View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by CiteSeerX

Ideal observer analysis of crowding and the reduction of crowding through learning

Gerald J. Sun

Department of Biological Sciences, University of Southern California, Los Angeles, CA, USA, &
The Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Susana T. L. Chung

School of Optometry, University of California, Berkeley, CA, USA

Bosco S. Tjan

Department of Psychology and the Neuroscience Graduate Program, University of Southern California, Los Angeles, CA, USA

Crowding is a prominent phenomenon in peripheral vision where nearby objects impede one's ability to identify a target of interest. The precise mechanism of crowding is not known. We used ideal observer analysis and a noise-masking paradigm to identify the functional mechanism of crowding. We tested letter identification in the periphery with and without flanking letters and found that crowding increases equivalent input noise and decreases sampling efficiency. Crowding effectively causes the signal from the target to be noisier and at the same time reduces the visual system's ability to make use of a noisy signal. After practicing identification of flanked letters without noise in the periphery for 6 days, subjects' performance for identifying flanked letters improved (reduction of crowding). Across subjects, the improvement was attributable to either a decrease in crowding-induced equivalent input noise or an increase in sampling efficiency, but seldom both. This pattern of results is consistent with a simple model whereby learning reduces crowding by adjusting the spatial extent of a perceptual window used to gather relevant input features. Following learning, subjects with inappropriately large windows reduced their window sizes; while subjects with inappropriately small windows increased their window sizes. The improvement in equivalent input noise and sampling efficiency persists for at least 6 months.

Keywords: peripheral vision, crowding, perceptual learning, ideal observer analysis

Citation: Sun, G. J., Chung, S. T. L., & Tjan, B. S. (2010). Ideal observer analysis of crowding and the reduction of crowding through learning. *Journal of Vision, 10*(5):16, 1–14, http://journalofvision.org/content/10/5/16, doi:10.1167/10.5.16.

Introduction

Human peripheral vision is limited by the visual system's inability to properly integrate features and segment scenes. This inability is exemplified by a phenomenon called crowding, where nearby items adversely impede one's ability to identify a target (e.g., Bouma, 1970; Flom, Weymouth, & Kahneman, 1963; Townsend, Taylor, & Brown, 1971). Despite the considerable amount of data on crowding, the mechanism of crowding remains undetermined (Levi, 2008; Pelli & Tillman, 2008). Crowding cannot be explained by the lower spatial resolution in peripheral vision. For example, crowding is present even when the spacing between the target and flanking letters is larger than the size of a letter, and the letters are above acuity (Pelli, Palomares, & Majaj, 2004). While the psychophysical properties of crowding share some similarities with overlap masking, they also differ qualitatively from overlap masking in several important ways (Chung, Levi, & Legge, 2001; Levi, Hariharan, & Klein, 2002; Pelli et al., 2004).

To date, one account for crowding is that the attention mechanism for peripheral vision lacks sufficient spatial resolution to discern a target from the surrounding clutter (He, Cavanagh, & Intriligator, 1996; Intriligator & Cavanagh, 2001; Leat, Li, & Epp, 1999; Strasburger, Harvey, & Rentschler, 1991; Tripathy & Cavanagh, 2002). A competing account attributes crowding to inappropriate feature integration at the lower level visual processing stages (Levi et al., 2002; Pelli et al., 2004). More recent work (Freeman & Pelli, 2007; Nandy & Tjan, 2007), nevertheless, does not see these as competing and mutually exclusive hypotheses—the end result is that wrong features and features from wrong locations are being integrated, resulting in a non-veridical percept. This feature integration error seems to be spatial (Nandy & Tjan, 2007) and apparently not caused by any inefficiency in the integration across spatial frequencies in the periphery (Nandy & Tjan, 2008).

doi: 10.1167/10.5.16 Received August 28, 2009; published May 11, 2010 ISSN 1534-7362 © ARVO

Downloaded from jov.arvojournals.org on 06/28/2019

Can we eliminate or reduce crowding? Chung (2007) showed that, following practice, accuracy of letter identification improves and the spatial extent of crowding is significantly reduced. Similar to the crowding effect, the mechanism underlying the reduction in crowding after learning is also not known. Understanding the mechanism of this improvement can be an important step toward understanding crowding.

The goals of the present study are to psychophysically determine the mechanism of crowding and that of the reduction of crowding through perceptual learning. We do so in the context of a simple observer model that attributes the limitation in visual performance to two sources: (1) the presence of noise or random spurious features that limit the precision of sensory measurements and (2) the reduction of the visual system's ability to make full use of the information available in the stimulus (Pelli, 1981; Pelli & Farell, 1999). We can quantify the former in terms of equivalent input noise and the latter in terms of sampling efficiency (Chung, Levi, & Tjan, 2005; Conrey & Gold, 2006; Gold, Bennett, & Sekuler, 1999; Legge, Kersten, & Burgess, 1987; Pelli & Farell, 1999; Tjan, Braje, Legge, & Kersten, 1995). This approach of observer modeling, sometimes referred to as the linear amplifier model, is a special case of the more elaborate observer models that include transducer non-linearity, signal-dependent noise, and signal uncertainty (Burgess & Colborne, 1988; Eckstein, Ahumada, & Watson, 1997; Lu & Dosher, 1999; Pelli, 1985; see also Lu & Dosher, 2008 for a review). Since both the linear amplifier model and the more elaborate perceptual template model of Lu and Dosher give qualitatively similar results for perceptual learning tasks in the periphery (Chung et al., 2005; Lu, Chu, Dosher, & Lee, 2005), we use the simpler of the two in the current study.

To preview, we found that crowding leads to an elevated equivalent input noise and a reduction in sampling efficiency both before and after perceptual training. Training reduces the effects of crowding but does not eliminate them. Following training, a significant reduction in crowding-induced equivalent input noise was observed for subjects with a high crowding-induced equivalent input noise prior to training, while a significant improvement in efficiency was found for individuals with a large crowding-induced deficit in efficiency. The increase in performance (reduction in crowding) due to learning was retained for as long as 6 months. The pattern of results suggests that subjects learn to optimize the size of a perceptual window for gathering input features.

Theory

The optimal strategy to maximize accuracy in an identification task is to select a response that is the most probable given the input (i.e., the one that maximizes the posterior probability). Following Pelli (1981), we added two limiting factors to such an ideal observer in order to model a human observer: (1) an additive white noise at the input with a constant power spectral density of N_{eq}, and (2) a "device" that reduces the available signal-to-noise ratio by effectively down-sampling the noisy input by a factor of η before giving it to the ideal observer for identification (Tjan et al., 1995). These two limiting factors, equivalent input noise (N_{eq}) and sampling efficiency (η), are macroscopic descriptors of a visual system since they encompass many possible mechanistic realizations.

The most common interpretation for N_{eq} is to see it as an aggregated quantity of the stochastic noises internal to the visual system that are independent of the target signal for a given task. Noise may exist at different stages of visual processing and it can also be caused by other stimuli in a visual scene. N_{eq} represents a fundamental limitation in the precision of measurements at different levels of abstraction.

A theoretical interpretation of η is that it represents the proportion of relevant information in the form of independent statistical samples that the visual system is able to use when making a perceptual decision. The mechanistic instantiation of η can be either deterministic or stochastic. Deterministic causes for $\eta < 1$ include computational steps that do not use an accurate or complete specification of the signals to be identified (imprecise template) or consider a variety of possible signals that actually do not appear in the task (invariance, uncertainty). Stochastic causes are forms of additive internal noise with a power spectral density proportional to the signal energy of the input. This type of noise is often called a "multiplicative" noise.

The sampling efficiency and equivalent input noise of a system can be determined using an external noise method (Pelli, 1981; Pelli & Farell, 1999). For our two-factor ideal observer model, which has a linear front end, the required contrast energy (E) of the target to reach a given accuracy criterion is linearly related to the power spectral density (N) of the external noise with a non-positive intercept at $E = 0$:

$$E = mN + E_0, \quad m > 0, \ E_0 \geq 0. \tag{1}$$

The contrast energy of the target is the square of its root mean square (rms) contrast multiplied by the image area. The power spectral density of the noise is the variance divided by the noise bandwidth, which is determined by the size of the noise pixels (Appendix B). The slope of the E vs. N (EvN) function is inversely proportional to sampling efficiency (η) and the negative of its horizontal intercept represents the amount of the

Downloaded from jov.arvojournals.org on 06/28/2019

equivalent input noise (N_{eq}) (Legge et al., 1987; Tjan et al., 1995). Specifically,

$$\eta = m_{\mathrm{ideal}}/m, \qquad N_{\mathrm{eq}} = E_0/m, \tag{2}$$

where m_{ideal} is the EvN slope of the ideal observer. This property of the ideal observer model presents a simple method for estimating η and N_{eq}: a target is masked with white noise, and the contrast threshold for identifying the target at a given accuracy criterion is measured at various levels of the masking noise. The absolute value of N_{eq} and a relative value of η are obtained by fitting a straight line to the EvN data set. The value of η is the ratio of the EvN slope of the true ideal observer[1] to that of the modeled human observer (Equation 2). We do not need to explicitly compute the EvN slope of the ideal observer to assess the effects of crowding because the ideal observer is not affected by crowding (Appendix C); nevertheless, we provided the ideal observer slope in Appendix C for future reference.

The main goals of the current study are to characterize crowding in terms of sampling efficiency and equivalent input noise and to study the effect of practice on these quantities. For this purpose, the external noise used to estimate η and N_{eq} masked only the target and not the flankers that closely flanked the target. Moreover, the flankers were presented at a fixed contrast, independent of target contrast. If a component of crowding is that erroneous spatial pooling causes random features from the flanker positions to be neurally superimposed on the target, then the presence of the flankers will behave like an additional source of noise, leading to an increase in N_{eq}. If flanker features interact with target features more selectively (e.g., a horizontal flanker feature suppresses the detection of a vertical target feature) or preferentially (e.g., a more reliably detected flanker feature is mistaken as a target feature), or if the visual system attempts to minimize crowding by being more stringent in its selection of target features, the equivalent number of target features utilized will be reduced when flankers are present, leading to a reduction in η. Changes in η and N_{eq} due to the presence of flankers as compared to the target-alone condition thus reveal different functional components of crowding.

It should be noted that we are interpreting the mechanism of crowding with respect to the additive-noise ideal observer model. η and N_{eq}, being macroscopic descriptors, do not uniquely correspond to a specific neural mechanism. The general notion of erroneous feature integration, for example, is not a precisely defined mechanism. An indiscriminate integration of flanker features at a later processing stage could yield a large N_{eq}, while a bias toward using features that are more reliably detected, whether appropriately from the target or inappropriately from the flankers, would lead to a decrease in η. As always, the most reasonable mechanistic interpretation depends on the overall pattern of the empirical findings and parsimony of the interpretation.

In this study, we are less interested in the actual values of η and N_{eq}, although both are available (Appendices A and C). Instead, we are more interested in how a subject's efficiency and equivalent noise are affected in the presence of crowding. Therefore, we will express a subject's efficiency and equivalent input noise in a target-flanked condition relative to those in a target-only (unflanked) condition. Specifically, we define the *efficiency ratio* (η_{r}) as the ratio of the sampling efficiency between the flanked and unflanked conditions:

$$\eta_{\mathrm{r}} = \frac{\eta_{\mathrm{flanked}}}{\eta_{\mathrm{unflanked}}} = \frac{m_{\mathrm{unflanked}}}{m_{\mathrm{flanked}}}, \tag{3}$$

where m is the slope of EvN line of Equation 1. We also define the *equivalent noise difference* (ΔN_{eq}) as the difference between the flanked and unflanked conditions:

$$\Delta N_{\mathrm{eq}} = N_{\mathrm{eq,flanked}} - N_{\mathrm{eq,unflanked}} = \frac{E_{0,\mathrm{flanked}}}{m_{\mathrm{flanked}}} - \frac{E_{0,\mathrm{unflanked}}}{m_{\mathrm{unflanked}}}. \tag{4}$$

An intuitive interpretation of η_{r} is the fraction of the quantity of target features used by the visual system in crowding relative to those used without crowding. Likewise, ΔN_{eq} can be thought of as the amount of the random flanker features, in units of noise, which are masking or mistaken for target features. Any changes in η_{r} and ΔN_{eq} after practice will inform us of the mechanistic nature of the reduction in crowding following learning.

Methods

Procedure

The main experiment comprised eight sessions, with one session per day. All subjects completed the eight sessions within 10 days. The first ("pre-test") and last ("post-test") sessions included eight experimental conditions: two flanking conditions (unflanked and flanked) crossed with four external noise levels. The intervening six ("training") sessions followed the procedures of Chung (2007), where subjects were trained to identify closely flanked letters in the absence of external noise. A follow-up test was performed 1 to 6 months after the post-test to assess any retention of learning.

During both testing and training, subjects identified letters presented at 10° in their lower right visual quadrant, midway between the horizontal and vertical

Subject	Letter size in x-height
AL	1.59°
BW	1.17°
CT	1.70°
LM	1.07°
MB	1.26°
SL	1.92°

Table 1. Stimulus size in x-height used for the six subjects.

meridians with a letter size (Table 1) corresponding to 2.5× the subject's single-letter acuity. Letter acuity at the test location was measured before commencing the experiment: subjects identified unflanked letters, randomly drawn from 26 lowercase letters, while the size of the presented letter was varied using the QUEST procedure (Watson & Pelli, 1983) to yield a threshold letter size corresponding to 79% identification accuracy. This threshold letter size was taken to represent the subject's letter acuity at the test location.

Pre- and post-tests consisted of eight experimental conditions, each tested five times in separate blocks, with 60 trials per block. The blocks and conditions were randomized with the constraint that block number (k + 1) of any condition was tested only after all the conditions had been tested at least k times.

For each experimental condition, a QUEST procedure, as implemented in the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997), was used to estimate the threshold contrast of the target letter that corresponded to an identification accuracy of 50%. The QUEST procedure was initialized identically for the first block of each experimental condition. For the other four blocks of repeated measurements of the same condition, QUEST was initialized to the threshold contrast estimated from the previous block; this procedure improves stability without prematurely committing to a threshold.

The follow-up test had the exact same design as the pre- and post-tests, except that only the noise levels, but not the flanking conditions, were blocked. Specifically, for a given noise level, the unflanked and flanked trials were first randomly mixed and repartitioned into 60-trial blocks. This was designed to determine whether the observed post-training improvement in the flanked condition was attained by learning a strategy specific for the flanked trials that were blocked.

Training sessions were structured after those in Chung (2007), which have proven to be effective. A training session consisted of 1000 trials divided into 10 blocks of 100 trials. Each subject completed six training sessions, scheduled over 6 days, for a total of 6000 trials. All subjects completed the pre-test, training, and post-test within 10 days.

Stimuli

The stimulus for each trial consisted of either a single letter for the unflanked condition or a letter flanked by two

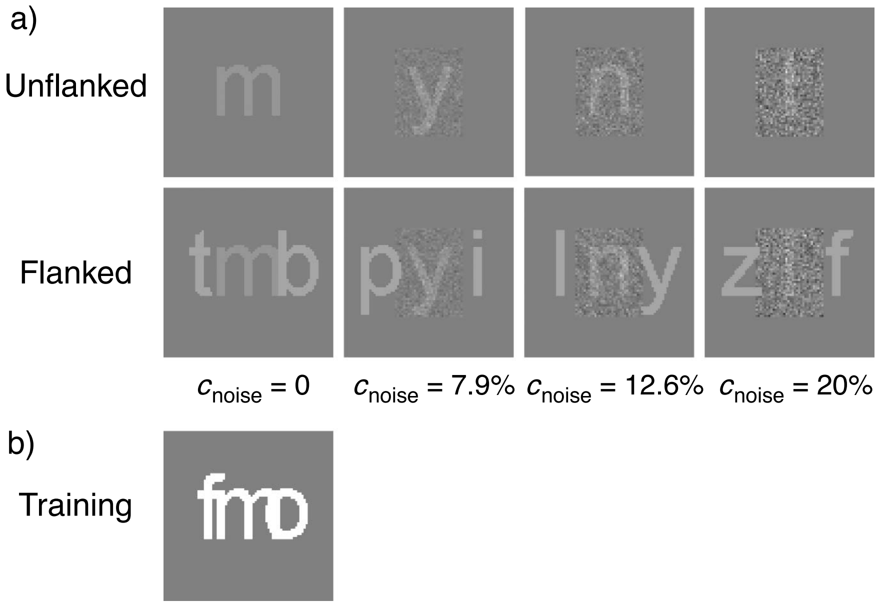


Figure 1. Stimulus examples. (a) Pre- and post-test stimuli in the unflanked and flanked conditions. The flankers, when presented, had a fixed Weber contrast of 33% and were placed from the target at a center-to-center distance of 1.0× x-height. The masking noise was confined to within the largest bounding box of a letter and centered at the target letter. Four noise levels, quantified here in units of rms Weber contrast, were used in the experiment. All target letters are shown at the same Weber contrast across the four noise levels for illustration. (b) An example of the training stimulus. The target letter and the flankers were presented at full contrast and the center-to-center distance between target and flanker was 0.8× x-height.

Downloaded from jov.arvojournals.org on 06/28/2019

other letters for the flanked condition. All letters were brighter than the mid-gray background. Target and flanking letters were randomly chosen from the set of 26 lowercase Roman letters from the English alphabet in Arial font (Mac OS 9), disregarding the proportional letter spacing associated with the font. Flanking letters, when present, were presented at either a center-to-center separation of 1.0× x-height (pre- and post-tests) or 0.8× x-height (training). The pixels of the flanking letters replaced those of a target letter in the regions where they overlapped; such overlaps occurred between letters of wider width ("m", "w") and were relatively rare. The Weber contrast of the flanking letters was fixed at 0.33 during the pre- and post-tests and 1.0 during training. The contrast of the target letter in the pre- and post-tests was adjusted with QUEST as described.

During the pre- and post-tests, a Gaussian (spectrally white) luminance noise field, equal to the largest bounding box of all target letters, was added to the target location. In order to produce an adequate level of noise spectral density, each stimulus pixel for both the letters and noise comprised 4 × 4 actual pixels on the CRT. The rms contrasts of the noise were 0, 0.079, 0.126, and 0.2. With a stimulus pixel size of 0.0777° at a viewing distance of 100 cm, the corresponding noise spectral densities were 0, 37.7, 95.8, and 241 $\times$ 10^{-6} deg^2. The mean luminance of the noise fields and the background luminance of the display were approximately 20 cd/m^2. Figure 1 depicts a sample of the stimuli in all eight experimental conditions of the pre- and post-tests, along with the condition used for training.

Stimuli were presented at the center of a calibrated Sony CRT screen. The calibrated CRT had a corrected gamma of 1.0 with 11 bits (2048 levels) of linearly spaced luminance levels, achieved with a passive video attenuator (Pelli & Zhang, 1991) and custom-built contrast calibration and control software implemented in MATLAB and ran on a Mac G4 (OS 9.2.2). Only the green channel of the monitor was used during the experiment.

The fixation mark was a green LED mounted at 10° to the upper left of the center of the target letter. For each trial, the target and flankers (if applicable) were presented simultaneously for 250 ms. Audio feedback followed the subject's response (a tone for correct trials or an announcement of the target letter for incorrect trials). The detailed timing of a trial is shown in Figure 2.

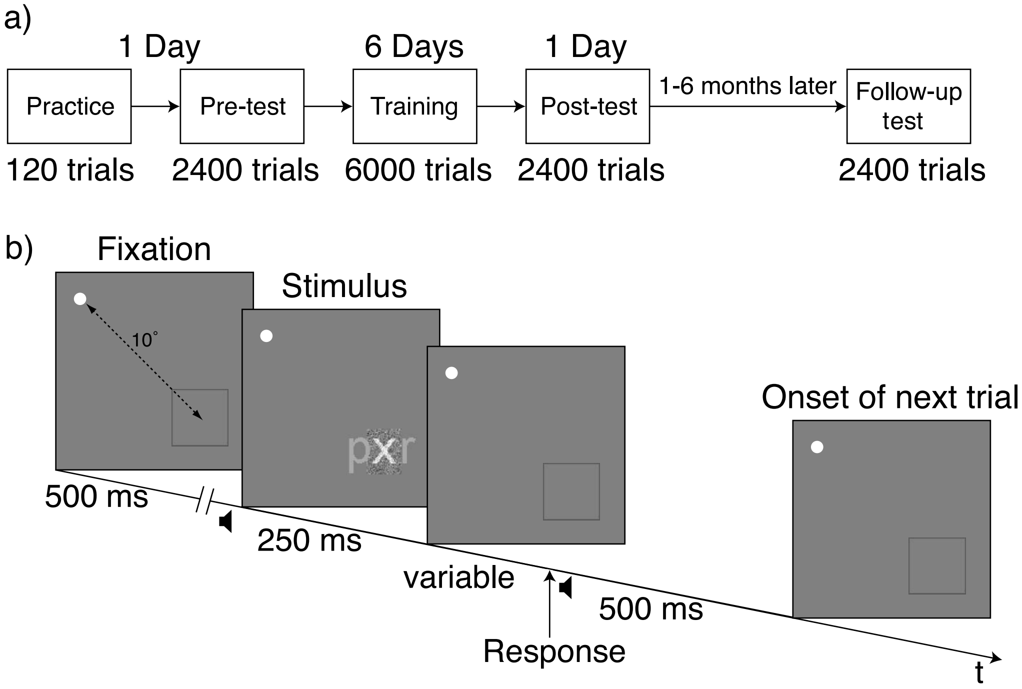
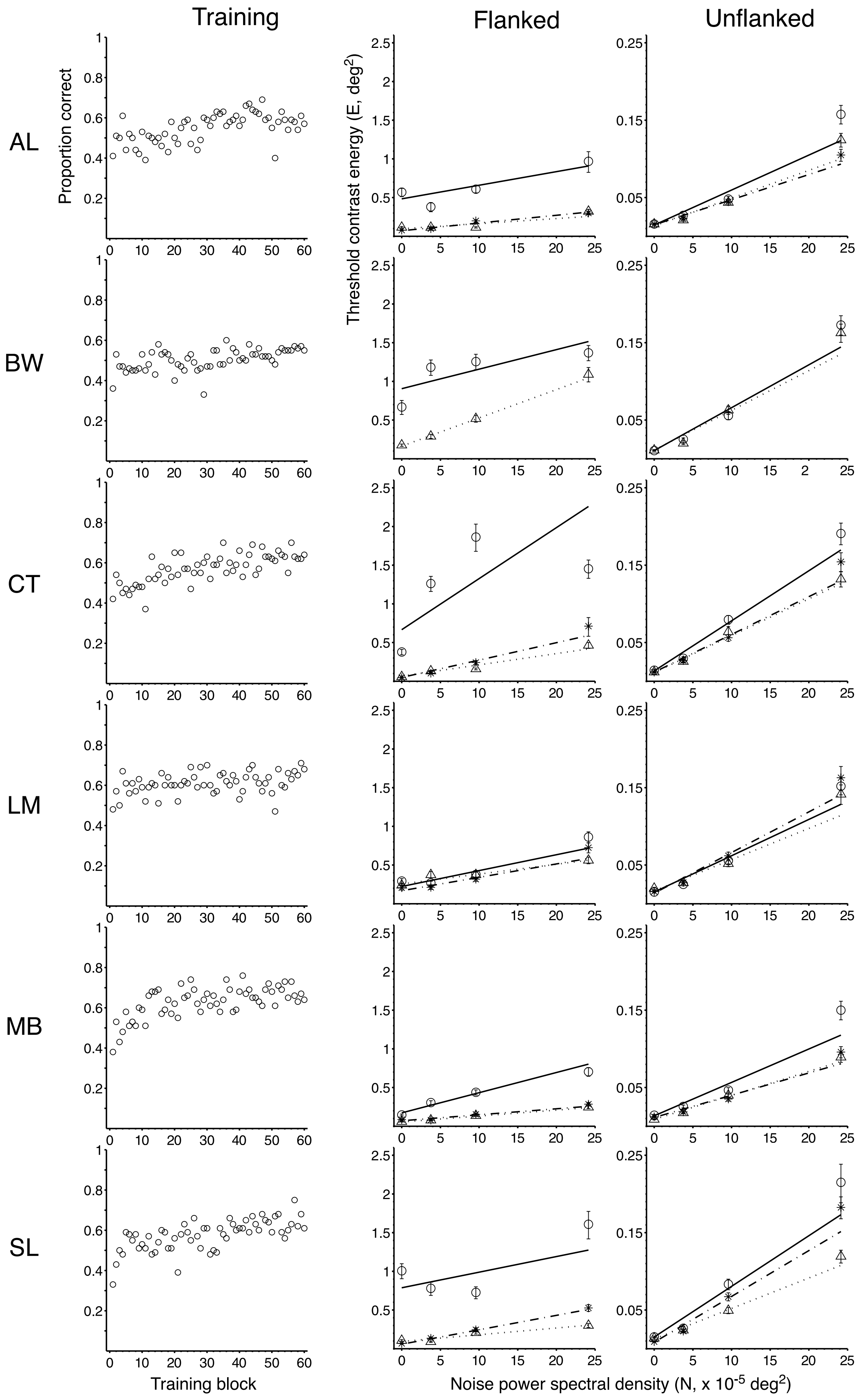


Figure 2. Experimental design. (a) Subjects participated in a brief practice session before commencing the pre-test. Over the following 6 days, subjects underwent training and were then tested for changes during a subsequent post-test. One to six months following the post-test, subjects participated in an additional session, which followed the same experimental routine as the pre- and post-tests except that the trials of unflanked and flanked letters were randomly interleaved for each noise level. (b) Stimulus timing in a given trial for the pre-, post-, and follow-up tests. A square target box marked by low-contrast dark lines and a size equal to five times the subject's letter acuity in x-height was shown on the screen in the absence of stimuli to indicate the expected target position in peripheral vision. After 500 ms, the stimulus was presented for 250 ms along with a brief tone; upon giving a response, a subject was provided with auditory feedback. The next trial followed after a 500-ms delay.

Downloaded from jov.arvojournals.org on 06/28/2019


Training
Flanked
Unflanked
AL
BW
CT
LM
MB
SL
Proportion correct
0.2
0.4
0.6
0.8
1
0
10
20
30
40
50
60
Training block
Threshold contrast energy (E, deg^2)
0.5
1
1.5
2
2.5
0.05
0.15
0.25
0
5
10
15
20
25
2
Noise power spectral density (N, x 10^{-5} deg^2)

Data analysis

The threshold contrast energy (E) in units of deg^2 is linearly related to the square of the measured threshold Weber contrast. The proportional constant (Appendix B) is the contrast energy, averaged over the 26 stimulus letters, when each letter is rendered with pixel luminance twice that of the background (a Weber contrast of 1.0).

The linear ideal observer model (Equation 1) was fitted to the empirical data of threshold contrast energy (E) vs. power spectral density of the masking noise (N) by minimizing the squared residuals defined in log(E) and scaled by the empirically determined standard error of log(E). This is because the measurement error of E generally increases with E, with a variance proportional to E^2. Bootstrapping was used to estimate the median and the 95% confidence intervals for quantities of interest.

Participants

Six subjects from the University of Southern California with normal or corrected-to-normal vision and naive to the purpose of the experiment participated with written informed consent and completed the main experiment. Five of the six subjects returned for the follow-up test.

Results

Figure 3 shows the results from the pre-, post-, and follow-up tests as well as the block-by-block accuracy for the training sessions. Crowding was substantial. Across the four noise levels, the threshold contrast required for identifying letters at an accuracy criterion of 50% was substantially higher in the flanked condition than in the unflanked condition for pre-, post-, and follow-up tests (pre-test: $F(1,5) = 211.7$, $p = 0.00003$; post-test: $F(1,5) = 86.4$, $p = 0.0002$; follow-up: $F(1,4) = 259.1$, $p = 0.00009$). The key numerical values extracted from Figure 3 are given in Table A1 in Appendix A. Training was effective in improving the accuracy of identifying letters in a flanked condition at full contrast without noise. The average accuracy during the training sessions improved from 50% in Day 1 to 61% in Day 6 (Table A2, $F(1,5) = 37.6$, $p = 0.002$), replicating a finding in Chung (2007).

Figure 3. Training performance (first column) and threshold contrast energy (E) versus noise power spectral density (N) for the flanked (second column) and unflanked (third column) conditions for each subject. Note that the ordinates for E differ by a factor of 10 between the flanked and unflanked conditions. Circles and triangles represent the threshold contrast energy measured during the pre- and post-tests, respectively; asterisks represent the threshold contrast energy during the follow-up test. Solid and dotted lines represent fits of Equation 1 to pre- and post-test data, respectively; dash-dotted lines represent fits to follow-up test data. Error bars are ±*SE*. Note that subject BW did not participate in the follow-up test.

Crowding

To reveal the mechanisms of crowding and the reduction of crowding due to perceptual training, we expressed the effects of crowding in terms of efficiency ratio and equivalent noise difference, as defined earlier. Figure 4 compares the efficiency ratio and equivalent noise difference measured before and after training. Both before and after training, the presence of flankers significantly reduced sampling efficiency and increased equivalent input noise for all subjects. When the data points of Figure 4 are projected to either the abscissa or the ordinate, the efficiency ratio never approached 1.0 (mean η_r was 0.22 before training and 0.36 after training), and the equivalent noise difference was never close to zero (mean ΔN_{eq} was 211×10^{-6} deg^2 before training and 133×10^{-6} deg^2 after training; in comparison, the strongest external noise used in the experiments was 241×10^{-6} deg^2 in power spectral density). The reduction in efficiency and increase in equivalent input noise due to crowding corresponds to the findings of Nandy and Tjan (2007) in that fewer appropriate features and more inappropriate features are being utilized in crowding, respectively.

Effect of practice on crowding

Across the group of six subjects, perceptual training over 6 days and 6000 trials appeared to yield only a marginal increase in the efficiency ratio (a mean of 0.22 pre-test vs. 0.36 post-test, $F(1,5) = 6.46$, $p = 0.052$) and no significant reduction in the equivalent noise difference (a mean of 211×10^{-6} deg^2 pre-test vs. 133×10^{-6} deg^2 post-test, $F(1,5) = 1.36$, $p = 0.30$). Such group analyses are misleading, however, because each individual subject did have a significant reduction in crowding as shown in Figures 3 and 4. Table 2 summarizes the improvements in the efficiency ratio and equivalent noise difference for individual subjects. With the exception of subjects AL and CT, who improved in both efficiency and equivalent input noise, a majority of the subjects (four out of six) improved in only one of the two quantities. Furthermore, if we perform a median split on the data, we find that all three subjects with equivalent noise differences above the median improved by reducing their equivalent input noise, while all three subjects with efficiency ratios below the median improved by increasing their efficiency. The data suggest that these two forms of improvements can be mutually exclusive. We postulate that the practice-induced

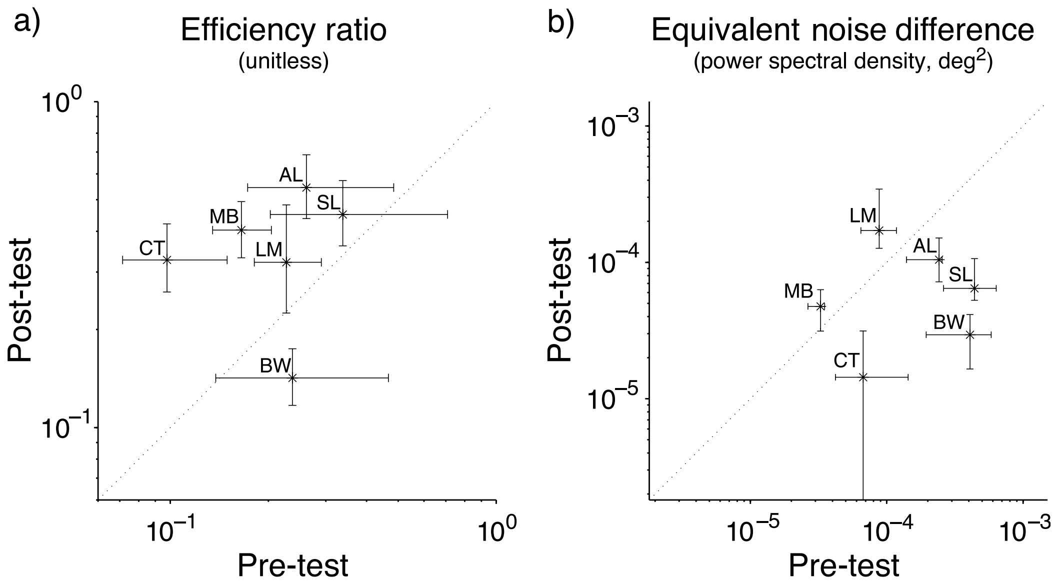


Figure 4. Effects of training on crowding. (a) Median efficiency ratio (η_r) in the post-test versus that in the pre-test. (b) Median equivalent noise difference (ΔN_{eq}) in the post-test versus that in the pre-test. Error bars are the bootstrapped 95% confidence intervals.

improvement may affect only a single perceptual factor. We shall return to this point in the Discussion section.

The effects of perceptual training on crowding, quantified above, are assessed relative to the respective unflanked condition. Consequently, our findings are less dependent on general task learning. To measure the amount of general task learning, we compare efficiency and equivalent input noise of the unflanked condition before and after the 6 days of training with flanked letters. Specifically, we calculated the efficiency ratio and equivalent noise difference of the unflanked condition after training with respect to the unflanked condition before training. Figure 5 shows that training with flanked letters improved sampling efficiency in the unflanked condition without any consistent effect on equivalent input noise. For the unflanked condition, this finding of improved efficiency, and not equivalent input noise, reproduces the pattern of results obtained with similar perceptual training tasks (Chung et al., 2005; Gold et al., 1999). This is, however, different from the effect of training on crowding, which improves efficiency or equivalent input noise in a subject-dependent manner.

Retention of learning

Figure 6 shows the results of the follow-up test as compared to pre-test. The improvements seen in the post-test are generally retained after 1–6 months. Chung, Legge, and Cheung (2004) reported a similar retention effect when they trained subjects to identify letter triplets in the periphery. Unlike during pre- and post-tests, we randomly interleaved the flanked and unflanked conditions during the follow-up test. The similarity in subjects' performance between the post- and follow-up tests shows that it is unlikely that subjects learned to use different perceptual strategies for the two conditions.

Discussion

Irrespective of any improvements due to learning, crowding reduces sampling efficiency and elevates equivalent input noise in peripheral vision. This is most evident in Figure 4 by projecting the data points to the abscissa and the ordinate. As compared to the unflanked condition, the efficiency ratio for identifying a flanked letter was significantly less than 1.0 and the equivalent noise difference was significantly greater than 0 for every subject before and after training. This agrees with the finding of Nandy and Tjan (2007) that crowding is caused by a reduction

	Improvement in	
Subject	η_r	ΔN_{eq}
AL	✓	✓
BW	×	✓
CT	✓	✓
LM	✓	×
MB	✓	×
SL	×	✓

Table 2. Training-induced improvements for each subject. A significant (per confidence intervals of Figure 4) increase in the efficiency ratio (η_r) or a significant decrease in the equivalent noise difference (ΔN_{eq}) is considered an improvement.

Downloaded from jov.arvojournals.org on 06/28/2019

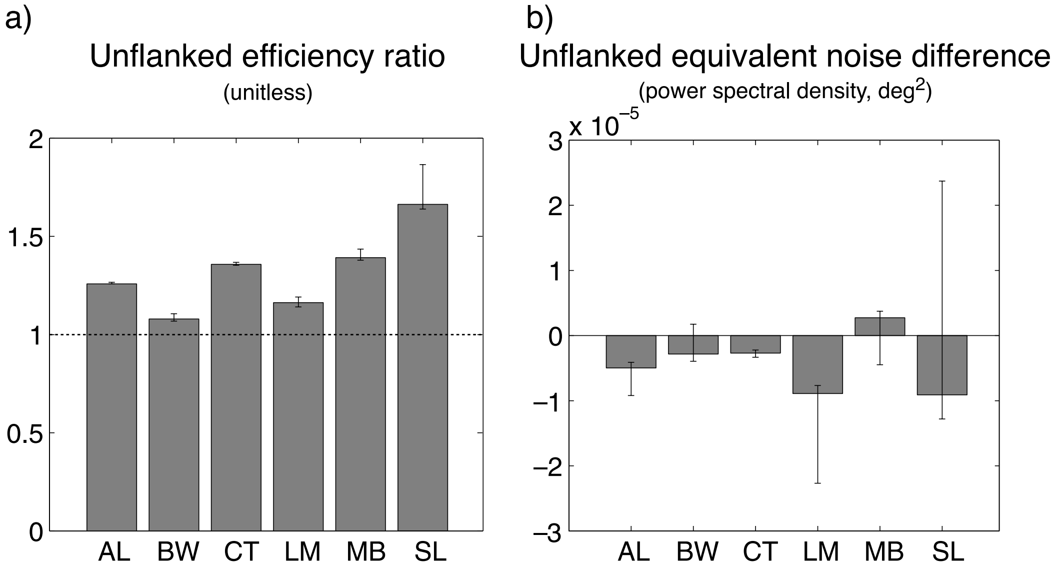
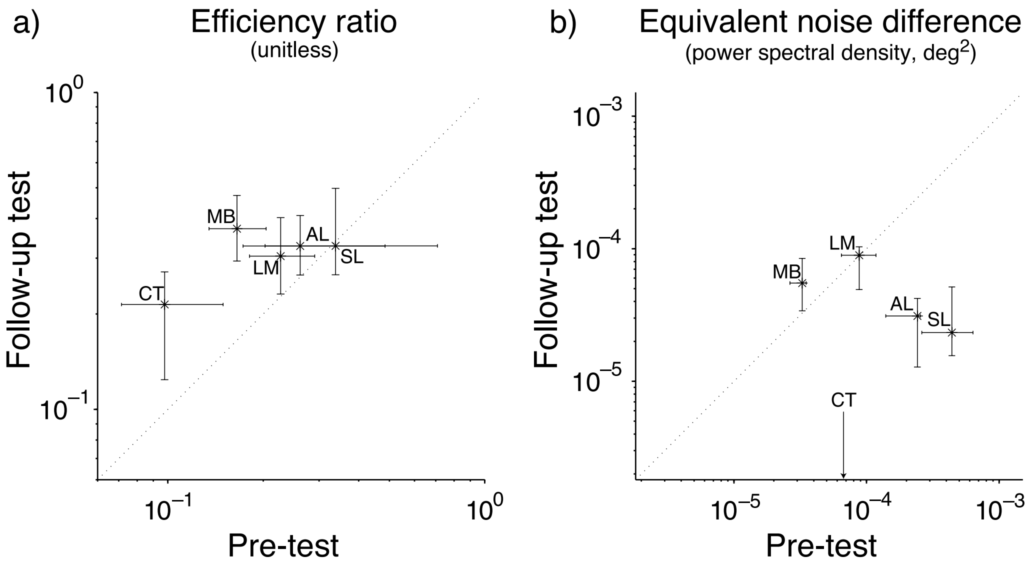


Figure 5. Effects of training on the unflanked condition. (a) Median unflanked efficiency ratio. (b) Median unflanked equivalent noise difference. Error bars are the bootstrapped 95% confidence intervals.

in the utilization of valid features and an increase in the utilization of invalid features. Nandy and Tjan reached their conclusion by comparing classification images between human and ideal observers, whereas our current finding was based on measuring contrast thresholds while masking the peripheral target with various levels of white noise.

The external noise used in the current study masked only the target, sparing the flankers. The utilization of any invalid features, either from the flanker locations due to



c)

Subject	AL	CT	LM	MB	SL
Days after post-test	33	36	178	41	97

Figure 6. Retention of learning. (a) Median efficiency ratio (η_r) in the follow-up test versus that in the pre-test. (b) Median equivalent noise difference (ΔN_{eq}) in the follow-up test versus that in the pre-test. Error bars are the bootstrapped 95% confidence interval for each value. (c) Number of days after post-test that the follow-up test was performed. Note that subject BW did not participate in the follow-up test, and CT, who had the lowest equivalent noise difference in the post-test, did not have a positive median value of equivalent noise difference for the follow-up test ($N_{eq,unflanked} > N_{eq,flanked}$).

Downloaded from jov.arvojournals.org on 06/28/2019

positional uncertainty or otherwise induced by the flankers, is analogous to adding another masking noise, thus revealing itself as an increase in equivalent input noise. A reduction in the use of valid features is equivalent to the reduction of target contrast energy in proportion to the power of the masking noise—hence a reduction in sampling efficiency. Our finding that crowding is caused by using fewer valid features and more invalid features is a modest elaboration of the commonly assumed faulty integration model of crowding (Levi, 2008; Levi et al., 2002; Pelli et al., 2004; Pelli & Tillman, 2008).

Learning reduces crowding. Chung (2007) showed that subjects improved in their accuracy following 6000 trials of identifying crowded letters at an eccentricity of 10° without any masking noise. We replicated this finding and showed that practice with noiseless stimuli transfers to conditions when the target was masked by luminance noise. This is consistent with the finding of Dosher and Lu (2005) that learning in the no-noise conditions transferred to the high-noise conditions.

Most importantly, we found that for the majority of our subjects, the effect of learning on reducing crowding is either due to an increase in sampling efficiency (by an average factor of 2.3) or a decrease in equivalent input noise (by an average factor of 3.7) but not both. Moreover, with an identical training regimen, those subjects who began with a lower sampling efficiency primarily improved in sampling efficiency, and those who began with a higher equivalent input noise primarily reduced their equivalent input noise. This was not due to any ceiling or floor effect since the unimproved dimension still had plenty of room for improvement. Training with no external noise does not disproportionally benefit the low-noise conditions—had that been the case, most of the improvements would have been found in equivalent input noise and not in sampling efficiency.

A parsimonious explanation of these findings, both in regard to crowding and the effect of learning on reducing crowding, is that a primary mode of learning is to adjust the size of a perceptual window. We assume that visual features within the window are integrated with little regard to their precise spatial location, and thus the visual system is unable to differentiate target features from surround features. Subjects with an inappropriately large window, which let in many flanker features in addition to the target features, had high equivalent input noise. These subjects improved by appropriately reducing their window sizes, which reduced the equivalent input noise without affecting sampling efficiency. Subjects with an inappropriately small or misplaced window, which did not have sufficient coverage of the target, had low sampling efficiency. They improved by increasing their window sizes appropriately to increase target coverage, leading to an increase in efficiency without changing the equivalent input noise. In theory, a subject can also improve by repositioning the perceptual window to maximize its coverage of the target while minimizing any coverage of the surround. This mode of learning would lead to both an increase in efficiency and a decrease in equivalent input noise. Our data show that this mode of learning is in the minority.

Our notion of the perceptual window is consistent with what Pelli et al. (2007) referred to as an “isolation field”, which was later renamed “combination field” (Pelli & Tillman, 2008). Pelli et al. showed that the shape and eccentricity of the isolation field, measured with non-reading tasks, determine reading speed in the periphery. They did not specify how feature integration is performed within the isolation field, except that feature integration within the field is mandatory regardless of whether the features are from the target or flankers. They also implied that the size of the combination field was fixed and possibly dictated by low-level cortical circuitry. Our data suggest that the size of this combination field is malleable to some extent via learning. This is consistent with the finding of Chung (2007) that perceptual training led to a 38% reduction of the spatial extent of crowding.

The current study was designed to characterize the effects of crowding in terms of sampling efficiency and equivalent noise before and after perceptual training by measuring contrast thresholds at various external noise levels. Thresholds improved as a result of practice, but threshold alone provides only a partial quantification of crowding. It has been argued that crowding should be described with two values: one that measures performance (accuracy or threshold) and another that measures the spatial extent of crowding (Chung & Bedell, 1995; Pelli & Tillman, 2008, online supplement). It is conceivable that practice improves performance without reducing its spatial extent. Nevertheless, data from the current study are such that a single parameter—the size of a perceptual window for feature processing—provides a concise explanation of an otherwise complex pattern of results. The current study, however, was not designed to directly measure the size of this window. A new study will be required to test this prediction and to refine the definition of such a perceptual window, including the relationship between the spatial extent of the perceptual window and that of crowding, which are not necessarily the same.

Conclusion

Consistent with the findings of Chung (2007), practicing identification of crowded letters in peripheral vision improves letter identification performance both in the flanked and unflanked conditions. We found that the improvement in the unflanked condition following practice was mostly attributable to an improvement in sampling efficiency, suggesting an increased utilization of valid letter features. Relative to the improved performance in

Downloaded from jov.arvojournals.org on 06/28/2019

the unflanked condition, improvement in the flanked condition was attributed to either an increase in the efficiency or a decrease in the equivalent input noise. This pattern of results is consistent with a corresponding adjustment in the spatial extent of a perceptual window for feature processing toward an optimal size. While crowding was significantly reduced after 6 days of training, the level of crowding remained substantial.

Appendix A

Estimated parameters for individual subjects

Tables A1 and A2.

Appendix B

Contrast energy and noise power spectral density

Contrast energy of a stimulus is defined as the sum of the squared pixel contrast over the signal region of the stimulus multiplied by the area of a stimulus pixel. The "signal region" for the current experiment is defined to be the same as the rectangular region masked by the external noise. The threshold contrast energy (E) reported in the current study is the contrast energy at threshold contrast (c) averaged over the 26 letter stimuli. Specifically,

$$E = c^2 \frac{1}{26} \sum_{j=1}^{26} \sum_{i \in S} t_{j,i}^2 \Delta x \Delta y, \tag{B1}$$

where $\Delta x = \Delta y = 0.0777°$ is the width and height of a stimulus pixel, S is the signal region, and $t_{j,i}$ is the contrast of pixel i of letter j when the letter is presented at a contrast of 1.0. For the letter stimuli used in the current experiment, the scaling constant $\left(\frac{1}{26}\sum_{j=1}^{26}\sum_{i \in S} t_{j,i}^2 \Delta x \Delta y\right)$ between E and c^2 varies with letter size and was numerically determined for each subject (in units of deg^2): 3.3441 (AL), 2.1874 (BW), 4.3254 (CT), 1.9462 (LM), 2.3135 (MB), and 5.2404 (SL).

Noise power spectral density (N) for the white noise used in the experiments (pixel-wise contrast noise of independent and identically distributed (iid) Gaussian with zero mean) is equal to the variance of a noise pixel divided by the 2-sided bandwidth of the noise; the 2-sided bandwidth of the noise is equal to the reciprocal of the area of a stimulus pixel. That is,

$$N = c_{\text{noise}}^2 \Delta x \Delta y, \tag{B2}$$

where c_{noise} is the rms contrast of the noise.

Appendix C

Ideal observer and the ideal observer EvN slope

The ideal observer for the task in the current study can be derived from first principles. Given an image I, the statistically optimal decision rule is to make the response that "center letter is r" for the most probable r:

$$\begin{aligned} r &= \arg\max_r \Pr(r|I) \\ &= \arg\max_r \Pr(r|I_{\text{target}}), \end{aligned} \tag{C1}$$

where I_{target} is the region of the image where the target letter is presented. Whether the target letter is flanked is irrelevant because the ideal observer does not have spatial uncertainty and the target letter and flankers do not overlap spatially. Applying Bayes' rule, ignoring scaling factors that do not depend on r, and knowing that (1) each letter is equally likely to be the target, and (2) the masking contrast noise comprises of an independent and identically distributed Gaussian on each pixel with a mean of 0 and a standard deviation of c_{noise}, we have

$$\begin{aligned} r &= \arg\max_r \Pr(r|I_{\text{target}}) \\ &= \arg\max_r \exp\left(-\frac{\|I_{\text{target}} - cT_r\|^2}{2c_{\text{noise}}^2}\right) \\ &= \arg\min_r \|I_{\text{target}} - cT_r\|^2, \end{aligned} \tag{C2}$$

where $\|\cdot\|$ is the L2 norm of a vector, c is the test contrast of the target letter, and T_r is the template of letter r at a contrast of 1.0.

The EvN function for this ideal observer is analytically a straight line passing through the origin (Tjan et al., 1995). To compute the EvN slope of the ideal observer, we used numerical simulation and binary search to look for c that led to letter identification accuracy of 50% with c_{noise} set at a convenient value of 1.0 (luminance is unbounded in numerical simulation, hence no issue of noise clipping). Using an efficient implementation of the

Downloaded from jov.arvojournals.org on 06/28/2019

Unflanked

		95% Confidence interval			95% Confidence interval	
Subject	m (unitless)	Lower bound	Upper bound	N_{eq} (deg^2) $\times 10^{-5}$	Lower bound $\times 10^{-5}$	Upper bound $\times 10^{-5}$
Pre-test						
AL	452	401	511	3.21	2.67	3.74
BW	555	489	844	1.89	−1.24	2.05
CT	647	582	720	2.06	1.95	2.54
LM	469	421	519	3.22	2.70	3.32
MB	432	383	510	3.11	2.50	3.90
SL	655	561	1115	2.30	−1.46	2.48
Post-test						
AL	359	319	402	3.71	3.25	4.66
BW	514	458	804	2.18	−1.41	2.45
CT	477	426	537	2.33	2.17	2.87
LM	404	353	462	4.11	3.71	5.59
MB	310	280	344	2.83	2.32	2.90
SL	394	351	444	3.22	2.74	3.76
Follow-up test						
AL	326	285	373	4.51	3.82	4.71
BW	NA	NA	NA	NA	NA	NA
CT	488	431	551	2.39	2.05	2.38
LM	526	457	601	2.59	2.10	3.06
MB	288	256	328	3.91	3.27	4.46
SL	593	524	867	1.40	−1.04	1.59

Flanked

		95% Confidence interval			95% Confidence interval	
Subject	m (unitless)	Lower bound	Upper bound	N_{eq} (deg^2) $\times 10^{-5}$	Lower bound $\times 10^{-5}$	Upper bound $\times 10^{-5}$
Pre-test						
AL	1756	725	2585	27.5	15.4	73.0
BW	2523	1312	3988	35.9	20.6	79.7
CT	6596	4154	9021	10.1	7.38	18.1
LM	2051	1645	2512	10.9	8.57	18.8
MB	2612	2133	3115	6.59	4.28	8.44
SL	2019	997	3109	39.0	25.7	99.4
Post-test						
AL	660	530	801	14.7	12.3	18.5
BW	3658	3097	4227	4.44	3.68	5.65
CT	1461	1196	1785	4.49	2.86	7.08
LM	1269	786	1709	20.2	15.8	40.8
MB	769	648	910	7.08	5.32	8.28
SL	871	686	1070	10.7	7.63	14.3
Follow-up test						
AL	998	826	1174	7.13	5.19	9.27
BW	NA	NA	NA	NA	NA	NA
CT	2247	1836	3768	2.17	−1.23	2.61
LM	1731	1334	2147	9.72	8.01	14.8
MB	775	640	927	9.13	7.11	11.0
SL	1865	1587	2166	3.12	2.90	4.37

Table A1. Estimated parameters of Equation 1 from data for each subject in each condition and test: slope of EvN line (m), equivalent input noise ($N_{eq} = E_0/m$), and their 95% confidence interval. The lines corresponding to these parameters are shown in Figure 3.

Downloaded from jov.arvojournals.org on 06/28/2019

	Average accuracy during training		
Subject	Day 1	Day 6	(Day 6) − (Day 1)
AL	0.488	0.563	0.075
BW	0.462	0.548	0.086
CT	0.474	0.630	0.156
LM	0.579	0.634	0.055
MB	0.514	0.672	0.158
SL	0.508	0.639	0.131

Table A2. Letter identification accuracy on the first and last days of training.

ideal observer described in Tjan and Legge (1998),[2] we ran 20 simulations, each consisting of 7800 trials to test each letter 300 times with different noise samples per simulation. Corresponding to the letter stimuli used with each subject in the experiment, the ideal observer slopes (m_{ideal}) were found to be (±SE): 8.126 ± 0.025 (AL), 8.336 ± 0.029 (BW), 8.192 ± 0.028 (CT), 8.488 ± 0.030 (LM), 8.103 ± 0.031 (MB), and 8.201 ± 0.025 (SL). The ratio m_{ideal}/m is the sampling efficiency for a human observer with an EvN slope of m. The subjects' EvN slopes from all the test conditions are provided in Table A1. (The same m_{ideal} for a subject applies to all conditions for that subject.)

Acknowledgments

We thank Paul Bulakowski for his comments on an earlier draft of this paper, and Denis Pelli, an anonymous reviewer, and James Elder (the editor) for their extensive and detailed comments. This work was funded by NIH/NEI Grants R01-EY017707 (BST), R01-EY012810 (STLC), and R01-EY016093.

Commercial relationships: none.
Corresponding author: Bosco S. Tjan.
Email: btjan@usc.edu.
Address: SGM 501, Los Angeles, California, USA.

Footnotes

[1] We distinguish between an ideal observer, which is the statistically optimal observer for the given stimuli and task, and an ideal observer model, which is a model of a human observer based on an ideal observer with respect to the stimuli, task, and the explicitly stated limiting factors, such as internal noise and down-sampling.

[2] The last line of Equation A3 in Tjan and Legge (1998) should read: $-2a^2XT - 2a\sigma NT + a^2TT$. The error was typographical and did not affect their implementation.

References

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature, 226,* 177–178. [PubMed]

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436. [PubMed]

Burgess, A. E., & Colborne, B. (1988). Visual signal detection. IV. Observer inconsistency. *Journal of the Optical Society of America A, Optics and Image Science, 5,* 617–627. [PubMed]

Chung, S. T. L. (2007). Learning to identify crowded letters: Does it improve reading speed? *Vision Research, 47,* 3150–3159. [PubMed] [Article]

Chung, S. T. L., & Bedell, H. E. (1995). Effect of retinal image motion on visual acuity and contour interaction in congenital nystagmus. *Vision Research, 35,* 3071–3082. [PubMed]

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, 33,* 695–709. [PubMed]

Chung, S. T. L., Levi, D. M., & Legge, G. E. (2001). Spatial-frequency and contrast properties of crowding. *Vision Research, 41,* 1833–1850. [PubMed]

Chung, S. T. L., Levi, D. M., & Tjan, B. S. (2005). Learning letter identification in peripheral vision. *Vision Research, 45,* 1399–1412. [PubMed]

Conrey, B., & Gold, J. M. (2006). An ideal observer analysis of variability in visual-only speech. *Vision Research, 46,* 3243–3258. [PubMed]

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

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, Optics, Image Science, and Vision, 14,* 2406–2419. [PubMed]

Flom, M. C., Weymouth, F. W., & Kahneman, D. (1963). Visual resolution and contour interaction. *Journal of the Optical Society of America, 53,* 1026–1032.

Freeman, J., & Pelli, D. G. (2007). An escape from crowding. *Journal of Vision, 7*(2):22, 1–14, http://journalofvision.org/content/7/2/22, doi:10.1167/7.2.22. [PubMed] [Article]

Downloaded from jov.arvojournals.org on 06/28/2019

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

He, S., Cavanagh, P., & Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature, 383,* 334–337. [PubMed]

Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology, 43,* 171–216. [PubMed]

Leat, S. J., Li, W., & Epp, K. (1999). Crowding in central and eccentric vision: The effects of contour interaction and attention. *Investigative Ophthalmology and Visual Science, 40,* 504–512. [PubMed]

Legge, G. E., Kersten, D., & Burgess, A. E. (1987). Contrast discrimination in noise. *Journal of the Optical Society of America A, Optics and Image Science, 4,* 391–404. [PubMed]

Levi, D. M. (2008). Crowding—An essential bottleneck for object recognition: A mini-review. *Vision Research, 48,* 635–654. [PubMed]

Levi, D. M., Hariharan, S., & Klein, S. A. (2002). Suppressive and facilitatory spatial interactions in peripheral vision: Peripheral crowding is neither size invariant nor simple contrast masking. *Journal of Vision, 2*(2):3, 167–177, http://journalofvision.org/content/2/2/3, doi:10.1167/2.2.3. [PubMed] [Article]

Lu, Z., Chu, W., Dosher, B. A., & Lee, S. (2005). Perceptual learning of Gabor orientation identification in visual periphery: Complete inter-ocular transfer of learning mechanisms. *Vision Research, 45,* 2500–2510. [PubMed]

Lu, Z., & Dosher, B. A. (2008). Characterizing observers using external noise and observer models: Assessing internal representations with external noise. *Psychological Review, 115,* 44–82. [PubMed]

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, and Vision, 16,* 764–778. [PubMed]

Nandy, A. S., & Tjan, B. S. (2007). The nature of letter crowding as revealed by first- and second-order classification images. *Journal of Vision, 7*(2):5, 1–26, http://journalofvision.org/content/7/2/5, doi:10.1167/7.2.5. [PubMed] [Article]

Nandy, A. S., & Tjan, B. S. (2008). Efficient integration across spatial frequencies for letter identification in foveal and peripheral vision. *Journal of Vision, 8*(13):3, 1–20, http://journalofvision.org/content/8/13/3, doi:10.1167/8.13.3. [PubMed] [Article]

Pelli, D. G. (1981). The effects of visual noise (Doctoral dissertation, Physiology Department, Cambridge University).

Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A, Optics and Image Science, 2,* 1508–1532. [PubMed]

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

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

Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision, 4*(12):12, 1136–1169, http://journalofvision.org/content/4/12/12, doi:10.1167/4.12.12. [PubMed] [Article]

Pelli, D. G., & Tillman, K. A. (2008). The uncrowded window of object recognition. *Nature Neuroscience, 11,* 1129–1135.

Pelli, D. G., Tillman, K. A., Freeman, J., Su, M., Berger, T. D., & Majaj, N. J. (2007). Crowding and eccentricity determine reading rate. *Journal of Vision, 7*(2):20, 1–36, http://journalofvision.org/content/7/2/20, doi:10.1167/7.2.20. [PubMed] [Article]

Pelli, D. G., & Zhang, L. (1991). Accurate control of contrast on microcomputer displays. *Vision Research, 31,* 1337–1350. [PubMed]

Strasburger, H., Harvey, L. O., Jr., & Rentschler, I. (1991). Contrast thresholds for identification of numeric characters in direct and eccentric view. *Perception & Psychophysics, 49,* 495–508. [PubMed]

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,* 3053–3069.

Tjan, B. S., & Legge, G. E. (1998). The viewpoint complexity of an object recognition task. *Vision Research, 38,* 2335–2350. [PubMed]

Townsend, J. T., Taylor, S. G., & Brown, D. R. (1971). Lateral masking for letters with unlimited viewing time. *Perception & Psychophysics, 10,* 375–378.

Tripathy, S. P., & Cavanagh, P. (2002). The extent of crowding in peripheral vision does not scale with target size. *Vision Research, 42,* 2357–2369. [PubMed]

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics, 33,* 113–120.

Downloaded from jov.arvojournals.org on 06/28/2019