UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Visual Learning for a Mid Level Pattern Discrimination Task

Permalink

https://escholarship.org/uc/item/31k3k5mx

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 22(22)

Authors

Fine, I.
Jacobs, Robert A.

Publication Date

2000

Peer reviewed

Visual Learning for a Mid Level Pattern Discrimination Task

I. Fine (fine@salk.edu)

Department of Psychology; University of California, San Diego, 9500 Gilman Drive
La Jolla, CA 92093-0109, USA

Robert A. Jacobs (robbie@bcs.rochester.edu)

Department of Brain and Cognitive Sciences, University of Rochester
Rochester, NY 14627, USA

Abstract

Our goal was to examine the plasticity of the human visual system at mid to high levels of visual processing. It is well understood that early stages of visual processing contain cells tuned for spatial frequency and orientation. However images of real-world objects contain a wide range of spatial frequencies and orientations. We were interested in how different spatial frequencies and orientations are combined. We used a pattern discrimination task - observers were asked to discriminate small changes in a "wicker-like" stimulus consisting of six superimposed sinusoidal gratings. Observers were asked to discriminate a 15% spatial frequency shift in two of these sinusoidal components, which were masked by four noise components. We found large amounts of perceptual learning for this task – over eight sessions of training observers' average percent correct increased by 31%, corresponding to their thresholds dropping to a third of their initial values. Further experiments suggest that learning was based on changes within a mid level stage of processing intermediate between low-level analyzers tuned for orientation and spatial frequency and high-level pattern matching or object tuned cells. This mid level stage seems to be "very roughly Fourier" and combines information from individual gratings using probability summation. This stage of processing is also remarkably plastic compared to earlier stages of processing.

Introduction

A great deal is known about low level visual pattern analyzers and their role in visual perception. At early stages of processing retinal input is represented by low level analyzers tuned for spatial frequency and orientation with receptive fields of limited spatial extent - properties very similar to simple cells in V1 (see Graham, 1989 for a review). However images of real-world objects contain a wide range of Fourier components, and therefore the combination of information across these low level analyzers is necessary to reliably recognize objects. Evidence suggests that there may be mid level mechanisms selectively pooling information across low level analyzers tuned for a wide range of spatial frequencies or orientations (e.g. Georgeson, 1992; Derrington & Henning, 1989; Burr & Morrone, 1994; Graham & Sutter, 1998; Olzak & Thomas, 1999).

It has been argued that relatively early stages of the visual system (V1) change with training (e.g. Ball & Sekuler, 1987; Fahle & Edelman, 1992; Sagi & Tanne,

1994; Ahissar & Hochstein, 1995,1996; Saarinen & Levi, 1995; Fahle & Morgan, 1996; Schoups & Orban, 1995). In addition, some learning effects have been noted (Olzak, personal communication, 1995; Fiorentini & Berardi, 1981) for tasks involving compound grating discriminations thought to involve mid level mechanisms.

The following experiments provide support for the existence of mid-level mechanisms pooling over analyzers tuned for spatial frequency and orientation. These mid level mechanisms are shown to be far more adaptable as a function of experience than low level analyzers.

Experiment 1

The purpose of Experiment 1 was to measure learning for a complex "wicker" stimulus that required observers to combine information over a wide range of spatial frequencies and orientations.
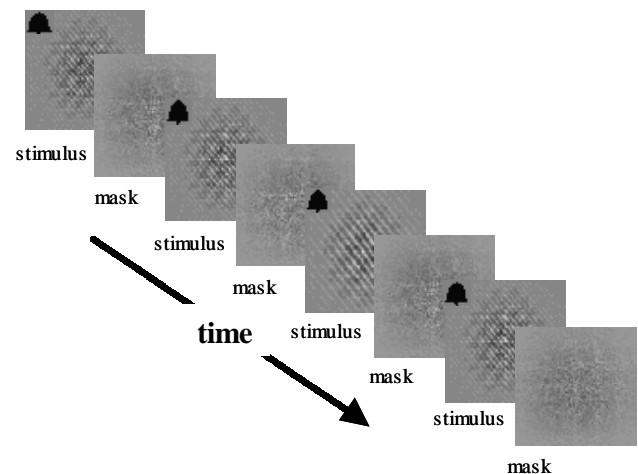
Methods

Figure 1: Diagram of the task used in the experiment.

Five observers were asked to perform a four alternative forced choice discrimination task (Figure 1). Four stimuli were presented sequentially in time. A two-dimensional white noise pattern was presented after each stimulus to reduce afterimage interference. Observers were asked to

indicate which of the four stimuli was different from the others using a key press. There are two important advantages of this four alternative forced choice procedure. First, the chance success rate was 25%, thereby providing more information per trial than a two alternative forced choice task. Second, such a task allows a same-different judgment without potential criterion effects (observers showing a bias towards responding same or different). The task was carried out using free-fixation. Observers were given auditory feedback and were self paced. Observers completed eight sessions, and completed 250 trials per session. Observers never carried out more than a single session in a day, and carried out three to five sessions a week.
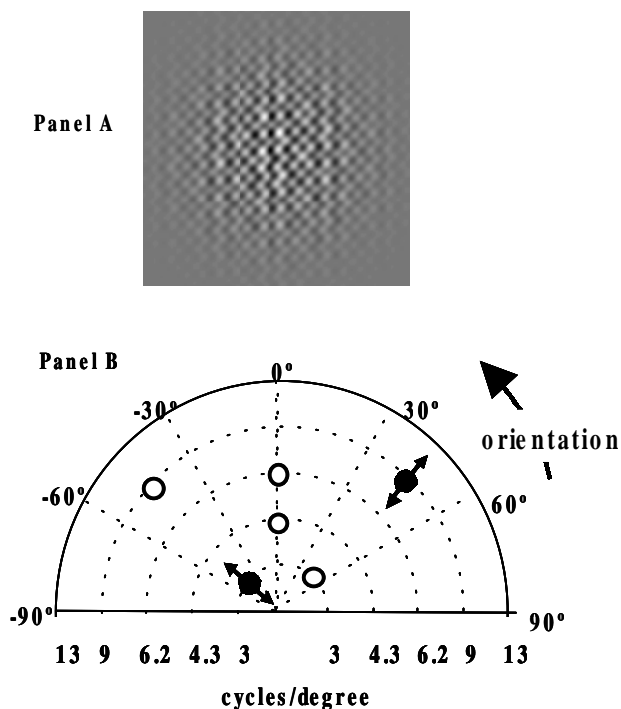

Panel A

Panel B

Figure 2: Panel A. Illustration of a typical stimulus. Panel B. Fourier representation of stimuli. The radius represents spatial frequency and the angle represents orientation. One signal component had a spatial frequency of 2.55 or 3.45 cycles/degree, an orientation of -45° and contrasts varying between 1.6-12.8%. The other signal component had a spatial frequency of 7.65 or 10.35 cycles/degree, an orientation of 45° and contrasts varying between 5.5-44%. There were four sinusoidal noise components, represented by empty circles: 1) spatial frequency of 9 cycles/degree, -45° orientation, 11% contrast 2) spatial frequency of 3 cycles/degree, 45° orientation, 3.2% contrast 3) spatial frequency of 4.3 cycles/degree, 0° orientation, 7.1% contrast 4) spatial frequency of 6.2 cycles/degree, 0° orientation, 7.1% contrast.

Figure 2 Panel A shows what a typical stimulus looked like. Each stimulus contained two signal compo-

nents and four sinusoidal noise components. Figure 2, Panel B represents the stimuli in Fourier space using polar coordinates. The radius represents spatial frequency and the angle represents orientation. The black filled circles represent the two possible signal components. These signal components were widely separated in orientation (at least 90° to each other) and widely separated in spatial frequency (approximately two octaves apart). One signal component was centered on 3 cycles/degree and had an orientation of –45 degrees and the other signal component was centered on 9 cycles/degree and had an orientation of 45 degree. Observers were asked to detect a 15% shift in the spatial frequency of the signal components, represented by the black arrows. The contrasts of the signal components were manipulated (based on pilot data) so each observer was presented with a range of difficulty levels. The empty circles in Figure 2 represent the four sinusoidal noise components that were added to the stimulus.

Stimuli were modulated spatially by a two dimensional Gaussian envelope with a sigma of 0.5693 degrees and temporally by a Gaussian envelope of sigma 0.237 seconds centered within a 0.67 second temporal window. The phases of the sinusoidal noise components were varied randomly across each interval of each trial. The phases of the signal components were varied randomly between each trial, and remained constant across the four intervals within each trial. Stimuli were presented using the VideoToolbox and Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997).

Observers were undergraduate or graduate students from the University of Rochester, varying in age between 19-28 years of age. Observers had normal or corrected to normal vision. Further details of this experimental procedure are described in Fine & Jacobs (2000).

Results and Conclusions

The black squares in Figure 3 show the percent correct as a function of session averaged across observers in Experiment 1 (black squares). All five observers showed a significant improvement in their performance over eight sessions. Observers' average percent correct increased by 31%, corresponding to a two-third decrease in their thresholds.

Most perceptual learning studies have been carried out using simple stimuli (grating discrimination or Vernier tasks). Learning effects for these low level tasks tend to be small or non-existent in the fovea (e.g. Fiorentini and Berardy, 1981, Beard, Levi and Reich, 1995). In contrast, we found large learning effects in the fovea, suggesting strongly that our task is mediated by a higher stage of processing than more simple tasks, and that this stage of processing is far more plastic than earlier stages.

These improvements in performance with practice were relatively long lasting, none of the observers showed any

decline in performance when retested more than a month after training.

It is worth noting that observers showed faster improvement for easier stimuli than for more difficult stimuli, suggesting possible bootstrapping from easy to difficult stimuli (Ahissar & Hochstein, 1997).

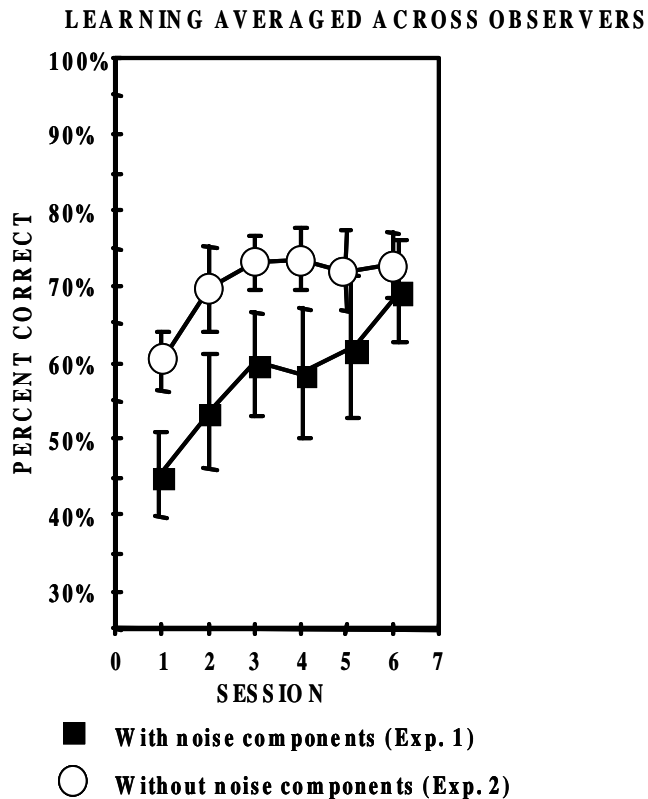**LEARNING AVERAGED ACROSS OBSERVERS**

Figure 3: Percent correct as a function of session averaged across observers with the sinusoidal noise components (Experiment 1 - black squares) and without the noise components (Experiment 2 - empty circles). The x-axis shows the session and the y-axis shows the percent correct. Standard error bars are shown.

Experiment 2

The extent of learning found in Experiment 1 suggests that performance in our task might be mediated by a mid level stage of processing rather than earlier stages. Experiment 2 was designed to exclude the possibilities that the learning found in Experiment 1 was due either to learning in low level mechanisms, or to improved non-visual cognitive strategies (such as learning the key press procedure, learning to fixate, learning the temporal structure of the task, etc.).

If the learning demonstrated in Experiment 1 was due to tuning changes within low level analyzers tuned for both spatial frequency and orientation then removing the sinusoidal noise components would not affect the amount of learning shown. The sinusoidal noise components in Experiment 1 were positioned so as to be invisible to analyzers tuned for the spatial frequency and orientation of the signal components (see Figure 2, Panel B) - every noise component differed from the signal components by at least 45 degrees of orientation or almost two octaves of spatial frequency. Estimates of the tuning of low level analyzers by other authors predict little low level masking between sinusoidal components separated by either two octaves of spatial frequency or 45 degrees orientation (Graham, 1989).

If the learning in Experiment 1 was due to non-visual cognitive strategies then we would expect an equal amount of learning in Experiments 1and 2 – the only difference between the two experiments was in the visual stimulus.

Methods

Display and task were identical to those used in Experiment 1. Only the stimulus differed in Experiment 2, in that the sinusoidal noise components (the empty circles of Figure 2) were no longer present - i.e. observers were asked to discriminate changes in spatial frequency within a simple plaid pattern.

Without the noise components the task would be trivially easy for the contrast levels and spatial frequency shifts used in Experiment 1. The difficulty of the task was adjusted by reducing the spatial frequency shift to between ±2.5% and ±12.5% (as opposed to 15% in Experiment 1) to avoid ceiling effects.

Three observers were given six sessions of training on the task.

Results and Conclusions

As shown by the empty circles in Figure 3, observers showed much less learning without the sinusoidal noise components. Observers showed some learning between sessions 1 and 2, but little learning after the second day. There was no significant drop in threshold across the three observers.

Differences in the amount of learning between Experiment 1 and 2 cannot be explained by ceiling effects. Initial performance was closely matched for the majority of subjects. In Experiment 1 three of the five observers in performed between 50-60% correct in the first session. In Experiment 2 two of the three observers performed between 50-60% correct in the first session. In addition, none of the observers' performance reached 90% correct by the end of training in either experiment.

There was some learning (~ 7%) between the first and second day in both Experiment 1 and Experiment 2. Given that we used naïve observers we think it likely that these learning effects are mainly due to non-visual factors - learning the key press procedure etc. However an alternative possibility is that this learning between the first and second day was due to learning in low level analyzers.

In any case, most of the learning shown in Experiment 1 was *after* the second session and cannot be due either to learning in low level analyzers or to learning better non-visual cognitive strategies.

Experiment 3

We were interested in how observers might be combining information from the two signal components. We have found that observers' performance in Experiment 1 can be well described using an independent probability summation model where observers correctly discriminate the "odd man out" if they detect a shift in either component (Fine & Jacobs, 2000). Experiment 3 was designed to further test whether observers' ability to combine information could be reasonably approximated using an independent probability summation model.

Other possible combination models include non-independent combination of information from the two signal components (as suggested by Olzak and Thomas, 1999) or some type of "pattern" or "template" matching.

The task carried out in Experiment 1 can be subdivided into two tasks, as shown in Figure 4. In the *same sign* task the "odd man out" was distinguished from the distracting stimuli by both signal components being shifted in the same direction in Fourier space. In half the trials both signal components were shifted higher in spatial frequency, as shown in Figure 4 Panel A. In the other half of the trials both signal components were shifted lower in spatial frequency.

In the *opposite sign* task the "odd man out" was distinguished from the distracting stimuli by both signal components being shifted in opposite directions in Fourier space. In half the trials the high spatial frequency component was shifted higher in spatial frequency, and the low spatial frequency component was shifted lower. In the other half of the trials, as shown in Figure 4 Panel B, the high spatial frequency component was shifted lower in spatial frequency, and the low spatial frequency component was shifted higher.

Independent probability summation implies that detecting a shift in the low spatial frequency signal component is unaffected by the direction of the shift in the high spatial frequency signal component, and vice versa. Any relationship between the directions of the spatial frequency shifts within the two signal components would be invisible to such a mechanism. Consequently, according to an independent probability summation model we would expect perfect transfer of learning from same sign to opposite sign tasks.

According to most non-independent models, including pattern matching, one would expect incomplete transfer between the two stimuli.

Methods

Display and task were identical to those used in Experiment 1, however observers were either exclusively trained with same sign stimuli, then tested with opposite sign stimuli, or were trained with opposite sign stimuli, then tested with same sign stimuli. Four observers were tested in all, two were trained with same sign stimuli and two were trained with opposite sign stimuli. Observers were given six sessions of training before being tested with the novel stimuli.

Panel A – SAME SIGN SHIFT

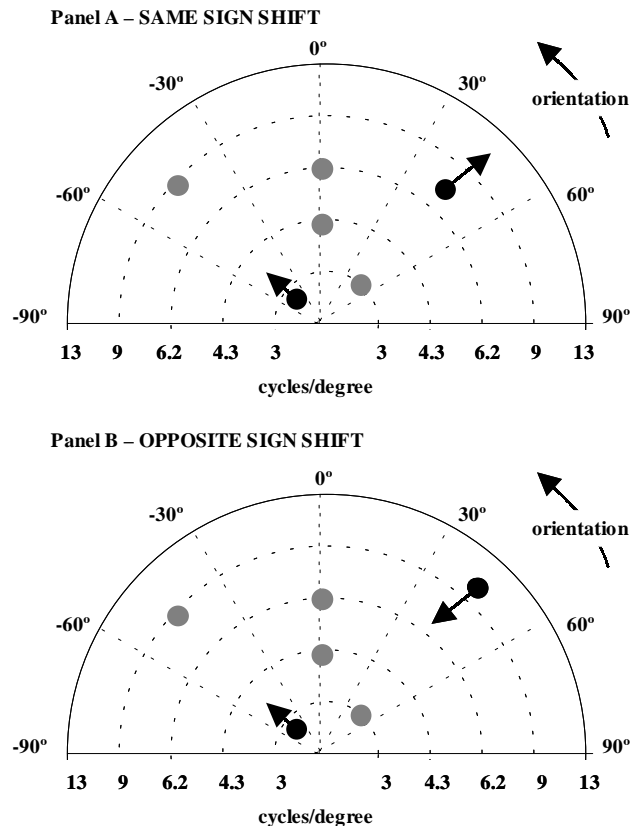**Panel B – OPPOSITE SIGN SHIFT**

Figure 4: Polar plot of the stimuli used for the transfer of task experiment. Panel A shows the same sign shift stimulus and Panel B shows the opposite sign shift stimulus.

Results and Conclusions

None of the four observers showed any drop in performance when tested with the novel stimulus. Interestingly only one of the four observers even noticed that the stimulus had changed. This perfect transfer of learning between same and opposite sign tasks is consistent with observers combining information independently, and is incompatible with most non-independent models (Olzak & Thomas, 1999), including pattern matching.

Interestingly, the shift in the signal components in the same sign task is compatible with a change of scale (as if both signal components moved closer or further away from the observer), while the shift in the signal components in the opposite sign task is compatible with a change in shape. The total transfer of learning between the two tasks suggests that "scale-invariance" may not yet be differentially encoded at this stage of processing.

General Conclusions

Our data support the existence of a mid level stage of processing intermediate between low and high levels of visual processing. This level of processing seems to be

"very roughly Fourier" in that it still represents stimuli in terms of their spatial frequency and orientation. Information from low level analyzers tuned for spatial frequency and orientation seems to be combined using probability summation. This stage may be responsible for beginning to selectively process information, extracting the combinations of spatial frequency and orientation that define meaningful objects. As our knowledge of the mechanisms underlying mid level visual tasks increases it should be possible to ask increasingly refined questions about the role of these mid level mechanisms, and in particular, the role adaptability plays in allowing such mechanisms to represent an unpredictable world. Interestingly, our studies show that this mid level stage of processing seems to be far more plastic than earlier stages.

As cells become more specific in what they represent, an increasing number of cells become necessary if all possible stimuli are to be represented. This is the paradox of the "grandmother cell" – not every possible object can have its own feature detectors in the brain without a prohibitive number of cells. Despite this apparent paradox, cells in the brain have been shown to be remarkably specific (e.g. Desimone, Albright, Gross, & Bruce, 1984; Logothetis, Pauls, Poggio, 1995). Neural plasticity may be a way of alleviating the trade-off between cell specificity and limited cell numbers. By dynamically changing neural representations as a function of experience cells can be preferentially allocated to represent behaviorally important stimuli. If this is the case, then we should find an intimate relationship between plasticity and specificity - as representations become more selective, they should also become more plastic.

References

Ahissar, M., & Hochstein, S. (1996). Learning pop-out detection: Specificities to stimulus characteristics. *Vision Research*, 36(21), 3487-3500.

Ahissar, M., Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, 387, 401-6

Ball, K., & Sekuler, R. (1987). Direction-specific improvement in motion discrimination. *Vision Research*, 276(6), 953-965.

Beard, B.L., Levi, D.L., & Reich, L.N. (1995). Perceptual learning in parafoveal vision. *Vision Research*, 35(12), 1679-1691.

Brainard, D. H. (1997) The Psychophysics Toolbox, Spatial Vision , 10:443-446.

Burr, D.C., & Morrone, M.C. (1994). The role of features in structuring visual images. In M. Morgan (Ed.), *Higher-Order Processing in the Visual System*. New York, New York: Wiley.

Derrington, A.M., & Henning, G.B. (1989). Some observations on the masking effects of two-dimensional stimuli. *Vision Research*, 29(2), 241-246.

Fahle, M., & Edelman, S. (1992). Long-term learning in vernier acuity: Effects of stimulus orientation, range and of feedback. *Vision Research*, 33(3), 397-412.

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

Fine, I., & Jacobs, R.A. (2000). Perceptual learning for a pattern discrimination task. *Submitted to Vision Research*.

Fiorentini, A., & Berardi, N. (1981). Learning in grating waveform discrimination: Specificity for orientation and spatial frequency. *Vision Research*, 21(7), 1149-1158.

Georgeson, M.A. (1992). Human vision combines oriented filters to compute edges. *Proceedings of the Royal Society of London*, 249B, 235-245.

Graham, N. (1989). *Visual pattern analysers*. Oxford, UK: Oxford Science Publications.

Graham, N., & Sutter, A. (1998). Spatial summation in simple (Fourier) and complex (non-Fourier) channels in texture segregation. *Vision Research*, 38(2), 231-257.

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 of the United States of America,* 88, 4966-4970.

Liu, Z. & Weinshall, D. (1999). Mechanisms of generalization in perceptual learning. *Advances in Neural Information Processing Systems*, MIT Press.

Logothetis, N.K., Pauls, J. & Poggio, T. (1995) Shape representation in the inferior temporal cortex of monkeys. *Current Biology,* 5(5):552-63.

Olzak, L.A., & Thomas, J.P. (1999). Neural recoding in human pattern vision: Model and mechanisms. *Vision Research*, 39(2), 231-256.

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

Saarinen, J., & Levi, D.M. (1995). Perceptual learning in vernier acuity: What is learned? *Vision Research*, 35(4), 519-527.

Sagi, D., & Tanne, D. (1994). Perceptual learning: Learning to see. *Current Opinion in Neurobiology*, 4(2), 195-199.

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