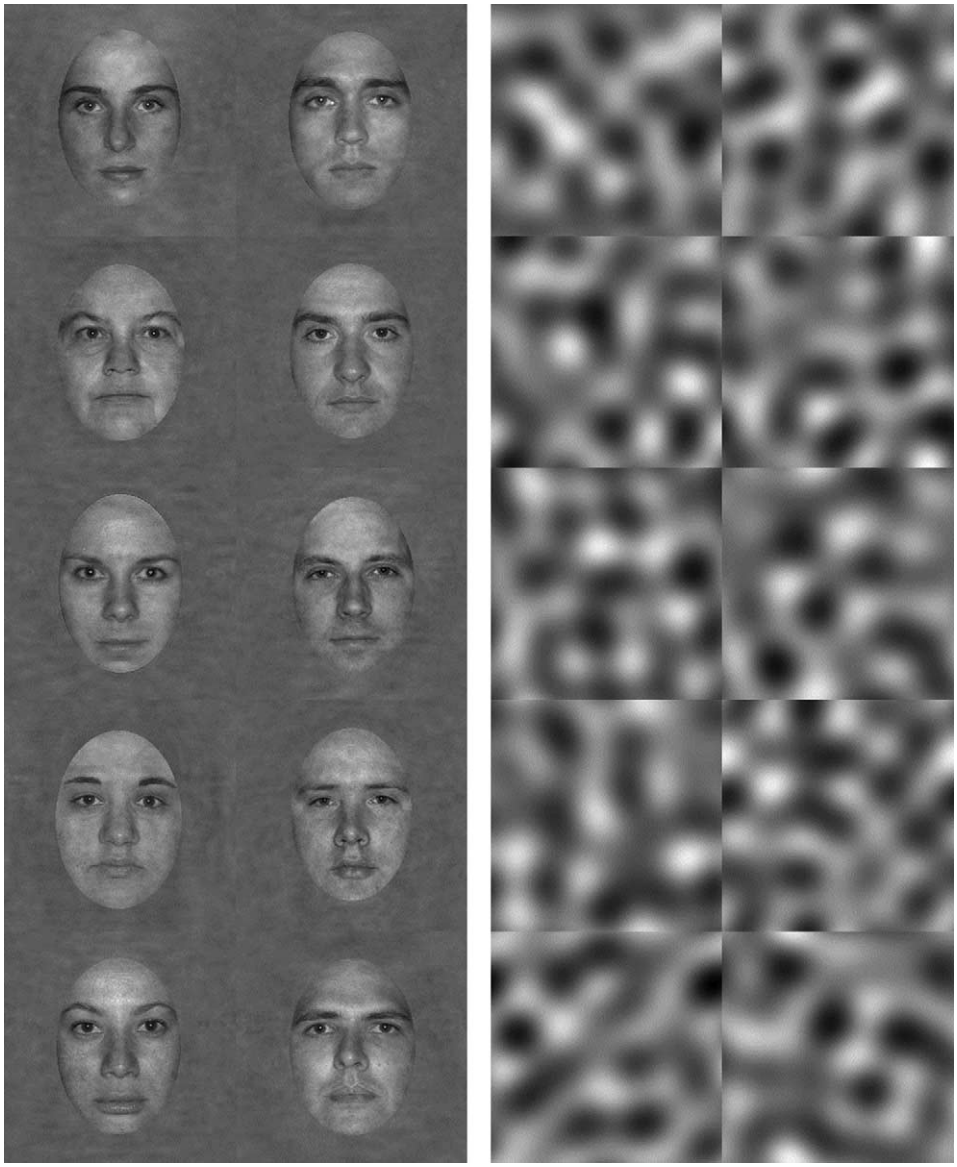

Fig. 3. The face (left) and texture (right) stimuli used in Experiments 1 and 2.

2.1.2.2. Textures. The texture patterns were produced by randomly generating 10 Gaussian white noise fields that were the same size as the face stimuli (256 × 256 pixels). The noise fields were converted into the spatial frequency domain, and filtered with a 2–4 cycle per image (c/image) circularly symmetric rectangular filter. The amplitude of all frequencies outside of the 2–4 c/image pass-band were set to zero, and the amplitude within the pass-band remained unchanged. The images were then converted back into the spatial domain. The filtering produced a set of 10 unique blob-like textures, shown on the right side of Fig. 3.

2.1.3. Display noise

The external noise added to the signal at each pixel was created by drawing a random sample from an independent Gaussian distribution of contrast values, with a mean of 0 and a variance as required by the condition. The noise on each trial was static (i.e., did not vary temporally during the course of a trial), white (i.e., pixels were independent of each other) and was matched to the size of the signal (256×256 pixels). Values beyond ± 2 standard deviations from the mean were discarded and replaced by random samples from the remaining contrast values. In each task, thresholds were measured in five different levels of external noise power spectral density. For the face task, the external noise power spectral density levels were: 0.04, 0.20, 1.02, 5.11 and 25.55×10^{-6} degree^2 (see Section 2.1.7 and Eq. (7) for a definition of noise power spectral density). Pilot studies suggested that equivalent noise was higher for the texture identification task, so the lowest external noise level was removed and replaced by a higher noise level of 51.10×10^{-6} degree^2. A unique noise field was generated on every trial.

2.1.4. Viewing conditions

Viewing was binocular through natural pupils, and a headrest/chinrest stabilized the observer's head throughout the session. The computer monitor supplied the only source of illumination during the experiment.

2.1.5. Human observers

All participants had normal or corrected-to-normal visual acuity (self-reported). Two observers participated in each task, with one observer (AMC) participating in both tasks. One observer was an author (JMG) and the remaining two observers were naive to the purposes of the experiment.

2.1.6. Tasks and procedure

Performance was measured using a single-interval, 1-of-10 identification task. The signal energy and power spectral density of the noise were varied according to the procedures detailed in the threshold estimation section below. Observers were familiarized briefly with high contrast versions of the stimuli before the beginning of the experiment. At the start of each trial, a small fixation point appeared at the center of the screen (3×3 pixels in size), and a brief tone indicated a trial could commence with a mouse click. After the mouse was clicked, the stimulus (signal + noise combination) appeared for 34 frames (approximately 500 ms). Next, the display was set to average luminance, and after a brief 100 ms pause, 100×100 pixel high contrast thumbnail versions of the 10 possible signals appeared on the screen surrounding the region where the stimulus had been displayed. Observers identified the stimulus by clicking the mouse on the appropriate image. Once an image was chosen, the displays were cleared and set to average luminance. Auditory feedback was given after each trial to indicate the accuracy of the response.

2.1.7. Threshold estimation

Identification thresholds at each of the five levels of external noise were measured by varying signal energy across trials. Signal energy was manipulated according to the method of constant stimuli. Pilot studies identified several signal energy levels that spanned the threshold range

for a typical unpracticed observer in each level of external noise power spectral density. Signal energy E is defined as

$$E = (c_{\mathrm{RMS}})^2 na \tag{6}$$

where n is the number of image pixels and a is the area of a single pixel, in degrees of visual angle squared. Noise power spectral density N is defined as

$$N = \sigma^2 a \tag{7}$$

where σ is the standard deviation of the noise, expressed in values of contrast. Signal energy levels were adjusted after each session for each observer as required by their rate of learning. Before each trial, a signal was chosen randomly to appear within the stimulus interval. There were 31 trials per stimulus energy level within each session, yielding a total of 775 trials (31 trials × 5 signal levels × 5 noise levels). The level of external noise, signal energy and the identity of the signal were chosen randomly on each trial. Each session was completed without breaks and lasted about 1 h. Only one session was completed each day. Observers in the face identification task completed six sessions within 10 days. Observers in the texture identification task completed four sessions within 7 days.

Psychometric functions were estimated by maximum-likelihood fits to a Weibull function. Threshold was defined as the signal energy yielding 50% correct responses. Confidence intervals for the threshold estimates were calculated by bootstrap simulations (Efron & Tibshirani, 1993). Each simulation consisted of at least 500 simulated data sets.

2.1.8. Ideal observer

The ideal decision rule for the 1-of-M pattern identification task described above is

$$\underset{j=1,\ldots,m}{\operatorname{argmax}} \left(\sum_{i=1}^{n} T_{ij} R_i \right) \tag{8}$$

where m is the number of possible signals, n is the number of pixels in each signal, T_{ij} is the ith pixel in the jth noise-free normalized signal, and R_i is the ith pixel in the noisy stimulus. This rule amounts to choosing the signal that yields the highest cross-correlation between the stimulus (i.e., signal + noise combination) and each of the M possible noise-free signal matrices (Green & Swets, 1966; Tjan et al., 1995).[1] Ideal observer thresholds were obtained in all conditions through Monte Carlo simulations, in which each noise-free signal matrix was compared to the stimulus at a range of signal energy values for each corresponding noise level tested with human observers. Ideal thresholds were estimated from psychometric functions that were fit to the data (using the procedure described above) from at least 10,000 simulated trials.

2.1.9. Equivalent input noise and calculation efficiency

Recall that an observer's equivalent input noise and calculation efficiency are estimated by measuring signal identification energy thresholds across a range of external noise power spectral density levels. Eq. (1) is fit to the thresholds, with the negative x-intercept, N_{i}, as the estimate of contrast–invariant internal noise and the slope, k, as an index of efficiency. As with the external noise, N_{i} is expressed in units of power spectral density.

It can be shown (Tjan et al., 1995) that the ideal observer's signal identification energy threshold in the tasks described above is a linear function of noise power spectral density, i.e.,

$$E_{\text{ideal}} = k_{\text{ideal}} N_{\text{e}} \tag{9}$$

where N_{e} is the power spectral density of the external noise. The slope parameter k_{ideal} varies with the set of signals and is directly related to the intrinsic difficulty of the task (i.e., the similarity of the signal matrices). The human observer's calculation efficiency η is defined as

$$\eta = \frac{k_{\text{ideal}}}{k} \tag{10}$$

Linear fits to both ideal and human noise masking functions were estimated by maximum-likelihood minimization, and bootstrap simulations (Efron & Tibshirani, 1993) provided confidence intervals for the fitted parameters (minimum 500 simulated experiments).

2.2. Results and discussion

Fig. 4 shows the 50% correct threshold signal energy levels as a function of external noise power spectral density for each observer in the face (top panels) and texture (bottom panels) identification tasks. Each symbol corresponds to a single threshold. The filled symbols correspond to the first half of the experiment, the open symbols the last half of the experiment. The r^2 values for each observer's maximum likelihood fit to Eq. (1) in each session (smooth curves) are summarized in Table 1, along with F statistics that indicate the significance of the fits. The average r^2 values across sessions for all observers were above .95, indicating that the effect of noise on thresholds in this task was well characterized by Eq. (1).

Table 1
Statistics for the Eq. (1) fits to the threshold data from Experiment 1

Session	AMC			CGB		
	r^2	F	p	r^2	F	p
Face identification						
1	.9681	91.07	3.92e−05	.9988	2,500.41	1.06e−08
2	.9968	951.75	1.18e−07	.9990	3,043.44	6.46e−09
3	.9969	976.75	1.10e−07	.9962	789.40	1.86e−07
4	.9955	664.08	2.88e−07	.9981	1,649.09	2.98e−08
5	.9999	23,762.60	3.80e−11	.9973	1,110.79	8.00e−08
6	.9957	703.09	2.50e−07	.9926	407.19	9.74e−07
Mean	*.9922*	*4,524.89*	*6.66e−06*	*.9970*	*1,583.39*	*2.15e−07*
Texture identification						
1	.9874	234.58	3.82e−06	.9726	71.05	1.90e−04
2	.9994	4,913.76	1.95e−09	.9969	954.28	1.17e−07
3	.9976	1,248.09	5.98e−08	.9716	106.57	2.67e−05
4	.9273	38.28	3.14e−4	.8976	26.30	7.51e−04
Mean	*.9779*	*1,608.68*	*7.94e−05*	*.9599*	*289.55*	*2.42e−04*

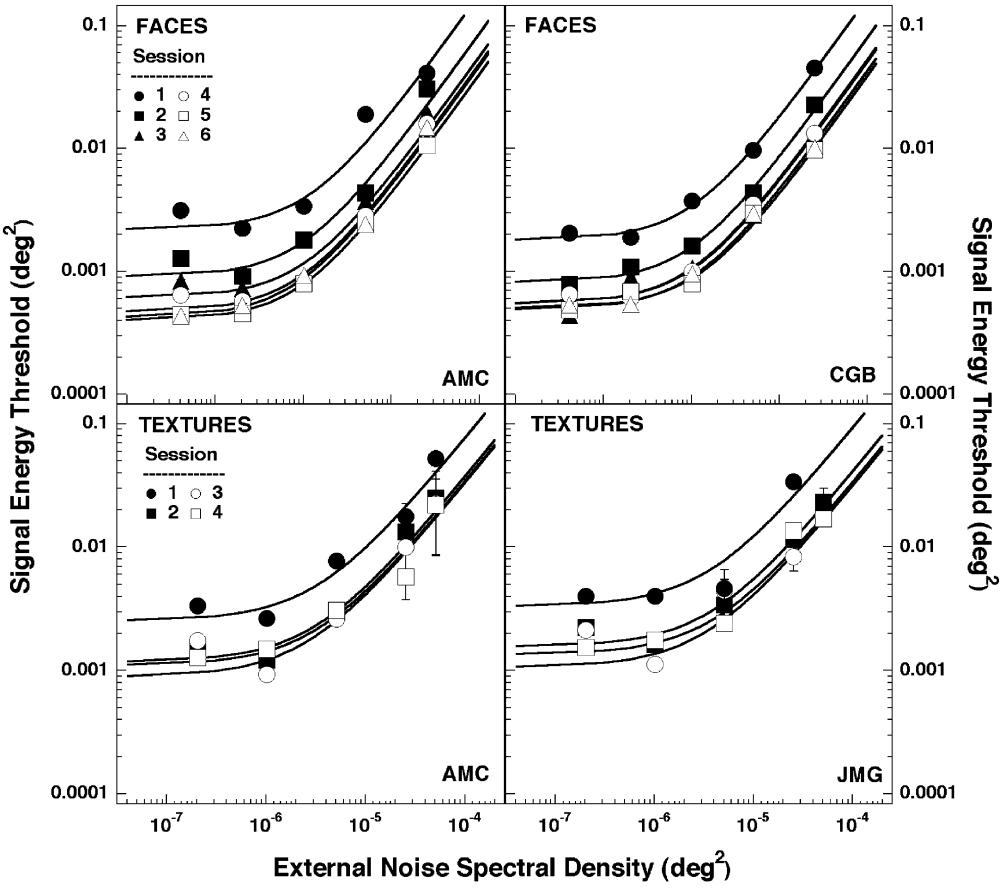

Fig. 4. Noise masking functions for the observers in the face (top panels) and texture (bottom panels) identification tasks from Experiment 1. Each panel plots an individual observer's signal energy thresholds as a function of external noise power spectral density. The filled symbols correspond to the sessions in the first half of the experiment, the open symbols the last half of the experiment (see legend). Solid lines correspond to maximum-likelihood fits to Eq. (1). Error bars on each symbol correspond to ±1 standard deviation. Often, the error bars are smaller than the symbols.

Inspection of Fig. 4 reveals a clear trend across sessions for all observers. Namely, the height of the functions shifts down uniformly across external noise levels with practice. However, the knee of the functions remains constant across sessions. This kind of shift in the noise masking function is consistent with the effects of increased calculation efficiency with no change in contrast–invariant internal noise. These effects can be seen more clearly in Fig. 5, which plots calculation efficiency (left panel) and equivalent input noise (right panel) for each observer, as a function of practice. Although one observer (AMC) did show a small but statistically significant decrease in the estimate of contrast–invariant internal noise from the first to the last session in the face identification task ($z = 2.89$, $p < .01$), there were no statistically significant changes across sessions for any of the other observers in either task. In contrast, all observers showed highly significant increases in calculation efficiency across

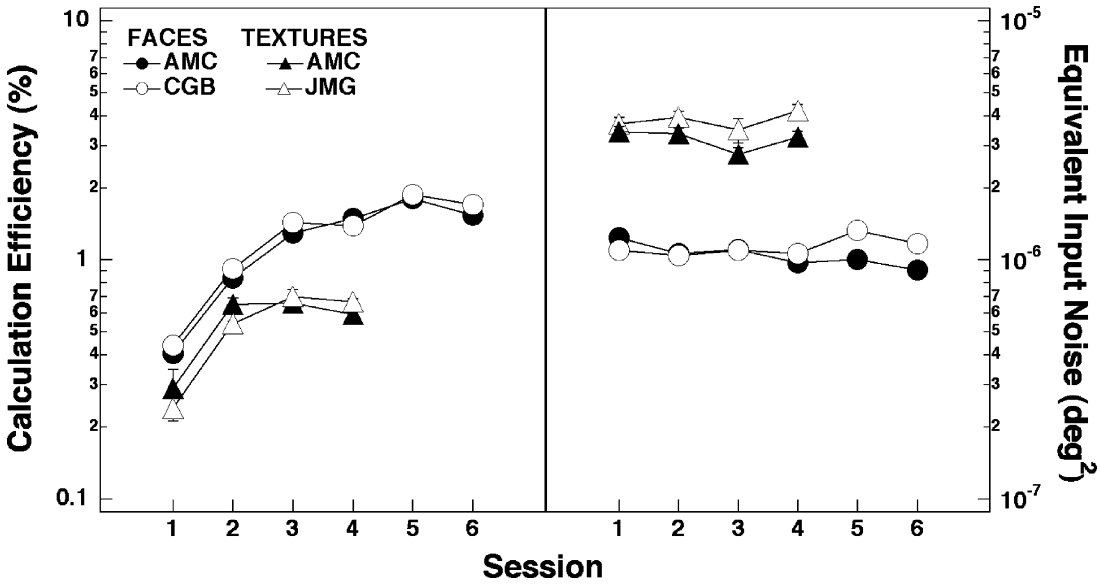

Fig. 5. Calculation efficiency (left) and equivalent input noise (right) as a function of experimental session for the observers in the face (circles) and texture (triangles) identification tasks in Experiment 1. Error bars on each symbol correspond to ±1 standard deviation.

sessions ($z = 4.3$, $p < 1.0\text{e}{-}05$ from the first to the last session for all observers), with efficiency increasing by about a factor of 4 in the face identification task (AMC: 4.4; CGB: 4.3) and by about a factor of 2–3 in the texture identification task (AMC: 2.3; JMG: 2.9). The absolute levels of calculation efficiency in the final sessions ranged between ∼1 and 2%, values that are similar to previous measures of efficiency for familiar face identification (Gold et al., 1999a).

Thus, the results from Experiment 1 indicate that learning served to increase calculation efficiency but had virtually no effect upon contrast–invariant internal noise. However, recall that the quantity of calculation efficiency incorporates limitations imposed by both the efficiency of the deterministic aspects of observers' calculations and the magnitude of any contrast-dependent internal noise. The data from Experiment 1 do not allow us to discriminate between these two constraints. Teasing these factors apart was the subject of Experiment 2.

3. Experiment 2

In Experiment 2, we used double-pass response consistency in a high level of external noise to isolate the effects of contrast-dependent internal noise on learning in our face and texture identification tasks. If decreases in contrast-dependent internal noise contributed to the increases in calculation efficiency found in Experiment 1, observers should become more consistent as learning takes place.

3.1. Methods

3.1.1. Stimuli

The stimuli used were the same sets of 10 faces and 10 textures used in the previous experiments. In each task, signal energy identification thresholds were measured in the highest levels of external noise used in the previous experiments (faces: 25.55×10^{-6} degree^2; textures: 51.10×10^{-6} degree^2). $N/2$ unique noise fields were generated for each experimental session, where N is the number of trials within a given session. The sequence of signal identities, signal energy levels, and seeds used to generate the noise fields during the first half of each session were saved before every trial to allow for the exact reproduction of the same sequence of stimuli during the second half of the session.

3.1.2. Procedure

Signal energy was varied across trials during the first half of each session according to two interleaved adaptive staircases. Staircases were used to remove the need to manually adjust the contrast levels of the signals in between sessions (as was the case with the method of constant stimuli used in Experiment 1). Two interleaved staircases were used to obtain measurements that spanned the range of the psychometric function. Signal energy levels were chosen that coarsely sampled a range of several log units around a rough threshold estimate (in step sizes of approximately 0.2 log units). The staircase shifted through these levels according to the accuracy of the observer's responses. One of the staircases used a 1-up-1-down rule and the other staircase used a 1-up-2-down rule. The staircases maintained this process throughout the first half of each experimental session, consisting of 200 trials per staircase (400 trials total). The second half of the session consisted of an exact replication of the first half of the session (i.e., an exact pixel-by-pixel reproduction of the sequence of trials shown during the first half of the session was shown again during the second half of the session), yielding a total of 800 trials per session. Note that this repetition meant that the signal energy was altered across trials in a manner that was contingent upon the observer's response during the first but not the second half of each session. Despite this difference, none of the observers reported being aware of this aspect of the experiment. Thresholds were estimated by fitting psychometric functions to the combined data from the two staircases. Each session was completed without breaks and lasted about 1 h. Only one session was completed each day. Each observer completed six sessions within 10 days. The observers were unaware that the first and second halves of each session were identical.

3.1.3. Observers

Two observers participated in the face identification task and two in the texture identification task. All observers had normal or corrected-to-normal visual acuity and were naive to the purposes of the experiment. None of the observers had participated in Experiment 1.

3.2. Results and discussion

Signal energy thresholds for all observers in the face (left panel) and texture (right panel) identification tasks are plotted as a function of session in Fig. 6. These data show that there was a clear improvement across sessions for all observers. In fact, three of the four observers exhibited

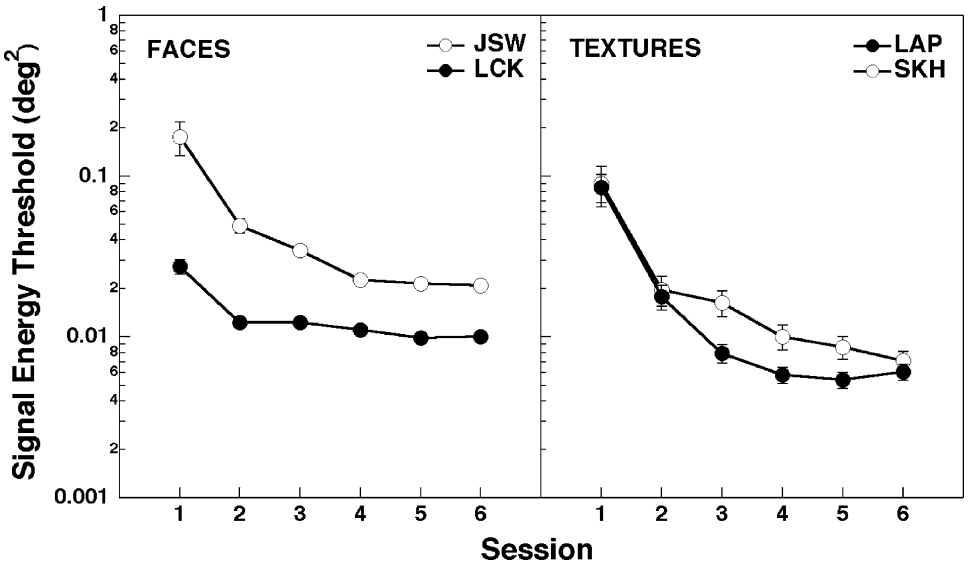

Fig. 6. Face (left panel) and texture (right panel) identification thresholds plotted as a function of session for the two observers in Experiment 2. The external noise power spectral density was set to the highest levels used in the corresponding tasks from Experiment 1. Error bars on each symbol correspond to ±1 standard deviation.

a markedly higher amount of improvement across sessions than was found in Experiment 1: observer JSW's contrast energy thresholds decreased by a factor of 8.2 in the face identification task, and observers LAP's and SKH's contrast energy thresholds decreased by factors of 14 and 12.5, respectively, in the texture identification task. A key difference between Experiments 1 and 2 was that the presentation of the different noise levels was completely randomized across trials in Experiment 1, yet was fixed throughout Experiment 2; similarly, signal energy was completely randomized across trials in Experiment 1, yet was correlated across trials in Experiment 2 by virtue of the use of adaptive staircases and the repetition of trials in the second half of the session. Thus, one possibility is that greater stimulus certainty tended to promote greater learning in Experiment 2. Regardless of the difference in absolute performance between Experiments 1 and 2, all observers showed large improvements with practice.

Fig. 7 shows the corresponding results of the consistency analyses for the face (top panels) and texture (bottom panels) identification tasks. Each panel plots the percentage of correct responses at each stimulus level as a function of the percentage of agreement between corresponding responses made in the two passes through the session for an individual observer. Each symbol corresponds to a single stimulus level within a given session. In all panels, the closed symbols correspond to the first three sessions, the open symbols the last three sessions. To understand how to interpret these plots, consider the performance of an observer with no internal noise. The response of such a noiseless observer would be exactly the same if a stimulus were repeated twice. Similarly, the responses of a noiseless observer would be perfectly correlated if it made two identical passes through an experiment. As a result, a noise-free observer's data in our task would fall along the rightmost side of the plot—performance would vary with signal energy (causing percent correct to vary), but the percentage of *agreement*

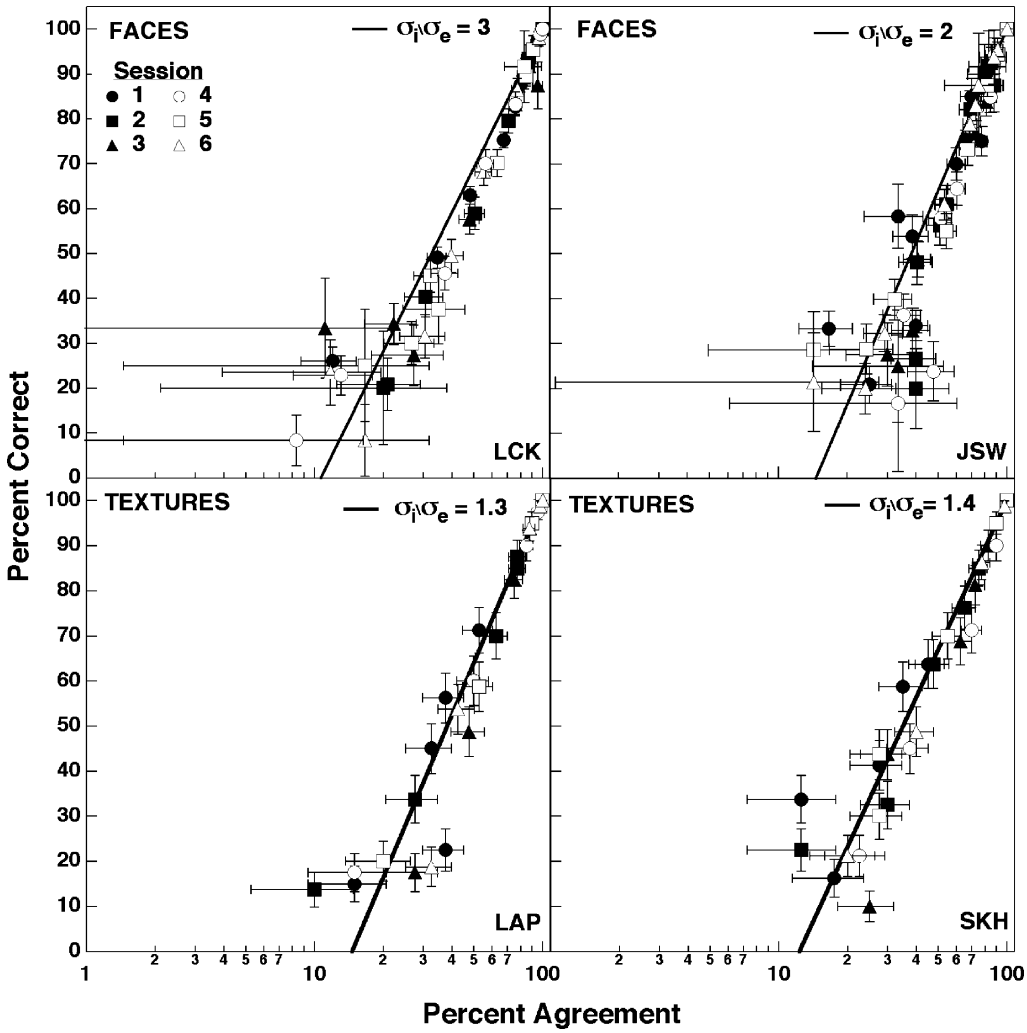

Fig. 7. Response consistency plots for all observers in the face (top panels) and texture (bottom panels) identification tasks from Experiment 2. Each panel plots percent correct as a function of percent agreement for each stimulus level tested. The filled symbols correspond to the first three sessions, the open symbols the last three sessions (see legend). Error bars on each symbol correspond to ±1 standard deviation. The large variation in error bar magnitude is due to the unequal number of trials at each data point (a result of the use of a staircase procedure). Solid lines in each plot correspond to the performance of a simulated observer with an internal/external noise ratio approximately equal to the average of the estimated internal/external noise ratios across sessions.

between corresponding responses in the two passes through the experiment would always be 100% for every stimulus level shown. Now consider the performance of an observer limited by internal noise (e.g., a human observer). The responses of a noisy observer will not be perfectly consistent for repeated trials. Instead, percent agreement will vary with percent correct and the ratio of the standard deviations of the internal (σ_i) and external (σ_e) noises (σ_i/σ_e) (Burgess & Colborne, 1988; Green, 1964). The relationship between percent correct, percent agreement

and σ_i/σ_e is well described by a straight line that passes through the top right corner of a plot such as those shown in Fig. 7. This line follows the form:

$$p_c = m \log_{10}\left(\frac{p_a}{100}\right) + 100 \tag{11}$$

where p_c is percent correct performance at a given level of signal energy, p_a is the percent agreement between the two passes through the experiment, and m is a free parameter corresponding to the slope of the line. The parameter m will vary systematically with σ_i/σ_e. Specifically, as σ_i/σ_e decreases, p_a for any given level of p_c will increase, causing the data in plots like those shown in Fig. 7 to shift to the right. Thus, we would expect to see a rightward shift in the data across sessions in Fig. 7 if contrast-dependent internal noise were responsible for some or all of the decrease in thresholds seen in Fig. 6. However, the consistency analyses show that there were no systematic changes in percent agreement across sessions for any of the observers. Instead, for each observer all of the data appear to fall along a single line. This conclusion was verified by fitting the data from each session to Eq. (11) to determine the slope parameter m. Monte Carlo simulations were then used to convert m into σ_i/σ_e estimates across sessions for each observer. Specifically, a simulated observer[2] built with different ratios of σ_i/σ_e was implemented to determine the relationship between σ_i/σ_e and m. In the tasks reported here, this relationship is well described by a function of the form

$$\frac{\sigma_i}{\sigma_e} = \alpha + \gamma_1 e^{-\beta_1 m} + \gamma_2 e^{-\beta_2 m} \tag{12}$$

where α, γ_1, γ_2, β_1 and β_2 are fitted parameters. Maximum-likelihood minimization was used to fit Eq. (11) to the human data, and bootstrap simulations (Efron & Tibshirani, 1993) were used to produce confidence intervals for m (minimum 500 simulated experiments). The fitted parameters[3] for Eq. (12) were then used to calculate σ_i/σ_e.

The σ_i/σ_e estimates for each observer are plotted as a function of session in Fig. 8. These data show that σ_i/σ_e ranged between ~1 and 1.5 for the texture identification task and about 1–4 for the face identification task. There are several noteworthy aspects to these data. First, the fact that σ_i/σ_e was significantly greater than zero in high-contrast external noise is consistent with previous results (Burgess & Colborne, 1988; Green, 1964; Spiegel & Green, 1981) and implies that it is appropriate to include a contrast-dependent internal noise component in the black-box model described above.

Second, although the data are somewhat noisy, there appear to be no systematic decreases in σ_i/σ_e with practice for any of the observers. Of course, it could be that, given the relatively small number of trials in each session, the response consistency measure is simply not sensitive enough to detect the magnitude of change necessary to produce the observed changes in performance that took place with learning. To test this possibility, we first estimated how much of a decrease in σ_i/σ_e would have been necessary to produce the total changes we observed in the slopes of the noise masking functions across learning sessions measured in Experiment 1. That is, we estimated how much σ_i/σ_e would have had to have decreased if all of the improvements in performance with learning were due to changes in internal contrast-dependent noise. This estimate corresponds to measuring the multiplicative constant m from Eq. (3). However, recall that the contribution of the contrast–invariant internal noise (N_i) will be negligible under conditions of high external noise. In addition, the contribution of the signal energy E will also

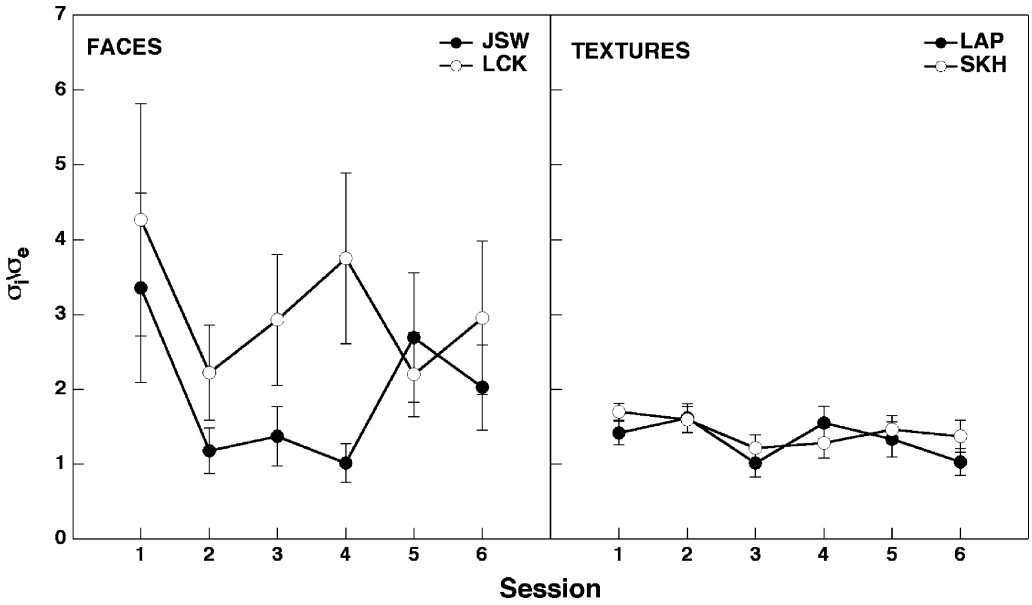

Fig. 8. Internal/external noise standard deviation ratio (σ_i/σ_e) estimates for all observers in the face (left panel) and texture (right panels) identification tasks from Experiment 2. Each panel plots σ_i/σ_e as a function of session. The different symbols correspond to different observers (see legend). Error bars on each symbol correspond to ± 1 standard deviation.

be negligible under thresholds conditions. So, under these conditions, Eq. (2) can be simplified to the following:

$$E = k(N_e + mN_e) = k(1+m)N_e = k'N_e \tag{13}$$

where $k' = k(1+m)$, the slope of the noise masking function (or, equivalently, $m = k'/k - 1$). Let k_a, k'_a and m_a equal k, k' and m after learning, and k_b, k'_b and m_b equal k, k' and m before learning. Then,

$$m_a = \frac{k'_a}{k_a - 1} \tag{14a}$$

$$m_b = \frac{k'_b}{k_b - 1} \tag{14b}$$

If we assume that all of the changes in k' are solely due to changes in m, then $k_a = k_b$. Solving for k_b in Eq. (14b) and substituting for k_a in Eq. (14a), we have

$$m_a = (m_b + 1)\left(\frac{k'_a}{k'_b}\right) - 1 \tag{15}$$

Because m_b is given by the squared empirical estimate of σ_i/σ_e obtained in the first session (recall that σ_i/σ_e was expressed as a ratio of standard deviations, whereas m is expressed as a ratio of variances), k'_a and k'_b can be computed from the threshold estimates (E) obtained in the first and last sessions, using the relationship $k' = E/N_e$, as defined in Eq. (13).

Table 2
Parameters and statistics for the response consistency sensitivity analyses from Experiment 2

	k'_a/k'_b	m_b	m_b *SE*	Empirical m_a	Empirical m_a *SE*	Predicted m_a	Empirical vs. predicted m_a (*p*-value)
Face identification							
JSW	0.3653	11.2587	2.5400	4.1099	1.1418	0.4619	0.0014
LCK	0.1193	18.1911	3.0992	8.7415	2.0548	6.0112	0.1839
Texture identification							
LAP	0.0710	2.0113	0.3118	1.0640	0.3539	−0.7863	1.71e−07
SKH	0.0796	2.8965	0.2207	1.8873	0.4292	−0.6897	1.93e−09

We used Eq. (15) to compute the predicted value of m_a for each observer in each condition of Experiment 2. We then compared these values to the empirical estimates obtained in the last session of the experiment. If response consistency was not a sensitive enough measure to detect changes in internal noise, the empirical and predicted values of m_a should not be significantly different from one another. The results of this analysis are shown in Table 2. For each observer, the first column in Table 2 shows the ratio of k'_a/k'_b. The next four columns show the empirical values of m_a and m_b, along with the corresponding estimates of standard error for each parameter (computed using bootstrap simulations as described previously). The final two columns show the predicted values of m_a and the results of a statistical comparison between the predicted and empirical values of m_a. For all observers, the predicted values of m_a were lower than the empirical estimates. The differences were most striking in the texture identification task, where the predicted values indicated that, in the absence of a change in efficiency, internal noise would have had to have been *negative* by the final session in order to produce the observed improvements in performance with learning!

With the exception of one observer in the face identification task (LCK), the differences between the predicted and empirical values of m_a were statistically significant.[4] These results suggest that response consistency is sensitive enough to rule out the possibility that reductions in contrast-dependent internal noise account for all of the learning in our face and texture identification tasks. Of course, small changes in contrast-dependent internal noise may have contributed to the improvements that took place with learning. Nevertheless, our results suggest that a significant portion of the learning effects in our experiments were due to deterministic increases in the efficiency of observers' calculations. Within the context of the black-box model outlined above, these results imply that practice should increase the similarity between the calculations used by human and ideal observers. This prediction of the model has been examined more closely in Experiment 3.

4. Experiment 3

The purpose of Experiment 3 was to extend the results of the first two experiments to include a more detailed description of the changes that take place with perceptual learning. Experiments

1 and 2 showed that the effects of perceptual learning in our pattern recognition tasks can be attributed to a gradual increase in the efficiency of the deterministic aspects of observers' calculations. However, the exact nature of those changes remains unspecified. Experiment 3 was designed to address this question by using the response classification technique. As was mentioned in the Introduction, the classification image will be proportional to the observer's template, or the calculation, used by a linear observer in a one of two pattern identification task (see Abbey et al., 1999; Murray et al., 2002). In the context of perceptual learning, response classification can be used trace the changes in an observer's calculations as learning takes place. In addition, human and ideal classification images can be compared to test the prediction that an observer's calculations should become more similar to the ideal as learning occurs (Murray et al., 2002).

4.1. Methods

4.1.1. Stimuli

The signals were two new faces and textures (Fig. 9), generated in the same fashion as described in Experiment 1. The signals were embedded in the highest levels of external noise used in the previous experiments (faces: 25.55×10^{-6} degree^2; textures: 51.10×10^{-6} degree^2).

4.1.2. Observers

Two observers participated in the face identification task and two in the texture identification task. All of the observers were naive to the purposes of the experiment and had normal or corrected-to-normal visual acuity. One observer (LCS) participated in both the face and

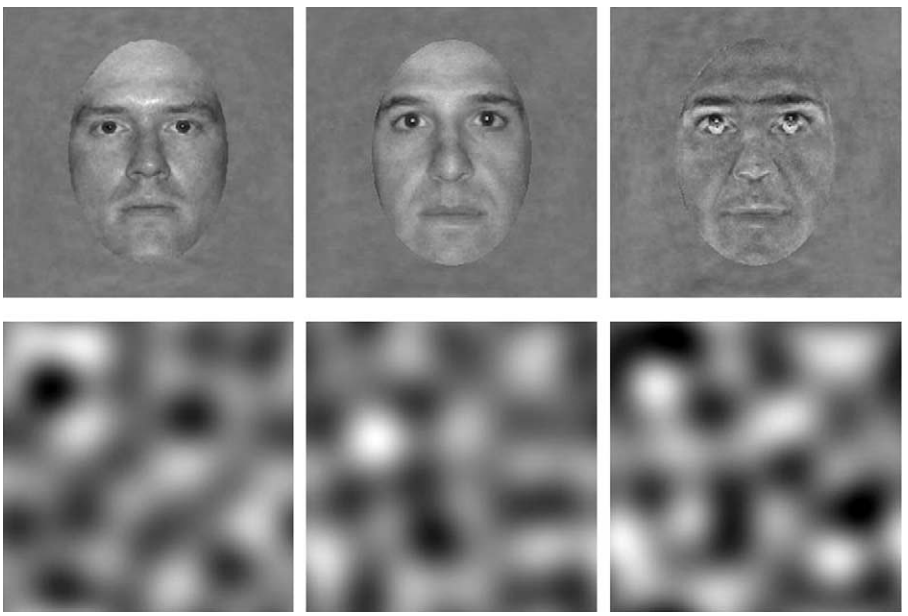

Fig. 9. The face (top, left two panels) and texture (bottom, left two panels) stimuli used in Experiment 3. The rightmost panel in each row is the difference between the two panels to the left. This image is also the classification image for an ideal observer after an infinite number of trials (see text for details).

texture identification tasks. The second observer (SKH) in the face identification task had also participated in the texture identification task in Experiment 2.

4.1.3. Procedure

In each task, signal energy was manipulated across trials according to a 1-up-2-down staircase to maintain roughly constant percent correct performance throughout the session. A unique noise field was generated before each trial. The sequence of signal identities, observer responses and seeds used to generate the noise fields were saved after every trial to allow for subsequent computation of the classification images. Each session consisted of 800 trials that were completed without breaks over the course of about 1 h. Only one session was completed each day. Each observer participated in a total of 12 sessions over the course of 16 days.

4.2. Results and discussion

Signal energy thresholds for the face (left panel) and texture (right panel) identification tasks are plotted as a function of session in Fig. 10. These data show a clear effect of learning, with the majority of learning occurring within the first 4–6 sessions. Over the course of the entire experiment, thresholds declined by about a factor of 3 in the face identification task (LCS: 3.1; SKH: 2.7) and about a factor of 5 in the texture identification task (AJR: 5.2; LCS: 5.1). Note that the magnitude of learning in Experiment 3 was similar to that found in Experiment 1, despite the use of a staircase procedure and single noise level as used in Experiment 2. This result suggests the increased learning effects found in Experiment 2 were not due to these two manipulations. However, the magnitude of learning may also depend upon the number of

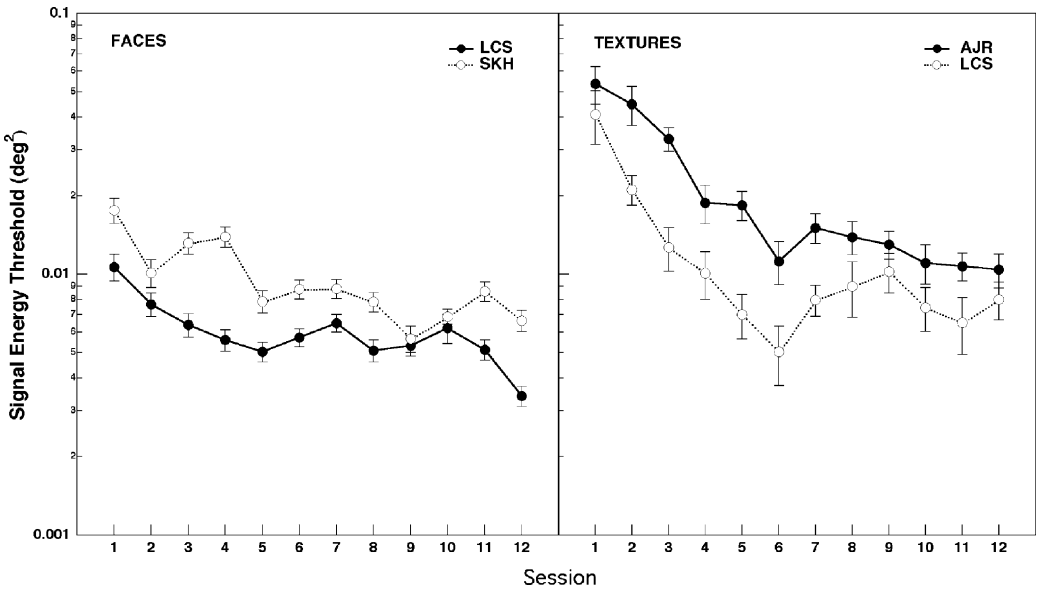

Fig. 10. Signal energy thresholds plotted as a function of session in the face (left panel) and texture (right panel) identification tasks from Experiment 3. The external noise power spectral density was set to the levels used in the corresponding tasks from Experiment 2. Error bars correspond to ±1 standard error.

alternatives (which was significantly smaller in Experiment 3), making it difficult to compare results across Experiments.

The corresponding classification images were computed for each observer in each session. These images were computed by averaging the noise matrices added to the signals point-by-point across trials according to each signal–response combination. In the case of only two signals, there are four signal (S)–response (R) combinations: S1R1, S1R2, S2R1 and S2R2. These four matrices may be combined to form the observer's *raw classification image*, *C*, as follows:

$$C = (\mathrm{S1R2} + \mathrm{S2R2}) - (\mathrm{S1R1} + \mathrm{S2R1}) \tag{16}$$

It has been shown (Ahumada, 2002; Murray et al., 2002) that this method of combining noise fields is statistically optimal in terms of maximizing the signal-to-noise ratio in the classification image for an unbiased linear observer. The contrast of a pixel in the resulting classification image corresponds to the correlation between the contrast of the noise at that pixel across trials and the observer's responses. When displayed as a picture, pixels that are brighter than mean gray indicate a positive correlation between the noise contrast at that pixel and the response 'image #2,' and those darker than mean gray indicate a negative correlation. (N.B.: The choice to positively correlate the noise with the response 'image #2' is a result of arbitrarily choosing the second image in the set to correspond to S2 and R2 in Eq. (16).)

Unfortunately, the classification images computed from individual sessions were too noisy to reveal any visible features. The major reason for this absence of salient features is the small number of trials used to compute each image. Typically, several thousand trials are needed to produce visible features in raw classification images (Abbey et al., 1999; Beard & Ahumada, 1998; Gold et al., 2000). However, the majority of learning in these tasks occurred within the first six sessions, enabling us to collapse the data across sessions within the first and second halves of the experiment. Collapsing across sessions in this fashion increases the number of trials in each classification image by a factor of 6 to 4,800. In addition, statistical analyses were used to test for global changes in the classification images. Specifically, the results from Experiments 1 and 2 imply that the observer's calculations become more similar to the ideal calculation with learning (thus increasing calculation efficiency). Given these results, a strong prediction of a linear version of our black-box model is that the observer's classification images should be more similar to the ideal observer's classification image after learning has taken place. For a one of two identification task, the ideal observer's classification image (template) is simply the difference between the two possible signals (Noreen, 1981). The 'goodness' of a human observer's classification image can be computed by cross-correlating it with the ideal classification image when both are normalized to unit energy (Murray, 2002). The resulting 'normalized cross-correlation' score will range between −1 (perfect negative correlation) and 1 (perfect positive correlation). The ideal classification images for the one of two face and texture identification tasks are shown in the rightmost panels of Fig. 9. The ideal observer's classification image can be thought of as an 'information map,' with the contrast at each pixel corresponding to its relative informativeness. If visual processing can be approximated by a linear template, and if learning improves the efficiency of this processing, then it follows that the calculations are becoming more similar to those used by the ideal observer. This should result in an increase in the similarity between the human and ideal classification images.

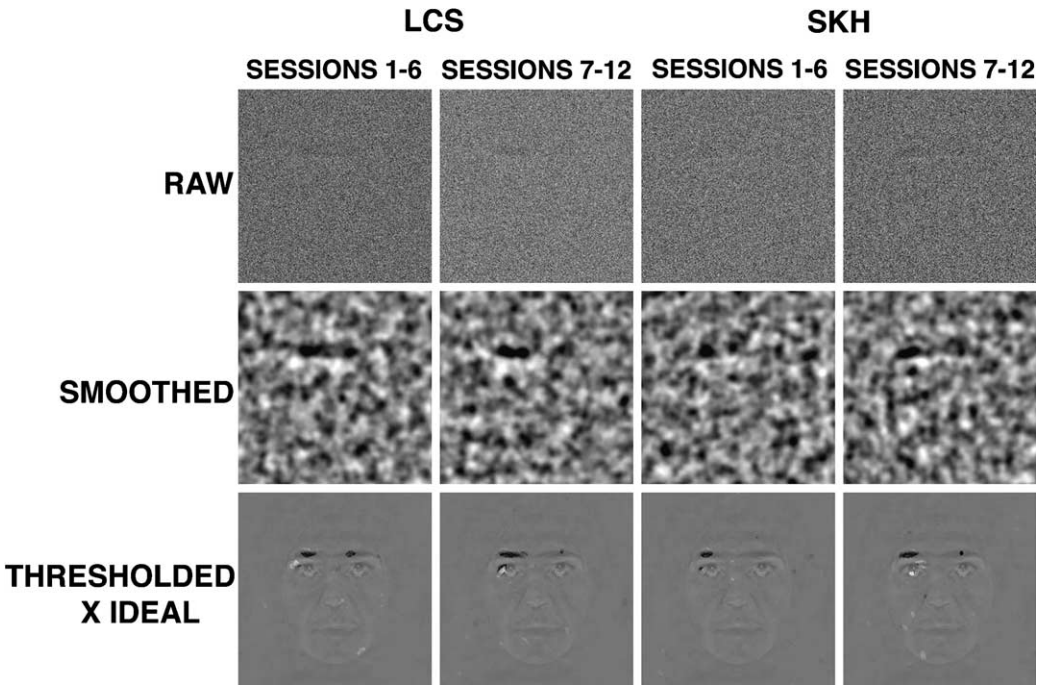

Fig. 11. The raw (top row), smoothed (middle row) and thresholded (bottom row) classification images for both observers in the face identification task from Experiment 3, pooled across either the first or last half of the experiment. In the thresholded images, all of the pixels that fell within a restricted range relative to mean gray in the smoothed classification image were set to a single low value and the remaining pixels were set to unity. The resulting image was then multiplied by the ideal template (see Fig. 9).

The top panels of Figs. 11 and 12 show the classification images from the first and second halves of the experiment for observers in the face and texture discrimination tasks, respectively. Although the images still appear noisy despite collapsing across sessions, close inspection reveals small 'hotspots' emerging from the background of noise. Fig. 13 shows the results of the ideal correlation analysis for each of the classification images. The left panel shows the results for the face identification task, the right panel the texture identification task. The height of each histogram bar corresponds to the correlation for an individual observer in either the first (black bars) or last (gray bars) half of the experiment. These data show that there was an increase in the correlation with the ideal template from the first to the last half of the experiment for each observer in each condition. A *z*-test was performed to determine whether each correlation differed significantly from what one would expect if there was no structure present in the image at all (i.e., compared to zero correlation with the ideal template). The results of this analysis for each image are shown above each bar, expressed as *p*-values. These data show that there was a significant amount of structure present within most of the combined classification image, and the amount of structure increased during the second half of the experiment for all of the observers. A second *z*-test was performed to determine whether the increase in correlation from the first to the last half of the experiment was statistically significant for each observer in each

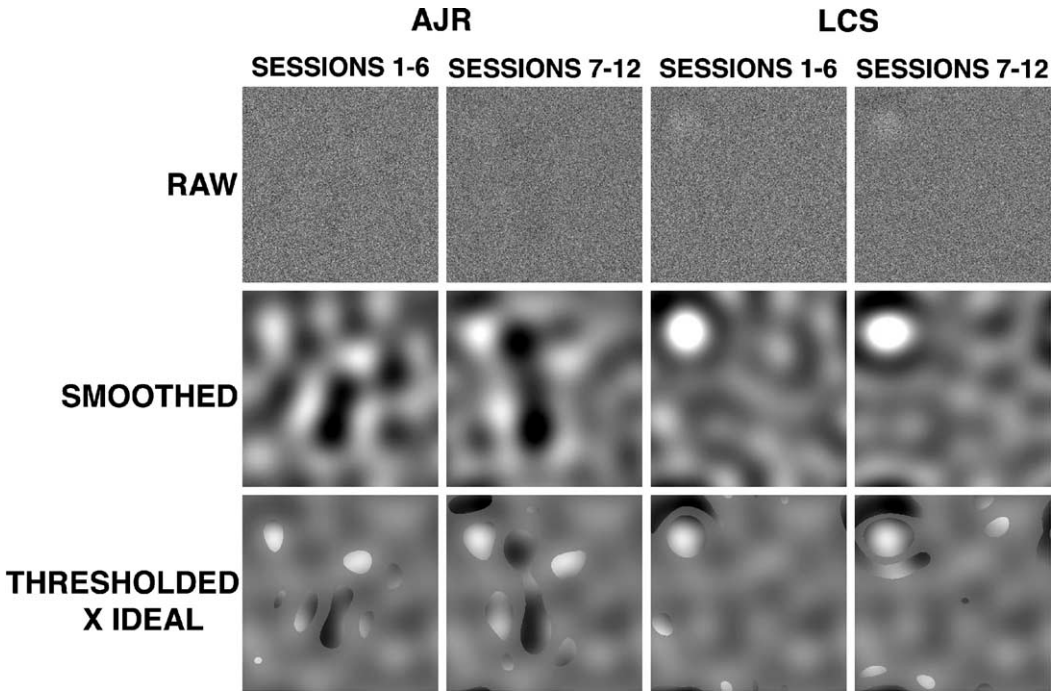

Fig. 12. The raw (top row), smoothed (middle row) and thresholded (bottom row) classification images for both observers in the texture identification task from Experiment 3, pooled across either the first or last half of the experiment.

condition. The results of this analysis are shown as *p*-values below each observer's data. These data show that the increase in correlation was statistically significant for all but one observer (LCS) in the texture identification task.

Despite the statistical significance of the analyses described above, the absolute levels of correlation were quite low. However, such low correlations are not surprising, given that the correlation analysis includes the entire stimulus; unlike the ideal observer, human observers undoubtedly use only a portion of the available stimulus information. As a result, the unused regions only contribute noise, which serves to reduce the correlation between the human and ideal templates. Therefore, it is of interest to examine the effects of removing some of the noise from the classification images that observers were unlikely to have used during the task. In the case of the face identification task, previous work has shown observers tend to make especially efficient use of spatial frequencies within a 2-octave wide band centered around 6 c/face, with efficiency gradually declining above and below the center frequency (Gold et al., 1999a; Nasanen, 1999). This finding suggests an appropriate filter to apply to the face classification images would be a 2-octave wide log-frequency filter centered at 6 c/face. In our analysis, we filtered the classification images (both human and ideal) by a log-Gaussian spatial frequency filter centered at 6 c/image with a bandwidth of 2 octaves (± 1 octave at half-height). In the case of the texture identification task, the signal was highly localized in the frequency domain (2–4 c/image). An optimal strategy in this task is to rely only upon frequencies that fall within

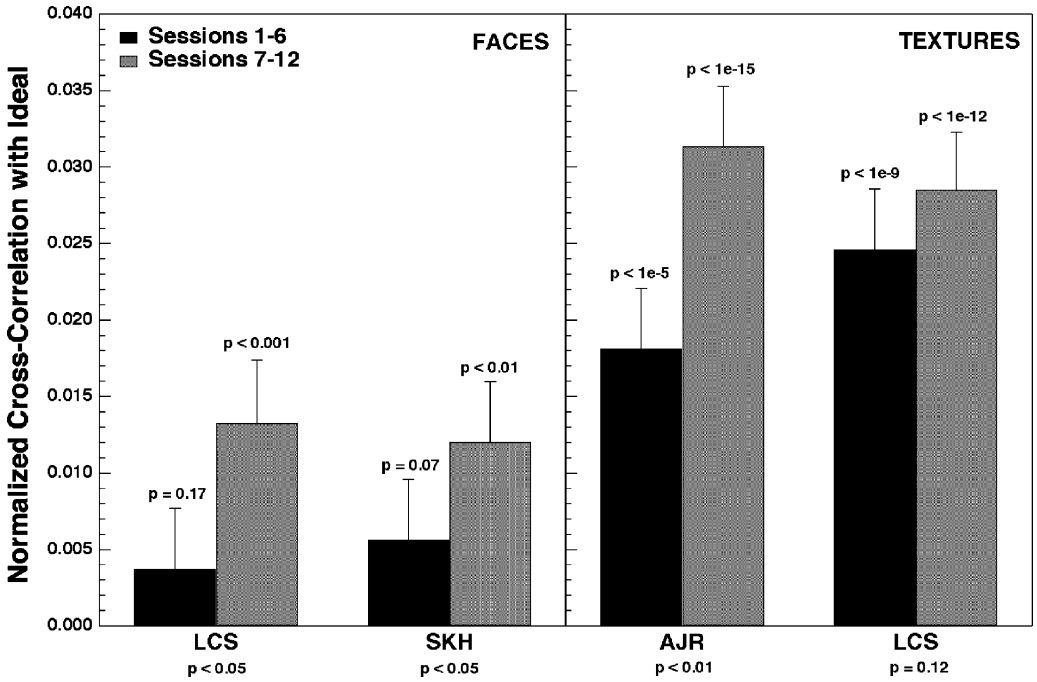

Fig. 13. Normalized cross-correlation between the human and ideal classification images in the first and last halves of the face (left panel) and texture (right panel) identification tasks from Experiment 3. The results of a statistical test comparing the correlation score to the score for a structureless classification image is shown above each bar. The results of a statistical test comparing the correlations for the first and last halves of the experiment are shown below each observer's data.

this limited range, and so an appropriate filter for the texture classification images would be one that is matched to the signal band. However, unlike with faces, there is no previous work suggesting observers might follow something similar to this optimal strategy. Because white noise is uncorrelated in both the spatial and the spatial frequency domains, we can compute a classification image in the spatial frequency domain in the same manner as a spatial classification image to see if amplitude in the signal band was correlated with observers' responses. This amounts to simply computing the Fourier transform of the spatial classification images.

The results of this spatial frequency analysis for one observer (LCS) in the texture identification task are shown in Fig. 14 (the results for the second observer were similar). This figure shows the power spectrum (i.e., the energy at each frequency) of the classification images from the first (left panels) and last (right panels) halves of the experiment, plotted in polar coordinates. In these plots, the origin corresponds to the DC component (i.e., average luminance), the distance from the origin spatial frequency, and the angle made by a vector drawn from the origin to a given point orientation. For clarity, only frequencies below 40 c/image are shown. The bottom panels are blurred versions of the top panels, which removed some of the masking produced by image pixelization.[5] These figures show a high contrast region within a 2–4 c/image ring around the center of each plot, indicating this observer relied almost exclusively upon these frequencies to perform the task. This result indicates that removing

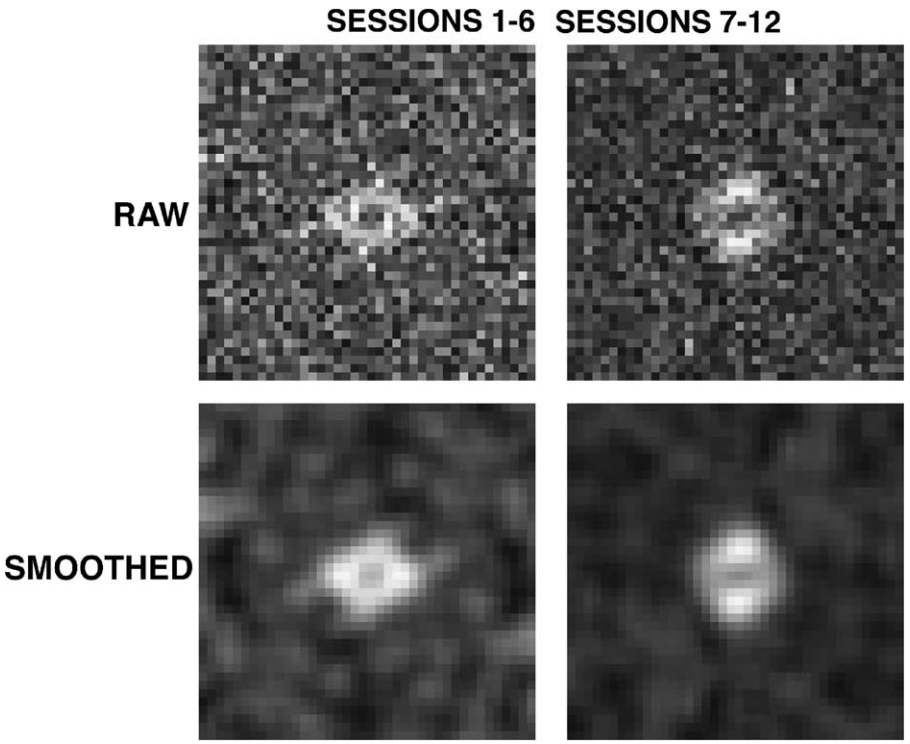

Fig. 14. Amplitude spectra of one observer's classification images (LCS) in the first (left column) and last (right column) halves of the texture identification task from Experiment 3, plotted in polar coordinates. In these plots, the origin corresponds to the DC component, the distance from the origin spatial frequency, and the angle made by a vector drawn from the origin to a given point orientation. Only frequencies below 40 c/image are plotted. The bottom two panels are blurred versions of the top two panels (convolved with a 5×5 pixel kernel).

frequencies outside of the signal band would greatly increase the signal-to-noise ratio in the texture classification images.

The middle panels of Figs. 11 and 12 show the classification images for the face and texture identification tasks, respectively, after being smoothed by the appropriate filters. The bottom row in each figure shows the smoothed images after being thresholded to remove all but the highest contrast pixels. Specifically, all of the pixels that fell within a restricted range[6] from mean gray were set to a single low value and the remaining pixels were set to unity. These thresholded images were then multiplied by the ideal classification image, resulting in an image where high contrast regions indicate the informative locations used by the observer to perform the task. In addition, the normalized cross-correlation between the human and ideal smoothed classification images was computed for each observer in each condition.

As before, the results of the cross-correlation analysis showed that the similarity between human and ideal classification images increased with learning (AJR textures: .39–.52; LCS textures: .36–.41; LCS faces: .04–.09; SKH faces: .04–.06). The smoothing process also served to increase the absolute levels of correlation between the human and ideal templates, especially in the texture discrimination task: one observer in the texture task (AJR) showed an increase of

a factor of 21.7 relative to the unfiltered correlation in the second half of the experiment. These effects can also be seen by visually inspecting the smoothed and thresholded classification images, where highly visible features now emerge from the background. In the face experiment, both observers concentrated upon the region around the left eye and eyebrow. In the texture experiment, AJR used an elongated region in the center, whereas LCS used a more localized blob in the top left corner. The thresholded classification images reveal that all of these locations were highly informative. Closer inspection of the images reveals that all of the observers appear to have used slightly larger regions in the second half of the experiment than in the first. This change in region size was verified by counting the percentage of pixels that exceed threshold in each half of the experiment. The percentage of pixels in the thresholded images increased by about a factor of 1.3 for both observers in the face identification task and by about a factor of 1.8 for both observers in the texture identification task. Although this increase in region size probably contributed to the increase in calculation efficiency, the use of more pixels does not necessarily insure that the calculation will be more efficient. For example, an observer could use twice as many pixels in the second half of the experiment, but if none of the pixels were informative, the calculation would be less efficient. It is therefore likely that a large portion of the improvement resulted from a re-weighting of pixels used in both halves of the experiment.

An aspect of the classification data worth emphasizing is the striking difference between the templates used by the two observers in the texture discrimination task. Inspection of the smoothed and thresholded classification images in Fig. 12 shows that the two observers adopted radically different strategies to discriminate between the textures: one observer (AJR) primarily relied upon a region in the center of each texture, whereas the other observer (LCS) primarily relied upon a region in the top left corner of each texture. Despite these differences, Fig. 10 shows that the performance of the two observers was actually quite similar throughout the last half of the experiment. These results demonstrate how the response classification analysis is able to reveal additional information not captured by the more gross measure of threshold or efficiency.

Do the observed changes in classification images account for the magnitude of learning? We examined this question by calculating face and texture discrimination thresholds for a simulated observer that used the filtered classification images as templates. If changes in calculation images were associated with perceptual learning, then the ratio of thresholds measured in sessions 1–6 and 7–12 ought to be similar in simulated and real observers. In fact, simulated and observed threshold ratios were quite similar: averaged across tasks and observers, thresholds in sessions 1–6 were 2.1 and 1.8 times higher than in sessions 7–12 in simulated and real observers, respectively. This result is consistent with the claim that the learning effect can be attributed largely to changes in the *linear* calculations that underlie response classification images.

Did changes in contrast-dependent internal noise affect our results? Although the previous experiments showed that internal noise does not decrease with learning in a 1-of-10 identification task, it still remains a possibility that the learning effects found in Experiments 1 and 2 do not extend to a one of two identification task. The importance of this possibility is made more apparent by noting that the changes in the classification images observed in the present experiment are also consistent with the effects of a decrease in contrast-dependent internal noise: such a reduction would increase the signal-to-noise ratio in the classification images, causing them to converge more quickly and thus increase the correlation with the ideal template and the number of significant pixels.

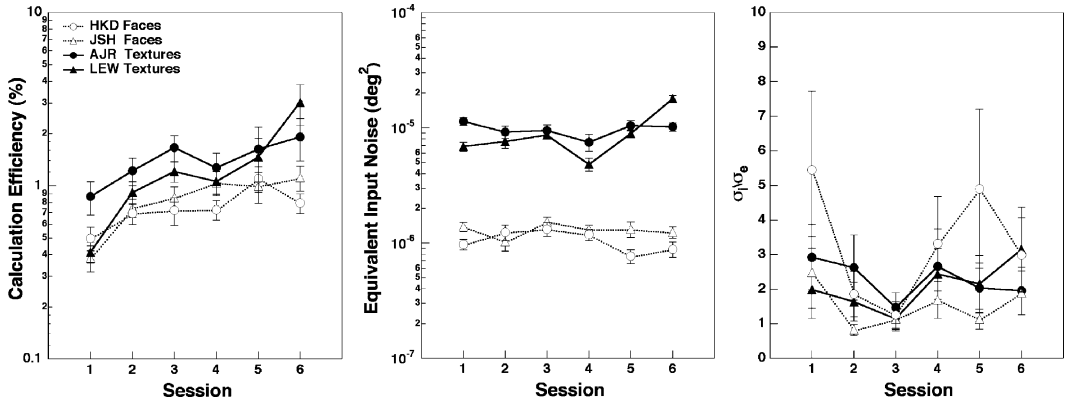

Fig. 15. Calculation efficiency (left panel), equivalent input noise (middle panel), and σ_i/σ_e estimates (right panel) as a function of experimental session for all observers in the one of two face (open symbols) and texture (closed symbols) identification tasks from Experiment 3. In all figures, error bars on each symbol correspond to ± 1 standard deviation.

This possibility was tested by measuring calculation efficiency and internal noise (both contrast–invariant and contrast-dependent) for the one of two identification tasks. Nearly all previous studies (including Experiment 1) have found a linear relationship between signal energy threshold and external noise power spectral density (Pelli & Farell, 1999). Based on the uniformity of previous results, we assumed that the noise masking functions in our one of two identification tasks would be linear, allowing us to measure thresholds in only two levels of external noise. This simplification increased the number of trials at each level of external noise, allowing us to concurrently measure equivalent input noise, calculation efficiency and response consistency in the same experiment. We used the highest and lowest external noise levels from Experiment 1, and the first and last halves of each session were identical to allow for the measurement of response consistency.[7] Each session (total of 6) consisted of 200 repeated trials per external noise level, for a total of 800 trials. Two new observers participated in both the face and texture discrimination tasks. The results of this experiment are shown in Fig. 15, which plots calculation efficiency (left panel), equivalent input noise (middle panel) and high noise σ_i/σ_e for each observer, as a function of practice. These data are very similar to the 1-of-10 identification task data of Experiments 1 and 2, with all observers showing increased calculation efficiency with little or no change in equivalent input noise or σ_i/σ_e. The one notable exception to this trend was observer LEW in the texture identification condition whose equivalent noise remained approximately constant during sessions 1–5 but *increased* during session 6. However, the overall results were consistent with the conclusion that an increase in the efficiency of the deterministic aspects of observers' calculations mediated the changes we found in the classification images with learning.

5. General discussion

Many experiments have demonstrated that practice in perceptual tasks produces substantial improvements in performance. Within the context of signal detection theory, these improve-

ments could be due to either increased signal strength or decreased internal noise (or some combination of the two). To discriminate between these two possibilities, we used a combination of psychophysical techniques involving the addition of external stimulus noise in conjunction with a black-box model of the human visual system. Experiment 1 was designed to discriminate between the effects of contrast–invariant internal noise and the efficiency of internal calculations as observers learned to identify two unfamiliar sets of patterns, human faces and band-pass filtered noise textures. Performance in both tasks was consistent with the model's prediction that signal energy threshold should be linearly related to the power of an externally added noise. This relationship also allowed us to infer the amount of contrast–invariant internal noise and the efficiency of the calculation as learning took place. The results of Experiment 1 indicated that contrast–invariant internal noise remained fixed while calculation efficiency increased systematically with practice.

Experiment 2 allowed us to disambiguate purely deterministic changes in observer calculations from changes in contrast-dependent internal noise, both of which could contribute to changes in calculation efficiency. Double-pass response consistency was used to measure independently any contributions made by contrast-dependent internal noise to the increases in calculation efficiency found in Experiment 1. Even though performance improved with practice, the consistency of responses made between identical passes through the experiment did not increase significantly with learning, indicating contrast-dependent internal noise did not play a large role in the increases in calculation efficiency found in Experiment 1.

The third and final experiment allowed us to specify directly some of the changes that take place in an observer's calculations as perceptual learning occurs. Response classification was used to measure the linear computations employed by observers as they learned to recognize unfamiliar faces and textures. The resulting classification images revealed some of the strategies used by observers in the different tasks, as well as some of the gross changes that took place between the first and last halves of the experiment. The classification images also were used to verify the prediction that observers' classification images should be more similar to the ideal classification image after learning has taken place.

The results of these experiments have important implications for current models of perceptual learning. In the next two sections, we discuss the constraints our results place on the neural mechanisms underlying perceptual learning and on theories of perceptual learning in general.

5.1. Relation to neural mechanisms

How is learning represented in the brain? Perceptual learning affects virtually every level of processing in the brain, from synaptic connections to global patterns of blood flow (for recent reviews, see Fahle & Poggio, 2002; Gilbert, Sigman, & Crist, 2001). At the level of single cell responses, several researchers have found changes in relative firing rates for visual stimuli as a function of experience (e.g., Logothetis, Pauls, & Poggio, 1995; Rainer & Miller, 2000; Rolls, Baylis, Hasselmo, & Nalwa, 1989; Schoups, Vogels, Qian, & Orban, 2001). For example, Rainer and Miller (2000) trained monkeys to recognize complex objects degraded by noise. After several days of practice, the monkeys performed significantly better with high levels of stimulus degradation. Individual neurons in the prefrontal cortex showed a related effect of experience. After practice, fewer neurons responded to the target stimuli, but the neurons

that did respond were more selectively tuned for those stimuli, and responded well over larger ranges of degradation. Although the details of their study differ from the experiments reported here, the sorts of changes in neural coding identified by Rainer and Miller could be responsible for the improvements we found in human observers with learning—the modification of response selectivity of neurons is one way the visual system could "fine tune" a perceptual template.

Of course, it is likely that our results are driven by response changes at the level of cell populations rather than individual neurons. Schoups et al. (2001) recently described how such a population level analysis might account for perceptual learning. Monkeys were trained to discriminate the orientation of a small grating from other similar orientations. Over the course of training, the monkeys' orientation discrimination thresholds decreased significantly, but only for gratings with similar orientation as the trained stimulus. After learning, Schoups et al. found that the slope of the tuning function measured at the trained orientation increased, but only for the sub-population of V1 neurons most likely to code the orientation difference detected by the monkey (i.e., those neurons with peak activity 12–20° from the trained orientation). This change in neuronal coding was apparent only when neurons were tested with the trained orientation. No such effect was found when tuning functions were measured at an untrained, orthogonal orientation, even when that untrained orientation had been passively viewed for an equal number of times as the trained orientation. A population model of the obtained response changes accounts well for the behavioral learning effects. Schoups et al.'s results, like those of Rainer and Miller (2000) and the present set of experiments, are consistent with the idea that perceptual learning is mediated by an increase in the efficiency of perceptual encoding. In the case of neuronal responses, increased efficiency is obtained through changes in the tuning functions of visual neurons. However, our results also specifically lead to the prediction that no change in noise accompanies the change in neuronal tuning functions. This prediction was tested directly by Schoups et al. They found that noise (as indexed by the normalized response variance of neurons) did not change as a function of learning, supporting the idea that perceptual learning changes the discriminative signal but not internal noise.

5.2. Mechanisms of learning

For our tasks, the ideal calculation is to correlate the input with copies of the stimulus alternatives (i.e., templates), and to select the item yielding the highest correlation. For example, in the two-item task in Experiment 3 the ideal rule is to transform the input, s, into a decision variable, v, using the rule

$$v = \sum_i w_{1i}s_i - \sum_i w_{2i}s_i = \sum_i (w_{1i} - w_{2i})s_i = \sum_i w_{\mathrm{d}i}s_i \tag{17}$$

where w_1 and w_2 are sets of weights comprising the templates for the two stimulus alternatives, w_d is the difference between the weights, and the summation is taken over all i pixels in the stimulus. For ideal performance, the w_1 and w_2 are the stimuli themselves—so w_d is simply the difference between the stimuli—and the ideal rule is to select alternative 1 if v is greater than 0 and alternative 2, otherwise. The task of learning is to select a set of weights that is as similar as possible to w_d to maximize response accuracy.

The problem of how to search a weight-space for an optimal solution to a categorization problem has been studied extensively in the neural network literature, resulting in a wide range of network architectures and weight-setting procedures that are useful in different situations (Arbib, 1995). Also, several neural network models have been proposed specifically to account for perceptual learning (Estes, 1994; Herzog & Fahle, 1998; Mato & Sompolinsky, 1996; McLaren, Kaye, & Mackintosh, 1989; Poggio et al., 1992; Weiss, Edelman, & Fahle, 1993). Our experiments used simple learning tasks, so it is likely that most, if not all, of these models could account for the decrease in threshold that occurred as a function of practice. Indeed, the stimuli in Experiment 3 are linearly separable, so even a simple perceptron (Rosenblatt, 1958) could learn to perform optimally in that task.

However, another aspect of our results—namely, that response consistency does not change with learning—does pose a challenge to some neural network models. Consider, the model described by McLaren et al. (1989). In this model, stimuli are represented as a large number of elements that are sampled randomly on each trial. During repeated presentation, associations are formed between elements that co-occur, and inhibitory connections are formed between elements that do *not* co-occur. The (positive) associations ensure that the complete set of elements representing a stimulus—rather than a small, variable subset—will be activated each time that item is presented, and the inhibitory connections ensure that the sets of elements activated by different stimuli will become more distinct. Thus, the model accounts for the basic finding that practice improves stimulus discrimination, as well as a variety of more subtle effects (Jones, Wills, & McLaren, 1998; McLaren, 1997; Wills & McLaren, 1998). Notice, however, that the model also predicts that response consistency should increase with learning. In the framework illustrated in Fig. 1, the random stimulus sampling that occurs at the start of learning would manifest itself as an internal noise. If, as posited, random stimulus sampling is a significant source of internal variability, then eliminating it by forming associations among co-occurring elements should reduce internal variability, and therefore increase response consistency. We found no evidence that response consistency changed with learning, suggesting that random stimulus sampling did not diminish with practice, or that it does not contribute significantly to internal variability.

More generally, the consistency data force us to consider how sources of noise fit into a model's architecture. The consistency data imply that the internal/external noise ratio at the level at which the decision variable is computed remains constant. For the model illustrated in Fig. 1, internal noise is added at the first processing stage, and the subsequent calculation cannot distinguish between the external and internal components of the total noise. Therefore, *any* adjustment of the template weights in Eq. (17) will leave the internal/external noise ratio (and response consistency) unchanged. This statement is true both for contrast-dependent and contrast-independent internal noise. Now consider a model in which internal noise is added to the decision variable after the calculation. The templates that yield optimal performance in the early-noise model will also yield optimal performance in the late-noise model. However, in the late-noise model, external noise is filtered through the template before it is combined with internal noise, and so altering the template might alter the internal/external noise ratio and response consistency. For the noise used in our tasks, it is easy to show that the proportion of noise variance passed by the linear template equals the sum of the squared weights (Bracewell, 2000). To account for the consistency data, the late-noise model therefore must assume that

either (a) the internal noise is entirely contrast-dependent, or (b) that the weights of the template are adjusted so that the sum of squared weights is constant with learning. The important point is not whether such constraints on the noise or weight-adjustment algorithms are easy to implement, but rather that they become necessary only when one considers the effects of learning on response consistency. Most network models have multiple layers consisting of noisy (i.e., probabilistic) non-linear elements, and it is difficult to predict how the consistency of such networks changes with learning. Measures of response consistency may be useful in distinguishing among competing neural network models.

5.3. Perceptual Template Model

Dosher and Lu (1998, 1999) conducted a set of experiments very similar to our Experiment 1. Their task involved training observers to judge the relative orientation of a peripherally presented sinusoidal grating in varying amounts of external noise. Their results were nearly identical to those we found with our face and texture identification tasks—practice produced a uniform shift down in log threshold as a function of log external noise. However, Dosher and Lu interpreted their results within the context of a slightly different black-box model, which they term the *Perceptual Template Model* or *PTM* (Lu & Dosher, 1998). The PTM is similar to the black-box model described in the introduction. However, one key difference is that it places all of the observer's internal noise (both contrast–invariant and contrast-dependent) after the calculation, rather than before it.[8] In the PTM framework, three processes can potentially contribute to perceptual learning. One process, termed *signal enhancement* or, equivalently, *additive noise reduction*, corresponds to either a reduction in contrast–invariant (i.e., additive) internal noise *or* an increase in the *gain* of the perceptual template (i.e., the sum of the squared weights of the template, as described in Eq. (17)). Lu and Dosher (1998) have shown that these two processes are mathematically equivalent in the PTM framework, and are empirically indistinguishable. Signal enhancement and additive noise reduction will lower thresholds primarily in conditions where external noise is low. The second process is termed *external noise exclusion*. Dosher and Lu illustrate this process by considering the effects of learning on a spatial-frequency tuned filter, or template, that is maximally sensitive to the target spatial frequency and has a bandwidth that narrows with learning. For such a calculation, learning would not alter the filter's response to the target frequency, but would reduce the filter's response to the external noise. Therefore, external noise exclusion reduces thresholds primarily in conditions when external noise is high. For the narrow-band targets used in their experiments, narrowing the filter's spatial frequency bandwidth also makes the template's shape more similar to the shape of the ideal filter. Therefore, one way of characterizing this process is to say that external noise exclusion corresponds to a combination of (i) an increase in the efficiency of the template (i.e., its similarity to the ideal, as described in the correlation analysis from Experiment 3); and (ii) a reduction in the total power passed by the template (i.e., a reduction in the template's gain, hence excluding external noise power that previously passed through the filter). Note that external noise exclusion incorporates two mechanisms, and that one of them (reduced template gain) has the opposite effect of signal enhancement. The third process is *internal noise reduction*. This process corresponds to a reduction in contrast-dependent internal noise. As in the early-noise model, the effect of internal

noise reduction is to reduce thresholds uniformly at all levels of external noise in log–log coordinates.

If we now consider the effects of perceptual learning from Experiments 1 and 2 in terms of these processes, the PTM appears to offer a different account of the data. One way the PTM could account for our results is in terms a reduction in contrast-dependent internal noise. However, this possibility is ruled out by the response consistency analysis in Experiment 2. Dosher and Lu (1999) also rule out this possibility for their learning data by comparing noise masking functions measured at two different d' criterion levels. A second way the PTM could account for our results is in terms of a combination of signal enhancement (reducing thresholds at low levels of external noise) and external noise exclusion (reducing thresholds at high levels of external noise). If these two mechanisms changed by exactly the same amount, thresholds would be reduced uniformly across all levels of external noise. This explanation is equivalent to the one adopted by Dosher and Lu to account for their results, and it appears to differ significantly from our claim that only the calculation changes with learning. However, recall that external noise exclusion is a combination of increased template efficiency and reduced template gain. This second component of external noise exclusion (reduced template gain) is, in mathematical terms, the inverse of signal enhancement. Thus, the claim that learning increases both external noise exclusion and signal enhancement is formally equivalent to the claim that learning increases only template efficiency but does not change template gain. In other words, if template efficiency can be changed without altering template gain, then the PTM could account for the effects of learning in terms of a single mechanism—increased template efficiency. Although it is possible that all three mechanisms (contrast–invariant internal noise, template gain, and template efficiency) changed in such a way as to mimic the effects of an increase in template efficiency, it is more parsimonious to conclude that practice produced changes in just one mechanism: template efficiency.

However, Lu and Dosher (2001b) have recently obtained other results that cannot be accounted for solely in terms of increased template efficiency. Specifically, they found that performance in a foveal grating orientation discrimination task improved with practice, but only in high levels of external noise. In terms of the curves shown in Fig. 2, this would produce a downward shift in thresholds only in high levels of external noise, producing a rightward shift in the knee of the noise masking function. In both models, this result can be accounted for in terms of a combination of increased template efficiency and increased contrast–invariant internal noise. However, a late noise model (like the PTM) can also account for the results in terms of a combination of increased template efficiency and decreased template gain (i.e., external noise exclusion). Thus, for a late noise model, a natural account of Lu and Dosher's foveal grating discrimination data is that learning served to narrow the bandwidth of an initially broadband template to better match the narrow-band grating pattern. In contrast, an early noise model is forced to account for this result in terms of increased template efficiency coupled with increased contrast–invariant internal noise. Although it is possible that internal noise might increase with perceptual learning, it is more parsimonious in this case to assume that the dominant internal noise occurred after the calculation and that internal noise remained fixed. One way of reconciling the different findings is to assume that the grating discrimination tasks used by Lu and Dosher (2001b) tap mechanisms at relatively early stages in visual processing, whereas our face and texture identification tasks tap later stages, and that the dominant internal

noise exists between the two stages. In the end, until it is possible to determine the location of internal noise relative to the template in a given task, choosing between an early and late noise model to describe one's data ultimately amounts to a question of parsimony.

6. Conclusions and future directions

In conclusion, our findings suggest that perceptual learning is associated with changes in the linear calculations that observers perform on a stimulus. Using the framework depicted in Fig. 1, learning is associated with changes in calculation efficiency, rather than reductions in internal noise. One obvious question that remains is whether this result generalizes to other stimuli and tasks. We obtained similar results using faces and random textures, both of which are spatially complex, but it is possible that different results might be obtained if observers were asked to identify simpler stimuli. Also, we used identification tasks, and it is an open question whether learning in detection tasks or various kinds of discrimination tasks (e.g., same–different or AB-X tasks) exhibits similar characteristics.

The classification image data demonstrated that observers base their responses on the information conveyed by only a small subset of pixels. Similar findings have been reported in other studies using spatially complex patterns (e.g., Gold et al., 2000; Sekuler et al., 2001). This result raises the question of what determines the parts of a stimulus that are used by an observer. Our results are consistent with the idea that observers learn to use pixels that are informative, but it is interesting to note that the informative parts of the faces and textures in our experiments also were relatively high in contrast. It is possible that parts of the stimulus were selected on the basis of detectability or salience, rather than the amount of discriminative information they contained. Additional experiments are needed to determine the relative contributions of these factors to feature selection.

Notes

1. It is worth noting that cross-correlation is the ideal rule here because all of the alternatives in each task have the same contrast energy. See Green and Swets (1966) for further discussion of ideal decision rules for other tasks and stimuli.
2. The simulated observers used the ideal decision rule (Eq. (8)). The use of this rule insured the performance of the simulated observer would eventually reach 100% correct at high levels of signal energy.
3. The values of the parameters fit to Eq. (12) in Experiment 2 were $\alpha = .114$, $\gamma_1 = 4.07$, $\beta_1 = .013$, $\gamma_2 = 75.907$ and $\beta_2 = .036$.
4. The failure to reach statistical significance for LCK was most likely due to the particularly large errors associated with the empirical estimates of m for this observer (see Fig. 8). The greater variability stems from this observer's high levels of contrast-dependent internal noise in combination with the nonlinearity of Eq. (12): lower values of m (which correspond to greater inconsistency) are located upon progressively steeper portions of the function, where σ_i/σ_e estimates are less reliable.

5. The plots were convolved with a 5×5 convolution kernel (the matrix product of [1 2 3 2 1] with itself transposed).
6. Specifically, the expected variance of a structureless classification image blurred by the appropriate convolution kernel was computed, and all pixels that fell within ± 2.57 standard deviations of the expected mean (zero correlation) were set to a single, low number.
7. New simulations were conducted to measure the slopes of the noise masking functions for the ideal observer in the one of two face and texture identification tasks. Additional simulations were conducted to obtain the values of the parameters for Eq. (12) in a one of two identification task ($\alpha = .0727$, $\gamma_1 = 1.076$, $\beta_1 = .003$, $\gamma_2 = 141.523$, $\beta_2 = .031$).
8. The PTM also includes an optional non-linearity after the template but before the internal noise. The inclusion of the non-linearity is often important for modeling various aspects of the shapes of psychometric functions (Eckstein et al., 1997; Legge et al., 1987; Lu & Dosher, 1999; Watson & Solomon, 1997). However, the qualitative differences between the two models in terms of the changes that take place in the shape of the noise masking function with changes in the different mechanisms are the same with or without the kind of non-linearity described in the PTM. For example, in the context of the PTM, including the non-linearity does not alter the conclusion that learning reduces internal noise and increases external noise exclusion (and similarly, including a non-linearity before the internal noise in the model illustrated in Fig. 1 would not alter the conclusion that learning increases calculation efficiency but has no effect on contrast–invariant internal noise). What the non-linearity *does* do is enable the PTM to account for the fact that the amount of internal noise reduction and external noise exclusion that is observed during learning depends upon the criterion used to define threshold (Lu & Dosher, 1999).

Acknowledgments

We would like to thank Richard Murray and three anonymous reviewers for helpful comments during earlier stages of this work. Parts of this research were published previously in *Nature* (Gold, Bennett, & Sekuler, 1999b), submitted to the University of Toronto in partial fulfillment of a Ph.D. in psychology (Gold, 2001), and presented at the Association for Research in Vision and Ophthalmology (Gold, Bennett, & Sekuler, 1999c) and Vision Sciences Society (Gold, Bennett, & Sekuler, 2002). This research was supported by NSERC grants OGP105494 and OGP0042133 to P.J.B. and A.B.S., and a University of Toronto Health award.

References

Abbey, C. K., & Eckstein, M. P. (2002). Optimal shifted estimates of human-observer templates in two alternative forced-choice experiments. *IEEE Transactions on Medical Imaging*, *21*(5), 429–440.

Abbey, C. K., Eckstein, M. P., & Bochud, F. O. (1999). *Estimation of human-observer templates in two-alternative forced-choice experiments*. Paper presented at the Proceedings of SPIE, San Diego, CA.

Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, *387*(6631), 401–406.

Ahumada, A. J. (1996). Perceptual classification images from vernier acuity masked by noise. *Perception*, *25*, 18.
Ahumada, A. J., Jr. (2002). Classification image weights and internal noise level estimation. *Journal of Vision*, *2*, 121–131.
Ahumada, A. J., & Lovell, J. (1971). Stimulus features in signal detection. *JASA*, *49*(6/2), 1751–1756.
Arbib, M. A. (1995). *The handbook of brain theory and neural networks*. Cambridge, MA: MIT Press.
Asaad, W. F., Rainer, G., & Miller, E. K. (2000). Task-specific neural activity in the primate prefrontal cortex. *Journal of Neurophysiology*, *84*(1), 451–459.
Ball, K., & Sekuler, R. (1987). Direction-specific improvement in motion discrimination. *Vision Research*, *27*(6), 953–965.
Banks, M. S., Geisler, W. S., & Bennett, P. J. (1987). The physical lmits of grating visibility. *Vision Research*, *27*, 1915–1924.
Banks, M. S., Sekuler, A. B., & Anderson, S. J. (1991). Peripheral spatial vision: Limits imposed by optics, photoreceptors, and receptor pooling. *Journal of the Optical Society of America A*, *8*(11), 1775–1787.
Barlow, H. B. (1956). Retinal noise and absolute threshold. *JOSA*, *46*, 634–639.
Barlow, H. B. (1957). Increment thresholds at low intensities considered as signal/noise discrimination. *Journal of Physiology*, *136*, 469–488.
Beard, B. L., & Ahumada, A. J. (1998). *Technique to extract relevant image features for visual tasks*. Paper presented at the SPIE, San Jose, CA.
Bennett, P. J., Sekuler, A. B., & Ozin, L. (1999). Effects of aging on calculation efficiency and equivalent noise. *Journal of the Optical Society of America A, Optics and Image Science Vision*, *16*(3), 654–668.
Bochud, F. O., Abbey, C. K., & Eckstein, M. P. (2000). Visual signal detection in structured backgrounds. III. Calculation of figures of merit for model observers in statistically nonstationary backgrounds. *Journal of the Optical Society of America A: Optics Image Science Vision*, *17*(2), 193–205.
Bracewell, R. N. (2000). *The Fourier transform and its applications* (3rd ed.). Boston: McGraw-Hill.
Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 443–446.
Buonomano, D. V., & Merzenich, M. M. (1998). Cortical plasticity: From synapses to maps. *Annual Review of Neuroscience*, *21*, 149–186.
Burgess, A. E., & Colborne, B. (1988). Visual signal detection. IV. Observer inconsistency. *Journal of the Optical Society of America A*, *5*(4), 617–627.
Burgess, D. G. (1990). High level decision efficiencies. In C. Blakemore (Ed.), *Vision: Coding and efficiency* (pp. 431–440). Cambridge, MA: Cambridge University Press.
Crist, R. E., Kapadia, M. K., Westheimer, G., & Gilbert, C. D. (1997). Perceptual learning of spatial localization: Specificity for orientation, position, and context. *Journal of Neurophysiology*, *78*(6), 2889–2894.
Croner, L. J., Purpura, K., & Kaplan, E. (1993). Response variability in retinal ganglion cells of primates. *Proceedings of the National Academy of Sciences United States of America*, *90*(17), 8128–8130.
Demany, L. (1985). Perceptual learning in frequency discrimination. *Journal of the Acoustical Society of America*, *78*(3), 1118–1120.
Dosher, B. A., & Lu, Z. (1998). Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. *Proceedings of National Academy of Science*, *95*, 13988–13993.
Dosher, B. A., & Lu, Z. L. (1999). Mechanisms of perceptual learning. *Vision Research*, *39*(19), 3197–3221.
Dosher, B. A., & Lu, Z. L. (2000). Mechanisms of perceptual attention in precuing of location. *Vision Research*, *40*(10–12), 1269–1292.
Dresslar, F. B. (1894). Studies in the psychology of touch. *American Journal of Psychology*, *6*(3), 313–368.
Eckstein, M. P., Ahumada, A. J., & Watson, A. B. (1997). Visual signal detection in structured backgrounds. II. Effects of contrast gain control, background variations, and white noise. *Journal of the Optical Society of America A*, *14*(9), 2406–2419.
Eckstein, M. P., Shimozaki, S. S., & Abbey, C. K. (2002). The footprints of visual attention in the Posner cueing paradigm revealed by classification images. *Journal of Vision*, *2*(1), 25–45.
Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.
Estes, W. K. (1994). *Classification and cognition*. Oxford: Oxford University Press.
Fahle, M., Edelman, S., & Poggio, T. (1995). Fast perceptual learning in hyperacuity. *Vision Research*, *35*(21), 3003–3013.

Fahle, M., & Morgan, M. (1996). No transfer of perceptual learning between similar stimuli in the same retinal position. *Current Biology*, *6*(3), 292–297.

Fahle, M., & Poggio, T. (2002). *Perceptual learning*. Cambridge, MA: MIT Press.

Fine, I., & Jacobs, R. A. (2000). Perceptual learning for a pattern discrimination task. *Vision Research*, *40*(23), 3209–3230.

Fiorentini, A., & Berardi, N. (1980). Perceptual learning specific for orientation and spatial frequency. *Nature*, *287*(5777), 43–44.

Fiorentini, A., & Berardi, N. (1997). Visual perceptual learning: A sign of neural plasticity at early stages of visual processing. *Archives of Italian Biology*, *135*(2), 157–167.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, *3*(2), 191–197.

Gauthier, I., & Tarr, M. J. (1997). Becoming a 'Greeble' expert: Exploring mechanisms for face recognition. *Vision Research*, *37*(12), 1673–1682.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nature Neuroscience*, *2*(6), 568–573.

Geisler, W. S. (1989). Sequential ideal observer analysis of visual discriminations. *Psychology Review*, *96*(2), 267–314.

Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York: Appleton.

Gilbert, C. D. (1994). Early perceptual learning [comment]. *Proceedings of the National Academy of Sciences United States of America*, *91*(4), 1195–1197.

Gilbert, C. D., Sigman, M., & Crist, R. E. (2001). The neural basis of perceptual learning. *Neuron*, *31*(5), 681–697.

Gold, J. M. (2001). *Signal and noise in perceptual learning*. Unpublished Ph.D. thesis, University of Toronto, Toronto.

Gold, J., Bennett, P. J., & Sekuler, A. B. (1999a). Identification of band-pass filtered letters and faces by human and ideal observers. *Vision Research*, *39*, 3537–3560.

Gold, J., Bennett, P. J., & Sekuler, A. B. (1999b). Signal but not noise changes with perceptual learning. *Nature*, *402*(6758), 176–178.

Gold, J. M., Bennett, P. J., & Sekuler, A. B. (1999c). Learning improves calculation efficiency for complex pattern identification. *IOVS*, *40*(4), S586.

Gold, J. M., Bennett, P. J., & Sekuler, A. B. (2002). *Visualizing perceptual learning*. Paper presented at the Vision Sciences Society, Sarasota, FL.

Gold, J. M., Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2000). Deriving behavioral receptive fields for visually completed contours. *Current Biology*, *10*, 663–666.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, *49*, 585–612.

Green, D. M. (1964). Consistency of auditory detection judgments. *Psychology Review*, *71*(5), 392–407.

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

Herzog, M. H., & Fahle, M. (1997). The role of feedback in learning a vernier discrimination task. *Vision Research*, *37*(15), 2133–2141.

Herzog, M. H., & Fahle, M. (1998). Modeling perceptual learning: Difficulties and how they can be overcome. *Biology Cybernetics*, *78*(2), 107–117.

Herzog, M. H., & Fahle, M. (1999). Effects of biased feedback on learning and deciding in a vernier discrimination task. *Vision Research*, *39*(25), 4232–4243.

Jones, F. W., Wills, A. J., & McLaren, I. P. (1998). Perceptual categorization: Connectionist modelling and decision rules. *Quartenary Journal of Experimental Psychology B*, *51*(1), 33–58.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*(11), 4302–4311.

Karni, A., Meyer, G., Rey-Hipolito, C., Jezzard, P., Adams, M. M., Turner, R., et al. (1998). The acquisition of skilled motor performance: Fast and slow experience-driven changes in primary motor cortex. *Proceedings of the National Academy of Sciences United States of America*, *95*(3), 861–868.

Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: Evidence for primary visual cortext plasticity. *Proceedings of the National Academy of Sciences United States of America*, *88*, 4966–4970.

Karni, A., & Sagi, D. (1993). The time course of learning a visual skill. *Nature*, *350*, 250–252.

Legge, G., Kersten, D., & Burgess, A. E. (1987). Contrast discrimination in noise. *Journal of the Optical Society of America A*, *4*(2), 391–406.

Levi, D. M., & Klein, S. A. (2002). Classification images for detection and position discrimination in the fovea and parafovea. *Journal of Vision*, *2*(1), 46–65.

Levi, D. M., Klein, S. A., & Carney, T. (2000). Unmasking the mechanisms for Vernier acuity: Evidence for a template model for Vernier acuity. *Vision Research*, *40*(8), 951–972.

Lillywhite, P. G. (1981). Multiplicative intrinsic noise and the limits to visual performance. *Vision Research*, *21*, 291–296.

Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology*, *5*(5), 552–563.

Lu, Z. L., & Dosher, B. A. (1998). External noise distinguishes attention mechanisms. *Vision Research*, *38*(9), 1183–1198.

Lu, Z. L., & Dosher, B. A. (1999). Characterizing human perceptual inefficiencies with equivalent internal noise. *Journal of the Optical Society of America A: Optics Image Science Vision*, *16*(3), 764–778.

Lu, Z. L., & Dosher, B. A. (2001a). Characterizing the spatial-frequency sensitivity of perceptual templates. *Journal of the Optical Society of America A: Optics Image Science Vision*, *18*(9), 2041–2053.

Lu, Z. L., & Dosher, B. A. (2001b). External noise exclusion as the mechanism of perceptual learning for orientation discrimination in fovea. *IOVS*, *42*(4), S316.

Lu, Z. L., Liu, C. Q., & Dosher, B. A. (2000). Attention mechanisms for multi-location first- and second-order motion perception. *Vision Research*, *40*(2), 173–186.

Mato, G., & Sompolinsky, H. (1996). Neural network models of perceptual learning of angle discrimination. *Neural Computation*, *8*(2), 270–299.

Matthews, N., Liu, Z., Geesaman, B. J., & Qian, N. (1999). Perceptual learning on orientation and direction discrimination. *Vision Research*, *39*(22), 3692–3701.

McLaren, I. P. (1997). Categorization and perceptual learning: An analogue of the face inversion effect. *Quartenary Journal of Experimental Psychology A*, *50*(2), 257–273.

McLaren, I. P., Kaye, H., & Mackintosh, N. J. (1989). An associative theory of the representation of stimuli: Applications to perceptual learning and latent inhibition. In R. G. M. Morris (Ed.), *Parallel distributed processing: Implications for psychology and neurobiology*. Oxford: Oxford University Press.

Mumford, W. W., & Schelbe, E. H. (1968). *Noise performance factors in communication systems*. Dedham, MA: Horizon House-Microwave, Inc.

Murray, R.F. (2002).Perceptual organization and the efficiency of shape discrimination. Unpublished Ph.D., University of Toronto, Canada.

Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2002). Optimal methods for calculating classification images: Weighted sums. *Journal of Vision*, *2*, 79–104.

Nasanen, R. (1999). Spatial frequency bandwidth used in the recognition of facial images. *Vision Research*, *39*(23), 3824–3833.

Neri, P., & Heeger, D. J. (2002). Spatiotemporal mechanisms for detecting and identifying image features in human vision. *Nature Neuroscience*, *5*(8), 812–816.

Neri, P., Parker, A. J., & Blakemore, C. (1999). Probing the human stereoscopic system with reverse correlation. *Nature*, *401*(6754), 695–698.

Noreen, D. L. (1981). Optimal decision rules for some common psychophysical paradigms. In S. Grossberg (Ed.), *Mathematical psychology and psychophysiology (SIAM-AMS proceedings; v. 13)* (p. x, 318). Providence, RI: American Mathematical Society.

Pelli, D. G. (1981). *Effects of visual noise*. Unpublished Ph.D. thesis, University of Cambridge, Cambridge.

Pelli, D. G. (1990). The quantum efficiency of vision. In C. Blakemore (Ed.), *Vision: Coding and efficiency* (pp. 3–24). Cambridge, MA: Cambridge University Press.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

Pelli, D. G., & Farell, B. (1999). Why use noise? *Journal of the Optical Society of America A: Optics Image Science Vision*, *16*(3), 647–653.

Perrett, D. I., Hietanen, J. K., Oram, M. W., & Benson, P. J. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philosophical Transaction of Royal Society of London B: Biological Science*, *335*(1273), 23–30.

Poggio, T., Fahle, M., & Edelman, S. (1992). Fast perceptual learning in visual hyperacuity. *Science*, *256*(5059), 1018–1021.

Raghavan, M. (1989). *Sources of visual noise*. Unpublished Ph.D. thesis, Syracuse University, Syracuse.

Rainer, G., & Miller, E. K. (2000). Effects of visual experience on the representation of objects in the prefrontal cortex. *Neuron*, *27*(1), 179–189.

Recanzone, G. H., Schreiner, C. E., & Merzenich, M. M. (1993). Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *Journal of Neuroscience*, *13*(1), 87–103.

Rolls, E. T., Baylis, G. C., Hasselmo, M. E., & Nalwa, V. (1989). The effect of learning on the face selective responses of neurons in the cortex in the superior temporal sulcus of the monkey. *Exp. Brain Res.*, *76*(1), 153–164.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychology Review*, *65*, 386–406.

Sathian, K., & Zangaladze, A. (1998). Perceptual learning in tactile hyperacuity: Complete intermanual transfer but limited retention. *Experimental Brain Research*, *118*(1), 131–134.

Schiltz, C., Bodart, J. M., Dubois, S., Dejardin, S., Michel, C., Roucoux, A., et al. (1999). Neuronal mechanisms of perceptual learning: Changes in human brain activity with training in orientation discrimination. *Neuroimage*, *9*(1), 46–62.

Schoups, A., Vogels, R., Qian, N., & Orban, G. (2001). Practising orientation identification improves orientation coding in V1 neurons. *Nature*, *412*(6846), 549–553.

Schoups, A. A., Vogels, R., & Orban, G. A. (1995). Human perceptual learning in identifying the oblique orientation: Retinotopy, orientation specificity and monocularity. *Journal of Physiology (London)*, *483*(Pt 3), 797–810.

Sekuler, A. B., Gold, J. M., Gaspar, C. M., & Bennett, P. J. (2001). The efficiency of upright and upside-down face recognition. *IOVS*, *42*(4), 3926.

Spiegel, M. F., & Green, D. M. (1981). Two procedures for estimating internal noise. *Journal of the Acoustical Society of America*, *70*(1), 69–73.

Tanner, W. P., Jr. (1961). Physiological implications of psychophysical data. *Annals of New York Academy of Science*, *89*, 752–765.

Tjan, B. S., Braje, W. L., Legge, G. E., & Kersten, D. (1995). Human efficiency for recognizing 3-D objects in luminance noise. *Vision Research*, *35*(21), 3053–3069.

Tolhurst, D. J., Movshon, J. A., & Dean, A. F. (1983). The statistical reliability of signals in single neurons in cat and monkey cortex. *Vision Research*, *23*(8), 775–785.

Tolhurst, D. J., Movshon, J. A., & Thompson, I. D. (1981). The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast. *Experimental Brain Research*, *41*(3/4), 414–419.

Tyler, C. W., Chan, H., Liu, L., McBride, B., & Kontsevich, L. (1992). Bit-stealing: How to get 1786 or more grey levels from an 8-bit color monitor. In B. E. Rogowitz & T. N. Pappas (Eds.), *Society of Photo-Optical Instrumentation Engineers. & IS&T—The Society for Imaging Science and Technology, 1998. Human Vision and Electronic Imaging III*: 26–29 January, 1998, San Jose, CA. Bellingham, WA: SPIE.

Vogels, R., Spileers, W., & Orban, G. A. (1989). The response variability of striate cortical neurons in the behaving monkey. *Experimental Brain Research*, *77*(2), 432–436.

Watson, A. B. (1998). Multi-category classification: Template models and classification images. *IOVS*, *39*(4), 912.

Watson, A. B., & Rosenholtz, R. (1997). A Rorschach test for visual classification strategies. *IOVS*, *38*(4), 1.

Watson, A. B., & Solomon, J. A. (1997). Model of visual contrast gain control and pattern masking. *Journal of Optical Society of America A*, *14*(9), 2379–2391.

Weiss, Y., Edelman, S., & Fahle, M. (1993). Models of perceptual learning in vernier hyperacuity. *Neural Computation*, *5*, 695–718.

Wills, A. J., & McLaren, I. P. (1998). Perceptual learning and free classification. *Quartenary Journal of Experimental Psychology B*, *51*(3), 235–270.