Available online at www.sciencedirect.com
ScienceDirect
Vision Research 47 (2007) 2170–2178
Vision Research
www.elsevier.com/locate/visres

Mechanisms of perceptual learning of depth discrimination in random dot stereograms

Liat Gantz [a], Saumil S. Patel [b], Susana T.L. Chung [a], Ronald S. Harwerth [a,*]

[a] College of Optometry, University of Houston, 505 J. Davis Armistead Building, Houston, TX 77204-2020, USA
[b] Department of Neurobiology and Anatomy, University of Texas Medical School at Houston, Houston, TX 77030, USA

Received 9 August 2006; received in revised form 17 April 2007

Abstract

Perceptual learning is a training induced improvement in performance. Mechanisms underlying the perceptual learning of depth discrimination in dynamic random dot stereograms were examined by assessing stereothresholds as a function of decorrelation. The inflection point of the decorrelation function was defined as the level of decorrelation corresponding to 1.4 times the threshold when decorrelation is 0%. In general, stereothresholds increased with increasing decorrelation. Following training, stereothresholds and standard errors of measurement decreased systematically for all tested decorrelation values. Post training decorrelation functions were reduced by a multiplicative constant (approximately 5), exhibiting changes in stereothresholds without changes in the inflection points. Disparity energy model simulations indicate that a post-training reduction in neuronal noise can sufficiently account for the perceptual learning effects. In two subjects, learning effects were retained over a period of six months, which may have application for training stereo deficient subjects.

Keywords: Perceptual learning, Random dot stereograms; Internal noise; Interocular decorrelation; Depth discrimination; Disparity energy model

1. Introduction

Perceptual learning is an improvement in performance as a result of training (with feedback) or practice (without feedback) (Chung, Legge, & Cheung, 2004; Fahle, 2005; Lu, Chu, Dosher, & Lee, 2005). Unlike sensitivity changes resulting from repetitive stimulation, such as desensitization or habituation, perceptual learning tends to persist over months, and in some cases, years (Fahle, 2005).

Perceptual learning has been shown to improve performance of visual tasks, including: Vernier acuity (Fahle, 1997; Herzog & Fahle, 1997; Saarinen & Levi, 1995; Westheimer, 2001), curvature acuity (Fahle, 1997), resolution acuity (Westheimer, 2001), bisection acuity (Westheimer, 2001), orientation discrimination (Dosher & Lu, 1999; Karni & Sagi, 1993; Polat, Ma-Naim, Belkin, & Sagi, 2004; Westheimer, 2001), reading speed (Chung et al., 2004), letter recognition (Chung et al., 2004), motion discrimination (Liu & Weinshall, 2000), contrast discrimination (Kuai, Zhang, Klein, Levi, & Yu, 2005), motion-direction discrimination (Kuai et al., 2005; Lu et al., 2005), waveform discrimination (Fiorentini & Berardi, 1980), feature detection (Ahissar & Hochstein, 1997), spatial phase discrimination (Berardi & Fiorentini, 1987), position discrimination (Li, Levi, & Klein, 2004; Li, Young, Hoenig & Levi, 2005) depth perception in random dot stereograms (Fendick & Westheimer, 1983; Kumar & Glaser, 1993; O'Toole & Kersten, 1992; Ramachandran, 1976; Ramachandran & Braddick, 1973; Skrandies & Jedynak, 1999; Sowden, Davies, Rose, & Kayne, 1996), and figure perception in random dot stereograms (O'Toole & Kersten, 1992).

Random dot stereograms (RDS), popularized as a research tool by Julesz in 1960 (Howard & Rogers,

* Corresponding author. Fax: +1 713 743 2053.
E-mail address: rharwerth@uh.edu (R.S. Harwerth).

0042-6989/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2007.04.014

2002), are a type of stimulus that has been used for human and primate stereopsis experiments (e.g. DeAngelis, 2000; Harwerth & Boltz, 1979; Julesz, 1960; Ohzawa, 1998; Schor, 1991; Walraven, 1975), with properties that are well suited for studies of perceptual learning (e.g. O'Toole & Kersten, 1992; Schmitt, Kromeier, Back, & Kommerell, 2002; Watanabe, Nanez, & Sasaki, 2001). The perception of depth in RDS arises by binocular combination of the monocular half views (O'Toole & Kersten, 1992). It has been suggested that learning plays an important role in the perception of depth in RDS because the time to perceive depth in RDS reduces with repeated observation (Fendick & Westheimer, 1983; O'Toole & Kersten, 1992; Ramachandran & Braddick, 1973; Saye & Frisby, 1975; Skrandies & Jedynak, 1999; Sowden et al., 1996). While many studies have addressed the perceptual learning of RDS in terms of features that are learned and/or transferred, to our knowledge, none have addressed the mechanisms that underlie the perceptual learning of RDS. This study was undertaken to investigate the mechanisms underlying the perceptual learning of depth discrimination in RDS.

Perceptual learning of RDS has been shown to be specific to retinal location (O'Toole & Kersten, 1992; Sowden et al., 1996). Since perceptual learning that is specific to retinal location and stimulus type is suggested to stem from plasticity changes in the neurons underlying the response (Dosher & Lu, 1999; Karni & Sagi, 1993), it is possible that learning of RDS depth discrimination is also spatially localized and stimulus dependent.

Changes in depth discrimination performance could occur as a result of changes in various internal noise sources including random neural firing rates, correspondence noise caused by imprecise combination of neural signals, signal transmission noise and noise in decision making processes. Alternatively, changes in depth discrimination performance may occur due to changes in the strength of the internal encoded relative disparity signal. This possibility is particularly relevant in cases where the binocular stimulus contains a disparity signal along with external noise. Lastly, the changes as a result of perceptual learning may alter the noise as well as the relative disparity signal in the disparity processing system. In this paper, we have used the disparity energy model (Ohzawa, 1998; Patel et al., 2003, Patel, Bedell, & Sampat, 2006; Qian & Zhu, 1997) to understand the neural substrate of the perceptual learning in RDS depth discrimination. Specifically, we sought to establish conditions within the model that are sufficient to explain the empirical data presented in this paper.

Interocular correlation in RDS describes the degree in which the elements of each monocular half view match each other (Cormack, Stevenson, & Schor, 1991). One hundred percent interocular correlation indicates that every element in one half-view is paired with a matched element in a corresponding location of the other half-view. When the half-views are generated to produce many unmatched elements, the interocular correlation is low. The unmatched elements of the RDS weaken the stimulus signal by reducing the number of elements in the stereoscopic depth plane, by adding noise to the disparity response mechanism, and by masking the disparity defined depth with ambiguous depth stimuli. Therefore, interocular correlation provides an ideal measure of signal strength for the sensory fusion of the two half views (Cormack et al., 1991).

This study measured the depth discrimination of RDS as a function of decorrelation strength before and after training. In the stimulus used for this study, *correlation* is defined as the proportion of elements of RDS that were forced to match in the two half views. The remaining unforced elements were paired randomly to be either matched (e.g. $\text{white}_{\text{LE}} \leftrightarrow \text{white}_{\text{RE}}$), anti-matched (e.g. $\text{white}_{\text{LE}} \leftrightarrow \text{black}_{\text{RE}}$) or unmatched (e.g. $\text{white}_{\text{LE}} \leftrightarrow \text{gray}_{\text{RE}}$). Similarly, *decorrelation* is defined as the proportion of unforced elements of RDS which are randomly paired. Pre- and post- training functions were compared to determine the magnitude of perceptual learning. Quantitative analyses were performed on the pre- and post- training functions to understand the mechanism underlying perceptual learning of RDS depth discrimination. In addition, the disparity energy model (Ohzawa, 1998; Patel et al., 2003, 2006; Qian & Zhu, 1997) was simulated to determine the neural mechanisms which would be sufficient to qualitatively explain the empirical data. On two subjects, post training tests were run after a period of six months to determine the long term retention of perceptual learning of RDS depth discrimination. Our results show a substantial reduction in stereothresholds as a result of the training. In most cases, the post training stereothresholds were a constant factor lower than pre- training strereothesholds for all decorrelation levels. The disparity energy model indicates that a reduction in firing rate noise in all the neurons in the model is sufficient to account for the empirical results.

2. Methods

2.1. Subjects

Seven healthy subjects, 20–36 years of age, inexperienced with RDS stimuli, were recruited. All subjects had at least 40 arcsec stereopsis as measured with a Titmus stereo test (Titmus Optical Company Inc., Petersburg, VA) in two directions (crossed and uncrossed). The research adhered to the tenets of the Declaration of Helsinki, and the experimental protocol was reviewed and approved by the University of Houston's Committee for the Protection of Human Subjects. Informed consent was obtained from the subjects and they received remuneration for their participation.

2.2. Apparatus and visual stimuli

The visual stimuli, generated using a VSG 2/3 graphics board (Cambridge Research Systems; U.K.), consisted of two vertically separated and horizontally aligned 13 × 5 arcdeg dynamic RDS, with a 0.022 arcdeg separation. The RDS, consisted of 6.7 arcmin × 6.7 arcmin elements, and different random-dot patterns, were presented on successive views at 60 Hz. Each element was black, white or gray. In each RDS pattern, half of the elements were gray and the remaining elements were black and

white with equal probability (i.e. 50% dot density). The stimulus was viewed through a liquid crystal optical shutter system that was synchronized with the monitor frame rate of 120 Hz.

The upper RDS served as a zero disparity reference of 0% decorrelation, while the lower RDS served as the test, of varying disparities. The task of the subjects was to discriminate whether the bottom half of the RDS was closer or farther, with respect to the top reference stimulus.

For 0% decorrelation, each element in one half-view was forced to have a corresponding matched element in the other half-view in every frame of the dynamic presentation. For 100% decorrelation, all elements were randomly paired in the two half-views in each frame of the dynamic stimulus. Horizontal disparities were introduced by displacing a portion of the elements in one half image with respect to the other (see Harwerth, Fredenberg, & Smith, 2003 for details).

Each session consisted of 200 trials. An auditory cue indicated the beginning of each trial, and a subject-mediated button press initiated a 500 ms stimulus presentation. During the response interval (1000 ms) that followed, subjects could either continue to press the button, to indicate that the bottom test was "far" relative to the zero disparity top reference, or release the button to indicate that the bottom test was "near" relative to the zero disparity top reference. A high frequency tone provided audio feedback for correct responses. Depending on the performance level, some subjects had five disparity magnitudes, presented as both "far" and "near" in each block of trials, and some had three. Generally, as subjects improved, the number of disparity magnitudes in each block of trials was reduced to three. The disparity range tested on each block of trials was monitored and adjusted relative to the subject's individual improvement. The decorrelation varied on pre- or post-training trials, and was fixed at 0% during the training sessions.

2.3. Design

Subjects completed a pre-training task consisting of five sessions of varying decorrelations, ranging between 0% and 80%, in 20% steps. The pre-training task was followed by training sessions of 0% decorrelation, until there was no further improvement in thresholds. In general, each training session consisted of five blocks of 200 trials. However, in the last training session, when thresholds were nearing asymptotic values, fewer blocks of 200 trials were required. The number of training trials varied individually, and ranged between 6600 and 11200 trials, with a mean of 9286 trials. Upon completion of the training, subjects completed a post-training trial, identical to the pre-training trial.

2.4. Data analysis

For each session, a psychometric function (Fig. 1) for depth discrimination was derived from the percentage of near responses as a function of disparity. Uncrossed disparities were arbitrarily assigned negative values for the purpose of constructing psychometric functions. Therefore, ideally, the psychometric function varied from zero near responses for the largest uncrossed disparities, to 100% near responses for the largest crossed disparities. Each set of data was fit with a logistic function (Berksen, 1972; Simpson, 1995) to determine the stereothreshold, taken as the semi-intraquartile range of the psychometric function (Fahle, 2005; Harwerth et al., 2003, Harwerth, Smith, Crawford, & von Noorden, 1997; Simpson, 1995).

Pre- and post-training stereothresholds (in arcmin) were plotted as a function of decorrelation on a log–log axes, and fit with the following equation:

$$y = a\left(\frac{1}{1-x} + b\right)^{z} \quad (1)$$

where y is the stereothreshold for a given decorrelation, x is the decorrelation, z represents the rate of degradation of the stereovision system as a function of decorrelation, a represents the minimum noise in the stereovision system, and b represents the tolerance of the stereovision system to decorrelation. This form of the equation was chosen to satisfy the condition that the stereotheshold approaches infinity as decorrelation tends to 100%. An inflection point was defined (arbitrarily) as the decorrelation level which yields a threshold that is $\sqrt{2}$ times higher than the threshold for 0% decorrelation.

A repeated measures ANOVA was performed to examine the effects of training on stereothresholds before and after the training for all decorrelations, and to test if the pattern of differences between thresholds for the pre- and post-training change with decorrelation (i.e. if the curves maintain the same shape before and after training).

A paired Student's T-test was performed to compare the pre-training and post training inflection points. Confidence intervals for the inflection points were calculated using a custom-written MATLAB (The MathWorks, Natick, MA) program. The program randomly selected a threshold for each decorrelation condition from the distribution of thresholds and their respective standard errors. Eq. 1 was fitted to each set of randomly selected thresholds. Only fits with a correlation coefficient greater than 0.8 were selected. For each fitted data set, the inflection point was calculated from the fitted parameters. This procedure was repeated 1000 times, and the 95% confidence intervals of the 1000 calculated inflection points were determined.

3. Results

Fig. 2 presents the learning curves of the subjects. All subjects showed a marked improvement (decreased by a factor of two or more) in stereothresholds with increasing training trials, and a decrease in the standard error of the stereothreshold measure. In other words, subjects demonstrated a consistent reduction in stereothreshold with practice.

Pre- and post-training functions relating stereothreshold to decorrelation (called *decorrelation functions* hereafter) for all subjects are shown in Fig. 3. It is clear by visual inspection of the data, that the post- learning curves are shifted downward with respect to the pre-learning curves, without a change in the shape of the curves. A two-way repeated measures ANOVA[1] confirmed this observation. As seen in Fig. 4b, log stereothresholds reduced after training and the main effect of training was significant ($F_{(\mathrm{df}=1,5)} = 35.92$, Huynh–Feldt corrected $P = .002$). There was no evidence of interaction between session (pre- vs. post-) and decorrelation ($F_{(\mathrm{df}=4,20)} = 0.24$, Huynh–Feldt corrected $P = .87$). Post-hoc comparisons of the logarithms of the pre-training to post-training stereothreshold for each decorrelation, with the alpha level maintained at 0.05, and Huynh–Feldt corrections, also demonstrated a significant decrease in the log stereothreshold after training (0%: $P = .0002$, 20%: $P = .002$, 40%: $P = .002$, 60%: $P = .0005$, 80%: $P = .0008$).

Due to the fact that the previous repeated measures ANOVA did not include all seven subjects, another repeated measures ANOVA was performed on all seven subjects for all but the 80% decorrelation condition. These results were similar to the results of the ANOVA that included six subjects for all decorrelation conditions. These

[1] The ANOVA included six (out of seven) subjects. Subject KC was excluded due to a partial pre-training data set. KC completed the 0–60% decorrelation conditions, but not the 80% decorrelation condition, due to time constraints.

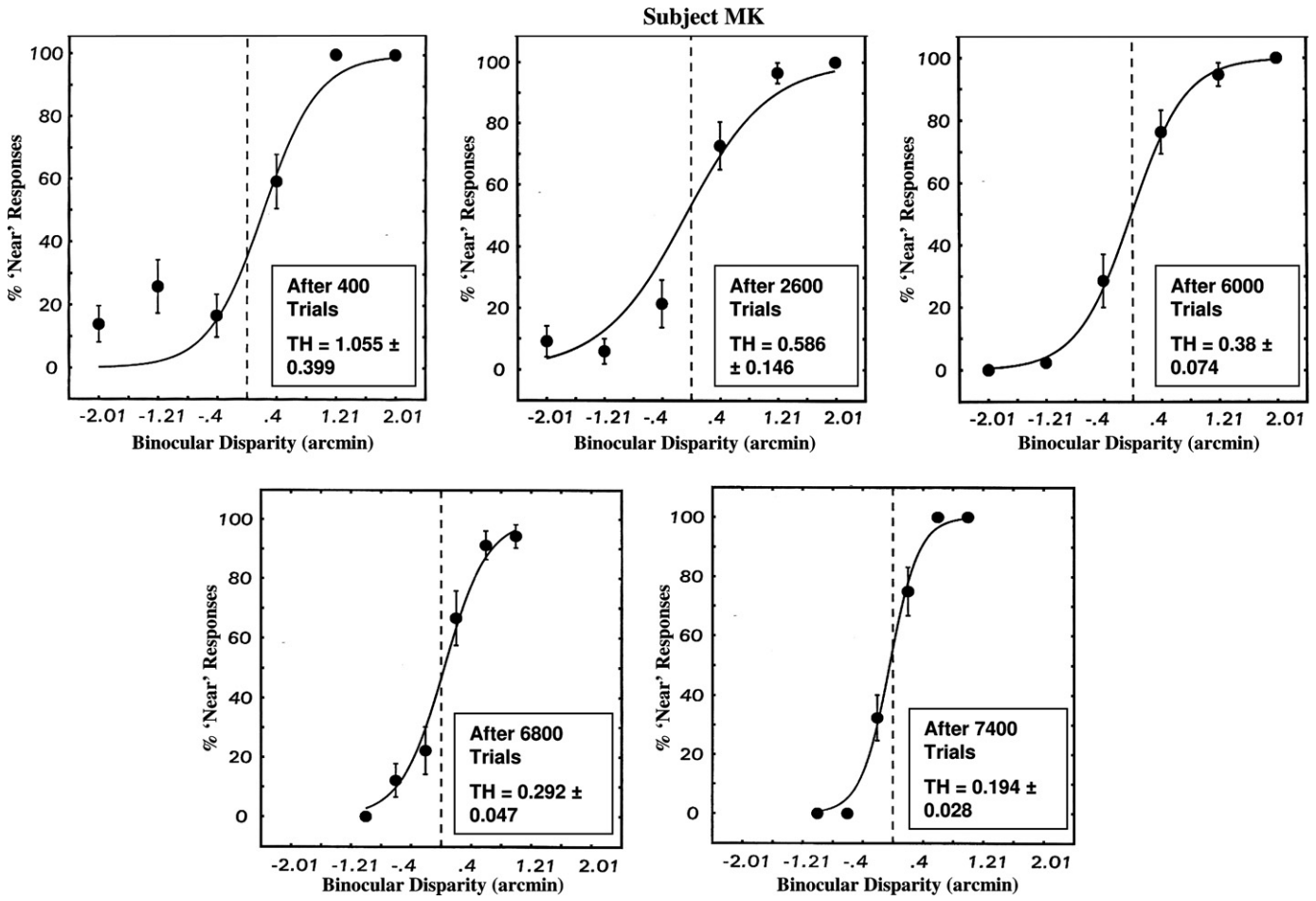

Fig. 1. Psychometric functions for performance on different sessions for subject MK demonstrate a systematic improvement with training. Slopes of the psychometric functions become steeper over a lower disparity range. The inset describes the amount of practice trials prior to obtaining the psychometric function, the threshold and standard error of the threshold measurement in units of arcminutes obtained from the psychometric function.

results imply that the training at 0% decorrelation transferred to all decorrelation values.

The inflection point of the curve, the point at which the stereovision system loses its tolerance to decorrelation, was determined arbitrarily by calculating the decorrelation which yielded a factor of $\sqrt{2}$ increase in stereothreshold. The logarithms of the pre- and post-training inflection points, as shown in Fig. 4b, were not different (Student's paired T-test $t_{(df=6)} = 0.655$; $p = .537$), suggesting that there was no evidence of a change in the stereovision system's tolerance to decorrelation as a result of training.

To examine the long term retention of perceptual learning of RDS depth discrimination, two subjects were retested using the post training paradigm six months after the completion of their training. These six month retention data are marked with inverted triangles in Fig. 3, for subjects KC and MW. Both subjects retained their post training threshold over a period of six months.

4. Discussion

It has been suggested that perceptual learning plays a role in the perception of depth in RDS because the time to perceive depth in RDS reduces with repeated observation (O'Toole & Kersten, 1992). This study was undertaken to determine the mechanisms underlying the perceptual learning of stereoscopic depth discrimination. Our results demonstrated a large change in the stereothreshold across decorrelation values, without a significant change in the stereovision system's tolerance to decorrelation. In order to understand the neural mechanism underlying the observed changes in stereothresholds, we used the disparity energy model (Ohzawa, 1998; Patel et al., 2003, 2006; Qian & Zhu, 1997) to generate simulated responses to stimuli with varying levels of decorrelation.

4.1. Simulations of disparity energy model

The disparity energy model described in Patel et al. (2006) was utilized here. There were two modifications made to the model. A noise source not utilized in previous models (Patel et al., 2006) was added to account for the tolerance of the stereovision system to low levels of decorrelation. This noise source was added in the form of a spatial jitter when simple cell signals are combined by the complex cells. In other words, a complex cell at

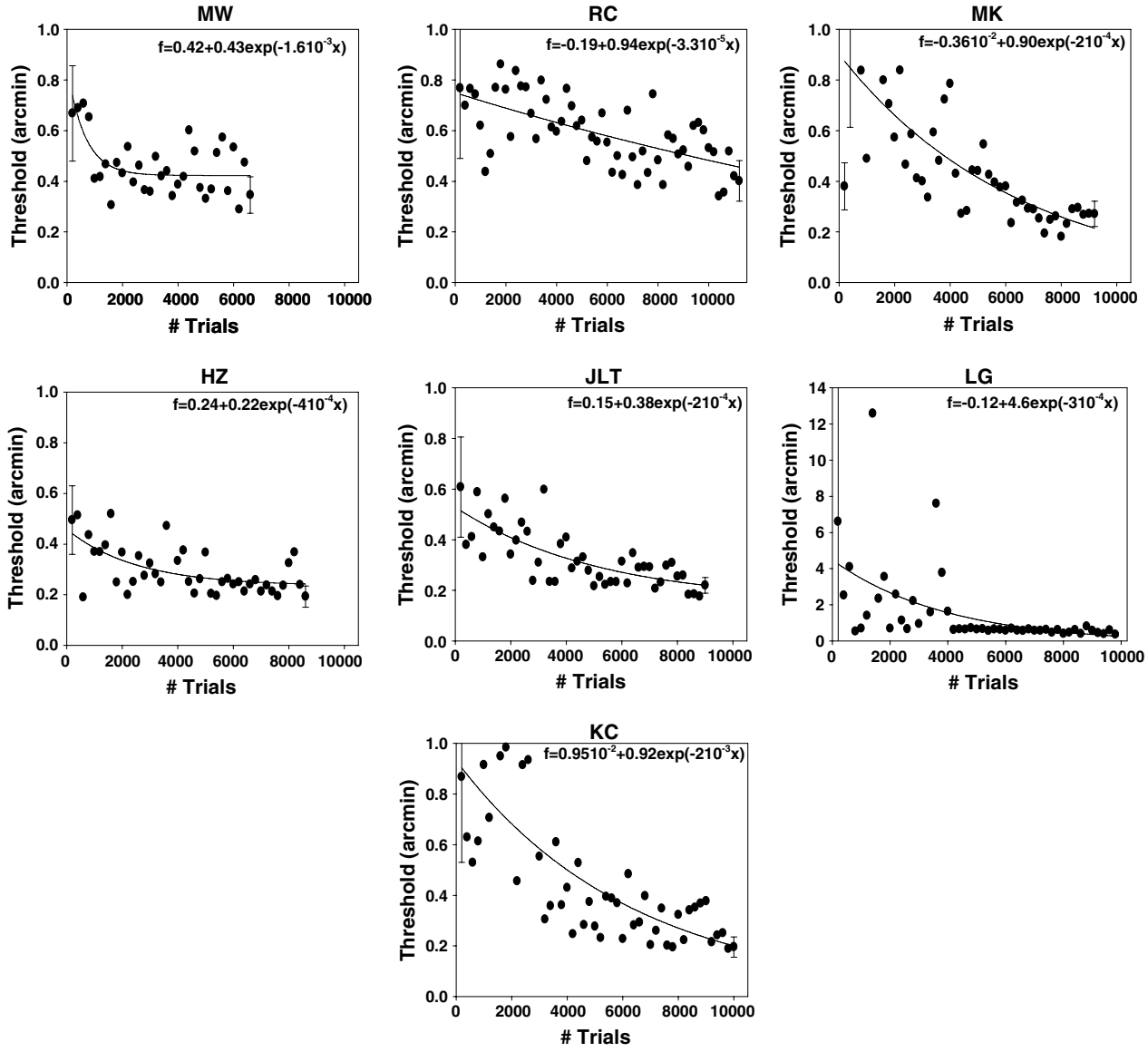

Fig. 2. Learning curves for all subjects, stereothreshold in units of arcminutes as a function of number of trials. Error bars represent standard errors of the threshold measurement of the first few training sessions and the last training session. Each data point represents a threshold obtained from 200 trials.

a spatial location (i,j) receives signals from simple cells at spatial locations $(i \pm m1,\ j \pm n1)$ and $(i \pm m2, j \pm n2)$, where $m1$, $m2$, $n1$ and $n2$ are random numbers. This type of noise is similar to the temporal jitter proposed to account for elevation of Vernier thresholds of moving stimuli (Bedell, Chung, & Patel, 2000). In addition, to reduce the simulation time, the receptive fields of monocular cells were one-dimensional horizontal versions of those used in Patel et al. (2006). Thus the monocular cells in the model used in this paper are isotropic as described by Qian and Zhu (1997). The details of simulation are available in the Appendix.

The simulation results are shown in Fig. 5. In each panel, the model's response is plotted as a function of decorrelation for different levels of neuronal noise represented by the noise multiplier α. Panels *a*, *b*, *c*, and *d* represent the signal, noise, inverse of signal to noise ratio $(1/d')$ and model fits to experimental data respectively. We assume that the psychophysical threshold is inversely related to the d′ and thus in Fig. 5d, we compare the average thresholds across observers computed from data in Fig. 3 to the scaled versions of the curves shown in Fig. 5c. As can be seen in Fig. 5d, the model simulates the decorrelation function reasonably well. In addition, the downward vertical shift of the decorrelation function as a result of training can also be explained quantitatively by a reduction in neuronal noise in the model. The model's noise multiplier parameter α was reduced from 0.02 to 0.005 to fit the pre- and post-training data respectively. We also found that systematic vertical shifts of the decorrelation function could not be obtained by any reasonable manipulations of neuronal tuning in the model. One point to be noted from

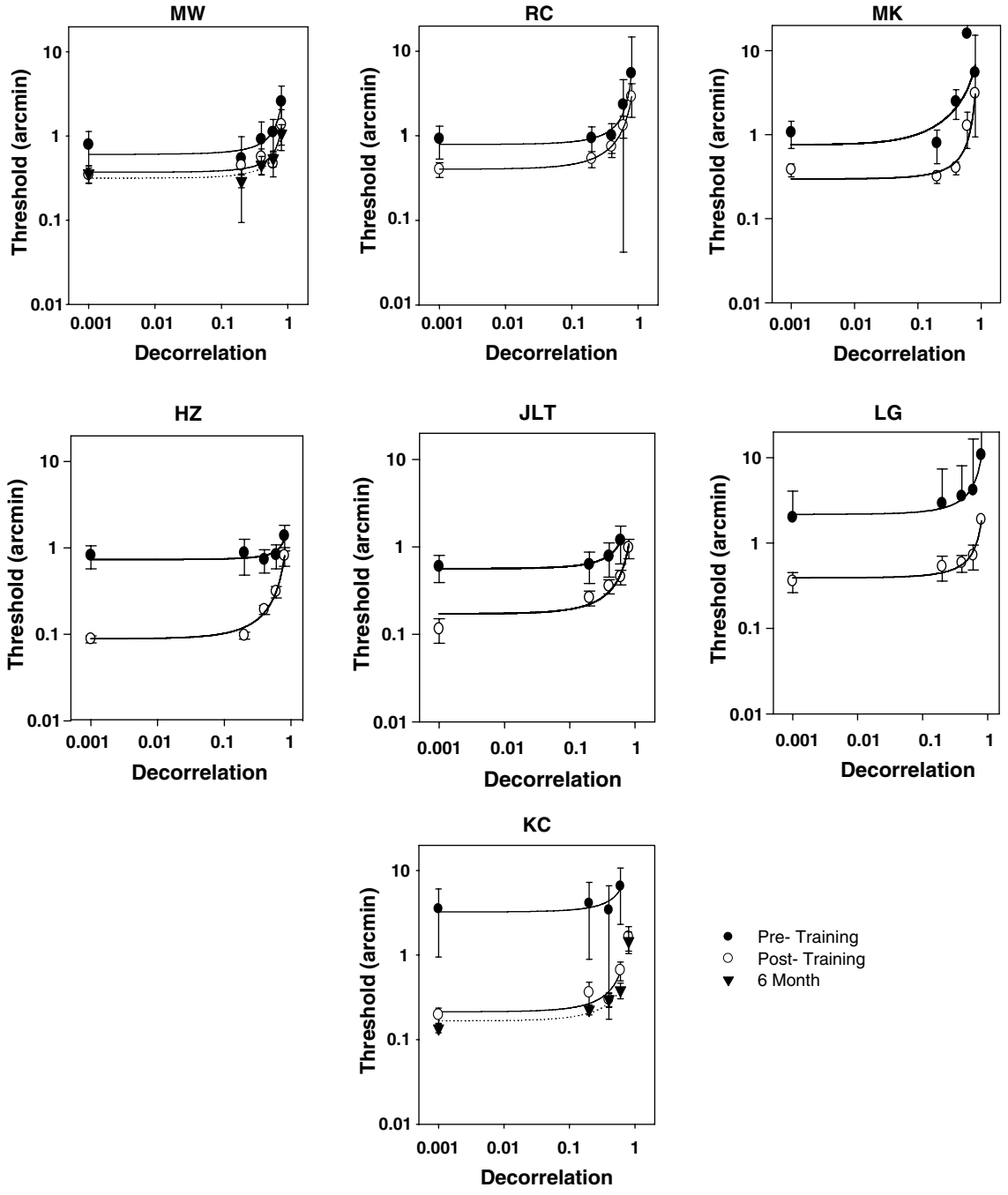

Fig. 3. Stereothresholds pre- and post-training are plotted as a function of decorrelation for each observer. The leftmost data point represents 0% decorrelation. Filled symbols represent the pre-training data, unfilled circles the post-training data, and triangles the six month data. See text for detailed explanation.

these simulation results is that a change in neuronal noise not only changes the noise in the disparity representation but it also changes the relative disparity signal. In other words, the mean and variance of the stochastic relative disparity signal are not always independent. Another point to be noted is that a multiplicative change in decorrelation function can be achieved by changes in additive neuronal noise. This occurs primarily due to the non-linear processing of complex cells in the model.

4.2. Relationship to other models of perceptual learning

In our study, the training stimuli comprised of 0% decorrelation of varying disparities. The pre- and post- training sessions included the same disparity range as the training condition, and only varied in the decorrelation. The fact that the improvement in stereothresholds at 0% decorrelation transferred to more specific and difficult conditions, comprised of higher decorrelation, is consistent

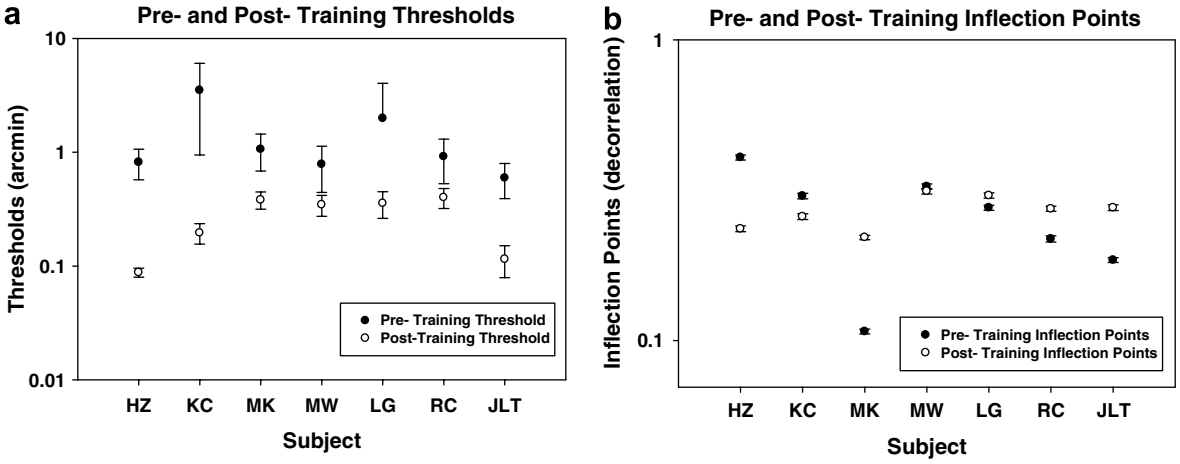

Fig. 4. (a) Pre-training (filled circles) and post-training (unfilled circles) stereothresholds for the 0% decorrelation condition are plotted for each subject. For all subjects, the post-training stereothresholds are lower than the pre-training stereothresholds. (b) Pre-training (filled circles) and post-training (unfilled circles) inflection points are plotted for each subject. The inflection points were calculated based on the fits of Eq. 1, as the decorrelation yielding a threshold that is $\sqrt{2}$ times higher than the stereothreshold for 0% decorrelation. The pre- and post-training inflection points were compared using a Student's paired T-test, and were not significantly different.

with other general models of perceptual learning (Ahissar & Hochstein, 1997, 2004; Liu & Weinshall, 2000). Based on the modeling results, it is possible that the training caused a reduction in noise in the neurons involved in disparity processing. It is premature to make any specific conclusions about where and what caused the perceptual learning of RDS depth discrimination, but we can say that a reduction of noise in early disparity processing is sufficient to account for the data presented here. There are other general learning mechanisms (Dosher & Lu, 2001, 2005; Lu et al., 2005; Saarinen & Levi, 1995) that could operate beyond the early disparity processing stages and these mechanisms may also contribute to the learning phenomenon reported here. A final possibility that cannot be excluded based on the results of this study, is that the improvement in depth discrimination is partly due to a general cognitive improvement in the ability to perform the task.

Karni and Sagi (1993) reported a 22 month and 32 month retention of performance in their texture discrimination task. The post- learning session in this study was retested in two subjects six months after they completed the study to examine the long term retention of learning. Both subjects showed remarkable six month retention of the learning. This long lasting learning phenomenon may be utilized to improve stereopsis in stereo deficient observers.

In summary, perceptual learning of depth discrimination of random dot stereograms is an adjustment within the brain which lasts at least six months. This learning does not alter the tolerance of the stereovision system to decorrelation in the stimulus but rather improves sensitivity uniformly across the entire range of decorrelation levels. Simulations of the disparity energy model suggest that a reduction in the firing rate noise in the early disparity processing neurons is sufficient to account for the empirical data.

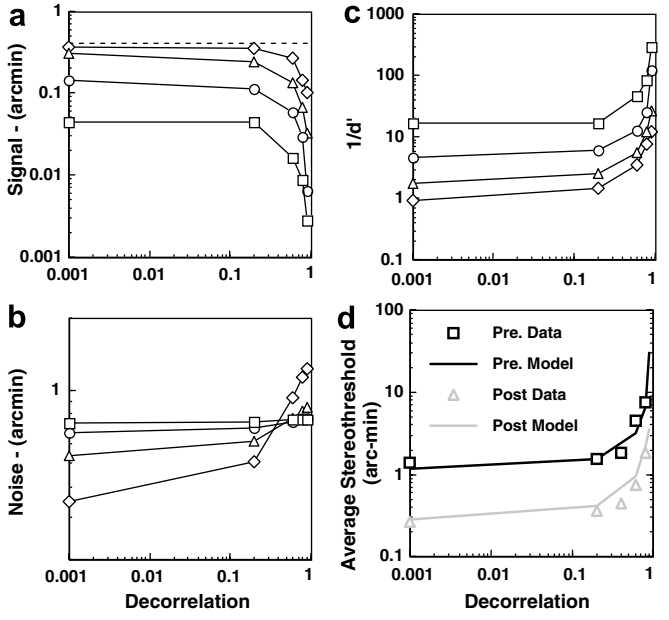

Fig. 5. Simulation results of the disparity energy model. (a) The signal present in the model's output (disparity map) as a function of decorrelation in the stimulus. The dotted horizontal line represents the disparity in the stimulus. Each curve in the panel corresponds to the internal neuronal noise multiplier α (diamond $= 0.001$, triangle $= 0.01$, circle $= 0.02$, square $= 0.04$). A larger value of α corresponds to higher firing rate noise in all the model neurons. (b) The noise present in the model's output as a function of decorrelation in the stimulus for various values of α (same symbol notation as in the top panel). (c) The inverse of the ratio of plot a (signal) to plot b (noise) and is termed $1/d'$. Details about the calculations of signal, noise and d' are available in the Appendix. (d) The stereothresholds averaged across all observers in the pre- (black squares) and post-training (gray triangles) conditions. The superimposed curves are scaled versions of the curves shown in (c). For the pre-training model curve, α is 0.02 and for post-training model curve, α is 0.005. The raw simulation results in (c) were multiplied by 0.25 to obtain the shown curves.

Acknowledgment

This work was supported by NIH Grants R01 EY01139, R01 EY12810 and P30 EY07751. We are grateful to Dr. Harold Bedell for helpful discussions.

Appendix A

A.1. Disparity energy model

The basic architecture of the model is similar to the phase disparity model that is described in Patel et al. (2006). There were two modifications made to the previous model. In the previously described model, the monocular neurons were orientation tuned. In the model used here, the monocular neurons had one-dimensional horizontal receptive fields and were thus isotropic (Qian & Zhu, 1997). This modification was done to reduce the simulation time and should not substantially affect the generality of the results. In the previously described model, the simple cells and complex cells were in complete spatial registration, meaning that a complex cell at (i,j) spatial location received signals from simple cells at (i,j) spatial location. In the model used here, a uniform random spatial jitter was introduced when signals from simple cells were combined by a complex cell. In other words, a complex cell at (i,j) spatial location received signals from simple cells at $(i \pm m1,\ j \pm n1)$ and $(i \pm m2, j \pm n2)$ spatial locations where $m1$, $n1$, $m2$ and $n2$ are uniformly distributed random numbers between -8 and 8. This noise source was added to the model to provide tolerance to low levels of decorrelation.

A.2. Stimulus used for simulations

Each binocular stimulus consisted of a pair of images (I_l and I_r) of 100×100 pixels. The size of a pixel was 2 arcmin. A pair of binary random-dot images called the seed images (S_l and S_r) were created first to yield a desired level of decorrelation. The desired level of decorrelation was set by randomly mismatching a percentage of dots in the two images. In other words, 100% decorrelation was set by having 50% of the dots matched and 50% of the dots antimatched. In order to specify a horizontal stimulus disparity of 0.2 arcmin, which is smaller than the size of a pixel, we used a simple weighting technique as given below:

$$\begin{aligned} I_l &= 0.1^{*}\mathrm{Shift}(S_l, 1) + 0.9^{*}S_l \\ I_r &= S_r \end{aligned} \tag{A1}$$

where, Shift(S_l, 1) horizontally and circularly shift S_l by 1 pixel. For convention purposes, assume that a shift of 1 (-1) pixel generates 2 arcmin crossed (uncrossed) disparity. We verified that for 0% decorrelation and in the absence of internal neuronal noise, the disparity energy model responds to this sub-pixel disparity stimulus with reasonable accuracy. In addition we also ran simulations with a stimulus that had a 1 pixel disparity, i.e. with $I_l = \mathrm{Shift}(S_l, 1)$ and $I_r = S_r$, and found that the results were qualitatively similar to those with sub-pixel disparity stimulus.

A.3. Simulation procedure

The simulations were run for various levels of decorrelation and neuronal noise multiplier α. All simulations were run for crossed as well as uncrossed stimulus disparity. For each value of decorrelation and α, the disparity energy model was run 120 times. A different set of images (I_l and I_r) was used for each run. For each disparity energy model run, the smoothed disparity map produced by the model was analyzed. The mean and standard-deviation of the disparity values in the map were computed and saved for later analysis.

A.4. Analysis of simulation data

For each value of decorrelation and α, three summary variables were computed by analyzing the mean and standard-deviation vectors of crossed $(\hat{\mu}_c, \hat{\sigma}_c)$ and uncrossed $(\hat{\mu}_u, \hat{\sigma}_u)$ disparities. Each vector had 120 elements. The summary variables computed are as below:

$$\mathrm{Signal} = \mathrm{mean}(\hat{\mu}_c - \hat{\mu}_u) \tag{A2}$$

$$\mathrm{Noise} = 0.5^{*}\mathrm{mean}(\hat{\sigma}_c + \hat{\sigma}_u) \tag{A3}$$

$$d' = 2^{*}\mathrm{mean}\left(\frac{\hat{\mu}_c - \hat{\mu}_u}{\hat{\sigma}_c + \hat{\sigma}_u}\right) \tag{A4}$$

where, mean is the arithmetic averaging function. The division operator in Eq. (A4) represents an element by element division. Signal, Noise and $1/d'$ are plotted in Fig. 5a, b, c, and d, respectively.

References

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

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Science, 8*, 457–464.

Bedell, H. E., Chung, S. T. L., & Patel, S. S. (2000). Elevation of vernier thresholds during image motion depends on target configuration. *Journal of the Optical Society of America A, 17*, 947–954.

Berardi, N., & Fiorentini, A. (1987). Interhemispheric transfer of visual information in humans: Spatial characteristics. *Journal of Physiology, 384*, 633–647.

Berksen, J. (1972). Minimum discrimination information, the no interaction problem, and the logistic function. *Biometrics, 28*, 443–468.

Chung, S. T. L., Legge, G. E., & Cheung, S. H. (2004). Letter recognition and reading speed in peripheral vision benefit from perceptual learning. *Vision Research, 44*, 695–709.

Cormack, L. K., Stevenson, S. B., & Schor, C. M. (1991). Interocular correlation, luminance contrast and cyclopean processing. *Vision Research, 31*, 2195–2207.

DeAngelis, G. C. (2000). Seeing in three dimensions: The neurophysiology of stereopsis. *Trends in Cognitive Sciences, 4*, 80–90.

Dosher, B. A., & Lu, Z. L. (1999). Mechanisms of perceptual learning. *Vision Research, 39*, 3197–3221.

Dosher, B. A., & Lu, Z. L. (2001). Mechanisms of perceptual learning. In L. Itti & G. Rees (Eds.), *Neurobiology of Attention* (pp. 471–476). San Diego: Academic Press.

Dosher, B. A., & Lu, Z. L. (2005). Perceptual learning in clear displays optimizes perceptual expertise: Learning the limiting process. *Proceedings of the National Academy of Sciences, 102*, 5286–5290.

Fahle, M. (1997). Specificity of learning curvature, orientation, and vernier discriminations. *Vision Research, 37*, 1885–1895.

Fahle, M. (2005). Perceptual learning: Specificity versus generalization. *Current Opinion in Neurobiology, 15*, 154–160.

Fendick, M., & Westheimer, G. (1983). Effects of practice and the separation of test on foveal and peripheral stereoacuity. *Vision Research, 23*, 145–150.

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

Harwerth, R. S., & Boltz, R. L. (1979). Stereopsis in monkeys using random dot stereograms: the effect of viewing duration. *Vision Research, 19*, 985–991.

Harwerth, R. S., Fredenberg, P. M., & Smith, E. L. (2003). Temporal integration for stereoscopic vision. *Vision Research, 43*, 505–517.

Harwerth, R. S., Smith, E. L., Crawford, M. L. J., & von Noorden, G. K. (1997). Stereopsis and disparity vergence in monkeys with subnormal binocular vision. *Vision Research, 37*, 483–493.

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

Howard, I. P., & Rogers, B. J. (2002). In: Seeing in depth (Vol. 2, pp. 547–549). Toronto: I. Porteous.

Julesz, B. (1960). Binocular depth perception of computer generated patterns. *The Bell System Technical Journal, 39*, 1125–1162.

Karni, A., & Sagi, D. (1993). A time course of learning a visual skill. *Nature, 365*, 250–252.

Kuai, S. G., Zhang, J. Y., Klein, S. A., Levi, D. M., & Yu, C. (2005). The essential role of stimulus temporal patterning in enabling perceptual learning. *Nature Neuroscience, 8*, 1497–1499.

Kumar, T., & Glaser, D. A. (1993). Initial performance, learning, and observer variability for hyperacuity tasks. *Vision Research, 33*, 2287–2300.

Li, R. W., Levi, D. M., & Klein, S. A. (2004). Perceptual learning improves efficiency by re-tuning the decision template for position discrimination. *Nature Neuroscience, 7*, 178–183.

Li, R. W., Young, K. G., Hoenig, P., & Levi, D. M. (2005). Perceptual learning improves visual performance in juvenile amblyopia. *Investigative Ophthalmology and Visual Science, 46*, 3161–3168.

Liu, Z. L., & Weinshall, D. (2000). Mechanisms of generalization in perceptual learning. *Vision Research, 40*, 97–109.

Lu, Z. L., Chu, W., Dosher, B. A., & Lee, S. (2005). Independent perceptual learning in monocular and binocular motion systems. *Proceedings of the National Academy of Sciences, 102*, 5624–5629.

Ohzawa, I. (1998). Mechanisms of stereoscopic vision: The disparity energy model. *Current Opinion in Neurobiology, 8*, 509–515.

O'Toole, A., & Kersten, D. (1992). Learning to see random dot stereograms. *Perception, 21*, 227–243.

Patel, S. S., Ukwade, M. T., Stevenson, S. B., Bedell, H. E., Sampath, V., & Ogmen, H. (2003). Stereoscopic depth perception from oblique phase disparities. *Vision Research, 43*, 2479–2492.

Patel, S. S., Bedell, H. E., & Sampat, P. (2006). Pooling signals from vertically and non-vertically tuned disparity mechanisms in human stereopsis. *Vision Research, 46*, 1–13.

Polat, U., Ma-Naim, T., Belkin, M., & Sagi, D. (2004). Improving vision in adult amblyopia by perceptual learning. *Proceedings of the National Academy of Sciences, 101*, 6692–6704.

Qian, N., & Zhu, Y. (1997). Physiological computation of binocular disparity. *Vision Research, 37*, 1811–1827.

Ramachandran, V. S. (1976). Learning like phenomena in stereopsis. *Nature, 262*, 392–394.

Ramachandran, V. S., & Braddick, O. (1973). Orientation specific learning in stereopsis. *Perception, 2*, 371–376.

Saarinen, J., & Levi, D. M. (1995). Perceptual learning in vernier acuity: What is learned? *Vision Research, 35*, 519–527.

Saye, A., & Frisby, J. P. (1975). The role of monocularly conspicuous features in facilitating stereopsis from random-dot stereograms. *Perception, 4*, 159–171.

Schmitt, C., Kromeier, M., Back, M., & Kommerell, G. (2002). Interindividual variability of learning in stereoacuity. *Graefe's Archive for Clinical and Experimental Opthalmology, 240*, 704–709.

Schor, C. M. (1991). Binocular sensory disorders. In D. Regan (Ed.), *Vision and Visual Dysfunction*. *Vol. 9* (pp. 179–218). Boston: CRC Press, lnc..

Simpson, T. L. (1995). Vision thresholds from psychometric analyses: Alternatives to probit analysis. *Optometry and Vision Science, 72*, 371–377.

Skrandies, W., & Jedynak, A. (1999). Learning to see 3-D: Psychophysics and brain electrical activity. *Neuroreport, 10*, 249–253.

Sowden, P., Davies, I., Rose, D., & Kayne, M. (1996). Perceptual learning of stereoacuity. *Perception, 25*, 1043–1052.

Walraven, J. (1975). Amblyopia screening with random dot stereograms. *American Journal of Opthalmology, 80*, 893–900.

Watanabe, T., Nanez, J. E., & Sasaki, Y. (2001). Perceptual learning without perception. *Nature, 413*, 844–847.

Westheimer, G. (2001). Is peripheral visual acuity susceptible to perceptual learning in the adult? *Vision Research, 41*, 47–52.