

REPORT DOCUMENTATION PAGE

AD-A268 711



Form Approved
OMB No. 0704-0188

Approved for release by the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

REPORT DATE

3. REPORT TYPE AND DATES COVERED

ANNUAL 15 Apr 92 TO 14 Apr 93

4. TITLE AND SUBTITLE

PSYCHOPHYSICAL ANALYSES OF PERCEPTUAL REPRESENTATIONS

5. FUNDING NUMBERS

AFOSR-90-0274
61103D
3484
HS

6. AUTHOR(S)

Dr Irving Biederman

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Univ of Minnesota
Dept of Psychology
75 East River Road
Minneapolis, MN 55455

AFOSR-TR

8. PERFORMING ORGANIZATION
REPORT NUMBER

93 0626

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFOSR/NL
110 Duncan Avenue, Suite B115
Bolling AFB DC 20332-0001

Dr John F. Tangney

10. SPONSORING / MONITORING
AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

DTIC
ELECTE
AUG 30 1993
S A D

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release;
distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Research during the year has been divided between studies at USC (Biederman and students) and Minnesota. Our research continues to focus on linking early sensory representations to higher-level perceptual representations. For this reason, we refer to our Center informally as the "Middle Kingdom." Studies outlined below have examined the sensory/perceptual "middle ground" in object recognition, depth perception, reading acuity, and auditory perception. Several of our studies have used ideal-observer analysis. The ideal-observer approach provides a means for quantifying the information available to perception and for evaluating the effectiveness with which humans use that information.

93-19916



14. SUBJECT TERMS

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION
OF REPORT

(U)

18. SECURITY CLASSIFICATION
OF THIS PAGE

(U)

19. SECURITY CLASSIFICATION
OF ABSTRACT

(U)

20. LIMITATION OF ABSTRACT

(UL)

ANNUAL TECHNICAL REPORT

Submitted to: Air Force Office of Scientific Research
AFOSR/NL
Building 410
Bolling A.F.B., D.C. 20332-6448

Dr. John Tangney, Program Manager

Title: *PSYCHOPHYSICAL ANALYSES OF PERCEPTUAL REPRESENTATIONS*

Grant No. AFOSR-90-0274

Period Covered: April 16, 1992 - April 15, 1993

Submitting Organization: University of Minnesota
Office of Res. & Technology Transfer Admin.
Suite 201
1100 Washington Ave. S.
Minneapolis, MN 55415

PI:¹ Professor Gordon E. Legge
(612) 625-0846

Gordon E. Legge
023-42-1595

Date of Report: 16 July 1993

Accession For	
NTIS CRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and/or Special
A-1	

DTIC QUALITY INSPECTED 3

¹During this project year, the grant was amended in order to subcontract funds to the University of Southern California for research under the direction of Professor Irving Biederman.

PROGRESS REPORT: 1992-1993

Introduction

Research during the year has been divided between studies at USC (Biederman and students) and Minnesota. Our research continues to focus on linking early sensory representations to higher-level perceptual representations. For this reason, we refer to our Center informally as the "Middle Kingdom." Studies outlined below have examined the sensory/perceptual "middle ground" in object recognition, depth perception, reading acuity, and auditory perception. Several of our studies have used ideal-observer analysis. The ideal-observer approach provides a means for quantifying the information available to perception and for evaluating the effectiveness with which humans use that information.

In the following paragraphs, we describe many projects supported by the grant. In some cases, projects that were described in detail in last year's report reached publication this year, and only the citation is given. A list of publications and conference presentations follows these descriptions.

Dynamic binding in a neural network for shape recognition. (Hummel & Biederman, 1992)

Upon exposure to a single view of an object, the human can readily recognize that object from any other view that preserves the parts in the original view. Experimental evidence suggests that this fundamental capacity reflects the activation of a viewpoint invariant structural description specifying the object's parts and the relations among them. This paper presents a neural network model of the process whereby a structural description is generated from a line drawing of an object and used for object classification. The model's capacity for structural description derives from its solution to the dynamic binding problem of neural networks: independent units representing an object's parts (in terms of their shape attributes and interrelations) are bound temporarily when those attributes occur in conjunction in the systems input. Temporary conjunctions of attributes are represented by synchronized (or phase locked) oscillatory activity among the units representing those attributes. Specifically, the model uses phase locking to: a) parse images into their constituent parts; b) bind together the attributes of a part; and c) determine the relations among the parts and bind them to the parts to which they apply. Because it conjoins independent units temporarily, dynamic binding allows tremendous economy of representation, and permits the representation to reflect the attribute structure of the shapes represented. The model's recognition performance is shown to conform well to empirical findings.

Size invariance in visual object priming. (Biederman & Cooper, 1992)

The magnitude of priming resulting from the perception of a briefly presented picture of an object in an earlier trial block, as assessed by naming reaction times (RTs), was found to be independent of whether the primed object was presented at the same or a different size as when originally viewed. In contrast, RTs and error rates for "same" responses for old-new shape judgments were very much increased by a change in object size from initial presentation. We conjecture that this dissociation between the effects of size consistency on naming and old-new shape recognition may reflect the differential functioning of two independent systems subserving object memory: one for representing the shape of an object and the other for representing its size, position, and orientation (metric attributes). With allowance for response selection, object naming RTs may provide a relatively pure measure of the functioning of the shape system. Both the shape and metric systems may affect the feelings of familiarity that govern old-new episodic shape judgments. A comparison of speeded naming and episodic recognition judgments may provide a behavioral, noninvasive technique for determining the neural loci of these two systems.

Unexceptional spatial memory in an exceptional memorist.
(Biederman, Cooper, Fox & Mahadevan (1992)

Rajan Mahadevan evidences an exceptional memory for arrays of digits. As Rajan does not appear to employ a mnemonic or imagery strategy, we tested whether Rajan is exploiting spatial memory. Rajan and eight control subjects each viewed 48 images of common objects and animals. Each image was shown either to the left or the right of fixation and was facing either left or right. Subjects had been instructed to remember the position and orientation of each picture. On a second block of trials with the same pictures but in different order, subjects judged whether the position and the orientation of each picture was the same or different than on the first block. Rajan's accuracy was reliably below the mean of the control subjects for both judgments. On a second testing, Rajan employed a strategy that essentially circumvented spatial memory. Rajan's exceptional memory capacity, so well evidenced for digits, apparently does not extend to spatial relations.

Metric invariance in object recognition: A review and further evidence. (Cooper, Biederman, & Hummel, 1992)

Phenomenologically, human shape recognition appears to be invariant with changes of orientation in depth (up to parts occlusion), position in the visual field, and size. Recent versions of template theories (e.g., Ullman, 1989; Lowe, 1987) assume that these invariances are achieved through the application of transformations such as

rotation, translation, and scaling of the image so that it can be matched metrically to a stored template. Presumably, such transformations would require time for their execution. We describe recent priming experiments in which the effects of a prior brief presentation of an image on its subsequent recognition is assessed. The results of these experiments indicate that the invariance is complete: the magnitude of visual priming (as distinct from name or basic level concept priming) is not affected by a change in position, size, orientation in depth, or the particular lines and vertices present in the image, as long as representations of the same components can be activated. An implemented seven layer neural network model (Hummel & Biederman, 1992) that captures these fundamental properties of human object recognition is described. Given a line drawing of an object, the model activates a viewpoint-invariant structural description of the object specifying its parts and their interrelations. Visual priming is interpreted as a change in the connection weights for the activation of: a) cells, termed geon feature assemblies (GFAs), that conjoin the output of units that represent invariant, independent properties of a single geon and its relations (such as its type, aspect ratio, relations to other geons), or b) a change in the connection weights by which several GFAs activate a cell representing an object.

Metric versus viewpoint invariant shape differences in visual object recognition. (Cooper & Biederman, 1993)

Purpose. Recognition by Components (Biederman, 1987) posits visual primitives (termed geons) for object recognition which are distinguished by viewpoint-invariant (i.e., nonaccidental) contrasts (VICs) of image edges (e.g. straight vs. curved, parallel vs. non parallel) rather than metric variations (e.g., length or degree of curvature). Is object recognition more sensitive to changes in VICs than to subjectively equal metric changes? **Method:** Subjects judged whether a pair of sequentially presented object images were the same or different in basic-level name (Exp. I) or physical shape (Exp. II). A pair of objects could differ either in the VICs for one of the parts (geon change) or in the length of one of the parts (metric change). For the metric changes, the variation could be either smaller than, equal to, or greater than the length that subjects judged to be subjectively equal to the geon change. In Exp. I, the members of a pair could have the same or different name. In Exp. II, members of a pair always had the same name. **Results:** Exp. I: Responses were reliably faster and less error prone for the objects that underwent metric changes, even when the change was subjectively greater than the geon change. Exp. II: It was more difficult to judge that two images represented different shaped objects when the differences between the objects were metric compared to viewpoint invariant. **Conclusion:** Object recognition is more sensitive to viewpoint invariant contrasts than to metric changes.

Binocular visual direction. (Mansfield & Legge, 1993)

We have been investigating how the binocular visual system combines information from the two eyes to determine the visual direction of features seen in stereoscopic depth. As we reported at ARVO in 1992 (Mansfield, Akutsu and Legge, 1992), we found that if a stereoscopically viewed feature has different contrasts in the left and right eyes, then its perceived direction is biased towards the visual direction seen by the eye with higher contrast. Depending on the viewing distance, this bias in visual direction can be quite large (more than 1 deg). This result rules out any theory of binocular visual direction that relies exclusively upon the geometry of binocular vision. Instead we have proposed that binocular visual direction is the most likely visual direction given the left and right eye's signals as independent estimates of the true visual direction to the target. This maximum likelihood theory explains the contrast induced shift in perceived direction because reducing the contrast of a feature usually results in an increase of the localization uncertainty associated with that feature. We have formally tested this theory by using a monocular vernier acuity task to measure the localization uncertainties caused by the different contrasts used in our original experiments. By inserting these values into our maximum likelihood model we were able to accurately predict both the change in perceived direction and the accuracy of binocular localization for targets with different contrasts in each eye. Moreover, our data demonstrate a clear distinction between the processes that determine binocular depth and binocular direction. The sensitivity for binocular depth judgments is adversely affected by reducing the contrast in one eye, whereas the sensitivity for judgments of binocular direction are largely unaffected. However, while interocular contrast differences alter the perceived direction of the binocular target, the perceived depth remains constant. This distinction suggests separate mechanisms for the perception of binocular depth and binocular direction.

MNREAD Acuity Charts. (Mansfield, Ahn, Legge & Lukber, 1993)

We have developed a new reading acuity chart that can be used to measure reading performance by individuals with either normal or low vision. Our new chart is an improvement on other charts in various ways: First, it covers a very wide range of print sizes, ranging from 1.3 logMAR (Snellen 20/400) right down to -0.5 logMAR (Snellen 20/6). The wide range of sizes ensures that an accurate measure of reading acuity can be obtained from nearly every individual. Second, the chart consists of a series of sentences, each of which has the identical length and layout (4 lines of text, each 13 character spaces long). This feature makes the chart ideal for measuring how reading speed (an objective measure of visual reading performance) is dependent on character size. The chart can be used to quickly determine the "critical character size for reading" --- the character size below which reading performance starts to deteriorate. This critical size is usually two to four times larger than the acuity limit. This chart has potential for precise measurement of near acuity, for direct evaluation of reading performance, and for evaluation of reading performance through visual

devices.

Surface orientation and illumination direction from shading. (Mamassian & Kersten, 1993)

Our previous studies of cast shadows have shown that observers make perceptual judgments consistent with a fixed light source. We have begun a new set of studies to determine the information provided by the gradient of shading on a smooth surface. We are interested in the ability to perceive the orientation of the tangent to the surface, for qualitatively different points, among different lighting conditions.

We completed some experiments using a "croissant"-shaped object, obtained from bending the long axis of an ellipsoid of revolution along a circular arc. The resulting object contains all three qualitatively different points existing on a smooth surface: elliptic, parabolic and hyperbolic. Twelve points were selected within each of these categories, and matched according to their slants. In addition to this curvature variable, we used four lighting conditions: the light source was either at the viewpoint, or at an angle from above, below or behind the object. The information provided by the occluding contour was kept constant by fixing the viewpoint. Subjects had to match the tangent plane to a textured sphere next to the object to the orientation of the surface at a designated point. The surface had a Lambertian reflectance. Subjects perceived surface orientation with an accuracy as high as 5 deg. for slant, and 10 deg. for tilt when the slant was larger than 20 deg. The manipulation of either the sign of the curvature or the position of the light source did not induce a biased percept. Finally, from Lambert's law, each measurement constrains the location of the light source to a circle on the Gaussian sphere. The illumination model implicitly used by the observer so derived comes close to the one used to generate the stimuli. Some implications can be drawn for certain assumptions commonly brought in computational studies to solve this underconstrained problem. In particular, we demonstrated the ability of the visual system to apprehend shape accurately unusual illumination conditions, and to perceive shape within regions that differ significantly from being spherical.

Multi-layer approach to surface segmentation. (Madarasmi, Kersten, & Pong, 1993)

Current Bayesian approaches to the interpolation and segmentation of range data commonly use line processes to describe the discontinuities between smooth regions and smoothness processes to constrain the surface interpretation within statistically correlated regions. This edge-based approach to segmentation is incomplete since there is no explicit grouping of image sites into the more meaningful, piece-wise smooth regions. Furthermore, psychophysical evidence such as illusory contour perception in the Kanizsa square suggests that grouping may be

performed simultaneously with boundary detection. We have continued to develop the multi-layer "depth planes" approach to surface segmentation. Each layer or depth plane represents a single smooth surface in the scene. The method outputs two types of intrinsic images: a label image assigning each image site to a depth plane (layer) and multiple depth images each containing the estimated scene depth values. We use a Bayesian framework to incorporate the various local constraints into probability distribution functions and compute the maximum (a posteriori) probability (MAP) corresponding to the solution. It is general approach to segmentation which we have applied to several types of problems.

This representation enables transparency computation since distant surfaces patches are grouped if they form a smooth global surface. In addition, partially occluded surfaces are filled in by an interpolation process within each layer. We have tested this multi-layer segmentation scheme on a range of images. We also use it as part of the regularization scheme for computing stereo disparity and applied it to random dot stereograms as well as natural gray-level images. Recently, we have extended this approach to develop an account of subjective contours as a "side effect" of partitioning the image into overlapping surfaces.

Object recognition and classification for human and ideal observers. (Liu, Knill & Kersten, 1992)

In '91/'92 we developed a novel paradigm for experimental studies of human object recognition (Liu, Kersten, and Knill, 1992). It was based on the use of ideal observer theory to estimate the statistical efficiency with which human subjects use stimulus information for performing a recognition task. We measured the statistical efficiency with which human observers made simple classification judgments of randomly shaped thick wire objects. We were able to show that human performance exceeded that of an ideal 2D template matching strategy, effectively eliminating the class of 2D template matching models as candidates for explaining the data. We also showed viewpoint dependent effects in that subjects' efficiencies were higher for learned views of objects than novel views, but that the effect decreased with increasing structure (e.g. symmetry, planarity) of the objects. Moreover, average efficiencies across all viewpoints increased with increasing regularity of objects, indicating that the visual system takes advantage of such regularities in storing object information and comparing it with image data for recognition. In the last year we have considerably extended the analysis to include a comparison of human performance with a class of computational models which have been proposed for object recognition known as Hyper Basis Function models. This involved the development of computer implementations of the models and their testing on the same task given to human subjects. We were able to show that even allowing for considerable generalizations of the models, human performance exceeded that of the models, eliminating them as candidate explanations for the results. As a result of these new analyses, we have revised the original paper completed in the previous grant year and submitted it to

Psychological Review (Liu, Knill, and Kersten, 1992).

Workshop on Visual Perception: Computation and Psychophysics. (Knill, 1992)

Over the course of the Summer and Fall, David Knill planned and organized a workshop on computational and psychophysical approaches to visual perception which was held in Chatham Bar, Massachusetts, Jan. 14-17. The workshop brought together researchers in computational vision and psychophysics to discuss ways of conceptualizing and modeling problems in visual perception. The workshop was a tremendous success and has resulted in a contract with Cambridge University Press for publication of a book with contributions from the participants. The costs of the workshop itself were paid by a separate grant from the AFOSR (AF/F49620-93-1-01-24).

Psychophysics of complex auditory signals. (Viemeister & Edwards)

The major focus of the work during the past year has been on temporal aspects of auditory perception and has focused on the detection and discrimination of modulation. Amplitude and frequency modulation exist in almost all complex auditory signals. A study of the way in which listeners can perform tasks involving modulation is essential to understanding the way in which they hear these types of signals. The following three experiments were run in order to investigate the mechanism by which modulation is encoded.

A. DETECTION AND DISCRIMINATION OF THREE-COMPONENT SIGNALS. Modulation detection thresholds were obtained for three-component signals. The amount of amplitude and frequency modulation (AM and FM) in these signals can be changed by simply changing either the relative phases or the relative amplitudes of the three components. A model was developed which used the detectability of AM and FM separately to predict the detectability of the three-component signals with AM and FM combined. Thresholds were obtained for discriminating between three component signals with different modulation depths and different modulation types. Discrimination thresholds were elevated from detection thresholds independent of the similarity between the modulation types. The data were well predicted by the model and implied a common decision statistic for both AM and QFM signals near threshold.

B. DISCRIMINATION BETWEEN AM AND FM SIGNALS. Discrimination thresholds were obtained for AM vs FM. This experiment was unique because instead of subjects discriminating between signals in which the modulation depth (AM) was equal to the modulation index (FM), subjects discriminated between AM and FM signals with equally detectable levels of modulation. Discrimination thresholds were

elevated from detection thresholds, indicating a certain amount of similarity between AM and FM signals previously unseen in this type of task. Psychometric functions were obtained for AM vs FM signals, and were found to be non-monotonic. In many cases performance was at chance even though the amount of AM and FM in the signals were 100% detectable. This indicates that the cue representing the FM signal at these levels was the same as the cue representing the AM signal. Discrimination performance at higher levels of AM and FM indicate that a second cue exists in the FM signal which is different from the AM cue.

C. DETECTION AND DISCRIMINATION OF TWO-TONE MODULATION.

Modulation detection and discrimination thresholds are currently being investigated for two-tone modulators in order to investigate the modulation encoding mechanism. Current modulation encoding theories attempt to explain the ability to detect or discriminate fluctuations in the envelope of a signal. These theories address modulation depth discrimination but not modulation frequency discrimination. Using two-tone modulators, experiments involving both these tasks can be created, with the hope of developing a theory that accounts for both modulation depth and modulation frequency discrimination performance. Preliminary investigations indicate that theories based on envelope power spectra do not account performance in these tasks.

PUBLICATIONS AND SUBMITTED MANUSCRIPTS

Biederman, I., & Cooper, E.E. (1992). Size invariance in visual object priming. Journal of Experimental Psychology: Human Perception and Performance, 18, 121-133.

Biederman, I., Cooper, E.E., Fox, P.W., & Mehadevan, R.S. (1992). Unexceptional spatial memory in an exceptional memorist. Journal of Experimental Psychology: Learning, Memory, and Cognition, 654-657.

Biederman, I. & Gerhardstein, P.C., (in press). Recognizing depth-rotated objects: evidence and conditions for 3D viewpoint invariance. Journal of Experimental Psychology: Human Perception and Performance.

Biederman, I., Gerhardstein, P.C., Cooper, E.E., & Nelson, C.A. (in press). High level object recognition without an inferior temporal lobe. Journal of Cognitive Neuroscience.

Cooper, E.E., Biederman, I., & Hummel, J.E. (1992). Metric invariance in object recognition: A review and further evidence. Canadian Journal of Psychology, 46, 191-214.

Edwards, B. & Viemeister, N.F. (1993). Modulation detection and discrimination with

three-component signals. Manuscript submitted for publication.

Fiser, J., Biederman, I., & Cooper, E.E. (1993). To what extent can matching algorithms based on direct outputs of spatial filters account for human shape recognition? Manuscript submitted for publication.

Hummel, J.E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. Psychological Review, 99, 480-517.

Liu, Z., Kersten, D. & Knill, D.C. (1993). Object classification for human and ideal observers. Manuscript submitted for publication.

Madarasmi, S., Kersten, D., & Pong, T.C. (1993). The computation of stereo disparity for transparent and for opaque surfaces. In C. L. Giles, S. J. Hanson, & J. D. Cowan (Ed.), Advances in Neural Information Processing Systems 5, San Mateo, CA: Morgan Kaufmann Publishers.

Madarasmi, S., Pong, T.C., & Kersten, D. (1993). An energy minimization approach to surface segmentation using a multi-layer representation. Manuscript submitted for publication.

Mansfield, J.S., Ahn, S.J., Legge, G.E., & Luebker A. (1993). A new reading-acuity chart for normal and low vision. In Ophthalmic and Visual Optics/Non-invasive Assessment of the Visual System Technical Digest 1993. (Vol. 3, pp. 232-235). Washington, DC: Optical Society of America.

*Viemeister, N.F. & Plack, C.J. (in press). Time analysis. In Yost, W., Popper, A., & Fay, R. (Eds.) Human Psychophysics (Vol. 1). New York: Springer-Verlag (Springer Series in Auditory Research).

Wurm L.H., Legge G.E., Isenberg, L.I. & Luebker A. (in press). Color improves object recognition in normal and low vision. Journal of Experimental Psychology: Human Perception and Performance.

*Review chapter partially supported by AFOSR.

CONFERENCE PRESENTATIONS AND ABSTRACTS

Cooper E.E. & Biederman I. (1993, May). Metric vs. viewpoint-invariant shape differences in visual object recognition. Poster presented at the meeting of the Association for Research in Vision and Ophthalmology, Sarasota, FL.

Edwards, B. & Viemeister, N.F. (1993). Detection and discrimination of mixed

modulation signals at low modulation frequencies. J. Acoust. Soc. Am., 93, 2411.

Liu, Z., Kersten, D., & Knill, D.C. (1992). Object discrimination for human and ideal observers, Investigative Ophthalmology and Visual Science, 33(Suppl.), 825.

Madarasmi, S., Kersten, D., & Pong, T. C. (1993, May). Depth from occlusion using multiple surface representations. Paper presented at the meeting of the Association for Research in Vision and Ophthalmology, Sarasota, FL.

Mamassian, P., & Kersten, D. (1993, May). Surface orientation and illumination direction from shading. Paper presented at the meeting of the Association for Research in Vision and Ophthalmology, Sarasota, FL.

Mansfield, J.S., Akutsu, H. & Legge, G.E. (1992). Interocular contrast differences introduce lateral shifts in the perceived location of binocular depth targets. Investigative Ophthalmology and Visual Science, 33(Suppl.), 1332.

Mansfield, J.S. & Legge, G.E. (1993). Binocular computation of direction and depth. Investigative Ophthalmology and Visual Science, 34(Suppl.), 1053.

Mansfield, J.S. & Legge, G.E. (1993, June). Vernier acuity for targets having different depth. Paper presented at the NATO Workshop on Binocular Stereopsis and Optic Flow, York University, Toronto, Canada.