PERGAMON

Vision Research 42 (2002) 3019–3030

Vision Research

www.elsevier.com/locate/visres

Perceptual learning of temporal structure

C. Aslin, R. Blake *, M.M. Chun

Vanderbilt Vision Research Center, Department of Psychology, Vanderbilt University, 111 21st Avenue S, Nashville, TN 37203, USA

Received 17 December 2001; received in revised form 13 May 2002

Abstract

We investigated the extent to which the ability to perceive spatial form from temporal structure (TS) improves with practice. Observers trained monocularly for a number of consecutive days on a shape discrimination task, with one group of observers judging shape defined by luminance contrast between target and background elements and another group judging shape defined by correlated TS (synchronized changes in motion direction between target and background elements). Substantial learning was found for both shape tasks, with complete interocular transfer of training. Observers trained on TS showed no transfer of learning to the luminance condition, but observers trained using the luminance display with incidental synchronized changes did show transfer to the TS task. Possible underlying neural changes are discussed.

1. Introduction

The work described in this paper concerns the ability of human observers to learn to perceive spatial structure (i.e., shape) based on temporal structure (TS) (i.e., synchronized change among stimulus elements). Our work begins with the observation that people can visually segregate a figure from its background on the basis of differential TS among stimulus elements defining figure and background. The human visual system's remarkable sensitivity to fine temporal information contained within dynamic visual displays is well established (Blake & Yang, 1997; Ross & Hogben, 1974; Westheimer & McKee, 1977). This sensitivity to temporal dynamics can be utilized in various forms of visual grouping. For example, a subset of dots (Ramachandran & Rogers-Ramachandran, 1991; Usher & Donnelly, 1998) or oriented contours (Fahle, 1993; Leonards, Singer, & Fahle, 1996) group together perceptually to form a boundary or a figure when those dots or those contours are rapidly flickered out of phase within a surrounding background of similar flickering elements. Grouping from common TS has also been demonstrated using displays in which elements change direction of motion irregularly over time, with points in time at which "figure" elements change direction differing from points in time at which "background" elements change direction (Farid & Adelson, 2001; Kandil & Fahle, 2001; Lee & Blake, 1999a). It should be noted that TS is not always effective as a grouping cue (Fahle & Koch, 1995; Kiper, Gegenfurtner, & Movshon, 1996), and there is disagreement about the details of the process responsible for time-based grouping when it does occur (Adelson & Farid, 1999; Lee & Blake, 1999b; Morgan & Castet, 2002).

We have observed that observers may initially experience difficulty segregating figure from ground based on TS alone, analogous to the notorious challenge that confronts observers when trying to decipher complex random-dot stereograms for the first time (Julesz, 1971). Still, just as binocular disparity provides sufficient information to segregate surfaces in depth, there is no denying that, at least under some circumstances, "timing" alone can provide reliable information about the spatial configuration of distributed stimulus elements. During work in our laboratory, we have observed that novice observers improve in temporal segregation tasks with repeated exposure to the dynamic stimulus sequences. It was this observation that motivated us to study the improvement in this ability with practice and the transfer of this improvement to perception of spatial structure defined by luminance.

* Corresponding author. Tel.: +1-615-343-7010; fax: +1-615-343-5027.
E-mail address: randolph.blake@vanderbilt.edu (R. Blake).

0042-6989/02/$ - see front matter
PII: S0042-6989(02)00386-3

The study of visual perceptual learning has a long history in perception psychology (Gibson, 1953), with interest in the problem accelerating in the last decade or so (Sagi & Tanne, 1994). It is well established that people get better with practice on a wide variety of visual tasks, including orientation discrimination (Matthews, Liu, Geesaman, & Qian, 1999), motion perception (Ball & Sekuler, 1982; Zanker, 1999), vernier acuity (Fahle, Edelman, & Poggio, 1995), spatial frequency discrimination (Fine & Jacobs, 2000), global stereopsis (Frisby & Clatworthy, 1975) and object recognition (Furmanski & Engel, 2000; Gauthier, Williams, Tarr, & Tanaka, 1998; Sinha & Poggio, 1996)—Fine and Jacobs (2002) provide an up to date, comprehensive overview of this literature. Many of these recent studies have been framed within the context of neural plasticity, the strategy being to document the degree to which improvement following training on a given set of stimulus conditions generalizes to other conditions. Thus we find instances where learning is highly specific for visual field location (Ahissar & Hochstein, 1996; Beard, Levi, & Reich, 1995; Fahle et al., 1995), for the trained eye (Karni & Sagi, 1991, 1993) and for specific orientations, spatial frequencies or directions of motion (Ball & Sekuler, 1982; Fahle, 1997; Fiorentini & Berardi, 1981; Karni & Sagi, 1991). Selective improvement on these kinds of "basic" visual tasks has been interpreted as evidence for plasticity within neural mechanisms early in visual processing, although the rationale underlying this interpretation has been questioned (Mollon & Danoliva, 1996). Conversely, visual learning that generalizes beyond the original training conditions has been attributed to plasticity within "high level" visual mechanisms (Ahissar & Hochstein, 1997; Sireteanu & Rettenbach, 2000; Wehrhahn & Rapf, 2001). There is some evidence that the specificity of perceptual learning varies with the difficulty of the visual task being mastered (Ahissar & Hochstein, 1997).

In our work, the focus is not on *where* learning occurs, i.e., early vs late in visual processing, but, instead, on *what* is learned as revealed by improvements in one's ability to use time-varying changes in the optical input to vision for visual grouping. We are interested, in other words, in the extent to which practice increases sensitivity to TS. We assume that changes in the optical input produce time-varying neural responses to that input. Thus performance on grouping tasks involving dynamic stimulation ("temporal structure" as we term it) must depend, at least in part, on the ability of observers to extract information about dynamics. We do not hypothesize the existence of unique mechanisms specialized for registering TS but, instead, we assume that TS is an inherent component of the visual system's response to structured, dynamic optical input. We assume, moreover, that improvement in visual grouping based on TS reflects refinements in the ability to extract and exploit the dynamic information contained in the time-varying neural activity within the visual pathways. Our study examines the extent to which practice promotes perceptual learning of shape recognition based on TS and the degree to which such learning transfers to stimulus conditions other than those utilized during training.

2. Experiment 1: temporal structure and dynamic luminance contrast

This experiment examined the extent to which practice improved performance on a shape discrimination task, where the shape was defined by TS or by dynamic luminance contrast (LUM). In addition, we evaluated transfer of training between these conditions. Observers were assigned to one of two training conditions, TS or LUM. The TS group was trained monocularly on a shape discrimination task for which the target shape was defined solely by TS. These observers were then tested for transfer of learning to the untrained eye, and for transfer to a comparable shape discrimination task (LUM display) in which the target was defined by LUM. Similarly, the LUM group was trained monocularly on the LUM display in which the target region was defined by a difference in the average luminance of the target region and the background region. These observers were then tested for transfer to the untrained eye, and for transfer of learning to the untrained TS display.

In this experiment, we have deployed TS animations that preclude the potential LUM cue discussed by Adelson and Farid (1999). These authors pointed out that low-pass temporal filtering could recover spatial structure from the sorts of displays devised by Lee and Blake (1999a), by creating brief episodes (i.e., single animation frames) during which the average LUM of a "figure" region would differ from the average LUM of the "background" region. This LUM cue could arise if a subset of equal contrast gratings all underwent extended periods of uninterrupted motion in one direction or if those gratings underwent several successive reversals in direction of motion. While it is arguable whether this potential cue is actually realized by the visual system (Lee & Blake, 1999b), it is straightforward to create modified TS displays in which this cue cannot arise from temporal integration—these modified displays contain no successive animation frames embodying "runs" or "jitter" and, in addition, the displays contain random variations in contrast and in luminance. With these revised displays, low-pass temporal filtering fails to uncover LUM correlated with target location (Lee & Blake, 1999b). Described in greater detail in the Methods section, these modified displays were used for the TS displays used in the perceptual learning experiments reported here.

3. Methods

3.1. Apparatus

The displays were presented on an iMac computer screen viewed in an otherwise dark room. The 14″ screen had a resolution of 600×800 and a refresh rate of 95 Hz. Observers wore a patch over one eye and used a chin rest to stabilize viewing distance. Responses were made on a standard computer keyboard.

3.2. Stimuli

The stimuli, created using Matlab and the Psychophysics Toolbox (Brainard, 1997), consisted of a square array of small circular elements appearing against a homogenous gray background (Fig. 1). At the viewing distance of 72 cm, the 576 (24×24) element array subtended 6.36×6.36 degrees of visual angle. Each element consisted of a one-dimensional sinusoidal grating viewed through a circular 'window' subtending 16′ visual angle. The orientation of each element was randomized for every trial, as was the luminance with an amplitude of ± 3.94 cd/m^2 around the background gray 70 cd/m^2. Contrast of each element was randomly modulated

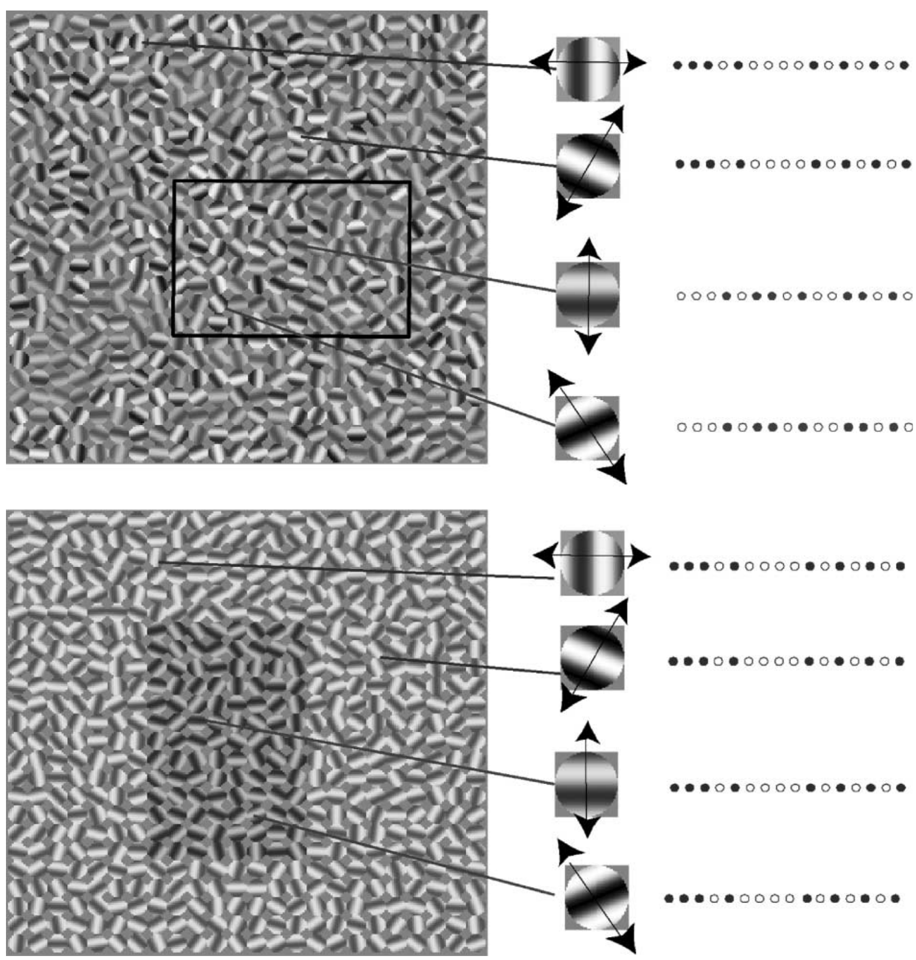

Fig. 1. The display comprised a square array of small circular elements. Each element was a phase-shifting sinusoidal grating, presented behind a circular window, with random orientation. In the TS display (top panel) a rectangular target region was defined by TS (target boundary shown here only for demonstration). The four elements shown at the right represent elements from the background and target regions. Within each region all elements had an identical TS determined by stochastic point processes (open and closed circles at right) that define the polarity of the direction of motion in each frame. The target and background point processes differed to varying degrees, creating differential TS that could mediate perception of a rectangular "target" if the difference in TS between figure and ground was sufficiently strong (where "difference" was indexed by the correlation between the two point processes). In the LUM display (bottom panel) the target region was defined by a difference in average luminance between gratings defining the target vs those within the background. In Experiment 1 all elements in the target and background regions obeyed the same point process (as illustrated on the right), creating coherent TS that was unrelated to the location and shape of the target. In Experiment 2, each grating had its own point process (not illustrated), thereby precluding coherent TS anywhere within the display.

during each trial on a frame by frame basis with a mean contrast of 0.5 and an amplitude of ± 0.2. Each element's sinusoidal grating was phase shifted by $2\pi/6$ radians per frame, a spatial displacement sufficient to produce smooth apparent motion of the grating within the circular aperture, with the direction of motion being in either of the two directions orthogonal to the grating's orientation. The direction of the phase shift in each frame was constrained so that no three consecutive frames contained either alternations between positive and negative phase shifts, or continuous phase shifts in a single direction. This constraint, as well as the randomization of contrast and luminance among elements, was implemented to preclude any potential LUM artifacts attributable to a temporal low-pass filter of the sort discussed by Adelson and Farid (1999) and by Lee and Blake (1999b).

The target in each task was a 13×8 element region whose long axis was oriented either horizontally or vertically; this "rectangular target" could appear anywhere within the central 20×20 element region of the 24×24 element array. The location and orientation of the target were random in each trial. Each trial was presented for 526 ms (50 video frames always synchronized to the video monitor refresh cycle). The target region was defined either by LUM or by TS—these two stimulus conditions are described in the next two subsections.

3.3. Temporal structure display

In the TS display the target and background regions were defined solely by differences in the points in time at which grating elements in the target vs the background reversed directions of motion. For any given grating we define these reversal times as a *point process*: a time series specifying the irregular, frame-to-frame sequence of motion directions of that grating. All gratings within the 13×8 element rectangular "target" region obeyed the same point process, meaning that all moving gratings reversed direction of motion at the same time. Likewise, all gratings within the background region shared the same point process, thus reversing their directions of motion simultaneously over time. It is important to stress that the orientations of all target and background elements, and thus the associated directions of motion, were random throughout the array—there existed no spatial structure specifying the location or the shape of the target relative to the background. Instead, target and background were distinguished solely by the degree of correlation between the two point processes associated with these regions. When the point process associated with the target elements is uncorrelated with the point process for the background elements, the target is maximally distinctive in location and shape. At the other extreme, when the two point processes are perfectly correlated (i.e., identical), the two regions share the same TS and, hence, the target is perforce invisible. Manipulating the correlation between the target point process and the background point process produces graded variations in the clarity of the target and, hence, in the observer's ability to judge the target's orientation on the 2AFC task ("horizontal" vs "vertical"). Operationally, different degrees of correlation were produced by generating a point process assigned to the target elements and then shuffling that point process to produce a specified correlation value (± 0.02 correlation units). This procedure does not introduce any kind of structured phase shift (i.e., uniform delay) between the two sets of elements but, rather, manipulates the percentage of direction reversals that are synchronized between target and background.

3.4. Luminance display

In the LUM display the average luminance of the "target" elements was different from the average luminance of the "background" elements. On half of the trials the target region had an average luminance higher than the background region, and on the remaining trials the target had a lower average luminance than the background. The magnitude of the luminance contrast, the difference in average luminance between the rectangular target and background, determined the visibility of the target. LUM ranged from 0.26 to 0.01. During each trial the difference in average luminance between the target and background regions ramped up to its maximum and back to zero following a gaussian distribution. This modulation was included to mimic the time-course of perception in the TS display. All elements in the display shifted in phase from frame to frame of the animation and reversed direction of phase shift irregularly during the animation, in the same fashion as the TS display; however, all elements throughout the entire LUM display obeyed the same point process, producing uniform coherent TS across the entire display. This TS in the LUM condition contained no information relevant to the location of the target.

3.5. Observers

Observers were paid volunteers, recruited from the graduate student population at Vanderbilt University; all gave informed consent to participate in this study, which was approved by the Vanderbilt University IRB. All observers had normal or corrected to normal visual acuity, and none had previous experience with TS though some had participated in unrelated psychophysical experiments. Each observer was randomly assigned to a given training condition, and none participated in more than one training condition.

3.6. Procedure

While seated in a darkened room, observers indicated by keypress the orientation of a "rectangular target" region within the square array of elements. Each experimental condition was explained to observers prior to the beginning of each session. Viewing was always monocular and stable head position was maintained with a chin rest at a viewing distance of 72 cm. The dominant eye, determined by a monocular sighting test, was assigned randomly to the trained or untrained condition for each observer. Observers were free to look anywhere within the display during each stimulus presentation. Each trial consisted of a timed presentation of the stimulus followed by a blank screen. Observers controlled the presentation of trials by pressing a key to initiate the next trial. No limitations were imposed on the inter-trial duration. Trial-by-trial feedback was not given during training or testing, but a graphical representation of performance was presented at the end of each block. Sessions lasted approximately 25 min and were completed once per day per training condition.

3.7. Design

During each daily session, observers completed four blocks of 60–80 trials administered in a standard 3:1 staircase procedure that converges onto the stimulus level yielding 81%-correct performance. Each staircase began with an easily recognized version of the target, with progressively difficult presentations being introduced contingent on the observer's performance on the 2AFC task. Specifically, following three consecutive correct responses, the staircase progressed to a more difficult stimulus, and following each incorrect response the staircase moved to an easier stimulus. Initial staircase steps were relatively large, but after the first incorrect response step sizes were reduced to smaller increments and decrements, with the staircase being terminated after twelve reversals in staircase direction; the average stimulus value associated with the last eight reversals was taken as the threshold estimate for that staircase run, and an overall threshold for that training session was computed from the average of the four blocks.

Observers began the experiment with 30 practice trials to gain familiarity with the task. Pre-training thresholds were then measured for the non-training conditions. Observers next completed a number of daily practice sessions with the "trained" condition (with the condition defined by the cue specifying target shape: "temporal structure" or "luminance"). Beyond a minimum of seven training sessions, no pre-arranged point of termination was set because we wished to maximize the amount of learning for each person. For several observers, day-to-day fluctuations in performance were substantial, and for these individuals more training sessions were administered. For all individuals, regardless of condition or day-to-day variability, training continued until improvements in performance leveled off, with three consecutive sessions showing no significant changes. At the conclusion of training observers were again tested for their thresholds in each condition, trained and untrained.

3.8. Results and discussion

Data from six observers were collected in Experiment 1, three observers in each condition. For the LUM condition, threshold was defined as the LUM associated with 81% correct performance; thus learning was characterized by lower LUM values at the end of training. For the TS condition, 81% correct thresholds were defined in terms of correlation between TS in target and background regions. We took the complement of this correlation value ($1 - \%$ correlation) as the performance index in order to match the direction of improvement with the LUM condition (and, thereby, insuring that for both conditions the percent change in performance with learning was expressed relative to the initial level of performance at the beginning of training—learning on both of the tasks is reflected as a decrease in threshold). To express "improvement" on a common scale, we normalized the performance measures by computing the difference in pre- and post-training thresholds divided by the sum of these two thresholds.

The ability to judge shape defined by TS and by LUM improved substantially with training: regardless of their initial level of performance, all six observers produced post-training thresholds that were significantly lower than their pre-training thresholds (see Figs. 2 and 3). Comparing the individual threshold estimates obtained during each daily training session (recall that each training session comprised four successive staircases), there is no evidence for performance improvement within a given session—thresholds for the last staircase of a session were, on average, 0.036 log-units higher than those for the first staircase of a session. This observation is not unprecedented (e.g., see Mednick et al., 2002), and it suggests the operation of some form of consolidation process for which sleep may be important (Karni & Sagi, 1993; Stickgold, 1998).

For some observers in both conditions, successive daily thresholds occasionally exhibited an "oscillatory" pattern, with average performance being poorer on one day than it was on the previous day. This kind of

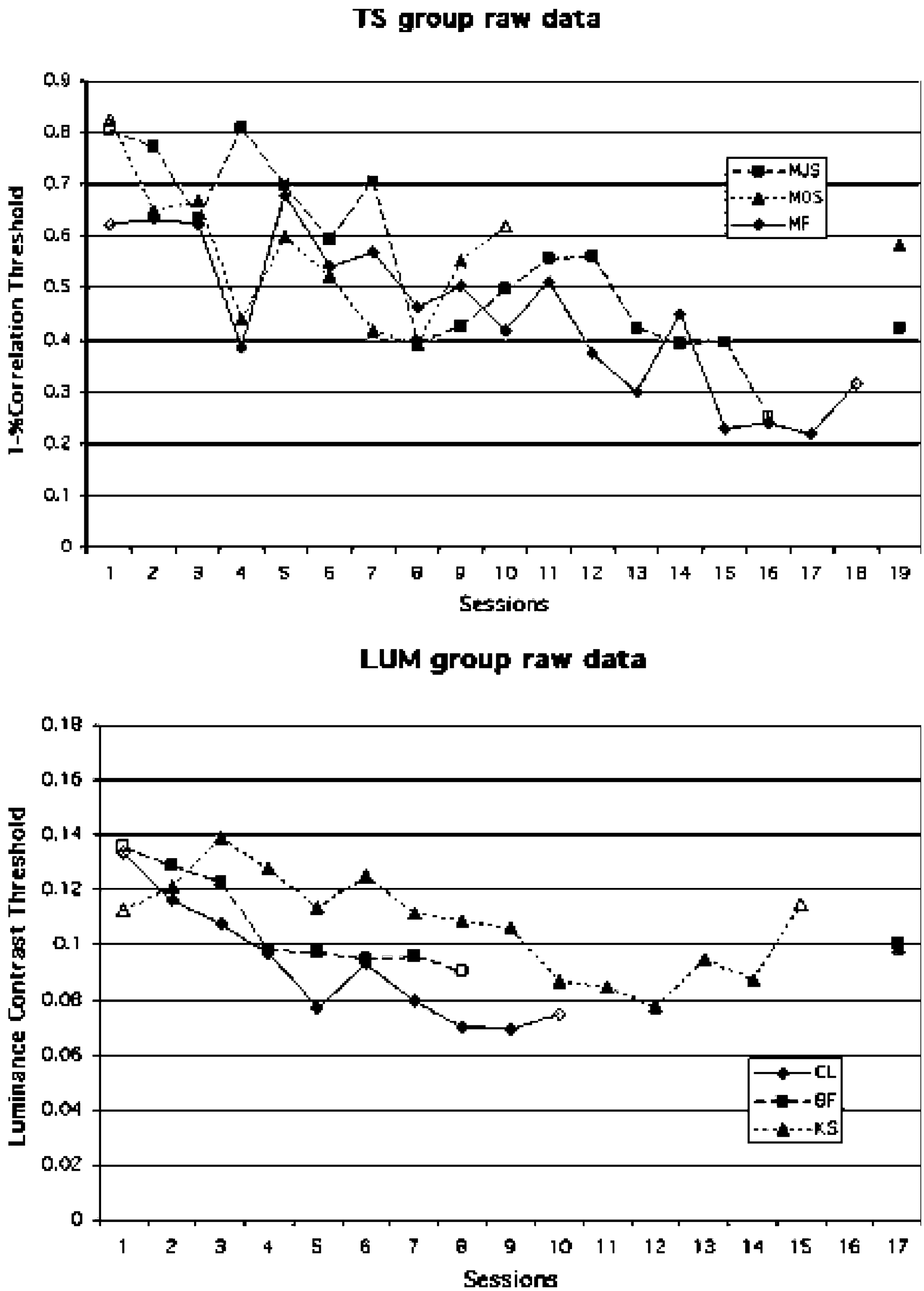

Fig. 2. Daily threshold estimates (average of four staircase runs per day) for observers in the TS training group (top plot) and LUM training group (bottom plot). All six observers in both groups improved with training on their respective training tasks. Filled symbols show threshold measurements during the training phase of the experiment; open symbols show the threshold measurements for the pre- and post-training tests with the untrained eye. Filled symbols to the far right show performance of two observers in each group 4–6 months after training.

behavior was also described by Herzog and Fahle (1999) in their study of vernier acuity and perceptual learning, under conditions where feedback was not provided or where feedback was purposefully incorrect. We did not provide trial-by-trial feedback, although observers did see a graphic summary of their results following each training session. Despite these occasional oscillations, however, the overall learning trends were abundantly evident in all observers trained on the TS and LUM conditions. Moreover, the performance improvements evidenced over sessions were quite enduring: two observers from both conditions were retested 4–6

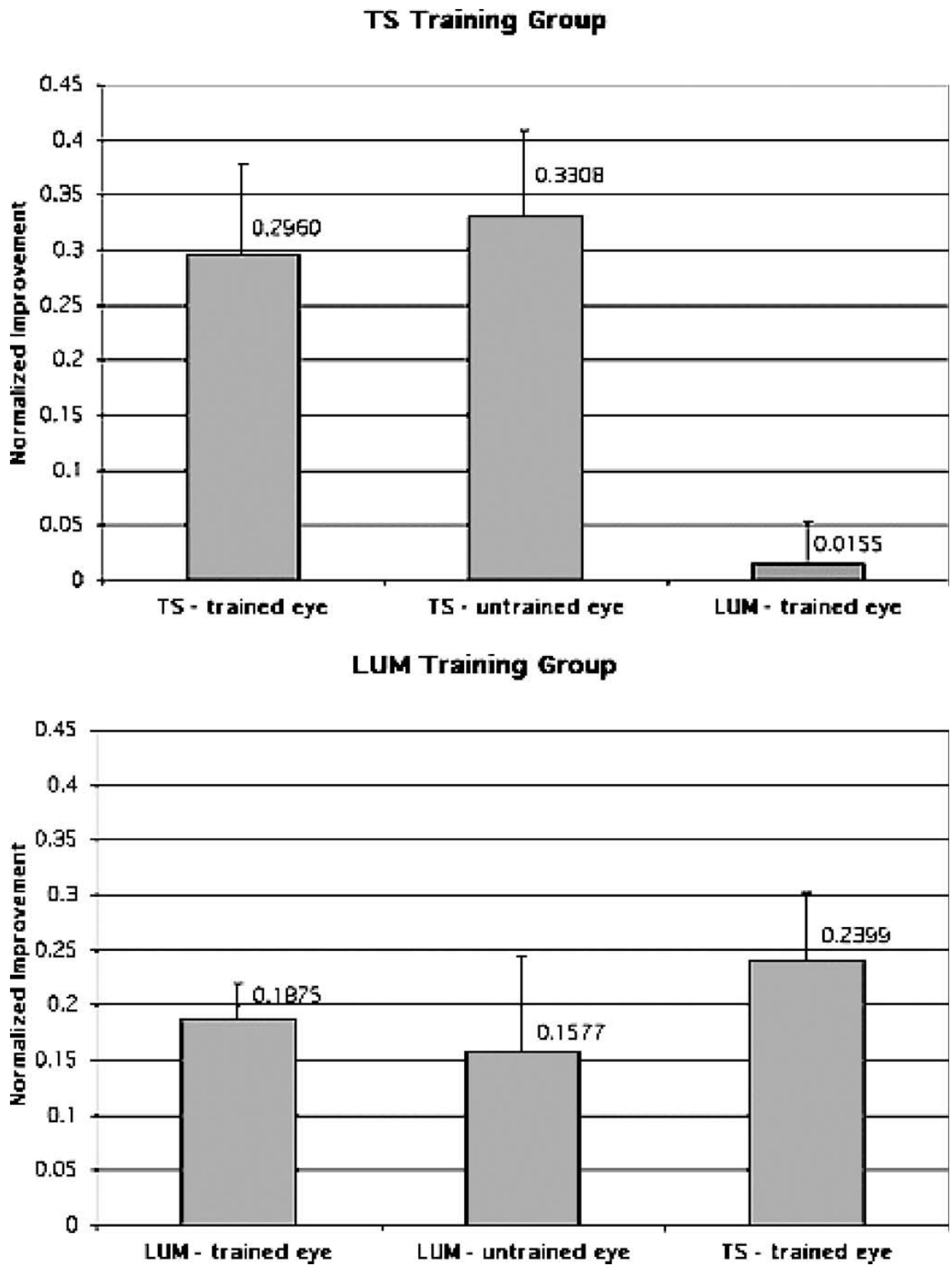

Fig. 3. Average percent improvement for the TS training group (top plot) and for the LUM training group (bottom plot). Improvement is expressed as the difference in pre- and post-training thresholds divided by the sum of the pre- and post-training thresholds. In the TS training group ($n = 3$) observers showed large improvements in the trained condition with both the trained eye ($t = 3.566$, $p < 0.05$) and the untrained eye ($t = 4.257$, $p < 0.05$). No significant improvement was found for the LUM task ($t = 0.398$, $p > 0.05$). In the LUM training group ($n = 3$) significant improvement was found for the trained condition with the trained eye ($t = 5.699$, $p < 0.05$). Improvement with the untrained eye was not statistically significant due to one observer who showed no interocular transfer ($t = 1.832$, $p > 0.05$). A very robust improvement occurred with the untrained TS condition ($t = 3.857$, $p < 0.05$). Error bars represent mean standard error.

months after the last training session, and their performance had not deteriorated significantly. When queried during the training period, observers reported no changes in their viewing strategies coincident with their improvements in performance—indeed, from their standpoint the task remained challenging throughout training because the staircase procedure guaranteed that most trials involved display conditions supporting good but not perfect performance. We see no reason to attribute these improvements with training to non-perceptual factors (e.g., motivation). Moreover, both training groups (with the exception of one observer [1] in the LUM condition) showed essentially complete interocular transfer of learning (Fig. 3, center histograms). Thus like many other forms of perceptual learning, the neural plasticity underlying performance on these tasks probably occurs at a binocular site in the visual system, meaning a locus beyond the predominantly monocular input layers in visual area V1 (assuming homology between human and macaque visual systems).

Turning next to the patterns of transfer between the two training regimes, there was an intriguing dissociation in the transfer of learning (Fig. 3, right-hand histograms): observers in the TS training group showed no post-training improvement with the LUM display, but observers in the LUM training group exhibited a very large post-training improvement with the TS display. What is the basis of this somewhat counterintuitive pattern of results?

It is possible that observers in the LUM condition, while relying on luminance differences to judge shape, were also benefiting from their repeated exposures to coherent TS in the LUM displays. Recall that all elements in the LUM displays were undergoing dynamic, correlated changes in direction of motion (although this coherent TS was unrelated to the shape of the target). Perhaps this prolonged exposure to coherent, dynamic displays throughout the training period enhanced sensitivity to TS, making it more salient during the post-training TS condition. Observers trained on the TS display, on the other hand, did not have incidental exposure to coherent luminance during training, which could account for the absence of transfer of TS training to LUM testing. From previous work it is known that incidental perceptual learning can occur under some conditions (Watanabe, Nañez, & Sasaki, 2001) but not under others (Shiu & Pashler, 1992).

If transfer from LUM to TS is indeed attributable to incidental exposure to coherent TS during LUM training, we would expect no transfer if the LUM training conditions were devoid of coherent TS. To test this hypothesis, we performed a second experiment.

4. Experiment 2: randomized luminance contrast

To test the contribution of TS in the LUM display to improvement on the untrained TS task, we created a new randomized luminance contrast (rLUM) display in which the motion reversal times associated with each grating patch were dictated by a point process randomly generated for each grating. Consequently, the average correlation among dynamic events within the display was zero—TS was incoherent throughout the display. A new group of observers was tested for transfer of learning to the untrained eye and the untrained TS display after training with the rLUM display.

Data were collected from four new observers with normal or corrected-to-normal visual acuity. All observers were naïve to the purpose of the study and had no prior experience with TS displays. The rLUM display itself was identical to the LUM display except for its TS. Each element in the rLUM display was assigned a random point process such that there was no coherent TS present in the display. Each element exhibited constant stochastic apparent motion, but the temporal pattern of this motion was completely random. All other aspects of the display were identical to the previously described LUM display, i.e., the target region was defined by a difference in average luminance from the background region.

4.1. Results and discussion

As can be seen in Fig. 4, three of four observers in the rLUM group showed modest improvements after training, with the exact amount of improvement varying considerably among observers, which is not unusual in perceptual learning experiments (Fine & Jacobs, 2002); again, evidence for interocular transfer of learning was found. We have no ready explanation for the more modest levels of improvement produced by training on the rLUM task, compared to the LUM display of Experiment 1.

Of relevance to the hypothesis under test, there was no significant improvement on the TS task after training with the rLUM display. One observer's performance actually decreased considerably after training, accounting for the large error bar seen in the right column of Fig. 5. However, even when this observer's data are removed from the analysis, no significant improvement ($M = 0.1005$; $t = 2.42$, $p > 0.05$) is seen on the untrained TS task.

One might argue that observers in the rLUM training group failed to exhibit transfer to the untrained TS task because their learning on the rLUM condition was too meager. We are disinclined to believe this, however. The observer with the largest improvement on the rLUM task actually showed the smallest improvement on the TS task, and similarly the observer with the largest improvement on the TS task exhibited the smallest improvement on the rLUM task. Thus the amount of learning on the rLUM task does not reflect the degree of improvement on the TS task. One might also argue that

[1] One observer showed no transfer of learning from the trained to the untrained eye, in marked contrast to our other observers. After completion of the experiment, we learned that this individual, although exhibiting normal corrected acuity in the two eyes, was a unilateral myope as a child and used one eye for near tasks and the other for distance tasks. This dissociation may have adversely affected binocular neural mechanisms, which are thought to affect interocular transfer (Banks, Aslin, & Letson, 1975).

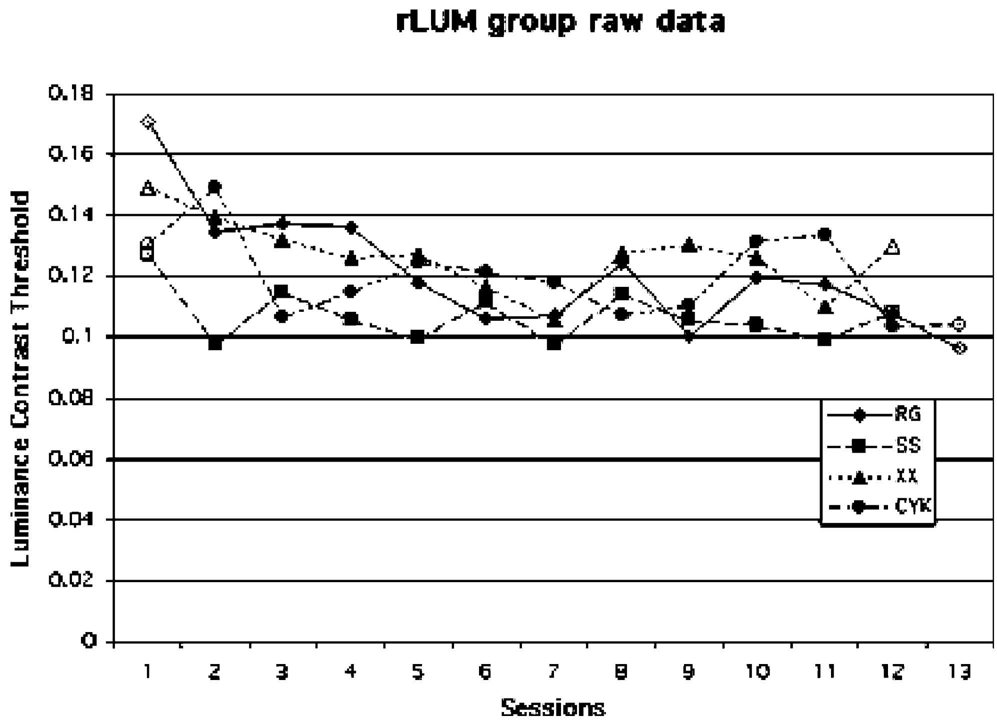

Fig. 4. Raw data for observers in the rLUM training group. Three out of four observers showed improvement on the trained task, though this learning was less robust than the LUM training group from Experiment 1. Filled symbols show threshold measurements during the training phase of the experiment. Open symbols show threshold measurements for the pre- and post-training tests with the untrained eye.

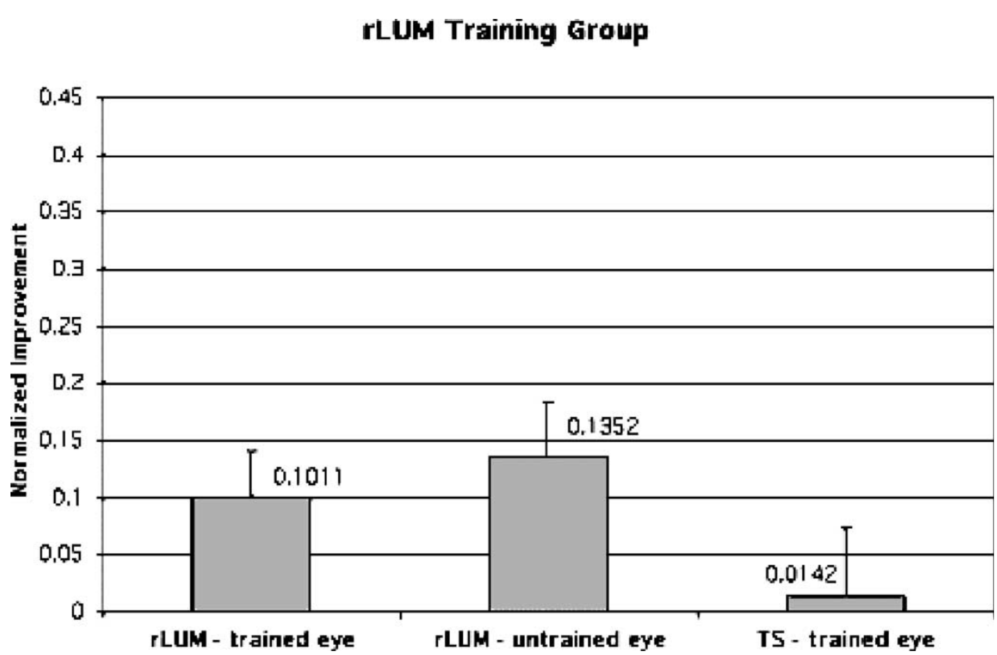

Fig. 5. Average percent improvement for the rLUM training group ($n = 4$). Improvement was measured as the difference in pre- and post-training thresholds divided by the sum of the pre- and post-training thresholds. Observers showed significant learning in the trained condition with the trained eye ($t = 2.566$, $p < 0.05$) and the untrained eye ($t = 2.788$, $p < 0.05$). No significant improvement was found for the untrained TS condition ($t = 0.238$, $p > 0.05$). Error bars represent mean standard error.

the rLUM task is much more difficult than the LUM task, because of the complexity associated with the random TS throughout the rLUM displays. And this greater complexity, the argument continues, somehow interferes with transfer to the TS task. However, comparison of the average pre-training LUM thresholds

(Fig. 2) with the average pre-training rLUM thresholds (Fig. 4) provides no support for this argument—both types of display yielded approximately equivalent pre-training thresholds (acknowledging, of course, that different observers participated in these two conditions).

All things considered, then, we are led to conclude that the absence of transfer from the rLUM condition to the TS condition in Experiment 2 stems from the absence of coherent TS in the rLUM display. Observers trained on the rLUM display were denied the incidental exposure to the information implicitly learned by people trained in the original LUM display of Experiment 1.

5. General discussion

The ability to discriminate shapes defined solely by TS improves with practice, with the magnitude of improvement being at least as great as that associated with comparable degrees of practice on shape discrimination based on luminance. This finding is not surprising, for people benefit from training on a host of visual tasks ranging from detection of spatial offsets between two lines (McKee & Westheimer, 1978) to object recognition of meaningful targets presented in noise (Gold, Bennett, & Sekuler, 1999). Indeed, it appears that the degree of improvement in performance with practice is related to the complexity of the task (Fine & Jacobs, 2002): greater learning is evidenced on more complex tasks. It is not immediately obvious where shape recognition based on TS falls on the "complexity" dimension, but it is worth noting that the slopes of the learning curves for the TS training in Experiment 1 are within the range of slopes produced by tasks characterized by Fine and Jacobs (2002) as "complex"—these were tasks in which external noise was utilized to mask detection or discrimination. While we did not explicitly manipulate noise in our displays, it is certainly the case that the TS sequences contained extraneous TS (random variations in contrast over time) that could constitute noise. Whether the degree of learning would be reduced with elimination of that potential source of noise remains to be determined. (Recall that random fluctuations in contrast were purposefully introduced to preclude possible luminance artifacts.)

For both shape cues—TS and LUM—learning transferred from the trained to the untrained eye, implying that the neural events underlying performance improvements on these two tasks transpire at a site in visual processing after information between the two eyes has been integrated. Again, this result is not unprecedented: interocular transfer has been found for other perceptual learning tasks, although exceptions do exist (e.g., Karni & Sagi, 1991).

Of particular significance for our purposes is the pattern of transfer between tasks. People trained on the shape discrimination based on TS showed no transfer of learning to the same task based on luminance-defined shapes. People trained on shape discrimination based on LUM, however, showed essentially complete transfer to the TS condition when the LUM display also contained coherent TS (i.e., synchronized changes in motion direction throughout the display) but essentially no transfer when the LUM display was devoid of coherent TS (i.e., unsynchronized changes in motion direction throughout the display). This pattern of learning transfer suggests that observers performing the TS task were not relying on some sort of luminance cue inadvertently created in the TS displays. If luminance cues were present and were creating shape information within the TS displays, then one might expect learning in the TS task to benefit performance on the LUM task; but this did not happen (Experiment 1). Nor did learning transfer from the LUM task to the TS task when the possibility of incidental learning of coherent TS was prevented (Experiment 2). These two results, together with our careful efforts to eliminate contrast and luminance artifacts in the TS displays themselves, suggest that luminance cues do not mediate shape detection in TS tasks.

Comparison of the LUM training conditions in Experiments 1 and 2 also indicates that incidental exposure to TS is sufficient to promote perceptual learning. This conclusion, too, is not without precedence. Watanabe et al. (2001) found that mere exposure to weak coherent motion—presented as the "background" for another task—was sufficient to increase observers' sensitivity for detection of coherent motion. At least on some tasks, in other words, visual perception can benefit from exposure to dynamic events even when those features were not the primary focus of attention during learning (but see Shiu & Pashler, 1992).

But what were our observers learning as they received extended exposure to these dynamic displays with synchronized changes in direction of motion? In other words, what experience-dependent changes occur within the central nervous system to support enhanced sensitivity to shape from TS? We can imagine at least two possible sources of this enhanced sensitivity. The first source is based on changes in the tuning properties of neurons mediating performance, changes that might be implemented by variations in the synaptic efficacy ("weighting") of connections among those neurons. Specifically, performance improvements may rely on experience-dependent increases in the temporal resolution of neural elements registering changes in direction of motion. Indeed, Lee and Blake (1999a, 2001) and Farid and Adelson (2001) have speculated that synchronized changes in motion direction of the sort used in our TS displays stimulate neurons selectively responsive to stimulus transients. Experience-dependent changes in the time constants of such neural elements could alter TS detection. However, improved fidelity of neuronal tuning may not be sufficient for improved TS detection, because the TS task additionally requires that the visual system extract the spatial distribution of those neural events with common (i.e., synchronized) TS. Accordingly, the second source of perceptual learning may rest on the improved ability to integrate synchronous activity across distributed neuronal populations that represent different regions of space. Experience-dependent enhancement in grouping efficiency should contribute to improved recognition of shape from TS.

Whatever the underlying bases of the learning effect we have documented, our findings further underscore the potential usefulness of TS in the optical input to vision, providing a robust source of information for spatial grouping. When we stop and think about it, such a role should not be surprising. After all, our eyes and brains evolved in a dynamic visual world in which objects move relative to one another and in which we ourselves are chronically moving our eyes and our heads. Why shouldn't vision include mechanisms to register this rich source of information, and why shouldn't we benefit from opportunities to exploit that information in order to perceive objects and events?

Acknowledgements

We thank Sang-Hun Lee for assistance with programming. This work was supported by a research grant (EY07760), a Training Grant (EY07135), and a Core Grant (EY01826) from the National Institutes of Health, and a research grant (BCS-0096178) from the National Science Foundation.

References

Adelson, E. H., & Farid, H. (1999). Filtering reveals form in temporally structured displays. *Science, 286*, 2231a.

Ahissar, M., & Hochstein, S. (1996). Learning pop-out detection: specificities to stimulus characteristics. *Vision Research, 36*, 3487–3500.

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

Ball, K., & Sekuler, R. (1982). A Specific and enduring improvement in visual motion discrimination. *Science, 218*, 697–698.

Banks, M. S., Aslin, R. N., & Letson, R. D. (1975). Sensitive period for the development of human binocular vision. *Science, 190*, 675–677.

Beard, B. L., Levi, D. M., & Reich, L. N. (1995). Perpetual learning in parafoveal vision. *Vision Research, 35*, 1679–1690.

Blake, R., & Yang, Y. (1997). Spatial and temporal coherence in perceptual binding. *Proceedings of the National Academy of Sciences: USA, 94*, 7115–7119.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 443–446.

Fahle, M. (1993). Figure–ground discrimination from temporal information. *Proceedings of the Royal Society of London. Series B, Biological Science, 254*, 199–203.

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

Fahle, M., Edelman, S., & Poggio, T. (1995). Fast perceptual learning in hyperacuity. *Vision Research, 35*, 3003–3013.

Fahle, M., & Koch, C. (1995). Spatial displacement, but not temporal asynchrony, destroys figural binding. *Vision Research, 35*, 491–494.

Farid, H., & Adelson, E. H. (2001). Synchrony does not promote grouping in temporally structured displays. *Nature Neuroscience, 4*, 875–876.

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

Fine, I., & Jacobs, R. A. (2002). Comparing perceptual learning across tasks: a review. *Journal of Vision, 2*, 190–203.

Fiorentini, A., & Berardi, N. (1981). Learning in grating waveform discrimination: specificity for orientation and spatial frequency. *Vision Research, 31*, 1149–1158.

Frisby, J. P., & Clatworthy, J. L. (1975). Learning to see complex random-dot stereograms. *Perception, 4*, 173–178.

Furmanski, C. S., & Engel, S. A. (2000). Perceptual learning in object recognition: object specificity and size invariance. *Vision Research, 40*, 473–484.

Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training "greeble" experts: A framework for studying expert object recognition processes. *Vision Research, 38*, 2401–2428.

Gibson, E. J. (1953). Improvement in perceptual judgments as a function of controlled practice or training. *Psychological Bulletin, 50*, 401–431.

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

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

Julesz, B. (1971). *Foundations of cyclopean perception*. Chicago, IL: University of Chicago Press.

Kandil, F. I., & Fahle, M. (2001). Purely temporal figure–ground segregation. *European Journal of Neuroscience, 13*, 2004–2008.

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

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

Kiper, D. C., Gegenfurtner, K. R., & Movshon, J. A. (1996). Cortical oscillatory responses do not affect visual segmentation. *Vision Research, 36*, 539–544.

Lee, S.-H., & Blake, R. (1999a). Visual form created solely from temporal structure. *Science, 284*, 1165–1168.

Lee, S.-H., & Blake, R. (1999b). Response to: Adelson and Farid. *Science, 286*, 2231.

Lee, S.-H., & Blake, R. (2001). Neural synergy in visual grouping: when good continuation meets common fate. *Vision Research, 41*, 2057–2064.

Leonards, U., Singer, W., & Fahle, M. (1996). The influence of temporal phase differences on texture segmentation. *Vision Research, 36*, 2689–2697.

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

McKee, S. P., & Westheimer, G. (1978). Improvements in vernier acuity with practice. *Perception & Psychophysics, 24*, 258–262.

Mednick, S. C., Nakayama, K., Cantero, J. L., Atienza, M., Levin, A., Pathak, A., & Stickgold, R. (2002). The restorative effect of naps on perceptual learning. *Nature Neuroscience, 5*, 677–681.

Mollon, J. D., & Danoliva, M. V. (1996). Three remarks on perceptual learning. *Spatial Vision, 10*, 51–58.

Morgan, M., & Castet, E. (2002). High temporal frequency synchrony is insufficient for perceptual grouping. *Proceedings of the Royal Society of London. Series B, Biological Science, 269*, 513–516.

Ramachandran, V. S., & Rogers-Ramachandran, D. C. (1991). Phantom contours: A new class of visual patterns that selectively activates the magnocellular pathway in man. *Bulletin of the Psychonomic Society, 29*, 391–394.

Ross, J., & Hogben, J. H. (1974). Short-term memory in stereopsis. *Vision Research, 14*, 1195–1201.

Sagi, D., & Tanne, D. (1994). Perceptual learning: learning to see. *Current Opinion in Neurobiology, 4*, 195–199.

Shiu, L.-P., & Pashler, H. (1992). Improvement in line orientation discrimination is retinally local but dependant on cognitive set. *Perception & Psychophysics, 52*, 582–588.

Sinha, P., & Poggio, T. (1996). Role of learning in three-dimensional form perception. *Nature, 384*, 460–463.

Sireteanu, R., & Rettenbach, R. (2000). Perceptual learning in visual search generalizes over task, locations, and eyes. *Vision Research, 40*, 2925–2949.

Stickgold, R. (1998). Sleep: off-line memory reprocessing. *Trends in Cognitive Science, 2*, 484–492.

Usher, M., & Donnelly, N. (1998). Visual synchrony affects binding and segmentation in perception. *Nature, 394*, 179–182.

Watanabe, T., Nañez, J. E., & Sasaki, Y. (2001). Perceptual learning without perception. *Nature, 413*, 844–848.

Wehrhahn, C., & Rapf, D. (2001). Perceptual learning of apparent motion mediated through ON- and OFF-pathways in human vision. *Vision Research, 41*, 353–358.

Westheimer, G., & McKee, S. P. (1977). Spatial configurations for visual hyperacuity. *Vision Research, 17*, 941–947.

Zanker, J. M. (1999). Perceptual learning in primary and secondary motion vision. *Vision Research, 39*, 1293–1304.