

**CONSTRUCTION AND EVALUATION OF A THREE-DIMENSIONAL DISPLAY  
FROM A TWO-DIMENSIONAL PROJECTION SURFACE  
BASED ON THEORETICAL CONSIDERATIONS OF  
METRIFICATION AND AFFINE SPACE**

by

**THOMAS D. WASON**

A dissertation submitted to the Graduate Faculty of  
North Carolina State University  
in partial fulfillment of the  
requirements for the Degree of  
Doctor of Philosophy

**PSYCHOLOGY**

Raleigh

1993

**APPROVED BY:**

\_\_\_\_\_  
\_\_\_\_\_

\_\_\_\_\_  
Chair of Advisory Committee

## **BIOGRAPHY**

Thomas Dimock Wason was born in Albany, New York, in 1942. He received his Bachelor of Science degree in Mechanical Engineering from the Massachusetts Institute of Technology in 1964. He received his Master of Science degree in Experimental Psychology from North Carolina State University in 1983. His thesis topic was "Auditory autocorrelation: an experimental study of central nervous system processing of direct and reflected speech sounds." This doctoral dissertation reflects his continued interest in time-structured processing in the central nervous system.

Mr. Wason worked for Texas Instruments in Dallas, Texas, and RCA in Plymouth, Michigan, in optoelectronics research and development. He was a consultant for many years in technology development. For eight years he was president of Allotech, Inc., Raleigh, North Carolina, which developed new concepts in the human-computer interface. Allotech received funding from the National Aeronautics and Space Administration, the National Institutes of Health, the Federal Aviation Administration, and the Department of Transportation. Recently he was Vice President of Development for Cardiovascular Diagnostics, Inc., Research Triangle Park, North Carolina. He lives with his wife, Marianne, and their cat in Raleigh, North Carolina.

## ACKNOWLEDGMENTS

This dissertation has had a long gestation. The patience and contributions of Dr. Donald H. Mershon, Dr. Sharolyn A. Converse, and Dr. Dennis R. Bahler are greatly appreciated. Without their suggestions and insistence on a proper experimental design, the results would have been forever murky.

Dr. James W. Kalat, my academic advisor, has been a continued source of encouragement and inspiration. He has been exceedingly patient. Dr. Kalat's careful editing has been invaluable; he has worked diligently to help me clarify the expression of these complex ideas. Dr. Kalat's broad knowledge of biological processes in psychology were particularly valuable.

I am particularly indebted to Dr. Joseph Lappin of Vanderbilt University. For many years I worked with him in his laboratory on the early stages of development of the affine shear-strain structure from motion display, under funding from NASA, the National Institutes of Health, and the U.S. Department of Transportation. His hospitality, insights, and discussions have had significant impact upon me and my research. It is a rare opportunity to be associated with one so skilled in both theoretical and experimental disciplines.

I am indebted to William C. Troxell of Kalix, Ltd, Glen Burnie, Maryland, who has been most helpful in resolving the complexities of the computer hardware and software. His continued encouragement is appreciated.

This dissertation would never have been started, much less completed, without my wife, Marianne D. Wason. She has encouraged, supported, and assisted me in many ways. She is a fine editor; I have relied upon her a great deal. She is the one who encouraged me to undertake an academic program to pursue my interests, and has stood by me, though thick and thin, all of these years.

## TABLE OF CONTENTS

	Page
TABLE OF CONTENTS .....	iv
LIST OF TABLES .....	vii
LIST OF ILLUSTRATIONS .....	viii
1. INTRODUCTION .....	1
1.1. Needs .....	2
1.2. System Requirements .....	2
1.3. Execution .....	3
1.4. Overview .....	3
2. STRUCTURE-FROM-MOTION .....	5
2.1. Aperture Model of Visual Perception .....	5
2.1.1. Models of the Central Nervous System.....	10
2.1.2. Aperture Principles.....	12
2.1.3. Multiple Subaperture (Array) Systems.....	17
2.1.4. An Aperture Model of the Early Visual System. ....	19
2.1.4.1. Optical Aperture .....	19
2.1.4.2. Neurophysiological Aperture .....	23
2.2. Concepts of Spatial Structure .....	30
2.2.1. The Nature of Euclidean Space .....	31
2.2.1.1. Orthonormal Space.....	32
2.2.1.2. Coupling of Axes.....	32
2.2.2. Affine Space .....	36
2.2.2.1. Decoupled Axes .....	37
2.2.2.2. "Unequal" Axes .....	40
2.2.3. Principle of Affine Equivalence .....	40
2.2.4. Perception of Space .....	49
2.3. Perception of Affine Structure from Motion.....	50
2.3.1. Perception of Motion.....	50
2.3.2. Affine Structure-from-motion .....	51
2.3.3. Spatial Structure .....	57
2.3.3.1. Affine Space Rescaling .....	57
2.3.3.2. Affine Shear-Strain .....	58
2.3.4. Motion .....	58

2.4.	Implementation of an Affine Space Display .....	61
2.4.1.	Multiple Subapertures with Micro-Shear-Strains .....	61
2.4.2.	The Pseudosaccade.....	66
2.4.3.	Gross Offset Affine Shear Strains .....	67
3.	SCALING OF AFFINE SPATIAL STRUTURES.....	68
3.1.	Scaling the Vectorfield.....	68
3.2.	Scaling Hierarchy .....	69
3.2.1.	Sequencing .....	69
3.2.2.	Ordination.....	70
3.2.3.	Cardination .....	71
3.2.4.	Metrification.....	74
3.2.5.	Absolute Scaling.....	75
3.3.	Structure .....	76
3.3.1.	Global Structure .....	76
3.3.2.	Symmetries .....	80
3.4.	Perspective.....	81
3.4.1.	Non-affine Transformation.....	83
3.4.2.	Affine Transformation.....	84
3.4.3.	Semi-affine Transformation .....	84
3.4.4.	Application of Affining Levels .....	85
4.	EXPERIMENTAL METHODS .....	86
4.1.	Experimental Design .....	86
4.2.	Display.....	97
5.	RESULTS.....	102
5.1.	Group Results .....	102
5.2.	Individual Results.....	113
6.	DISCUSSION.....	115
6.1.	Conclusions .....	115
6.2.	The Affine Intermediate in Displays .....	119
6.3.	Topics For Future Research .....	121
7.	REFERENCES .....	123

8. APPENDIX .....	138
Programmed Instructions to Subject .....	138

## LIST OF TABLES

Table	Page
Table 3.1. Scaling hierarchy. ....	71
Table 4.1. Metric level risers and offsets. ....	91
Table 5.1. Responses compared to stimuli presented. ....	103
Table 5.2. Parameter estimates and significance tests. ....	109
Table 5.3. Analysis by metric level. ....	110
Table 5.4. Normalized pyramid image versus scale height. ....	112
Table 5.5. Image variable in ANOVA model. ....	112
Table 5.6. Repeated measures ANOVA. ....	113
Table 5.7. Individual subjects' responses. ....	114

## LIST OF ILLUSTRATIONS

Figure	Page
Figure 1.1. Vector-scalar product. ....	4
Figure 2.1. Simple lens system. ....	13
Figure 2.2. Behavior of two simple optical systems. ....	14
Figure 2.3. Radar beam. ....	17
Figure 2.4. Mask with two subapertures (holes). ....	20
Figure 2.5. Cross-section of binocular aperture. ....	21
Figure 2.6. Single vector. ....	22
Figure 2.7. Vernier offset. ....	27
Figure 2.8. Normal Cartesian coordinates for Euclidean space. ....	33
Figure 2.9. Rotational coupling in Cartesian coordinates. ....	35
Figure 2.10. Necker cube. ....	39
Figure 2.11. Shear strain deformation. ....	40
Figure 2.12. Modular processing model. ....	45
Figure 2.13. Ames room. ....	48
Figure 2.14. Rotation in non-orthogonal space. ....	49
Figure 2.15. Figure with equal height and width. ....	53
Figure 2.16. Filled spaces appear longer than empty ones. ....	54
Figure 2.17. Mid-section of lines are equal. ....	54
Figure 2.18. The Zöllner illusion. ....	55
Figure 2.19. Cubic figure with lines. ....	55
Figure 2.20. Disparity drive decay. ....	60
Figure 2.21. Centerline-based geometry. ....	62
Figure 2.22. Rotation for aperture positions. ....	62

Figure 2.23. Shear strain for aperture positions. ....	64
Figure 2.24. Parallel shifts of lines in depth. ....	64
Figure 2.25. Circle of apertures. ....	66
Figure 3.1. Recursion figure from Lappin and Wason (1991). ....	73
Figure 3.2. Self-scaling "Rat Cage." ....	74
Figure 3.3. Cube of recursing elements. ....	77
Figure 3.4. Dot tower. ....	78
Figure 3.5. Two-level structural hierarchy. ....	79
Figure 3.6. Three-level structural hierarchy. ....	79
Figure 3.7. Perspective depth span relative to vectorfield. ....	83
Figure 4.1. The top view of a pyramid. ....	87
Figure 4.2. A typical display stimulus. ....	88
Figure 4.3. Profiles of the stimulus pyramids in affine space. ....	89
Figure 4.4. The four metric structures. ....	90
Figure 4.5. Metric shear-strain offsets. ....	91
Figure 4.6. A shear-strain shift for producing affine structure. ....	93
Figure 4.7. The same affine pyramid in two affine spaces. ....	94
Figure 4.8. Response scale. ....	96
Figure 4.9. Examples of pixel aliasing. ....	99
Figure 5.1. Mean responses to stimuli with $\pm 1$ std. dev. ....	102
Figure 5.2. Metric and affine means of all responses. ....	105
Figure 5.3. A metric = 0 image. ....	106
Figure 5.4. Response means for each metric level. ....	107
Figure 5.5. Response means for each affine level. ....	108
Figure 6.1. Perception of metric structure. ....	117

## 1. INTRODUCTION

**Space** and **time** are commonly regarded as the **forms** of existence of the real world, **matter** as its **substance**. A definite portion of matter occupies a definite part of space at a definite moment of time. It is in the composite idea of **motion** that these three fundamental conceptions enter into intimate relationship.

Hermann Weyl  
*Space Time Matter* (1921)

The amount and complexity of information we must deal with is constantly accelerating. The general introduction of computing power into the hands of ordinary people has not simplified this problem; it has compounded it. We can now collect and calculate enormous bodies of data that we must somehow assimilate. This information ranges from the esoteric to the ordinary: astrophysics, nuclear physics, fluid dynamics, meteorology, demographics, geographic information systems, systems management, marketing, finances, and entertainment.

Traditionally, graphics have been used to depict information—from cave walls to computer screens. We have outstripped the capabilities of flat displays to present information (e.g., Friedhoff & Benzon, 1989), and are moving into the third dimension in space for display.

The general model for three-dimensional displays is stereoscopic, in which slightly different images are presented to each eye to produce the perception of space. Three-dimensional stereoscopic displays are available in a number of forms (cf., Okoshi, 1976). These typically require special glasses which many users find cumbersome. Knowledge of how we perceive space can be used to generate the perception of space from a two-dimensional surface without glasses. A useful projection surface is a computer-driven cathode ray tube (CRT).

Depth displays are frequently developed by using rules of thumb and trial-and-error to make design decisions. Descriptions of space, and of the visual system that perceives space, can be restructured in order to be understood more readily. Using this restructuring I propose to demonstrate that a useful spatial display can not only be created from a flat screen, but that the perceived depth can be controlled by design, using commercially available equipment.

### 1.1. NEEDS

The accessibility of any technology can be described as follows:

**GOOD**

**FAST**

**CHEAP**

**Pick Any Two.**

This description applies at several levels: development, purchase, and use. The operating target for the system to be developed is that it 1) be relatively inexpensive to purchase, 2) provide adequate but not necessarily photorealistic three-dimensional spatial renderings in close to real time, and 3) be usable by those not skilled in computer graphics.

### 1.2. SYSTEM REQUIREMENTS

The display objective is to develop a process that will operate with:

- a conventional personal computer
- a conventional non-interlaced CRT display (pixel based)
- no special optics or glasses
- a graphics accelerator circuit card.

The model user is a person who sits at a desk or workstation with a keyboard and CRT in front of him/her. The system is useful for normally sighted people in a normal office environ-

ment with low to moderate light levels. It requires no more room than a conventional personal computer and operates in essentially the same environment on a desk top.

### 1.3. EXECUTION

The objective of this study was to develop a working three-dimensional display and to evaluate the possibility of predicting and controlling people's perception of spatial structure, based on theoretical considerations. Given the many possible applications of such a display, there was no optimal display; guidelines for developing a good display for different situations are presented later in this dissertation. These guidelines are based on a reformulation of the description of the early visual system and of the structure of three-dimensional space. The display software was developed by the investigator.

### 1.4. OVERVIEW

This spatial display process is based on the phenomenon of the perception of structure-from-motion. The phenomenon was reported by Wallach and O'Connell (1953) and was given the name "kinetic depth effect" (KDE). This concept has been incorporated into the more generalized expression structure-from-motion (SFM), which includes, in addition to KDE, the stereokinetic effect (SKE, Musatti, 1924), structure from optic flows, and motion parallax. These concepts can be included under a description of spatial geometries and an aperture model of the visual system, which together formulate the three-dimensional display process.

The perception of depth from stereopsis results from binocular parallax. Neurophysiologically, this process converges on some central nervous system (CNS) structures common with those active in structure-from-motion (Nawrot & Blake, 1989, 1991; Tittle & Braunstein, 1993; Cornilleau-Pérès & Droulez, 1993). Parallax (e.g., structure-from-motion) provides an unscaled vectorfield in which the relative vector lengths and signs correspond to sequence in depth for the features in the visual scene. This vectorfield is scaled independently

from the vector generation process, through monocular factors. The product of the vectorfield and the scalarfield produces the perception of spatial structure (Fig. 1.1).

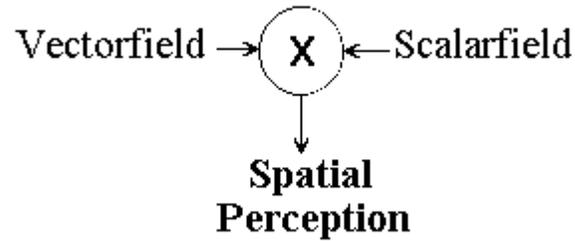


Figure 1.1. Vector-scalar product.

How the vectorfield is developed in the perceptual system through structure-from-motion will be presented in four parts: 1) an aperture model of the brain's visual system; 2) the nature of space itself; 3) perception of affine spatial structures; and 4) the vectorfield display that generates structure from motion.

A goal of this disertation is to develop a means of displaying information on a two-dimensional screen that people will perceive as three-dimensional, and to predict the nature of that perception.

## 2. STRUCTURE-FROM-MOTION

I have divided the discussion of the perception of spatial structure-from-motion parallax into three sections:

1. A model of the visual perceptual system
2. Concepts of spatial structure
3. Implementation of a structure-from-motion display.

### 2.1. APERTURE MODEL OF VISUAL PERCEPTION

I shall present a model of perception based on the scaling of affine structures which will lead to a three-dimensional display. The display creates the perception of spatial structure through the independent generation of affine structure and structural metrics. An affine structure is an unscaled structure. Affine transformations of Euclidean space, and of the structures which lie in it, will be discussed below. The displayed space is mapped into an affine space created by structure-from-motion (SFM) with decoupled axes. Affine, or unscaled, space is discussed below (Section 2.2.). The perception of affine structure is generated through SFM. Structural metrics are generated through ordination, cardination, and metrification. These metric terms will be defined below (Section 3). The hypothesis is that structure and relative scale (metrics) can be independently manipulated under carefully prescribed conditions, demonstrating the independence of the processes of the perception of affine structure and metrification. The experiment independently varies the depth vectorfield and depth scalar, demonstrating that uniform variation of the vectorfield (ratio-ed change over the visual field) produces no change in perceived depth, but that changes in the scalar fields control the perceived relative depth.

The perception of spatial structure employs eight principles, which will be explained in the course of this dissertation:

1. Structures are perceived as affine transformations of Euclidean space ( $E^3$ ) in all directions, not just depth.
2. Stable affine transformations of the same space are all perceptually equivalent, although geometrically different.
3. Orthogonality is perceived through rotational coupling and the inherent coordinates of object structure(s).
4. Affinely perceived structures require metric scaling.
5. There is a hierarchy of affine structure scaling levels: sequence, ordination, cardination, metrification, and absolute scaling. These will be discussed below.
6. Affine structures are subsequently scaled principally through the congruence of recursive elements and/or motion.
7. Perspective can be approximated locally as affine scale and shear-strain transformations of Euclidean space.<sup>1</sup>
8. The same hierarchies of metric structure perception that apply to an object within an affinely defined subspace can also apply between subspaces that are individually locally affine transformations of Euclidean space.

The understanding of visual perception is by no means complete. Visual perception is more than a geometric process in which a scene is re-mapped onto the retina. The concept of reaching out with the eye to grasp the external world begs the question of how visual

---

<sup>1</sup>Locally, perspective space can be modeled as multiplied by a scale factor for distance and a shear-strain deformation to compensate for being off the visual line-of-sight.

perception occurs. Under normal circumstances, the structure and scaling of visually perceived objects is relative to the object, rather than to the observer. How does an object—or a scene—scale itself?

Pre-17th century models of visual perception presumed that the external world was directly grasped by the central nervous system through the sense receptors. It was the Arab scientist Ibn al-Haytham (or Ibn al-Haitham, Latinized to Alhazen, 965-1041) who first proposed that vision was the process of *admitting* light into the eye, not the process of *emitting* something (Gray, 1989, pp. 44-45). A central question of his time was how one is able to perceive something large through an aperture as small as the pupil of the eye. al-Haytham was the first to employ geometrical principles in the study of vision. He was not able to reconcile the inverted image on the retina with veridical erect perception; hence the lens was an attractive organ of sensation. He concluded that the lens was the sensitive organ, and that vision was based on the reinforcement of optical rays striking the lens perpendicular to the lens' curved surface that allowed directionality to light sensitivity. He recognized that if the direction from which light emanated could not be discerned, it would not be possible to perceive the external structure.<sup>2</sup>

Leonardo da Vinci (A. D. 1452-1519) was also unable to reconcile the inverted image problem. He conceived of an eye that maintains the erect image through refraction, projecting it into the (presumed) hollow optic nerve to be carried to the liquid-filled ventricles of the brain, where sensation occurred. da Vinci formulated an eye as a passive *camera obscura* that "piped" the image up into the brain. Some sort of homunculus viewed the image.

It was not until 1583 that Felix Plater (Crombie, 1964, p. 4) proposed that the retina, not the lens, was the photosensitive organ. Kepler ushered in the 17th century and what Crombie

---

<sup>2</sup>The compound eye of the insect solves the problem in a manner not dissimilar from that proposed by al-Haytham.

terms the "mechanistic hypothesis" (Crombie, 1964) with the concept that the eye can be studied as an optical instrument. Kepler followed the processes of visual perception to the point where the image is projected onto the retina, leaving it to the "natural philosophers" to deduce that which was beyond mathematics (i.e., beyond geometry). Geometry relates to images, not to perception.

The transition from a holistic view to a reductionist view was, in the opinion of Crombie (1964), fundamental to the development of the scientific method. In the study of visual perception this transition was embodied as a change from theories of the direct perception of the external world to a commitment to reduce vision and other natural phenomena to mechanistic processes. This change, which occurred at the end of the Renaissance in the 17th century, made science as we know it possible, for it provided an organized structure for investigation wherein one could deduce that which was, and was not, knowable with current methods. Descartes reduced the processes of perception to the extreme, maintaining that in animals there was no perception at all: stimulation produced motion directly. Perception was, if you will, an epiphenomenon occurring only in humans.

We are currently in the midst of blending the reductionist viewpoint with a modified holistic viewpoint. Chaos theory maintains that the behavior of the whole cannot always be accurately predicted from the behavior of the parts. The ensemble may behave in an unexpected way (e.g., Ruelle, 1989; Stewart, 1989). Stewart (1989) argues that until recently we studied only those problems that could be reduced to (linear) parts. The rest were considered "special cases."

This historical sketch puts the current work in visual perception in perspective. Gibson (1950) maintained that one could not deduce the behavior—perception—of the whole from the sum of reduced-case visual phenomena. Perception, he argued, was of the whole. He did not provide an effective working alternative for the analysis of visual perception, however.

The development of an understanding of the processes of visual perception has entailed understanding the underlying neural processes. The retinal image is no longer envisioned as "projected" over the optic nerve back into the central nervous system. The retina performs significant local transformations on the patterns and changes in light. As outlined below, the ganglion cells are effectively overlapping transforming sub-apertures with differing filter or gain characteristics. The image itself is virtually an epiphenomenon; we do not perceive the retinal image. Instead, perception involves coordinated transformations through many subapertures.

Just how the perception of spatial structure occurs is still largely a mystery. As a first step, the external structure is optically mapped onto the retina. This is straightforward. Mapping from one coordinate system to another is conceptually analogous to this piping function. Visual perception does not simply entail processes of re-mappings from an external, 3-dimensional coordinate space onto a 2-dimensional retinal coordinate space. It appears that retinal coordinates are of limited use in the central nervous system (CNS). For instance, we can fuse images on the two retinae even when corresponding points do not fall on homologous parts of the retinae (Burt & Julesz, 1980). This fusion is necessary for the perception of three-dimensional structure. Fusion follows three-dimensional organization. Using a binocular display with disparity to produce depth perception, Green and Odom (1986) found that corresponding points that were matched in depth in two successive images fused in apparent motion. Subjects presented with two image correspondence alternatives that were equidistant retinally formed same-depth correspondence (versus different-depth) 100% of the time. On the basis of retinal coordinates alone, correspondences of targets at the same and different depths were equally likely. Image correspondence was linked to 3-D proximity via a disparity metric in the binocular displays. The visual coordinates are not Euclidean, but affine (stretchy).

The perception of spatial structure of an object becomes a three component process:

1. The perception of affine structure
2. Perceptual scaling of a scene or object relative to itself (this is the perception of metric structure; "metric" will mean relative scale within the object in this discussion)
3. Perception of true size or the absolute scaling of the relatively scaled object.

A coordinate system is an invention, a convenient method of describing a space. There is, of course, no absolute reference system. In the study of perception, one can consider several possible references: the observer's eye, the scene, and the coordinates relative to each object in the scene. A self-referenced, egocentric coordinate system appears under reduced cue conditions. To maintain such a reference system operationally in the real world, visual phenomena such as displays, movies, and photographs require compensating "computations" in order to achieve veridical perception. In spite of many models, the CNS does not appear to be a "computationally driven" system, at least in the most literal sense of the term. It is a system of thresholds, diffusions, and neurochemistry.

#### 2.1.1. Models of the Central Nervous System

The operation of nervous tissue, e.g., neuropile, is generally modeled as either computational or as properties of materials. In a computational model, neuropile is considered to perform algorithms. In a material model the neuropile behaves in a certain manner as a result of its properties. The computational model is currently the predominant one, its explorations attempting to discern the algorithms that underlie the processes of the central nervous system (e.g., Marr, 1982; Poggio & Girosi, 1990). Modeling the behavior of a system with an algorithm does not mean that the system executes an algorithm. The conceptual difference between a computational model and a properties model is exemplified in a planetary system. The planets move in orbits about the sun because of the behavior of masses in motion in gravita-

tional fields. There is no algorithm being performed. The behavior, i.e., orbital motion, can be computationally modelled. Physicists do not presume that their computational model describes a computational operation that produces planetary motion. Similarly, the operation of a lens can be described mathematically, but the lens does not perform a mathematical algorithm. A lens behaves in a certain manner with respect to light based upon the physical properties of the light and lens. CNS properties models are more likely to include biochemistry, diffusion, electrochemical potentials, ions, and glial cells in addition to neurons and synapses (e.g., Nobili, 1987; Xu & Li, 1986).

The difference between computational and properties models is sometimes reduced to semantics. Dictionary definitions of "computational" specifically refer to mathematical processes and algorithms. Mead (1989) proposes a new definition of "computational" to bridge this difference. His definition relates to the behavior of non-linear analog circuits. His models are local mathematical models, i.e., at the level of the neuron. He then develops silicon subcircuits to represent neuronal properties in an attempt to develop ". . . an ideal synthetic medium in which neurobiologists can model organizational principles found in various biological systems" (p. xii). Mead's models are essentially "properties" models of nervous tissue which incorporate the organization of its constituent units.

The value of an aperture model is that it permits algorithmic or computational modelling of processes without mistaking the algorithm for the actual operational behavior of the (nervous) system. Neuropile is an extremely anisotropic and non-linear medium, as is a computer chip. This gross similarity is not to be mistaken for operational similarity. An aperture model is essentially a "properties of a medium" model.

We can consider a visual display as projecting the components of spatial information to the apertures of the visual system. Two types of apertures will be considered:

- 1) the optical aperture synthesized from the two eyes and

2) the neurophysiological aperture.

First, we shall consider the general principles of apertures, and then apply them to the human visual system.

An aperture is a "hole through which information passes." The concept is used in optical systems, radar, and radiotelescopy, among others. Significantly, the concept of an aperture is used in systems that transmit radiant information (e.g., slide projectors), receive information (e.g., cameras) and those that do both (e.g., radar systems). An aperture may gather and rearrange information, and it may filter out information, but, strictly speaking, it does not add information to the flux. I shall review briefly the principles as they relate to the aperture model. First, it is useful to describe the essential components of a simple aperture.

### 2.1.2. Aperture Principles

The general principles of apertures can be demonstrated with a simple idealized thin lens system. The information flux (i.e., light) flows *through* the lens aperture; consequently, the aperture is generally *normal* to the flow of information. Most apertures perform some transformation on the flux transmitted through it. For instance, a convex lens focuses light by bending the light toward the lens axis: the farther from the axis in the aperture the light falls, the more that it bends the light. Another way to describe the transformation performed by a convex lens is to say that it produces a variable phase delay in a wave front falling across the aperture. The phase delay decreases across the aperture from the center (maximum delay) to the edge (minimum delay).

A thin lens system is depicted in Fig. 2.1. At the base of an arrow an object point source of light,  $P_o$ , at a distance  $f_o$  from a lens,  $L$ , projects light to the lens aperture,  $\mathbf{A}$ , which is focused by the lens to an image point,  $P_i$ , which is a distance  $f_i$  behind the lens. The center of the lens is the nodal plane,  $P_n$ , of the lens. The diameter of the lens defines its aperture,  $\mathbf{A}$ , as a

circle of diameter  $D$ . For a thin lens with a focal length of  $f$ , the relationship between focal length and image and object positions is:

$$\frac{1}{f} = \frac{1}{f_i} + \frac{1}{f_o} \quad (\text{Eq. 2.1})$$

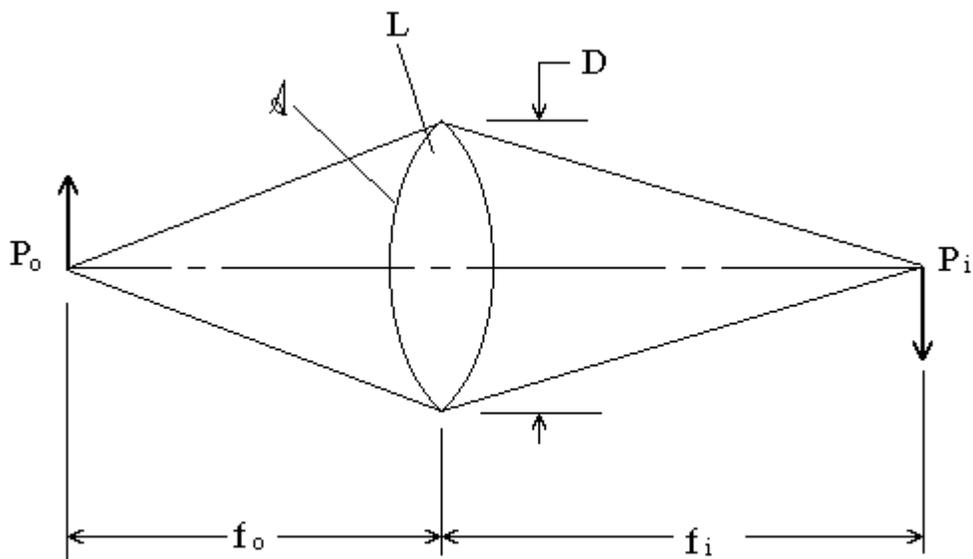


Figure 2.1. Simple lens system.

Now consider two systems each with two object points, one point,  $P_f$ , which the system focuses as an image on a screen, and another,  $P_r$ , behind  $P_f$ , which the system focuses in front of the screen (Fig. 2.2). Of the two systems, the system with the larger aperture (Fig. 2.2a) more sharply focuses the images, producing a sharper image,  $P_f'$ , of  $P_f$  on the screen, while also focusing the image,  $P_r'$ , of  $P_r$  away from the screen, producing a larger

blur circle for it on the screen. Thus we can say that the larger system aperture provides better spatial resolution not only of the points, but also between them in depth. This ability of a system to sharply define sources (e.g., to image a feature, to resolve between two features) is a measure of the qualities of the aperture. One can think of the larger aperture diameter as being able to *resolve* better than the smaller one. This is also reflected in the  $f$ -stops of camera lenses. The  $f$ -stop is the focal length of the lens divided by the diameter. The lower the  $f$ -stop number, the sharper the image in space and the shallower the depth of field.

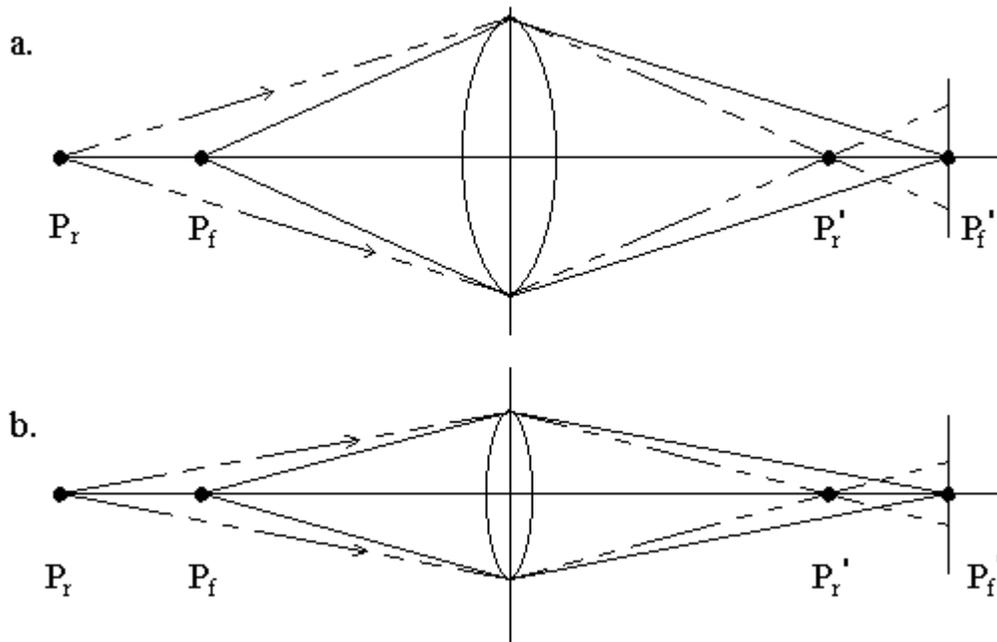


Figure 2.2. Behavior of two simple optical systems.

System a) with a larger aperture, b) smaller aperture.

A lens forms an image by focusing the light falling across its aperture  $\mathbf{A}$  onto a screen. Several things can affect the quality of the image. As mentioned above, the size and shape of the aperture will have direct effects. For high quality optics, the larger the aperture, the better

the image. The light emanating from a point is coherent and in phase, meaning that all the wavefronts are aligned, or synchronous. [This is true even for a point source of "incoherent light"; the coherence path length is just short.] If the light waves arrive in focus with the phase relationship maintained, the image quality will be limited by the diffraction characteristics of the aperture.

*Coherence* is a measure of the degree to which spatially or temporally separated phenomena have a stable known or knowable *phase relationship* with each other. Phenomena can be *coherent* without being *synchronous*. This distinction is significant. A laser operating in the proper mode (TEM<sub>00</sub>) produces photons that are all in phase; thus they are not only coherent, the wavefronts are also *synchronous*—meaning at all points across the waveform, all of the wave points are *simultaneously* of the same phase. For many purposes the output of a single mode laser can be considered as one large photon. The raw, unmodulated, continuous laser beam itself carries virtually no information. A hologram produced with a laser records the variations in the phase relationships of the light reflected from (or transmitted by) an object. The surface configuration of the object *modulates* the wavefront. The wavefronts in the hologram plane are still coherent, but they are no longer synchronous. A great deal of information is carried in the phase relationships between the points in space.

Coherence is a description of the relationships among events separated in space and/or time. The ability to preserve that relationship is the *coherence function* (Gagliardi & Karp, 1976, pp. 419-426). We can speak of an aperture coherence function, a spatial coherence function, and of the coherence function of the medium. Over any given region of space and/or time there is a spatial coherence function describing the relationship of fields, or potential fields, across that space. Frequently a spatial coherence function is made across a plane approximately normal to the direction of propagation of the flux (e.g., light, microwaves). The flux propagates through some medium which can distort and scatter. The coherence function of

a medium is a description of its ability to propagate a spatial coherence function. One medium that is of particular importance is an aperture.

Returning to our thin lens model, an ideal convex lens will transform plane waves falling essentially normal to its aperture,  $\mathbf{A}$ , into converging bowl-shaped wavefronts. If the coherence function of the lens is good, i.e., it has no little bubbles or variations in refractive index, the coherence of the wavefront will be preserved, producing a sharp well-defined image. If the coherence function across the lens is not good, a poor image will result (i.e., the Hubble telescope). The lens can perform transforms other than focusing. For example, it can filter out some wavelengths, or it can have different focal lengths for different wavelengths, causing chromatic aberrations in the image.

An aperture is described in terms of parameters, e.g., geometry, frequencies (spatial and temporal), polarization, integration times, etc. The performance of an aperture is a function of those parameters. An aperture does not have to be round. For example, radar antennae are typically rectangular. The large width provides a high resolution for position of both the outgoing beam and the return reflections in the horizontal plane. The small height provides poor angular resolution for elevation. The projected beam is a vertical fan or plane (Fig. 2.3) that will intersect aircraft over a wide range of elevations, and that will provide good horizontal angular resolution for position. A symmetrical, small beam would have to be scanned both vertically and radially. As this fan-beam is swept only radially, it provides faster searching at the expense of vertical resolution. Other means are used to determine aircraft elevation.

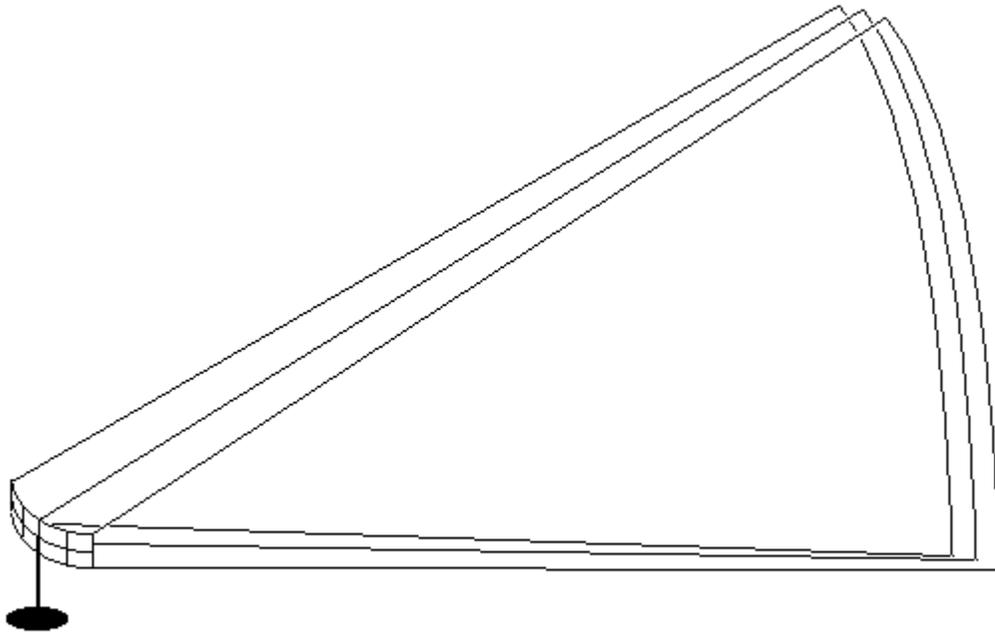


Figure 2.3. Radar beam.

Thus far we have considered apertures of single elements: a lens and a (microwave) reflective radar antenna. These concepts have direct application to the visual system. Apertures also can be synthesized out of smaller subapertures. We turn to them now.

### 2.1.3. Multiple Subaperture (Array) Systems.

A simple lens aperture can be divided in two. If we place a narrow strip of occluding material, such as black tape, across the lens, almost the same image will be formed as with the unoccluded lens. A small amount of light will be lost—the amount occluded by the tape. If instead of using tape, we physically cut the lens in two and set the two halves in exactly the same relationship as they had in the original undivided lens, the light from each half will combine as in the image plane as before. Following this logic, the original aperture, **A**, could be divided into subapertures. This approach has been used in some of the new large-aperture

mirror telescopes, which are made up of a number of sub-mirrors which manufacturers figure (grind) to form a single large mirror. To be effective, these submirrors must preserve the spatial coherence function of the incoming light.

Smaller subapertures may not completely fill the aperture space. The subapertures may be spatially separated to sample the information from different vantage points. The information from the separated subapertures can be combined to *synthesize* a single large aperture. Spaced subapertures can produce fine resolution limits in such *distributed arrays*. Radio telescopes synthesize a large aperture from a collection of smaller radio telescope dishes. Rather than attempt to locate each subdish in some idealized geometry, the position of each dish is compensated for by time- or phase-lagging the signals from each by an appropriate amount to maintain the *aperture coherence function* within desired limits, typically a small fraction of a wavelength. Some radar arrays do not move at all, but synthesize the entire aperture including its movement with a number of phase-adjusted elements (small antennae). These are called "phased-array radars" (Steinberg, 1976).

Multi-element array apertures may not completely fill the total aperture extent, as described. An extreme example is very-long-baseline (VLBL) radiotelescopy, in which there may be thousands of kilometers between the individual radio dishes. For maximum resolution, the effective coherence function across the aperture must be maintained. For some array apertures, the locations of the elements may not be precisely known, or the propagation delays in the cables may be less than ideal or subject to variability. Scientists and engineers use techniques in the design of multiple element aperture systems to establish and maintain the aperture coherence function. Such systems are called adaptive or *self-cohering*. Typically they use signals of known coherence characteristics (Steinberg 1976, pp. 212-252). We will see that such techniques may be valuable as applied to the visual system. Frequently array apertures use an aperiodic or random placement of elements. This reduces artifacts which may result

from a regular, periodic arrangement (Steinberg 1976, pp. 123, 139). Multi-element arrays can have coupling between the array elements, which may be beneficial or detrimental, depending upon the specifics of the situation. Mutual coupling may increase the sensitivity to low level signals at the cost of local phase resolution and subsequent spatial resolution.

#### 2.1.4. An Aperture Model of the Early Visual System.

The most obvious application of the concept of apertures is to the eye itself. It is useful to first apply aperture concepts to a single eye so that the visual system can be more easily understood under this rubric. The aperture and element concepts applied to the eye, and to the receptive fields in the visual cortex, provide a useful context for discussing the perception of image features in motion and under stereopsis. From this, we can define how the display will project onto these apertures to produce the perception of spatial structure.

##### 2.1.4.1. Optical Aperture

The obvious aperture components of the eye are the cornea, lens, and pupil. The cornea and lens form the initial light flux transforming components, and the pupil, formed by the opening in the iris, determines the aperture size of the individual. The optical system produces a nearly diffraction-limited image, meaning that it passes those spatial frequencies that could theoretically be passed by an aperture of the pupil's size.

The two eyes acting together can be considered a single synthesized aperture (Schneider & Moraglia, 1992). As we have discussed, an aperture serves to gather information at a distance. Binocular vision can be considered to have a spatially distributed aperture. Consider the description of the lens that was subdivided into two subapertures by placing a strip of black tape across the center of the lens. Let us extend that tape into a mask that covers the entire lens, but has two small holes near opposite edges (Fig. 2.4). Light emanating from a point ( $P$ ) will fall across the entire lens mask, including the two subapertures. The light will be focused

to a point ( $P'$ ). Either aperture alone will produce a focused image of the point (Fig. 2.5). The diameter of the apertures—the pupils of the eyes—is considerably smaller than the spacing between two eyes (2###). As a result, the ability of either of the two apertures ( $A_1, A_2$ ) alone to localize the image of the point in depth is much less than that of the two apertures together.

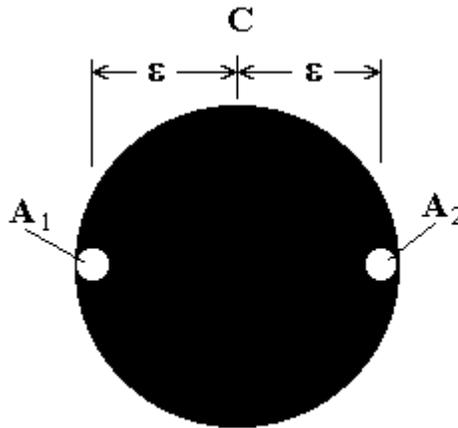


Figure 2.4. Mask with two subapertures (holes).

The two apertures,  $A_1$  and  $A_2$ , spaced at distance ### on either side of the center,  $C$ , of the lens can be considered subapertures of the entire binocular aperture,  $A$  (Fig. 2.5). The depth-resolving power of the pair is significantly greater than of either of the single subapertures if the coherence of the two images (retinal ganglion cell transforms, actually) can be maintained. Consider a system with two apertures that images a central point,  $P$ , flanked by two points at a different depth plane. The projected image of the central point is  $P'$ .

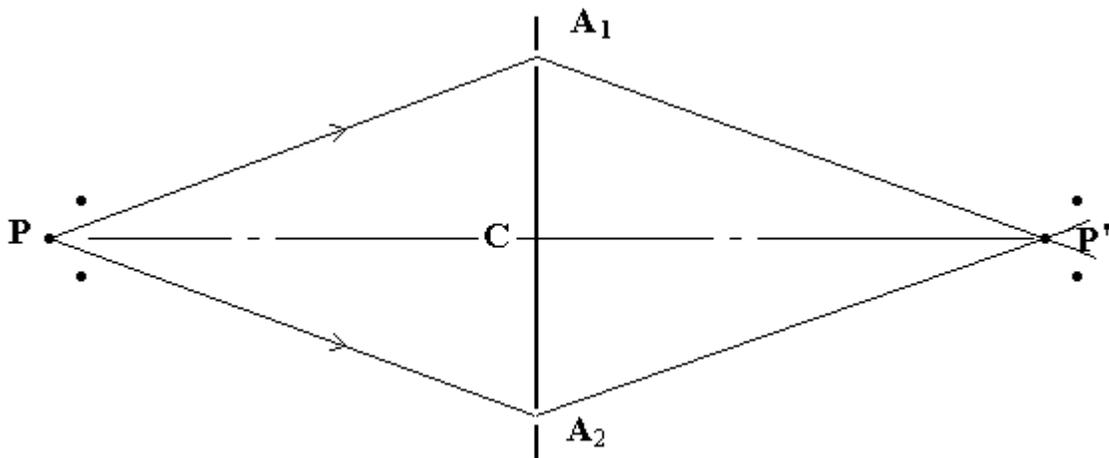


Figure 2.5. Cross-section of binocular aperture.

The binocular aperture,  $A$ , can be used to demonstrate the generation of the depth vector-field,  $V$ . The image of the three points, with  $P$  in the middle, projected from the two subapertures is depicted in Fig. 2.6 that depicts the image planes for each of the apertures separately. The upper trio of dots is projected by subaperture  $A_1$ , the lower by subaperture  $A_2$ . Note that the central point,  $P'$ , has a different position relative to the two flanking points in the two images. The difference between the two locations of the central  $P'$  point relative to the flanking points can be represented as a vector,  $V_{1,2}$ . The length of this vector is proportional to the difference in depth between the two flanking points and the center point.

For this limited group of points, only the relative positions are important. The actual retinal coordinates are of little consequence, as long as the correspondence between the points is maintained. Thus, the pattern of dot spacings needs to be generated in each visual system subaperture. The patterns from the two eyes together must then have a coherence that is maintained by the visual system in order for the correspondences and patterns to be compared to create the local vector. As discussed below (2.1.4.2.), the patterns within the subapertures, and

the subsequent coherence between the patterns from the subapertures, are possibly saccadically and/or motion-initiated coherent changes in the retinal images. The patterns can be modeled from wavelet transformations created by the retino-cortical neural systems. The retinal coordinates are less important than the relationships represented in the local phase relationships.

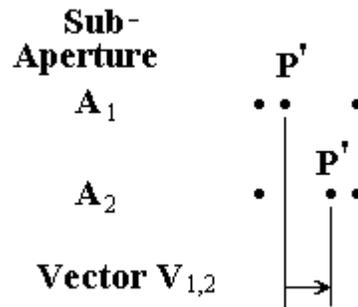


Figure 2.6. Single vector.

Optically, each eye subaperture can be modeled as a lens with a prism, since the base separation is considerably larger than the pupil. Each eye performs a transformation that creates its "image structure" field. The two fields are then overlaid, analogous to projecting the images through prisms so that they overlap with some plane of correspondence between images. Overlaying the two transformed images provides a map of the local differentials. This constitutes a *disparity vectorfield* (hereafter referred to as the *vectorfield,  $V$* ). In the visual system it is not the images that are overlaid, but the transformations of the images through the retino-cortical system apertures into local patterns (e.g., wavelets, as discussed by Mallat, 1991; Mallat & Zhong, 1992).

The disparity vectorfield,  $V$ , can be generated through stereopsis and/or through motion parallax. Instead of transformed images gathered simultaneously through two subapertures,

images can be "acquired" serially in time. The differences between temporally separated images that are due to relative motions between scene and observer can similarly give rise to vectorfields correlated to scene depth spatial structure. Thus, we can consider vision to acquire information through a *spatiotemporal aperture*.

#### 2.1.4.2. Neurophysiological Aperture

The retina is ontologically part of the central nervous system (CNS), since it is developmentally an evagination of the brain. The light flux is projected back onto the retina, which has rods and cones as the primary light-to-ionic activity transducers (photoreceptors). Transducers *transform* flux or activity of one kind into another kind. The rods are sensitive to low light levels. Because all rods have the same spectral response characteristics, they do not discriminate colors. The cones have a lower sensitivity to light and fall into three categories with different spectral responses; therefore, the cone system provides color information at higher illumination. The photoreceptors are not evenly distributed over the retina. The cones are more centrally concentrated. The greater the distance from the fovea, the lower the concentration of cones. The distribution has some radial asymmetry and individual variability (Curcio, Sloan, Packer, Hendrickson, & Kalina, 1987). Thus, the transduction aperture does not have uniform characteristics over its extent. The ionic, hence electrotonic, coupling between the rod and cone systems varies as a function of light level (Uehara, Matthes, Yasumura, & LaVail, 1990; Yang & Wu, 1989; Guth, 1991; Lamb & Pugh, 1990).

The activity from the photoreceptors within the retina modulate the activity of ganglion cells through a matrix of interconnecting cells. Each ganglion cell projects a single axon back within the optic nerve to the lateral geniculate nucleus (LGN). The photoreceptors are subaperture elements for the ganglion cells. There are 100 times more photoreceptors than ganglion cells, indicating convergence. The convergence is not a simple one; there is considerable processing and coupling in the retina (Dowling, 1987, pp. 42-80). One can consider each ganglion

cell as a subaperture of the entire retinal aperture. These subapertures are overlapping, not simply tessellated; thus each photoreceptor will influence more than one ganglion cell. The retinal area, or aperture, which can affect the behavior of a ganglion cell is called its *receptive field*. These are complex subapertures with specific sensitivities to the spatio-temporal characteristics of the light upon the photoreceptors (Dowling, 1987, pp. 33-41, 93-123; Lee, Pokorny, Smith, Martin, & Valberg, 1990; Ratliff, 1965; Shapley & Victor, 1986). Thus, like a lens, the aperture created by a ganglion cell receptive field performs a transform. The greater the distance from the fovea, the larger the receptive field of a ganglion cell. The ganglion cell axons project to the LGN, whose neurons project activity back to the primary visual cortex (Mignard & Malpeli, 1991; Shou & Leventhal, 1989; Gazzaniga, 1989; Maunsell, Nealy, & DePriest, 1990).

Recall that in multi-element electronic arrays, the preservation of phase structure is important. Phase is the timing relative to a particular frequency. Temporal structure in perceptual information processing, and the concomitant importance of phase, is highly important. It is worthwhile to discuss the importance of coherence and synchrony from a perceptual and biological standpoint. Three questions can be raised:

1. Is the time structure of image events maintained in the early visual system?
2. Can differences in the time structure of an image event be perceived?
3. Is the maintenance of time structure important to perception?

The answers to Questions 1 and 3 appear to be Yes. The answer to Question 2 is No. We can show that the time structure of image events is maintained in the early visual system, and that time structure cannot be readily perceived directly, although it does have perceptual consequences.

In response to Question 1—whether temporal structure is maintained in the early visual system—there is good evidence supporting a positive response. Temporal coherence can be

maintained, and there is information encoded in the temporal structures. The early visual system does provide a medium with a high coherence function. At the optic disk, synchrony for image changes as reflected in ganglion cell activity has been disrupted by the trans-retinal traverse time differences. The path lengths of the ganglion cell axons from the somas to the optic disc differ significantly. The propagation times of the neural action potentials (spikes) across the retina to the disc subsequently differ by up to 4 ms. Yet, by the time the optic nerve has terminated in the LGN, the spikes are back in phase to within less than 100 microseconds difference (Stanford, 1987). The resynchronization over the optic tract requires different propagation speeds over different axons, which is possible (Waxman, 1980; Mirsky & Jessen, 1990), and which would be aided by mechanisms that maintain synchronization or phase-lock once it is achieved, such as was demonstrated by Katz and Schmitt (1940) through manipulations of electrotonic coupling between axons. Whether this medium coherence function is developed genetically or through experience is open to question (Kuffler, 1953; Meister, Wong, Baylor, & Shatz, 1991). The frequent saccades of the visual system would provide the type of synchronous stimulus useful in a self-cohering system. Saccades provide simultaneous, retina-wide transients which can be used for establishing and maintaining coherence in the visual pathway (Reynolds & Skinner, 1964).

There is evidence that the retinal-LGN system provides a medium with a coherence function that maintains important perceptual information. McClurkin, Optican, Richmond, and Gwane, (1991) have demonstrated that the temporal structure of information flowing from the retina to the LGN is of significance. They propose that the visual system uses multiplexed temporal codes to carry and process visual information (Gwane, McClurkin, Optican, & Richmond, 1988; McClurkin, Gwane, Richmond, Optican, & Robinson, 1988; Richmond, McClurkin, Gwane, & Optican, 1988). Multiplexed temporal codes is a process in which in-

formation is transformed into time-based codes that are superimposed in time over the same channel.

To Question 2, there are few examples that demonstrate that fine temporal structures can be perceived directly. At a gross level, one can note that with no change in visual image on the retina, perception fades (Kelly, 1969, 1981; Saleh, Tulunay-Kessy, Ver Hoeve, & Hom, 1991). Diamond (1979) demonstrated that the human observer is sensitive to irregularities in flicker patterns. This is a sensitivity to phase within a subaperture, not between subapertures, so it does not completely answer the question. The perception is one of "difference," not of temporal structure. If one considers Diamond's (1979) work in the context of overlapping ganglion subapertures of different response characteristics, it is positive evidence, though weak.

Burr (1979) demonstrated perceptual consequences of temporal structure with vernier acuity. Vernier acuity is a measure of the ability to perceptually resolve a misalignment of two line segments. Due to hyperacuity, for stationary stimuli this resolution is finer than the spacing between photoreceptors (Kulikowski, 1978; Stigmar, 1971; Westheimer, 1975). Vernier acuity is typically demonstrated by presenting subjects with line segments that are not quite aligned (Fig. 2.7). Burr used line segments in apparent motion. The corresponding segments were aligned in space, but one was delayed slightly relative to the other (less than 1 ms). Subjects perceived a vernier offset during apparent motion. This could only be due to temporal structure, as the stimulus had no physical offset in the display.



Figure 2.7. Vernier offset.

The work of Lappin and his colleagues (Lappin & Bell, 1976; Lappin, Wason, & Akutsu, 1987; Mowafy, Blake, & Lappin, 1990) on the perception of correlation between rapidly moving, retinally separated dots indicates that perception is sensitive to temporal structure.

Question 3 follows logically from the second: even if one cannot directly perceive temporal structural differences in the retinal image, does temporal structure have perceptual consequences? The very fact of the perception of structure-from-motion argues strongly for the importance of the preservation, perhaps in some other form, of temporal event relationships within the visual image. Observing this effect does not explain how it occurs, however. Apparent motion of a single dot produced by the sequential short illumination of a line of two or more dots is the perceptual result of temporally structured visual events (Morgan & Thompson, 1975; Morgan, 1976; Hogben & Di Lollo, 1985; Chang & Julesz, 1983a, 1983b; Navon, 1983). Apparent motion is governed by more than the temporal events, however. A Ternus display is a line of dots in which the central dots are always on, and the end dots flash in alternation. People perceive either a central stationary set of dots with a single dot which moves between the ends of the line, or a line that moves laterally one space as a single unit. Petersik and Rosner (1990) could change the perceived apparent motion in a Ternus display by manipulating the context. The perception could be altered from that of a group of dots moving back and forth one dot space to that of a stationary central group with a single dot moving from one end of the group to the other by the visual linkages made from the central dots to other fixed dots. If the links moved to indicate that the entire group moved, that was the perception; if the links were stationary, the central group was stationary. Perception involves all of the visual elements, not a few in isolation (e.g., Dick, Ullman, & Sagi, 1987; Stoner & Albright, 1993). This is consistent with an aperture model that is responsive to the relationship of events across the aperture. Other models combine inputs from modules to achieve the same results. The latter involves considerable computational complexity for complex scenes; as the number

of elements in a scene increase linearly, the number of relationships among elements rises exponentially. Aperture models improve in performance as more correlated information appears across the aperture.

Wilson and Anstis (1969) demonstrated a change in visual delay as a function of illuminance. As an image is comprised of different luminances, it would appear that the timing to the optic disk may be subject to luminance-delay effects that could cause image elements to lose synchronicity. As the retina performs differential transforms, it is more responsive to contrasts than luminance, however. Luminance-caused differential timing delays between the eyes produce the Pulfrich effect (Krekling, 1973; Williams & Lit, 1983), causing a plumb bob swinging in a fronto-parallel plane to appear to follow an elliptical trajectory when a dark glass is placed in front of one eye. Mere delays between the eyes without concurrent motion do not produce such an effect, as there are CRT-based stereoscopic systems which present the images to the eyes in alternation. The dark glass used to produce the Pulfrich effect decreases the image intensity to the entire eye, presumably delaying the entire image. The slight time delay of one image relative to the other with differences between the simultaneous images supports the importance of temporal structure in visual perception. We can safely say that there are perceptual consequences to the temporal structure of events in the retinal image.

The image events on the retina are transformed and passed on to the LGN, where they are subsequently transformed and passed back to the primary visual cortex. Significant work has been done to explore the nature of the relationships between optical activity on the retina and the location and nature of the response in the visual cortex. Pioneering work was done by Hubel and Wiesel (1962, Wiesel & Hubel, 1965) who discovered the receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. Neurons in the visual cortex are responsive to specific stimuli over specific receptive fields on the retina. The cortical neurons can be considered to have apertures that perform specific transforms on patches of the

retinal image, with the transforms distributed throughout the cells of the retina, LGN, and cortex.

As discussed in the context of lenses, transformation apertures can attenuate some information in order to accentuate other information. A lens may filter out certain wavelengths of light to provide better contrast in the image. Cortical receptive fields may be considered in the same manner (e.g., Stork & Wilson, 1990; Robson, 1966; Burr, 1980; Hess, Pointer, & Watt, 1989; Burr, 1980; Harvey, Rentschler, & Weiss, 1985; Snyder, Bossomaier, & Hughes, 1986). Wavelet transform theory (Daubechies, Mallat, & Willsky, 1992; Mallat, 1991; Mallat & Hwang, 1992; Mallat & Zhong, 1992) may provide a useful model of how some receptive field subapertures, each with particular spatiotemporal response characteristics, may be combined. [c.f. Loeb, White, & Merzenich, 1983; Reichart, 1971; Koskol, 1991; Foley, 1991; Goodman & Russell, 1971] There are parallels in the auditory system (e.g., Green, Richards, & Onsan, 1990; Moore, Glasberg, & Schooneveldt, 1990; Henning & Gaskell, 1981). This pattern of repeated sequences of transformations through apertures is frequently found in optical systems.

It is worthwhile at this point to restate the three principal questions relating to the importance of temporal structuring in visual perception, and to summarize the conclusions with respect to structure-from-motion displays:

1. Is the time structure of image events maintained in the early visual system?  
Yes.
2. Can differences in image event time structure be perceived? No.
3. Is the maintenance of time structure important to perception? Yes.

A SFM-based display must have an accurately controlled, well defined temporal structure. The implementation of the three-dimensional display must include careful specification and control of spatio-temporal structures.

## 2.2. CONCEPTS OF SPATIAL STRUCTURE

Since the display is intended to produce spatial perception, let us first discuss the nature of space. This material is quite complex and has been dealt with more completely by others (e.g., Gray 1989; Weyl, 1921). My intent is to provide a common set of definitions and relationships that will be useful later.

Affine spaces are perceptually equivalent.

Euclidean space is an affine space, but not

all affine spaces are Euclidean.

THEREFORE, an affine representation of a

Euclidean space is perceptually

equivalent to the Euclidean space.

We live in an Euclidean space;

THEREFORE, affine representations of that

space are perceived as Euclidean.

The discussion will explain the following concepts:

### 2.2.1. The Nature of Euclidean Space

The normal world in which we live is locally Euclidean space. This is frequently represented in a Cartesian coordinate system having three orthogonal directions with unit vectors of equal lengths in each of the directions. Generally we think of one plane parallel to the ground,

defined by two axes at right angles to each other, and a third axis orthogonal (perpendicular) to the ground. This is purely a convention of convenience, of course, as Cartesian axes need only be at right angles to each other. I mention this "ground based" orientation to make a point: we live on a sphere. The rules of spherical geometry approach those of plane geometry (Cartesian) only when the patch of ground is small compared to the size of the globe.

An important concept in geometry is that of parallel lines. The "problem of parallels" (namely, a proof for them) has been explored extensively over the centuries (Gray, 1989). Euclid's fifth postulate relates to a definition of parallel lines:

If a straight line falling on two straight lines makes the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which are the angles less than the two right angles.

Gray, *Ideas of Space*, 1989, p. 28

Attempts to use this postulate to prove the existence of parallel lines have not been wholly successful; they typically use a "seesaw" proof, wobbling the orientation of one line with respect to the other until only one orientation can be found that does not produce an intersection. (Indeed, some "non-Euclidean" geometries deny the existence of parallel lines.)

The existence of parallels is an important concept in the study of spatial perception. Gray (1989, p. 29) states that "without parallels it is hard to do very much geometry at all, because parallels are needed to transport equal angles about." And, by extension, transporting equal angles about in space also refers to transporting objects about in space without distortion. Underlying the perception of spatial structure is the ability to relate one region to another. This, we shall see, is related to moving objects from one region to another, either directly or by inference.

Thomas Reid in 1764 (Gray, 1989, p. 71) set forth what he purported to be a new geometry, a hemispherical geometry, centered around the observer's eye. His geometry reflected his

interests in vision. In actuality, he was offering a different coordinate system, not an alternative description of space. One must be careful to differentiate between alternative descriptive geometries and alternative spatial manifolds. Einstein, for instance, offered an alternative description of space itself (Weyl, 1921) using a Riemannian geometry.

#### 2.2.1.1. Orthonormal Space

Normal Euclidean vector space ( $E^3$ ) can be described as an *orthonormal* basis of three vectors (Wylie & Barrett, 1982). These vectors are orthogonal (in  $E^3$ , perpendicular), meaning that none of the three vectors can be derived from the addition of components of the other vector(s), and the basis set has equal lengths for each of the unit vectors. This is the normal Cartesian coordinate system.

#### 2.2.1.2. Coupling of Axes

There is an implied coupling among the axes in Euclidean space ( $E^3$ ). The location of any point in space can be unambiguously defined with three values in a Cartesian coordinate system describing  $E^3$ . (Hereafter, unless noted, I will use  $E^3$  to denote Euclidean space described with a Cartesian coordinate system.) There are, however, six degrees of freedom for an object in  $E^3$ : three for position (using some feature to represent the position of the entire object) and three for orientation or rotation relative to the axes.

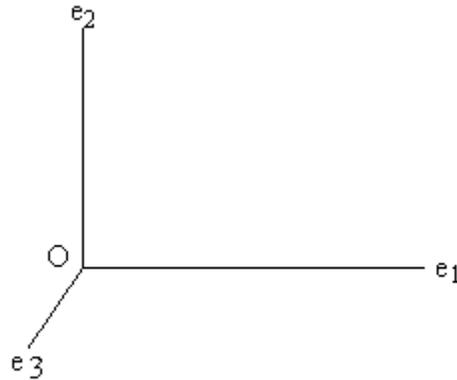


Figure 2.8. Normal Cartesian coordinates for Euclidean space. An orthonormal space.  $O$  is the origin of this coordinate system.

An important concept in mathematics is the **transformation**. A transformation is a mathematical operation that changes a description of some state or process into another form (Bracewell, 1990). For example, the rotation of a rigid object in  $E^3$  is a linear transformation that preserves the absolute values of, and the angles between, vectors (Korn & Korn, 1968, pp. 471-472). This can be considered the "rotational coupling" between pairs of axes.

Consider an  $E^3$  space with axes  $e_1$ ,  $e_2$ , and  $e_3$ .  $e_3$  projects up, out of the plane. A stick of length  $l$  is lying in space with one end anchored at the origin,  $O$ , as illustrated in Fig. 2.9a. The projection of the stick onto axis  $e_2$  is  $l \sin \alpha_1$ . The projection of the stick onto  $e_1$  is  $l \cos \alpha_1$ . Because the stick is lying in the  $e_1 - e_2$  plane, the projection up onto  $e_3$  is zero ( $0$ ). If the stick rotates around the origin in plane  $e_1 - e_2$ , to an angle  $\alpha_2$ , as shown in Fig. 2.9b, the stick now projects onto  $e_1$  an amount  $l \cos \alpha_2$  and onto  $e_2$  by an amount  $l \sin \alpha_2$ . Thus we can say that in a rigid space,  $e_1$  is **coupled** into  $e_2$  through rotation around the  $e_3$  axis. When the stick rotates from  $\alpha_1$  to  $\alpha_2$ , the change in projection along  $e_1$  is  $l (\cos \alpha_2 - \cos \alpha_1)$ , and the

change in projection along (and consequently depth along)  $e_2$  is  $l (\sin \alpha_2 - \sin \alpha_1)$ . When the stick is lying parallel to either axis, instead of at an intermediate angle between the two, the rotational coupling is the smallest. Consider the coupling into  $e_2$ . For a unit length (to make the analysis easier), the projection onto  $e_2$  can be written:

$$e_2 = \sin \alpha . \tag{Eq. 2.2}$$

The first derivative of  $e_2$  will indicate the *rate* at which  $e_2$  changes as a function of small changes in angle  $\alpha$  at any given angle:

$$\frac{de_2}{d\alpha} = \cos \alpha . \tag{Eq. 2.3}$$

Similarly, the rate of change of the  $e_1$  projection for changes in  $\alpha$  is:

$$\frac{de_1}{d\alpha} = -\sin \alpha . \tag{Eq. 2.4}$$

We can see that the effect of the rotational coupling of small changes of  $e_1$  ( $de_1$ ) into changes in  $e_2$  ( $de_2$ ) is not the same for all points of angular rotation ( $\alpha$ ). The changes can be expressed in terms of the effect due to a change in  $\alpha$  ( $d\alpha$ ) at a particular  $\alpha$ . When  $\alpha$  is small ( $\alpha \rightarrow 0$ ) then the rates of change approach as limits:

$$\frac{de_2}{d\alpha} \rightarrow 1 \tag{Eq. 2.5}$$

and

$$\frac{de_1}{d\alpha} \rightarrow 0. \tag{Eq. 2.6}$$

When  $\alpha$  approaches  $90^\circ$  (the stick is parallel with axis  $e_2$ ) then

$$\frac{de_2}{d\alpha} \rightarrow 0 \quad (\text{Eq. 2.7})$$

and

$$\frac{de_1}{d\alpha} \rightarrow -1. \quad (\text{Eq. 2.8})$$

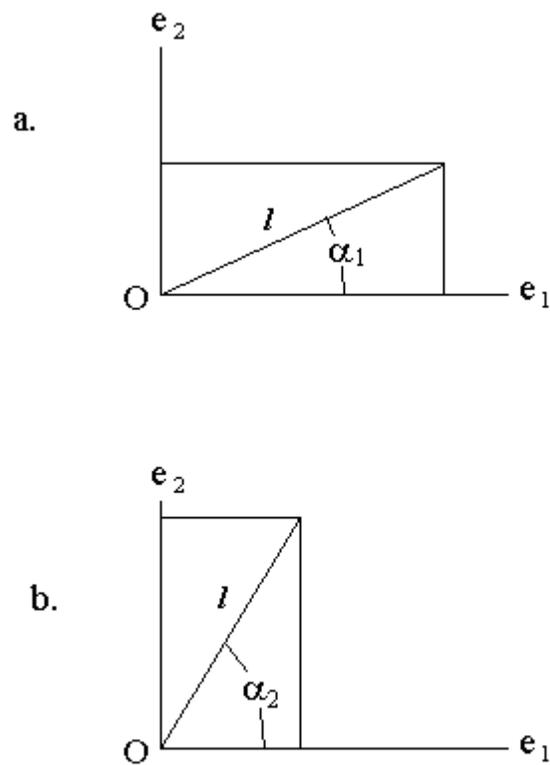


Figure 2.9. Rotational coupling in Cartesian coordinates.

The stick can be translated away from its original position touching the origin (O) to lie on a line parallel with the original in the plane  $e_1 - e_2$ , and this property of coupling through rotations around axes parallel to the  $e_3$  axis will be maintained. [Thus, we see the importance of the concept of *parallel*.] In a rigid orthonormal space, therefore, the axes are *rotationally coupled* with unitary vectors. Proffitt, Rock, Hecht, and Schubert, (1992), made a similar analysis relative to the stereokinetic effect (SKE) versus the kinetic depth effect (KDE) with geometric forms.

A rigid object, with its own Cartesian coordinate system, that is rotating in space will rotationally couple the scaling between axes. Thus, a rotating object could self-scale for the observer, simultaneously defining (in this case, scaling) the space around itself. The rotational coupling between  $e_1$  and  $e_2$  means that as a unit vector rotates in a plane its projections onto  $e_1$  and  $e_2$  change relative to each other. The relative changes in these projections of (above) can be expressed as

$$\frac{de_2}{de_1} = -\tan \alpha. \quad (\text{Eq. 2.9})$$

As a tangent function is nonlinear, this is a nonlinear relationship. This means that the derivation of the metric of a structure in Euclidean space from rotational coupling would be computationally intensive because the coupling is not constant, but is a function of the angle  $\alpha$  (e.g., Hoffman & Bennett, 1986, 1985; Bennett, Hoffman, Nicola, & Prakash, 1989).

### 2.2.2. Affine Space

It is possible to mathematically define spaces which are not orthonormal. Of particular interest are those spaces which are *affine* transformations of Euclidean space.

**Affine** (*adj.*): of, relating to, or being a transformation (as a translation, a rotation, or a uniform stretching) that carries straight lines into straight lines and parallel lines into parallel lines but may alter distance between points and angles between lines (geometry).

*Webster's Ninth New Collegiate Dictionary*, 1987.

Notice that the concept of parallel lines is central to the definition of an affine transformation. A unit vector is a vector that is one unit long in a particular direction. In orthonormal space (Euclidean), all unit vectors,  $u_i$ , are equal: all unit vectors have been *normalized* to the same length. As an example of an affine transformation, each of the unit vectors can be multiplied by a different constant ( $\alpha_1, \alpha_2$ ) to define new unit vectors ( $u'_i$ ):

$$u'_1 = \alpha_1 u_1, \quad (\text{Eq. 2.10})$$

$$u'_2 = \alpha_2 u_2. \quad (\text{Eq. 2.11})$$

In this example,  $u'_1, u'_2$  no longer form an orthonormal *basis*. A *basis* is a minimum vector set which covers, or spans, the space.  $u'_1, u'_2$  form a basis, but are *not normal* (i.e., unit vector lengths are no longer all the same.)

An example of a non-affine transformation is the changing of one unit vector length according to the position along another axis:

$$u'_2 = \alpha(e_1 u_2). \quad (\text{Eq. 2.12})$$

In this case a line plotted parallel to the  $e_1$  axis will not remain parallel to it after the non-affine transformation, but will diverge from it and will intersect the origin (when  $e_1 = 0$ ).

First, we will consider de-coupling between axes rather than the rotational coupling discussed above. Then we will consider cases in which the unitary vectors are not equal.

#### 2.2.2.1. Decoupled Axes

Consider a Nekker cube (Fig. 2.10). It is normally rendered with the front and back faces lying in *frontoparallel* planes. If this were a true orthonormal space, the side lines (edges) would be either invisible or would converge to a slightly smaller rear face (due to perspective). Yet the Nekker cube is invariably perceived as a cube, with perceptual ambiguity about which

face is the front and which is the back. Mathematically this cube can be considered a 3-dimensional parallelepiped in a 3-dimensional vectorfield manifold (Weyl, 1921, p20). We will find the concept of a *vectorfield* to be important in discussions of visual perception. A "3-dimensional vectorfield manifold" is an *affine space*. An affine space is defined by linear dimensions; a unit vector parallel to the  $x$ -axis has the same length no matter where in the space it is located. For our purposes, an affine space can be transformed into a Euclidean space through only affine transformations. Euclidean space is a specific member of the family of affine spaces; all three of its unit vectors are equal and at right angles to each other.

Alternatively, the deformation of the cube can be considered a *shear-strain* (Fig. 2.11) affine transformation of the *space*.

The *shear-strain* component is specified with respect to two axes which are perpendicular in the undeformed body and is designated by the symbol  $\gamma$  with two subscripts to indicate these axes. Shear strain is defined as the tangent of the change in angle between these two originally perpendicular axes.

Crandall & Dahl, 1959, p. 147

The shear strain (Synge & Schild, 1949) transformation of the Necker cube space can be resolved into two components, an  $e_2 - e_1$  component and an  $e_2 - e_3$  component. The  $e_2 - e_1$  component ( $###_{2,1}$ ) can be considered an uncoupled rotation of axis  $e_2$  around  $e_3$  with no rotation of the  $e_1$  axis. Similarly, the  $e_2 - e_3$  component can be considered a rotation of axis  $e_2$  around the  $e_1$  axis. In such a transformed space, a stick lying along the  $e_2$  axis still has no projection onto the  $e_1$  axis. If the space were to undergo repeated or continuous shear-strain transformation(s), the rotation of the  $e_2$  axis in space would not be coupled into the other axes. We can consider such a series of affine transformations as equivalent to a decoupling between axes. The manner in which rotations couple between axes serves to scale the axes relative to each other. Under shear-strain transformations there is no rotational coupling of the shear-strain induced rotation of an axis; therefore, there is no relative scaling among some of the axes. The

object undergoes shear strain with the space; therefore, an object in a shear-strain transforming space, i. e., undergoing a continuous shear strain transformation, may not self-scale in all axes (e.g., DeLucia & Hochberg, 1991).

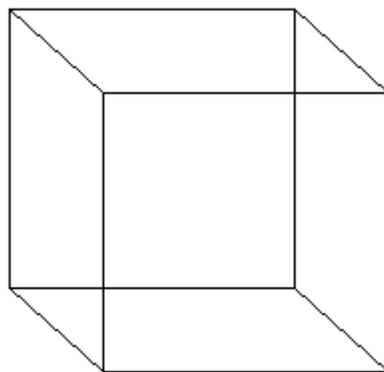


Figure 2.10. Necker cube.

### 2.2.2.2. "Unequal" Axes

Thus far we have considered spaces in which the unitary vectors,  $u_i$ , are of equal length.

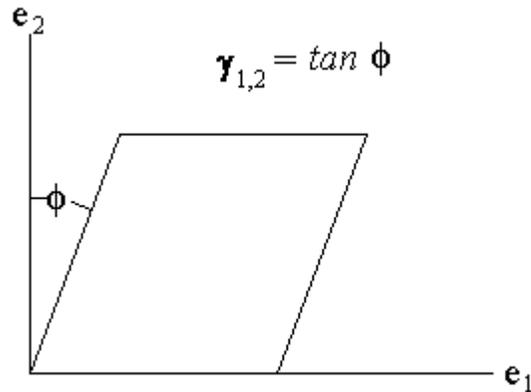


Figure 2.11. Shear strain deformation.

In an affinely transformed or transforming space, this is not necessary. In the shear strain deforming space, the scale in the  $e_2$  direction is uncoupled from the other two. Under an affine transformation, the individual basis vectors may be multiplied by different scalars. Objects are scaled in basis units, thus if  $|e_1| \neq |e_2|$ , the object will rescale as it rotates in the transformed space (Todd & Bressan, 1990).

### 2.2.3. Principle of Affine Equivalence

The nature and motion of objects define the space in which they are perceived. In some cases, motions of an object can be considered as affine (or near-affine) transformations of the space associated with the object. It is parsimonious to propose that: within limits, certain affine (and near-affine) transformations of Euclidean space are perceptually equivalent. These spaces

can be mapped into perceptual equivalence through Lie transformations only of the fronto-parallel image plane (changes in the flat image, e.g., rotation about the line of sight, radial expansion, translation; see Dodwell, 1983).

However, it cannot be said that spaces that are affine transformations of each other are geometrically equivalent. They are not. The significance of this concept is that the perceptual system does not map one affine transformation into another; the affine transformations of the 3-space manifold are perceptually equivalent. In other words, many 3-dimensional spaces are equivalent for the perceptual system, much as the modality from which one gathers information, such as reading or listening, is separate from the information. The principle of the perceptual equivalence of affine transformations of Euclidean space arises from the apparent fact that affine spaces are perceptually equivalent. Since, by our definition, affine spaces are affine transformations of Euclidean space, then all affine spaces are perceptually equivalent to the Euclidean space in which we live. Our experience in Euclidean space makes it the common referent among affine transformations of space.

The principle that affine transformations of a space are perceptually equivalent may not be "intuitively obvious to the casual observer." Sitting in the front row, side aisle, at a motion picture produces a highly distorted image of the scene on the retina and yet the perception is not one of a distorted scene. Experiments by Cutting (1987, 1991) and by Busey, Brady, and Cutting's, (1990) clearly demonstrate the perceptual equivalence of objects in affinely transformed space. Cutting explored the ability of an observer to veridically perceive scenes at a movie when the observer is located at a position not analogous to the camera position. Of particular interest is the ability to perceive rigidity under motion. Through computers, he mathematically projected objects upon a screen at different slants and under differing projections. Use of the computer "projection" method removed any perception of the slant of an actual screen. He generated images under orthographic and polar projections at angles of  $0^\circ$ ,  $22\frac{1}{2}^\circ$ ,

and  $45^\circ$  from normal. The apparent screen angle changed sinusoidally by  $\pm 5^\circ$ . Subjects were normal to the computer display screen. An orthographic projection onto a slanted screen produces an affine transformation. The affine transformation at  $22\frac{1}{2}^\circ$  produces an 8% width decrease  $[(100 \times (1 - \sin(90 - \theta)))]$ ; at  $45^\circ$  the decrease is 29%. A polar projection simulates perspective in one axis. A polar projection onto a slanted screen produces a non-affine transformation; the top and bottom lose parallelism, and the horizontal scale is not constant. The transformations above did not change during each trial. Subjects viewed images of rotating rigid and non-rigid rectangular solids and attempted to discriminate rigid from non-rigid solids. At  $0^\circ$  and  $22\frac{1}{2}^\circ$ , objects in both projections were perceived as rigid. A cube rotating in space is still perceived as a cube. When the screen angle increased to  $45^\circ$ , the polar projection was no longer perceived as rigid, whereas the orthographic projection was. The polar projection non-affine distortion will be slight at  $22\frac{1}{2}^\circ$ , but significant at  $45^\circ$ . The affine (and nearly affine) transformations at  $0^\circ$  and  $22\frac{1}{2}^\circ$ , and at  $45^\circ$  for the orthographic projection, produce equivalent perceptions. The non-affine transformation produced by the polar projection at  $45^\circ$  did not. Norman and Todd (1993) did find that subjects were sensitive to changes in affine stretching in the fronto-parallel plane, but not in depth. Subjects perceived non-rigidity under such changes. The changes were rapid and occurred on a frame-to-frame basis, and are significantly different from the slow changes an observer would experience in normal motion. The rapid picture plane changes undoubtedly interfered with the ordinal and cardinal scaling processes, described below. Thus, the affine equivalency hypothesis is not refuted.

Affine space is scaled in perception. We have an extremely wide range of scale adjustment that spans orders of magnitude. It is remarkable that models of real objects, such as trains, planes, automobiles, molecules, and houses, can be so compelling, given the orders of magnitude of scale error. This (re-)scaling is so ordinary as to be commonly overlooked. Yet this is also what happens when perceiving a distant object: the affine space is scaled. Gross

scale adjustment appears to be a facility of the perceptual system. This is consistent with the perceptual equivalence of affinely transformed spaces. A geometrical description of an image on a screen does not address the most significant aspects of the visual perception of spatial structure.

Coupling with non-unitary axis vectors satisfies the apparent problem raised by the veridical perception of space projected onto a movie screen if one considers such an affine (or near affine) transform of orthonormal space as a member of the perceptually equivalent geometries. Considerable evidence has been developed in recent years that the visual perception of space involves the perception of affine spaces rather than Euclidean space.

Perceptual equivalence means that affine transformations of a space do not affect the perceived affine structure. This apparent tautology can be resolved by restating: perception of structure is not perception of Euclidean structure. An affine structure has no dimensions associated with it. This is difficult to envision, as we perceive real things with real sizes. We do not separate an object from the space it occupies (e.g., Killing, 1892). Therefore, it is useful to discuss affine transformations of the space occupied by the object. If we define that object in terms of its space, and the space undergoes transformations, so does the object.

As discussed above, perception is the product of two spaces: a vector space and a metric space. The vector space is one of relative distances within a direction. One can think of this as a "rubber space" with well-defined rules for its deformation. A simple experiment will demonstrate an affine perception. Hold a nearly featureless book at arm's length on the palm of your hand. The more nearly featureless the book cover (e.g., an old mathematics book), the better the effect. Hold the book with one end directly toward you so that you cannot see either side. Close one eye. Now tilt the book around an axis normal to your line of sight by flexing your wrist. Note that it is difficult to judge the length of the book. You can still perceive that the

book is a parallelepiped, but its length is difficult to judge correctly. The book's *affine* structure is perceived, but not its true *metric* structure.

Underlying this concept of the perception of affine structures is the separateness of the perception of an object and of its place. This is the difference between the "**What**" of an object and the "**Where**" of an object (Sagi & Julesz, 1985).

When we perceive a book, we perceive what it is—its shape, size, and color—irrespective of where it is located with respect to our selves. In a recent review of mechanisms of visual perception, Van Essen, Anderson, and Felleman (1992) have enumerated the subdivision of processes in the early (retinae and LGN) and mid (visual and medial temporal cortices) visual system. They have presented a model of modular processing that is summarized in Fig. 2.12. This is a summary of the modular processing in some 32 visual areas. P represents the parvocellular neurons in the lateral geniculate nucleus (LGN) and their receptive fields in the retina; M represents the magnocellular system. The parvocellular and magnocellular systems are apertures with specific characteristics. Sary, Vogels, and Orban (1993) have also demonstrated the separateness of What and Where by recording responses in the primate inferotemporal cortex to shapes. Cells responsive to a specific geometric shape, for instance a square, were responsive over a large retinal region, and responded about equally to a black square on a white background as to a white square on a black background, or to a square of dots moving across a field of stationary dots. We can conclude that at the level of "Visual Tasks" in Fig. 2.12 that the *What* and *Where* of an object are separated.

This separation of *What* and *Where* is maintained into working memory in the prefrontal cortex. Recent work by Wilson, O'Scalaidhe, and Goldman-Rakic (1993) demonstrates the anatomical separation of the perception of an object and its spatial location. Monkeys were trained to look to the left or right after a 2.5 second delay in response to a particular form presented on a display. The same monkeys were trained to look, after a delay, at the locus of a

stimulus presented to the left or right of a fixation point. The object-based, or form-based, task caused increased activity in the inferior convexity of the prefrontal cortex, but not in the dorso-lateral prefrontal cortex. The reverse was true for the location task, although both tasks used the same forms and required the same responses.

The *place invariance* of object perception requires what Van Essen et al. (1992) refer to

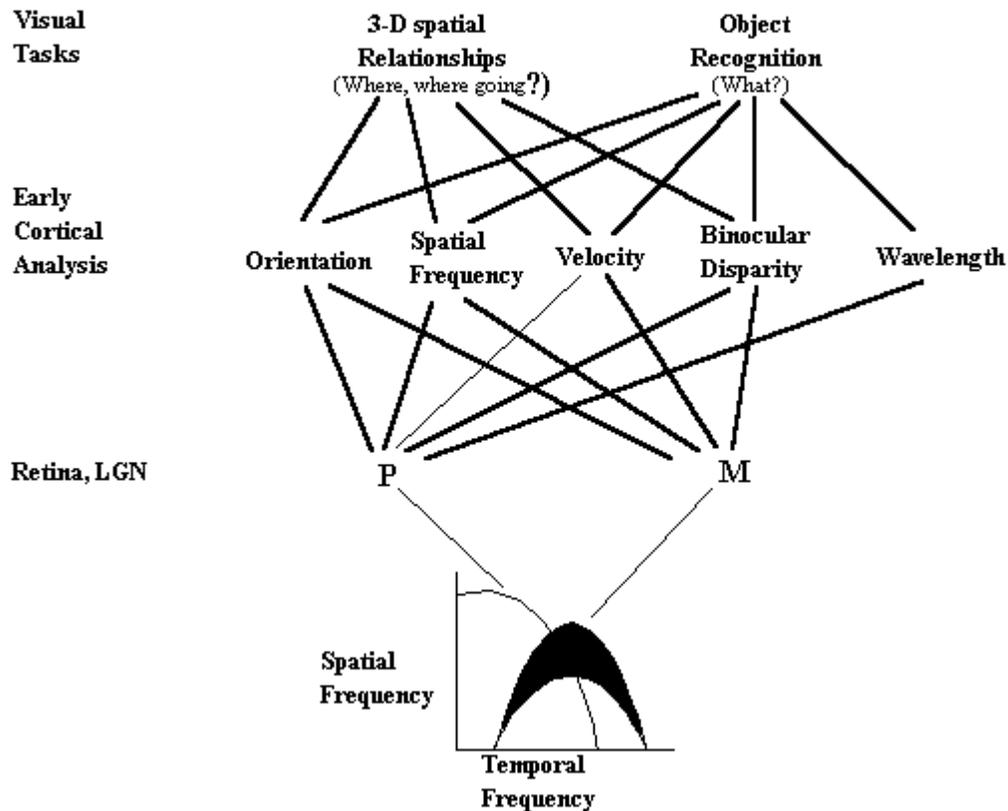


Figure 2.12. Modular processing model.  
Adapted from Van Essen et al., 1992.

as "scale invariance." This *size constancy* of perception means that the object is perceived at its own scale and is not usually scaled by some computational process that adjusts perceived size according to spatial location. This is in contrast to Gogel's (1990) phenomenological approach to the perception of size and structure, whereby the perception of the observer's direction and distance (i.e., location), and motion relative to the object's must be incorporated for veridical perception of the object's metric structure.

Let us return to our book. If we now hold the book at arm's length at an oblique angle, such that we now see a corner formed by three surfaces, we find that changing the tilt of the book does not substantially change the perceived length, the metric structure. This is because the axes of the object are rotationally coupled. This rotational coupling is the equivalent of the element recursion, which is equivalent to moving the element through space to provide scaling, as described above. The rotational coupling is due to the coordinate system inherent in the object. This object "self metrifies." Ames windows and rooms violate self-scaling by providing a false orthogonal coordinate system. An Ames room, depicted below (Fig. 2.13, from Kaufman, 1974) is encountered as a "crazy room" in amusement parks. It is a non-orthogonal room (i.e., walls and ceiling are at non-right angles) that appears orthogonal from one specific vantage point. The relative sizes of people in the room are misperceived. The misperception of these Ames constructs will not survive affine transformations, however, as the apparent parallels will not maintain parallelism. An Ames figure cannot withstand an affine transformation of its space. The inherent rotational coupling of the book, however, is consistent with the true coupling under affine transformations, such as rotation. Rotational coupling serves to delineate the true orthogonality of axes—even if affinely transformed.

Movement of an object in space implies congruence between the spaces it occupies (Killing, 1892). By extension, movement—including rotation—of an object in space serves to provide relative scales between areas of space or between axes in the same locale of space

(Lappin & Wason, 1991). Perceptually, the three-dimensional space will be defined by the way objects move in the space—either by direct movement or by implied rotations due to the symmetries or regularities of the objects. If a point on an object rotating at a constant angular velocity about the intersection of a transformed space (Fig. 2.14) traversed each of the four quadrants (I, II, III, IV) in the same amount of time, one could presume that for a human observer the perception would be of a normal space. This is an example of equivalence of affinely transformed spaces. Therefore, these three-dimensional spaces are perceptually the same. Only by placing one in the context of another would the differences be evident.

This perceptual affine equivalence has significant implications for what is involved in the perception of space and how it is accomplished—and represented—in the central nervous system (CNS). With the equivalency of affine and near-affine transformations, size constancy is one consequence, as local perspective distortions can be considered as local affine transformations. Extended perspective space projected into a display can be considered a non-affine space. Locally, however, one can approximate perspective as an affine transformation of Euclidean space. The off (visual) axis space will include shear strain components. The analysis of perspective space is covered more fully under "Monocular Factors in Spatial Perception: 3.4. Perspective."

Figure 2.13. Ames room.  
From Kaufman, 1974, p. 345.

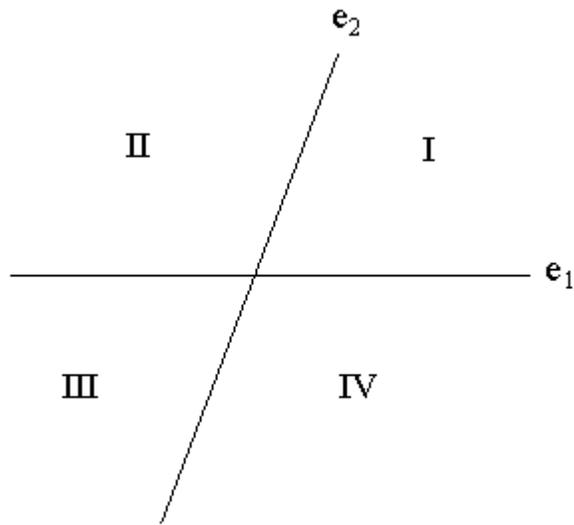


Figure 2.14. Rotation in non-orthogonal space.

#### 2.2.4. Perception of Space

Computer-generated displays permit the development of experimental viewing conditions that can be manipulated to create stimuli which would not be experienced in the normal environment. In the Western tradition, much of this work is reductionist, but there are those who bear in mind that the elucidation of properties of the visual system under minimal conditions may not be a good predictor of the performance of the system in the perceptually rich Gibsonian real world. A discussion of the perception of space must first address the nature of the experimental presentation (e.g., Gibson, 1950). The display uses motion parallax to create the perception of affine structure.

### 2.3. PERCEPTION OF AFFINE STRUCTURE FROM MOTION

Structure-from-motion (SFM) will be used to create an unscaled vectorfield,  $V$ , in the observer's visual system. Kaiser and Proffitt (1992) have demonstrated that such displays are feasible, as has the investigator under previous, unpublished work funded by various federal agencies. An effective display must consider both spatial structure and motion.

#### 2.3.1. Perception of Motion

Most current studies of the perception of motion use apparent motion that is produced by the sequential presentation of a series of dots (light or dark). Hue, brightness, and form are generally preserved under apparent motion (Navon, 1983). Thus most research is performed with sequences of spaced presentations, rather than a continuously moving stimulus. As long as the spacing and timing falls within broad limits, this produces no problems.

Given that apparent motion is equivalent to real motion, one can address the question of the relative motions of separated points. A unified structure moving in space will have all points moving in a coherent, synchronous manner. If one is to detect structure-from-motion, it is evident that the coherence of such motion must be detectable. It is (Lappin & Bell, 1976; Bell & Lappin, 1973). The visual system is responsive to the coherence of motion of separated points. Lie transformations are changes in a planar image. These include radial expansions, rotations, and translations of the image. Lappin et al., (1987) displayed a triangle of three moving dots on a CRT. The dots either jumped under a Lie transform (e.g., radial expansion, translation, rotation) or jumped independently in random directions. With jumps occurring every 20 ms, the coherence or lack of coherence was clearly evident. Others have reported similar results (e.g., Mingolla, Todd, & Norman, 1992; Mowafy et al., 1990). The perception of motion is complex and can involve several processes operating in parallel. All such processes appear sensitive to the coherence of the motion of separated points, as would be required for the perception of a unified structure (Dick et al., 1987; Hogben & Di Lollo, 1985;

Livingston & Hubel, 1988; Stoner & Albright, 1993). Such motion information can produce the perception of spatial structure (Doner, Lappin, & Perfetto, 1984; Eby, 1992; Hoffman & Bennett, 1985, 1986; Kaiser & Proffitt, 1992; Lappin, Doner, & Kottas, 1980; Ono & Steinbach, 1990; Petersik, 1987; Richards, 1985; Saidpour, Braunstein, & Hoffman, 1992; Tittle & Braunstein, 1993; Todd, 1984; Wallach & Centrella, 1990).

### 2.3.2. Affine Structure-from-motion

A brief description of how an affine structure can be defined in a vector space is appropriate at this point. Few objects and scenes are oriented to the observer's "natural" coordinate system of a *picture plane* normal to the line of sight and depth along the line of sight. Real objects and scenes do not have "depth" separated from "width" and "height." To discuss the depth dimension as separate from the other dimensions is therefore arbitrary. Generally objects and scenes structure themselves in our perception. How is this so? Let us return to a basic Cartesian coordinate system. For any point in that space, there are three degrees of freedom. For any object, however, there are six: three of position, three of orientation. If an object is self-structuring, then it is necessary to translate relative dimensions (relative scalars) in one axis into the others. For many objects, this is relatively simple, in that they carry a "natural" coordinate system. For instance, most man-made objects have high degrees of symmetry and orthogonality. As a general case, though, we must "carry" the relative metric from one dimension into the other. Let us suppose that we do not care about the exact orientation of the object's coordinate system relative to our own; all we want to know is the object's structure. Therefore, we are no longer interested in mapping the object's image onto our retina; we are interested in visually mapping the object's separate axes metrics onto each other!

First, we can define features' axis positions relative to other features. This is a *sequencing* operation. I have discussed how this can be achieved as a vectorfield in space, but let us carry that to all directions. Previously we have considered the vectorfield to be a description of

sequence in depth. For a moment, let us revert to an observer-oriented Euclidean coordinate system ( $E^3$ ). As before,  $e_3$  is the line of sight and  $e_1$  and  $e_2$  define a fronto-parallel plane (a picture plane). Each point on the object has three coordinate values, an  $e_1$ , an  $e_2$ , and an  $e_3$ . Consequently there are three vectorfields:  $e_1$ - $e_2$  with  $e_3$ -direction vectors (or  $e_3$  to denote a vector);  $e_2$ - $e_3$  ( $e_1$ -vectors,  $e_1$ ); and  $e_1$ - $e_3$  ( $e_2$ -vectors,  $e_2$ ). For simplicity, I shall denote each vectorfield by its vectors, e.g., the  $e_1$ - $e_2$  field of  $e_3$  is the " $e_3$ -field." Thus each point on the object can be represented with three unscaled vectors:  $e_1$ ,  $e_2$ ,  $e_3$ . Stereo or motion disparity can produce an  $e_3$  sequence of features of which feature is next to which. The  $e_1$ - and  $e_2$ -fields are easy to establish: you can tell where things are in the  $e_1$ - $e_2$  plane if you are sighting down the  $e_3$  axis. This is not to say that the  $e_2$  and  $e_3$  scales are equal relative to each other, but that one can tell the relative sequence of points along the  $e_2$  and  $e_3$  ordinates. (No surprises here.) The  $e_2$ - and  $e_3$ -mapped locations of points can be considered, again, as *sequencing vectorfields*. The vectors are not parallel with the lines of sight, but are normal to it. Thus, we can envision a structure as defined by three vectorfields, ultimately of local differentials. This is, of course, an affinely defined structure. The observer-based vector space was used to simplify the explanation of how the structure of an object is defined in a three-vector space system. This vector structure is central to our use of affined space in a display. Remapping this observer-based three-vectorspace description of the object into an object-referenced three-vector manifold ( $O^3$ ) is straightforward. The point is, we are considering all dimensions, including those in the fronto-parallel plane, to be vectors, not metrics. Additionally, we are not imposing a uniform vector metric over all of space for a given axis, but only for local relationships.

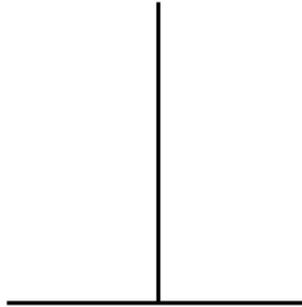


Figure 2.15. Figure with equal height and width.

The relative scales of  $e_2$  and  $e_3$  are not particularly well matched, as might be implied from a spatial-to-retinal mapping scheme. Visual illusions provide hints of the low level of metric precision in the retinal array. For example, the classic inverted "T" (Fig. 2.15) with equal base width and height produces the illusion that it is taller than it is wide.

The partially or completely filled space on the left (Fig. 2.16) appears longer than the open one on the right (Luckiesh, 1922, p. 49).

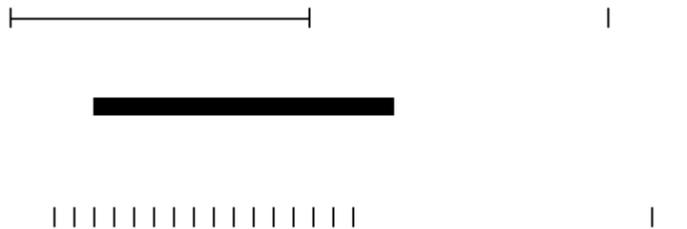


Figure 2.16. Filled spaces appear longer than empty ones.

Although the middle sections of the two lines in Fig. 2.17 are of equal length (Luckiesh, 1922, p. 54), they are generally not perceived as equal.



Figure 2.17. Mid-section of lines are equal.

Similarly, local effects can override global metrics in the retinal image. In the Zöllner illusion (Fig. 2.18), the opposing local slants cause the global parallelism of the vertical lines to be misperceived as non-parallel (Luckiesh, 1922, p77).

The cubic figure (Fig. 2.19) from Luckiesh (1965, p.58) is particularly interesting. The figures on the faces contain oblique and right angle lines intersecting a vertical line. The perceived slant of the surfaces affects the perception of the angles between the lines. The lines that are at right angles on the page are perceived as oblique, and the oblique lines are perceived as at right angles to the vertical. The retinal image is a poor place to do scalar geometry.

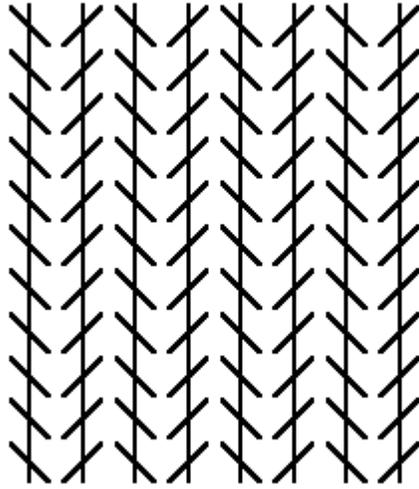


Figure 2.18. The Zöllner illusion.

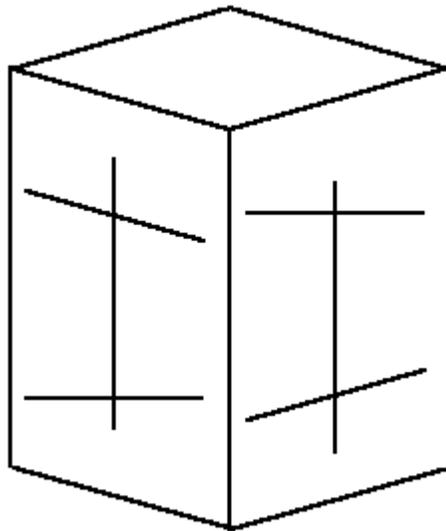


Figure 2.19. Cubic figure with lines.

As outlined above, much of the affine structure of the object is available in the fronto-parallel plane. Motion—or stereoscopic disparity—provides a method for the perception of affine structure. This has been roughly outlined above. Koenderink and van Doorn (1991) provide a lucid conceptual and mathematical demonstration of the ability to recover affine structure-from-motion via apparent motion from two views. This they term the "affine structure-from-motion" theorem. The resulting structure is defined in affine geometry, meaning that the coordinates of the object are subject to scaling and the coordinate system is subject to shear-straining. We shall return to the issue of structuring below.

As discussed above, Cutting (1987, 1991) experimentally supports this conclusion of the perception of spatial structure based on the perception of affine structure with investigations of affine shear transformations and non-affine (perspective) transformations in motion pictures viewed off-axis. He found that the affine transformations produce no change in the rigidity or form of the objects displayed in motion, whereas strong perspective (i.e., non-affine) transformations reduce the perception of rigidity.

It has been widely demonstrated that the structure perceived from motion is not accurately scaled for depth; thus it is the perception of an affine space (e.g., Cutting, 1987; Hildreth, Grzywacz, Adelson & Inada, 1990; Koenderink & Van Doorn, 1991; Lappin & Love, 1992; Lappin, 1990, in press; Proffitt, Rock, Hecht & Schubert, 1992; Proffitt & Kaiser, 1991). As an example, Todd and Bressan (1990) presented subjects with tasks that required either the perception of affine or Euclidean structure-from-motion. As an example of perception of Euclidean structure-from-motion, subjects viewed two bars joined at one end. The structure rotated rigidly in space. The two bars differed in length; subjects reported which bar they perceived as longer. The subjects had to make metric comparisons between two different orientations. This is a Euclidean task, and subjects required differences of approximately 30% to differentiate the lengths reliably. On the other hand, subjects viewed several bars in (SFM) space.

One bar changed length during rotation while the others had constant lengths; hence the changing bar was not a rigid structure. Subjects reported which bar changed in length. Subjects could reliably detect changes of only 3% over 3 frames. This task required only affine perception of space.

### 2.3.3. Spatial Structure

A brief review of the transformations and structure of the space to be displayed are in order. As the vectorfields are largely *local differential vectors*, special consideration must be given to the overall spatial structure of the display. This may require the addition of spatially structuring components or systems to the scene (see Section 3.3.). These components should refer local spatial structures to more global structures through a hierarchy. To the degree possible, affine (rather than non-affine) transformations of space should be used. Mappings of perspective space into the vectorfield can frequently be made through affine approximations coupled with Lie transformations of the image. For instance, a volume of space off the line of sight can be approximated as a shear-strained space with the image translated off axis.

#### 2.3.3.1. Affine Space Rescaling

An affine space has been defined above (Section 2.2.2.). Essentially, in transformations between affine spaces all parallels are maintained, although angles and distances may change. As noted, the normal space we live in, Euclidean space, is an affine space. Since affine spaces are perceptually equivalent, transformations of object space into spatial displays should be limited to the affine or as nearly-affine as possible. An affine space has the same scale at all distances. Perception of space from disparity, whether from SFM or stereopsis, is the perception of affine structures. This is the same as saying that disparity-based perception is the perception of structure in affine spaces. Disparities, and consequently relative depth vectors, have different scales at different distances. A unit vector of disparity resulting from the depth

spacing of two nearby objects represents a greater depth spacing between two distant objects. Mapping of perspective space into the vectorfield is inherently a non-affine transformation as the affine scale is not constant over distance, but there are near-affine approximations that can be used (Section 3.4.). The principal affine transformation other than affine shear-strain will be a uniform (as opposed to perspective) rescaling of the depth ( $e_3$ ) dimension to normalize it to the vectorfield. It is decoupled from rescaling in the  $e_1$  and  $e_2$  axes, which are parallel to the fronto-parallel plane. Although  $e_1$  and  $e_2$  may be scaled differently, particularly to approximate aspects of perspective, shear-strain can provide much of the perspective transformation. As discussed below, Section 3., Affine Scaling, affine space must be locally scaled.

#### 2.3.3.2. Affine Shear-Strain

Moderate shear strain deformations (as described above, Section 2.2.2.1.) can be considered to be perceptually equivalent. Large shear strain deformations of space are permissible as long as the motions of objects are consistent with the space. As such, object motions should be considered to be first plotted in Euclidean space on a frame-by-frame basis, and those coordinates subsequently transformed into shear-strained space. Direct calculation of motion in shear-strained space is subject to error or misinterpretation, potentially producing the perception of a distorting scene.

#### 2.3.4. Motion

Although the depth vectorfield is based on structure-from-motion (SFM), there should be no net translation of observer position relative to the display space. This can be achieved by using a limited range of motion about some nominal fixed position, small enough for no significant translation, small enough for apparent motion, and large enough to create a vectorfield (Chang & Julesz, 1983).

One can presume that an instantaneous lateral shift in the observer's position relative to the display space will provide a certain level of coherent, simultaneous retinal image drive, produced by motion, for the creation of a disparity parallax vectorfield. This instantaneous vectorfield provides a level of drive,  $d_d$ , for the perception of depth. Let us consider that a certain minimal level of **disparity drive**,  $d_{dm}$ , is required to create a perception of depth. The vectorfield is a **differential field** ; the larger the change and/or the faster the change, the larger the disparity drive. An energy decay model is appropriate for the decay of the disparity drive following an abrupt change. Let us assume that the disparity drive level,  $d_d$ , decays exponentially with time:

$$d_d = d_{d0} e^{-t/\tau} , \quad (\text{Eq. 2.13})$$

for which  $d_{d0}$  is the initial drive level at the instant,  $t = 0$ , that the vectorfield is generated, and  $\tau$  is the decay time constant for the drive.  $t$  is the time since the generation of the vectorfield. As long as the disparity drive level is above some threshold,  $d_{dm}$ , the perception of depth is maintained. The values of  $d_{dm}$  and  $\tau$  are probably not constant, but can vary within limits for the observer and the conditions. Thus, as the observer becomes attuned to the display, the size of the shifts in observer position may be decreased, or the time interval between shifts may be increased, reducing computational load and observer awareness of the process. This would reflect either a change in threshold,  $d_{dm}$ , or a change in the decay time constant,  $\tau$ . Changes in ### are more likely. The size of the disparities, and the amount in connected space, together provide a disparity energy level which dissipates exponentially (Fig. 2.20). Thus images with more contiguous spatial complexity will produce a higher level of disparity drive energy, which will take longer to decay down to the disparity drive threshold for spatial parallax. One would expect a richer, fuller vectorfield to produce a stronger drive than a depleted one (i.e., if there is more stuff in the visual field, the perception of depth will last longer).

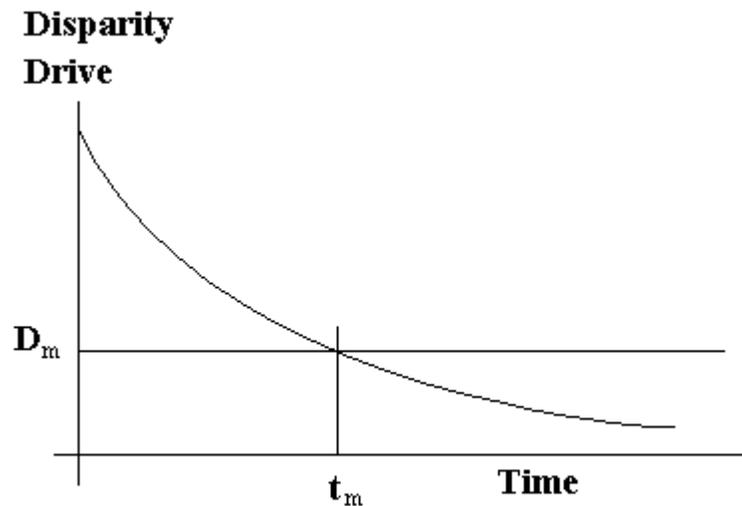


Figure 2.20. Disparity drive decay.

The vectorfield is scaled independently from its actual amplitude, as it is scaled relative to itself (self-normalized); consequently, one would expect the perception of spatial depth to be maintained continuously until it abruptly ceases. This is a direct result of the perceptual equivalence of affine depth rescaling.

The exponential decay of the *instantaneous vectorfield* may be the explanation for the findings of Todd and Bressan (1990), Braunstein, Hoffman, Shapiro, Andersen, and Bennett, (1987), and Braunstein, Hoffman, and Pollick, (1990) that there is little improvement in the perception of spatial structure from using more than two frames in apparent motion. Each new vectorfield "overwrites" the previous one, as the physical location in the CNS is the same (Hogben & Di Lollo, 1985). Since scientists consistently use motion that is either linear or circular (i.e., the instantaneous vectorfields are nearly the same), there may be some small residual effect from the summation of decaying successive vectorfields that results in the small (3%) but consistent improvement in depth when more than two frames are used.

Motion which has a consistent pattern will eventually produce predictive eye movements. As the motions are small, the predictive movements can be expected to overshoot the actual changes, resulting in an increased perception of the movement. A random motion (in direction and time) to the movements should minimize this effect. Now let us turn to the proposed implementation of these concepts.

#### 2.4. IMPLEMENTATION OF AN AFFINE SPACE DISPLAY

Implementation of the structure-from-motion (SFM) display has several facets. These include subapertures, pseudosaccades, and gross offset affine shear-strains.

Since the SFM vectorfield is unscaled for depth, display contents should always span the maximum possible vector range to make full use of the affine space, which will be created by a small affine shift. Use of less than the full affine depth reduces the potential spatial resolution. This reduces the effectiveness of the display. The vectorfield will be scaled independently. This may be counterintuitive. The contents of a display should be normalized to the maximum extents of the affine space. The affining level should be set to suit the worst case depth span expected (greatest  $z_s$ ). Methods developed for scaling the SFM vectorfields can also be applied to stereoscopic displays.

##### 2.4.1. Multiple Subapertures with Micro-Shear-Strains

In Section 2.1.4.1., "Optical Aperture," a subaperture model of the binocular visual system was presented (Figs. 2.4, 2.5). Each eye was shown to constitute a subaperture of a potential circular aperture. We can expand the selection of subapertures from two to as many as we wish within a defined aperture, and select them in some defined sequence. The orientation of space with respect to these subapertures can be shown. We shall first describe the space relative to the two binocular subapertures—the eyes—as shown in Fig. 2.21. This depicts the  $e_1$ - $e_3$  plane.

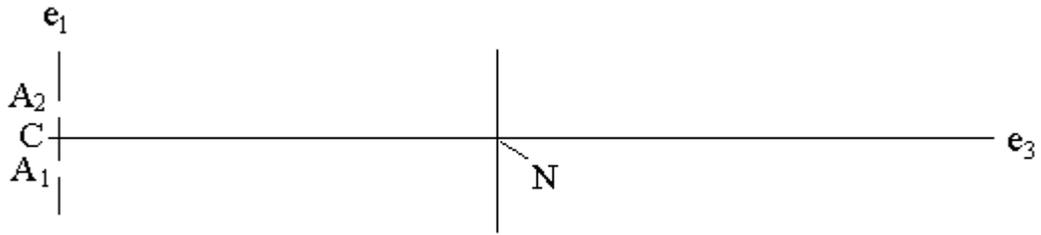


Figure 2.21. Centerline-based geometry.

$A_1$  and  $A_2$  represent the two subapertures of the two eyes depicted above in Figs. 2.4 and 2.5.  $C$  is the center of the aperture.  $A_1$  and  $A_2$  lie at either side of the center.  $e_3$  projects out in front of the observer.  $N$  is the nodal point, the point at which the axes of the two subapertures (i.e., eyes) converge. The  $e_3$  axis passes through  $C$  and the nodal point. A nodal plane is shown passing through the nodal point perpendicular to  $e_3$ .

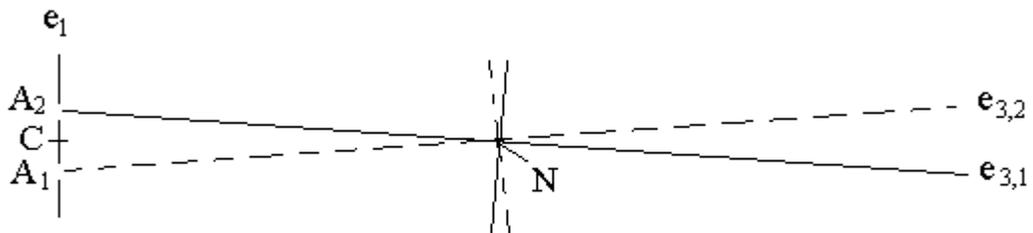


Figure 2.22. Rotation for aperture positions.

One can consider each subaperture to act as the centerpoint ( $C$ ) when it is individually accessed. In other words, if  $C$  is the Cyclopean eye (Julesz, 1971),  $A_1$  and  $A_2$  can be considered two views from the common point. Thus we can consider a stationary viewpoint with a

space which rotates about the nodal point,  $N$ , as the equivalent from an aperture standpoint (Fig. 2.22). When the space is viewed from subaperture  $A_1$ , it is equivalent to rotating  $e_3$  about the nodal point,  $N$ , to orientation  $e_{3,1}$ . Similarly, space viewed from  $A_2$  is equivalent to viewing from  $C$  and rotating space such that  $e_3$  has the orientation  $e_{3,2}$ . Note that the normals through  $N$  rotate accordingly. The local differentials between the two images constitute the local disparities, be they binocular or apparent motion.

Instead of rotating  $e_1$ - $e_3$  space about  $N$ , the orientation of the nodal plane relative to  $C$  (the line  $C - N$ ) can be held constant by subjecting the space to a small shear strain of  $e_3$  relative to  $e_1$ , as shown in Fig. 2.23. The nodal plane  $N_p$  now has a constant orientation relative to the observer. As the shear strain is an affine transformation, there will be no perceived distortion of space for modest transformations. The actual aperture shift will be small, and thus the shear-strain will be small. Larger shear-strains can be introduced for other purposes (see Section 2.4.3. below). The shear-strain transformation is more evident in Fig. 2.24, in which parallels to  $e_{3,1}$  and  $e_{3,2}$  have been included.

The local differential between the views of space from the two aperture positions, whether Euclidean or affinely transformed, generates the vectorfield,  $V$ . The disparity vector between the two aperture views is a linear function of the distance in the  $e_3$  direction from the nodal point,  $N$  (Euclidean), or nodal plane,  $N_p$  (affine). The vector sign is a function of whether the point is in front of or behind the nodal plane. It is readily apparent that the magnitude of shift of a point resulting from an aperture shift will be linearly related to that point's distance from the nodal plane. This shift between aperture positions provides the differentials that form the depth vectorfield.

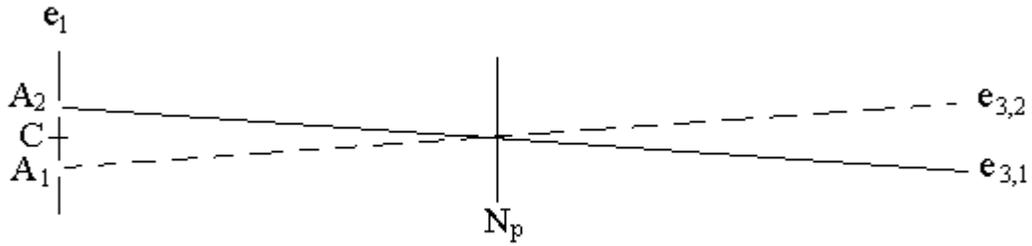


Figure 2.23. Shear strain for aperture positions.

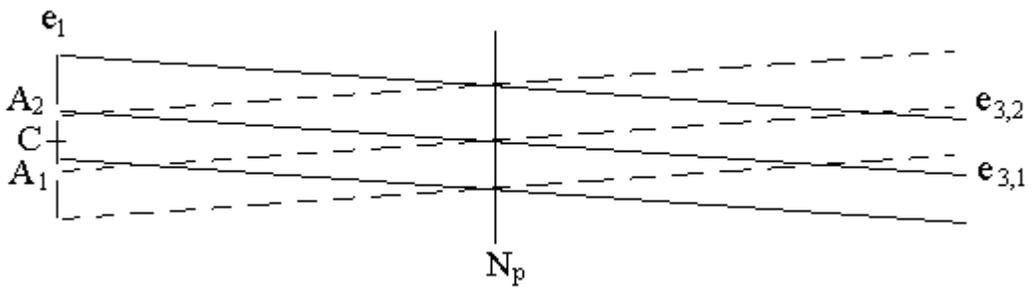


Figure 2.24. Parallel shifts of lines in depth.

At this point we are in danger of being overwhelmed by nomenclature. To simplify matters, standard terms will be adopted. As the depth of the vectorfield is unscaled, shear-strain angles (###) have no meaning. Therefore, the plane of subaperture points will also serve to define the shear strains, although not in angles as with the normal convention. The null (0,0) aperture micro shear strain is thus the centerpoint,  $C$ , which is coincident with the center (defined as  $e_1$  &  $e_2 = 0$ ) of the nodal plane, the nodal point ( $N$ ), when  $e_3$  is normal to the nodal plane. The aperture position will be defined as a displacement relative to the center ( $C$ ) as a

fraction of the width ( $W$ ) of the screen at the front surface of the screen, which is the front of the vectorfield. By this definition, the (sub)aperture position,  $A_p$ , and the shear strain are synonymous, even though they could be considered  $180^\circ$  out of phase. (### = eccentricity of aperture positions relative to  $C$ .)

Thus far we have considered subaperture positions that lie at either edge along a horizontal diameter of a round (super) aperture as in Fig. 2.4. This is analogous to a binocular arrangement. Actually, subapertures can be located anywhere within the superaperture (henceforth, the term "aperture" refers to a subaperture,  $A_p$ ). One possible arrangement is in a line along the diameter. If these apertures are presented in sequence, with no accompanying shear strain, we have the equivalent of the kinetic depth effect (KDE) described by Wallach and O'Connell (1953). The linear sequence with the accompanying shear strain is the same as the alternating linear stereokinetic effect (SKE) described by Proffitt et al. (1992). Defining the superaperture as a line extending continuously by rotating around the nodal point in a circle or part of a circle describes a more extreme kinetic depth effect. Translation along a line, a continuously moving aperture, with subsequent equal translation of the nodal point, produces normal motion parallax, which is a KDE with a radius of infinity.

Apertures could be located around the rim of the superaperture, forming a circle (Fig. 2.25). If presented in a continuous circular sequence, e.g.,  $A_1, A_2, A_3, \dots$  with corresponding shear strains, the stereokinetic effect (SKE) will be produced (Musatti, 1924). Thus, it is apparent that selecting the appropriate aperture positions, with or without relevant affine shear strains, can describe the conventional SFM effects. This same system can be used to define an SFM display.

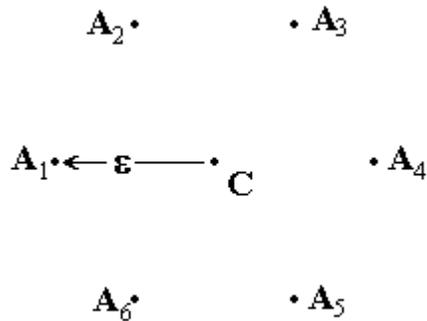


Figure 2.25. Circle of apertures.

#### 2.4.2. The Pseudosaccade

An observer can become fatigued by watching a continually oscillating or rotating SFM display. Presumably this is due to the generation of predictive eye movements (Bittencourt, Smith, Lloyd, & Richens, 1982; Kelly, 1990). It has been assumed here that the disparity drive ( $d_d$ ) is a function of the size of the differential and decays exponentially with time (Eq. 2.13). The circular SKE has disparity shifts of equal amplitude spaced at equal times, but this need not be the case. It is possible to make an unpredictable display shift pattern.

As discussed above, saccades can provide the retinal events that facilitate maintenance of the coherence within and between the retino-cortical apertures. Also, saccades do not normally occur at regular intervals, but are quasi-random (Mates, 1978). A good model for such randomness is a Poisson distribution. Saccades also do not normally occur with a regular spatial pattern. By selecting aperture positions quasi-randomly from a circular pattern, shifts of different perceptually random lengths and directions will occur. The differing lengths will produce differing drive levels, which will allow differing times before decaying down to the

drive threshold ( $d_{dm}$ ), so differing times between shifts can be used. These related but variable aperture shift lengths, directions, and times can be used to produce a pseudo-random sequence of "pseudosaccades." A true saccade results in a change in the position of the image on the retina, but no change in the structure within the image itself (a Lie transformation with uniform, continuous differentials between images). The motion is synchronous across the entire retina. Motion parallax results in a structural change within the image. The pseudosaccade is a combination of the two, using the abrupt, synchronous transretinal change to provide coherence to the structural image changes that provide disparities. It accesses the depth-from-disparity transformation capabilities of the CNS through a non-disparity mechanism (i.e., saccadic abrupt synchronous image changes).

#### 2.4.3. Gross Offset Affine Shear Strains

The SFM aperture shifts will be accompanied by affine shear strain shifts such that the nodal plane will remain parallel to the aperture plane, which is the surface of the display. Large stationary shifts can also be imposed as needed to provide static coupling through the structure of the display (such as the vertical and horizontal offsets of the Necker cube, Fig. 2.10). These shear strains can be measured with the same system used to specify the aperture shifts, based on  $C$ . Such affine shifts could be up to 20% of the screen width ( $W$ ), and can provide a method for producing effective modest rotations of the display space. As these would be shifts of the vectorfield, the actual perceived angle would be a function of depth scaling.

### 3. SCALING OF AFFINE SPATIAL STRUCTURES

#### 3.1. SCALING THE VECTORFIELD

The process of visual perception of spatial structure requires the scalar field-vectorfield product ( $\mathbf{S} \times \mathbf{V}$ ). Monocular factors develop the scalar field,  $\mathbf{S}$ , which scales the vectorfield,  $\mathbf{V}$ . In the absence of a scalar field, there are "system defaults," e.g., the specific distance tendency and the equidistance tendency (Gogel, 1965, 1977), which serve as the scalar field,  $\mathbf{S}$ . For instance, the equidistance tendency provides a local scale that is based on the proximities of the image components without adequate depth information necessary to spatially separate image components. Similarly, in the absence of an adequate vectorfield, a strong scalar field can produce the perception of depth, due to the  $\mathbf{S} \times \mathbf{V}$  product. The visual system attempts to remove ambiguity (e.g., one cannot perceive both states of a Necker cube simultaneously). This display depends on the interaction of the two components ( $\mathbf{S}$ ,  $\mathbf{V}$ ) in the generation of a perceptual 3-manifold. The scalar field,  $\mathbf{S}$ , does not need to contain a scaling value for each point in the vectorfield,  $\mathbf{V}$ . If the spatial structure adequately integrates the space (as discussed below, Section 3.3.), then a relatively small number of points needs to be scaled, since the remainder of the vector points will be scaled ratiometrically.

As discussed above (e.g., Nawrot & Blake, 1989, 1991), the SFM and binocular stereopsis processes converge. Both produce vectorfields, and both have a limited range of disparities over which they can produce vectorfields. The disparities range from a minimum threshold ( $d_{\min}$ ) to a maximum ( $d_{\max}$ ) before loss of correspondence between image features in motion (Wehrhahn & Rapf, 1992; Chang & Julesz, 1983a, b) or stereopsis (Julesz, 1971). The techniques of monocular scaling presented below can be used in either type of display. For clarity in this paper, specific reference is made only to SFM vectorfields.

Accommodation and convergence are not significant cues in a visually rich environment, and they are not of concern in well-designed displays. Image focus may be of concern,

however, as it is determined by the display. Spatial frequency manipulations in the display can direct the attention of the observer to a particular depth.

### 3.2. SCALING HIERARCHY

How is an affine structure scaled? How are the axes of the object scaled relative to each other? How are they scaled absolutely? To address these issues we can create a scalar hierarchy:

1. Sequencing
2. Ordination
3. Cardination
4. Metrification
5. Absolute Scaling.

Each level is more specific and requires the previous level(s) in order to operate. The roles of generating factors in the scaling hierarchy can be arranged as in Table 3.1.

We can organize a discussion of this matrix according to scaling levels or scaling factors. I shall briefly explain each scaling level and how it is generated by the relevant scaling factors. Some factors provide more than one level of scaling.

#### 3.2.1. Sequencing

*Sequencing* is simply a matter of establishing order without sign: which points are next to which. The relative and absolute spacings are not necessarily well specified. The vectors from structure-from-motion (SFM) and from the static fronto-parallel plane projections provide more than order; local relative spacing information is available in terms of the local relative vector lengths. Unless provided in the structure, global relative spacings are not available. Recall that affine structures are unscaled in all three dimensions. Sequence does not inherently define order, e.g., which point is nearer the observer. Obviously the axes projecting onto the picture

plane have a left-right specified. An ordinal sign change applied to the vectors reverses the perception.<sup>3</sup> In structure-from-motion analyses (e.g., Bennett et al., 1989), this reversal problem arises frequently.

### 3.2.2. Ordination

The *ordinal* nature of the sequence is an ordered ranking (e.g., from low to high, front to back, left to right, first, second, third . . . ) supplied from several sources, less intensity or contrast with depth, texture gradients, element size (which is a larger scale texture gradient that I shall return to), interposition, and linear perspective. These give weak dimensional scaling but serve well to remove ambiguity about depth sequence. An ordinal ranking does not specify a scale or relative distance, merely order.

The ordinal scaling can be local. A mask viewed from the inside at a modest distance, gently rotated, will often appear to be a face in relief, instead of the true itaglia (carved into the surface, the opposite of bas relief) (Klopfer, 1991). The mask is viewed stereoscopically in a normal environment. The rest of the environment does not reverse in depth at the same time. In fact, as discussed below (Section 3.3.), one of the tasks in 3-dimensional display design is the integration of all local structures into a gross structure. This is done through a hierarchical structure.

---

<sup>3</sup>The separateness of *sequence* from *order* may be reflected in the ease of perception of mirror images and symmetry. This also argues for the sequence aspect of vectors in the fronto-parallel plane in the CNS, as they can be separately reverse-ordered as fields.

Table 3.1. Scaling hierarchy.

Scaling Factors	Scaling Level				
	Sequence	Ordination	Cardination	Metrification	Absolute Scaling
Disparity Vectors	S		V		
Interposition	S	S			
Perspective	W	M			
Intensity	M	S			
Texture Gradient		M			
Element Recursion		M	S	S	
Isometric Motion			V	S	
Familiar Size	V	M	W	V	S
Ref. to Familiar Size		W			M

S = Strong

W = Weak

M = Moderate

V = Variable

Blank = Normally not applicable

### 3.2.3. Cardination

*Cardinal* scaling establishes the relative distance between points, not simply sequence and order. Cardinal metrics are relative scales within an axis or orientation, not between. One can consider it a "rubber ruler." One can stretch it to span the space required, and the relative lengths of the intervals between points will maintain a constant ratio. The vectors of the vectorfield provide local cardination without ordination, or sign. Element size also provides cardination; often cardination and structure metrification are served by the same factors. We could take a single element and move it over the object as a little ruler, providing a metric for

the structure. Killing (1892) used *congruence* to describe both cardination and metrification, in which space and objects are intimately related:

Every object covers a space at every time. The space covered by one object cannot simultaneously be covered by another object.

Every object can be moved. If an object covers the space of a second object at any time, then the first object can cover the space covered by the second object at any (other) time.

Every space (object) can be partitioned. Each part of a space (object) is again a space. If A is part of B and B is part of C, then A is part of C, where A, B, and C may be either spaces or objects.

This model of movement of an element does not often fit the real world, however. As discussed in Lappin and Wason (1991), illustrated in Figure 3.1 below, the multiple presence of the same or similar elements within the scene can provide translational symmetry through *recursion* of the same or nearly the same elements. In the real world these may be bricks in a facade, shingles on a roof, siding on a house, leaves on a tree, etc., which serve as natural metrics. In technological displays (e.g., mechanical engineering stress-strain displays), the finite elements can serve to provide the cardinal metrics. The elements described above can be perceived clearly. The elements that make up texture are difficult to perceive individually, serving primarily in the spatial frequency domain.

In the real world the boundary between *cardination* within an axis and *metrification*, relative scaling among axes, is not sharp. Analogous to moving a single element over the object, or tessellating the object with uniform tiles, one can move the rigid object. Rotation is angularly the same throughout a rigid structure. Translation is the same throughout a rigid structure. As

discussed previously, the coherence of the motion is critical. Motion, then, can serve as a recursive element, providing a basis for metrics of the structure.

Figure 3.1. Recursion figure from Lappin and Wason (1991).

#### 3.2.4. Metrification

***Metrification*** is scaling among orientations. This is a critical operation in the spatial perception of metric structures. Essentially, a metric from one axis or orientation must be "carried" into another. A well-defined affine structure with metrically or statistically uniform elements distributed throughout the structure can provide structural metrification. Surfaces can constitute good structures for translating metrics between orientations. Many structures inherently provide relative scaling between axes. An extreme example is the "rat cage" in Fig. 3.2.

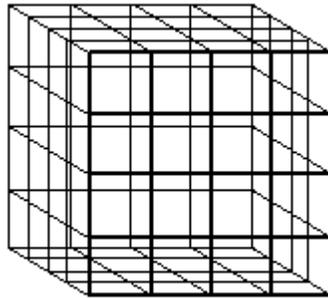


Figure 3.2. Self-scaling "Rat Cage."

In order to provide metrification through elements, the elements must be structurally or proximally related. Element recursions can have exact locations (e.g., bricks and tiles) or statistical characteristics. For example the statistics of the distributions of sizing and spacing of trees in a forest are relatively constant. Tree leaves are effective elements. They are of relatively uniform size, structurally related, and have many different orientations. For example, stand near a tree or a shrub. Place yourself a distance away from the nearest leaf roughly equal

to the span of the tree's canopy width. Look at the leaves that are near, mid-range, and distant. The immediate automatic perception is that the leaves are all the same size. It is only by conscious effort that one can see that the linear extents in the visual field of the leaves on the far side are one half of those at the near side. The visual field areas of the far side leaves are one quarter those of the near. These repeated, semi-randomly recursive visual elements in a continuous structure provide the metrics to scale the affine structure; they are not sized by it. Molecular models depicted as links or sticks and balls have rather uniform elements that are often randomly distributed in space and orientation.

As discussed above, motion can serve to carry metrics among axes. This is particularly true of rotations. Rotations and translations have been discussed extensively in Lappin and Wason (1991).

### 3.2.5. Absolute Scaling

*Absolute scaling* is made through reference, typically familiarity. An affine structure does not scale among axes. A metric structure does scale among axes but does not have an inherent absolute scale. Familiarity of objects or scenes provides scales either directly, i.e., the object in question is familiar, or through reference by proximity to and/or structural relationship with familiar objects. The scale associated with an object does not change as we change our position relative to it, as the structure and scaling are object-based.

An analogy to the independence of absolute scaling from structure is holography. A hologram can be recorded with light of one wavelength and reconstructed with light of a different wavelength. The reconstruction holograph will have a different size than the original; the reconstruction size is a function of the recording and reconstruction wavelengths. The structure will be unchanged, however.

### 3.3.STRUCTURE

#### 3.3.1. Global Structure

We cannot perceive space directly, only through the objects that occupy it. **Global structure** is the relationship between local structures in a global space. This is the most important aspect of spatial structuring. Recall that the specifics of retinal coordinates are not maintained in the visual system transformations of the retinal images. As the vectorfields are essentially local differentials, the display structure must be maintained with a global structure if the spatial relationships among widely separated scene elements are to be perceived correctly (Toet, 1987; Burbeck, 1992; Lasaga & Hecht, 1991; Wertheimer, 1938). This can be accomplished with a hierarchy of structures from local to intermediate to global. Aspects of global spatial perception can be considered an extension of the same mechanisms that provide local structure. In a display, an affine structure needs to be established. This structure is generated by the coherence of motion in separated spatial regions and through the connection and/or proximity of components. Many scenes inherently contain the necessary structure.

Toet (1987) has described a **hierarchical** perception of spatial order, from local to global. This concept can be incorporated into a hierarchy of spatial reference in the design of displays. First, the features within an object must have a perceived spatial relationship. This has been discussed above. Separated local structures do not inherently have a well-perceived relationship to each other. This relationship is established through a hierarchy of structures. First, there are local relationships. These local relationships in turn relate to a higher level of structure. The entire structural "tree" relationship can be considered a hierarchy of structures. A spatial structure that supports the relationship of local structures to a global structure should be only as complex as necessary. The objective of a well-designed 3-D display is the reduction of complexity while creating a metrically defined spatial structure. If the scene displayed does not contain the related levels necessary, the spatial perception will be inaccurate.

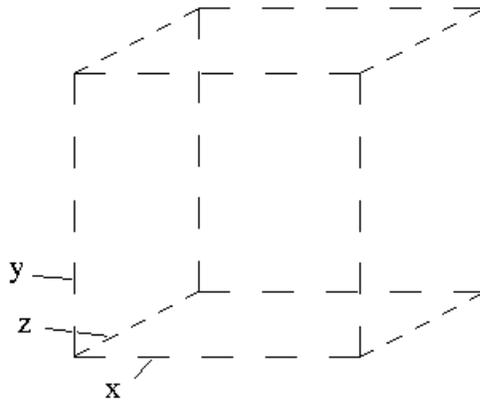


Figure 3.3. Cube of recurring elements.

Structure-from-motion, or stereo disparity, serves to create the affine structures at all levels of the hierarchy. The structures should be designed to create clear disparities regardless of the direction of the disparity (horizontal, vertical, or oblique). Long line segments can provide disparities only normal to their extension, and thus are less effective than a line defined by short segments. The ends of the lines, or line "terminators," are features that have good spatial location in all directions. A consistent element size can provide cardinality and metrification on local, intermediate, and global scales. As an example, the cube depicted in Fig. 3.3 has equal element sizes for all three axes ( $x$ ,  $y$ , &  $z$ ). It is a simple, unified structure with little visual clutter. The edges of the cube are formed from dashed lines, providing uniform elements throughout. Although the cube is affinely shear-strained in an orthographic projection, the result is a robust perception of a cube in depth, which is ambiguous in depth, periodically reversing.

An example of a hierarchy of spatial structures will be useful. At the lowest level, an isolated object, such as the small cube, needs to have an observable relationship with a local spatial feature. This feature has a spatial relationship with the next level upward on the scale of

a structural spatial hierarchy. A small dot is a minimal local feature. It has a spatial relationship with other dots arranged in a line spanning the depth of the display (Fig. 3.4). These dots in space have been called "towers," since on a dark display they are similar in appearance to

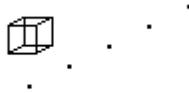


Figure 3.4.  
Dot tower.

radio towers seen from a low-level aircraft at night.

The dot towers produce little spatial clutter. The tower can be located in space relative to a larger structure; for example, the dashed cube described above serves as a useful spatial structure (Fig. 3.6). It spatially defines a significant volume.

The cube can be the top level of a structural hierarchy, spanning the entire volume of interest. Cubes can be assembled into larger structures (e.g., Fig. 3.5), thus providing greater spatial resolution of the enclosed volume.

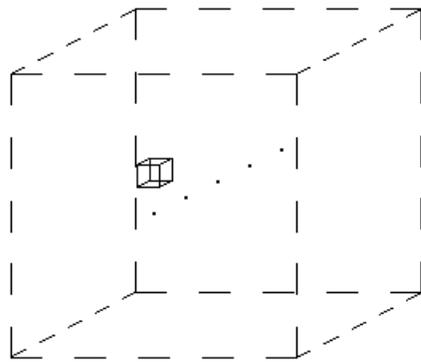


Figure 3.6. Two-level structural hierarchy.

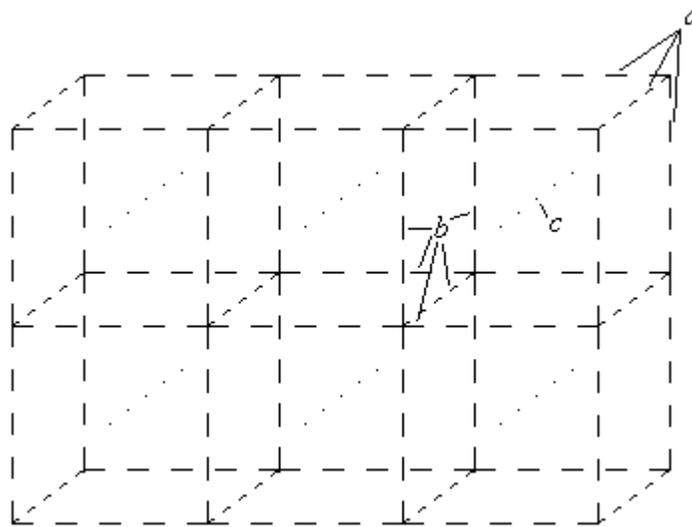


Figure 3.5. Three-level structural hierarchy.

A rule in the development of a display spatial structure is "less is better." The structure of space must be well-defined, however. As discussed above, objects will often embody enough structure with adequate metric elements to obviate the need for other structures. An analysis of the structure desired, as outlined above, will provide a basis for the design of displays. Ellis, Smith, Grunwald, and McGreevy (1991) reported a study of the perception of the location of aircraft symbols in a computer-simulated airspace. Aircraft were placed over a coarsely gridded ground field. One or two vertical risers projected from the field to each aircraft. Perception varied considerably as the line of sight relative to the ground varied. Addition of a second riser reduced, but did not remove, the variability in perceiving the angle and altitude of the symbols. Yeh and Silverstein (1991) reported similar errors in monocular and binocular displays of geometric figures over a similar field. As the angle of the observer's line of sight became more parallel with the ground, the errors in perception of location and altitude of the figures increased. Both studies yielded poor results in the subjects' perception of 3-D spatial relationships. In both studies the volumes depicted were poorly defined from a structural standpoint. The entire volumes were not structured, but only the ground. Symbols were referenced to a structure that was not hierarchically continuous, there being only two levels of structure, local and a partial global. An objective of this dissertation is the production of well-defined space within which objects can be accurately localized and scaled.

### 3.3.2. Symmetries

Certain structures perform rotational coupling due to their symmetries. Regular geometric figures, e.g., cubes, tetrahedrons, tessellated spheres, rectangles, and structures with well delineated equal scales of grids on the surfaces, have scales that are common among orientations. These regularities can be the basis of unifying structures. Additionally, structures that contain interconnected elements of about the same size that are semi-randomly oriented provide effective rotational coupling. Although this may seem to be a difficult condition to meet, it is

quite common. For instance, stick-and-ball molecular models and finite-element delineated structures are self-scaling in depth.

### 3.4. PERSPECTIVE

For most implementations of perspective in spatial displays, there have been no consistent results in producing reliable, correct perceptions of spatial location and orientation of elements in space when perspective is used as a cue to depth (Ellis, Smith, & Hacisalihzade, 1989; Yeh, & Silverstein, 1992). As a general rule, SFM space, and, for that matter, stereopsis space, should have less—one-half to one-quarter—of the perspective present in the actual distances and locations involved. A reduced representation of perspective is consistent with the perception of an affinely structured space, since the visual system scales the depth from other information and can benefit from the decrease in non-affine transformation inherent in the perspective mapping. The efficacy of reduced perspective in displays is consistent with the finding of Cutting (1991) that the visual system is quite insensitive to distortions due to affine transformations, including shear strains, but is sensitive to the non-affine distortions of perspective.

For a perspective view to be effective, there needs to be a continuity in depth that provides a continuum of near-affine transformations. In motion pictures, although the objects move in  $E_3$ , the camera, projector, and viewer position subject that space to affine and non-affine transformations. With no depth continuum such as a telephoto lens shot with no objects at intermediate depths, the continuity is missing, and one affine space appears to abut a significantly different space that is not a near-affine transformation, resulting in perceptual distortions. Extreme perspective, produced with the use of a short focal-length lens, creates significantly non-affine transformations, and should produce a less compelling perception of rigid, undistorted space in projection (Cutting, 1991). Let us move to a more formal discussion of perspective transformations that can be used in the development of a display.

The display depth space is normalized to the vectorfield. The use of perspective requires a method of mapping perspective space into a vectorfield,  $\mathbf{V}$ . To do this, mapping functions,  $f_m$ , can be developed. We can consider that the vectorfield,  $\mathbf{V}$ , is to span some specific depth in perspective space (Fig. 3.7).  $\mathbf{V}$  has a span of  $+\mathbf{V}_m$  to  $-\mathbf{V}_m$  for which  $\mathbf{V}_m$  defines the maximum vector length. The nodal plane,  $N_p$ , lies at the zero length vector. In a binocular system,  $N_p$  would be considered the plane of convergence, and  $+\mathbf{V}_m$  might not equal  $-\mathbf{V}_m$ , as depth of field for some people is greater in the crossed than in the uncrossed disparities.

The general format of the perspective mapping function can be modeled in three ways:

1. As a non-affine transform.
2. Locally, as an affine transform in all directions.
3. Intermediate, as a semi-affine transformation whereby z-axis is affine, and x and y have perspective compression with depth. This is a good model that will be useful for most displays.

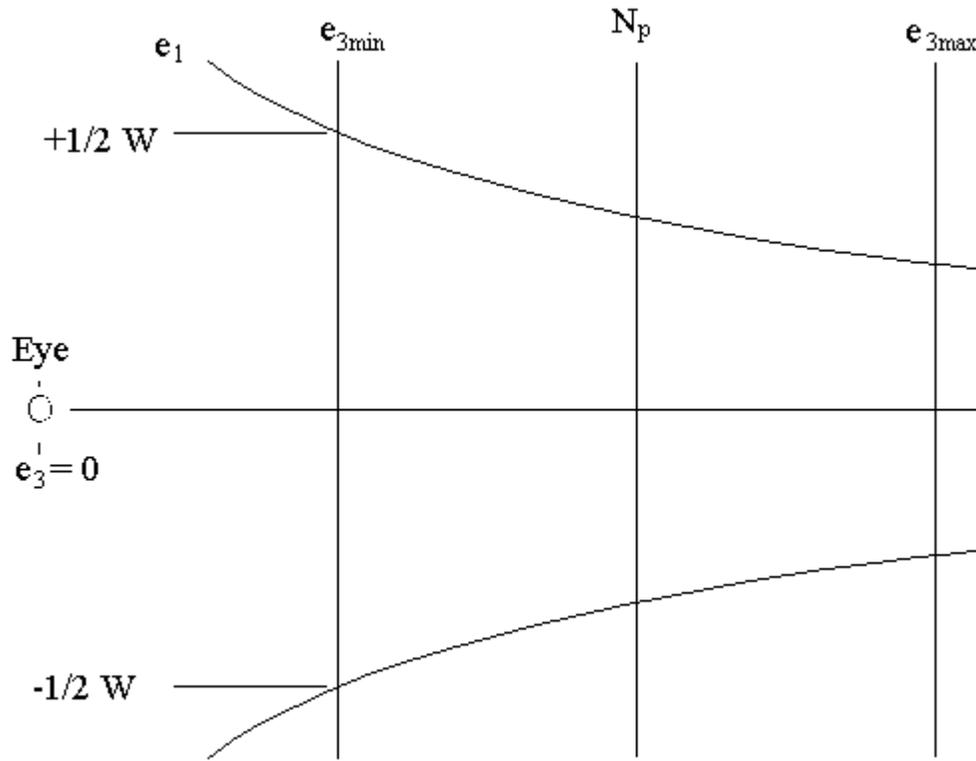


Figure 3.7. Perspective depth span relative to vectorfield.

### 3.4.1. Non-affine Transformation

A non-affine transformation for perspective is a full modeling of perspective in which all dimensional unit vectors,  $u_i$ , are compressed as depth ( $e_3$ ) increases. The equation below is written with the origin at the same location as the observer. A defined subspace can be placed in the perspective space. The form of the mapping function transformation is:

$$u_i' = u_i \beta_i / e_3, \quad i = 1, 2, 3 \text{ for Euclidean } (E^3) \text{ space, a } \mathbf{unit\ vector}. \quad (\text{Eq. 3.1})$$

$$\beta_i = k, \quad i = 1, 2, 3. \quad (\text{Eq. 3.2})$$

$u_i$  = unit vector in Euclidean space,  $u_i'$  = transformed vector.

### 3.4.2. Affine Transformation

By looking at the perspective depth span in Figure 3.7, one can see that for short depth spans relative to the distance between the observer and the window,  $W$ , over which there is no significant change in  $e_1$  (x),  $e_2$  (y), or  $e_3$  (z), a simple affine scaling function can be used. This is a local model of perspective with no perspective transformation or compression of the x and y dimensions with depth. The z-axis transformation is affine:

$$u_i' = u_i \beta_i, \quad i = 1, 2, 3 \text{ for Euclidean } (E^3) \text{ space.} \quad (\text{Eq. 3.3})$$

$$\beta_i = k. \quad (\text{Eq. 3.4})$$

### 3.4.3. Semi-affine Transformation

For some intermediate depth spans, the  $e_1$  and  $e_2$  directions have a significant perspective contraction with respect to  $e_3$ , but depth can maintain a linear scale for normal affine scaling by the monocular scalar fields. This is an intermediate model of perspective. The model can be considered a semi-affine transformation (z-axis affine, with x and y having perspective compression with depth). This model can be used for most display situations. Its rigorous form is:

$$u_i' = \frac{u_i \beta_i}{e_3}. \quad (\text{Eq. 3.5})$$

$$\beta_i = k, \quad i = 1, 2 \text{ and} \quad (\text{Eq. 3.6})$$

$k$  is an appropriate constant. A linear approximation can be made:

$$u_i' = u_i \beta_i, \quad \text{where} \quad (\text{Eq. 3.7})$$

$$\beta_i = \frac{\Phi/L - ke_3}{W}, \quad i = 1, 2. \quad (\text{Eq. 3.8})$$

For either case there is no compression in the depth dimension:

$$\beta_i = 1, \quad i = 3 \quad (\text{this is an affine transformation in depth}). \quad (\text{Eq. 3.9})$$

The second form (Eq. 3.8) is the preferred one, as it more closely approximates an affine transformation through the use of a small value of  $k$ . As noted, perspective should be under-represented to approximate an affine transformation of space.

#### 3.4.4. Application of Affining Levels

These three levels of representation of perspective can be applied in roughly three ranges of depth spans ( $z_s$ ) in terms of the major screen dimension (usually width),  $W$ :

1. non-affine transform (full representation):  $z_s < W$
2. affine transform:  $z_s < 1/2 W$
3. semi-affine:  $z_s < W$

There is overlap in these *affining level* ranges. Obviously the full non-affine transformation can always be used, but it is computationally more complex and often not required. One of the forms of the semi-affine transformation can be applied instead of the non-affine transformation for any  $z_s$  of less than  $1/2 W$ . For depths of less than  $1/2 W$ , the simple affine transformation can be used, with no compression of  $x$ ,  $y$ , and  $z$  for depth. The semi-affine transformation will be used in most applications, and can be considered the "default" condition since it defines a roughly cubical space.

#### **4. EXPERIMENTAL METHODS**

This dissertation has three objectives:

1. Creating the perception of 3-dimensional space from a 2-dimensional surface
2. Testing the hypothesis of the perceptual equivalence of affine spaces
3. Demonstrating that metric structuring largely controls the perception of affine spaces.

These objectives are related. A 3-dimensional display was generated by producing affine structure-from-motion (SFM). The perceptual equivalence of different affine spaces was demonstrated within the display. The perception of the sizes of structures within the affine space display was controlled with metric factors.

The experimental methods have two principal components: 1) the experimental design, and 2) the display generation. Since the specifications for the experimental stimulus display provide the basis for the hardware selected and the software design, the experimental design will be discussed first.

##### 4.1. EXPERIMENTAL DESIGN

According to the hypothesis, the metric scaling of an affine structure is the primary determinant of spatial structure. An hypothesis is that under appropriate conditions, one can generate an affine space independently of the scaling, or metrification, of that structure. A test of the independence of metric processes in perception requires the independent manipulation of affine structure and metrification. The independent variables of this experiment must include an affine manipulation and a metric factor. A dependent response that is variable and potentially responsive to perceived metric structure is required.

The response variable should reflect a subject's perception of metric space. It should be quantifiable. A stimulus is required that does not have an inherent metric structure. A cube

has an inherent metric structure. A pyramid does not have an inherent ratio of height-to-width, or *aspect ratio*. The height of a pyramid relative to the width of its base is indeterminate when viewed monocularly from the top (Fig. 4.1). This view projected onto a flat surface provides inadequate information to determine the pyramid's aspect ratio, except (as in Fig. 4.1) to indicate that if the four faces are not equal, the structure cannot be flat.

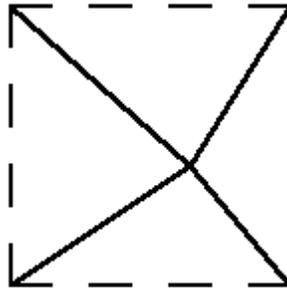


Figure 4.1. The top view of a pyramid.

The pyramid was placed in an affine space and metric factors were applied. The metric structure was a square-front-faced rectangle of dashed lines. The stimulus was comprised of a pyramid centered in the metric structure, such as in Figure 4.2. The entire space was shear-strained along the  $x$  and  $y$  directions to make the dashed risers visible. Affine space was generated using shear-strain structure from motion (SFM) as described previously.

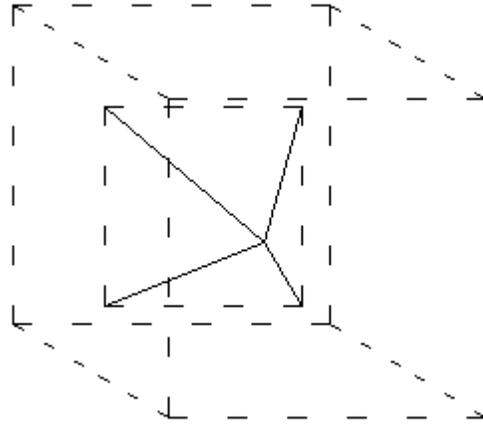


Figure 4.2. A typical display stimulus.

The affine space was created with 256 levels in depth by defining the depth dimension as an eight-bit variable ( $2^8 = 256$ ). Three affine pyramids were used, with affine spans of 102, 160 and 240 levels. The spans were adjusted to match the ratios achievable for the metric structures. These pyramids had affine height ratios of approximately  $2/3 : 1 : 1-1/2$  or  $1 : (1 \times 1.5) : (1 \times 1.5 \times 1.5)$  (Fig. 4.3). This is a logarithmic series, producing a constant perceptual difference among pyramid levels according to Weber's law of proportionality of detectability.

Four levels of metric structures were used. These are illustrated in Fig. 4.4. The front and back planes are squares of 6.3 cm, with 6 dashes per edge. The space between dashes was twice the length of the dash. The front face was 16 dash units wide and high. Metric level 1 had 4 dashes in depth, or a depth of 10 dash units. Metric level 2 had 6 dashes in depth (16 dash units), and metric level 3 had 9 dashes in depth (25 dash units). Metric levels 1 - 3 spanned the entire affine space of 256 levels. The metric level 0 had no depth; it was a square lying at the nodal plane.

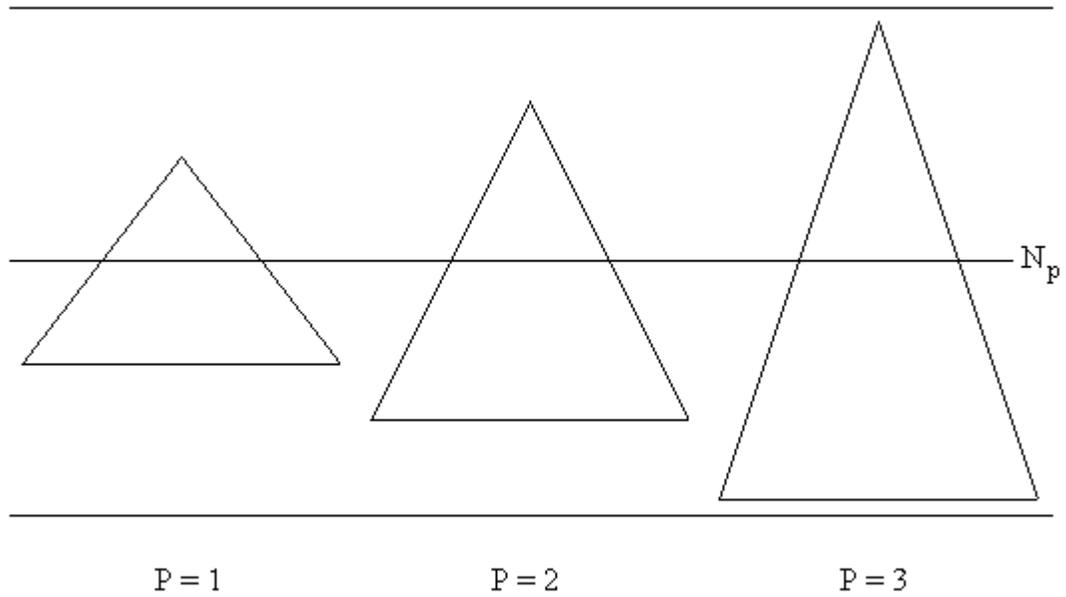


Figure 4.3. Profiles of the stimulus pyramids in affine space.

Gross static affine shear-strains in the  $x$  (OX) and  $y$  (OY) directions were applied to the affine space (Fig 4.5). These offsets were adjusted for each metric level such that the angle of the metric structure risers would be constant in the display. The display had a screen resolution of  $1024 \times 768$  pixels. The affine space with the metric level 2 was subjected to gross offset affine shear-strains of  $OX = 36$ ,  $OY = -20$ . The metric level 1 shear-strains were  $2/3$ , and the metric level 3 shear-strains were  $1\frac{1}{2}$  of metric level 2 (Table 4.1). This maintained the same logarithmic series used in the metric structure and in the pyramid affine aspect ratios. The pyramid and metric structures were both subjected to the same gross offset shear-strains.

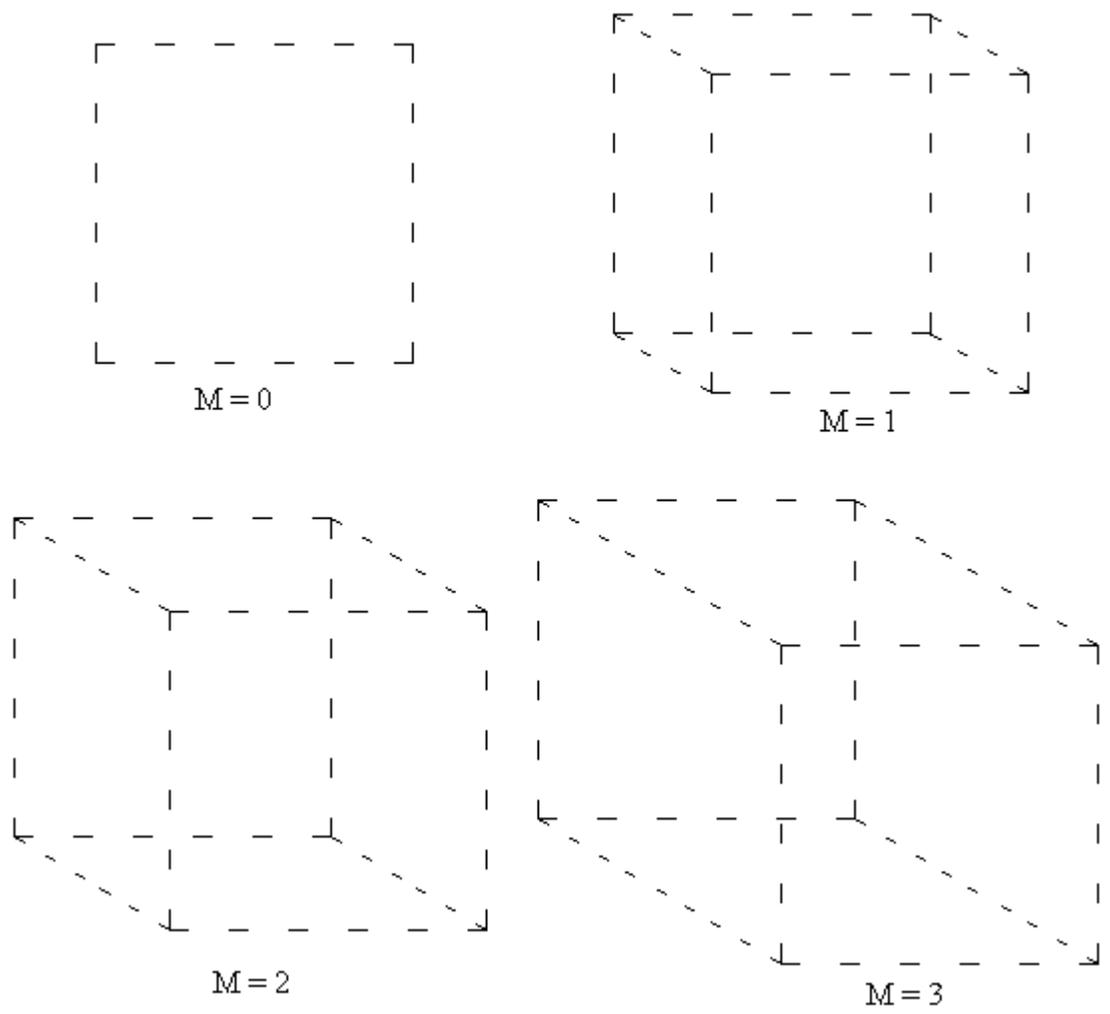


Figure 4.4. The four metric structures.

Table 4.1. Metric level risers and offsets.

Metric Level	Riser Units	OX	OY
0	0	0	0
1	6	23	-13
2	9	36	-20
3	25	59	-31

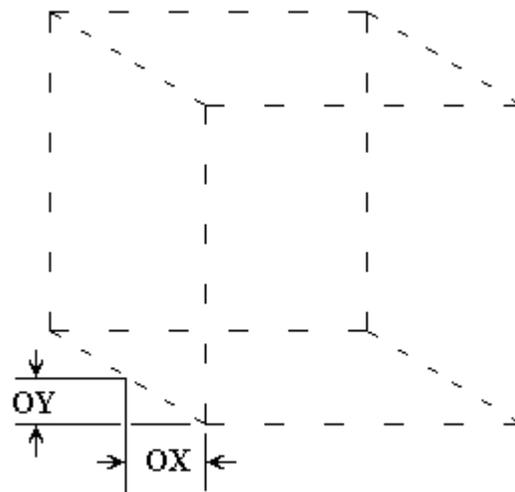


Figure 4.5. Metric shear-strain offsets.

All structures were subjected to a semi-affine perspective transformation. The  $x$  and  $y$  dimensions were reduced with increasing depth. The same perspective was applied to all experimental conditions. The perspective was applied to affine space, as opposed to metric space. The affine space was transformed as if its depth were 7.5% of the distance from the

subject to the screen. This is a very slight perspective transformation that served to prevent the display from appearing distorted by a lack of perspective.

To provide clear separation of the metric structure and the pyramid, each was displayed in a different color. The pyramid was displayed in green, the metric structure in yellow. Intensity decreased with increasing affine depth. The same intensity depth-cueing was applied to all affine spaces. The background was black.

Affine space was generated through small affine shifts. The small affine shifts deformed the space in a manner shown in Fig. 4.6. These shifts were randomized in direction and timing. A six-position aperture circle of radius  $\epsilon$  was used, with the aperture positions equally spaced around the perimeter (Fig. 2.25). This pattern produced three aperture shift lengths:  $\epsilon$ ,  $1.732\epsilon$ , and  $2\epsilon$ . The size of the affine vectors was proportional to the size of  $\epsilon$ . Four affine shift radii were used: 0, 2, 3, and 3.5 pixels, corresponding to affine levels 0, 1, 2, and 3. Affine levels 1, 2, and 3 maintained the same 1:1½ ratio among levels as was used in the pyramid levels and the metric levels.

Changing the aperture shift radius  $\epsilon$  changed the affine space. A larger radius increased the vector lengths, consequently increasing the affine space depth. The affine depth of the structures embedded in the space changed with the space, as shown in Figure 4.7.

A relative shift duration of 1 follows an aperture shift of 1 unit. A shift two positions around the aperture set has a relative magnitude of 1.732, and subsequently a relative time duration of 1.732. A three position shift, across the diameter, has a relative shift magnitude of 2 and a resulting duration of 2. If one considers saccades to occur randomly in time (Mates 1978), then the intervals between saccades can be modeled as having a Poisson distribution. The timing of the pseudosaccades followed an approximately Poisson distribution of 3:3:2 for the relative shift durations of 1, 1.732, and 2 respectively. A random sequence was generated from a total of 48 shifts. The computer program tested random sequences until it produced one

in which all aperture positions were occupied an equal number of times,  $\pm 1$ . Using the exponential decay model of the disparity drive, the duration of time after a shift was made proportional to the magnitude of the shift. The shortest shift used in the experimental trials was 180 ms. The mean duration between shifts was 272 ms.

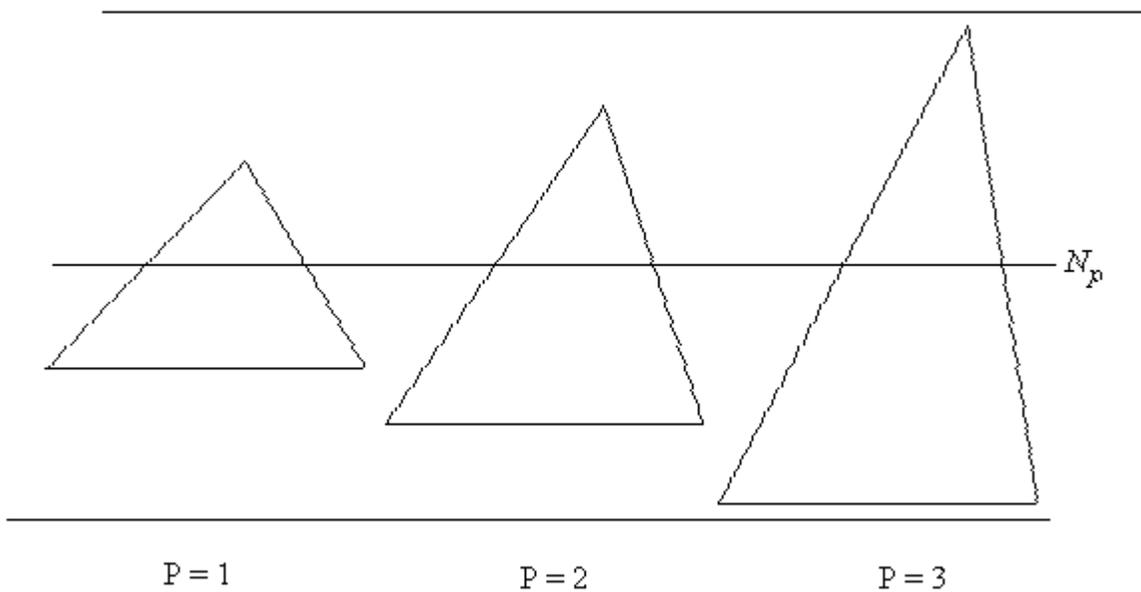


Figure 4.6. A shear-strain shift for producing affine structure.

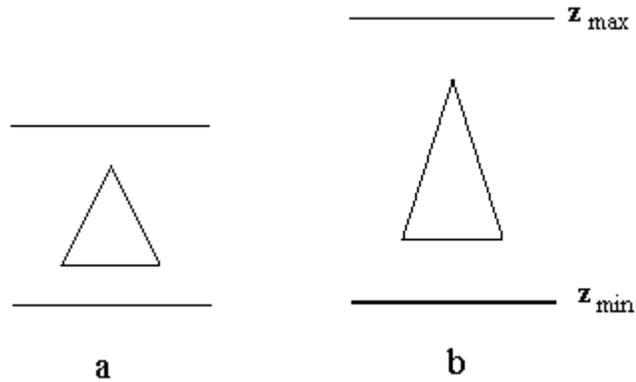


Figure 4.7. The same affine pyramid in two affine spaces.

Conceptually the gross offset shear-strains, depth intensity cueing, perspective, and SFM pseudosaccadic affine shifts were applied to the affine space, and consequently to all structures depicted in that space.

Subjects were asked to judge the aspect ratio of the pyramids. The response scale (R) was a set of figures of ten pyramid profiles (Fig. 4.8). The aspect ratio of the profiles from 1 to 9 spanned the same range as would be predicted for the stimuli presented. A pilot study with a larger span indicated that an extended span was not necessary, as responses fell within the predicted span. Scale values 1, 3, 5, 7, and 9 corresponded to the figures actually presented. Scale values 2, 4, 6, and 8 followed the logarithmic nature of the scale. Scale figure 0 was a line, indicating a flat figure. If subjects used the largest triangular face of a pyramid as the basis for their judgments, the result would be to increase the mean response, as the  $M = 0$  view of a pyramid face has the appearance of a scale level 2. This judgment base would also slightly collapse the scale.

The theoretical response (R) scale value of a pyramid can be calculated as:

$$R = 2P + 2M - 3. \tag{Eq. 4.1}$$

The scale value of stimuli with metric levels of 0 is indeterminate; consequently, it was declared as 0. Based on the vector-scalar product model, an affine level of 0 also provides inadequate spatial information for aspect ratio determination. Note that a scale level can be produced by several different combinations of pyramid and metric levels.

The use of the same ratio among pyramid levels, metric levels, and affine levels produced stimuli that had the same display image for different conditions. For instance, A pyramid of level 2 in a metric space of 3 had the same image as a pyramid of level 3 in a metric space of 2. Similarly, a pyramid of level 2 subjected to an affine shift of level 3 underwent the same image transformations as a pyramid of level 3 subjected to level 2 affine shifts. This strategy of common multiples was intended to prevent the subjects from using the flat display image in a meaningful way in their judgments.

The three independent variables were used to produce a total of 48 conditions, each represented once per block of trials:

- 4 affine aperture shift levels (0, 2, 3, 4.5)
- 4 metric levels (0, 4, 6, 9)
- 3 affine pyramid levels (102, 160, 240).

Subjects were first shown a demonstration of the SFM display. Three objects were presented. The first was a green monochrome cube with all corners connected. The cube spanned an affine space of level 2. Subjects were encouraged to visually explore the space. The second presentation was of a stick model of a glucose molecule in a "chair" configuration. Two colors, green and yellow, were used. Finally, a full hierarchical structure of  $2 \times 3$  dashed yellow cubes with a green pyramid in the lower center cube was presented. The subject was instructed that he or she would be making judgments of the aspect ratio of the pyramid. All subjects readily reported the spatial structures as three-dimensional, but not all subjects saw all pyramids as being completely within the metric structure.

Figure 4.8. Response scale.  
The scale for reporting the aspect ratio of the pyramid profiles.

Following the introductory demonstration, the subjects performed a practice block which included instructions (see sec. 8, Appendix). This practice ensured that all subjects received the same instructions. Subjects were instructed to use the profile scale that was placed on a stand below the CRT display, facing the subject, to estimate the aspect ratio of the pyramid in the display. Subjects responded by pressing an appropriate number key on a standard computer keyboard. Non-numeric entries were not accepted; a low tone sequence denoted an in-

appropriate response. Immediately after the keypress the screen was blanked, and a new presentation was loaded into the computer memory and displayed. This required 10 seconds. The trials were self-paced by the subject, with no time limit. The computer automatically recorded the conditions and results.

Each subject responded to 3 blocks of 48 trials each. The trials were individually randomized by the computer with no repeated blocks. Subjects were told at the start of the experiment that they would have 3 blocks of trials with a five-minute break between blocks. The program instructed the subject to take the break at the end of each block. They were told that a block ordinarily took about 20 minutes to perform, but that they were under no time constraints.

The experiment was conducted in a quiet office environment with lowered light levels. The subject was alone in the office; I was in the next room. The adjoining door was open.

Twelve unpaid volunteer subjects were drawn from my acquaintances, eight males and four females, ranging in age from 32 to 65 years old. All use personal computers. All were currently enrolled in, or had completed, college. All subjects used appropriate visual correction, if required. All subjects reported normal color vision. All subjects successfully completed the task. Use of human subjects was approved by the Institutional Review Board on Research Involving Human Subjects at North Carolina State University.

#### 4.2. DISPLAY

A display system was developed to implement the experimental design. The display requirements include:

- The ability to present figures that are made up of lines, or *vectors*, (wireforms) using any number of vectors
- Depth intensity cueing (4 bits)

- The ability to move line positions and orientations by small amounts ( $< .5\%$  of the display width)
- Generation of gross and fine affine shifts in displayed structures
- Synchronous image change over the entire display
- Randomized presentation of six image frames
- Pseudorandom timing of shifts
- Automatic data logging.

There are two general classes of display requirements: 1) image characteristics and 2) timing. These considerations were addressed in the selection of the hardware and the design of the software. Although the principles developed above can be applied to forms with colored surfaces, this study used simple vector drawings.

Producing an effective small-motion affine SFM display requires the ability to plot forms as continuous lines that change in position and orientation by small amounts. The pixel resolution of a standard computer display would have unduly limited the resolution needed for the experiment. Normal computer CRT displays are made up of dots in a rectilinear matrix of columns and rows. Images are plotted by changing the illumination level of each dot, or *pixel*. Thus the image resolution appears to be equal to the size of the dots. This limits the characteristics of a display due to *aliasing*.

Aliasing produces an irregular appearance in lines drawn at any angle other than the vertical or horizontal. Fig. 4.9 illustrates the problem. Lines generated at non-axis angles are produced as a series of steps. As the angle of the line changes, the pitch of the steps will

change. Thus, if the line undergoes a small rotation, small features are generated on the lines that change in a manner that does not reflect the actual behavior of the line.



Figure 4.9. Examples of pixel aliasing.

The amount by which a point can be shifted in such a pixel-plotted image is limited by the size of the pixel. This is the pixel resolution of the display. There are two ways to improve the resolution of the display and also decrease the effects of aliasing: 1) increase the screen pixel resolution, 2) use antialiasing techniques. Both were used to create an effective display.

Antialiasing produces increased perceptual resolution, particularly under motion. Antialiasing utilizes the fact that perceptual spatial resolution is greater than is specified by spatial frequency (Burr, 1980). Antialiasing is achieved by plotting the center of brightness of a line. Flanking pixels are plotted at reduced intensity to decrease the discrete steps. This center-of-brightness plotting technique permits the movement of a line by less than the size of a pixel. This was used to achieve the small affine SFM image transformations. The screen pixel resolution was  $1024 \times 768$ . The computer program used a 2X subpixel plotting technique to achieve antialiasing (Covington, 1990). Lines were plotted with a modification of Bresenham's line algorithm (Annino & Driver, 1986). The modification plotted in three dimensions to allow intensity modulation with depth, or intensity depth cueing. All images were precomputed and stored on a hard disk.

The two colors that were used, green and yellow, were stored in the graphics card digital-to-analog converter (DAC) look-up table. Each color was formed from a mixture of the phosphor colors, red (R), green (G) and blue (B). Each phosphor color was defined with 6 bits, providing 63 intensity levels for each phosphor. The display colors were generated by mixing the colors to provide different intensities with no hue change. Green was formed from a ratio of R:G:B of 1:2:1; yellow was a ratio of 1:1:0. To provide gamma compensation for phosphor response the summed intensities followed a logarithmic curve:  $I = K * \ln((DAC \text{ mod } 64) + O) + M$ . A color calculation program took as input the level, M, for the lowest intensity, and an offset, to create the curve between the minimum intensity and the maximum intensity (127). A minimum of 45 and an offset of 10 provided good antialiasing and good depth-cueing. All figures were plotted as lines (vectors) in depth, with decreasing intensity for increased depths. Each pixel was defined by 1 byte: 6 bits of intensity, 2 bits of color. Only two colors are used, using 127 of the 256 DAC values.

The display was presented on a NEC 4FGe 15" CRT color monitor. Because this monitor does not have an etched screen, it provides a sharper image. The image size on this monitor can be adjusted, as can the color balance. This is a good monitor for psychophysical experiments. The display was fitted with a modified anti-glare filter. The optically coated antiglare filter was mounted in a bezel which slides onto the monitor. The filter had a transmission of 31%. To increase the display brightness ratio (DBR), the bezel was painted flat black. A 6"-deep visor was attached to the bezel. The inside surface of the visor was also painted flat black. The visor reduced the light falling on the bezel, increasing the DBR further, and on the screen, increasing the pixel contrast ratio (PCR). The resulting image was sharp and clear. The display surface was nearly invisible in an office with low light.

The program was written in Borland Pascal 7.0 for an AT-compatible 486DX personal computer with a clock speed of 66MHz with 16MB of RAM. The computer had an ATI UltraGraphics video card with 2MB of video RAM.

The timing requirements of the display process were unusual. As pseudosaccades with pseudorandom timing were used for the aperture shifts, precise timing was not necessary. On the other hand, in order to maintain image change coherence over the entire screen, the entire display image had to change at one time. These two requirements together created an opportunity for a cost-effective display. The image generation and the image display changes can proceed asynchronously.

Pixel-based computer video systems produce an image by reading a buffer memory over and over for each frame refresh. Rather than attempt to produce a new image in a frame buffer within one frame refresh interval, one can produce the images in separate frame buffers and then switch the frame buffer during the vertical retrace. This strategy provided adequate time to transfer an image into a frame buffer without loss of coherence in the image change in the displayed frame.

## 5. RESULTS

### 5.1. GROUP RESULTS

A total of 1728 reponses were collected from 12 subjects. All subjects successfully completed three blocks of 48 trials each. The subjects' responses correlated with, but did not match, the calculated heights of the pyramids (Fig. 5.1, Table 5.1). The subjects tended to overestimate the heights of the pyramids, especially at the low end of the scale. The intended mean height was 3.75; the subjects' mean response was 5.545 ( $\sigma = 1.824$ , Std. Error [S.E.] = 0.044).

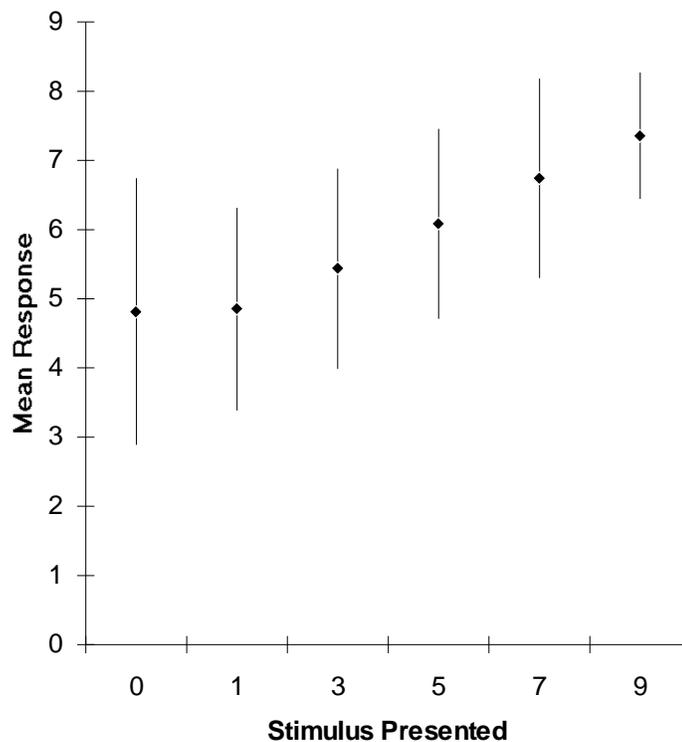


Figure 5.1. Mean responses to stimuli with  $\pm 1$  std. dev.

Table 5.1. Responses compared to stimuli presented.  
Data from all subjects combined.

	Stimulus Level Presented						Responses
	0	1	3	5	7	9	
Response 0	10	0	0	0	0	0	10
1	19	0	0	2	0	0	21
2	56	2	2	0	1	0	61
3	87	18	12	5	0	0	122
4	168	23	37	18	3	0	249
5	156	39	70	92	18	1	376
6	127	13	62	90	70	45	372
7	62	5	14	77	79	45	282
8	37	6	5	17	26	34	125
9	34	2	14	23	19	18	110
Presentations	756	108	216	324	216	108	1728

Most strikingly, the subjects' responses did not reflect the large number of stimuli with metric (M) levels of 0. A metric level of 0 has a theoretical pyramid profile of 0 on the response scale. The mean of responses to stimuli with non-zero metric levels was 5.964 ( $\sigma = 1.577$ , S.E. = 0.044). This is larger than the mean of the non-zero metric level stimuli (mean = 5.00,  $\sigma = 2.310$ , S.E. = 0.064). Clearly the metric = 0 stimuli did not produce an equivalent number of responses near 0. The mean response for those stimuli with non-zero affine (A) levels was 5.733 ( $\sigma = 1.724$ , S.E. = 0.048), again, considerably higher than the stimulus mean.

The main effect was that metric level and pyramid level were monotonically related to mean response at all affine (A) levels. An increase in either variable, M or P, produced an increase in mean response level as measured with the pyramid profile scale. This significance was reflected in the ANOVA (F test) shown in Table 5.2, which indicates very low probabilities that chance phenomena could account for the result. Affine level was only dichotomously effective, with significance only between a zero and a non-zero state.

Averaging the responses across all pyramid levels for the means of the responses by metric levels and by affine levels reveals that the metric level affects the perception of the

height of the pyramid compared to its base, as shown in the "Metric Means" plot in Fig. 5.2. The greater the level of metric structuring, the greater the perceived height. Affine level produced its greatest effect only between the zero and non-zero levels (Fig. 5.2). Converting the affine level to a dichotomous variable of 0 and 1 by substituting 1 for all non-zero levels did not change the significance in the factor analysis of variables (ANOVA) from the multilevel ( $p < 0.0001$ ,  $F_{3,1719} = 29.72$ ) to the dichotomous ( $p < 0.0001$ ,  $F_{1,1718} = 82.41$ ) affine variable, and did produce a significant small M\*P interaction component ( $p < .005$ ,  $F_{3,1718} = 4.31$ ).

The subjects' responses to the metric = 0 condition (Fig. 5.3) were also different from those to higher metric levels. With the 0 metric condition (Fig. 5.3), the response differences among pyramid levels was small, though significant (see Table 5.2). Each point in Figures 5.4 and 5.5 represents one of the 48 stimulus conditions and is the mean of 36 responses. The effect of affine levels was also significant and larger than that of the pyramid level in the metric = 0 condition. The estimated regression parameter for affine level (0.4852) was also larger than for pyramid level (0.2847) at the 0 metric level. This was not true at higher metric levels, where both the F value and the estimated regression parameter were larger than for affine level (Table 5.2). Clearly the effects of zero affine and metric levels were different from those for the non-zero levels. A model using each of the parameters as a main effect was statistically significant ( $p < 0.0001$ ,  $F_{8,1719} = 105.6$ ), however.

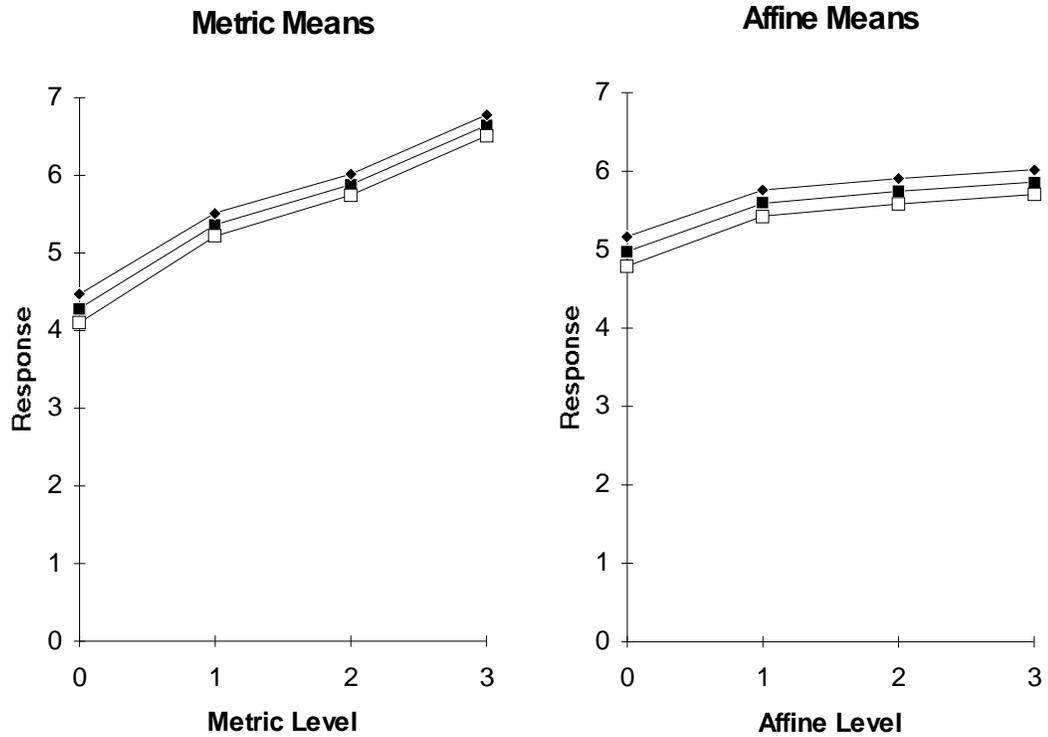


Figure 5.2. Metric and affine means of all responses. Metric means are collapsed across affine levels, affine means are collapsed across metric levels. Upper and lower 95% confidence interval of the means shown. Each point is the mean of 432 responses collapsed across pyramid levels and subjects.

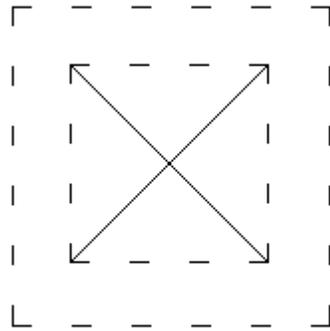


Figure 5.3. A metric = 0 image.

An analysis of a subset of records that contain neither metric = 0 nor affine = 0 levels (972 records) allows us to look at the effects of affine level and metric level, the main subject of this dissertation. Selection of this subset is a reflection of both the dichotomous nature of the affine level and of the atypical appearance of displays with a 0 metric level. This subset produces significant effects for both the metric level ( $p < 0.0001$ ,  $F_{2,965} = 73.16$ ) and the pyramid level ( $p < 0.0001$ ,  $F_{2,965} = 97.01$ ). As discussed in the experimental design, the purpose of multiple levels of pyramid aspect ratios was to provide stimuli that the subject could not recognize on the basis of their affine shifts. The metric and pyramid regression parameters were both significant (Table 5.2). With this subset of data, however, the affine parameter was no longer significant ( $p < 0.2037$ ,  $F_{2,965} = 1.59$ ) in the determination of the subjects' responses. This was also reflected in the small estimated affine regression parameter. The  $M = 0$ ,  $A = 0$  condition reduced the effect of the pyramid independent variable to insignificance ( $P > 0.7876$ ,  $F_{2,105} = 0.24$ ), indicating that shading alone was inadequate to separate pyramid levels.

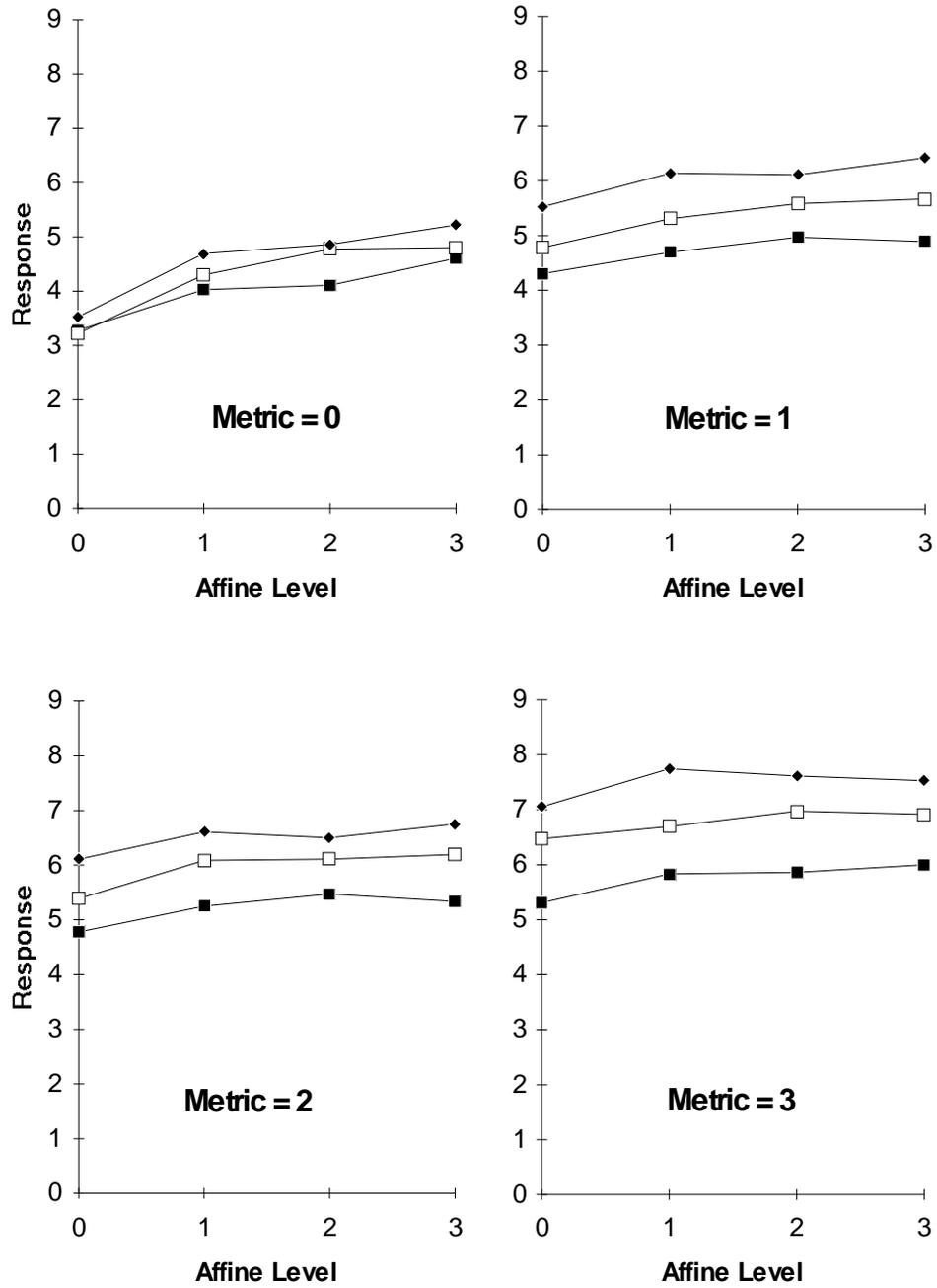


Figure 5.4. Response means for each metric level. Each plot has responses separated into the 3 pyramid levels. The lower line is means for pyramid 1, the upper line is means for pyramid 3.

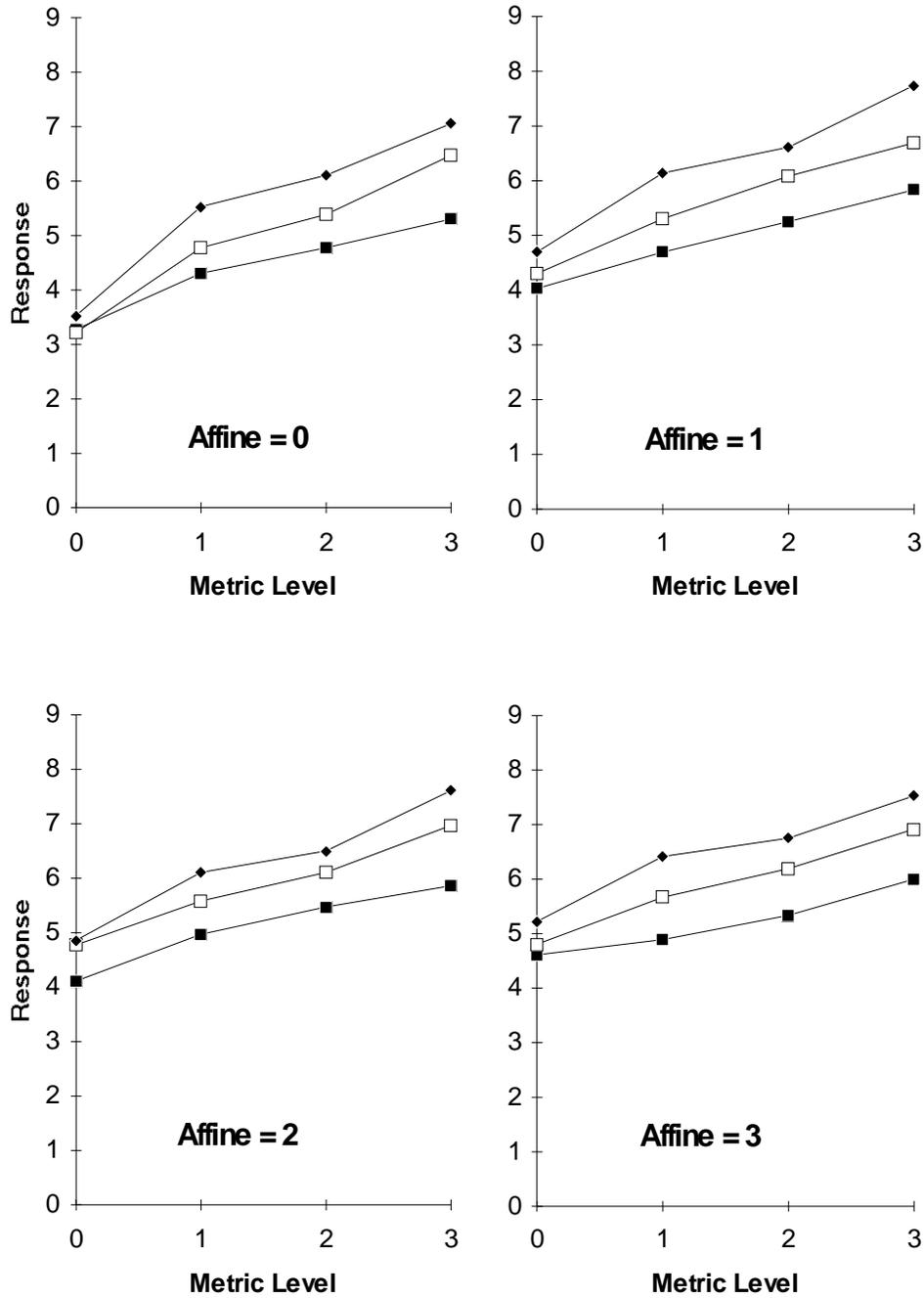


Figure 5.5. Response means for each affine level. Pyramid levels plotted separately. In each plot, the lower line is for pyramid level 1, upper for pyramid level 3. These are the same data as in Fig. 5.4, arranged differently.

Table 5.2. Parameter estimates and significance tests.  
 Analysis by metric groups and for all data grouped together.  
 Linear regression parameter estimates with T tests. F tests of  
 ANOVAs. Data from all subjects pooled.

<b>METRIC</b>	0	1	2	3	All	All w/o Metric = 0 or Affine = 0 records
Parameter	-	-	-	-	0.7586	0.6173
$p < T$	-	-	-	-	0.0001	0.0001
$p < F$	-	-	-	-	0.0001 $F_{3,1719} = 188.62$	0.0001 $F_{2,965} = 73.16$
<b>PYRAMID</b>						
Parameter	0.2847	0.6667	0.6424	0.8333	0.6068	0.7130
$p < T$	0.0092	0.0001	0.0001	0.0001	0.0001	0.0001
$p < F$	0.0340 $F_{2,420} = 3.41$	0.0001 $F_{2,420} = 29.46$	0.0001 $F_{2,420} = 33.95$	0.0001 $F_{2,420} = 64.47$	0.0001 $F_{2,1719} = 94.89$	0.0001 $F_{2,965} = 97.01$
<b>AFFINE</b>						
Parameter	0.4852	0.2537	0.2046	0.1759	0.2799	0.0895
$p < T$	0.0001	0.0001	0.0004	0.0012	0.0001	0.0814
$p < F$	0.0001 $F_{3,420} = 144.07$	0.0005 $F_{3,420} = 6.07$	0.0007 $F_{3,420} = 5.79$	0.0045 $F_{3,420} = 4.41$	0.0001 $F_{3,1719} = 29.72$	0.2037 $F_{2,965} = 1.59$

A more detailed analysis of the effects of pyramid level and affine level within each metric level is presented in Table 5.3. To illustrate the dichotomous nature of affine level, these data do not include the A = 0 records. Without the A = 0 data, affine level was not

statistically significant even within the  $M = 0$  level ( $p < 0.0891$ ,  $F_{2,319} = 2.44$ ). These results can be compared with those of  $M = 0$  which include the  $A = 0$  data in Table 5.2. The  $M = 0$ ,  $A = 0$  condition does not allow a reliable separation of pyramids by aspect ratio ( $p < 0.0219$ ,  $F_{2,319} = 3.87$ ). This result is not surprising, as there was inadequate information to make such judgments. The only available relative depth information was the slight difference in intensity spans among pyramid aspect ratios. The effects of  $P$  and  $A$  were quite different between the  $M = 0$  and the  $M > 0$  metric levels.

Table 5.3. Analysis by metric level.  
Affine levels greater than 0.

Metric Level	0	1	2	3
<b>Pyramid</b> Param. Est.	0.3380	0.6852	0.6343	0.8194
$p < T$	0.0057	0.0001	0.0001	0.0001
$p < F$	0.0219 $F_{2,319} = 3.87$	0.0001 $F_{2,319} = 25.16$	0.0001 $F_{2,319} = 27.88$	0.0001 $F_{2,319} = 47.49$
<b>Affine</b> Param. Est.	0.2685	0.1389	0.0556	0.0741
$p < T$	0.0278	0.1502	0.5164	0.3808
$p < F$	0.0891 $F_{2,319} = 2.44$	0.3483 $F_{2,319} = 1.06$	0.8089 $F_{2,319} = 0.21$	0.5997 $F_{2,319} = 0.51$

The experimental design contains a confound. Each of the three non-zero metric levels has  $X$  and  $Y$  shear-strain offsets ( $OX$ ,  $OY$ ) that are directly proportional to the metric depth, as can be seen in Table 5.1 of the experimental design. This proportionality was intentional, as it served to maintain a constant angle for the risers of the metric structure. The use of the same

riser angle at all times served to make the number of riser dashes, and the offsets between the front and rear planes of the metric structures, effective metric factors. Thus, the offset slant over the affine space of the metric structure is directly related to the metric level.

The pyramid top is offset from the center of the pyramid bottom in the image on the screen by the slant of the metric offset. For each metric slant, the amount of offset of the pyramid top is a function of the pyramid level. The three non-zero metric levels and the three pyramid levels are each in multiples of about 1.5. Multiplying the metric slant, normalized for the level 1 slant, by the pyramid level, normalized for pyramid level 1, produces a pyramid top offset value. This is an image variable, and not a function of affine level. The relationship between the image variable and the calculated pyramid height (Eq. 5.1) is summarized in Table 5.4. The correlation between the image variable and the scaled height of the pyramid is 0.960. The height scale is logarithmic. The correlation between the log of the image variable and the scaled pyramid height is 0.998. Thus, the image variable is a reliable predictor of pyramid aspect ratio.

The image variable is a significant factor in a model for the response when analysed by an ANOVA (Table 5.5). The affine variable is dichotomous. As when the metric and pyramid variables were used, the affine variable is significant for the  $A > 0$  levels (all "No  $A = 0$ " data in Table 5.5) only when the  $M = 0$  level is included in the analysis. The  $A > 0$  levels are only significantly different in effect when there is no other basis (a non-zero metric or image variable) for scaling the pyramid height. This is consistent with a hypothesis of the equivalence of affine spaces.

Table 5.4. Normalized pyramid image versus scale height.  
H is the scaled height of the pyramid presented. Pyramid is the pyramid level; Metric is the metric level.

**Pyramid**

<b>Metric</b>	<b>1</b>	<b>H</b>	<b>2</b>	<b>H</b>	<b>3</b>	<b>H</b>
<b>1</b>	1.000	1	1.569	3	2.353	5
<b>2</b>	1.565	3	2.455	5	3.682	7
<b>3</b>	2.565	5	4.024	7	6.035	9

Table 5.5. Image variable in ANOVA model.  
The significance of the affine and pyramid variables when the image variable is included in the model. The metric factor is never significant.

**Variables**

	<b>Image</b>		<b>Affine</b>		<b>Pyramid</b>	
<b>DATA</b>	<b>F-Value</b>	<b>P &gt;</b>	<b>F-Value</b>	<b>P &gt;</b>	<b>F-Value</b>	<b>P &gt;</b>
<b>ALL</b>	153.64 F <sub>5,1715</sub>	0.0001	29.99 F <sub>3,1715</sub>	0.0001	5.75 F <sub>2,1715</sub>	0.0033
<b>No A = 0 or M = 0</b>	84.15 F <sub>4,962</sub>	0.0001	1.59 F <sub>2,962</sub>	0.2042	1.89 F <sub>1,962</sub>	0.1700
<b>No A = 0</b>	108.36 F <sub>5,1284</sub>	0.0001	3.76 F <sub>2,1284</sub>	0.0235	6.62 F <sub>2,1284</sub>	0.0014
<b>No M = 0</b>	109.65 F <sub>4,1285</sub>	0.0001	16.31 F <sub>3,1285</sub>	0.0001	1.88 F <sub>1,1285</sub>	0.1709

The ANOVA above is based on a pooling of all subject data. Each subject responded to each of the 48 conditions three times; therefore, a repeated measures ANOVA is appropriate. As can be seen in Table 5.6, this analysis does not significantly change the results. Affine level is a dichotomous variable. Affine level is not significant for those conditions in which affine and metric levels are both greater than 0.

Table 5.6. Repeated measures ANOVA.

	<b>All Data</b>	<b>Affine &gt; 0 Metric &gt; 0</b>
<b>Affine</b>	P > 0.0001 F <sub>3,22</sub> = 25.23	<b>P &gt; 0.2478</b> F <sub>2,22</sub> = 1.49
<b>Image</b>	P > 0.0001 F <sub>5,22</sub> = 129.24	P > 0.0001 F <sub>4,22</sub> = 78.66

## 5.2. INDIVIDUAL RESULTS

The shortest time within which a subject completed a block of trials was 12 minutes, 53 seconds (CCH). The longest time was 55 minutes, 34 seconds (subject SAJ). The responses of these subjects were not markedly different. A full factor ANOVA of individual subject responses to all stimulus conditions and to only those conditions in which metric and affine levels were both greater than zero is summarized in Table 5.7.

Affine level was significant ( $p < .05$ ) for 9 of the 12 subjects across all of the data records (Table 5.7). Responses of only one subject (JWK) demonstrated significance in affine level when the data for the  $M = 0$  and  $A = 0$  stimulus conditions were excluded from the analysis. One subject demonstrated a significant effect of affine level in conditions where neither metric nor affine level were zero ( $F_{2,54} = 10.07, p < .0002$ ). The regressed parameter estimate

for this subject's affine level effect, 0.3333 (T test:  $p < 0.0002$ ), was greater than for the entire subject population (0.0895,  $p < 0.08$ , N.S.) but still considerably less than the estimated parameters for this subject's metric (0.7407,  $p < 0.0001$ ) and pyramid (0.9259,  $p < 0.0001$ ) levels.

Seven of the subjects demonstrated a significant interaction ( $p < .05$ ) between metric and pyramid levels when all data were included; only three subjects showed significant interactions in the reduced subset. One subject (SAK) produced all possible interactions when all data were included, but had significance for only M, P, and the MP interaction in the reduced dataset. Two subjects (GDN, WLL) produced no affine effects at any level of the full factor model, even in the full data set.

Table 5.7. Individual subjects' responses.  
Results of an ANOVA with all factors in the model. Analysis of full data set and subset of data with no Metric = 0 or Affine = 0 stimulus records. X indicates a significance of  $p < .05$ . M = Metric, P = Pyramid, A = Affine; MP = Metric\*Pyramid, MA = Metric\*Affine, PA = Pyramid\*Affine, and MPA = Metric\*Pyramid\*Affine interactions.

Subject	All Data							No Metric = 0 or Affine = 0 Records						
	M	P	MP	A	MA	PA	MPA	M	P	MP	A	MA	PA	MPA
CCH	X	X	X	X		X		X	X					
CSJ	X	X	X	X				X	X					
GDN	X	X	X					X	X					
JSK	X	X		X				X	X					
JWK	X	X	X	X			X	X	X		X			
MBH	X	X	X	X	X			X	X	X				
RLD	X	X		X	X			X	X					
RNJ	X	X				X		X	X					
SAC	X	X	X	X				X	X	X				
SAJ	X	X		X				X	X					
SAK	X	X	X	X	X	X	X	X	X	X				
WLL	X	X						X	X					

## 6. DISCUSSION

### 6.1. CONCLUSIONS

The hypothesis in this dissertation is that the magnitude of the affine shift will not be the major factor in determining the subject's perception of depth. Instead, affine structures will be independently scaled by metric factors. This hypothesis was reflected in the results. For the full set of data, which included both affine ( $A = 0$ ) and metric ( $M = 0$ ) conditions, the affine variable was significant. When the data for the  $A = 0$  and  $M = 0$  conditions were removed, leaving three different non-zero affine levels, the affine variable was no longer significant, while the metric variable remained significant. The results support the hypothesis. The presence of any affine depth produced by SFM improves the perception of depth.

The pyramid was scaled by the metric factors relative to the entire affine depth span. Therefore, if the affine space was reliably scaled for depth, subjects should be able to differentiate among the pyramid levels. When  $M > 0$ , they could. However, when  $M = 0$ , subjects made much weaker and less accurate discriminations among pyramid levels. When  $M = 0$ , the regression coefficient for P was small (0.2847); when  $M > 0$ , the coefficient was considerably larger (0.7141). The pyramid (P) variable was significant in both cases. In the  $M = 0$  condition, information is inadequate to scale the space. Evidently, subjects were influenced by the size of the affine SFM movement in making judgments as reflected by the higher affine regression coefficient (for  $M = 0$ , coeff. = 0.4852; for  $M > 0$ , coeff. = 0.2054).

The experimental display was designed to produce strong metrics, using both the surrounding structure and the shear-strain offset to provide metric information. The product of the metric offset variable and pyramid variable created a confounding image metric variable which does not allow us to differentiate between the image metric and structure metric factors. The combined metric factors are significant.

In this experiment a constant depth slant angle was maintained to provide a consistent metric system. Norman and Todd (1993) found that affine changes in the picture plane produced perceived distortions in the structure, although changes in the affine depth did not. Although the principle of affine equivalence is presumed to apply in all directions, sensitivity to changes in affine structure is not the same in all directions. When taking one's seat in the movie theater at the front row, side aisle, it takes a while to become accustomed to the affine distortions.

Tod and Norman's (1993) finding of sensitivity to affine transformations in the picture plane is at variance with an observation with the affine SFM display of this experiment. When the psuedosaccade shifts were operating, one can move  $75^\circ$  off of the normal to the screen without a perceived distortion of the structure. Yet such a motion does produce a large affine transformation of the retinal image.

The purpose of this dissertation was to investigate the relative effectiveness of affine level and metric level in the perception of spatial structure in displays. Clearly the roles of metric and affine variables differ. The metric level is more important than affine level in controlling the perception of the scale of a spatial structure. Any non-zero affine level enhances the perception of spatial structure, but does not control its metrification. The dichotomous effect of the affine variable supports the hypothesis that non-zero affine levels are equivalent. Affine level alone, in the absence of metric factors, was not adequate to produce a reliable scaling of depth. The perception of spatial structure is non-Euclidean; there is no disparity-based scale for depth.

The model of the perception of metric structure (Fig. 6.1) parallels the intial vectorfield-scalar field model of Fig. 1.1. The significance is that the perception of metric structure involves an intermediate perception of affine structure which is subsequently metrified. The use of an *affine intermediate* is conceptually simpler than attempting to derive metric structure

from disparity information alone. It allows the CNS to perceive a structure independent of its location relative to the observer, given adequate information to scale the affine structure. This emancipates the observer from the computational difficulties inherent in the perception of structure from an egocentric viewpoint. One result of this perceptual process is the ability to perceive metric structure correctly in motion pictures viewed from the "front row, side aisle" (Cutting, 1987) and in pictures viewed at a slant (Cutting, 1990).

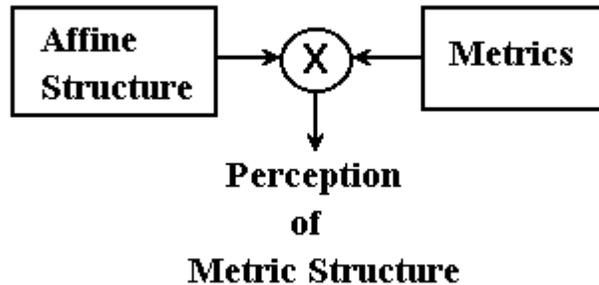


Figure 6.1. Perception of metric structure.

That the standard deviations of the subjects' responses to the non-zero metric stimuli were smaller than the standard deviations of the stimuli themselves indicates a tendency to collapse responses to some mean value. This follows from the very small number of "zero" responses. Both of these results can be considered as analogous to a specific distance tendency (Gogel, 1977).

When the metric level and/or the affine level are 0, the information does not adequately specify a pyramid aspect ratio, yet subjects' responses indicate a non-zero perception of depth. The stimuli had intensity modulation, being darker with greater depth. The 63 levels of gray

scale in each color produced a perceptually continuous shading of lines that receded in depth, such as the edges of the pyramid. The intensity span was constant for the affine depth and did not vary for different metric levels. Thus, intensity provides a very weak differentiation among the affine pyramids. Shading is not generally a strong factor for the perception of metric shape, however (Erens, Kappers, & Koenderink, 1993; Bülthoff & Mallot, 1988).

One cannot analyze the results of one particular experimental condition separately from the context of the entire experiment. The  $M = 0$  condition was presented in the context of other pyramids which had clearly perceived depths. Subjects were predisposed to perceive pyramids as opposed to flat figures.

The original model of the perception of spatial structure was the product of a vectorfield and a scalar field (Fig. 1.1). If either vectorfield or scalar field is 0, then the perceived depth of a structure would be zero. Clearly, when either  $A = 0$  or  $M = 0$ , the subjects' responses were not zero. I propose that a zero affine level is not the same as a zero vectorfield, and that a zero metric level is not the same as a zero scalar field.

It is useful to ask "What would a zero level of vectorfield or scalar field mean?" The model of the CNS used here is a properties model, as opposed to a computational model. Such a model is based on the physiological activity of neuropile. Physiological activity has a base level of randomness, i.e., of noise. This noise is a significant determinant of the threshold level of detection. A zero vectorfield or scalar field would imply a zero noise level. This is not possible. The threshold field level is, therefore, non-zero. The threshold level determines the effective residual field level, which is not zero. Thus, one can envision a non-zero ***threshold vectorfield*** and a non-zero ***threshold scalar field***. The vectorfield and scalar field are apertures.

In this experiment, the experience of viewing pyramids leaves a residual noise pattern in the CNS consistent with a pyramidal structure. As What and Where are separately located in

the CNS, this pattern of residual activity, superimposed on a random physiological noise level, provides a residual vectorfield and/or scalar field. Thus, the  $M = 0$  and  $A = 0$  experimental conditions do not result in the perception of a flat surface. If the surface of the display screen were apparent due to dirt or glare, a zero metric field would be generated, probably suppressing the perception of depth to some degree.

From a theoretical standpoint the perception of spatial structure is different under "reduced cue" conditions that do not provide adequate information for determining spatial structure. Although the principal focus of this dissertation is the perception of spatial structure from adequate information, it is interesting to note that subjects did perceive depth when information was inadequate, and therefore that the effects of the affine factor are quite different from its effects when adequate information is available. This summary is consistent with Gibson's (1950) contention that one cannot draw conclusions about the nature of perception from conditions that do not provide adequate information. In this dissertation I have investigated how adequate information is used to perceive spatial structure. This information can be directly applied to the display of spatial structures.

## 6.2. THE AFFINE INTERMEDIATE IN DISPLAYS

Effective spatial displays can be developed by using an intermediate affine structure stage. The design of a 3-dimensional display can be achieved in three steps:

1. define the affine display space
2. map the spatial structure(s) into the affine space
3. develop metrics to scale the affine structure.

The affine space is defined in terms of the disparity that will generate it. In the proposed shear-strain SFM display, the affine space is defined by the front and back planes of maximum shift. In a stereoscopic display, it is defined by the maximum crossed and uncrossed stereo disparities. Note that effective eye separation is not the principal variable. The unimportance of the

amount of eye spacing in stereoscopic displays is a result of the affine equivalence principle, whereby stereo disparity and convergence do not provide metric information in the presence of other metric information. A stereo display is equivalent to the  $A > 0$  condition in the experiment. The affine space is then scaled linearly by metric factors within these limits. The spatial structure is mapped into the affine space by normalizing the depth to the limits of the affine space, thus causing the spatial structure to span the entire affine depth.

The affine structure is then scaled through metric factors. These have been described in Section 1, above. They are:

- Sequencing
- Ordination
- Cardination
- Metrification
- Absolute Scaling

Frequently displays do not have any affine structure *per se* but must rely on the non-zero threshold vectorfield implied from the random physiological noise in the CNS. In other words, if the display is a flat screen with no disparity, strong image characteristics which define its spatial structure must be used for an effective display. Using a strong metric system in a zero affine, or zero disparity, display will produce the perception of spatial structure, since the metric scalar field will interact with the non-zero threshold vectorfield to produce a perceived non-zero metric structure. A display which does not incorporate all levels of metric scaling is apt to provide an unreliable perception of spatial structure and should be avoided. Thus the list above provides a checklist for evaluating a proposed display. Any missing metric levels will degrade the reliability of the perception of structure in the display. A display with affine structure but no metric structure is ambiguous and will lead to a wide variety of perceptions of spatial structure by the viewer. A display with inadequate continuity in the structural

hierarchy can also produce misperceptions. The experimental display applied the metric to the display space, but not to the pyramid directly. This reduced cue condition allowed some subjects to occasionally see the pyramid as in front of the metric structure. This lack of tight coupling between the metric structure and the pyramid may also be partially responsible for the under-reporting of the range of pyramid heights.

The importance of display metrics has been clearly demonstrated. A method for creating an affine structure that is scaled by those metrics in a display has been demonstrated.

### 6.3. TOPICS FOR FUTURE RESEARCH

Research invariably raises more questions than it answers. The parameters for the pseudosaccade were selected for appearance. A more complete study of the optimal base time as a function of display complexity, and of the optimal number and spatial distribution of aperture positions would be fruitful. The pseudosaccades are based on affine shifts—what would be the effect of simultaneous animation of the display? As animation would be asynchronous and thus incoherent with the shifts: the two should proceed without conflict. Preliminary studies indicate this to be the case.

The perception of spatial structure in the affine SFM display is quite robust. This perception persists even when the display is viewed off-axis. Moving one's viewpoint 75° off-axis does not produce a distortion in the perceived structure. Instead, the orientation of the structure relative to the display surface appears to change. This striking and unusual effect bears further investigation for the information it may provide about visual processes. This effect is consistent with a model of visual spatial perception via an affine intermediate.

Although not as useful as a display, *per se*, an exploration of the independent metrification is of interest. This could be achieved by populating the display volume with randomly oriented elements. For instance, randomly oriented line segments of equal lengths could define the volume without producing the more effective cubic structures used in this experiment.

These metric elements would not require the addition of a gross shear-strain offset to the space, removing the image confound.

Perspective is an promising area for further study. What is its effect? Which perspective transformation is appropriate? As the depth extent of a display is increased, a larger perspective transformation in mapping Euclidean space into the affine space is required. The perspective transformation is inherently nonaffine, potentially reducing the effectiveness of an affine shear-strain SFM display. It will be interesting to find the depth of a display structure that produces conflict between the requirements for affine transformations and the requirements for perspective. One can presume that the conflict will result in the perception of a nonrigid structure.

These areas for future study have a common theme: the parallel growth in understanding of the processes of perception and the development of effective interfaces between humans and technology. This is a valuable collaboration of interests. The power of knowledge is its universality of applicability.

## 7. REFERENCES

- Annino R., & Driver, R. D. (1986). Computer graphics. *Scientific and engineering applications with personal computers*. New York: John Wiley & Sons.
- Bell, H. B. & Lappin, J. S. (1973). Sufficient conditions for the discrimination of motion. *Perception & Psychophysics*, *14*, 45-50.
- Bennett, B. M., Hoffman, D. D., Nicola, J. E., & Prakash, C. (1989). Structure from two orthographic views of rigid motion. *Journal of the Optical Society of America A*, *6*, 1052-1069.
- Bittencourt, P. R., Smith, A. T., Lloyd, D. S., & Richens, A. (1982). Determination of smooth pursuit eye movement velocity in humans by computer. *Electroencephalography & Clinical Neurophysiology*, *54*, 399-405.
- Bracewell, R. N. (1990). Numerical transforms. *Science*, *248*, 697-704.
- Braunstein, M. L., Hoffman, D. D., & Pollick, F. E. (1990). Discriminating rigid from nonrigid motion: minimum points and views. *Perception & Psychophysics*, *47*, 205-214.
- Braunstein, M. L., Hoffman, D. D., Shapiro, L. R., Andersen, G. J., & Bennett, B. M. (1987). Minimum points and views for the recovery of three-dimensional structure. *Journal of Experimental Psychology: Human Perception & Performance*, *13*, 335-343.
- Bülthoff, H. H., & Mallot, H. A. (1988). Integration of depth modules: Stereo and shading. *Journal of the Optical Society of America A*, *5*, 1749-1758.
- Burbeck, C. A. (1992). Separation discrimination with embedded targets. *Vision Research*, *32*, 2295-2302.
- Burr, D. C. (1979). Acuity for apparent vernier offset. *Vision Research*, *19*, 835-837.
- Burr, D. C. (1980). Motion smear. *Nature*, *284*, 164-165.
- Burr, D. C. (1980). Sensitivity to spatial phase. *Vision Research*, *20*, 391-396.

- Burt, P., & Julesz, B. (1980). Modifications of the classical notion of Panum's fusional area. *Perception, 9*, 671-682.
- Busey, T. A., Brady, N. P., & Cutting, J. E. (1990). Compensation is unnecessary for the perception of faces in slanted pictures. *Perception & Psychophysics, 48*, 1-11.
- Chang, J. J., & Julesz, B. (1983a). Displacement limits, directional anisotropy and direction versus form discrimination in random-dot cinematograms. *Vision Research, 23*, 639-646.
- Chang, J. J. & Julesz, B. (1983b). Displacement limits for spatial frequency filtered random-dot cinematograms in apparent motion. *Vision Research, 23*, 1379-1385.
- Cornilleau-Pérès, V., & Droulez, J. (1993). Stereo-motion cooperation and the use of motion disparity in the visual perception of 3-D structure. *Perception & Psychophysics, 54*, 223-239.
- Covington, M. A. (1990). Smooth views. *Byte, 15*, 279-283.
- Crandall, S. H., & Dahl, N. C. (1959). Definition of strain components. *An introduction to the mechanics of solids*. New York: McGraw-Hill.
- Crombie, A. C. (1964). Early concepts of the senses and the mind. *Scientific American*, May 1964. Reprint #184. San Francisco: W. H. Freeman.
- Curcio, C. A., Sloan, Jr., K. R., Packer, O., Hendrickson, A. E., & Kalina, R. E. (1987). Distribution of cones in human and monkey retina: individual variability and radial asymmetry. *Science, 236*, 579-582.
- Cutting, J. E. (1987). Rigidity in cinema seen from the front row, side aisle. *Journal of Experimental Psychology: Human Perception & Performance, 13*, 323-334.
- Cutting, J. E. (1991). On the efficacy of cinema, or what the visual system did not evolve to do. In Ellis, S. R. (Ed.), *Pictorial communication in virtual and real environments* (pp. 486-495). London: Taylor & Francis.

- Daubechies, I., Mallat, S., & Willsky, A. (1992). Introduction: Wavelet transforms and multiresolution signal analysis. [Special issue]. *IEEE Transactions on Information Theory*, **38**, 529-531.
- DeLucia, P. R., & Hochberg, J. (1991). Geometrical illusions in solid objects under ordinary viewing conditions. *Perception & Psychophysics*, **50**, 547-554.
- Diamond, A. L. (1979). Microsecond sensitivity of the human visual system to irregular flicker. *Science*, **206**, 708-710.
- Dick, M., Ullman, S., & Sagi, D. (1987). Parallel and serial processes in motion detection. *Science*, **237**, 400-402.
- Dodwell, P. C. (1983). The Lie transformation group model of visual perception. *Perception & Psychophysics*, **34**, 1-16.
- Doner, J., Lappin, J. S., & Perfetto, G. (1984). Detection of three-dimensional structure in moving optical patterns. *Journal of Experimental Psychology: Human Perception & Performance*, **10**, 1-11.
- Dowling, J. E. (1987). *The retina: an approachable part of the brain*. Cambridge, Massachusetts: The Belknap Press of Harvard University Press.
- Eby, D. W. (1992). The spatial and temporal characteristics of perceiving 3-D structure from motion. *Perception & Psychophysics*, **51**, 163-178.
- Ellis, S. R., Smith, S., Grunwald, A., & McGreevy, M. W. (1991). Direction judgement error in computer generated displays and actual scenes. In Ellis, S. R. (Ed.), *Pictorial communication in virtual and real environments* (pp. 505-526). London: Taylor & Francis.
- Ellis, S. R., Smith, S., & Hacısalihzade, S. (1989). Visual directions as a metric of virtual space. In *Proceedings of the 33rd Annual Meeting of the Human Factors Society*.
- Erens, R. G. F., Kappers, A. M. L., & Koenderink, J. J. (1993). Perception of local shape from shading. *Perception & Psychophysics*, **54**, 145-156.

- Foley, J. T. (1991). Effect of an aperture on the spectrum of partially coherent light. *Journal of the Optical Society of America A*, *8*, 1099-1105.
- Friedhoff, R. M. & Benzou, W. (1989). *Visualization*. New York: Harry N. Abrams.
- Gagliardi, R. M., & Karp, S. (1976). *Optical communications*. New York: John Wiley & Sons.
- Gazzaniga, M. S. (1989). Organization of the human brain. *Science*, *245*, 947-952.
- Gibson, J. J. (1950). *The perception of the visual world*. Boston, Massachusetts: Houghton-Mifflin.
- Gogel, W. C. (1965). Equidistance tendency and its consequences. *Psychological Bulletin*, *64*, 153-163.
- Gogel, W. C. (1977). The metric of visual space. In W. Epstein (Ed.), *Stability and constancy in visual perception: Mechanisms and processes* (pp. 129-181). New York: John Wiley & Sons.
- Gogel, W. C. (1990). A theory of phenomenal geometry and its applications. *Perception & Psychophysics*, *48*, 105-123.
- Goodman, J. W. & Russell, F. D. (1971). Non-redundant arrays and postdetection processing for aberration compensation in incoherent imaging. *Journal of the Optical Society of America A*, *61*, 182-191.
- Gray, J. (1989). *Ideas of space: Euclidean, non-Euclidean and relativistic* (2nd ed.). Oxford: Clarendon Press.
- Green, M., & Odom, J. V. (1986). Correspondence matching in apparent motion: evidence for three-dimensional spatial resolution. *Science*, *233*, 1427-1429.
- Green, D. M., Richards, V. M., & Onsan, Z. A. (1990). Sensitivity to envelope coherence. *Journal of the Acoustical Society of America*, *87*, 323-329.

- Guth, S. L. (1991). Model for color vision and light adaptation. *Journal of the Optical Society of America A*, *8*, 976-993.
- Gwane, T. J., McClurkin, J. W., Optican, L. M., & Richmond, B. J. (1988). Lateral geniculate nucleus neurons in awake behaving primates. III: Successful predictions of a multi-channel model. *Abstracts, Society for Neuroscience, 18th Annual Meeting*, #127.10, 309.
- Harvey, L. O., Jr., Rentschler, I., & Weiss, C. (1985). Sensitivity to phase distortions in central and peripheral vision. *Perception & Psychophysics*, *38*, 392-396.
- Henning, B. G. (1988). Spatial-frequency tuning as a function of temporal frequency and stimulus motion. *Journal of the Optical Society of America A*, *5*, 1362-1373.
- Henning, B. G. & Gaskell, H. (1981). Monaural phase sensitivity with Ronken's paradigm. *Journal of the Acoustical Society of America*, *70*, 1669-1673.
- Hess, R. F., Pointer, J. S., & Watt, R. J. (1989). How are spatial filters used in the fovea and parafovea? *Journal of the Optical Society of America A*, *6*, 329-339.
- Hildreth, E. C., Grzywacz, N. M., Adelson, E. H., & Inada, V. K. (1990). The perceptual buildup of three-dimensional structure from motion. *Perception & Psychophysics*, *48*, 19-36.
- Hoffman, D. D., & Bennett, B. M. (1985). Inferring the relative three-dimensional positions of two moving points. *Journal of the Optical Society of America A*, *2*, 350-353.
- Hoffman, D. D., & Bennett, B. M. (1986). The computation of structure from fixed axis motion: Rigid structures. *Biological Cybernetics*, *54*, 71-83.
- Hogben, J. H., & Di Lollo, V. (1985). Suppression of visible persistence in apparent motion. *Perception & Psychophysics*, *38*, 450-460.
- Hubel, D. H. & Wiesel, T. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, *160*, 106-154.

- Julesz, B. (1971). *Foundations of Cyclopean perception*. Chicago: University of Chicago Press.
- Kaas, J. H., Krubitzer, L. A., Chino, Y. M., Langston, A. L., Polley, E. H., & Blair, N. (1990). Reorganization of retinotopic cortical maps in adult mammals after lesions of the retina. *Science*, **248**, 229-231.
- Kaiser, M. K. & Proffitt, D. R. (1992). Using the stereokinetic effect to convey depth: Computationally efficient depth-from-motion displays. [Special Issue: Visual Displays]. *Human Factors*, **34**, 571-581.
- Katz, B., & Schmitt, O. (1940). Electric interaction between two adjacent nerve fibers. *Journal of Physiology*, **97**, 471-488.
- Kaufman, L. (1974). *Sight and mind: An introduction to visual perception*. Oxford: Oxford University Press.
- Kelly, D. H. (1969). Flickering patterns and lateral inhibitions. *Journal of the Optical Society of America A*, **59**, 1361-1370.
- Kelly, D. H. (1981). Disappearance of stabilized chromatic gratings. *Science*, **214**, 1257-1258.
- Kelly, D. H. (1990). Moving gratings and microsaccades. *Journal of the Optical Society of America A*, **7**, 2237-2244.
- Killing, W. (1892). Über die grundlagen der geometrie. *Journal für die reine und angewandte Mathematik*, **109**, 128.
- Klopfer, D. S. (1991). Apparent reversals of a rotating mask: a new demonstration of cognition in perception. *Perception & Psychophysics*, **49**, 522-530.
- Koenderink, J. J., & Van Doorn, A. J. (1991). Affine structure from motion. *Journal of the Optical Society of America A*, **8**, 377-385.
- Korn, G. A., & Korn, T. M. (1968). *Mathematical handbook for scientists and engineers* (2nd ed.). New York: McGraw-Hill.

- Koskol, J. B. (1991). Effect of spatiotemporal jitter on time-sequentially sampled imagery. *Journal of the Optical Society of America A*, *8*, 883-892.
- Krekling, S. (1973). Some aspects of the Pulfrich effect. *Scandinavian Journal of Psychology*, *14*, 87-90.
- Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, *16*, 37-68.
- Kulikowski, J. J. (1978). Limit of single vision in stereopsis depends on contour sharpness. *Nature*, *275*, 126-127.
- Lamb, T. D., & Pugh, E. N., Jr. (1990). Physiology and adaptation in rod and cone photoreceptors. *Seminars in the Neurosciences*, *2*, 3-13.
- Lappin, J. S. (1990). Perceiving the metric structure of environmental objects from motion and stereopsis. In Warren, R. & Wertheim, A. H. (Eds.), *Perception and the control of self-motion* (pp. 541-547). Hillsdale, New Jersey: Erlbaum.
- Lappin, J. S. (in press). Seeing structure in space-time. Publication in honor of G. Johansson's 80th birthday.
- Lappin, J. S., & Bell, H. H. (1976). The detection of coherence in moving random-dot patterns. *Vision Research*, *16*, 161-168.
- Lappin, J. S., Doner, J. F., & Kottas, B. (1980). Minimal conditions for the visual detection of structure and motion in three dimensions. *Science*, *209*, 717-719.
- Lappin, J. S., & Love, S. R. (1992). Planar motion permits perception of metric structure in stereopsis. *Perception & Psychophysics*, *51*, 86-102.
- Lappin, J. S. & Wason, T. D. (1987). The perception of geometrical structure from congruence. In Ellis, S., Kaiser, M., & Grunwald, A. (Eds.), *Spatial displays and spatial instruments* (NASA Conference Publication 10032, pp. 18-1 - 18-15). Washington, D.C.: National Aeronautics and Space Administration.

- Lappin, J. S., & Wason, T. D. (1991). The perception of geometric structure from congruence. In Ellis, S. R., & Kaiser, M. K., *Pictorial communications in virtual and real environments* (pp. 425-448). London: Taylor & Francis.
- Lappin, J. S., Wason, T., & Akutsu, H. (1987). Visual detection of common motion of spatially separate points. *Bulletin of the Psychonomic Society*, **25**, 343.
- Lasaga, M. I., & Hecht, H. (1991). Interactions of local features as a function of global goodness and spacing. *Perception & Psychophysics*, **49**, 201-211.
- Lee, B. B., Pokorny, J., Smith, V. C., Martin, P. R., & Valberg, A. (1990). Luminance and chromatic modulation sensitivity of macaque ganglion cells and human observers. *Journal of the Optical Society of America A*, **7**, 2223-2236.
- Livingston, M. & Hubel, D. (1988). Segregation of form, color, movement, and depth: anatomy, physiology, and perception. *Science*, **240**, 740-749.
- Loeb, G. L., White, M. W., & Merzenich, M. M. (1983). Spatial cross-correlation. *Biological Cybernetics*, **47**, 149-163.
- Luckiesh, M. (1922). *Visual illusions: Their causes, characteristics and applications*. Reprint. New York: Dover Publications, 1965.
- Malik, J. & Perona, P. (1990). Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America A*, **7**, 923-932.
- Mallat, S. (1991). Zero crossings of a wavelet transform. *IEEE Transactions on Information Theory*, **37**, 1019-1033.
- Mallat, S., & Hwang, W. (1992). Singularity detection and processing with wavelets. *IEEE Transactions on Information Theory*, **38**, 617-643.
- Mallat, S., & Zhong, S. (1992). Characterization of images from multiscale edges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **14**, 710-732.
- Marr, D. (1982). *Vision*. San Francisco: W. H. Freeman & Co.

- Mates, J. W. B. (1978). Eye movements of African chameleons: Spontaneous saccade timing. *Science*, **199**, 1087-1089.
- Maunsell, J. H. R., Nealy, T. A., & DePriest, D. D. (1990). Magnocellular and parvocellular contributions to responses in the middle temporal visual area (MT) of the Macaque monkey. *Journal of Neuroscience*, **10**, 3323-3334.
- McClurkin, J. W., Gwane, T. J., Richmond, B. J., Optican, L. M., & Robinson, D. L. (1988). Lateral geniculate nucleus neurons in awake behaving primates. I: Responses to B&W 2-D patterns. *Abstracts, Society for Neuroscience, 18th Annual Meeting*, #127.8, 309.
- McClurkin, J. W., Optican, L. M., Richmond, B. J., & Gwane, T. J. (1991). Concurrent processing and complexity of temporally encoded neuronal messages in visual perception. *Science*, **253**, 675-677.
- Mead, C. (1989). *Analog VLSI and neural systems*. Reading, Massachusetts: Addison-Wesley.
- Meister, M., Wong, R. O. L., Baylor, D. A., & Shatz, C. J. (1991). Synchronous bursts of action potentials in ganglion cells of developing mammalian retina. *Science*, **252**, 939-943.
- Mignard, M., & Malpeli, J. G. (1991). Paths of information flow through the visual cortex. *Science*, **251**, 1249-1251.
- Mingolla, E., Todd, J. T., & Norman, J. F. (1992). The perception of globally coherent motion. *Vision Research*, **32**, 1015-1031.
- Mirsky, R., & Jessen, K. R. (1990). Schwann cell development and the regulation of myelination. *Seminars in the Neurosciences*, **2**, 423-435.
- Moore, B. C. J., Glasberg, B. R., & Schooneveldt, G. P. (1990). Across-channel masking and comodulation masking release. *Journal of the Acoustical Society of America*, **87**, 1683-1694.
- Morgan, M. J. (1976). Pulfrich effect and the filling in of apparent motion. *Perception*, **5**, 187-195.

- Morgan, M. J., & Thompson, P. (1975). Apparent motion and the Pulfrich effect. *Perception*, **4**, 3-18.
- Morgan, M. J., Ward, R. M., & Hole, G. J. (1990). Evidence for positional coding in hyperacuity. *Journal of the Optical Society of America A*, **7**, 297-304.
- Mowafy, L., Blake, R., & Lappin, J. S. (1990). Detection and discrimination of coherent motion. *Perception & Psychophysics*, **48**, 583-592.
- Musatti, C. L. (1924). Sui fenomeni stereocinetici [On the phenomenon of a stereokinetic effect]. As in Proffitt, D. R., Rock, I., Hecht, H., & Schubert, J., (1992) *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 3-21.
- Navon, D. (1983). Preservation and change of hue, brightness, and form in apparent motion. *Bulletin of the Psychonomic Society*, **21**, 131-134.
- Nawrot, M. & Blake, R. (1989). Neural integration of information specifying structure from stereopsis and motion. *Science*, **244**, 716-718.
- Nawrot, M. & Blake, R. (1991). The interplay between stereopsis and structure from motion. *Perception & Psychophysics*, **49**, 230-244.
- Nobili, R. (1987). Ionic waves in animal tissues. *Physics Review A*, **35**, 1901-1922.
- Norman, F. J., & Todd, J. T. (1993). The perceptual analysis of structure from motion for rotating objects undergoing affine stretching transformations. *Perception & Psychophysics*, **53**, 279-291.
- Okoshi, T. (1976). *Three-dimensional imaging techniques*. New York: Academic Press.
- Ono, H., & Steinbach, M. J. (1990). Monocular stereopsis with and without head movement. *Perception & Psychophysics*, **48**, 179-187.
- Petersik, J. T. (1987). Recovery of structure from motion: Implications for a performance theory based on the structure-from-motion theorem. *Perception & Psychophysics*, **42**, 355-364.

- Petersik, J. T., & Rosner, A. (1990). The effect of position cues on the appearance of stimulus elements in a bistable apparent movement display. *Perception & Psychophysics*, *48*, 280-284.
- Poggio, T., & Girosi, F. (1990). Regularization algorithms for learning that are equivalent to multilayer networks. *Science*, *247*, 978-982.
- Proffitt, D. R., & Kaiser, M. K. (1991). Perceiving environmental properties from motion information: minimal conditions. In Ellis, S. R., & Kaiser, M. K., (Eds.), *Pictorial communications in virtual and real environments* (pp. 45-60). London: Taylor & Francis.
- Proffitt, D. R., Rock, I., Hecht, H., & Schubert, J. (1992). Stereokinetic effect and its relation to the kinetic depth effect. *Journal of Experimental Psychology: Human Perception & Performance*, *18*, 3-21.
- Ratliff, F. (1965). *The Mach bands: Quantitative studies on neural networks in the retina*. New York: Holden-Day.
- Reichart, W. (1971). Autocorrelation: A principle for the evaluation of sensory information by the central nervous system. In Dodwell, P. C. (Ed.), *Perceptual Processing* (pp. 130-144). New York: Appleton-Century-Crofts.
- Reynolds, G. & Skinner, T. (1964). Mutual coherence function applied to imaging through a random medium. *Journal of the Optical Society of America A*, *54*, 1302-1309.
- Richards, W. (1985). Structure from stereo and motion. *Journal of the Optical Society of America A*, *2*, 343-349.
- Richmond, B. J., McClurkin, J. W., Gwane, T. J., & Optican, L. M. (1988). Lateral geniculate nucleus neurons in awake behaving primates. II: Temporal modulation is more important in LGN neurons than ganglion cell fibers. *Abstracts, Society for Neuroscience, 18th Annual Meeting*, #127.9, 309.
- Robson, J. G. (1966). Spatial and temporal contrast sensitivity functions of the visual system. *Journal of the Optical Society of America*, *56*, 1141-1142.

- Ruelle, D. (1989). *Chaotic evolution and strange attractors*. Cambridge: Cambridge University Press.
- Sagi, D. & Julesz, B. (1985). "Where" and "what" in vision. *Science*, **228**, 1217-1219.
- Saidpour, A., Braunstein, M. L., & Hoffman, D. D. (1992). Interpolation in structure from motion. *Perception & Psychophysics*, **51**, 105-117.
- Saleh, B. E. A., Tulunay-Kessy, U., Ver Hoeve, J. N., & Hom, M. (1991). Dynamics of adaptation for vision with a stabilized image. *Journal of the Optical Society of America A*, **8**, 1172-1181.
- Sary, Vogels, & Orban (1993). Shape responses in the inferior temporal cortex. *Science*, **260**, 995-997.
- Schneider, B., & Moraglia, G. (1992). Binocular unmasking with unequal interocular contrast: the case for multiple Cyclopean eyes. *Perception & Psychophysics*, **52**, 639-660.
- Shapley, S. & Victor, J. (1986). Hyperacuity in cat retinal ganglion cells. *Science*, **231**, 999-1002.
- Shou, T., & Leventhal, A. G. (1989). Organized arrangement of orientation-sensitive relay cells in the cat's dorsal lateral geniculate nucleus. *Journal of Neuroscience*, **9**, 4287-4302.
- Snyder, A. W., Bossomaier, T. R. J., & Hughes, A. (1986). Optical image quality and the cone mosaic. *Science*, **231**, 499-501.
- Stanford, L. R. (1987). Conduction velocity variations minimize conduction time differences among retinal ganglion cell axons. *Science*, **238**, 358-360.
- Steinberg, B. S. (1976). *Principles of aperture & array system design: Including random and adaptive arrays*. New York: John Wiley & Sons.
- Stewart, I. (1989). *Does God play dice?: The mathematics of chaos*. Oxford: Basil Blackwell.

- Stigmar, G. (1971). Blurred visual stimuli. II: The effect of blurred visual stimuli on vernier and stereo acuity. *Acta Ophthalmologica*, *49*, 364-379.
- Stoner, G. R., & Albright, T. D. (1993). Image segmentation cues in motion processing: Implications for modularity in vision. *Journal of Cognitive Neuroscience*, *5*, 129-149.
- Stork, D. G. & Wilson, H. R. (1990). Do Gabor functions provide appropriate descriptions of visual cortical receptive fields? *Journal of the Optical Society of America A*, *7*, 1362-1372.
- Synge, J. L., & Schild, A. (1949). *Tensor Calculus*. Reprint. New York: Dover Publications (corrected ed.), 1978.
- Tittle, J. S., & Braunstein, M. L. (1993). Recovery of 3-D shape from binocular disparity and structure from motion. *Perception & Psychophysics*, *54*, 157-169.
- Tobias, L., Volckers, U., & Erzberger, H. (1989). *Controller evaluation of the descent advisor automation aid*. (NASA Technical Memorandum, #102197). Washington, D.C.: National Aeronautics and Space Administration.
- Todd, J. T. (1984). The perception of three-dimensional structure from rigid and nonrigid motion. *Perception & Psychophysics*, *36*, 97-103.
- Todd, J. T., & Bressan, P. (1990). The perception of 3-Dimensional affine structure from minimal apparent motion sequences. *Perception & Psychophysics*, *48*, 419-430.
- Toet, A. (1987). *Visual perception of spatial order*. Unpublished doctoral dissertation, University of Utrecht, Holland.
- Uehara, F., Matthes, M. T., Yasumura, D., & LaVail, M. M. (1990). Light-evoked changes in the interphotoreceptor matrix. *Science*, *248*, 1633-1636.
- Van Essen, D. C., Anderson, C. H., & Felleman, D. J. (1992). Information processing in the primate visual system: An integrated systems approach. *Science*, *255*, 419-423.
- Wallach, H., & Centrella, N. M. (1990). Identity imposition and its role in a stereokinetic effect. *Perception & Psychophysics*, *48*, 535-542.

- Wallach, H., & O'Connell, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology*, **45**, 205-217.
- Waxman, S. G. (1980). Determinants of conduction velocity in myelinated nerve fibers. *Muscle Nerve*, **3**, 141-150.
- Wehrhahn, C., & Rapf, D. (1992). ON- and OFF-pathways form separate neural substrates for motion perception: Psychophysical evidence. *Journal of Neuroscience*, **12**, 2247-2250.
- Wertheimer, M. (1938). *A Source Book of Gestalt Psychology*. (W. D. Ellis, Trans.). London: Routledge & Kegan Paul.
- Westheimer, G. (1975). Visual acuity and hyperacuity. *Investigative Ophthalmology and Visual Science*, **14**, 570.
- Weyl, H. (1918/1921). *Space Time Matter* (4th ed.), (H. L. Brose, Trans.) Reprint. New York: Dover Publications, 1952.
- Wiesel, T. N., & Hubel, D. H. (1965). *Journal of Neurophysiology*, **28**, 1029.
- Williams, J. M., & Lit, A. (1983). Luminance-dependent visual latency for the Hess effect, the Pulfrich effect, and simple reaction time. *Vision Research*, **23**, 171-179.
- Wilson, F. A. W., O'Scalaidhe, S. P., & Goldman-Rakic, P. S. (1993). Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science*, **260**, 1955-1958.
- Wilson, J. A., & Anstis, S. M. (1969). Visual delay as a function of luminance. *American Journal of Psychology*, **82**, 350-358.
- Wylie, C. R., & Barrett, L. C. (1982). *Advanced engineering mathematics*. New York: McGraw-Hill.
- Xu Jing-Hua, & Li Wei (1986). The dynamics of large scale neuron-glia network and its relation to the brain functions (I). *Communications in Theoretical Physics* (Beijing, China), **5**(4), 339-346.

Yang, X., & Wu, S. M. (1989). Modulation of rod-cone coupling by light. *Science*, *244*, 352-354.

Yeh, Y. Y., & Silverstein, L. D. (1992). Spatial judgments with monoscopic and stereoscopic presentation of perspective displays. *Human Factors*, *34*, 583-600.

## 8. APPENDIX

### PROGRAMMED INSTRUCTIONS TO SUBJECT

#### INSTRUCTIONS

The purpose of this experiment is to explore the effects of viewing conditions on the perception of size. In this experiment, there are no "right" answers.

Press any number to continue...  
(Key press pages to next screen)

This experiment will present different pyramids and other structures. You will be asked to estimate the ratio of the height of each pyramid to the width of its base. You are not judging the absolute size, just the ratio of height to width

Press any number to continue...  
(Key press pages to next screen)

The rating scale is a set of pyramid profiles in a numbered sequence. Estimate which pyramid most nearly approximates the height-to-width ratio of the displayed pyramid.

Enter your estimate by pressing the appropriate number key (either set).

Each block of trials is randomized differently.

Press any number key when you are ready to begin.

Press any number to continue...  
(Key press clears screen, starts practice trials.)