Spring 2010

Shape Representation in Primate Visual Area 4 and Inferotemporal Cortex

Thomas M. Murphy
University of Pennsylvania, murphyt@seas.upenn.edu

Shape Representation in Primate Visual Area 4 and Inferotemporal Cortex

Abstract

The representation of contour shape is an essential component of object recognition, but the cortical mechanisms underlying it are incompletely understood, leaving it a fundamental open question in neuroscience. Such an understanding would be useful theoretically as well as in developing computer vision and Brain-Computer Interface applications. We ask two fundamental questions: "How is contour shape represented in cortex and how can neural models and computer vision algorithms more closely approximate this?" We begin by analyzing the statistics of contour curvature variation and develop a measure of salience based upon the arc length over which it remains within a constrained range. We create a population of V4-like cells – responsive to a particular local contour conformation located at a specific position on an object's boundary – and demonstrate high recognition accuracies classifying handwritten digits in the MNIST database and objects in the MPEG-7 Shape Silhouette database. We compare the performance of the cells to the "shape-context" representation (Belongie et al., 2002) and achieve roughly comparable recognition accuracies using a small test set. We analyze the relative contributions of various feature sensitivities to recognition accuracy and robustness to noise. Local curvature appears to be the most informative for shape recognition. We create a population of IT-like cells, which integrate specific information about the 2-D boundary shapes of multiple contour fragments, and evaluate its performance on a set of real images as a function of the V4 cell inputs. We determine the sub-population of cells that are most effective at identifying a particular category. We classify based upon cell population response and obtain very good results. We use the Morris-Lecar neuronal model to more realistically illustrate the previously explored shape representation pathway in V4 – IT. We demonstrate recognition using spatiotemporal patterns within a winnerless competition network with FitzHugh-Nagumo model neurons. Finally, we use the Izhikevich neuronal model to produce an enhanced response in IT, correlated with recognition, via gamma synchronization in V4. Our results support the hypothesis that the response properties of V4 and IT cells, as well as our computer models of them, function as robust shape descriptors in the object recognition process.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Bioengineering

First Advisor

Brian Litt, M.D.

Second Advisor

Leif H. Finkel, M.D., Ph.D.

Third Advisor

Gershon Buchsbaum, Ph.D.

Keywords

shape representation, object recognition, IT, inferotemporal cortex, V4, macaque

SHAPE REPRESENTATION IN PRIMATE VISUAL AREA 4 AND INFEROTEMPORAL CORTEX

Thomas Michael Murphy

A DISSERTATION

in

Bioengineering

Presented to the Faculties of the University of Pennsylvania

in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

2010

_____

Brian Litt, M.D.
Supervisor of Dissertation

_____

Susan Margulies, Ph.D.
Graduate Group Chairperson

Leif H. Finkel, M.D., Ph.D.	Advisor
Gershon Buchsbaum, Ph.D.	Dissertation Committee Chairman
Larry A. Palmer, Ph.D.	Dissertation Committee Member
Jianbo Shi, Ph.D.	Dissertation Committee Member
William C. Stacey, M.D., Ph.D.	Dissertation Committee Member

Shape Representation in Primate Visual Area 4 and Inferotemporal Cortex

Dedication

For Louise & Leif

Thanks for so much…

Acknowledgment

Thanks to my friends and colleagues throughout the university, particularly in the Finkel and Litt labs, for innumerable useful and enjoyable conversations.

Thanks to Brian for guidance and support during the most trying of times.

Thanks to Gershon, Larry, Jianbo and Bill for helpful advice throughout this process.

Thanks to Maciej for encouragement and feedback at critical junctures.

Thanks to Leif for inspiration.

Thanks to Robin, Alex and Madeline, the loves of my life, for making it all worthwhile.

ABSTRACT

SHAPE REPRESENTATION IN PRIMATE VISUAL AREA 4
AND INFEROTEMPORAL CORTEX

Thomas Michael Murphy

Brian Litt, M.D.

The representation of contour shape is an essential component of object recognition, but the cortical mechanisms underlying it are incompletely understood, leaving it a fundamental open question in neuroscience. Such an understanding would be useful theoretically as well as in developing computer vision and Brain-Computer Interface applications. We ask two fundamental questions: "How is contour shape represented in cortex and how can neural models and computer vision algorithms more closely approximate this?" We begin by analyzing the statistics of contour curvature variation and develop a measure of salience based upon the arc length over which it remains within a constrained range. We create a population of V4-like cells – responsive to a particular local contour conformation located at a specific position on an object's boundary – and demonstrate high recognition accuracies classifying handwritten digits in the MNIST database and objects in the MPEG-7 Shape Silhouette database. We compare the performance of the cells to the "shape-context" representation (Belongie *et al*., 2002) and achieve roughly comparable recognition accuracies using a small test set. We analyze the relative contributions of various feature sensitivities to recognition accuracy and robustness to noise. Local curvature appears to be the most informative for shape

recognition. We create a population of IT-like cells, which integrate specific information about the 2-D boundary shapes of multiple contour fragments, and evaluate its performance on a set of real images as a function of the V4 cell inputs. We determine the sub-population of cells that are most effective at identifying a particular category. We classify based upon cell population response and obtain very good results. We use the Morris-Lecar neuronal model to more realistically illustrate the previously explored shape representation pathway in V4 – IT. We demonstrate recognition using spatiotemporal patterns within a winnerless competition network with FitzHugh-Nagumo model neurons. Finally, we use the Izhikevich neuronal model to produce an enhanced response in IT, correlated with recognition, via gamma synchronization in V4. Our results support the hypothesis that the response properties of V4 and IT cells, as well as our computer models of them, function as robust shape descriptors in the object recognition process.

Table of Contents

List of Tables

List of Illustrations

Chapter 1

General Introduction

1.1 The Ventral Stream and Models of Object Recognition

Detection, categorization and identification are generally agreed upon as being the three major components of object recognition. Salience would seem to be a prerequisite of detection and recognition. The perceptual salience of an object, the degree to which it pops-out from the background and captures attention, determines the difficulty of locating it in search tasks and the speed of recognizing it in rapid presentations. It has been well covered in the literature. The Gestalt psychologists identified several properties, such as continuity, colinearity or cocircularity, and closure, which confer

salience upon objects, although their relative contributions to overall salience and manners of integration remain unclear.

Ullman has proposed the idea that salience is a global property that integrates the Gestalt factors across an entire object (Shashua and Ullman, 1988; Ullman, 1996). Closure is itself a global property (Yen and Finkel, 1998), and Kovács and Julesz have shown, using roughly circular contours, that closure leads to a marked increase in salience (Kovács and Julesz, 1993). In Ullman's original algorithm, salience is determined by the length, continuity and curvature of the contour. Long, smooth contours with little change of curvature and no gaps are calculated to be the most salient. Curvature covariation has been investigated (Chapter 3), and in a detailed study of Ullman's methodology, Alter and Basri have found that, for many images, their algorithm robustly predicted salience values in accord with human perception (Alter and Basri, 1996). In this dissertation, we consider closed 2-dimensional bounding contours primarily, and concern ourselves with downstream functionalities, namely representation and recognition.

An accepted animal model for object recognition studies is the rhesus macaque monkey (*Macaca mulatta*). Its cortical area that is predominantly visual in function accounts for 52% of the cerebral cortex (compared to 27% in humans) (Van Essen, 2003). Visual area 4 (V4) of the macaque, an important intermediate hierarchical stage of visual form processing of the ventral cortical stream (the "what?" pathway, thought to be responsible for shape recognition) between the primary visual cortex (V1) and the inferior temporal complex (inferotemporal cortex) (IT) (Van Essen and Gallant, 1994), has received a

significant amount of research attention. Virtually all visual sensory information in the ventral pathway passes through extrastriate area V4 on its way to the inferotemporal areas (Ungerleider and Mishkin, 1982), with V4 receiving inputs from visual area 2 (V2) and providing the major source of input to IT (Felleman and Van Essen, 1987; Felleman and Van Essen, 1991).

Selectivity for simple dimensions, such as orientation, spatial frequency, length and width, has been demonstrated in V4 (Desimone and Schein, 1987), but these properties have also been found in earlier stages of form processing, such as area V1. The Van Essen group, however, has been successful in finding selectivities in V4, thought to be form-related, that have not been found in earlier stages. They have found that nearly all V4 neurons, while biased towards polar and hyperbolic stimuli over Cartesian stimuli, convey information about all three stimulus classes, with most having tuning curves in multiple classes (Gallant *et al.*, 1996). This, along with their positional invariance, perhaps suggests that V4 cells are neither simple feature detectors nor simple filters, but rather nonlinear filters broadly tuned along several form-related dimensions. V4 cells responsive to polar stimuli could facilitate the perception of curvature, an important feature in natural image understanding, as well as mediate view invariance.

Responses within V4 are dependent upon the spatial relationship between the stimulus position in the classical receptive field and the direction of attention (Connor *et al.*, 1997). This differential modulation implies that an object-specific attention-based representation of the position of visual features, or some intermediate form leading to

this, may be present. Object-centered coordinates such as these within a local reference frame could yield translation and scale invariance, desirable in any recognition system, biological or artificial. This is analogous and consistent with the idea of human navigation being largely dependent upon a continuously updated egocentric representation of object locations, with a lesser dependence upon an enduring allocentric map of environmental shape (Wang and Spelke, 2000). Note that a neural code of human spatial navigation based on cells that respond at specific locations and cells that respond to views of landmarks has been identified in the hippocampus and parahippocampal region (Ekstrom *et al*., 2003).

Research by Connor and colleagues has identified cells in area V4 that are selective for both the local shape and the global position of segments of object borders, with most cells strongly responsive to a particular type of boundary conformation at a specific position within a larger shape (Pasupathy and Connor, 2001; Pasupathy and Connor, 2002). Specifically, these neurons are selective for both the magnitude and direction of curvature of a stimulus, and individual V4 cells appear to encode moderately complex boundary information at specific locations within larger shapes.

Contour curvature detection, with neurobiological correlates in area V4, is sensitive to noise and provides a feature that is important for recognition, with lower level curvature calculations proving useful in higher level object recognition. Pasupathy and Connor have pointed out that complex shape representation in area V4 is parts-based (since contour segments are defined by conformation and position) as well as distributed (since

individual cells encode smaller parts of larger objects) (Pasupathy and Connor, 2001). If shapes are represented as combinations of primitive features then shape recognition could be seen as a hierarchical process. With areas V2 and V4 selective both for the magnitude and direction of curvature, area V1 could provide the input to a population of local curvature detectors while area V4 could perform global matching between the curvature detectors. It is conceivable that feedback from area V4 to area V1 provides some top-down control of salience. A number of studies have proposed neural and computational mechanisms for computing curvature (Dobbins *et al.*, 1989). These lower level curvature calculations are consistent with mechanisms, such as end-stopping, available in primary visual cortex.

Simoncelli and Olshausen have reported that sensory neurons are evolutionarily and developmentally adapted to the statistical properties of the stimuli to which they are exposed (Simoncelli and Olshausen, 2001). It is therefore not surprising that the difference between a salient closed contour and a non-salient background is determined by a statistic of the contour itself. Elder and Goldberg consider contour grouping to be equivalent to the recovery of sequences of tangents. They have found that the statistical dependencies between neighboring tangents on a contour are much greater than those between distant tangents (Elder and Goldberg, 2002). Since local cues ultimately lead to global criteria, and since cocircularity is part of the Gestalt principle of good continuation, local curvature consistency on a contour could ultimately enhance global contour salience.

The surprising finding by von der Heydt and associates – that a majority of cells in V2 and V4, and a smaller number of cells in V1, carry information about how local features belong to objects – is significant. Specifically, these neurons were seen to code the side to which a border in a figure belongs (Zhou *et al.*, 2000). This response was seen to be generated within the visual cortex, not projected down from higher levels, and clearly represents global image context integration. Perception tends to assign contrast borders to objects, according to the Gestalt psychologists, and the von der Heydt results show that this is accomplished at an earlier cortical level than previously thought.

The research by Hegdé and Van Essen suggests that neurons in visual area V2 encode information about many complex shape and contour characteristics and are involved in form analysis to an extent not previously realized (Hegdé and Van Essen, 2000). The responsiveness of cells in V2 to complex stimuli, such as angles, arcs, circles, intersecting lines, and non-Cartesian (hyperbolic and polar) gratings, while not as pronounced as in V4, is apparent. They have found that most V2 cells showed differential responsiveness to these stimuli, suggesting that V2 cells explicitly represent complex shape information. They imply that V2 cells may sample the grating and contour stimuli space widely, yielding a simplified, low-dimensional visual representation with perceptually relevant information preferentially left intact (Hegdé and Van Essen, 2003). Again, the primitive features that compose the complex shapes could be addressed in area V1.

Both the von der Heydt as well as the Hegdé / Van Essen results give credibility to a hierarchical feedforward flow of visual information model, from V1 / V2 to V4 to IT, to support visual form processing. These results can be considered in the context of information theory, with quantification of the amount of stimulus information carried by neural responses (Borst and Theunissen, 1999).

The model of Poggio and colleagues has been both consistent with physiological data and successful in object recognition tasks. It is based on simple hierarchical feedforward architectures and assumes that both invariance to position and scale as well as feature specificity are built up through separate mechanisms (Riesenhuber and Poggio, 1999b). Following the paradigm that the average feature complexity as well as receptive field size increase from V1 to IT, the model consists of linear, template matching units and non-linear, pooling units. Poggio's main conclusion is that the assignment of different features to the correct object (the "binding problem") does not require complex oscillation or synchronization mechanisms. The model is bottom-up, without the absolute requirement of an explicit top-down (attentional or otherwise) signal, and does not require an explicit segmentation stage (Riesenhuber and Poggio, 1999a). Poggio proceeds to caste this entire model as a view-based module, which is then incorporated into his "Standard Model". Here, view- and component-tuned units represent the output of the view-based module, and are subsequently used to create view invariant (object-tuned) units. These are input to task-related units, performing such visual tasks as identification / discrimination or object categorization (Riesenhuber and Poggio, 2003).

Feedback pathways for top-down modulation of responses and support of learning may also be incorporated.

The work by Hinkle and associates provides some insights into the global significance of area V4. The finding that stereoscopic disparity tuning (conventionally associated with the dorsal pathway) is prevalent in area V4 positions it as a major source of disparity information for IT and emphasizes the importance of stereoscopic depth cues in the ventral pathway (Hinkle and Connor, 2001). It also suggests that 3-dimensional shape information is processed in V4 based on these cues. A bias towards certain disparities has also been observed, possibly reflecting the ventral pathway's particular emphasis on foreground objects or parts of objects projecting towards the viewer. The finding that neurons in area V4 are tuned for 3-dimensional orientation (Hinkle and Connor, 2002) supports these ideas and further suggests that the initial stages of shape analysis utilize depth cues. Three-dimensional orientation tuning facilitates the computationally difficult 3-dimensional position invariance and is compatible with both viewpoint-invariant and viewpoint-dependent models.

While we consider only 2-dimensional boundary elements in our present effort, a complete model of V4 might have to incorporate 3-dimensional information, reflecting the internal cortical representation of the real world.

V4 cells respond to a variety of stimulus features – including color (McKeefry and Zeki, 1997), orientation (Desimone and Schein, 1987; Hinkle and Connor, 2002), disparity

(Hinkle and Connor, 2001), and complex spatial patterns (Gallant *et al.*, 1996). Extrastriate cells show selectivities to several aspects of form (Gallant *et al.*, 1996) and border ownership (Zhou, Friedman and von der Heydt, 2000). In addition, V4 cell receptive field properties are strongly modulated by attention (Reynolds and Desimone, 2003; Bichot, Rossi and Desimone, 2005; McAdams and Maunsell, 2000; Motter, 1994; Connor *et al.*, 1997), and the presence of a small feature within the large receptive field can drive cellular response.

Gray and McCormick have observed the synchronous rhythmic firing in the gamma frequency band of chattering cells in response to visual stimulation (Gray and McCormick, 1996). The functional significance of these oscillations might be to integrate low level visual features. By recruiting large populations of curvature sensitive cells in V4 into synchronous firing, cells representing consistent curvatures on contours could be bound together via the similarity of their firing rates. Hopfield and Brody have proposed a mechanism in which groups of cells with similar firing rates synchronize (Hopfield and Brody, 2001). Synchronization occurs naturally in the types of cortical architectures studied by Beierlein and colleagues (Beierlein *et al.*, 2000). Hopfield and Brody make the point that in a large ensemble of cells, a large fraction of cells firing at the same rate is statistically unlikely. Thus, a set of connected V4 cells, each sensitive to magnitude and direction of curvature, that are coupled by horizontal connections, could rapidly synchronize.

An intriguing idea has emerged that suggests that area V4, as a key stage in a network of cortical and subcortical areas working in concert, contains a retinotopic salience map that guides saccadic eye movements during free viewing (Mazer and Gallant, 2003). It is seen that bottom-up, visually driven activity in V4 predicts the direction of subsequent saccades and is modulated by top-down, feature attention-related signals. Information about the spatial distribution of activity in V4 could be used in downstream areas to guide subsequent exploratory eye movements toward interesting positions in the visual field. Future efforts might include modeling the significant contribution of top-down information from higher cortical areas. This not only can localize the regions of interest, but also can make adjustments to the iso-curvature segments as well.

Much research has focused on the general neural selectivity of neurons in the inferotemporal cortex (Freedman *et al*., 2003; Baker *et al*., 2002; Tsunoda *et al*., 2001; Op de Beeck *et al*., 2001; Booth and Rolls, 1998; Rolls *et al*., 1997; Gallant *et al*., 1996; Logothetis *et al*., 1995; Kobatake and Tanaka, 1994; Fujita *et al*., 1992; Young, 1992; Felleman and Van Essen, 1991; Gross *et al*., 1972). Other work by Connor and colleagues has focused on selectivity for 2-dimensional boundary shape (perhaps the kind that actually dominates responses to realistic objects) in the inferotemporal cortex. It has been found that IT neurons integrate specific information, such as curvatures, orientations, and relative positions, about the shapes of multiple contour fragments (typically 2–4) (Brincat and Connor, 2004). Explicit signals that code structural relationships between parts are generated, useful for high-level object representation, and supporting the idea of parts-based shape representation. This once again may support a

hierarchical feedforward implementation, with IT input originating in V4. It is related to the fragment-based approach of Ullman and colleagues. Here, fragments (contour segments producing responses in V4) are the component building blocks used to represent a large variety of objects belonging to a common class (Ullman *et al.*, 2001).

Other research has focused on specific aspects of IT coding. In a departure from controlled viewing tasks, DiCarlo and Maunsell have seen that most IT neuronal responses are unaffected by free viewing, as when a primate behaves naturally and visually explores cluttered environments by changing its direction of gaze (DiCarlo and Maunsell, 2000). Tsunoda and colleagues have reported that an object is represented in IT by a combination of cortical feature columns, each representing a visual feature, with combinations of active and inactive columns used for individual features (Tsunoda *et al*., 2001). Baker and colleagues have trained monkeys to discriminate among stimuli consisting of discrete parts (Baker *et al*., 2002). After training, responses to learned images, though not stronger, are enhanced in selectivity for parts and wholes, indicating a possible neural mechanism for holistic effects. Kiani and colleagues have found that the categorical structure of objects (animate and inanimate with further hierarchical subdivisions) is represented by the pattern of IT population activity, with objects of the same category clustered and evoking similar response patterns (Kiana *et al*., 2007). Zoccolan and colleagues have seen that IT neurons with sharp selectivities for unique combinations of diagnostic object features typically have low tolerance to variations in position, size, illumination and clutter, and vice versa (Zoccolan *et al*., 2007). Op de Beeck and colleagues have found evidence for a large-scale, highly reproducible and

stable, map of shape selectivity in IT that is largely independent of object class familiarity and behavioral task (Op de Beeck *et al*., 2008). McMahon and Olson have reported that the influences of shape and color sum linearly in most IT neurons, with neither conjunction selectivity nor a specialized feature binding process necessary for representation (McMahon and Olson, 2009).

As with V4, a complete model of IT might have to incorporate 3-dimensional information. Yamane and colleagues have found evidence in IT for an explicit neural code for complex 3-dimensional shape, with widespread tuning for the spatial configurations of surface fragments (Yamane *et al*., 2006; Yamane *et al*., 2008). Sereno and colleagues have suggested that 3-dimensional shape representations are highly localized, yet widely distributed in occipital, temporal, parietal and frontal cortices, with distributed networks intersecting both "what" and "where" processing streams (Sereno *et al*., 2002).

Although our approach considers only implementations with neurobiological correlates, some insights might be gained by considering state-of-the-art segmentation techniques from computer vision. For example, future successful models might choose to employ the image-based algorithm of Shi and Malik (2000). In their methodology, the perceptual grouping problem is solved by extracting the global impression of an image. Here, image segmentation is treated as a graph partitioning problem. A class-based segmentation method (Borenstein and Ullman, 2002), guided by a stored representation of the shape of objects within a general class, has similarities to human vision. It emphasizes the role of

high-level information by using class-specific criteria. Another technique, useful in segmenting an image into foreground and background (Yu and Shi, 2003), employs parallel processes: one for low-level pixel grouping for feature saliency and another for high-level patch grouping for object familiarity.

The state-of-the-art category-level recognition system of Malik and colleagues (Belongie *et al*., 2002) provides a shape description based on the distances between all pairs of points on the object's bounding contour. Shape context log-polar histograms are computed for each point on the contour. The collection of histograms fully characterizes each shape. Malik solves the correspondence problem between two shapes using optimal assignment. He estimates the aligning transform using these correspondences and regularized thin-plate splines. Finally, he measures similarity between the shapes as a function of matching errors between corresponding points and aligning transform magnitude.

Another model, that of LeCun and colleagues (LeCun *et al.*, 1998), has been successful in several applications, including handwritten character recognition. It applies machine learning techniques to multilayer neural networks, trained with the backpropagation algorithm. It relies more on automatic, gradient-based learning and less on hand-designed heuristics. Specifically, it employs convolutional neural networks, specifically designed to handle 2-dimensional shape variability. It uses a learning paradigm to globally train all the modules – feature extractors and classifiers – and optimize a global performance criterion.

A future successful model of V4 would have to incorporate other findings as well. For instance, the position and variability of the color center and retinotopic organization within V4 has been investigated (McKeefry and Zeki, 1997). Also, direction-of-motion selectivity after adaptation has been confirmed (Tolias *et al.*, 2005) and has great significance.

It seems that recognition might require curvature measurement comparisons over a significant image region, the segmentation of contours into target vs. background, and possibly the discrimination of the direction of figure. This suggests the need for both horizontal as well as top-down information. This is very likely accomplished at the level of V4. The possibility of cortical hypercolumns receiving "matched" input from multiple other local and distant hypercolumns could be investigated. It may be the case that the synchronization of chattering bursts signifies cliques of connected hypercolumns. These might possibly implement some form of Bayesian inference related to Mumford's framework, with recurrent feedforward and feedback loops integrating top-down context and bottom-up stimulation (Lee and Mumford, 2003).

1.2 Computational Neuroscience – Necessity and Benefits

The somewhat controversial nature of computational neuronal modeling is exemplified by the recent "cat brain" debate (Adee, 2009) between Dharmendra Modha, a computer scientist on DARPA's SyNAPSE project who presented a simulation that his team claimed approached the scale of a cat's brain, and Henry Markram, a neuroscientist who claimed that the simulation was a hoax, calling into question the legitimacy of brain simulation research. Modha's position was that the simulation was not a cat brain, but rather on the scale of a cat's brain, in terms of the number of neurons and synapses. Markram objected to the impoverished detail in the simulation's point neurons and regarded it as trivial.

We feel that the acquisition of physiological and anatomical data alone is insufficient for a complete understanding of neural processing. To learn how the brain works, experimental studies of animal and human nervous systems must be coupled with computational brain models. At this stage, neuroscience requires a quantitative framework to integrate and manage its enormous amount of experimental data. Computational neuroscience can provide this.

Computational models can aid in conceptualizing experiments and can help to interpret experimental results. The parameter manipulations that are possible within a model can

far exceed what is biologically practical. A model provides access to mechanisms with levels of sensitivity and specificity that are unavailable experimentally, as well as the ability to record all parameters simultaneously. Computational protocols may be run repeatedly, with little or no preparatory work, without concern for loss or sacrifice of specimens.

It is reasonable to think that a sufficiently detailed neuronal model, imbedded within a realistic network, will produce realistic behavior. A good model should be able to reproduce experimental results and it should allow us to generate predictions and test hypotheses at the appropriate level of detail. It even has the capacity to contradict experimental expectations.

Simplified models may sacrifice biological accuracy for computational efficiency, but models do not have to be perfect to be useful. Therefore, we create models with as much physiological and anatomical fidelity as is available to us in the published literature, while remaining tractable. This represents a compromise between the requirements of simple computation and biological realism. We must, however, understand the simplifications if we are to understand the connections between our computational models and nature.

A typical criticism of computational neuroscience is that it does not generate enough predictions. In fact, some journals deem models that do not make testable predictions to be unpublishable. Also, some parameter choices may be biologically unrealistic and

some assumptions made to fill knowledge gaps may be incorrect. None of these shortcomings are insurmountable.

We feel that computational neuroscience should ideally augment, guide and complement experimental neuroscience, rather than replace it. This dual approach should provide intuitions and a deeper understanding for both sub-disciplines. Most importantly, computational simulations might lead to interesting and novel insights and unexpected results that can later be validated experimentally.

1.3 MATLAB Models

We have used GENESIS (see http://genesis-sim.org/), NEURON (see http://www.neuron.yale.edu/neuron/), Java (see http://java.sun.com/), python (see http://www.python.org/), etc., to create computational neural models.

Our primary tool, however, is MATLAB. Its modular design, high-level code, flexible interface options, device independence, numerical algorithms and data visualization capabilities make it ideal for our purposes. Unsatisfactory execution time has simply never been an issue for us.

With many third-party toolboxes available (such as Bayes Net Toolbox for Matlab: http://www.cs.ubc.ca/~murphyk/Software/BNT/bnt.html), it remains a very popular

choice for computational neuroscientists, as well as scientists and engineers in general, worldwide.

We have consistently used the latest versions of MATLAB, including MATLAB toolboxes, which were available. Naturally, these have evolved over the course of our investigations. The particular versions used for each piece of the research are noted in the Chapters.

Chapter 2

Research Overview

2.1 Chapter Contents

We begin with a curvature-related investigation in Chapter 3. The salience of a contour depends upon several factors, including continuity, closure and curvature consistency. We analyze the statistics of curvature variation using a single image from Shimon Ullman's (Shashua and Ullman, 1988) original work on contour salience. We develop a measure based on the arc length of a contour segment over which curvature variation remains within a constrained range. Locally, all contours in the image are similar with respect to curvature consistency. However, when the entire contour is considered, the most salient contours are found to have the most consistent curvatures. This finding reinforces Ullman's point that salience is a global property of the object. We interpret

these results in view of Rosenholtz's (Rosenholtz, 1999) model of salience as a statistical measure of outliers from a population. In addition, we speculate on the visual cortical mechanisms in striate and extrastriate cortex required to carry out salience measurements on this class of images. A portion of this material has been previously published (Murphy *et al.*, 2003).

In Chapter 4, we continue with an investigation of curvature-sensitivity in macaque V4. Cells in extrastriate visual cortex have been reported to be selective for various configurations of local contour shape (Pasupathy and Connor, 2001; Hegdé and Van Essen, 2003). Specifically, Pasupathy and Connor found that in area V4 most cells are strongly responsive to a particular local contour conformation located at a specific position on the object's boundary. We use a population of "V4-like cells" – units sensitive to multiple shape features modeled after V4 cell behavior – to generate representations of different shapes. Standard classification algorithms (earth mover's distance, support vector machines) applied to this population representation demonstrate high recognition accuracies classifying handwritten digits in the MNIST database and objects in the MPEG-7 Shape Silhouette database. We compare the performance of the V4-like unit representation to the "shape-context" representation of Belongie and colleagues (Belongie *et al*., 2002). Results show roughly comparable recognition accuracies using the two representations when tested on portions of the MNIST database. We analyze the relative contributions of various V4-like feature sensitivities to recognition accuracy and robustness to noise – feature sensitivities include curvature magnitude, direction of curvature, global orientation of the contour segment, distance of

the contour segment from object center, and modulatory effect of adjacent contour regions. Among these, local curvature appears to be the most informative variable for shape recognition. Our results support the hypothesis that V4 cells function as robust shape descriptors in the early stages of object recognition. A portion of this material has been previously published (Murphy and Finkel, 2007).

In Chapter 5, we continue our investigation of how contour shape is represented in cortex. We extend our earlier results to consider the recognition properties of a population of cells modeled after those found in IT, which nonlinearly integrates specific information about the 2-dimensional boundary shapes of multiple contour fragments (V4 cell inputs) with tuning functions on the shape × position domain (Brincat and Connor, 2004; Brincat and Connor, 2006). Using nonlinear least squares optimization and genetic algorithms to fit parameters, we create selective IT-like cell populations with similar response patterns. We are principally interested in the number of constituent Gaussian terms in the IT-like cells' total response equations, the linear and nonlinear parts of these equations, the amount of nonlinearity and how these aspects relate to the shapes of objects (as opposed to their orientations and scales). We evaluate the performance of our IT populations on a set of real images as a function of the V4-like cell inputs. The stimuli (2-dimensional closed contours representing object boundaries) evoke a pattern of activity across the population of IT cells. Shape recognition is evaluated by demonstrating that the patterns of activity across the units to members of a particular object class resemble each other to a higher degree than they resemble members of any other class. We examine cell response space in more detail using principal components

analysis and a 2-dimensional and 3-dimensional non-classical non-metric multidimensional scaling analysis. We find the correlation coefficients of the observations (cell responses) and variables (images) and determine the sub-population of cells that are most effective at identifying a particular category. We use a support vector machine, as well as a tree-based model, for classification based upon cell population response. In general, we obtain very good results across a wide range of parameter values and implementation strategies, comparable to those obtained previously with the digit database. Our results suggest that curvature- and position-sensitive units, as described by Brincat and Connor in IT, can function as robust shape descriptors.

We concentrate on realistic biological models of cells – those that have a high biological plausibility without burdensome implementation costs – within our networks in Chapter 6. We use the Morris-Lecar neuronal model, with genetic algorithms utilized for parameter fitting, to more realistically illustrate the previously explored shape representation pathway in V4 – IT while remaining faithful to IT cell response patterns. As an aside, we demonstrate biologically-based object recognition using spatiotemporal patterns within a self-organized winnerless competition neural network with FitzHugh-Nagumo model neurons. We conclude with an examination of gamma synchronization in the V4 – IT circuit. We use the Izhikevich neuronal model and demonstrate that an initially out-of-phase network's inherent characteristics and dynamics can induce synchronized responses in V4 via PING mechanisms by applying current input to the network. Additionally, we show that a response amplification in IT, correlated with recognition, results from the synchronized spiking in V4 and roughly coincides with the

onset of synchronization. Our results suggest that realistic biological models of cells with curvature- and position-sensitive response properties, as described by Pasupathy and Connor in V4 and Brincat and Connor in IT, can function as robust shape descriptors.

We summarize our results in Chapter 7 and discuss future steps in Chapter 8.

2.2 Acknowledgements

This work was supported by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Office of Naval Research under Grant N00014-01-1-0625 as well as grant 9873463 from the NSF KDI program.

The work was also supported by generous Research Fellowships from the University of Pennsylvania.

2.3 Motivation, Personal Objectives and Research Goals

This dissertation was *proposed* with very specific research goals. We sought to develop a population model of V4-like cells and investigate their ability to represent contour shape. We sought to evaluate the performance of this V4 network on the MNIST database of

handwritten digits. We sought to develop a model of IT response based on V4 inputs and evaluate the performance of an IT population on a set of real images. We sought to investigate the recognition performance within the IT network as a function of the number of V4 subunits and their nonlinear combinations. Finally we sought to explore the connections between our model and current psychological models of object categorization. (Incidentally, all of these objectives have been accomplished.)

Perhaps, though, our motivation and goals should be expanded slightly.

The representation of contour shape is an essential component of object recognition, but the cortical mechanisms underlying shape analysis and object recognition are incompletely understood, leaving it a fundamental open question in neuroscience. We hope to approach such a neurobiological understanding, which would be useful theoretically as well as in developing or improving computer vision and Brain-Computer Interface (BCI) methodologies and applications.

Some computer vision approaches to object recognition have begun to achieve impressive levels of accuracy and robustness, yet lack a clear connection to known cortical constructs. We hope to narrow the divide between the theoretical computational neuroscience and the biological neurophysiology by creating theoretical models of abstract computations as well as neural implementations. Both types of models might be able to generate experimentally testable predictions.

We ask two fundamental questions: “How is contour shape represented in cortex and how can neural models and computer vision algorithms more closely approximate this?” Our primary objective is to determine why the response properties of V4 and IT cells (i.e., their receptive fields), and in particular their sensitivities to curvatures and contour positions, are useful. We hope to establish a clear connection between a computer model of a recognition system and known cortical constructs within a biologically realistic network architecture.

This type of research – in search of an improved theoretical understanding – is essential to our field, important in that it is a model problem, interesting in that it resides at the center of the mind-brain issue. Leif Finkel and I began our research together by investigating and developing models of striate cortex. Although none of that earlier work is reflected in this dissertation, my personal objective is to continue the recognition work that Leif and I started in V4 and IT.

Chapter 3

Curvature Covariation as a Factor in Perceptual Salience

3.1 Introduction

The perceptual salience of an object is the degree to which it pops-out from the background and captures attention. The salience of a target determines the difficulty of locating it in search tasks, and the speed of recognizing it in rapid presentations. The Gestalt psychologists identified several properties that confer salience upon objects, such as continuity, colinearity or cocircularity, and closure. However, the relative degree to which each of these properties contributes to overall salience remains unclear, as does the manner in which these factors are integrated.

As Ullman pointed out, salience is a global property that integrates the Gestalt factors across an entire object (Shashua and Ullman, 1988; Ullman, 1996). In Figure 3.1, taken from Ullman's original paper, the three circular contours pop out and are more salient than the background squiggles. We are interested, in this paper, in understanding quantitatively what factors render the circles salient.

One Gestalt factor that distinguishes the circular contours from the background is closure. Closure is itself a global property (Yen and Finkel, 1998), and Kovács and Julesz have shown, using roughly circular contours, that closure leads to a marked increase in salience (Kovács and Julesz, 1993).

In Ullman's original algorithm, salience was determined by the length, continuity, and curvature of the contour. Long, smooth contours with little change of curvature and no gaps are calculated to be most salient. In a detailed study of this algorithm, Alter and Basri found that, for many images, it robustly predicted salience values in accord with human perception (Alter and Basri, 1996). To our knowledge, a detailed analysis of the original Ullman image (Figure 3.1) has not been performed.

We sought to determine, in this image, the degree to which the circular contours are more or less cocircular than the background squiggles, over a range of spatial scales. In other words, to what degree does curvature covary across the circular contours as compared to across the squiggles. This study therefore represents an attempt to gather image statistics on a single image.

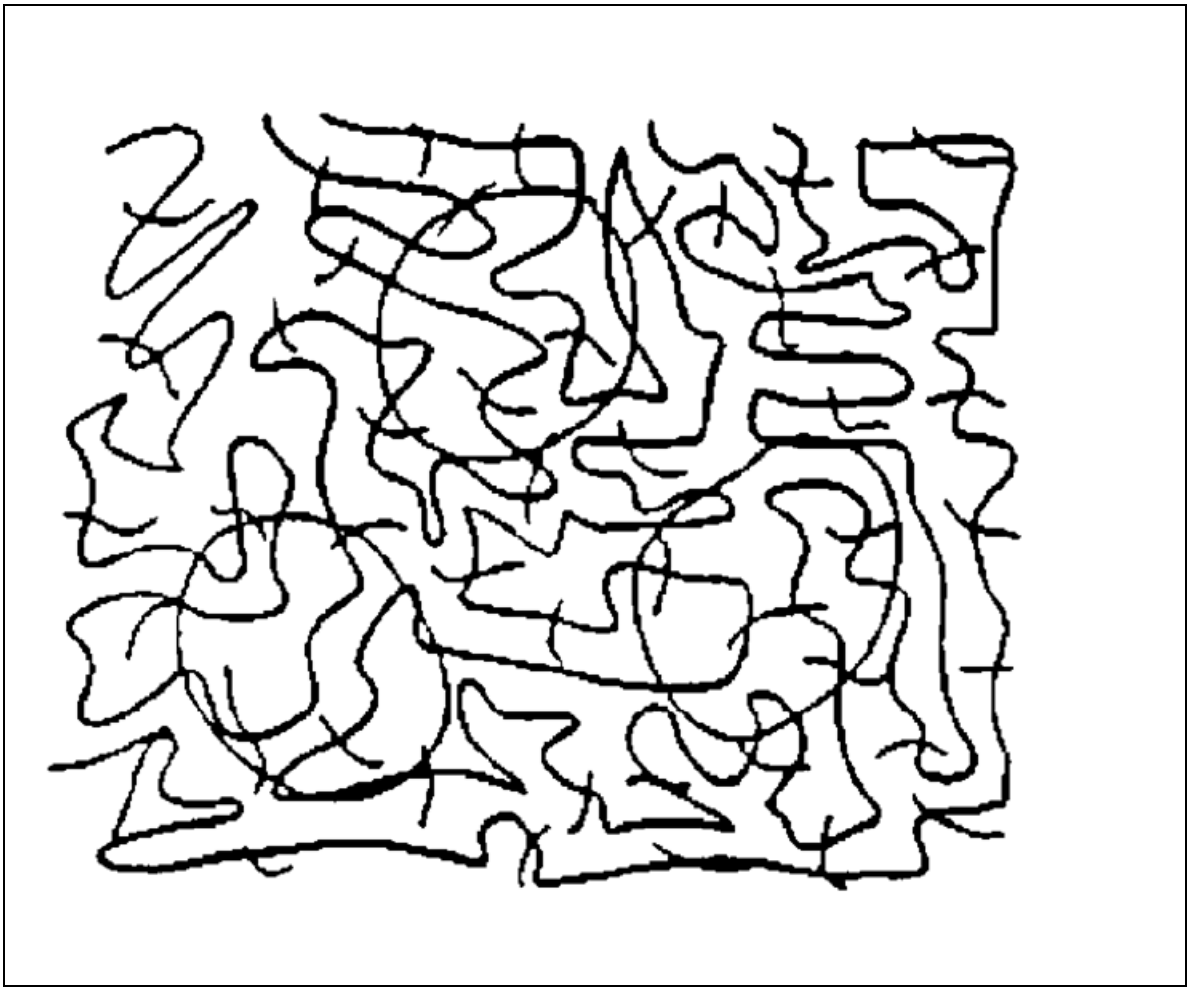

Figure 3.1 – Shashua & Ullman image.

3.2 Methodology

By deliberate construction, the three closed contours in Figure 3.1 are physically similar to the background contours in terms of contrast, thickness, and range of orientations. The background squiggles are not all continuous, but terminate at the borders of the figure, forming three open contours. The closed contours are shorter (239, 244, 243 pixels) than the background contours (1238, 2011, 880 pixels), but it is not clear that the terminations of the background contours are perceptually significant.

An erosion algorithm was first employed to reduce image contours down to single-pixel widths (using the MATLAB bwmorph function). We were then able to decompose the image into closed and background contours and investigate each separately.

Orientations were computed using Freeman & Adelson's G2 / H2 steerable filters (Freeman and Adelson, 1991). At locations where contours cross, we ascertained that the correct orientation was assigned to each contour.

Curvature is a scale-dependent quantity. Therefore, as illustrated in Figure 3.2, we used a weighted average of the curvatures computed over several arc lengths, and assigned the averaged value as the curvature at each pixel. This weighted average was chosen to provide some minimal smoothing while retaining the true nature of the contours. The

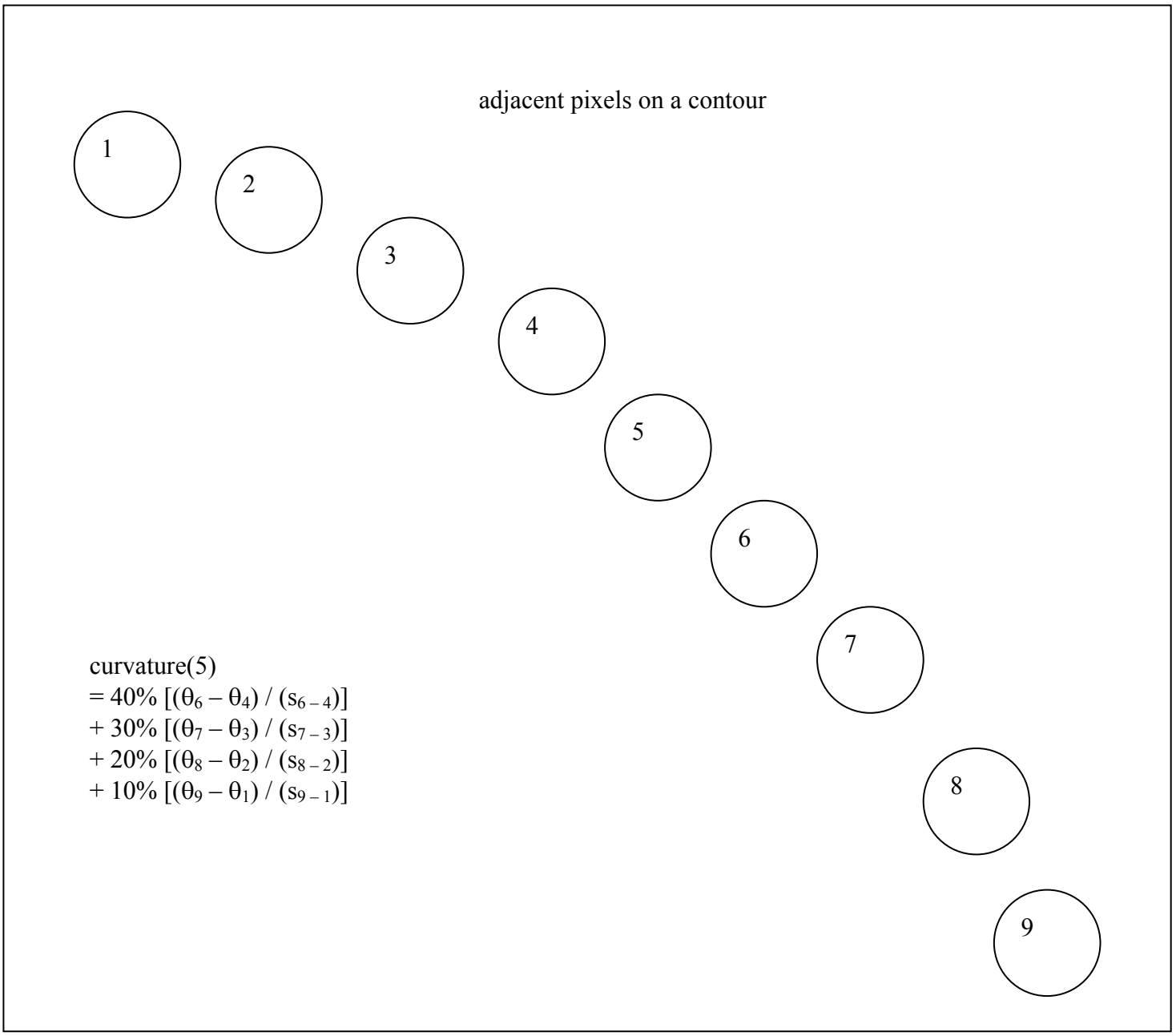

Figure 3.2 – Method of calculation of curvature at each pixel.

angle subtraction formulas are corrected modulo 360° so that, for example, the difference between a 350° orientation and a 10° orientation is 20°, not 340°.

Once the curvature had been calculated at each pixel, cocircular segments were defined. A range (max curvature − min curvature) of curvature values for all contours was computed. Adjacent pixels on the same contour were deemed cocircular if the difference in their curvatures was below a set threshold, expressed as a percentage of this total curvature range.

All simulations were carried out using the MATLAB application development environment (version 6 R12) and the associated Image Processing Toolbox (version 3.1).

3.3 Results

We first investigated a number of technical image processing issues. We carried out the measurements with and without an initial contour-thickening step in an attempt to smooth out the thinned contours. Results were similar in both cases.

Our algorithm is parameterized to allow subsampling of the contours (for example, start at pixel x of the contour and consider only every y pixels). However, the results that we present represent a consideration of all pixels. We also considered calculating curvatures using the slope of the line connecting the pixels for θ (instead of the steerable filter result)

and the Euclidean distance between pixels for s (instead of the arc length). Both of these approaches were abandoned because of poor results. Other qualitatively poorer results were obtained using Gabor filters instead of steerable filters, performing calculations with overlapping segments of the contours, and mid-pixel averaging if greater than threshold. All of these techniques were subsequently dropped.

Decomposition of the image yielded three closed contours, three long background contours, and a number of short, open cross-hatches. The mean and standard deviation of the curvatures for the closed and background contours were computed. As expected, the mean curvature for each closed contour (0.0202, 0.0204, 0.0185) was close to the inverse of its approximate radius (0.0263, 0.0258, 0.0259, respectively), computed by dividing the contour length by 2π. The sign of the curvature is generally determined by the direction of contour traversal (clockwise or counter-clockwise). The standard deviation of the curvature over the closed contours (0.0284, 0.0310, 0.0343) was significantly less than that of the background contours (0.0840, 0.0869, 0.0877). Thus, overall, the circular contours are more cocircular than the background contours. However, it remained to be determined whether portions of the circular contours were more cocircular than portions of the background. This question can be assessed by determining, for each pixel, how far one can move along the contour until the curvature deviates past a fixed threshold. For example, in Figure 3.3, the color of the pixel (and accompanying color bar) indicates the number of consecutive adjacent pixels on each contour that remain within a threshold of 45% of the total curvature range of all of the circular contours. It is apparent that at this

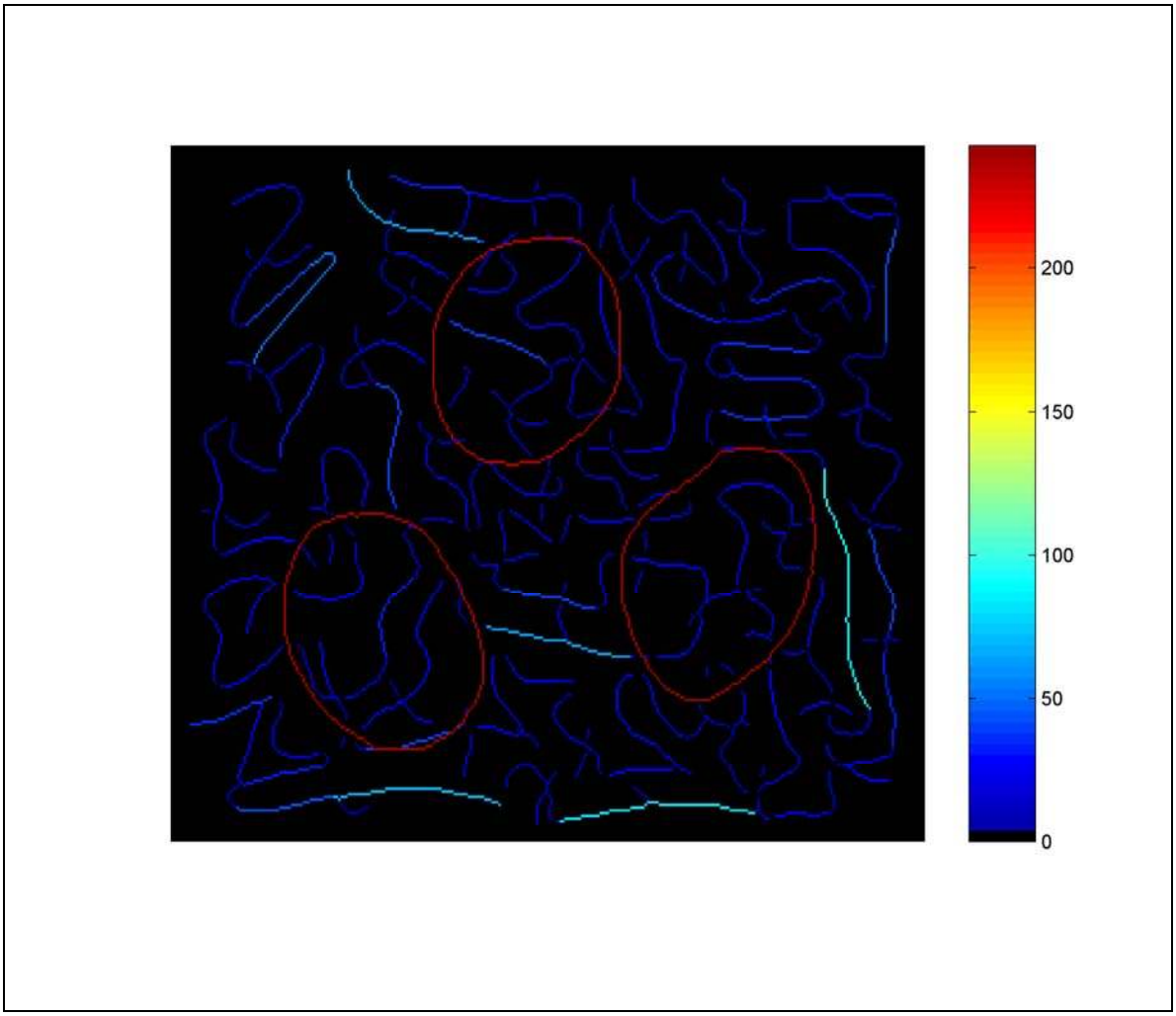

Figure 3.3 – Threshold at 45% of curvature range.

threshold the closed contours contain much longer stretches of cocircularity, compared to the background contours and cross-hatches.

Figure 3.4 provides a similar computation when the threshold has been set to 35% of the maximal curvature range. This stricter threshold cuts down on the extent of contours conforming to the curvature constraints. Although the closed contours still contain longer segments meeting this criterion, the difference between the closed contours and background squiggles is less dramatic.

Figure 3.5 shows the results with a threshold set at 15% of the maximal curvature range. As the definition of cocircularity becomes stricter, i.e., curvature is constrained to a narrower range of values, the differences between the circular and background contours diminishes. These results therefore show that at a local scale, all contours in the image are similar. What distinguishes the circular contours (apart from closure) is that they maintain a similar curvature over a much longer extent, compared to the squiggles, where "similar" can be quantitatively defined. In other results, not shown here, we have found that further loosening of the curvature constraints (i.e., allowing variation over 50% or more of the maximum range), yields similar results to the 45% case, up to a limit. At this point (within a threshold of approximately 75% of the maximal curvature range), the definition of cocircularity is so loose that its discriminatory power begins to diminish. Large portions of all contours (circular and background) begin to appear cocircular.

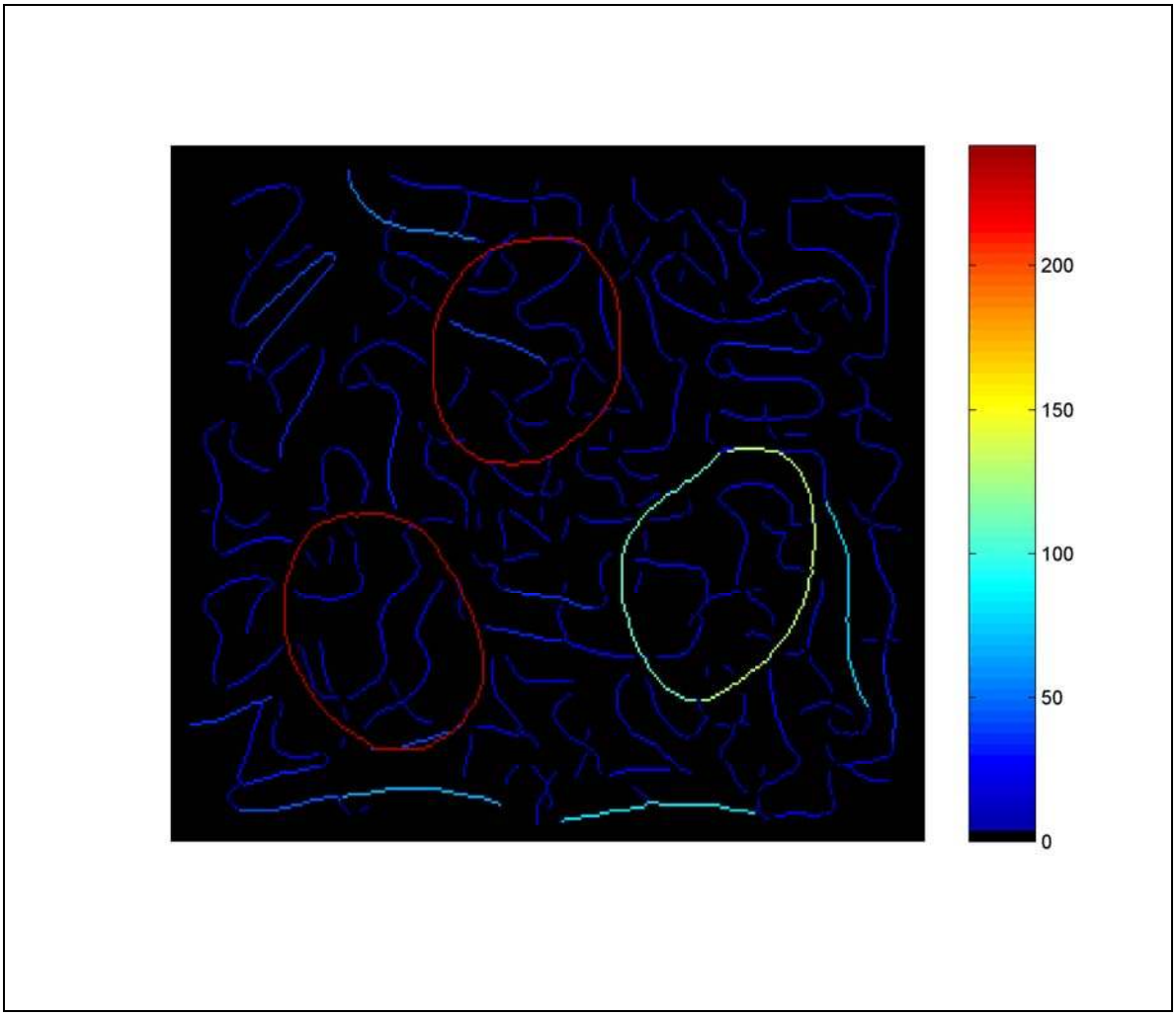

Figure 3.4 – Threshold at 35% of curvature range.

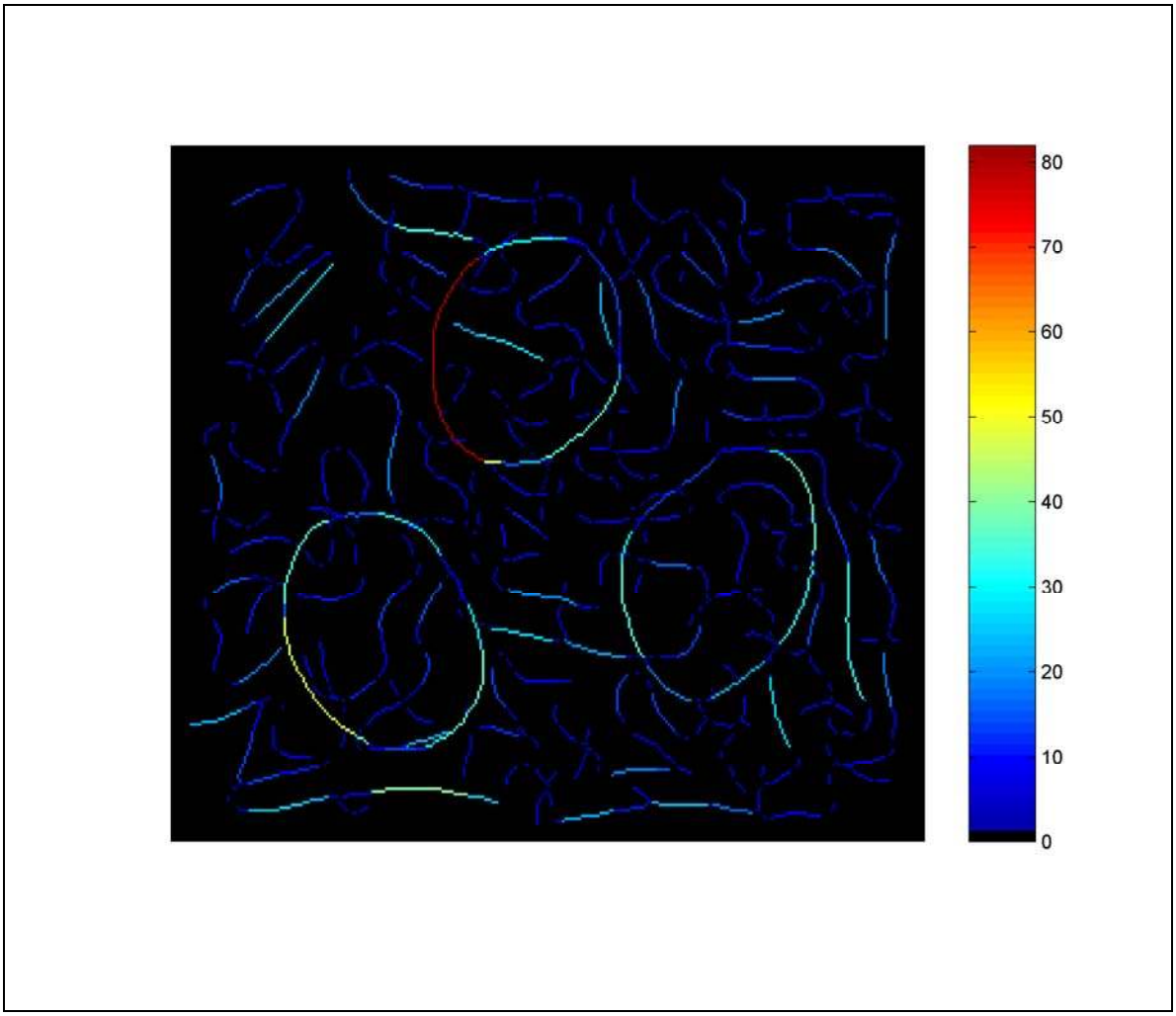

Figure 3.5 – Threshold at 15% of curvature range.

3.4 Discussion

If curvature covariation is a factor in determining perceptual salience then Figure 3.3 argues that the circular contours should be more salient than the background. However, as Figures 3.4 and 3.5 demonstrate, the degree to which curvature covariation contributes to salience will depend upon the mechanisms and the scale over which curvature information is computed in visual cortex.

A number of studies have proposed neural and computational mechanisms for computing curvature (Dobbins *et al.*, 1989). These lower level curvature calculations are consistent with mechanisms, such as end-stopping, available in primary visual cortex. At higher cortical levels, Van Essen, Connor and colleagues have shown that cells in areas V2 and V4 are selective both for the magnitude and direction of curvature (Hegdé and Van Essen, 2000; Pasupathy and Connor, 2001).

Our results show that curvature values on the circles are much more similar over longer extents. A cortical mechanism for distinguishing the circles could thus be based on the similarity of firing rates of curvature sensitive cells in V4. For example, Hopfield & Brody (Hopfield and Brody, 2001) have proposed a mechanism in which groups of cells with similar firing rates synchronize. Synchronization occurs naturally in the types of cortical architectures studied by Connors and colleagues (Beierlein *et al.*, 2000). Hopfield and Brody make the point that in a large ensemble of cells, a large fraction of

cells firing at the same rate is statistically unlikely. Thus, a set of connected V4 cells, each sensitive to magnitude and direction of curvature, which are coupled by horizontal connections, would rapidly synchronize in response to the circles, but to a much lesser degree to the background squiggles.

Salience depends upon the degree to which a target differs from the background. Rosenholtz (Rosenholtz, 1999) has proposed that a possible metric for salience is to consider the Mahalanobis distance between a target and the background. Thus, salient targets are those whose features are statistical outliers from the background population.

For the Ullman figure, the circular contours are statistical outliers by virtue of their consistent curvature relative to the fluctuating curvatures of the background. The contours are also outliers with respect to closure, as discussed above. One might ask, however, why the background squiggles are not salient since they statistically differ compared to the circles. One possible answer relates to Rosenholtz's (Rosenholtz, 1999) explanation of salience asymmetries. As illustrated in Figure 3.6, the standard deviation of curvatures on a circle is much less than on a squiggle. Thus, curvatures on the circle are more consistent, and lie several (circle) standard deviations away from the mean curvatures of the squiggles. But the mean curvature of the squiggle lies close (in terms of the standard deviation of the squiggle distribution) to the mean of the circle curvatures. The relative consistency of the circle curvatures distinguishes them from the background.

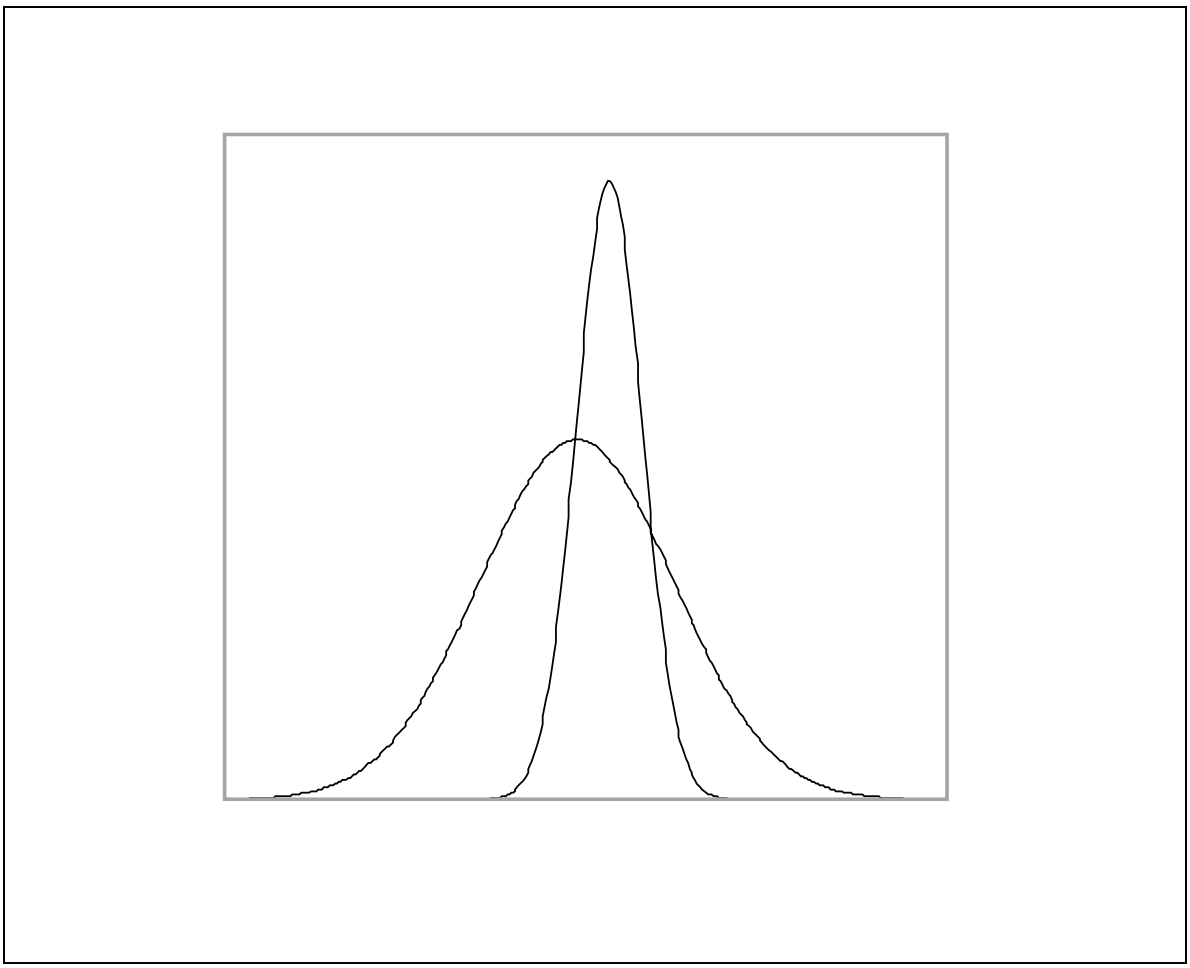

Figure 3.6 – Gaussian model of circle vs. background curvatures.

Area V4 is well situated to carry out computations of the type envisioned here. Connor and colleagues have found cells in macaque V4 that respond selectively to the magnitude and direction of curvature. V4 contains an extensive network of horizontal connections, which span significant portions of the visual field (Amir *et al.*, 1993). Through these connections, possibly together with top-down information from higher temporal areas, V4 is thought to mediate contextual population-based interactions within the scene, such as those required in color constancy. Finally, V4 is known to play a critical role in determining salience. Lesions of V4 render animals incapable of detecting a less salient target in the presence of more salient distracters (Schiller, 1993).

3.5 Conclusion

For the single image considered here, the consistency of curvature on a contour is correlated with increased perceptual salience. Salience appears to depend upon the statistical distribution of feature properties, such as curvature, over the object, in comparison to the background. These findings support the earlier work of both Ullman and Rosenholtz. The need to compare curvature measurements over a significant image region, to segment contours into target vs. background, and possibly to discriminate the direction of figure, suggest the need for both horizontal as well as top-down information, possibly at the level of V4, in determining contour salience.

Chapter 4

Shape Representation by a Network of V4-like Cells

4.1 Introduction

The representation of contour shape is an essential component of object recognition. Human observers are exquisitely sensitive to changes in contour shape (Zana and Cavalcanti, 2005), and this capacity develops in early infancy (Norcia *et al.*, 2005). However, the cortical mechanisms underlying shape analysis are incompletely understood.

Visual cortical neurons in monkey extrastriate cortex are sensitive to the fine structure of contour shape. Units in area V2 respond differentially to a variety of local contour configurations including angles, arcs and intersections (Hegdé and Van Essen, 2000;

Hegdé and Van Essen, 2003). In area V4, Pasupathy and Connor have described cells that are selective for a particular local shape configuration at a particular location on the contour (Pasupathy and Connor, 2001). For example, one neuron might prefer contours with a sharp concavity at the northwest corner; another unit might be selective for contours with a shallow convexity at the southernmost tip. The response of these cells is further modulated by the local contour configurations at neighboring locations on the contour. Thus, the unit preferring a sharp concavity at the NW corner might be potentiated by a sharp concavity located immediately clockwise along the contour, or suppressed by a convexity located immediately counterclockwise. Pasupathy and Connor have demonstrated that a population of such V4 units can provide a detailed description of the contour shape (Pasupathy and Connor, 2002).

We seek to evaluate the ability of a simulated population of these "V4-like cells" to recognize objects in a standard object recognition test bed — the MNIST database of handwritten digits. We construct V4-like units with sensitivities to local curvature, relative location of the contour configuration on the contour, distance of the contour configuration from the center of mass, convexity / concavity, and modulatory effect of adjacent contour locations. Stimuli (2-dimensional closed contours representing object boundaries) evoke a pattern of activity across the population of V4-like cells. Shape recognition is evaluated by demonstrating that the patterns of activity across the units to members of a particular object class, e.g., all images of the digit "2", resemble each other to a higher degree than they resemble members of any other class, e.g., images of the

digit "3". This measure corresponds to that reportedly used by humans and monkeys in object classification (Sigala *et al*., 2002).

For the purposes of this study, we segment each contour into iso-curvature regions, and dedicate a V4-like unit to describing each region (its curvature, orientation, distance from object center, etc.). This approach is computationally tractable and connects conceptually to previous models of "parts-based" recognition (Biederman, 1987; Riesenhuber and Poggio, 1999; Ullman *et al*., 2001). However, there is no theoretical need for this initial segmentation stage – a large population of V4-like cells could describe overlapping regions of the contour in an over-complete representation.

As an initial test, we directly compare the accuracy of recognition using V4-like units to the "shape-context" approach of Belongie and colleagues (Belongie *et al*., 2002). Belongie's algorithm describes contour shape by computing, for each point on the contour, the distance and direction to every other point on the contour. The shape context at each point is represented by a histogram of distances to other points versus direction (distance and direction are each digitized into ~5 and 12 bins, respectively). Shape matching is accomplished through a sophisticated process of image registration – optimal point correspondence followed by a distance measure. To allow a direct comparison of approaches, we use Belongie's registration and correspondence routines, and simply substitute the V4-like curvature-based estimate in place of the shape context histogram. Recognition accuracy using the two descriptions (shape context vs. curvature) is found to be roughly comparable.

We then test the performance of a population of V4-like cells on the MNIST database and the MPEG-7 Shape Silhouette database (Jeannin and Bober, 1999) and evaluate accuracy of classification. We analyze the contribution of sensitivities to different contour features (position, curvature at neighboring locations, etc.), and find that local curvature is the most critical component of the shape description.

4.2 Methodology

To establish the feasibility of using V4-like units for shape recognition, we directly compare the V4-like shape representation with Belongie and colleagues' "shape context" algorithm – a state-of-the-art category-level recognition system (Belongie *et al*., 2002). Belongie and colleagues propose a shape description based on the distances between all pairs of points on the object's bounding contour. Shape context log-polar histograms are computed for each point on the contour. The collection of histograms fully characterizes each shape. Belongie and colleagues solve the correspondence problem between two shapes using optimal assignment. They estimate the aligning transform using these correspondences and regularized thin-plate splines. Finally, they measure similarity between the shapes as a function of matching errors between corresponding points and aligning transform magnitude. We employ Belongie's exact algorithms, but substitute V4-like curvature measurements in place of Belongie's shape context descriptors.

We next develop a stand-alone model of V4-like units to evaluate recognition performance. Simple image processing techniques are used to decompose the image into closed contours, each of which is analyzed independently (e.g., the inner and outer contours of a digit "9" (see Figure 4.2)). For each image, we extract contours using the numerical gradient, and determine a set of boundary points with oriented tangents. The result is a parametric description (x(t), y(t), and tangent(t)) of each contour. For each point along the contour, we compute its angle (0° – 360°) relative to and distance (in pixels) from the image's center of mass. For consistency, we consider the contour's points such that each contour begins at approximately 3 o'clock and proceeds counterclockwise.

Using the parametric forms x(t) and y(t), where $0 < t < L$, L being the length of the curve, the curvature is given by:

$$\kappa(t) = ((dx/dt) (d^2y/dt^2) - (d^2x/dt^2) (dy/dt)) / ((dx/dt)^2 + (dy/dt)^2)^{3/2} .$$

Several approaches have been developed for extracting and representing curvature information. The curvature scale space (CSS) shape representation for planar curves developed by Mokhtarian and Mackworth (1986, 1992) is based on identifying inflection points on the curve at varying levels of detail. This is accomplished by convolving several Gaussian kernels of different space constants with the curvature function. This technique has proven successful in computer vision applications that recognize curved objects (Mokhtarian, 1995). The curvature-tuned smoothing (CTS) method of Dudek and

Tsotsos (1997) is another technique for shape representation and recognition of objects. Based on smoothed multiscale curvature information and using a dynamic programming matching strategy, it is related to the CSS representation. Wuescher and Boyer's (1991) formulation is something of a departure from Mokhtarian. Following their methodology, we convolve x(t) and y(t) with the derivative of a Gaussian to both smooth (regularize) and differentiate the functions. This results in the discrete curvature, parameterized by the Gaussian space constant:

$$\kappa(t, \sigma) = ((x(t) * g'(t, \sigma)) (y(t) * g''(t, \sigma)) - (x(t) * g''(t, \sigma)) (y(t) * g'(t, \sigma))) / (((x(t) * g'(t, \sigma))^2 + (y(t) * g'(t, \sigma))^2)^{3/2} ,$$

where

$$g(t, \sigma) = (1 / \sigma\sqrt{2\pi}) \exp(-t^2 / 2\sigma^2) ,$$

$*$ is the convolution operator, and $'$ is the differentiation operator.

It should be noted that the curvature of the pre-digitized object cannot be calculated exactly. It can only be estimated, with a lower bound on the error of the achievable measurement (Worring and Smeulders, 1993).

Following Pasupathy and Connor (2001), this curvature may optionally be squashed:

$$\kappa_{squashed} = (2.0 / (1 + \exp(- 0.125 * \kappa))) - 1.0 .$$

We define the *direction of curvature* to be orthogonal to the tangent and to point towards the interior of the closed contour (Sajda and Finkel, 1995). It is computed using the orientation and inverse tangent.

Next, we segment each contour into iso-curvature regions using one of two methods. The first method is simplistic. We choose a standard size (in number of boundary points) for each region. Remaining points are evenly distributed. We choose a starting point on the contour (and therefore the starting point for each region) based upon the arrangement that yields the lowest average standard deviation of curvature for each region. Ideally, each iso-curvature segment has a zero standard deviation of curvature.

The second method, following Wuescher and Boyer's (1991) curvature voting technique, considers segments of constant curvature (within a curvature tolerance t_c) and segments of rapidly changing curvature. Curvature is quantized into bins of a specified width (typically $1/2t_c$ bins for every span of 0.10 in curvature), with peaks of the resulting histogram representing the curvature values most likely to fit the longest segments. Continuing with their methodology, the longest contiguous, constant curvature segments in the most prevalent curvature ranges are repeatedly extracted. The leftover portions of the contour are either absorbed by adjoining segments or become segments themselves. The pre-determined minimum segment length (l_{min}) reflects the amount of segment detail.

In the experiments described below, unless otherwise noted, a standard size (6 boundary points) for each iso-curvature region is chosen for segmentation. In addition, all contours are smoothed using a Gaussian function with a standard deviation of 2 pixels (a sigma value of 2).

For each iso-curvature segment of each image, we create feature vectors composed of various combinations of features chosen from the following: mean polar angle of contour region (e.g., 3 o'clock), mean curvature of the region, mean curvature of the clockwise adjacent region, mean curvature of the counterclockwise region, mean direction of curvature of the region, mean distance from the center of mass of the region, an indication of whether the region belongs to an inner vs. outer contour, and Gaussian averaged response of the region – the simulated response of a population of optimally positioned neurons across the curvature × position domain (see Figure 4.5 below). These features all have approximate neurobiological correlates in area V4 and other extrastriate areas (Desimone and Schein, 1987; Kobatake and Tanaka, 1994; Gallant *et al.*, 1996; Pasupathy and Connor, 1999; Wilkinson *et al.*, 2000; Zhou *et al*., 2000; Pasupathy and Connor, 2001; Pasupathy and Connor, 2002).

It is worth noting several general differences between "curvature context", as used for the direct comparison with the shape context algorithm, and the implemented V4-like units used later in this paper. First, in the direct comparison with shape context, radial distance is not considered in the computation of "curvature context" (only curvature and angular

direction are employed). In our stand-alone framework of V4-like cells considered later, however, radial distance is one of the features that is examined. Secondly, "curvature contexts" are curvature / angular direction histograms and, like shape contexts, are computed with respect to every point on the contour. In our stand-alone framework of V4-like cells, however, curvature, angular direction, distance, etc., are computed with respect to the center of mass.

Recognition, both identification and classification, requires some measure of matching – that is, some way of describing two objects / images as being more or less similar to each other. We have used various methodologies for comparing groups of segments, including histogram cross correlations for curvature-angular position histograms, minimum sum of squared differences, image cross correlation, appearance-based parametric eigenspaces, and Support Vector Machines (SVMs) (Müller *et al.*, 2001; Schölkopf *et al.*, 2001). The results described here are obtained using the Earth Mover's Distance (EMD) comparison, as elaborated by Rubner and colleagues (Rubner *et al*., 2000; Rubner *et al.*, 2001). EMD considers two distributions, represented by signatures (sets of weighted features), and is defined as the minimal work or cost to transform one signature into the other (i.e., filling the "collection of holes" of one distribution with the "properly spread mass of earth" of another distribution). Because the concept of work is based upon the user-defined ground distance, which in turn should be based upon perceptually meaningful distance measures between individual features, this technique is reported to match perceptual similarity judgments better than other distance measures. For the computation of the EMD between color images, for instance, the ground distance between individual colors should correlate

strongly with human color discrimination performance (i.e., distance approximately matches human perception of the differences between those colors) (Rubner *et al*., 2000; Wyszecki and Stiles, 1982; Tversky, 1977). Another advantage is that angle comparisons (for example, 358° is very close to 2°) can easily be handled.

All simulations were carried out using the MATLAB application development environment (version 7.3.0.267 R2006b) and the associated Image Processing Toolbox (version 5.3).

4.3 Results

Two example digits (each a "2") from the extensive MNIST database of digitized handwritten digits are shown in Figure 4.1 (A–B). Calculated "curvature contexts" are shown (Figure 4.1 (E–F)) for the circled sample points on the image contours. The histograms have 5 bins for the curvature values and 12 for angular direction. Note the similarity between the histograms for the two example digits. Two additional digits (each a "9") are shown in Figure 4.1 (C–D). Note the visual dissimilarity between the histograms for the "2's" (Figure 4.1 (E–F)) versus the "9's" (Figure 4.1 (G–H)), which suggests that curvature context may be a useful shape representation.

Using Belongie's algorithm, but substituting curvature context for shape context, the degree of match between two "2" digits from Figure 4.1 is determined. We calculate the

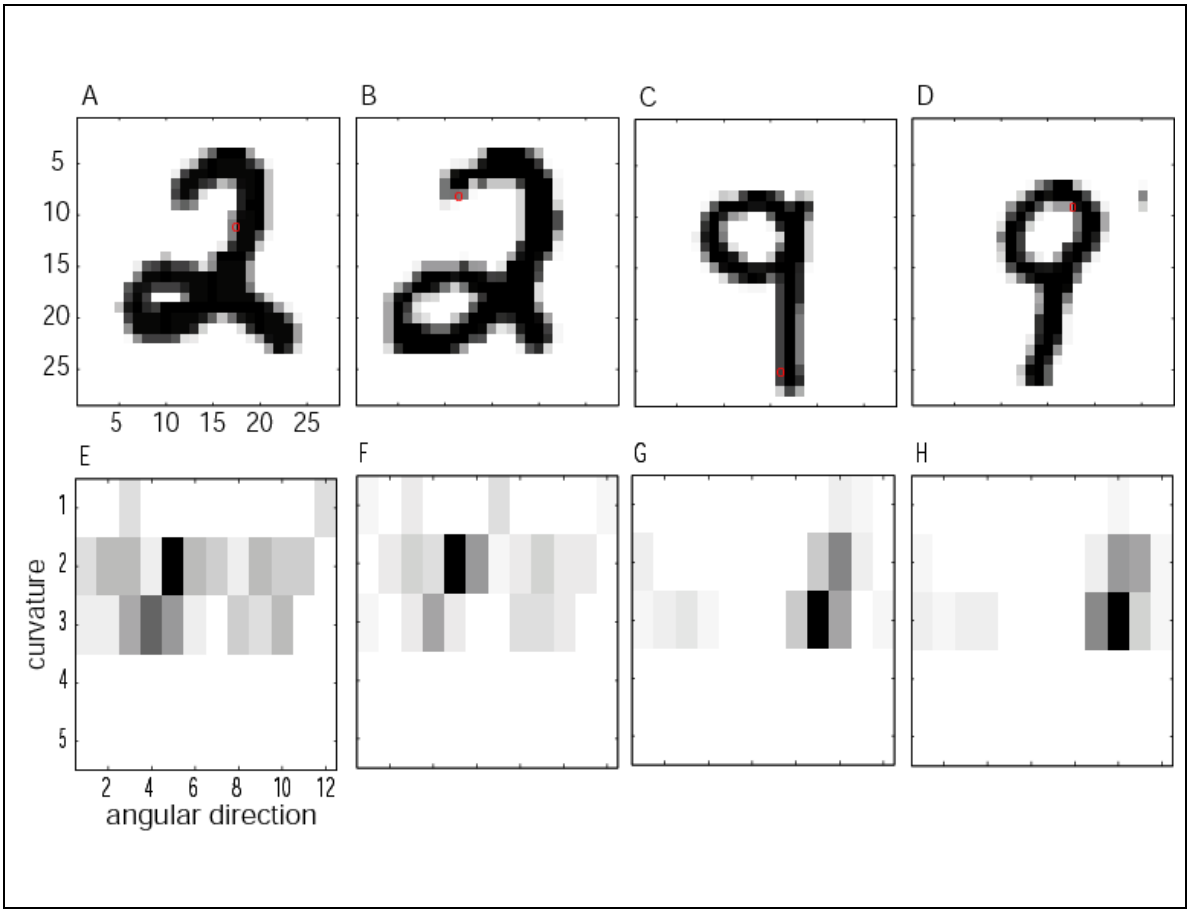

Figure 4.1 – Curvature histograms. We employ the framework of Belongie and colleagues, but substitute our curvature measurements for his shape context descriptors. (A–B) Two different stimuli, each representing the digit "2". (C–D) Two different stimuli, each representing the digit "9". Curvature contexts are computed with respect to the circled sampled points. (E,F,G,H) The corresponding histograms for the stimuli. Histograms have 5 bins for the curvature value and 12 for the angular direction.

untransformed correspondences between the points of each of the two images using optimal assignment (Figure 4.2 (A)). We then estimate the aligning transform between the two images using these correspondences and regularized thin-plate splines, measuring affine and matching costs (Figure 4.2 (B)). We present a representation of the first image after it has been warped into the second image (Figure 4.2 (E)). Finally, we compute the local sum of squared differences and average local sum of squared differences of this warping (Figure 4.2 (F)). Figure 4.2 (C,D,G,H) shows the same calculations for matching the two "9" digits from Figure 4.1.

We compare our results, after only one iteration, with those reported by Belongie and colleagues. For each numeric digit ("1's", "2's", etc.), two example images are randomly selected from our test set and compared to each other. We use Belongie's shape context algorithm to compute the matching cost and SSD and compare the results to those obtained using curvature contexts. In both measures, lower values correspond to better matches. As shown in Table 4.1, "shape context" and "curvature context" shape representations achieve comparable recognition accuracies.

The values obtained for the closeness of match depend quantitatively upon the details of segmentation and curvature estimation. The basic contour parameters: x, y, and tangent, the discrete and smoothed curvature, and the curvature direction, are shown in Figure 4.3 for a sample digit image. Both methods of iso-curvature segmentation are illustrated in Figure 4.4 for the same digit image, with different parameter value choices shown for each method. Note that the iso-curvature segments of an inner contour, when present, are

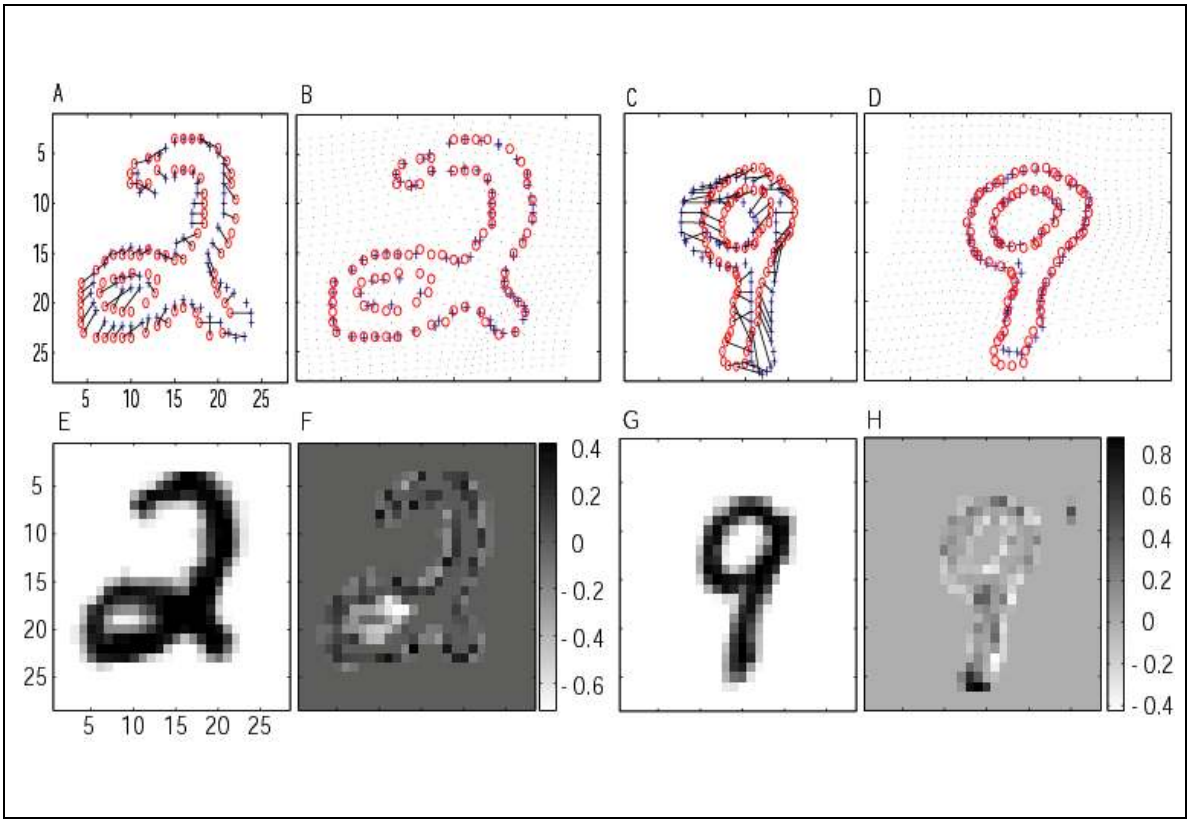

Figure 4.2 – Corresponding points, matching costs, and sum of squared differences. Belongie's algorithm is used, with curvature context substituted for shape context. (A) The untransformed correspondences between the images of Figure 4.1 (A–B). (B) The transformation, with affine cost = 0.23553 and matching cost = 0.10746. (C) The untransformed correspondences between the images of Figure 4.1 (C–D). (D) The transformation, with affine cost = 0.26059 and matching cost = 0.18963. (E) A representation of the stimulus of Figure 4.1 (A) after warping to the stimulus of Figure 4.1 (B). (F) The local sum of squared differences (=0.047503) and average local sum of squared differences (=0.04682) of this warping. (G) A representation of the stimulus of Figure 4.1 (C) after warping to the stimulus of Figure 4.1 (D). (H) The local sum of squared differences (=0.048906) and average local sum of squared differences (=0.048653) of this warping.

digit	shape context matching cost	shape context SSD	curvature context matching cost	curvature context SSD
"0"	0.0555	0.0414	0.0946	0.0474
"1"	0.0616	0.0426	0.1427	0.0493
"2"	0.0703	0.0446	0.0783	0.0494
"3"	0.0576	0.0391	0.0635	0.0445
"4"	0.0652	0.0422	0.1065	0.0483
"5"	0.1048	0.0685	0.0894	0.0528
"6"	0.0769	0.0473	0.0892	0.0568
"7"	0.0877	0.0429	0.0921	0.0373
"8"	0.0569	0.0389	0.0635	0.0288
"9"	0.0570	0.0390	0.1146	0.0395

Table 4.1 – Comparison of results. Two example images were randomly selected for each numeric digit. Matching costs and sum of squared differences for the shape context vs. curvature context comparison are shown.

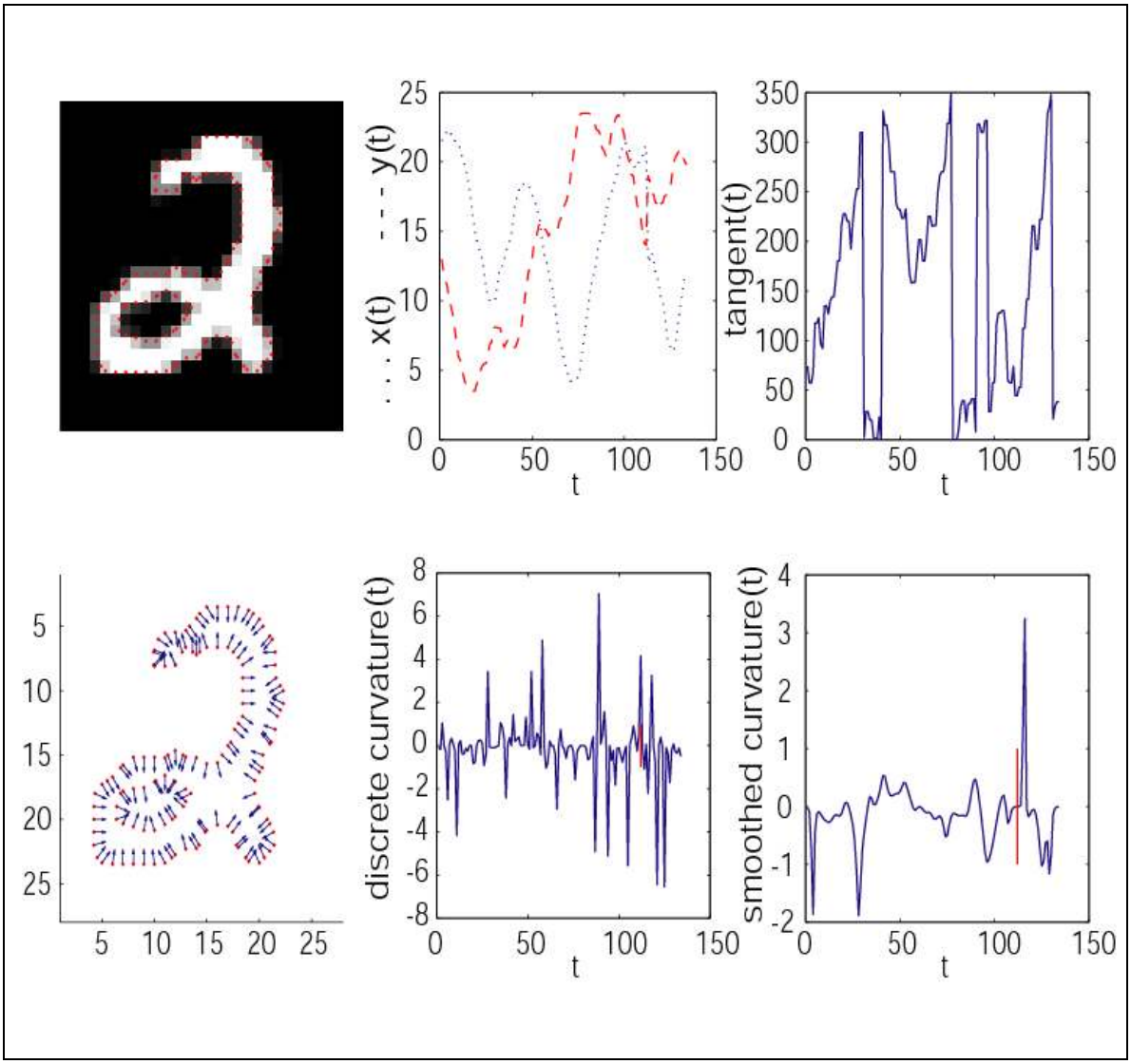

Figure 4.3 – x, y, tangent, and curvature values. The sample points of the original image are shown, as are the computed curvature directions (arrows). For the smoothed curvature, the vertical line separates outer and inner contours. A sigma value of 2 is used for smoothing.

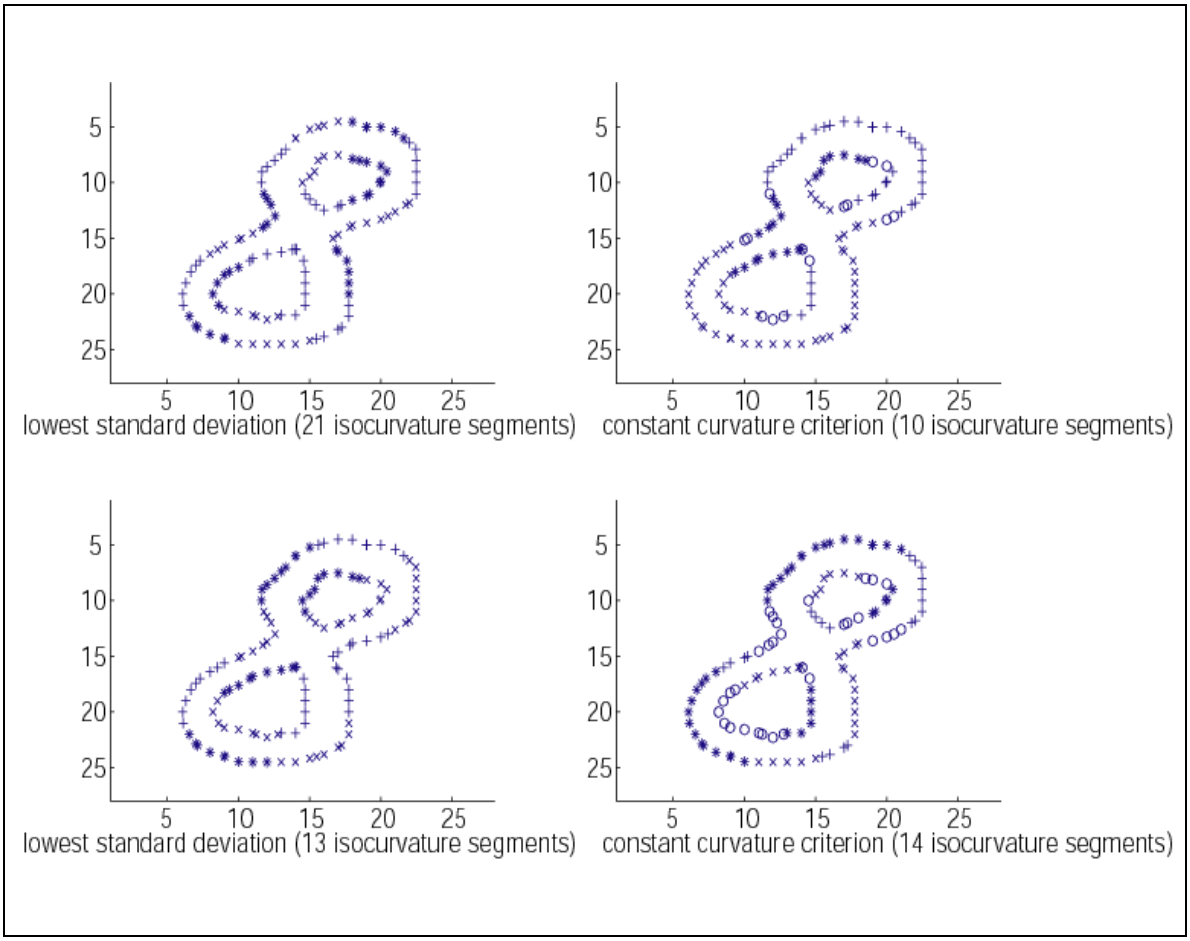

Figure 4.4 – Segmentation. Two different methods, each with 2 sets of parameter values, are shown. Individual iso-curvature regions (segments) are shown as different symbols. For the lowest standard deviation method (the left column), different region sizes are used (top left, region size = 6, bottom left, region size = 9). For the constant curvature criterion method (the right column), open circles indicate areas "filled in" (1 - coverage). For the top right, the curvature tolerance = 0.2, the bin size = 0.04, and the minimum segment length = 5, resulting in a coverage (areas not “filled in”) of 88.7%. For the bottom right, the curvature tolerance = 0.1, the bin size = 0.02, and the minimum segment length = 4, resulting in a coverage of 75%. The resulting number of iso-curvature segments for each digit is indicated.

not required to align with a corresponding outer contour segment. In the bottom left image of Figure 4.4, for example, an inner contour segment begins at approximately 9 o'clock and ends at approximately 1 o'clock, while its corresponding outer contour segment begins at approximately 9 o'clock and ends at approximately 11 o'clock. The standard region size with lowest standard deviation method is shown in the left column. The constant curvature criterion method is shown in the right column, with open circles representing leftover portions. The segmentations appear qualitatively reasonable, approximating the expected results if the segmentation were performed "by eye". For the images analyzed, we find that a segmentation region size of 6 pixels results in the greatest number of iso-curvature segments (21), while the curvature tolerance 0.2 results in the least (10). The degree to which the segmentation affects system performance is considered below.

The simulated response of a population of optimally positioned neurons (corresponding to Pasupathy and Connor's (2002) Figure 2) across the curvature × position domain is shown in Figure 4.5. The responses are modeled as 2-dimensional Gaussian functions, centered ("+" symbol) at the mean squashed curvature and angular position of each iso-curvature segment. The standard deviation values that are used are those found by Pasupathy and Connor (2001) in V4 neurons. The color-coded amplitude in the resulting summed population activity corresponds to the likelihood that a similar curvature / angular position combination is present, with peaks representing the salient boundary features. Two example image digits (each a "5") are shown in the left panels. Note the similarity between the digits. Two additional digits (each a "7") are shown in the right

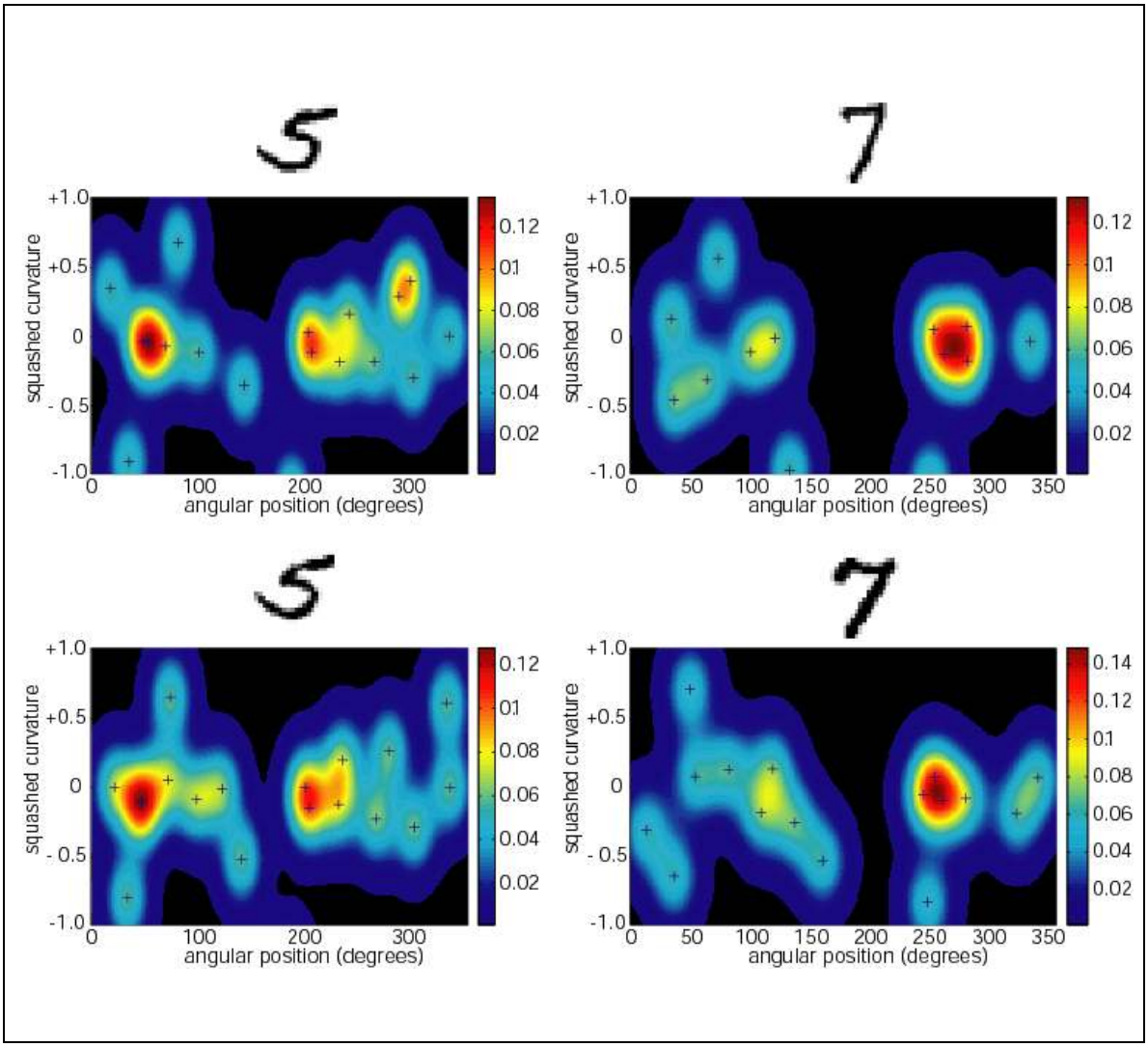

Figure 4.5 – 2-dimensional Gaussian population responses. The left column contains two examples of the digit “5”. The right column contains two examples of the digit “7”. The means (Gaussian peaks marked with a "+") are from the iso-curvature segments. The standard deviations are from Pasupathy and Connor. The overlays are the actual images that elicited the population responses.

panels. Note the visual dissimilarity between the left and right panels, again suggesting that some level of category discrimination is achieved in the curvature × position domain.

A 4-dimensional Gaussian tuning function (corresponding to Pasupathy and Connor's (2001) Figure 3), for three adjacent boundary elements, is shown in Figure 4.6. Here, each individual surface plot represents the 2-dimensional Gaussian response of the central boundary element in the squashed curvature × angular position domain. The rows and columns of plots represent the modulatory effects of adjacent contour elements. The rows correspond to different arbitrarily assigned clockwise-adjacent curvatures, while the columns correspond to different arbitrarily assigned counterclockwise adjacent curvatures. A hypothetical cell responding optimally would exhibit strong tuning for a slight concavity towards the bottom of the image, flanked clockwise by a slight concavity and flanked counterclockwise by a slight convexity. The highlighted portion of the "2" digit exhibits these characteristics.

Recognition Performance of V4-like Units

To initially explore the performance capabilities of the V4-like population, we select 100 digit images from the MNIST database of handwritten digits as stimuli. 100 images is a relatively small data set compared to those used in many character recognition studies (Amit *et al*., 1997; LeCun *et al.*, 1998; Belongie *et al*., 2002; Sebastian *et al*., 2003; Grigorescu and Petkov, 2003) but provides an indication of performance capabilities. (Below we will present results using all 10,000 images in the database.) The 100-image set contains several different example images of each digit. The feature vector for each

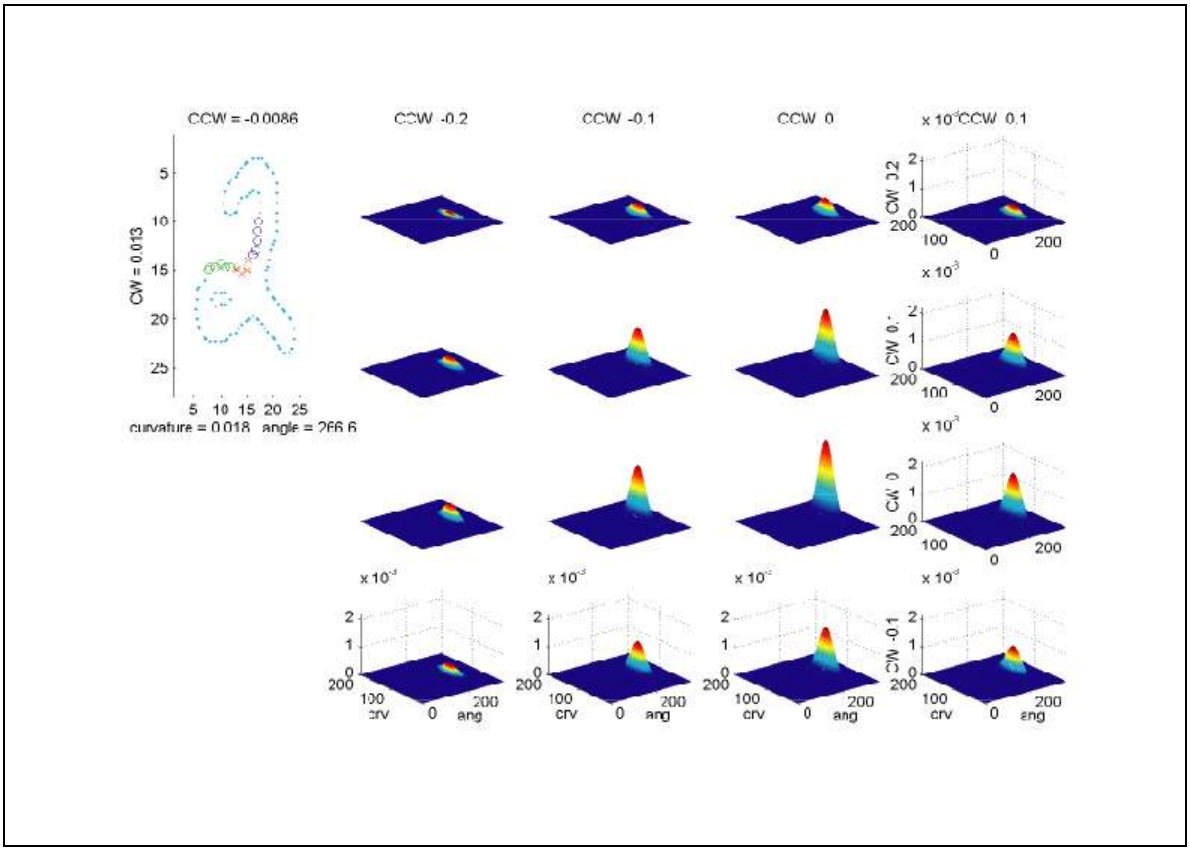

Figure 4.6 – 4-dimensional Gaussian responses, with arbitrarily assigned clockwise- and counterclockwise-adjacent curvatures. Counterclockwise values (left to right): -0.2, -0.1, 0, 0.1 . Clockwise values (top to bottom): 0.2, 0.1, 0, -0.1 . Each individual surface plot represents the 2-dimensional Gaussian response of the central boundary element in the squashed curvature × angular position domain. The rows and columns of plots represent the effects of hypothetical adjacent contour elements. Inset: an example digit, with an iso-curvature segment's (the "x"'s) curvature and angle given. The curvature values of the iso-curvature segment (the roughly vertical segment of open circles) that is clockwise on the contour as well as the iso-curvature segment (the roughly horizontal segment of open circles) that is counterclockwise on the contour are shown.

image is compared to that of every other image, and the average distance from one digit to all other examples of the same digit is determined. This average (earth mover's) distance calculation is motivated by human perceptual studies – evidence suggests that categorization is based on an average fit approximation to the class, rather than on exact matches to prototypes (Kahana and Sekuler, 2002). If an image's lowest average distance is to a group of images representing the same digit then a match is said to have occurred. Otherwise, an invalid classification results.

Various combinations of parameter values and feature vector arrangements are tried in different experiments. The experiment shown in Figure 4.7, for instance, employs feature vectors (one for each iso-curvature segment) composed of the mean angle of the region, the mean curvature of the region, the mean direction of curvature of the region, and the mean distance from the center of mass of the region. Figure 4.7 is the matching matrix, representing the number of images that are classified correctly. The blocks of same digits are read 0 to 9, bottom left to top right. It can be seen that only one image is incorrectly classified – an image of the digit "6" is classified as a "0" – resulting in a 99% correct classification performance. All of the "0" and "6" digits used, including the one outlier "6" digit (in the middle of the figure) that was misclassified, are shown in Figure 4.8. Visual inspection shows that the misclassified "6" could easily be confused for a "0".

As another example, the experiment shown in Figure 4.9 employs feature vectors (one for each iso-curvature segment) composed of the mean angle of the region, the mean curvature of the region, the mean curvature of the clockwise-adjacent region, and the

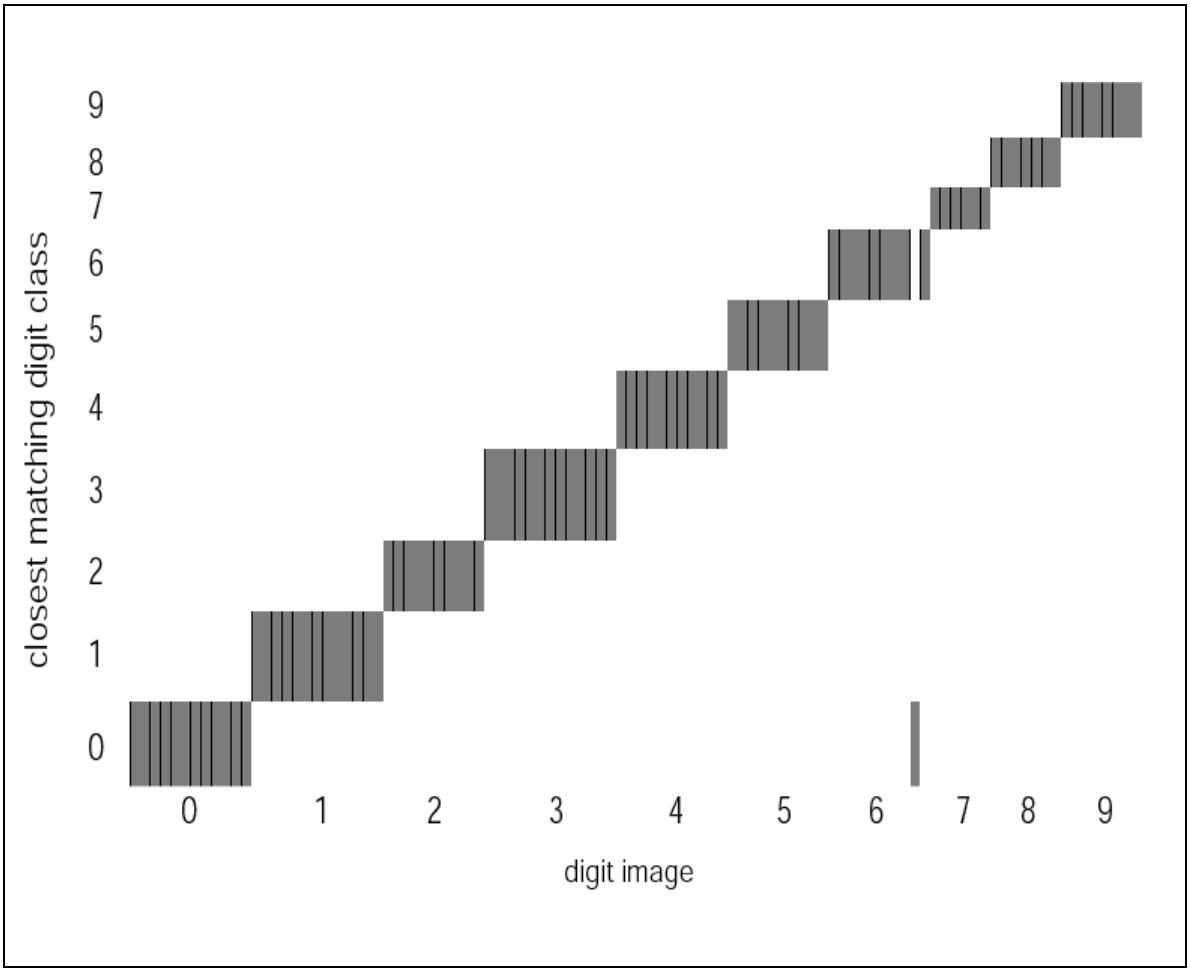

Figure 4.7 – Matching matrix. For each digit image, the closest match amongst the average values for each digit is determined. Correct matches appear as a contiguous rectangle, with width related to the number of sample images for the corresponding digit. Digit values are on the axes. Using the angle, curvature, direction, and distance features, with a region size of 6 and a sigma value of 2, the total matching to average is 99%.

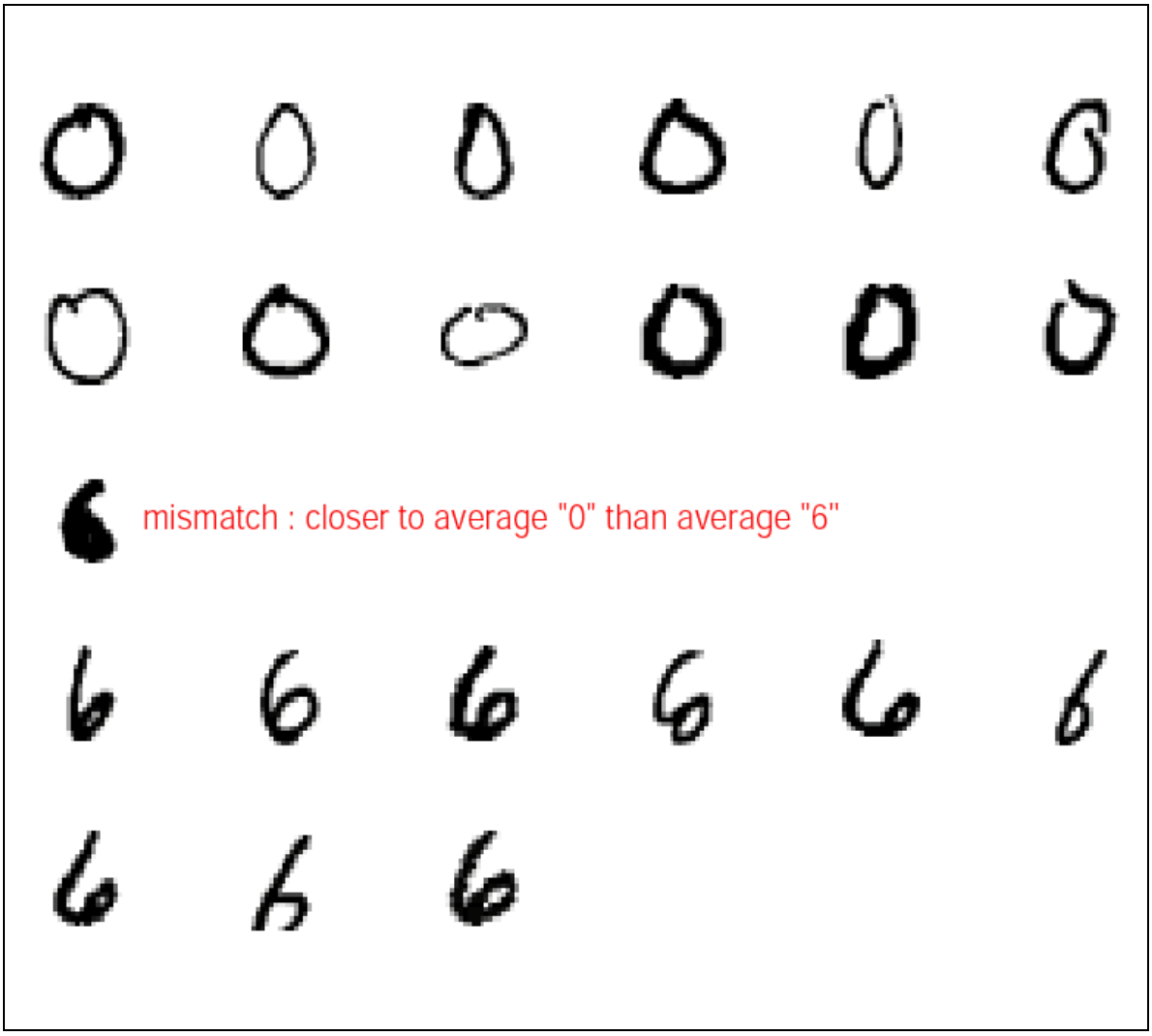

Figure 4.8 – Misclassified digit. The identified "6" digit is classified as a "0". All other digits are classified correctly.

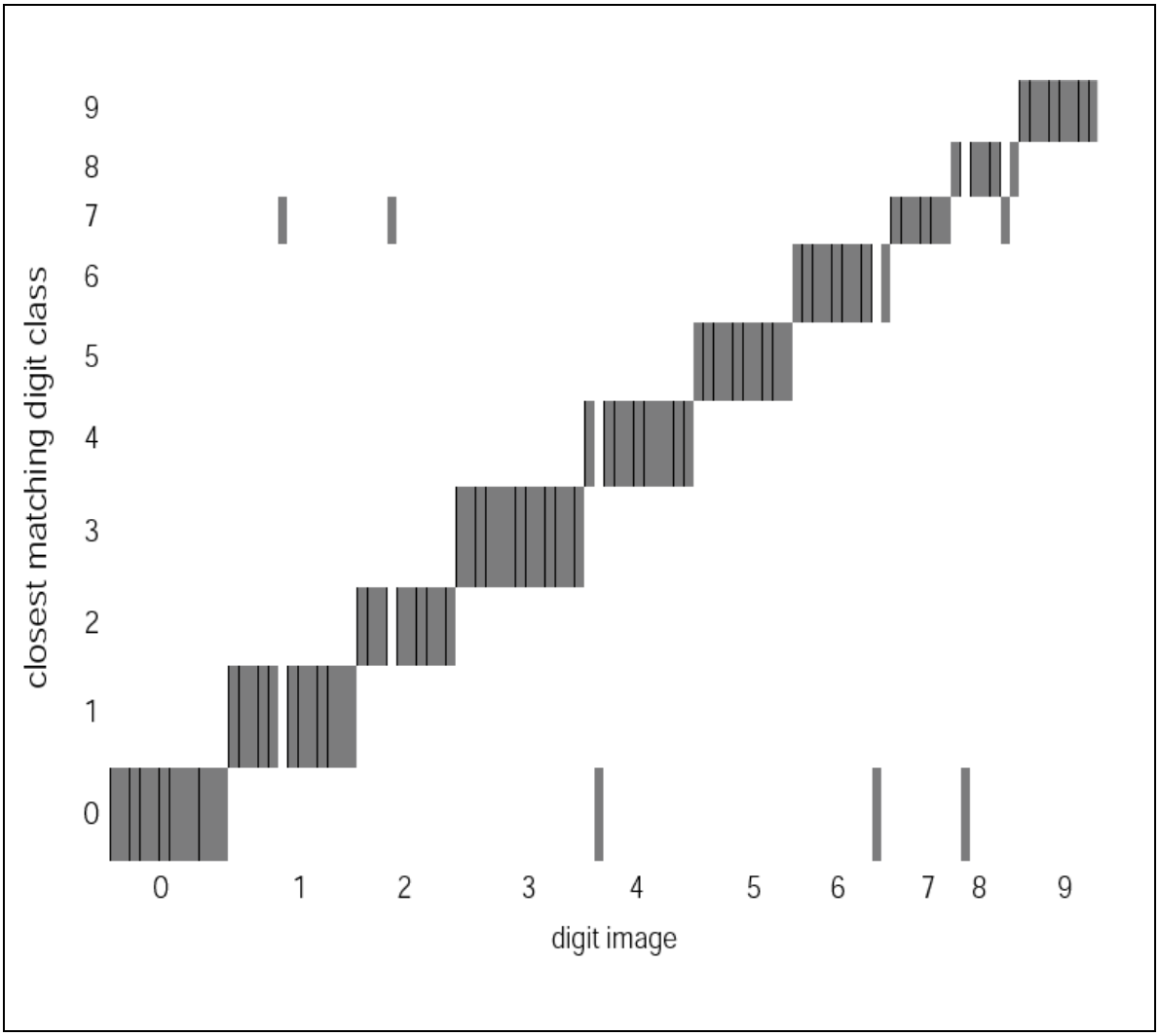

Figure 4.9 – Matching matrix. For each digit image, the closest match amongst the average values for each digit is determined. Correct matches appear as a contiguous rectangle, with width related to the number of sample images for the corresponding digit. Digit values are on the axes. Using the angle, curvature, clockwise curvature, and counterclockwise curvature features, with a region size of 6 and a sigma value of 2, the total matching to average is 94%.

mean curvature of the counterclockwise-adjacent region. Figure 4.9 is the matching matrix, representing the number of images that are classified correctly. It can be seen that a 94% correct classification performance results from this combination of parameters and features. The six incorrectly classified digits include the unusual "6" that is misclassified in Figure 4.7. It would appear that the choice of parameter values and feature vector arrangements used in this experiment is somewhat inferior to the choice made for the experiment corresponding to Figure 4.7.

We attempt to identify patterns of parameter values and feature vector arrangements that consistently result in superior performance. Several parameter comparisons are made in Figure 4.10. The top graph gives correct matching performance for several sigma (Gaussian smoothing) values and several region sizes for fixed segmentation. The bottom graph gives matching performance for several curvature factors. We define any feature's "factor" as the degree to which that feature influences similarity judgments between images. In the distance measure, the "factor" is multiplied by the feature's value, and thus provides the scaling coefficient between different features in the N-dimensional space. In one case, a distinction is made between inner and outer contours. In another case, it is not. The results suggest that the value of sigma is less important than the choice of region size, and that it is not advantageous to consider the distinction between inner and outer contours. Several more parameter comparisons are made in Figure 4.11. The top graph gives matching performance for several sigma (Gaussian smoothing) values and several region sizes for fixed segmentation. The composition of the feature vector differs from that used in Figure 4.10. The bottom graph gives

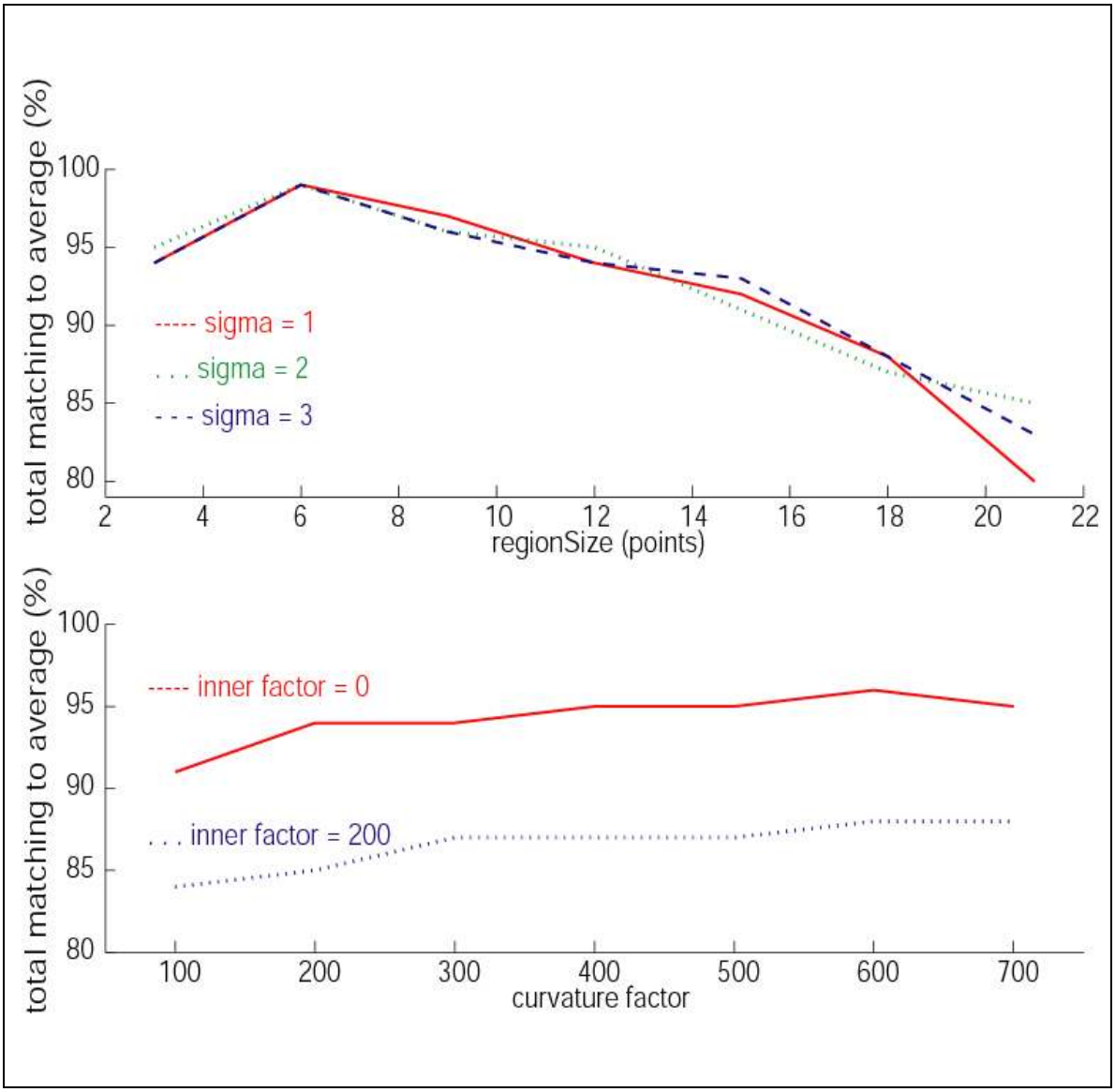

Figure 4.10 – Parameter comparisons. Different line types and abscissa values signify different parameter values. Performance (as the percentage of correct classifications) is on the ordinate. In the top graph, the region size (in points) is varied with different sigma values, using the angle, curvature (with a factor of 600), direction, and distance features and an inner contour factor of 0. In the bottom graph, the curvature factor is varied with different inner contour factor values, using the angle, curvature, direction, and distance features and a region size of 9 and a sigma value of 2.

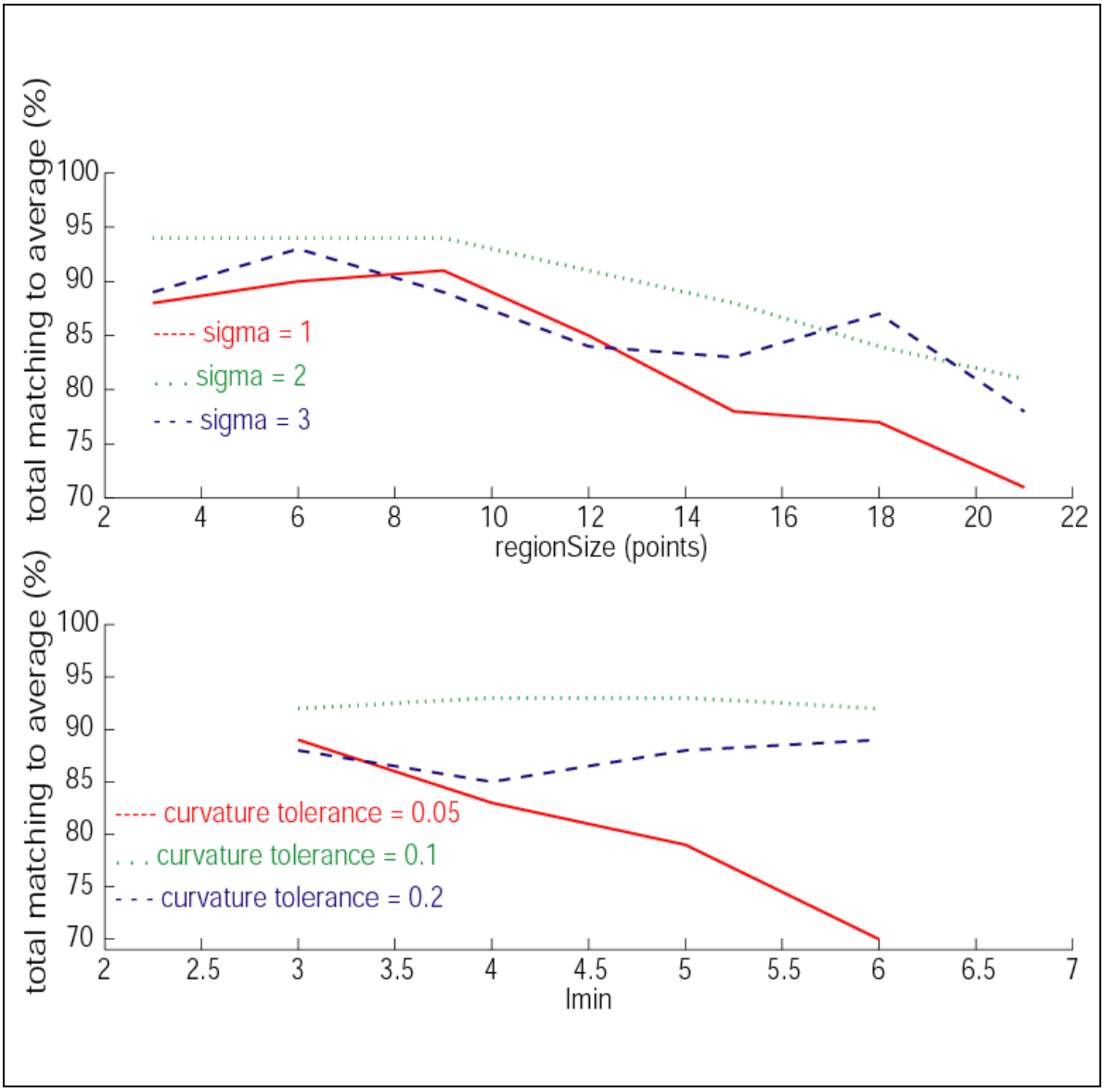

Figure 4.11 – Parameter comparisons. Different line types and abscissa values signify different parameter values. Performance (as the percentage of correct classifications) is on the ordinate. In the top graph, the region size (in points) is varied with different sigma values, using the angle, curvature, clockwise curvature, and counterclockwise curvature features. In the bottom graph, the minimum segment length is varied with the curvature tolerance, using the angle, curvature (unsquashed), clockwise curvature, and counterclockwise curvature features and a sigma value of 2.

performance for several curvature tolerances and several minimum segment lengths, both important in segmentation using the curvature voting and grouping methodology. It again appears that region size is more important than the value of sigma that is chosen. In addition, curvature tolerance seems to have an optimal operating range.

It can be seen in Figure 4.10 that the method of iso-curvature segmentation using the lowest standard deviation arrangement with a standard region size of six points results in the best performance, with performance falling off as the region size is increased. The constant curvature criterion method of iso-curvature segmentation, shown in Figure 4.11, yields somewhat inferior performance results. However, using a curvature tolerance of 0.1, performance is relatively high and stable for a wide range of minimum segment lengths. Performance falls off rapidly with increasing minimum segment length using a curvature tolerance of 0.05.

The experiment shown in Figure 4.12, with performance for the angle and direction features compared with performance for the angle and curvature features, for different region sizes, represents a more extensive test on the region size versus accuracy trade-off. As expected, for smaller region sizes, performance for both sets of features is comparable. A divergence in performance, with angle and curvature the superior feature vector, is seen for larger region sizes. It appears that the method of iso-curvature segmentation using the lowest standard deviation arrangement with a standard region size is optimal for a particular size. This region size is image-dependent. The larger the size, the less curvature variation captured, but the more "representational" of the contour's

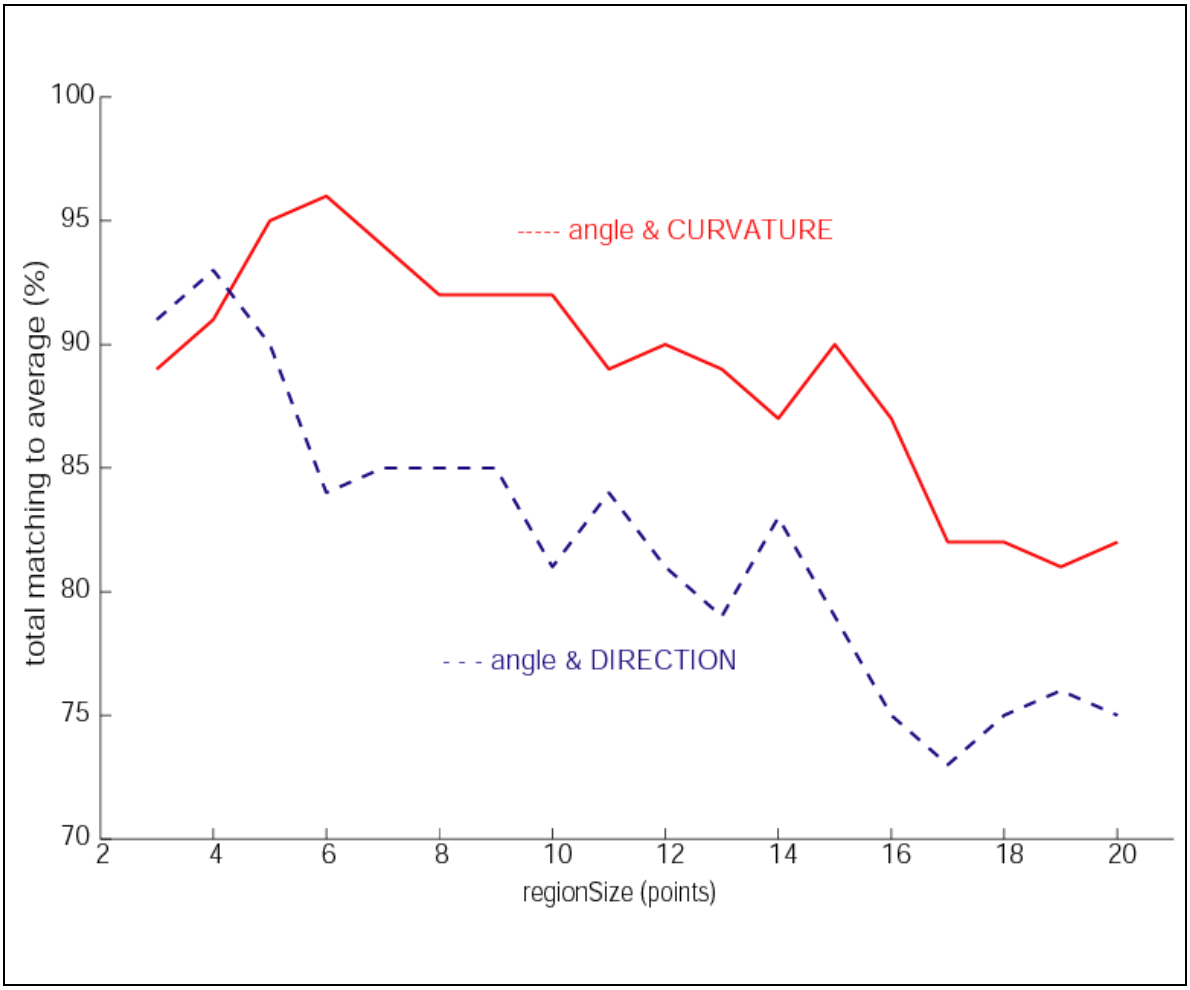

Figure 4.12 – Parameter comparisons. Different abscissa values signify different region sizes. Performance (as the percentage of correct classifications) is on the ordinate. For the solid line, only the angle and curvature are used as features. For the dashed line, only the angle and direction are used as features.

essential characteristics. An image with rapidly changing contour curvature values at a particular scale would be best characterized by a small region size (and therefore more segments) to acquire sufficient detail. Here, direction and angle (or distance and angle, as in the shape context algorithm) might prove as useful as curvature and angle. (Consider the discussion of the MPEG-7 Shape Silhouette database below.) A larger region size (and fewer segments) would be sufficient for an image with broadly, gently changing contour curvature values at the same scale. Here, curvature and angle are superior. As always, curvature of the pre-digitized object cannot be calculated exactly and is particularly sensitive to quantization error.

A summary of feature vector arrangements for the V4-like units that are considered in various experiments, along with the percentage of digit images that are correctly matched, is given in Figure 4.13.

A system must perform robustly in the presence of noise to be an appropriate model for visual recognition. To study the robustness of features to signal degradation, noise is added to various feature components. The experiment shown in Figure 4.14, for instance, employs feature vectors (one for each iso-curvature segment) composed of the mean angle of the region, the mean curvature of the region, the mean direction of curvature of the region, and the mean distance from the center of mass of the region. The correct matching performance for several values of Gaussian noise (as a multiplier of the segment's standard deviation) added individually to each of the features is given, with error bars. It appears that curvature is the feature that is the most sensitive to noise.

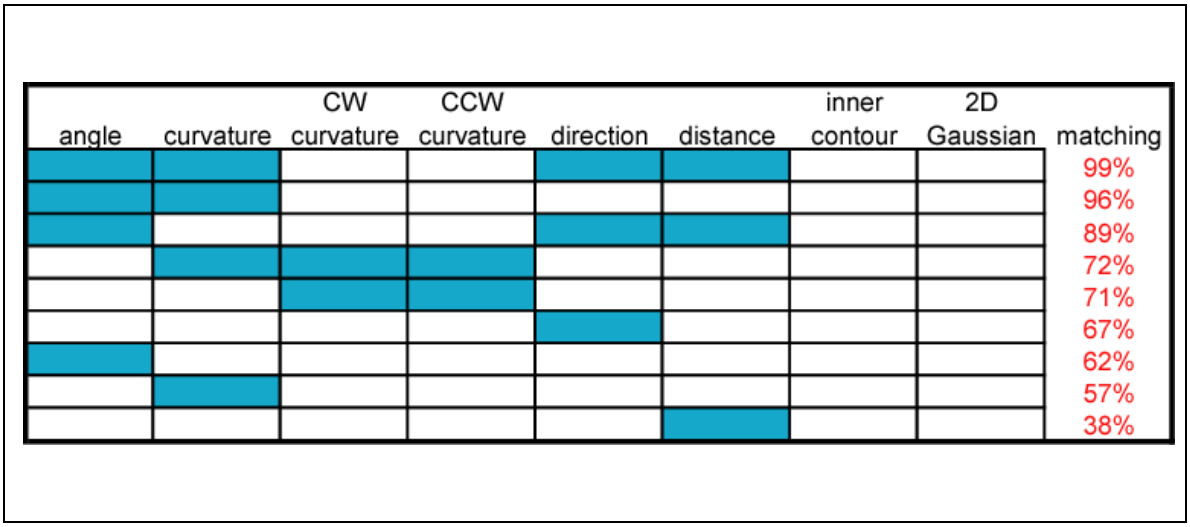

Figure 4.13 – Feature inclusion summary. Many different arrangements of features for the V4-like units were considered. For each particular combination, the performance of the system (as the percentage of correct classifications) is specified in the "matching" column. The inclusion of a particular feature within the combination is indicated with a dark box.

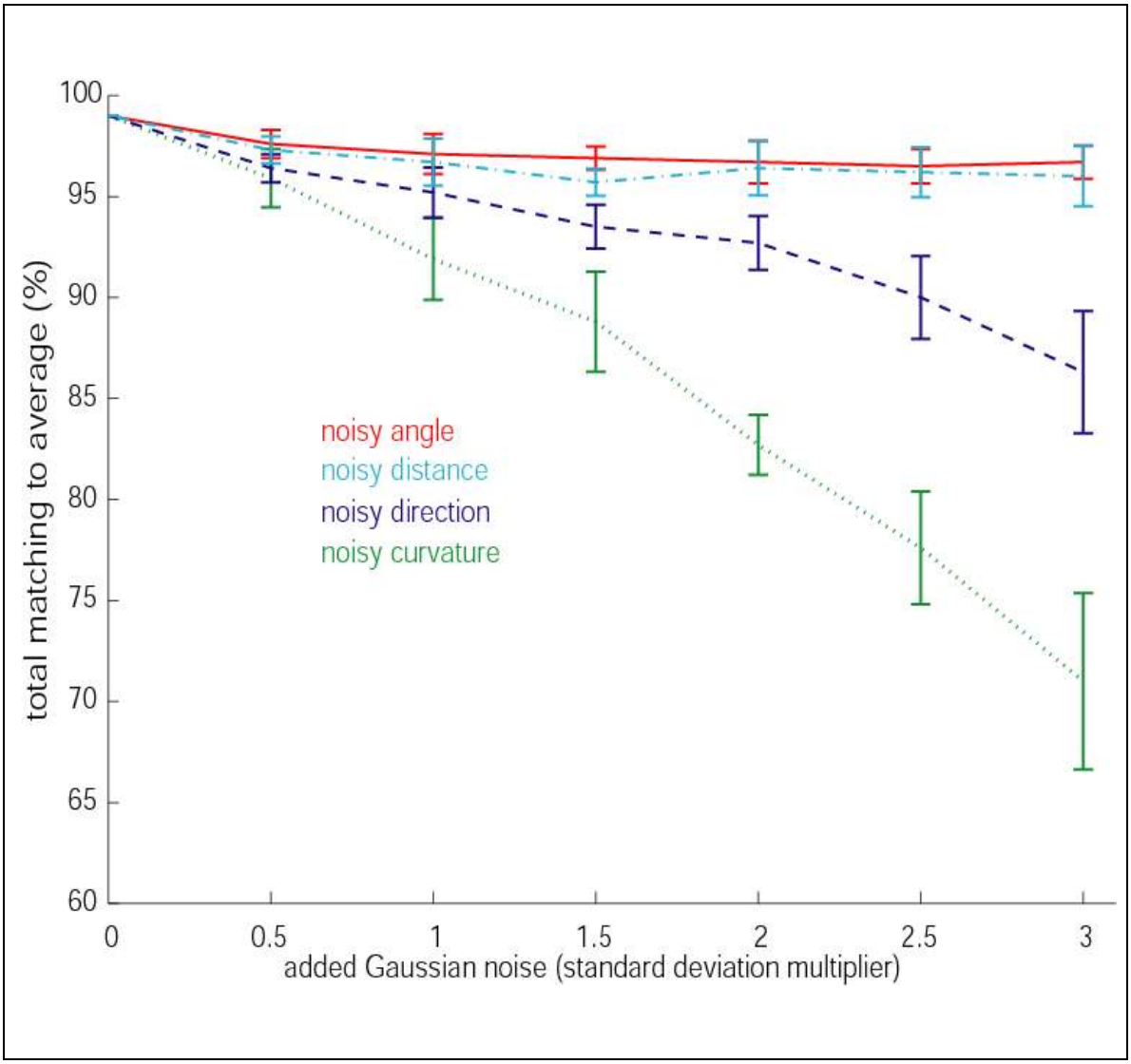

Figure 4.14 – Noisy features. Gaussian noise (as the abscissa multiple of standard deviation) is added individually to each of the features, represented by different line types. Performance (as the percentage of correct classifications) is on the ordinate. Noisy performance using the angle, curvature (with a factor of 600), direction, and distance features and an inner contour factor of 0, a region size of 6, and a sigma value of 2 is shown.

In general, one would expect the features most predominantly used by the visual system for recognition to be the most robust in the face of noise. Our finding that curvature is the most informative feature would then suggest it should be most insensitive to noise. It is important to note that our additional finding, that noise affects curvature most strongly, is not inconsistent with this point. This follows because noise is added to the measurements of this feature and is not inherently part of the feature itself. Thus, the result that degrading curvature information has the largest effect on recognition suggests that curvature is the most salient feature for shape recognition.

The experiment shown in Figure 4.15 employs feature vectors (one for each iso-curvature segment) composed of the mean angle of the region, the mean curvature of the region, the mean curvature of the clockwise-adjacent region, and the mean curvature of the counterclockwise-adjacent region. The correct matching performance for several values of Gaussian noise (as a multiplier of the segment's standard deviation) added individually to each of the features is again given, again with error bars. It again appears that curvature (including that of the clockwise- and counterclockwise-adjacent regions) is the feature that is the most sensitive to noise.

As a more extensive evaluation, we employ the entire 10,000 digit images from the MNIST "Test Set" database of handwritten digits as stimuli. This image set contains approximately 1,000 different example images of each digit. Using an average distance calculation for categorization, various combinations of parameter values and feature

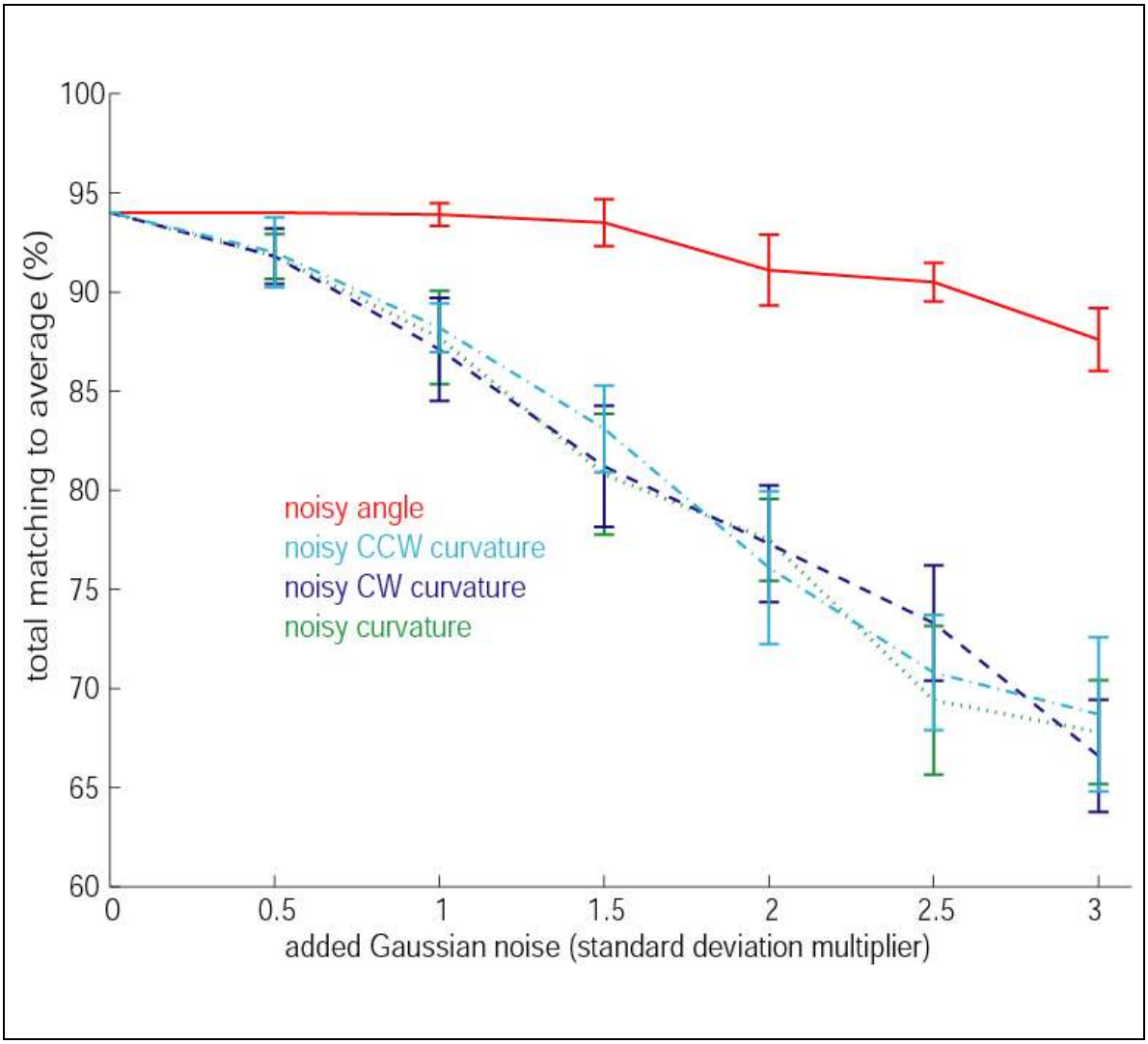

Figure 4.15 – Noisy features. Gaussian noise (as the abscissa multiple of standard deviation) is added individually to each of the features, represented by different line types. Performance (as the percentage of correct classifications) is on the ordinate. Noisy performance using the angle, curvature, clockwise curvature, and counterclockwise curvature features and a region size of 6 and a sigma value of 2 is shown.

vector arrangements are again tried in different experiments. As shown in Table 4.2, an experiment employing feature vectors (one for each iso-curvature segment) composed of the mean angle of the region, the mean curvature of the region, the mean direction of curvature of the region, and the mean distance from the center of mass of the region (similar to the experiment represented by Figure 4.7) results in a 97.03% correct classification performance. Figure 4.16 is the matching matrix, representing the number of images that are classified correctly. This result is only slightly inferior to the best results achieved by other researchers within their full computer vision applications (99.37% [Belongie *et al*., 2002], 99.3% [LeCun *et al.*, 1998]).

To experiment with other, non-digit objects, we choose images from the MPEG-7 Shape Silhouette database (Jeannin and Bober, 1999) as stimuli. One example image from each of several categories is shown in Figure 4.17. In an experiment employing 20 of these categories and feature vectors similar to those used in the experiment represented by Figure 4.7 (the mean angle of the region, the mean curvature of the region, the mean direction of curvature of the region, and the mean distance from the center of mass of the region), a 93% correct classification performance is achieved. (Specifically – apple: 19 of 20 matched; bone: 20 of 20 matched; bottle: 20 of 20 matched; cellular phone: 18 of 20 matched; cup: 20 of 20 matched; elephant: 17 of 20 matched; face: 20 of 20 matched; flatfish: 18 of 20 matched; fork: 18 of 20 matched; fountain: 20 of 20 matched; heart: 20 of 20 matched; key: 19 of 20 matched; lizard: 16 of 20 matched; pencil: 18 of 20 matched; personal car: 17 of 20 matched; ray: 17 of 20 matched; shoe: 20 of 20 matched; teddy: 20 of 20 matched; tree: 18 of 20 matched; watch: 17 of 20 matched.)

digit	actual matches	possible matches	percent correct
"0"	977	980	99.6939 %
"1"	1,124	1,135	99.0308 %
"2"	1,014	1,032	98.2558 %
"3"	978	1,010	96.8317 %
"4"	930	982	94.7047 %
"5"	852	892	95.5157 %
"6"	944	958	98.5386 %
"7"	989	1,028	96.2062 %
"8"	928	974	95.2772 %
"9"	967	1,009	95.8375 %
total	**9,703**	**10,000**	**97.03 %**

Table 4.2 – Summary of results for the entire 10,000 digit images from the MNIST "Test Set" database. The angle, curvature, direction, and distance features are employed, with a region size of 6 and a sigma value of 2 For each digit image, an average distance calculation is used for categorization.

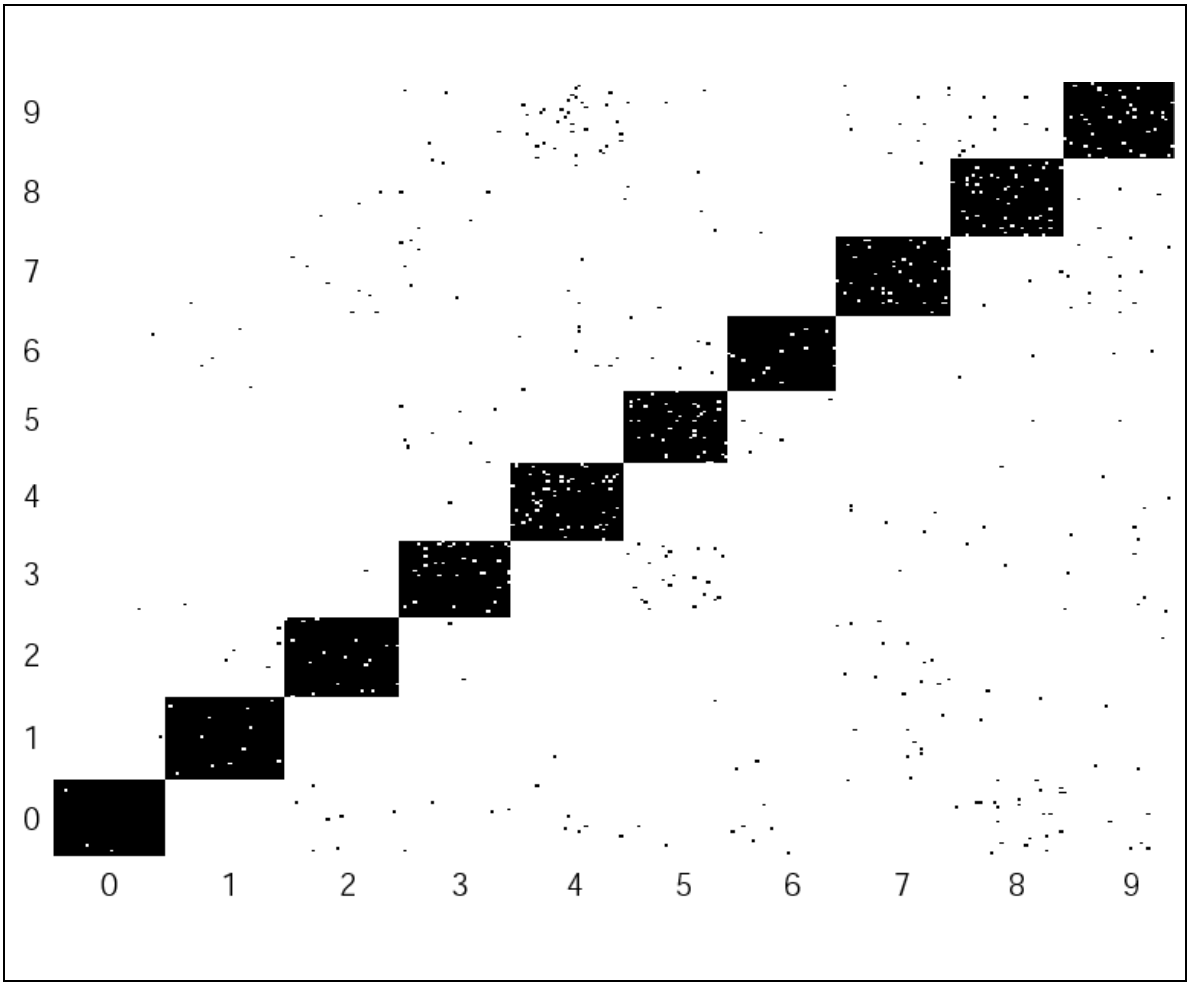

Figure 4.16 – Matching matrix for the entire 10,000 digit images from the MNIST "Test Set" database. For each digit image, an average distance calculation is used for categorization. Correct matches appear as dots within the large rectangles on the diagonal. Incorrect matches appear outside the large rectangles. The area of each rectangle corresponds to the number of sample images for the corresponding digit in the Test Set. Digit values are on the axes. Using the angle, curvature, direction, and distance features, with a region size of 6 and a sigma value of 2, the total matching to average is 97.03%.

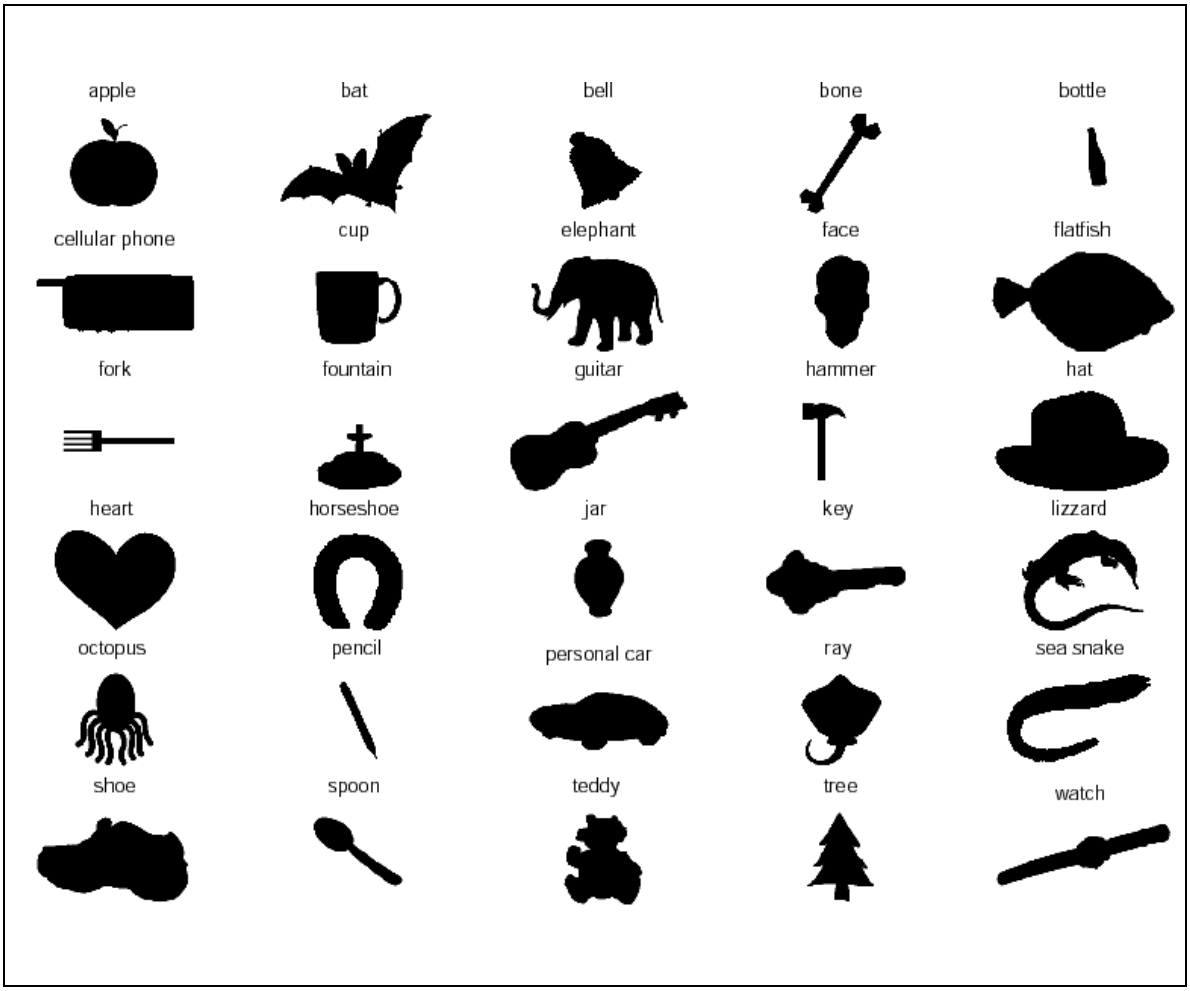

Figure 4.17 – MPEG-7 Shape Silhouette database. One example image from each of several categories is shown.

Although the region size used in the shape silhouette studies (20) is larger than that used in the digit studies (6), it should be noted that the scale is different – the shape silhouette images are much larger. In addition, unlike the MNIST data, the MPEG-7 images are not uniformly scaled or oriented, making categorization more difficult. Our results do not achieve the performance accuracy achieved in computer vision models, and are also slightly inferior to estimates of human visual performance (LeCun *et al.*, 1995; Simard *et al.*, 1993), which is not surprising as contour shape provides only one source of information used in object recognition. Nonetheless, the high level of performance achieved based solely on contour shape supports the hypothesis that shape representation by populations of V4 neurons plays a key role in recognition.

4.4 Discussion

Since Marr (1982), it has been taken that intermediate-level visual cortical areas compute a representation of object shape that is used by higher areas for recognition. The discoveries of Pasupathy and Connor (2001), Hegdé and Van Essen (2003), and others reveal a previously unsuspected degree of sophisticated shape processing in extrastriate visual cortex. To a large degree, the properties of V4 units represent a biological solution to shape description: if you know "what" the contour looks like at each position along its extent, and "where" every kink and bulge is located, the shape is described fairly well.

The importance of contour curvature in both human and computer vision has long been apparent. In his seminal work relating information theory to visual perception, Attneave (1954) found that information is concentrated along contours – specifically at those points on a contour where its direction changes most rapidly, such as curvature peaks. This finding has been validated for natural objects (Norman *et al.*, 2001), with contour-based object identification and segmentation itself being later corroborated empirically (De Winter and Wagemans, 2004). Hoffman and Richards (Hoffman and Richards, 1984; Richards and Hoffman, 1985), arguing that the visual system decomposes a shape into a hierarchy of parts, focus on part boundaries rather than part shapes and segment a bounding contour into parts with curvature-extrema-defined endpoints. While Leyton's (1989) rules governing the perception of shape are based upon the symmetry and curvature structure of the shapes, he proposes a more cognitive view of parts, considering them temporal or causal processes.

Some researchers have extended this work to incorporate curvature sign, finding the human visual system to be substantially more sensitive to changes in concave regions of a bounding contour than to changes in convex regions (Cohen *et al.*, 2005; Feldman and Singh, 2005). Others have explored the consequences of this asymmetry in perceptual figure-ground assignment (Hoffman and Singh, 1997), visual search (Xu and Singh, 2002) and speed and accuracy of visual comparisons (Barenholtz and Feldman, 2003). In contrast, Bertamini and Lawson (2006) fail to find evidence that concave targets are inherently more salient, instead suggesting that concavities are processed preferentially as the result of an early computation of part structure. Some earlier work has tried to

characterize psychophysically the cortical mechanisms that underlie the discrimination of very small curvatures in a stimulus (Kramer and Fahle, 1996) and has considered how the magnitude and direction of curvature affect the strength of long-range interactions between neurons (Pettet, 1999). Researchers have studied the problem of quantifying human intuition regarding shape similarity and local deformations (Basri *et al.*, 1998) and have defined similarity metrics based on intrinsic properties of the curve, such as length and curvature (Sebastian *et al.*, 2003).

Computer algorithms have been developed to extract and recognize the shape of silhouettes from the curvature extrema of their bounding contours (Chien and Aggarwal, 1989). A computer vision model of contour fragment grouping from contour junction information in a 2-dimensional image has been developed (Bergevin and Bubel, 2003). In addition, in a Bayesian model of contour grouping, the bounding contour of an object in an image is used in conjunction with some prior knowledge about the object (Elder *et al.*, 2003). Similar probabilistic models of perceptual grouping for contour integration have been developed for human visual perception (Feldman, 2001), with the importance of interpolation in segmentation and grouping processes that act on fragmentary contours being demonstrated (Kellman, 2003).

The need for a robust and stable organization and representation of object parts in the presence of rigid transformations, occlusions and changes in viewing geometry has led some recognition researchers to argue for a role for a *skeleton-based descriptor* along with the bounding contour (Siddiqi and Kimia, 1995). Others have demonstrated the

psychophysical relevance of these descriptors (Siddiqi *et al.*, 1996) or have stated that representing objects in terms of their absolute edge locations, as in a *contour-based descriptor*, is seriously flawed, primarily because of the difficulty of finding edges in conditions of low signal-to-noise ratio or occlusion or because of the independence of the scale of the object in the representation (Burbeck and Pizer, 1995). It has been proposed that skeletal-based descriptors (such a shock-graphs or medial axes) capture the spatial arrangement of parts that leads to distinct shapes, thus overcoming a potentially major drawback of contour-based descriptors (Sebastian and Kimia, 2005). A recognition framework based upon matching skeletons of 2-dimensional shape outlines has been developed by considering these descriptors to be curve-based representations with paired contours and additional (often an order of magnitude greater) computational requirements (Sebastian *et al.*, 2001; Sebastian *et al.*, 2003).

In a review of shape representation and description techniques, Zhang and Lu (2004) point out that contour-based methods fail to capture global shape features and are sensitive to noise, as boundary variations can cause significant local effects. However, skeletal-based descriptors also require shape contour information, and are therefore somewhat sensitive to effects such as quantization error. Considering these arguments, we attribute our success with contour-based descriptors to methodology and choice of application. Sebastian and Kimia (2005) find that the increased computational complexity of skeleton matching is justified by increased robustness in the presence of articulation or rearrangement of parts. However, in applications where variations are smaller (such as handwritten character recognition), contour matching is superior because

of its relative simplicity and roughly equivalent recognition rate. More importantly, we include such quantities as polar angle of the region, curvature of adjacent regions, direction of curvature, distance from the center of mass, etc. (all with approximate neurobiological correlates), in our feature vectors. This, coupled with our contour comparisons using the earth mover's distance, provides us with sensitivity to local curve-based features, as well as some of the advantages of a representation of the spatial relationships among the iso-curvature segments of each image, thus allowing us to perform well when shape defects are present.

Our results do not address how cells in V4 achieve their selectivity to contour configuration and location. V4 is retinotopically organized, and sensitivity to a configuration at a particular location on the contour (e.g., 3 o'clock) suggests an object-centered representation. Several models of cortical curvature detection have been put forward aimed primarily at curvature sensitivity in V1 arising from end-stopped receptive field structure (Wiesel and Gilbert, 1989; Dobbins *et al*., 1989). However, research suggests that cells in V2, V4 and related areas carry out a more general, broader computation than just estimating curvature. Riesenhuber and Poggio (1999) have proposed that V4 selectivities can emerge from a series of linear and nonlinear integrations of early visual responses. Their model (Riesenhuber and Poggio, 2003) accounts for a range of reported receptive field characteristics in V4, and can account for how a V4 neuron becomes selective for a particular location on the contour. Their model is bottom-up, without the requirement of an explicit top-down (attentional or otherwise) signal, an explicit segmentation stage, or complex synchronization mechanisms for

binding elements of the contour together. It should be noted that V4 cells respond to a variety of stimulus features – including color (McKeefry and Zeki, 1997), orientation (Desimone and Schein, 1987; Hinkle and Connor, 2002), disparity (Hinkle and Connor, 2001), and complex spatial patterns (Gallant *et al.*, 1996). Extrastriate cells show selectivities to several aspects of form (Gallant *et al.*, 1996) and border ownership (Zhou *et al*., 2000). In addition, V4 cell receptive field properties are strongly modulated by attention (Reynolds and Desimone, 2003; Bichot *et al*., 2005; McAdams and Maunsell, 2000; Motter, 1994; Connor *et al.*, 1997), and the presence of a small feature within the large receptive field can drive cellular response. However, our findings suggest that curvature, the feature that is the most sensitive to noise as seen in Figures 4.14 and 4.15, would seem to be the most important for shape recognition.

A second issue that remains to be answered, concerns whether and how a global description of object shape emerges. Descriptions of shape based on local characteristics can miss, or fail to emphasize, important global or "structural" differences between objects. For example, the digit "3" and the digit "5" are very similar locally: they have similar bottoms and a similar top, but differ in the direction of the upper convexity. Focusing on the "upper convexity" as opposed to a smaller segment indicates a parts-based description. Higher visual areas may learn to be selective to differences, at a variety of hierarchical levels, which distinguish categories of objects. Top-down effects might then focus the response of V4-like units onto those regions or features whose local configurations are critical for the class distinction (Hochstein and Ahissar, 2002; Sigala and Logothetis, 2002). Pasupathy and Connor (2001) point out that complex shape

representation in area V4 is parts-based (since contour segments are defined by conformation and position) as well as distributed (since individual cells encode smaller parts of larger objects). It is thought that a parts-based coding system, using either a finite number of primitives or a continuous part representation with graded tuning, has the combinatorial power and representational capacity to encode a virtually infinite variety of objects (Pasupathy and Connor, 2002; Biederman, 1987; Tsunoda *et al.*, 2001; Rolls *et al*., 1997). A question raised by Pasupathy and Connor's results is how many such units would be necessary to represent all possible shapes. If each 5^{o} receptive field region requires units for roughly 12 orientations, 5 distances, 5 curvatures plus 5 other shapes (sharp corners, etc.) and 2 directions of curvature, that suggests ~1200 different V4-like units per hypercolumn – well within the number of cells per layer in extrastriate cortex responding to a 5^{o} receptive field region (Van Essen, 2003; Bullier, 2001; Motter, 2003).

We have shown that a population of units modeled after Pasupathy and Connor's description of V4 cells can classify images from the MNIST database at high levels of accuracy. As the matching matrices of Figures 4.7, 4.9, and 4.16 demonstrate, features such as curvature that have neurobiological correlates in or near area V4 are capable of functioning as robust shape descriptors in the early stages of shape recognition. The performance of our system depends upon extracting reasonable estimates of local curvature. In digitized images, curvature estimation is a challenging problem. We have used the approach of Wuescher and Boyer (Wuescher and Boyer, 1991; Worring and Smeulders, 1993) which far surpasses results we were able to achieve using local

curvature filters or direct computation of changes in tangent angle. The visual system exhibits hyperacuity to curvature (Watt and Andrews, 1982; Wilson, 1985) and is extremely sensitive to direction of curvature (Cohen *et al.*, 2005) – particularly in regards to vernier acuity and other capabilities. However, our results (and those of Belongie and colleagues (Belongie *et al*., 2002)) indicate that high precision in the curvature estimate is not essential – binning the curvature values into a few (3-5) bins (very low, low, mid, high, and very high) yields only a small change in overall recognition accuracy. Most critical is the choice of the relevant scale at which to measure curvature and the necessity of avoiding large discontinuities in fine-scale curvature estimates that can arise from digitization.

Successful object recognition requires a number of processes that have not been dealt with in this study. We begin with a 1-pixel thick segmented contour (i.e., no background pixels) and provide the (x, y) coordinates of each pixel. Thus, our results cannot be compared to stand-alone recognition systems, such as that of LeCun and colleagues (LeCun *et al.*, 1998) which must handle earlier stages of processing as well. By using the MNIST database of digits, we have largely bypassed issues of scale, rotation and translation invariance. The degree of variability in scale, rotation and translation of the images is modest, and is within the range of sensitivity to scale and rotation observed in V4 cells (Logothetis *et al*., 1995). Use of V4-like units implies that shape responses are sensitive to object orientation (the NW corner is not the NW corner anymore after rotation). However, most cortical expert recognition systems (e.g., face recognition (Viola and Jones, 2001b; Schneiderman and Kanade, 2004), biological motion

recognition (Song *et al*., 2003; Casile and Giese, 2005; Giese and Poggio, 2003)), and reading, are sensitive to object orientation as well.

Recognition of objects embedded in more natural image scenes introduces many additional complexities, including the need to identify which V4-like units are responding to the same contour. However, initial fast-pass recognition might be plausible based merely on the distribution of V4-like cells activated in a region. Suppose each unit is sensitive to local and neighboring curvatures on a segment of contour, as well as to the location of this configuration within the visual field region (e.g., a 5^{o} window). The distribution of responses of various such V4-like units within the spatial window might provide an initial match to a shape class. Top-down effects could then aid in segmentation and refinement of the contour representation.

Results from Sigala and colleagues (Sigala *et al*., 2002) suggest that humans categorize by comparing objects to well-known members of alternative categories, either directly or based on class boundaries (as in SVM). These investigators also found that humans and monkeys learn which features are most diagnostic for distinguishing particular categories. In the results presented here, the responses of V4-like units depended with equal weight on each feature sensitivity: e.g., curvature, magnitude, direction of curvature, location on the contour. In other words, each feature had equal weight (after initially multiplying the feature by its "factor") in the shape matching calculation using the Earth Mover's Distance (EMD) comparison (Rubner *et al*., 2000; Rubner *et al.*, 2001), our most consistent distance measurement. However, in a two alternative forced choice

recognition task – e.g., is it a "0" or a "6" – certain features are most informative for the decision. It may be known *a priori* that a "6" usually has a sharp convexity at 1 o'clock (the stem) and a "0" usually does not. Examination of the response of V4-like units to this 1 o'clock segment show that, depending upon their feature sensitivities, some units respond roughly equally to both "0's" and "6's" (e.g., average curvature and average distance from the center of mass, as seen in Figure 4.18). However, some units respond very differently (e.g., averaged direction of curvature, as seen in Figure 4.19), yielding separability. If we weight the units that respond differently more strongly than those that respond similarly, we can recompute the performance of the model for discriminating all "0's" and all "6's". Employing this modification, the model performs at 100% – in comparison to the 99% performance in the untrained model shown in Figure 4.7. It should be noted that a naive and inappropriate over-weighting of the 1 o'clock iso-curvature segment across the entire population diminishes the system's overall performance slightly (3%). Thus, initial responses of V4 units might produce activation of both "0" and "6" detectors in higher visual areas. Faced with this decision, top-down inputs could alter the gain or sensitivity of particular V4 cells, thereby effectively changing the weighting of curvature features so as to distinguish the "0" vs. "6". Example digits, with 1 o'clock segments highlighted, are shown in Figure 4.20. The "6" digit image in the top left corner had previously been misclassified (see Figure 4.7 and Figure 4.8).

These results suggest that top-down inputs can improve classification accuracy by re-weighting the contributions of intermediate-level units. This process can aid in

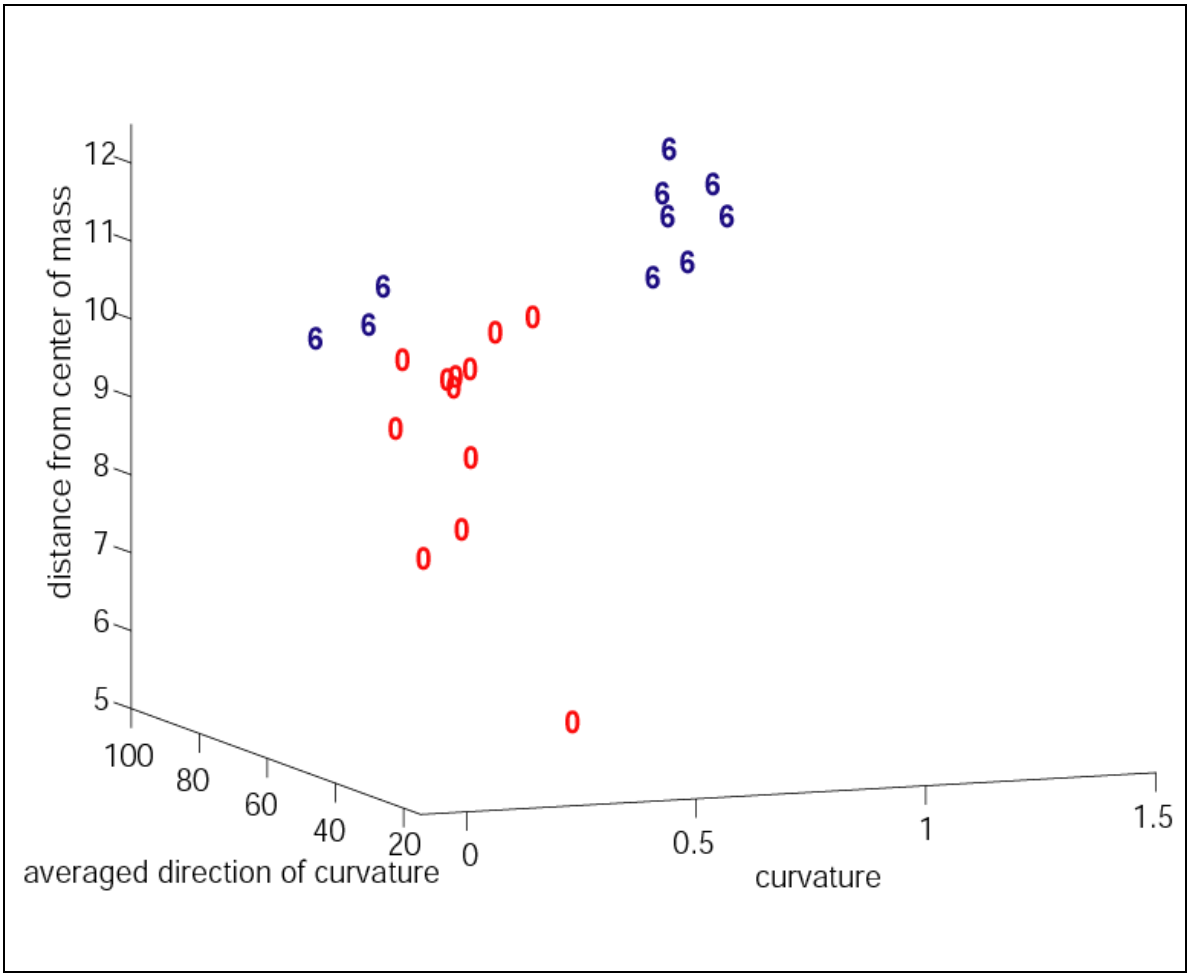

Figure 4.18 – Dimensional inseparability of segment features of "0" and "6" images. The 1 o'clock iso-curvature segments of each of the "0" and each of the "6" digits is shown in the curvature × direction × distance space. The "0"'s and "6"'s are not separable in the curvature or distance from center of mass dimensions.

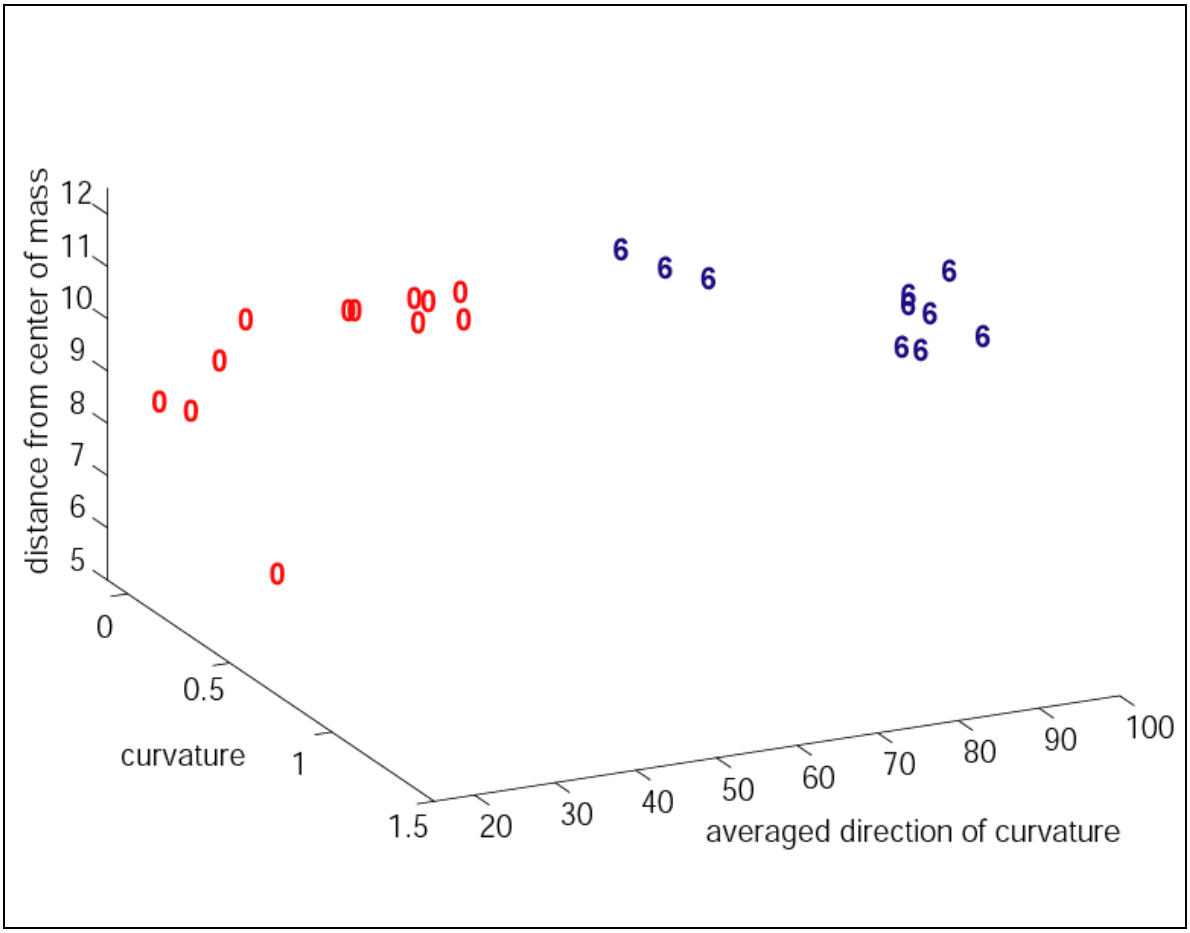

Figure 4.19 – Dimensional separability of segment features of "0" and "6" images. The 1 o'clock iso-curvature segments of each of the "0" and each of the "6" digits is shown in the curvature × direction × distance space. The "0"'s and "6"'s are separable in the averaged direction of curvature dimension.

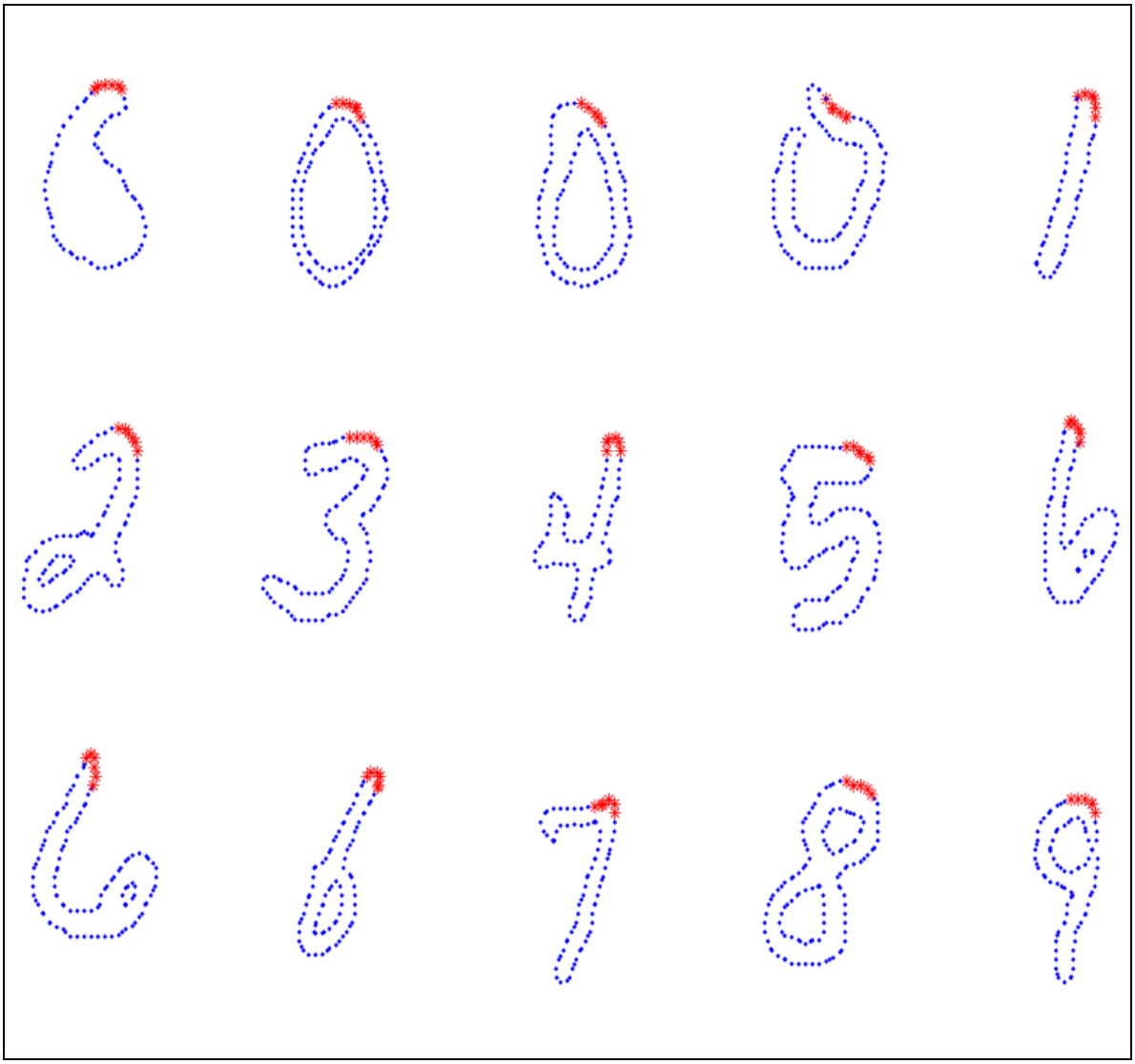

Figure 4.20 – 1 o'clock segments. Some sample 1 o'clock iso-curvature segments across the population are shown as open circles.

identifying regions of interest, and can make adjustments to the iso-curvature segmentation as well.

4.5 Conclusion

Our results suggest that curvature- and position-sensitive units, as described by Pasupathy and Connor in area V4, can function as robust shape descriptors. We have demonstrated shape categorizations based on curvature representations and established a connection between state-of-the-art recognition systems and known cortical mechanisms.

Chapter 5

Shape Representation and Object Recognition in the Inferotemporal Cortex (IT)

5.1 Introduction

The representation of contour shape is an essential component of object recognition, but the cortical mechanisms underlying shape analysis and object recognition are incompletely understood, leaving it a fundamental open question in neuroscience. Work in monkey extrastriate cortex shows that visual cortical neurons are sensitive to the fine structure of contour shape. In visual area 4 (V4), an intermediate stage in the ventral (shape recognition) pathway in the occipital and temporal lobes extending hierarchically from primary visual cortex (V1) to inferotemporal cortex (IT), Connor and colleagues have described cells that are selective for a particular local shape configuration at a

particular location on a contour within a larger shape. The response profiles of these cells are further modulated by the local contour configurations at neighboring locations on the contour (Pasupathy and Connor, 2001). The Connor group has also demonstrated that a population of these V4 units can provide a detailed description of the contour shape (Pasupathy and Connor, 2002). These discoveries reveal a previously unsuspected degree of sophisticated shape processing in extrastriate visual cortex.

Other research has focused on neural selectivity for complex 2-dimensional boundary shape (perhaps the kind that actually dominates responses to realistic objects (Kovács *et al.*, 2003)) in macaque inferotemporal cortex (TEO / PIT and posterior TE / CIT) (Brincat and Connor, 2006; Brincat and Connor, 2004; Freedman *et al.*, 2003; Baker *et al.*, 2002; Tsunoda *et al.*, 2001; Op de Beeck *et al.*, 2001; Booth and Rolls, 1998; Rolls *et al.*, 1997; Gallant *et al.*, 1996; Logothetis *et al.*, 1995; Kobatake and Tanaka, 1994; Fujita *et al.*, 1992; Young, 1992; Felleman and Van Essen, 1991; Gross *et al.*, 1972). It has been found that IT neurons integrate specific information, such as curvatures, orientations, and relative positions, about the shapes of multiple contour fragments (typically 2 – 4). Explicit signals that code structural relationships between parts are generated and could be useful for high-level object representation, supporting the idea of parts-based shape representation. Processing by these cells thus bears some similarity to the fragment-based approach to recognition of Ullman and colleagues (Ullman *et al.*, 2001) and the components-based approach of Biederman (1987), with fragments / components functioning as building blocks used to represent a large variety of objects belonging to a common class.

Some computer vision approaches to object recognition have begun to achieve impressive levels of accuracy and robustness, yet lack a clear connection to known cortical constructs. There is no accepted neurobiological theory as to how object recognition, and the underlying analysis of shape, is accomplished. Such an understanding would be useful theoretically as well as to develop or improve computer vision and Brain-Computer Interface (BCI) methodologies and applications. Two fundamental questions might be put forth: "How is contour shape represented in cortex and how can neural models and computer vision algorithms more closely approximate this?"

Here, we continue our investigation of how contour shape is represented in cortex. In doing so, we hope to narrow the divide between the theoretical computational neuroscience and neuroengineering and the biological neurophysiology. Our earlier results (Chapter 4) support the hypothesis that curvature- and position-sensitive V4 cells – evaluated against the standard MNIST database of handwritten digits and the MPEG-7 Shape Silhouette database (Jeannin and Bober, 1999) – function as robust shape descriptors in the early stages of object recognition. We demonstrated good shape categorizations based on a particular local contour conformation located at a specific position on the object's boundary (curvature representations) and established a connection between state-of-the-art recognition systems and known cortical mechanisms.

We extend these results to consider the recognition properties of a population of cells modeled after those found in IT, which nonlinearly integrates specific information about the 2-dimensional boundary shapes of multiple contour fragments (V4 cell inputs) with tuning functions on the shape × position domain (Brincat and Connor, 2004; Brincat and Connor, 2006). Using nonlinear least squares optimization and genetic algorithms to fit parameters, we create selective IT-like cell populations with similar response patterns. We are principally interested in the number of constituent Gaussian terms in the IT-like cells' total response equations, the linear and nonlinear parts of these equations, the amount of nonlinearity and how these aspects relate to the shapes of objects (as opposed to their orientations and scales). We evaluate the performance of our IT populations on a set of real images as a function of the V4-like cell inputs. The stimuli (2-dimensional closed contours representing object boundaries) evoke a pattern of activity across the population of IT cells. Shape recognition is evaluated by demonstrating that the patterns of activity across the units to members of a particular object class resemble each other to a higher degree than they resemble members of any other class. This measure corresponds to that reportedly used by humans and monkeys in object classification (Sigala *et al.*, 2002).

We examine cell response space in more detail using principal components analysis and a 2-dimensional and 3-dimensional non-classical non-metric multidimensional scaling analysis. We find the correlation coefficients of the observations (cell responses) and variables (images) and determine the sub-population of cells that are most effective at

identifying a particular category. We use a support vector machine, as well as a tree-based model, for classification based upon cell population response.

Our primary objective is to determine why the response properties of V4 and IT cells (i.e., their receptive fields), and in particular their sensitivities to curvatures and contour positions, are useful. We answer this question by constructing a robust model of V4 and IT cell properties and functionality and demonstrating its utility at shape recognition and categorization tasks. In general, we obtain very good results across a wide range of parameter values and implementation strategies, comparable to those obtained previously with the digit database.

5.2 Methodology

In our current work, we focus on natural imagery – in the form of 360 × 360 JPEG images – kindly supplied by Drs. Kanwisher and Grill-Spector (Grill-Spector and Kanwisher, 2005). These are significantly more complex – in both information content and image processing techniques required – than the MNIST digits used in Chapter 4. We select images from this dataset belonging to one of seven different categories (axes, cats, fish, guitars, handsaws, hats, and scissors), with ten randomly selected samples from each category.

As in our previous work, simple image processing techniques and hand-segmentation are used to decompose the images into closed-contour silhouettes. We represent the shapes with sets of points sampled from the shape contours. For each image, we extract contours using the numerical gradient and determine a set of boundary points with oriented tangents. The result is a parametric description (x(t), y(t), and tangent(t)) of each contour. For each point along the contour, we compute its angle (0° – 360°) relative to and distance (in pixels) from the image's center of attention (center of mass, centroid, center of image).

Following Wuescher and Boyer's (1991) methodology, we convolve x(t) and y(t) with the derivative of a Gaussian to both smooth (regularize) and differentiate the functions. This results in the discrete curvature, parameterized by the Gaussian space constant σ:

$$\kappa(t, \sigma) = ((x(t) * g'(t, \sigma)) (y(t) * g''(t, \sigma)) - (x(t) * g''(t, \sigma)) (y(t) * g'(t, \sigma))) / (((x(t) * g'(t, \sigma))^2 + (y(t) * g'(t, \sigma))^2)^{3/2} ,$$

where

$$g(t, \sigma) = (1 / \sigma\sqrt{2\pi}) \exp(-t^2 / 2\sigma^2) ,$$

$*$ is the convolution operator, and $'$ is the differentiation operator.

We define the direction of curvature to be orthogonal to the tangent and to point towards the interior of the closed contour (Sajda and Finkel, 1995). It is computed using the orientation and inverse tangent.

We define **iso-curvature segments** as contiguous portions of the bounding contour, composed of individual boundary points, with identical or nearly identical curvature values.

We segment each contour into iso-curvature regions using one of two methods. In the first method, we choose a standard size (a number of boundary points) for each region. Remaining points are evenly distributed. We choose a starting point on the contour (and therefore the starting point for each region) based upon the arrangement that yields the lowest average standard deviation of curvature for each region. Ideally, each iso-curvature segment has a zero standard deviation of curvature. The second method, following Wuescher and Boyer's (1991) curvature voting technique, considers segments of constant curvature (within a curvature tolerance t_c) and segments of rapidly changing curvature. Curvature is quantized into bins of a specified width (typically $1/2t_c$ bins for every span of 0.10 in curvature), with peaks of the resulting histogram representing the curvature values most likely to fit the longest segments. Continuing with their methodology, the longest contiguous, constant curvature segments in the most prevalent curvature ranges are repeatedly extracted. The leftover portions of the contour are either absorbed by adjoining segments or become segments themselves. The pre-determined minimum segment length (l_{min}) reflects the amount of segment detail. In the experiments

described below, unless otherwise noted, a standard size (18 boundary points) for each iso-curvature region is chosen for segmentation. In addition, all contours are smoothed using a Gaussian function with a standard deviation of 2 pixels (a sigma value of 2).

Once an iso-curvature segment is defined, all points on the segment can be thought of as having the same curvature value. An alternative, or refining, method of finding this value, after iso-curvature segment assignment, is by using the osculating circle technique of differential geometry (Gray, 1997). Here, the osculating circle – the best circle that approximates the iso-curvature segment, with the same tangent and curvature, at its midpoint – is found. This circle's radius (r) is then related to the segment's curvature by:

$$r = 1 \; / \; | \kappa(t) | \; .$$

For each iso-curvature segment of each image, we create feature vectors composed of mean polar angle of contour region (e.g., 3 o'clock is assigned to be 0 radians, 12 o'clock is assigned to be π / 2 radians, etc.), mean curvature of the region, mean direction of curvature of the region and mean distance from the center of mass of the region. These features all have approximate neurobiological correlates in area V4 and other extrastriate areas (Desimone and Schein, 1987; Kobatake and Tanaka, 1994; Gallant *et al.*, 1996; Pasupathy and Connor, 1999; Wilkinson *et al.*, 2000; Zhou *et al.*, 2000; Pasupathy and Connor, 2001; Pasupathy and Connor, 2002).

Recognition, both identification and classification, requires some measure of matching – that is, some way of describing two objects / images as being more or less similar to each other. We have used various methodologies for comparing groups of segments, but the results described here are obtained using the Earth Mover's Distance (EMD) comparison, as elaborated by Rubner and colleagues (Rubner *et al.*, 2000; Rubner *et al.*, 2001). EMD considers two distributions, represented by signatures (sets of weighted features), and is defined as the minimal work or cost to transform one signature into the other (i.e., filling the "collection of holes" of one distribution with the "properly spread mass of earth" of another distribution).

We compute earth mover's distances – normalized average-to-average, total matching to average, closest matching – as well as a percentage of total distance comparison between images and between image categories. We also incorporate some degree of affine transformation invariance in our models.

Some more extensive methodological descriptions can be found in our previous work (Chapter 4).

We create populations of cells (from six to several hundred), modeled after those found in IT, which integrate specific information (curvatures, orientations, relative positions) about the 2-dimensional boundary shapes of multiple contour fragments (V4 cell inputs) and which have tuning functions on the shape × position domain (Brincat and Connor,

2004; Brincat and Connor, 2006). We fit each cell's response pattern with a nonlinear Gaussian-based tuning function.

Each IT-like cell has up to six (but typically three) **Gaussian constituents** (A, B, C, …) in its total response equation, each receiving inputs directly from the V4-like cells (responding to an image's iso-curvature segments) and together used to compute the nonlinear response of the IT cell. The Gaussian constituents are essentially variables in the total response equations of the IT-like cells. Each term of these equations may include one or more (multiplied together) Gaussian constituents and may also be multiplied by a coefficient. Positive coefficients imply an excitatory contribution to the total response; negative coefficients imply an inhibitory contribution to the total response. Each Gaussian function is n-dimensional (where "n" is the number of features considered). Any of the image's iso-curvature segments can contribute to the Gaussian constituent function's response, with response magnitude proportional to the distance between the "n" features of the iso-curvature segment and the center of the n-dimensional Gaussian constituent. For a maximal response, the image's features (in the case of four features: the angle, curvature, direction of curvature and distance of each of the iso-curvature segments) would have to be perfectly aligned with those of the Gaussian constituents. Otherwise, a sub-optimal response would result, determined by Gaussian falloffs from the means at rates proportional to the specified standard deviations. The mean and standard deviation of each Gaussian constituent's 4-feature vector (angle, curvature, direction of curvature, and distance) are determined in a variety of ways, including averaging iso-curvature segments from a number of actual images, employing

iso-curvature segments from a single prototype image, or deriving them entirely from data fitting (and random optimization) by performing nonlinear least squares optimization (Coleman and Li, 1996; Dennis, 1977) or with genetic algorithms (Goldberg, 1989). The coefficients (for the linear and nonlinear components A, B, C, AB, AC, BC, and ABC) are determined in a similar manner.

The desired responses are chosen to be either relative (in the proportion of average-to-average Earth Mover's Distances, as in the normalized EMD values to be seen in Figure 5.6) or, alternatively, absolute (constant high for the active category and constant low for all other categories: 20-30 Hz within-category, 1 Hz out-of-category).

Note that each IT-like cell is tuned (trained) with a variable number of images from the Kanwisher dataset belonging to one of the seven different selected categories (i.e., a subset of the images that will later be used to evaluate (test) the cells). Rigorous evaluations of classification errors and the segregation of training and testing data are not essential for our present work, as they were in our previous work (Chapter 4). Similarly, we do not concern ourselves with large datasets as we did previously. Here, we are primarily interested in functional behavior.

For a 3-Gaussian-constituent model of an IT cell of the form (A + B + C + AB + AC + BC + ABC), a total of 31 parameters must be selected or determined and the distribution of all parameter values for all such cells is considered. For an alternative model of the

form $(A + B + C)^2 = A^2 + B^2 + C^2 + 2AB + 2AC + 2BC$, 30 parameters must be selected.

We evaluate the performance of our IT population on a set of real images as a function of the V4-like cell inputs. The stimuli (2-dimensional closed contours representing object boundaries) evoke a pattern of activity across the population of IT cells. Shape recognition is evaluated by demonstrating that the patterns of activity across the units to members of a particular object class resemble each other to a higher degree than they resemble members of any other class, corresponding to the measure reportedly used by humans and monkeys in object classification (Sigala *et al.*, 2002). We examine the cell population's response to each image and the response of each cell from the population to each of our test images.

To further study the cell response space, we perform a principal components analysis (Jolliffe, 2002) and a 2-dimensional and 3-dimensional non-classical non-metric multidimensional scaling analysis (Borg and Groenen, 2005; Cox and Cox, 2001). We find the correlation coefficients of the observations (cell responses) and variables (images) and determine the sub-population of cells that are most effective at identifying a particular category. We represent the p-values used for testing the hypothesis of no correlation. We construct a support vector machine (SVM) for classification based upon cell population response (Müller *et al.*, 2001; Schölkopf *et al.*, 2001). We train it with six randomly-chosen examples from each category and test with two randomly-chosen examples from each category. We also fit a tree-based model for classification,

employing Gini's diversity index criterion for choosing a split (Breiman *et al.*, 1984). Finally, representations of particular cells' Gaussian constituent shape selectivity models are rendered on top of actual test images.

All simulations were carried out in a Microsoft Windows XP Professional SP2 environment on an Intel® Pentium® 4 CPU running at 2.80 GHz with 3.00 GB of RAM. All models were constructed using the MATLAB application development environment (version 7.9.0.529 R2009b) and the associated Curve Fitting Toolbox (version 2.1), Genetic Algorithm and Direct Search Toolbox (version 2.4.2), Image Processing Toolbox (version 6.4), Neural Network Toolbox (version 6.0.3), Optimization Toolbox (version 4.3), Signal Processing Toolbox (version 6.12), Statistics Toolbox (version 7.2) and Wavelet Toolbox (version 4.4.1).

5.3 Results

Example Kanwisher natural images from each of the seven sampled categories (axes, cats, fish, guitars, handsaws, hats, and scissors) in a typical trial sequence are shown in Figure 5.1. Iso-curvature segments are shown in red or blue. A blue highlight indicates an iso-curvature segment within one of the constituent parts (e.g., an axe handle), a red highlight indicates an iso-curvature segment within the other (e.g., an axe blade). The within-category variation in this real image population, including scale and orientation, is apparent. Constituent parts and iso-curvature segments within constituent parts for one

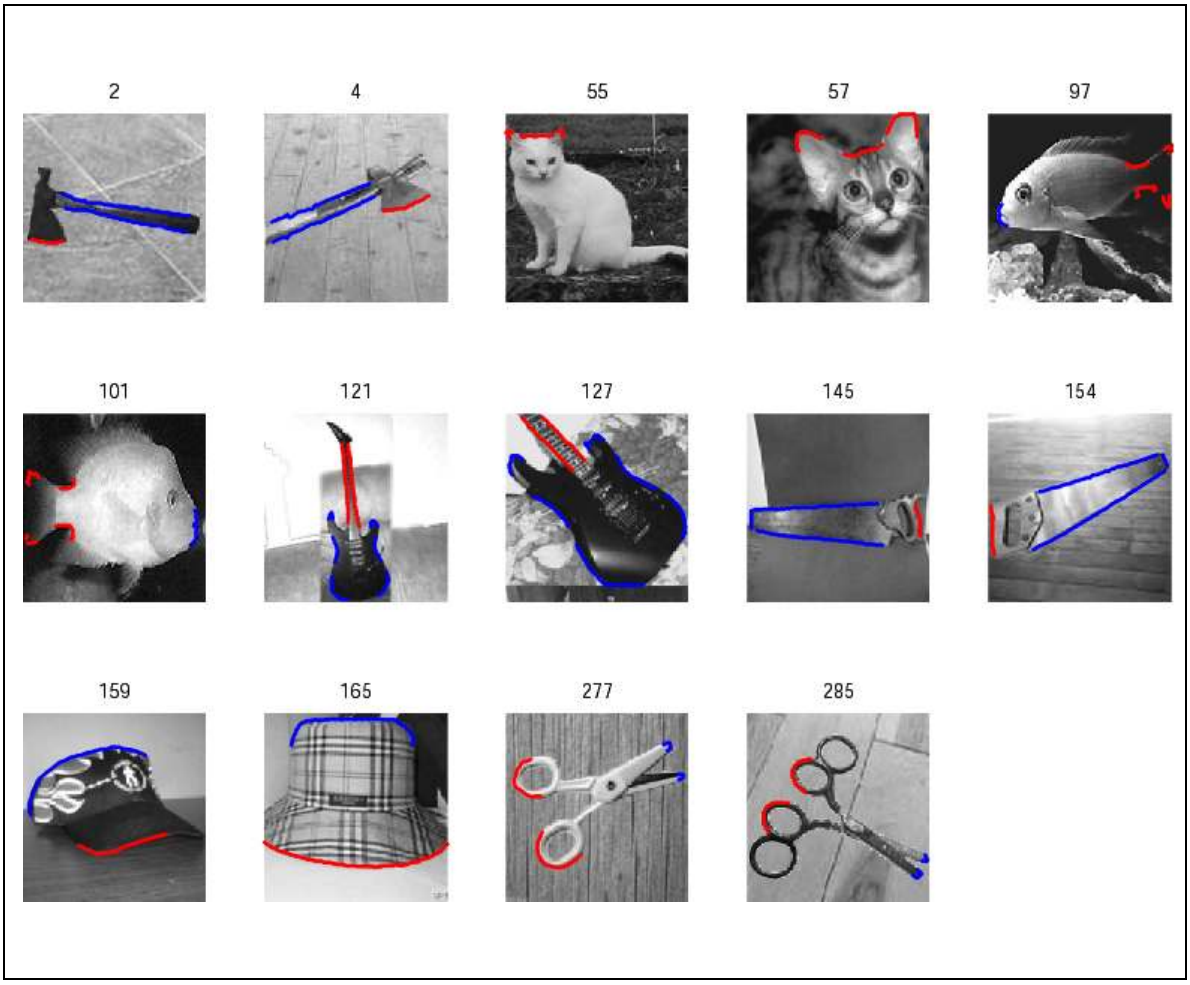

Figure 5.1 – Example images. Image numbers are given. All 7 categories (axes, cats, fish, guitars, handsaws, hats, and scissors) are represented. Iso-curvature segments within constituent parts are highlighted in red or blue.

image are shown in Figure 5.2. Other segment characteristics, such as curvature direction, smoothed curvature, and curvature parameters estimated with an osculating circle, are also illustrated. Note that concave curvatures are considered negative, convex curvatures positive. In Figure 5.3 we demonstrate the image normalization performed as a prelude to our recognition process. We expand or contract along the image's longest axis to fill the frame and orient vertically. We make the claim of affine transformation (scale, rotation, translation) invariance. The degree of variability in scale, rotation and translation of the images is modest, and is within the range of sensitivity to scale and rotation observed in V4 cells (Logothetis *et al.*, 1995) and this standardization of stimuli before the analysis essentially eliminates concerns about rotation and scale invariance. In Figure 5.4 we illustrate the feature vector values for each of the iso-curvature segments and for both constituent parts of a single 360 × 360 image. The features are: angle (made by a line to the mid-point of the iso-curvature segment from the center of mass, measured counter-clockwise from 3 o'clock), curvature of the iso-curvature segment, direction of this curvature and distance (of the mid-point of the iso-curvature segment) from the center of mass. Together, these figures illustrate the simple image processing techniques used to decompose each image into closed contours, each consisting of a number of independent iso-curvature segments.

To establish a baseline, we initially explore the performance capabilities of a V4-like cell population, identical to our earlier work (Chapter 4), and equivalent to an IT network without nonlinear components. We select ten Kanwisher natural images from each of the seven sampled categories (axes, cats, fish, guitars, handsaws, hats, and scissors) as

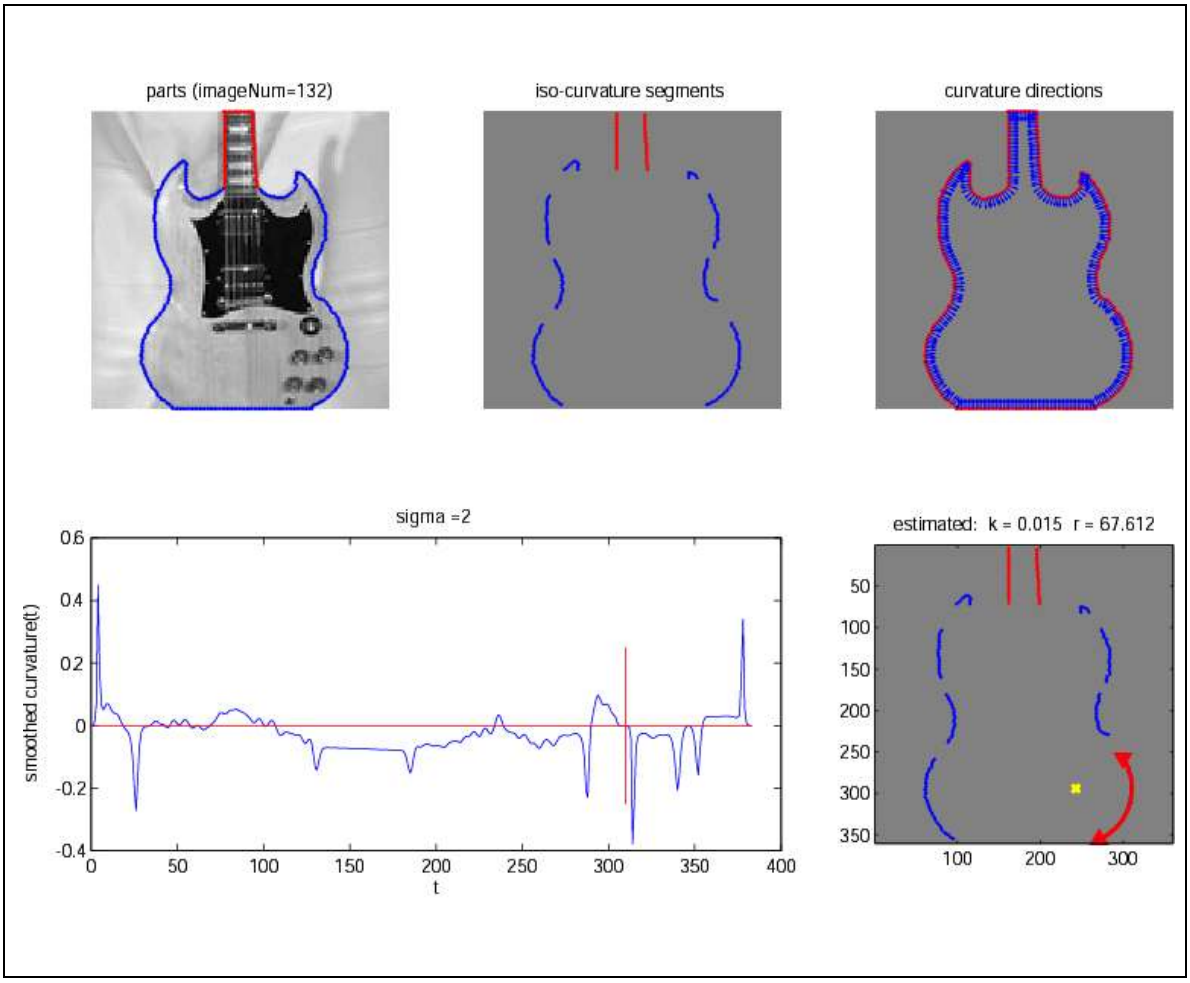

Figure 5.2 – Parts, segments, curvatures, directions, etc., for one image. Constituent parts (2) and iso-curvature segments within constituent parts (2 red and 8 blue) for one image are highlighted. All curvature directions point "inside" the figure. The vertical red line in the bottom left smoothed curvature plot represents the boundary between the first and second constituent part. The curvature is smoothed with a specific parameter (sigma = 2). Note that concave curvatures are considered negative, convex curvatures positive. The red arrowed segment in the bottom right plot represents the estimation of curvature (k) with an osculating circle of a certain radius (r). The yellow "x" represents the center of this osculating circle.

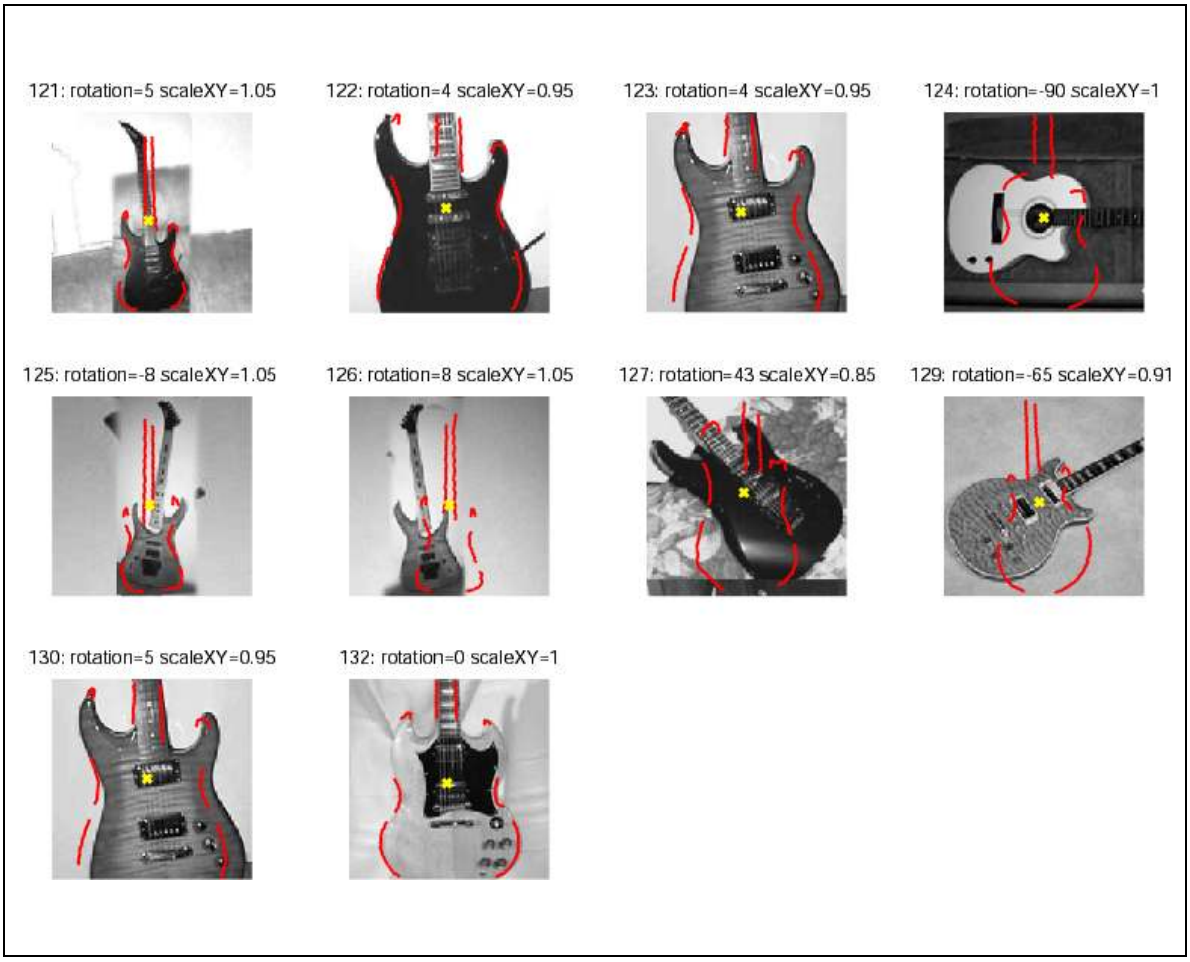

Figure 5.3 – Image normalization. Each image (with image numbers given) in every category is normalized by orienting (degrees of rotation value) vertically and expanding or contracting (scale value) to fill the frame. The center of mass is indicated with a yellow "x". The red highlights represent the "new" (after normalization) iso-curvature segments (without regard for constituent parts).

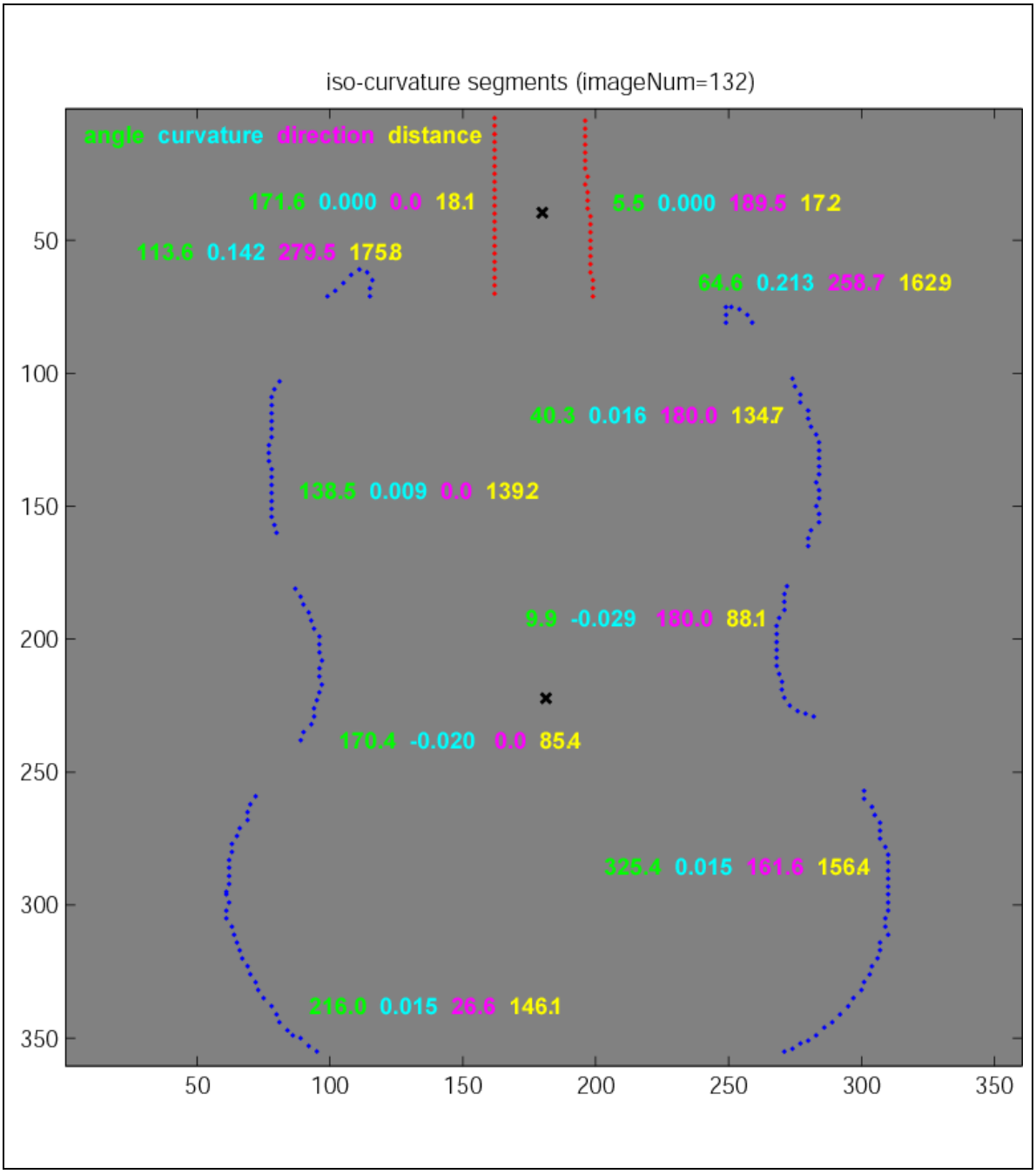

Figure 5.4 – Feature vector values for one image. Constituent parts and iso-curvature segments within constituent parts for one image are highlighted. The center of mass of each constituent part is shown with a black "x". The values of the feature vector components (angle, curvature, direction of curvature and distance from the center of mass) for each iso-curvature segment are shown.

stimuli. The feature vector for each image is compared to that of every other image, and the average distance from one image to all other same-category example images is determined. This average (earth mover's) distance calculation is motivated by human perceptual studies – evidence suggests that categorization is based on an average fit approximation to the class, rather than on exact matches to prototypes (Kahana and Sekuler, 2002). If an image's lowest average distance is to a group of images representing the same category then a match is said to have occurred. Otherwise, an invalid classification results. The experiment shown in Figure 5.5 employs feature vectors (one for each iso-curvature segment) composed of the mean angle of the region, the mean curvature of the region, the mean direction of curvature of the region and the mean distance from the center of mass of the region. Figure 5.5 is the matching matrix, representing the number of images that are classified correctly. The blocks of the same categories are read bottom left to top right. Several images are incorrectly classified, resulting in an 85.7143% correct classification performance. The total matching to average methodology is similar to prototype methods. The affine transformation (scale, rotation, translation) invariance of our model is an advantage here. Alternatively, a closest matching methodology could have been employed, resulting in a 98.5714% correct classification performance. Our 85.7% performance is below current image processing standards, but encouraging in that it precedes the nonlinear integration of boundary components seen in IT. In our previous work (Chapter 4) we have demonstrated the utility of a population of cells without nonlinearity in categorization tasks. Perhaps it is precisely the nonlinear integration component of the IT cells' functionality that facilitates recognition at the highest level.

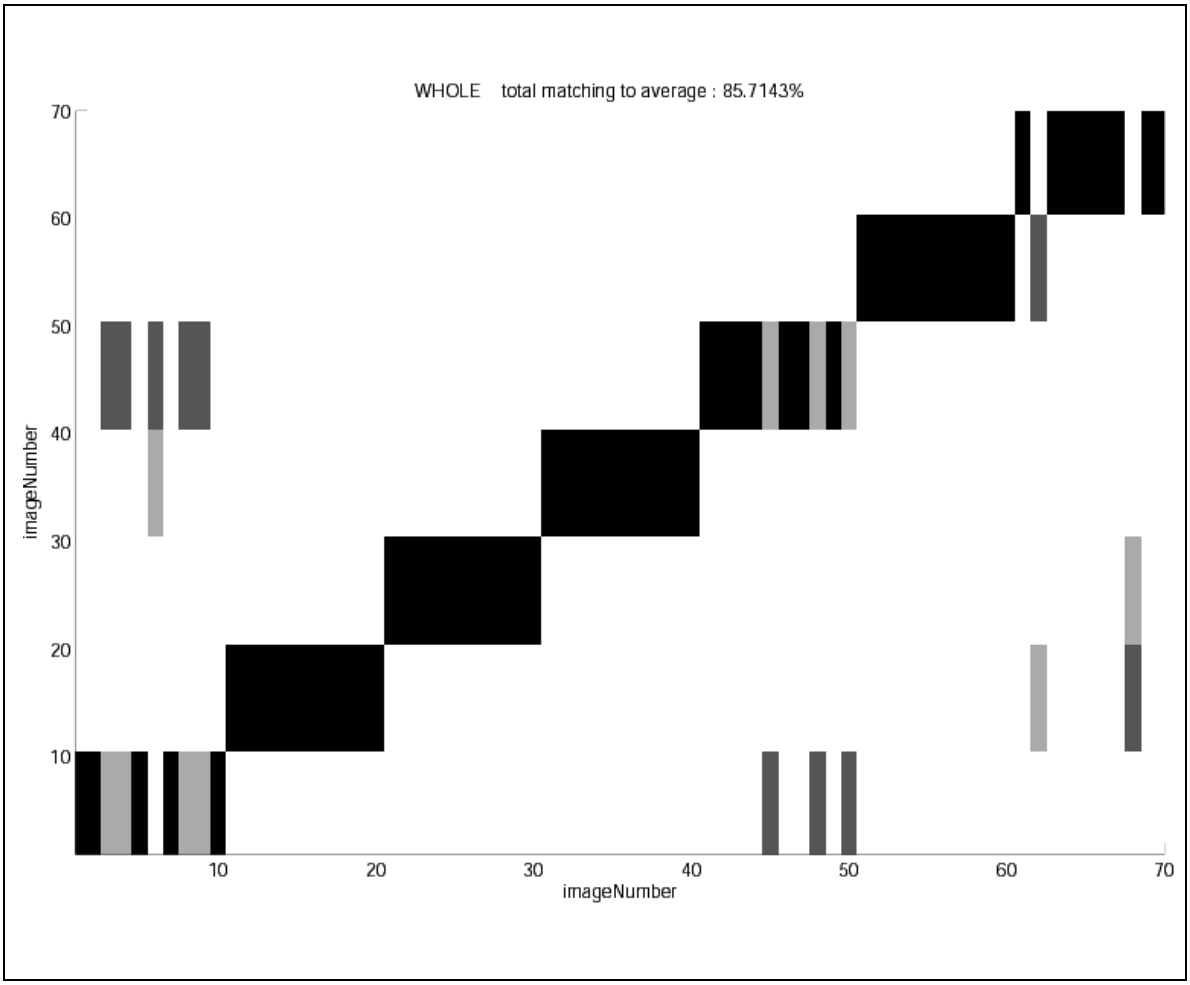

Figure 5.5 – Matching matrix. For each image, the closest match amongst the average image values for each category is determined. Correct matches appear as contiguous rectangles on the diagonal, with widths related to the number of sample images for the corresponding category. Mismatches appear off the diagonal. Image numbers (with 10 sample images from each of the 7 categories) are on the axes. The abscissa can be thought of as representing the stimulus categories; the ordinate can be thought of as representing the response categories. Darker bars (black represents the maximum) indicate a greater number of matches. Using the angle, curvature, direction and distance features, with a region size of 18 and a sigma value of 2, the total matching to average result is 85.7%. Only whole images are considered (i.e., no constituent parts).

Continuing with the same feature vectors, Figure 5.6 illustrates the normalized average-to-average (earth mover's) distance between the image categories with our model's affine transformation (scale, rotation, translation) invariance once again an advantage. It is not surprising that the same-category distances – along the diagonal – are the smallest. Note that symmetry would exist only without affine rotations. (Distance is symmetric, but closest match is not.) These normalized distance values are later used in the creation of portions of the cell populations. Figure 5.7 illustrates the percentage of total (earth mover's) distance comparison for each image and each image category (including component parts). The upper figure pertains to the handsaw "blade" component part, the lower figure to the handsaw "handle" component part. Each contains ten sections along the abscissa – one for each handsaw image. The height of each bar, with values on the ordinate, represents the inverse of distance as a percentage of the total distance from the handsaw part ("blade" or "handle") to the other part. Not surprisingly, each handsaw blade is, in general, closest to the population of handsaw blades and each handsaw handle is, in general, closest to the population of handsaw handles. Where this is not the case (for handsaw handle number 5, for example), we speculate that the combination of component parts (blades plus handles) would facilitate recognition, using, for example, a "mixture of experts" classifier with a Hierarchical Mixture Model (HMM) (Jacobs *et al.*, 1991; Jordan and Jacobs, 1994; Titsias and Likas, 2002).

Each image in a single category ("guitar") is shown in the top 2 rows of Figure 5.8. After image normalization (as in Figure 5.3) and traversal of the bounding contour in the

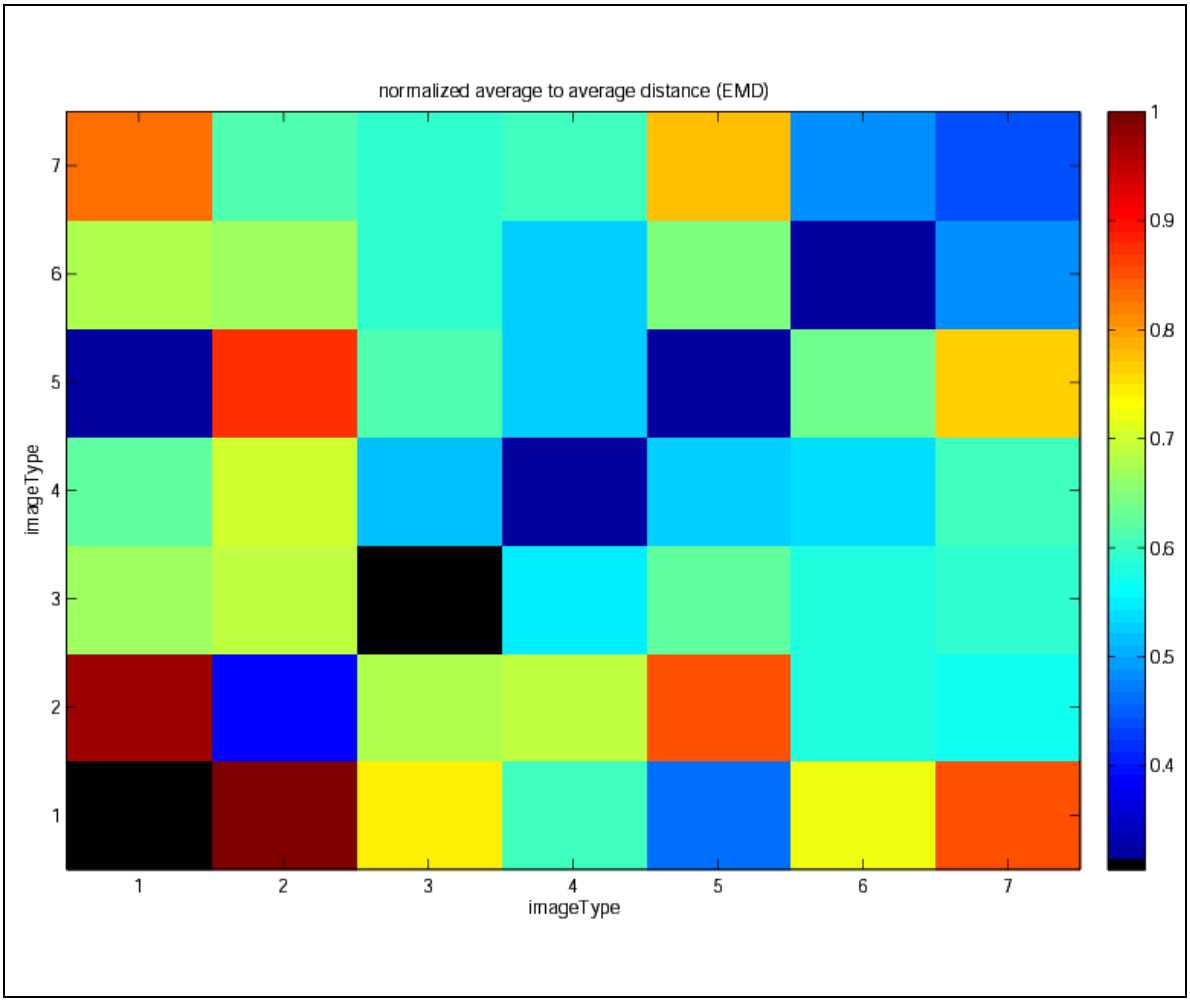

Figure 5.6 – Normalized average-to-average Earth Mover's Distance. The 7 image types (categories) are on the axes. The abscissa can be thought of as representing the stimulus categories; the ordinate can be thought of as representing the response categories. Cooler colors (see the color bar) represent closer distances between the categories.

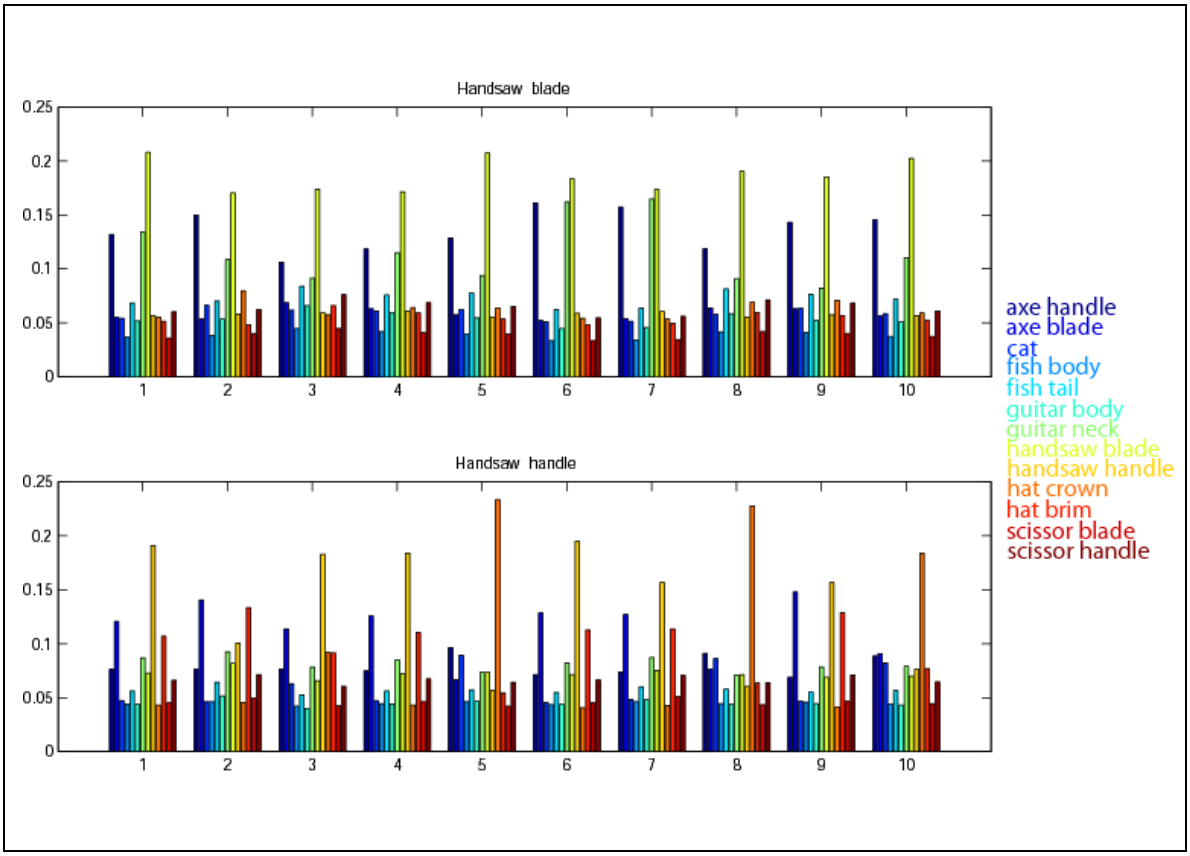

Figure 5.7 – Percentage of total Earth Mover's Distance comparison. Each image and each image category (including component parts) are considered. The upper figure pertains to the handsaw "blade" component part, the lower figure to the handsaw "handle" component part. Each contains 10 sections along the abscissa – one for each handsaw image. Bar colors correspond to component parts, listed on the right. The height of each bar, with values on the ordinate, represents the inverse of distance as a percentage of the total distance from the handsaw part ("blade" or "handle") to the other, color-coded part.

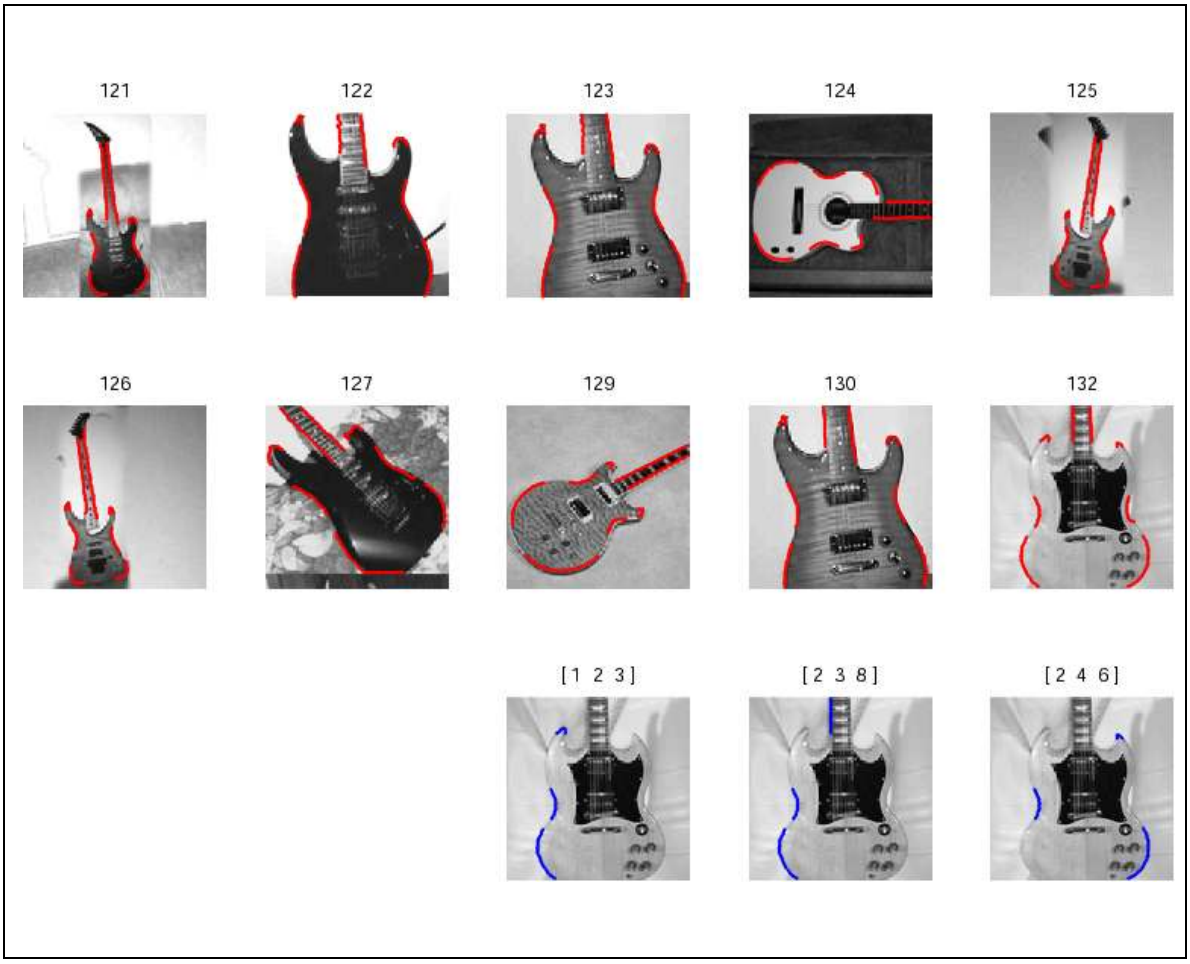

Figure 5.8 – Iso-curvature segments. Each image (with image numbers given) in a single category ("guitar") is shown in the top 2 rows. After image normalization (as in Figure 5.3) and traversal of the bounding contour in the counter-clockwise direction (starting at approximately 11 o'clock in each normalized image), 8 ordered iso-curvature segments can be found for each image. These iso-curvature segments are highlighted in red. Only whole images are considered (i.e., no constituent parts). The bottom row shows 3 different selections of 3 iso-curvature segments each for a single image, highlighted in blue, with selected segment numbers given.

counter-clockwise direction (starting at approximately 11 o'clock in each normalized image), 8 ordered iso-curvature segments can be found for each image. The bottom row shows 3 different selections of 3 iso-curvature segments each for a single image. These selected iso-curvature segments are later used to create the IT cell population through averaging, nonlinear least-squares optimization, etc. The essential components of these IT-like cells are known as Gaussian constituents (A, B, C). They receive their inputs directly from the V4-like cells (i.e., their responses are based upon the evaluation of iso-curvature segments). The first 2 recognition tasks can be seen as object detection and object categorization. The third (and perhaps final) recognition task – within-category identification – is well-illustrated in Figure 5.8. Consider, for example, the difference between the identification of an electric guitar and the identification of an acoustic guitar.

Figures 5.9 and 5.10 illustrate the distribution of parameters for a 6-cell population, with each cell created using one of the investigated techniques. The mean and standard deviation of each of the cell's four features (angle, curvature, direction of curvature, distance) for three Gaussian constituents (A, B, C) (for a total of 24 parameters) are derived using different techniques (one for each cell): using the average of iso-curvature segments (the selected [1 2 3] segments for Figure 5.9; the selected [2 3 8] segments for Figure 5.10) from multiple images, using a single prototype image's iso-curvature segments, or entirely from data fitting (and optimized randomly). The desired response can be relative (as in the normalized average-to-average Earth Mover's Distance value) or absolute (20-30 Hz within-category, 1 Hz out-of-category). Figure 5.9 represents the arrangement "A + B + C + AB + AC + BC + ABC" with 7 additional (fitted) coefficient

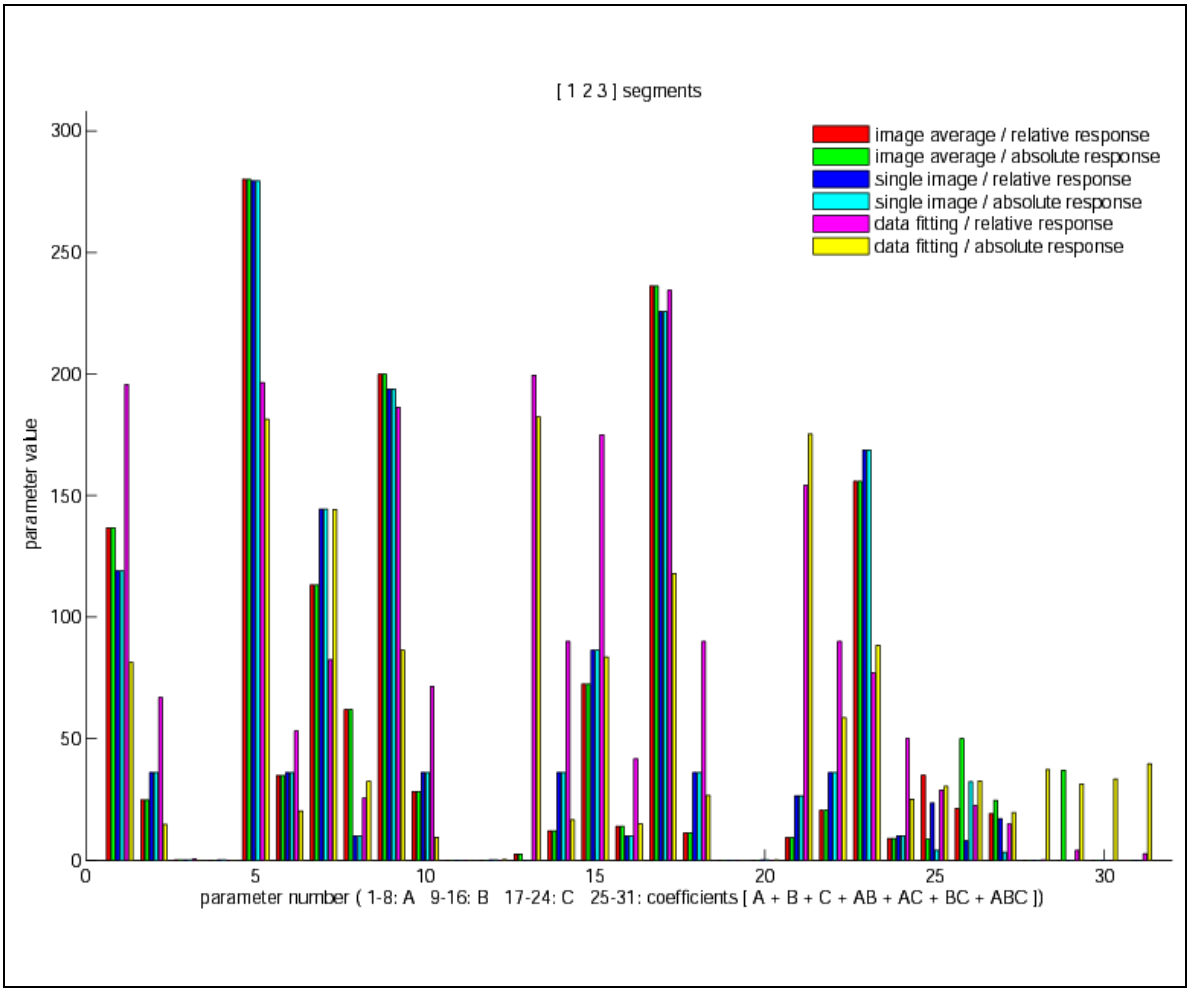

Figure 5.9 – Distribution of parameters for a 6-cell population. Parameter numbers, in groups of 6, are given along the abscissa. The height of each bar, with values on the ordinate, represents the parameter value. The mean and standard deviation of each of the cell's four features (angle, curvature, direction of curvature, distance) for three Gaussian constituents (A, B, C) (for a total of 24 parameters) are derived using different color-coded techniques (one for each cell): using the average of iso-curvature segments (the selected [1 2 3] segments for the case illustrated) from multiple images, using a single prototype image's iso-curvature segments, or entirely from data fitting (and optimized randomly). The desired response can be relative (as in the normalized EMD value of Figure 5.6) or absolute (20-30 Hz within-category, 1 Hz out-of-category). The nonlinear arrangement "A + B + C + AB + AC + BC + ABC" with 7 additional (fitted) coefficient parameters required (for a total of 31) is represented. Parameter 1 represents A's angle mean, parameter 2 represents A's angle standard deviation, parameter 3 represents A's curvature mean, parameter 4 represents A's curvature standard deviation, parameter 9 represents B's angle mean, etc.

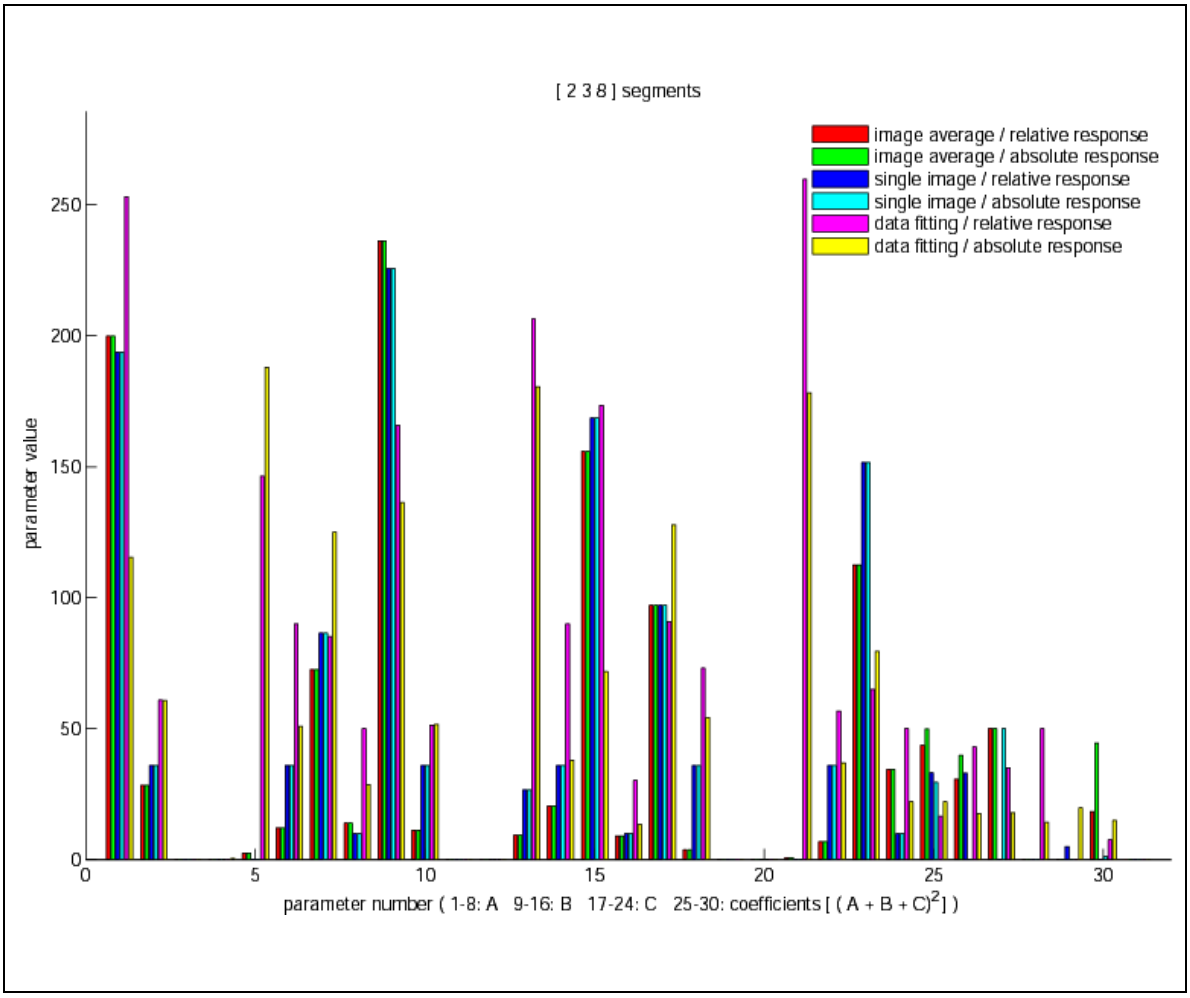

Figure 5.10 – Distribution of parameters for a 6-cell population. Parameter numbers, in groups of 6, are given along the abscissa. The height of each bar, with values on the ordinate, represents the parameter value. The mean and standard deviation of each of the cell's four features (angle, curvature, direction of curvature, distance) for three Gaussian constituents (A, B, C) (for a total of 24 parameters) are derived using different color-coded techniques (one for each cell): using the average of iso-curvature segments (the selected [2 3 8] segments for the case illustrated) from multiple images, using a single prototype image's iso-curvature segments, or entirely from data fitting (and optimized randomly). The desired response can be relative (as in the normalized EMD value of Figure 5.6) or absolute (20-30 Hz within-category, 1 Hz out-of-category). The nonlinear arrangement "$(A + B + C)^2 = A^2 + B^2 + C^2 + 2AB + 2AC + 2BC$" with 6 additional (fitted) coefficient parameters required (for a total of 30) is represented.

parameters required (for a total of 31). Figure 5.10 represents the arrangement "(A + B + C)^2 = A^2 + B^2 + C^2 + 2AB + 2AC + 2BC" with 6 additional (fitted) coefficient parameters required (for a total of 30). The wide variety of parameter values used in the cell construction – both between techniques and between nonlinear arrangements and segment selections – is apparent. In Figure 5.11 we show the normalized distribution of all parameter values for all cells in a 42-cell population, derived with various techniques in the manner of previous smaller populations. Again, the wide variety of parameter values is apparent.

In the top row of Figure 5.12, we show two particular cells' Gaussian constituent shape selectivity models rendered on top of two actual test images. With affine transformation (scale, rotation, translation) invariance and normalized average-to-average EMD distance symmetry, these cells would probably still respond well to these images, even though they are not ideally aligned. In the bottom row of Figure 5.12, we show two images' iso-curvature segments. For an IT-like cell to respond maximally to this image, V4-like cells (inputs to IT) would have to be perfectly aligned with these contour segments. Gaussian constituents are properties of cells, whereas iso-curvature segments are properties of images. In Figure 5.13 we show the (A, B, C) Gaussian constituents of a particular cell, along with the mean and standard deviation for each element of the feature vector (angle, curvature, direction of curvature, distance) for each Gaussian function. Again, for a maximal response from this cell, this image's features (the angle, curvature, direction of curvature and distance of each of the iso-curvature segments) would have to be perfectly aligned with those of the Gaussian constituents. Otherwise, as is the case for this

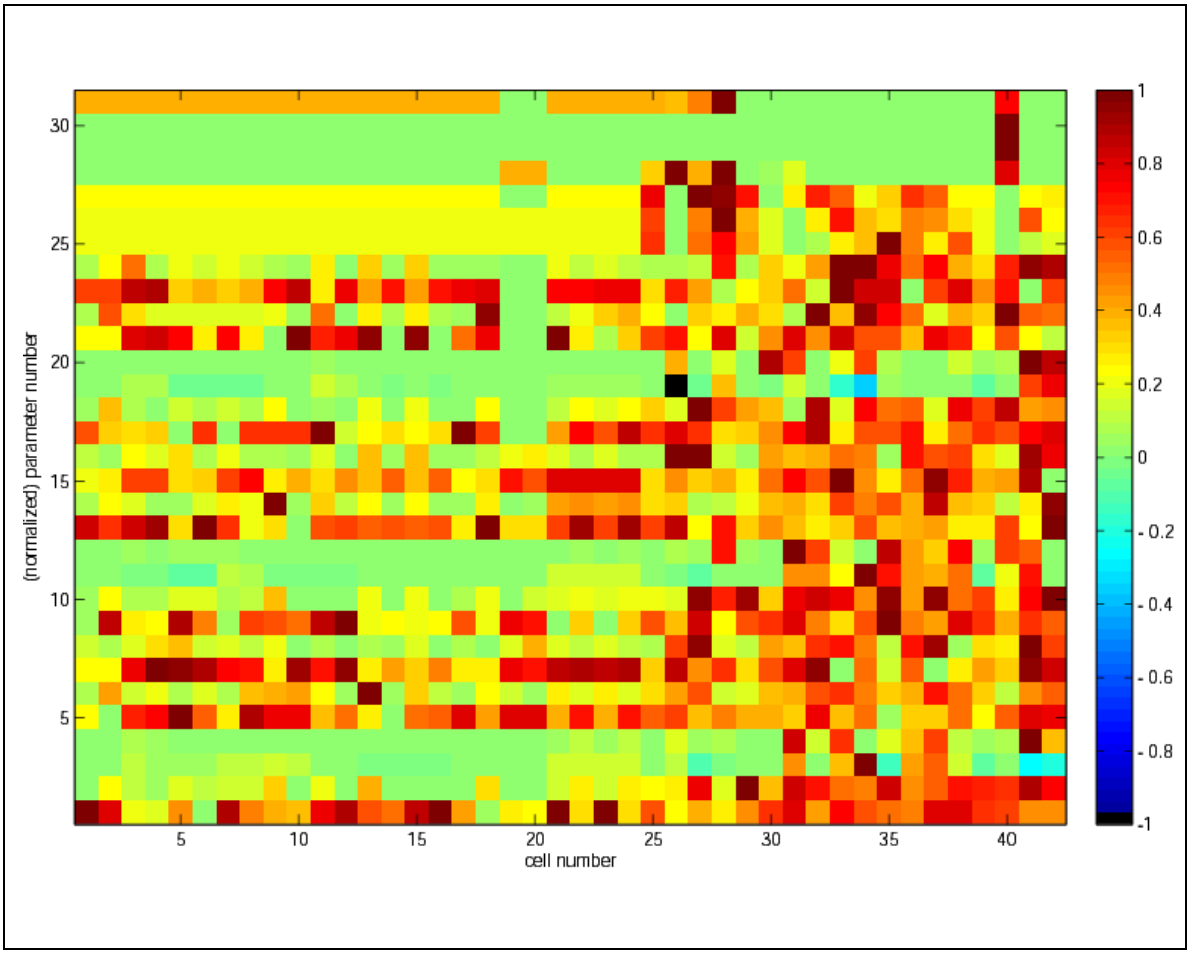

Figure 5.11 – Normalized distribution of all parameter values for all cells. Cell numbers in a 42-cell population are on the abscissa. Parameter numbers are given on the ordinate. The parameter values are normalized to range from −1 to +1 (see the color bar). The nonlinear arrangement "A + B + C + AB + AC + BC + ABC" (requiring 31 parameters) was the standard for all cells in the population, with the various derivation techniques used.

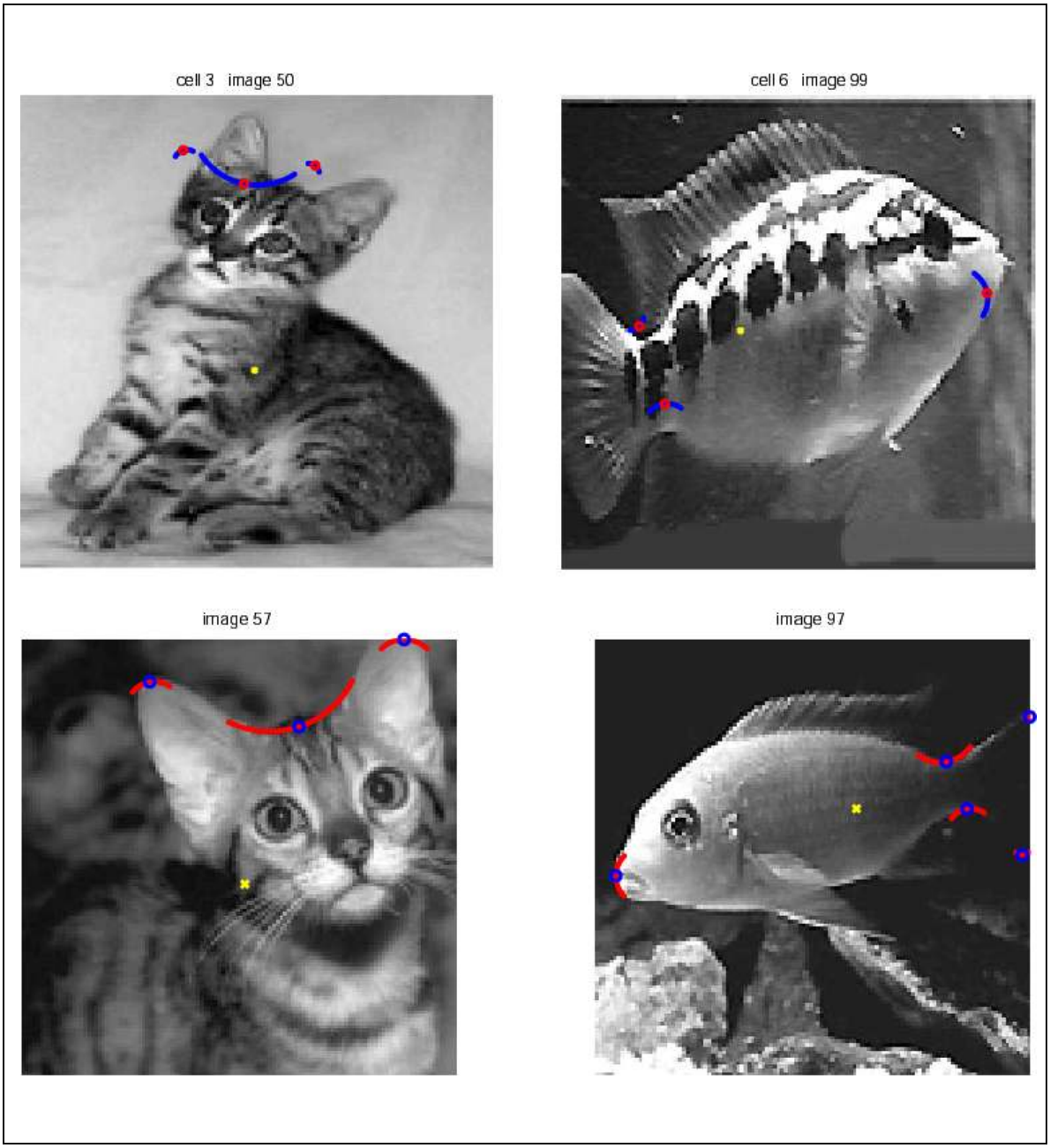

Figure 5.12 – Gaussian constituents and iso-curvature segments. The Gaussian constituent shape selectivity models (properties of cells) are shown in the top row (blue contours with red-circled centers) for 2 particular cells and rendered on top of 2 particular images. The iso-curvature segments (properties of images) are shown in the bottom row (red contours with blue-circled centers) for 2 particular images. The center of mass is indicated with a yellow "x".

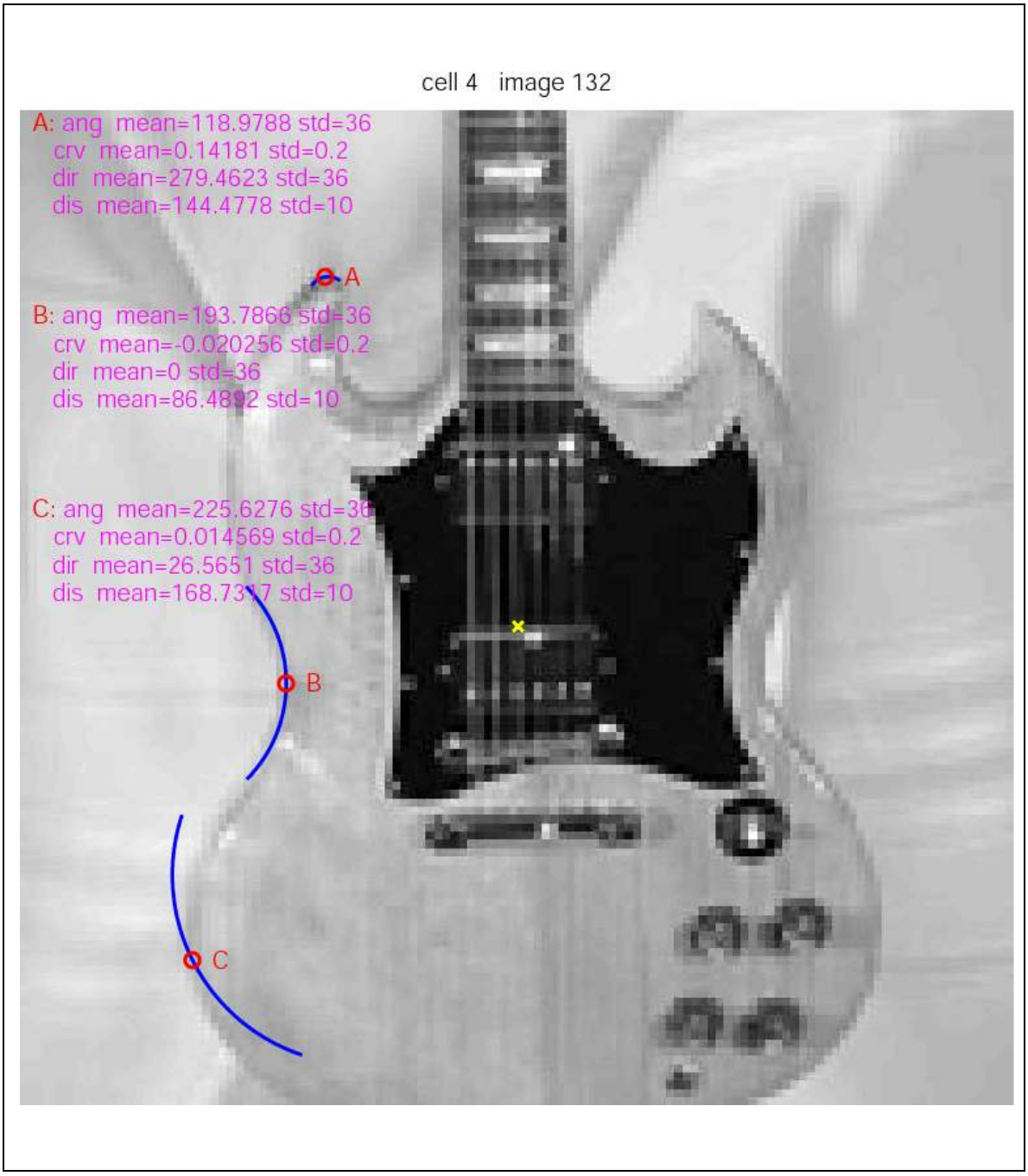

Figure 5.13 – Gaussian constituents. The (A, B, C) Gaussian constituents (blue contours with red-circled centers) for a particular cell are rendered on top of an actual image, with the mean and standard deviation for each element of the feature vector (angle, curvature, direction of curvature, distance) for each Gaussian constituent provided. The center of mass is indicated with a yellow "x".

particular image, a sub-optimal response would result, determined by Gaussian falloffs from the means at rates proportional to the specified standard deviations.

In Figure 5.14 we introduce the histogram of IT cell responses. Here, we show the response of one cell from the population to each image. Clearly, this cell responds preferentially to guitars. This is reminiscent of the study of a single unit in the left posterior hippocampus / medial temporal lobe of epilepsy patients with depth electrodes by Cristof Koch (Quiroga *et al.*, 2005). Here, the cells were activated exclusively by different views of Jennifer Aniston, for example, and not Julia Roberts, etc. In the same manner, we show in Figure 5.15 the responses of four cells from the population to each image. Although there is wide variability in the response of the cells (both to images within- and out-of-category), each clearly favors fish. Note that these cells are not designed to be, nor do they behave like, "grandmother" cells. They simply respond well to cells within a single category, at the exclusion of others. We have created many IT-like cells similar to these that respond preferentially to each of the seven categories in our sample space.

Principal components analysis (PCA) is used for data visualization and dimensionality reduction. It is a linear method concerned with the correlations between the data dimensions, and essentially a coordinate transformation. It is, however, an engineering approach to a data problem, and not at all necessary in the visual system. Nevertheless, it is an efficient means of data compression and representation and is a useful tool for analysis, so we utilize it. The responses of all of the cells in our 42-cell population to all

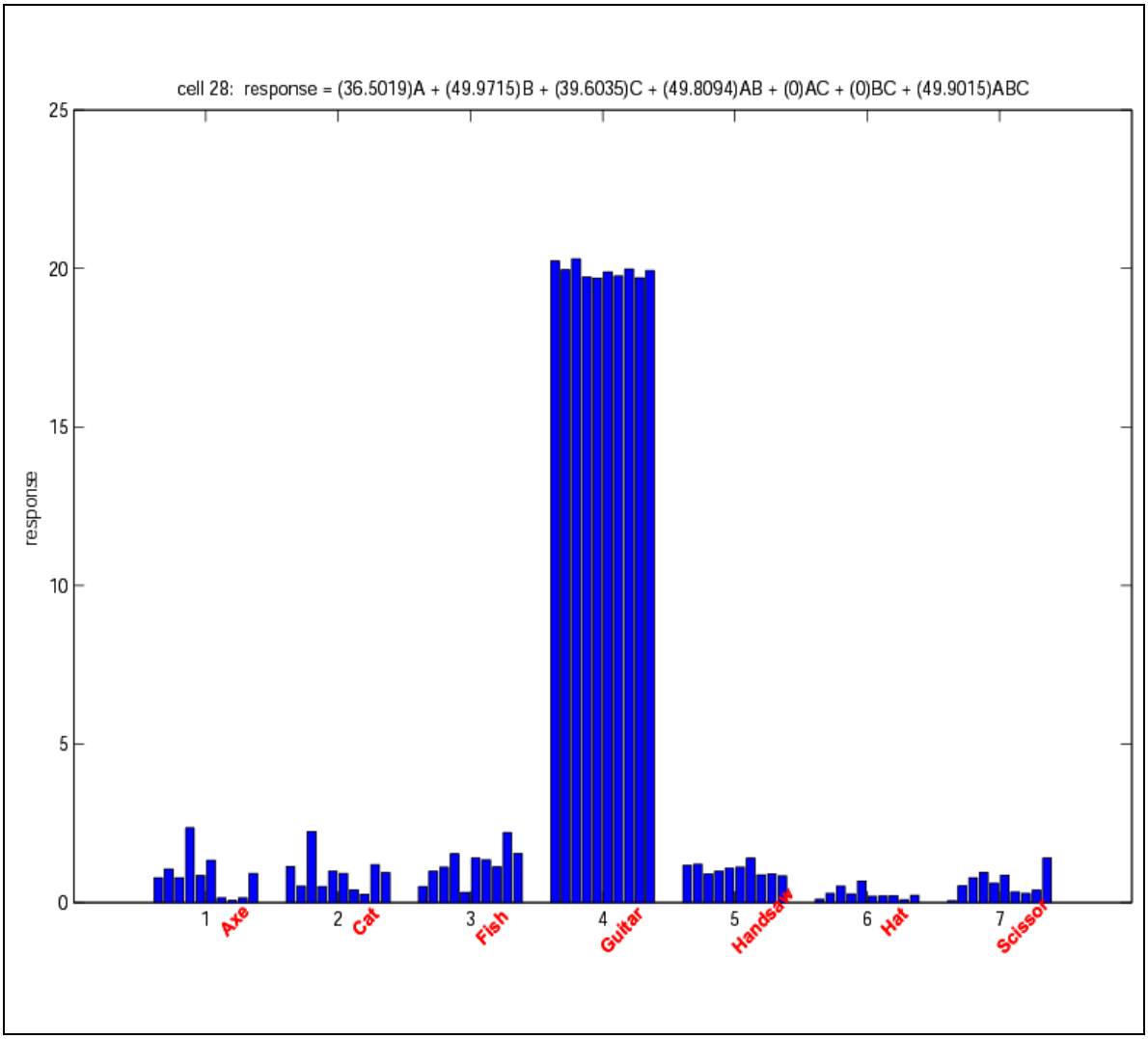

Figure 5.14 – Histogram of IT cell responses. This represents the response of one cell from the population to each image. The 70 images are grouped into 7 categories. The height of each bar represents the response of the cell to each of these images. The cell's full response equation, including fitted coefficients, is shown above. This cell responds preferentially to guitars.

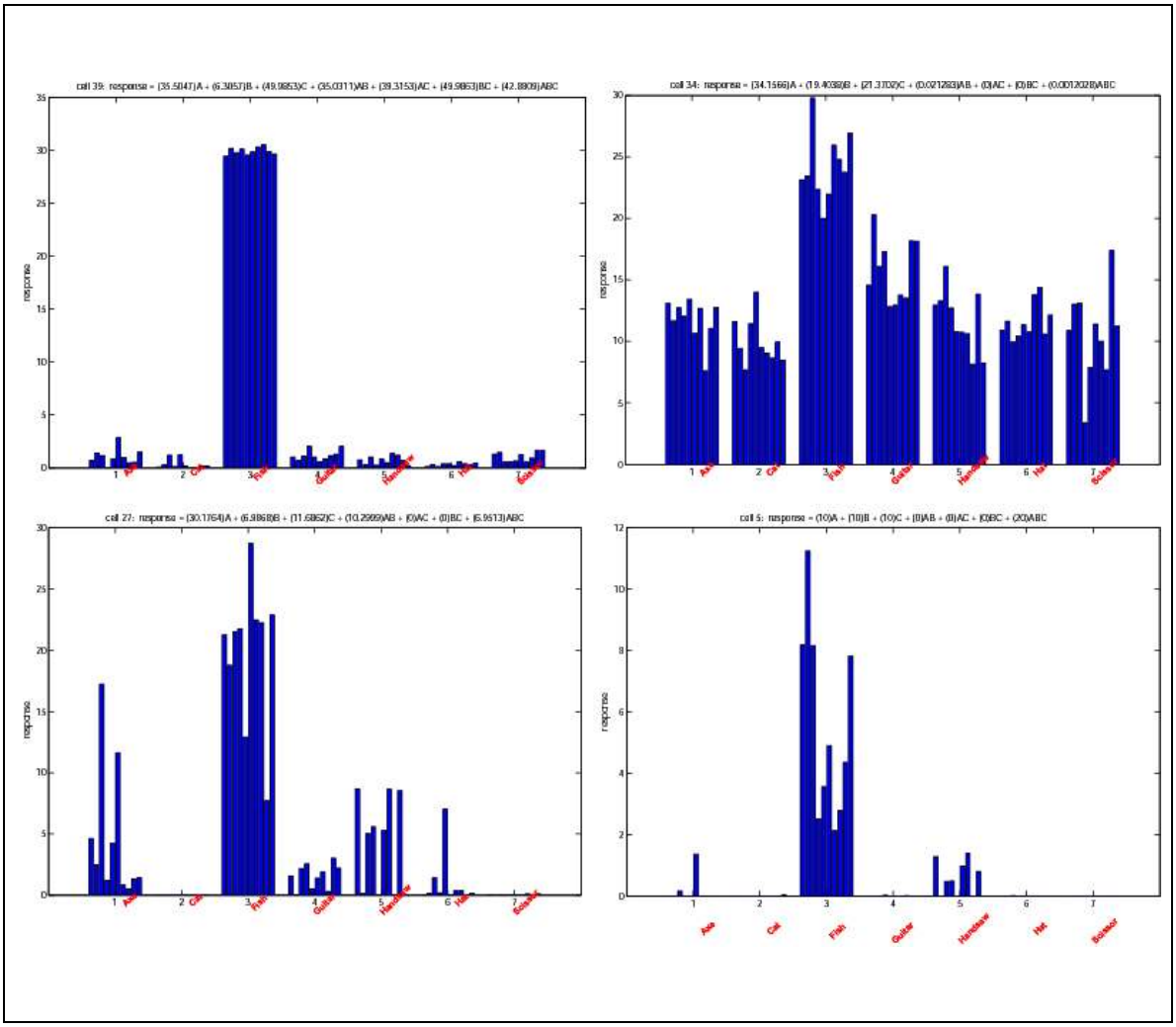

Figure 5.15 – Response of four cells from the population to each image. The 70 images are grouped into 7 categories in each of the 4 panels. The height of each bar in each panel represents the response of each cell to each of these images. Each cell's full response equation, including fitted coefficients, is shown above its panel. Each of these cells responds preferentially to fish.

of the images are subjected to principal components analysis. In Figure 5.16, all of the sample images are projected onto the first two principal components. A 4-Nearest Neighbor classification attempt (Mitchell, 1997) results in a lackluster 15.7 % error. Note that this error is appropriately considered a training error, not a testing error. A biplot allows us to show the magnitude and sign of the contribution of each variable (the responses of the 42 cells) to the first two principal components, and to visualize how each observation (the 70 images – ten from each of seven categories) is represented in terms of those components. We provide this in Figure 5.17. A Pareto chart (also known as a scree plot) is given in Figure 5.18. These help us to visualize the percentage of total variability explained by each principal component. The lack of an obvious "elbow" indicates that all dimensions, or at least more than the first two dimensions, although correlated, contribute to the representation. This fact, coupled with our poor classification performance, lead us to seek a more appropriate method of data visualization and dimensionality reduction for our cell response space.

We use three-dimensional non-classical non-metric multidimensional scaling (MDS) analysis to better visualize our cell response space. This technique shares some of the advantages of the nonlinear dimensionality reduction technique of locally linear embedding (LLE), such as a preservation of topology (Roweis and Saul, 2000). The data representing the dissimilarities between the responses of all of the cells in the same 42-cell population to all of the images is subjected to three-dimensional non-classical non-metric MDS analysis and we present a Shepard plot in Figure 5.19. As is typical, our goal is to create a configuration (to be seen in Figure 5.20) of points in 3 dimensions with

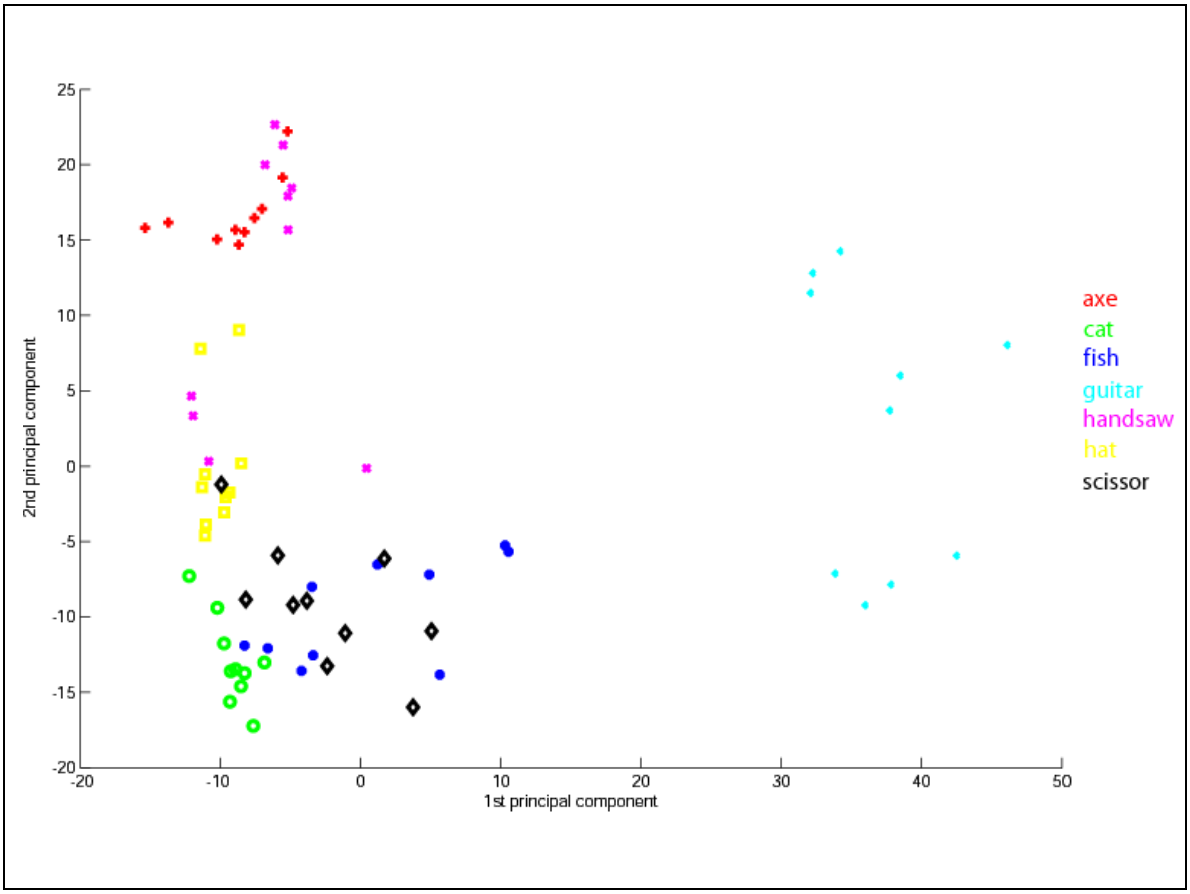

Figure 5.16 – Principal Components Analysis (PCA). The responses of all of the cells in the 42-cell population to all of the images are subjected to principal components analysis. The 10 sample images from each of the 7 categories – coded by color and shape – are projected onto the first 2 principal components axes. Classification (4-Nearest Neighbor) at this level results in an error of 15.7%.

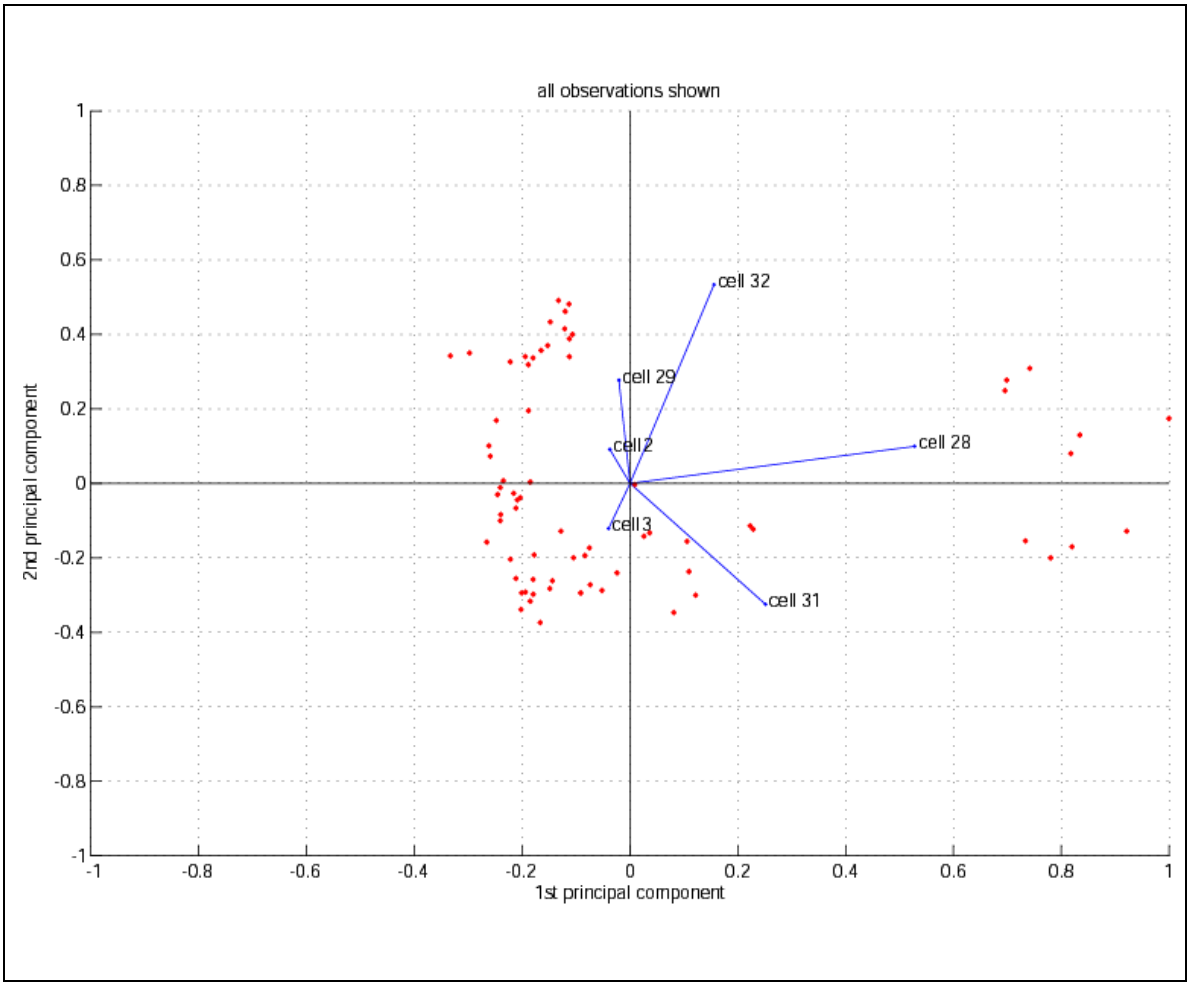

Figure 5.17 – Biplot. As in Figure 5.16, principal components analysis was applied to the cell responses. Of the 42 cells (variables) in the population, 6 representative cells were chosen. The direction and length of each blue vector corresponds to the variable's contribution to the first 2 principal components. The red dots indicate observations (the 70 images). These are the principal component scores – the representations of the observations in principal component space.

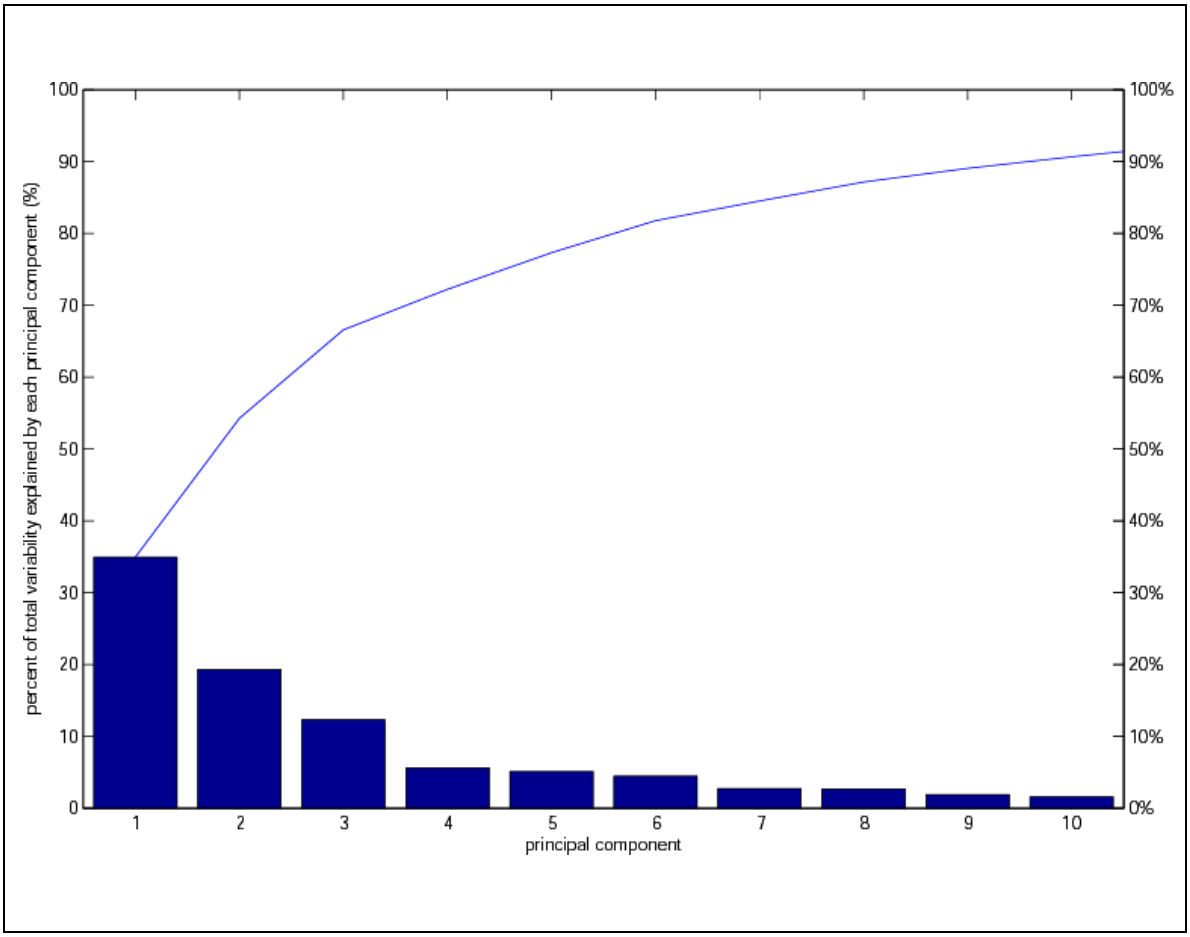

Figure 5.18 – Pareto chart. As in Figure 5.16, principal components analysis was applied to the cell responses. Principal component numbers are given on the abscissa. The height of each bar, with values on the ordinate, represents the percentage of total variability explained by each principal component. The blue line displays the cumulative sum of percentage.

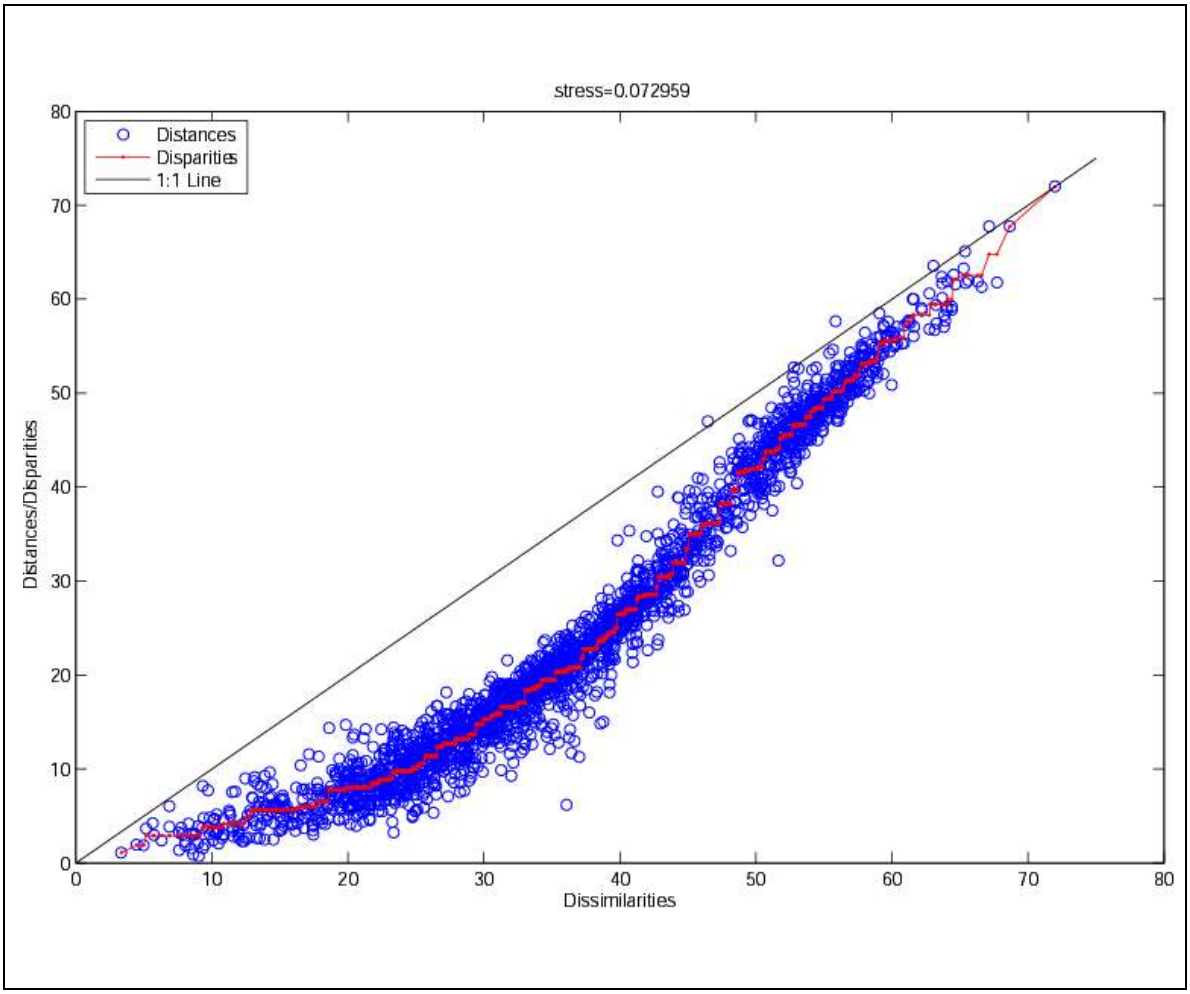

Figure 5.19 – Shepard plot. The data representing the dissimilarities between the responses of all of the cells in the 42-cell population to all of the images is subjected to three-dimensional non-classical non-metric multidimensional scaling (MDS) analysis. Blue circles represent distances, the red line represents disparities, with dissimilarities on the abscissa and the ratio of distances to disparities on the ordinate. Stress (7.3%) measures how well the solution recreates the dissimilarities.

inter-point distances close to the original dissimilarities. But here, instead of trying to approximate the original dissimilarities themselves, disparities are used as nonlinear, monotonic transformations of the dissimilarities, with distances approximating the disparities. The small scatter of blue circles about the red line shows how the distances approximate the disparities well. The nonlinear but increasing red line shows that the disparities reflect the ranks of the dissimilarities. Its concave shape reminds us that the fit tends to contract small distances relative to the corresponding dissimilarities and that only relative distances between points, and not absolute distances, should be taken literally. The stress is a measure of how well the solution recreates the dissimilarities, with smaller values indicating a better fit. A stress of 7.3% is very good, lending confidence to this method.

In Figure 5.20, again using our 42-cell population, all of the sample images are projected onto the three dimensions resulting from the non-classical non-metric MDS analysis. A 4-Nearest Neighbor classification attempt results in a superior 2.9% error, suggesting that this type of analysis makes the data most amenable to clustering. Note again that this error is appropriately considered a training error, not a testing error. The structure of the data at this reduced dimensionality is apparent.

In Figure 5.21 we illustrate the hypothesis of no correlation. The correlation coefficients of the observations (cell responses) and variables (image categories) are derived and the p-values used for testing the hypothesis of no correlation are obtained. The black rectangles represent a p-value of < 0.05, indicating a significant correlation between the

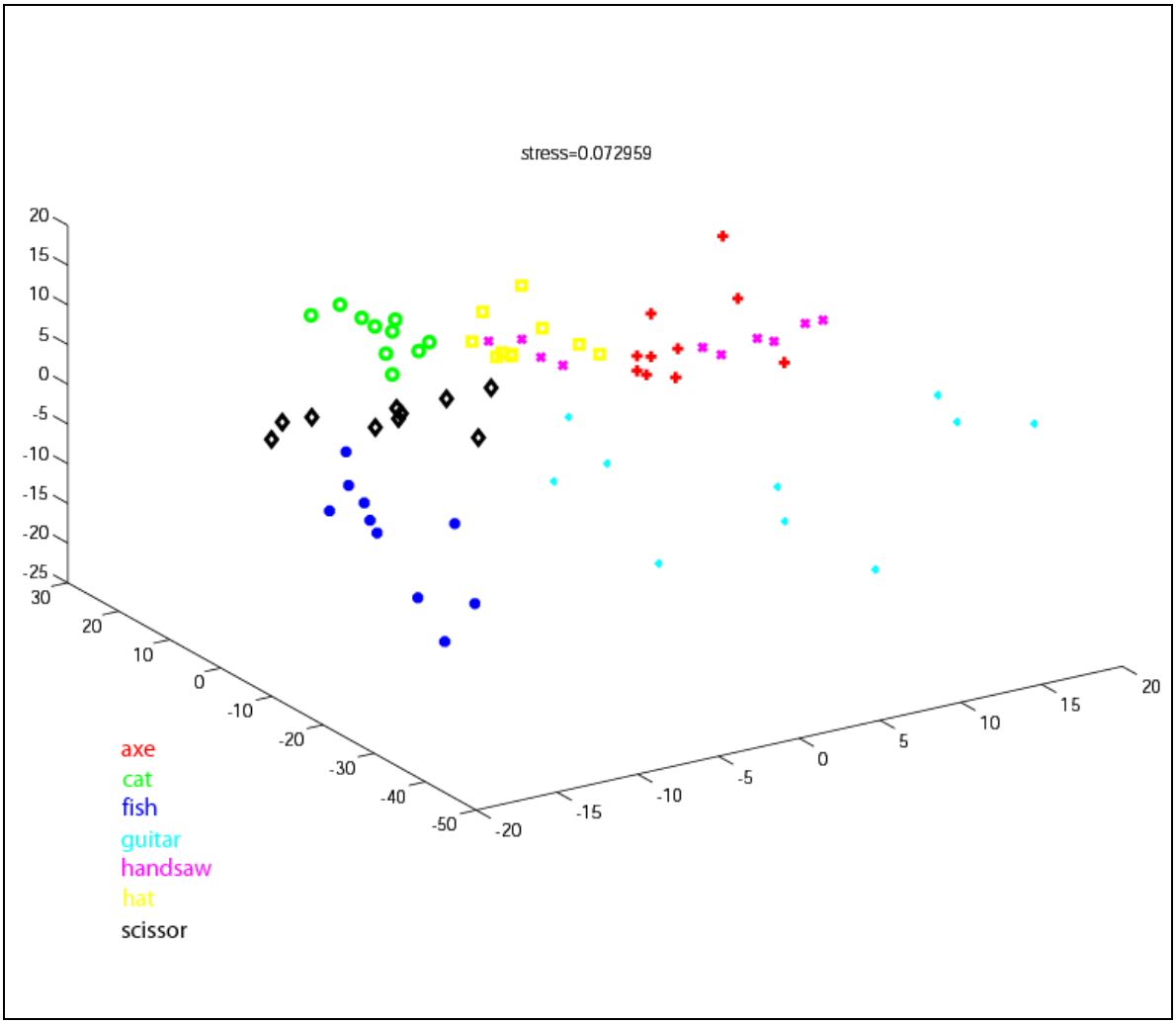

Figure 5.20 – Three-dimensional non-classical non-metric multidimensional scaling (MDS) analysis. As in Figure 5.19, three-dimensional non-classical non-metric MDS analysis was applied to the dissimilarities between the cell responses. On this scatter plot, the 10 sample images from each of the 7 categories – coded by color and shape – are projected onto the 3 dimensions resulting from the analysis. Classification (4-Nearest Neighbor) at this level results in an error of 2.9 %.

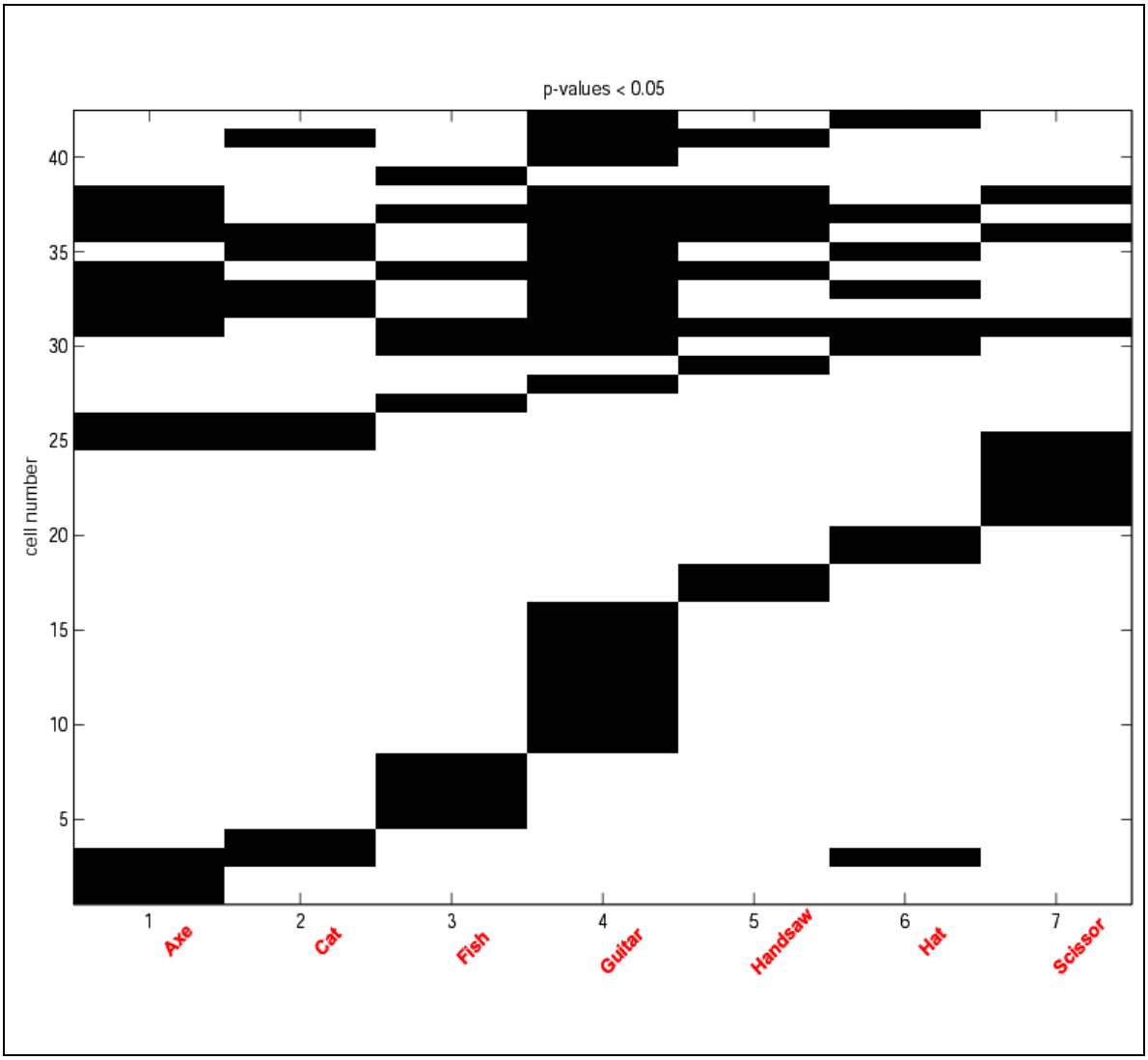

Figure 5.21 – Hypothesis of no correlation. The 70 images are grouped into 7 categories, given along the abscissa. The 42-cell population, with values on the ordinate, is used for evaluation of correlation coefficients. The black rectangles represent a p-value of < 0.05, indicating a significant correlation between the observations (cell responses) and variables (image categories).

observations and variables (i.e., cells correlated with category determination). This could potentially lead us to a sub-population of cells that are most effective at identifying a particular category. (We will explore this concept further in Figure 5.29 and 5.30.) Note that any structure or pattern discerned in the figure's rectangles is artifactual – simply a result of a non-random ordering of the cells. Also note that there are significantly correlated cells for each of the seven categories – some important for just one category, indicating a highly discriminating cell, others spanning multiple categories, indicating a less exclusive, but otherwise useful, cell.

As a broad demonstration of the discriminatory power of our network, we show the cell population responses to each image in two categories in Figure 5.22. The responses (representing IT cell excitation) of each of the cells in the 42-cell population are given for each of the ten "guitar" images and for each of the ten "cat" images (on the right). The differences between the populations for each of the categories are visually apparent.

We also construct a support vector machine (SVM) for classification based upon cell population response. In a fairly simple test, we train it with six randomly-chosen examples from each category and test it with two randomly-chosen examples from each category. We achieve 100% correct categorization. In Figure 5.23 we demonstrate the use of a decision tree for classification. This tree-based model employs a set of simple rules to achieve classification of image categories. Using our 42-cell population, we see that only 6 cells are necessary for 100% performance. These several classification

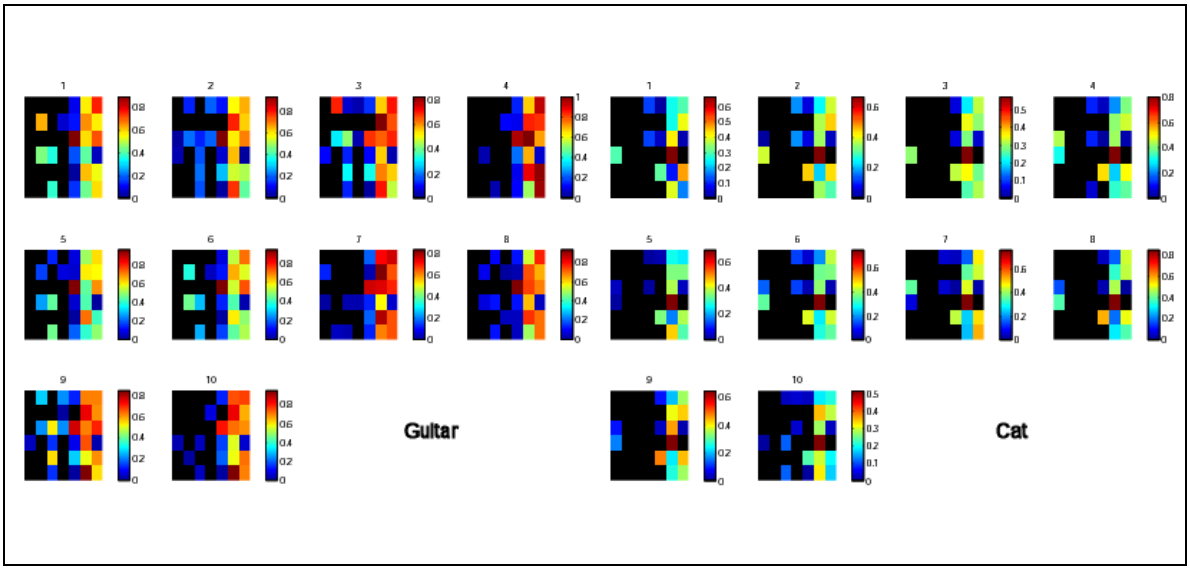

Figure 5.22 – Cell population responses to each image in two categories. The responses of each of the cells in the 42-cell population are given for each of the 10 "guitar" images (on the left) and for each of the 10 "cat" images (on the right). Hotter colors (see the color bars) represent a larger degree of IT cell excitation. The differences between the populations for each of the categories are visually apparent.

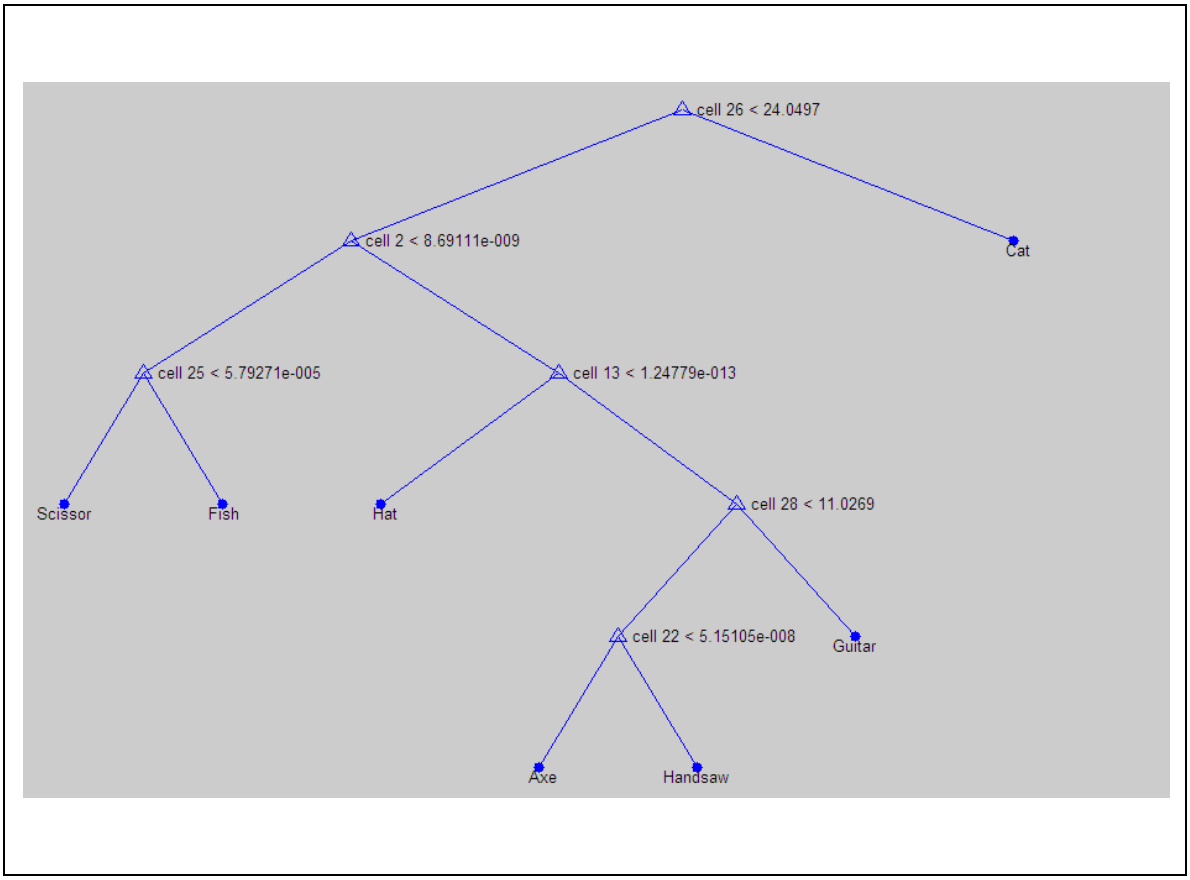

Figure 5.23 – Decision tree for classification. This tree-based model employs a set of simple rules to achieve classification of image categories. At each non-terminal node, the left branch is taken in the case of a "true" decision, the right branch is taken in the case of a "false" decision. Eventually, a terminal node (classification) is reached. Using our 42-cell population, we see that only 6 cells are necessary for 100% performance.

techniques demonstrate that our network can properly categorize a number of real images in a variety ways.

We present the responses of six alternatively-constructed cells to each image in Figure 5.24. Each of the six cells in this figure utilizes the parameter set of one of the cells in Figure 5.9, with matching construction methodologies. The mean and standard deviation of each of the cell's four features (angle, curvature, direction of curvature, distance) for three Gaussian constituents (A, B, C) are derived using different techniques (one for each cell): using the average of iso-curvature segments (the selected [1 2 3] segments for the case illustrated) from multiple images, using a single prototype image's iso-curvature segments, or entirely from data fitting (and optimized randomly). The desired response can be relative (as in the normalized EMD value) or absolute (20-30 Hz within-category, 1 Hz out-of-category). The nonlinear arrangement "A + B + C + AB + AC + BC + ABC" is represented. Each of these cells responds preferentially to guitars. Figure 5.26 is similar to Figure 5.24, but with each of the six cells in this figure utilizing the parameter set of one of the cells in Figure 5.10, with matching construction methodologies. Here, the average of iso-curvature segments from multiple images is based upon the selected [2 3 8] segments and the nonlinear arrangement "$(A + B + C)^2 = A^2 + B^2 + C^2 + 2AB + 2AC + 2BC$" is represented. Although each of the cells in Figure 5.24 and Figure 5.26 responds preferentially to guitars, the wide variety of response profiles – both between techniques and between nonlinear arrangements and segment selections – is apparent. This again illustrates the lack of dependence upon specific cellular configurations required to achieve superior performance.

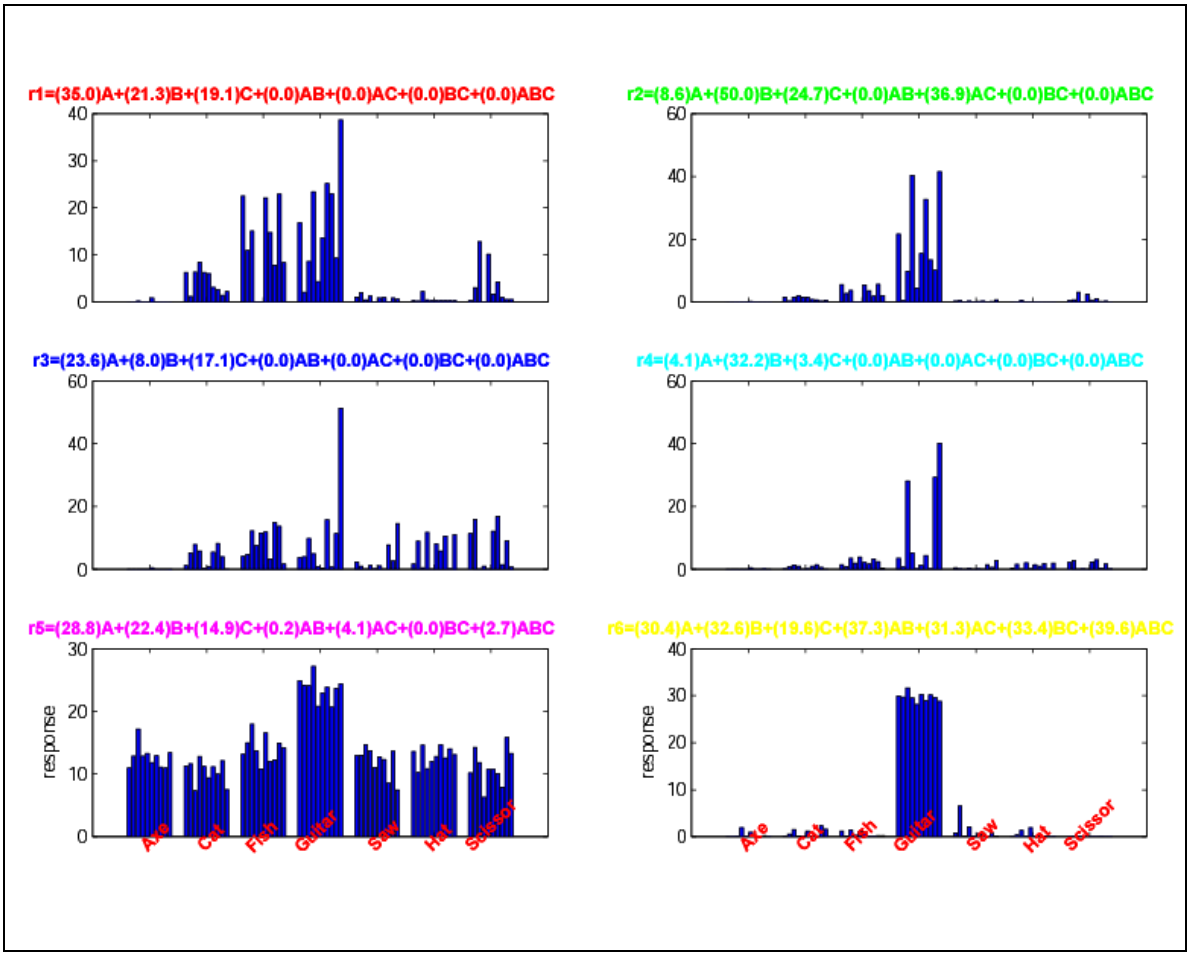

Figure 5.24 – Responses of six alternatively-constructed cells to each image. The 70 images are grouped into 7 categories in each of the 6 panels. The height of each bar in each panel represents the response of each cell to each of these images. Each cell's full response equation, including fitted coefficients and color-coded to match the construction methodology used for Figure 5.9, is shown above its panel. The mean and standard deviation of each of the cell's four features (angle, curvature, direction of curvature, distance) for three Gaussian constituents (A, B, C) are derived using different color-coded techniques (one for each cell): using the average of iso-curvature segments (the selected [1 2 3] segments for the case illustrated) from multiple images, using a single prototype image's iso-curvature segments, or entirely from data fitting (and optimized randomly). The desired response can be relative (as in the normalized EMD value of Figure 5.6) or absolute (20-30 Hz within-category, 1 Hz out-of-category). The nonlinear arrangement "A + B + C + AB + AC + BC + ABC" is represented. Each of these cells responds preferentially to guitars.

In Figure 5.25 we show the Gaussian constituent response contributions of one of the cells of Figure 5.24 to the ten images in the category to which the cell responds optimally ("guitar"). The components of the response are not surprising, since the cell's construction was based on the selected [1 2 3] iso-curvature segments. Figure 5.27 is similar to Figure 5.25, but with the Gaussian constituent response contributions of one of the cells of Figure 5.26 given to the ten images in the category to which the cell responds optimally ("guitar"). The components of the response are also not surprising, since the cell's construction was based on the selected [2 3 8] iso-curvature segments. Note that in each case, the Gaussian constituent response contributions are subsequently combined nonlinearly and multiplied by coefficients to achieve the total response of each of the IT-like cells. Figures 5.25 and 5.27 illustrate the fact that different images may activate the IT-like cells, and specifically their Gaussian constituents, differently, yet produce similar total responses. This type of cellular behavior is necessary to achieve the robust discriminatory power of our network. Consider, for example, the first (image 121) and seventh (image 127) images of Figure 5.27. The iso-curvature segments of image 121 contribute moderately to all three ("A", "B", "C") Gaussian constituents of cell 2, whereas the iso-curvature segments of image 127 contribute significantly to Gaussian constituent "B" only. However, both images 121 and 127 produce a total IT response close to 40 Hz (as seen in the top right panel of the Figure 5.26).

We continue our investigation of the iso-curvature segments' contributions to Gaussian constituents' responses in Figure 5.28, but in greater detail. The responses (representing

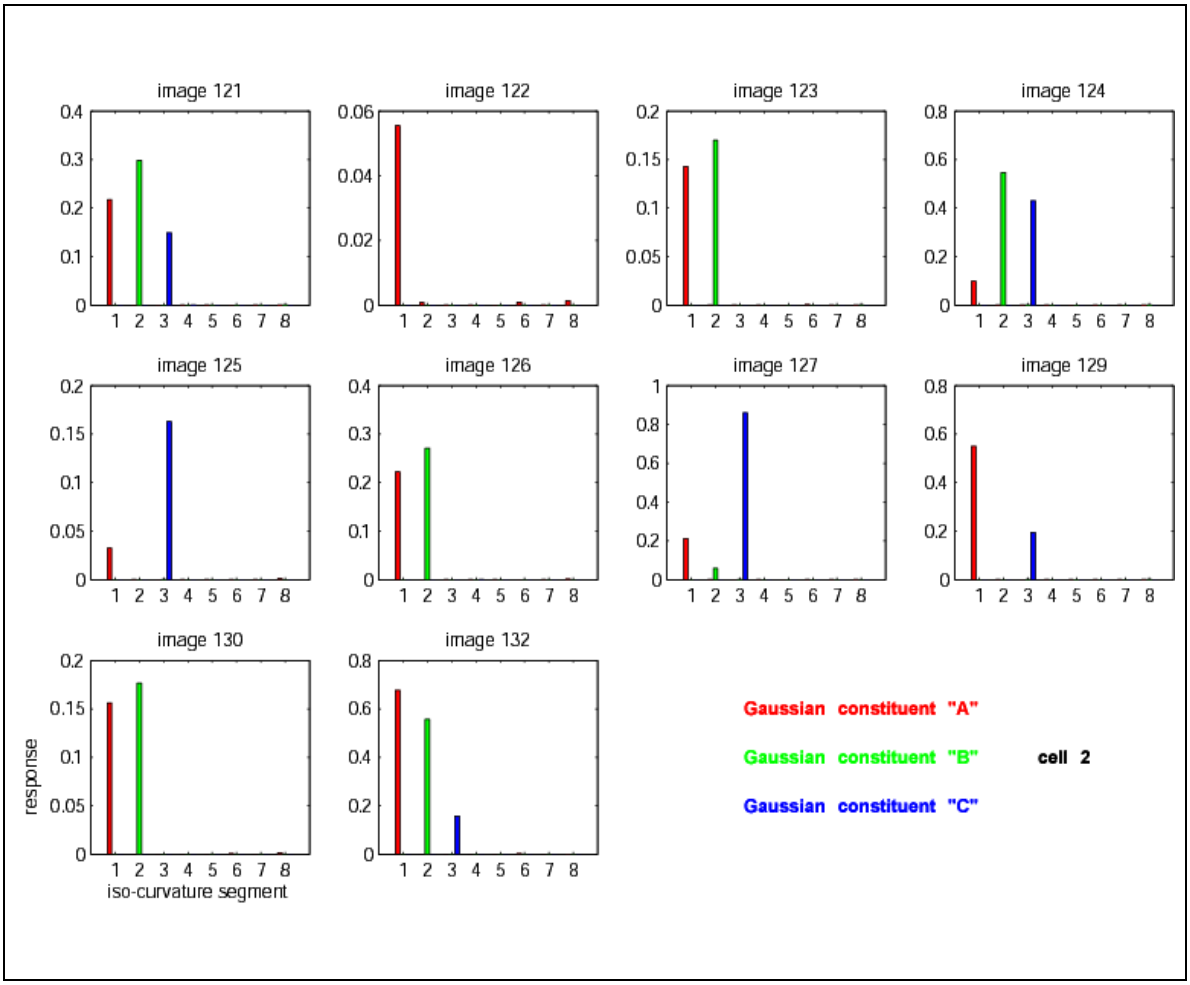

Figure 5.25 – Response contributions of one cell. The Gaussian constituent response contributions of one of the cells of Figure 5.24 (cell 2, in the top right panel of the figure, constructed using the average of the selected [1 2 3] iso-curvature segments) to the 10 images (with image numbers given) in the category to which the cell responds optimally ("guitar") is shown. Iso-curvature segment numbers are given along the abscissas. The height of each color-coded bar, with values on the ordinates, represents the corresponding Gaussian constituent's (A's, B's, or C's) contribution to the cell's total response to the image, with maximum value corresponding to the n-dimensional Gaussian peak at 1 for each constituent.

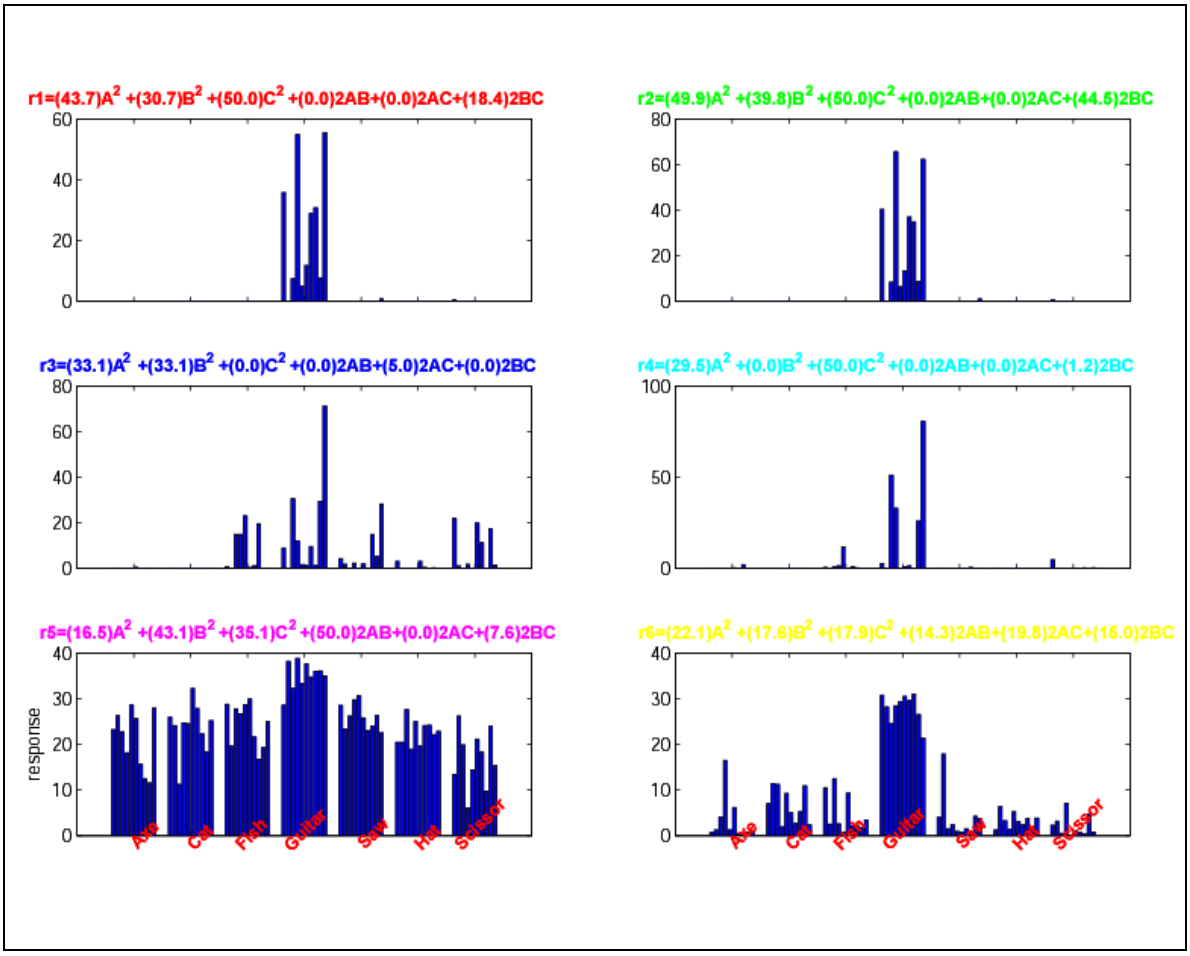

Figure 5.26 – Responses of six alternatively-constructed cells to each image. The 70 images are grouped into 7 categories in each of the 6 panels. The height of each bar in each panel represents the response of each cell to each of these images. Each cell's full response equation, including fitted coefficients and color-coded to match the construction methodology used for Figure 5.10, is shown above its panel. The mean and standard deviation of each of the cell's four features (angle, curvature, direction of curvature, distance) for three Gaussian constituents (A, B, C) are derived using different color-coded techniques (one for each cell): using the average of iso-curvature segments (the selected [2 3 8] segments for the case illustrated) from multiple images, using a single prototype image's iso-curvature segments, or entirely from data fitting (and optimized randomly). The desired response can be relative (as in the normalized EMD value of Figure 5.6) or absolute (20-30 Hz within-category, 1 Hz out-of-category). The nonlinear arrangement "$(A + B + C)^2 = A^2 + B^2 + C^2 + 2AB + 2AC + 2BC$" is represented. Each of these cells responds preferentially to guitars.

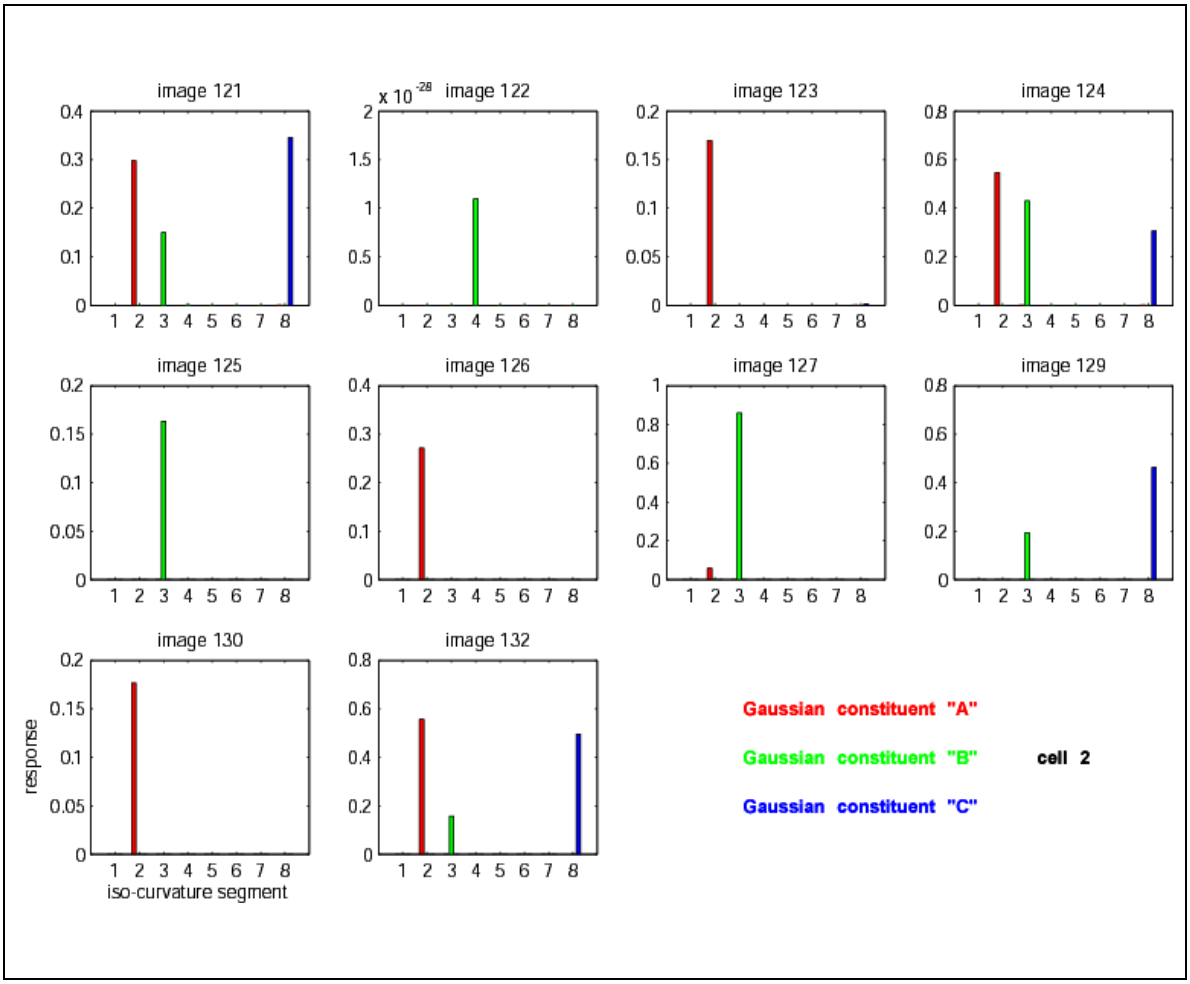

Figure 5.27 – Response contributions of one cell. The Gaussian constituent response contributions of one of the cells of Figure 5.26 (cell 2, in the top right panel of the figure, constructed using the average of the selected [2 3 8] iso-curvature segments) to the 10 images (with image numbers given) in the category to which the cell responds optimally ("guitar") is shown. Iso-curvature segment numbers are given along the abscissas. The height of each color-coded bar, with values on the ordinates, represents the corresponding Gaussian constituent's (A's, B's, or C's) contribution to the cell's total response to the image, with maximum value corresponding to the n-dimensional Gaussian peak at 1 for each constituent.

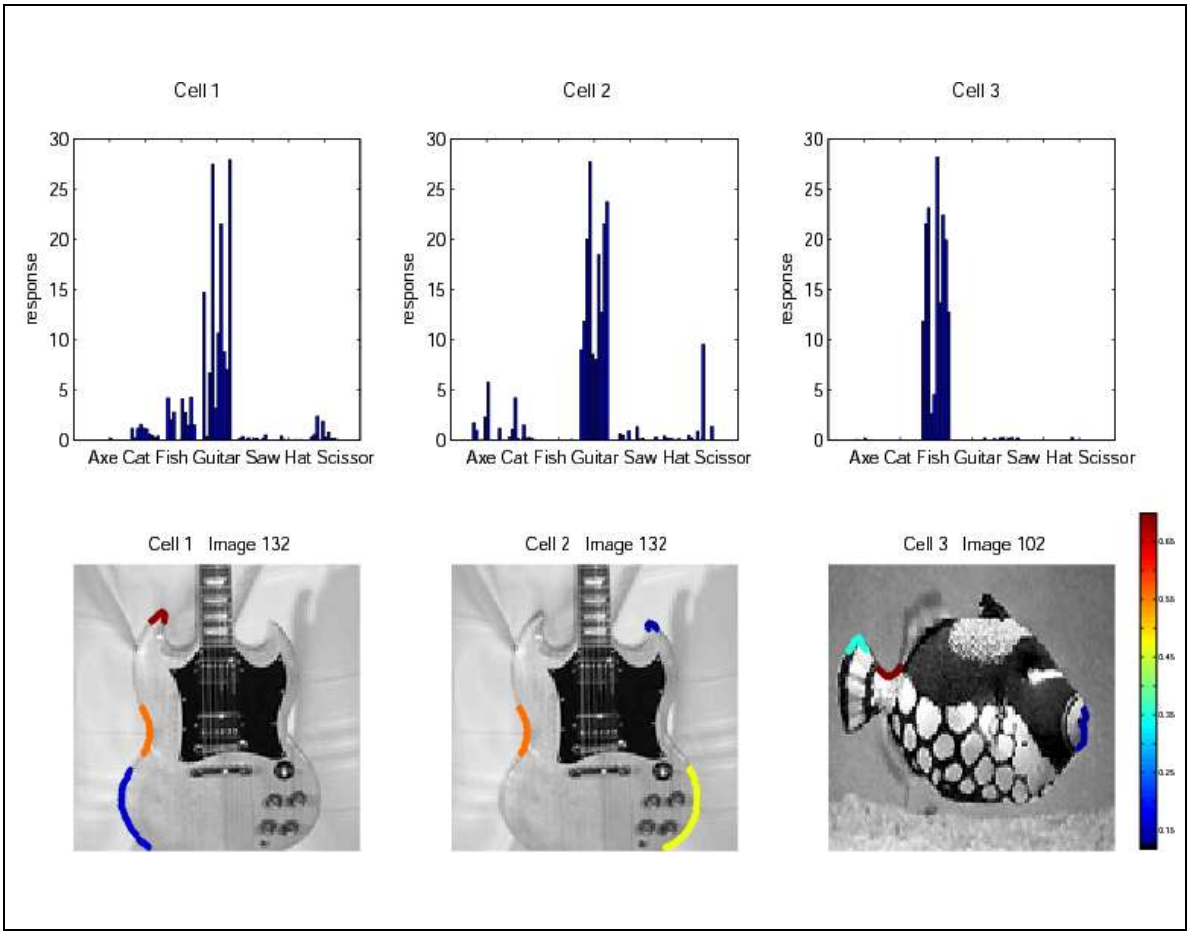

Figure 5.28 – Iso-curvature segments' contributions to Gaussian constituents' responses. The top row represents the responses of 3 cells from the population to each image. The 70 images are grouped into 7 categories in each of the 3 top panels. The height of each bar in each panel represents the response of each cell to each of these images. The left panel cell (tuned with the selected [1 2 3] segments) responds preferentially to guitars. The middle panel cell (tuned with the selected [2 4 6] segments) responds preferentially to guitars. The right panel cell (tuned with the selected [1 2 3] segments) responds preferentially to fish. The bottom row represents the contributions of the iso-curvature segments, singly or in combination, from each of these 3 particular images, to the overall responses of the corresponding cells. These responses are proportional to the alignment (proximity) of the iso-curvature segments with the cells' Gaussian constituents. Hotter colors (see the color bar) represent a larger degree of IT cell excitation.

IT cell excitation) of three cells from the population to each image are shown in the top row of this figure. The left panel cell (tuned with the selected [1 2 3] segments) responds preferentially to guitars. The middle panel cell (tuned with the selected [2 4 6] segments) responds preferentially to guitars. The right panel cell (tuned with the selected [1 2 3] segments) responds preferentially to fish. The bottom row of the figure represents the contributions of the iso-curvature segments, singly or in combination, from each of these three particular images, to the overall responses of the corresponding cells. These responses are proportional to the alignment (proximity) of the iso-curvature segments with the cells' Gaussian constituents.

In Figure 5.29 we consider sub-populations – subsets of the full IT cell population. The height of each bar in each panel is proportional to the average firing rate of all cells to all of the images in each of the categories. The left panel corresponds to a full, 100-cell population. The seven panels on the right correspond to seven (presumably different) 10-cell subsets of the full population, chosen to yield the "best" average response (i.e., the most differential response with the "cleanest" histogram) to each of the seven categories. Clearly, there are subsets of cells in the population that are more adept at certain categorizations.

Overall accuracy is investigated and visualized in Figure 5.30. This figure is similar to the right panels of Figure 5.29, but instead of selecting the ten "best" cells from the population, we select a variable number of cells, starting with the full 120-cell population (100%) and steadily decreasing. The percentage of "best" cells selected from the full

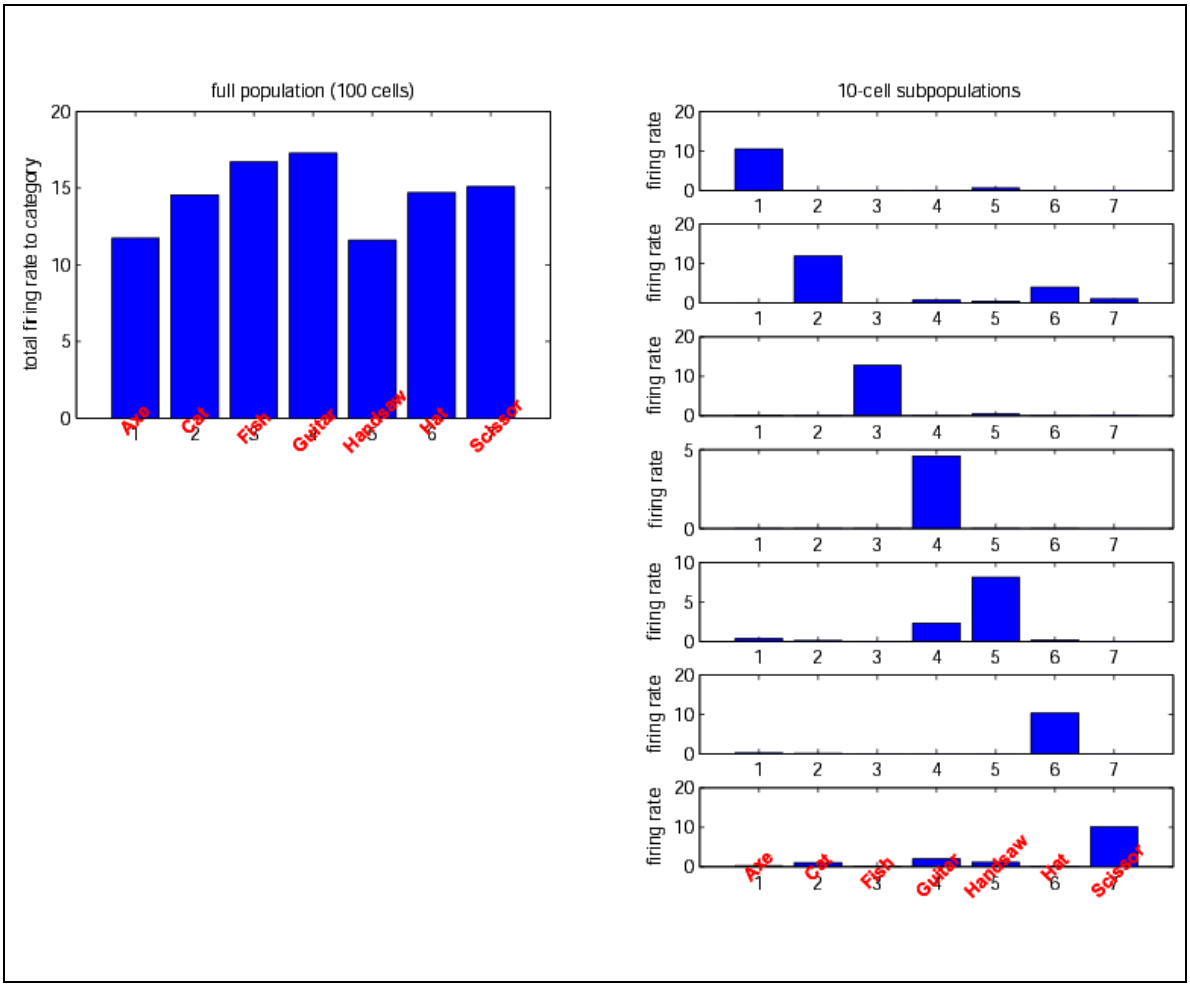

Figure 5.29 – Sub-populations. The 70 images are grouped into 7 categories in each of the panels. The height of each bar in each panel represents the average firing rate of all cells to all of the images in each of the categories. The left panel corresponds to a full, 100-cell population. The 7 panels on the right correspond to 7 (presumably different) 10-cell subsets of the full population, chosen to yield the "best" average response (i.e., the most differential response with the "cleanest" histogram) to each of the 7 categories.

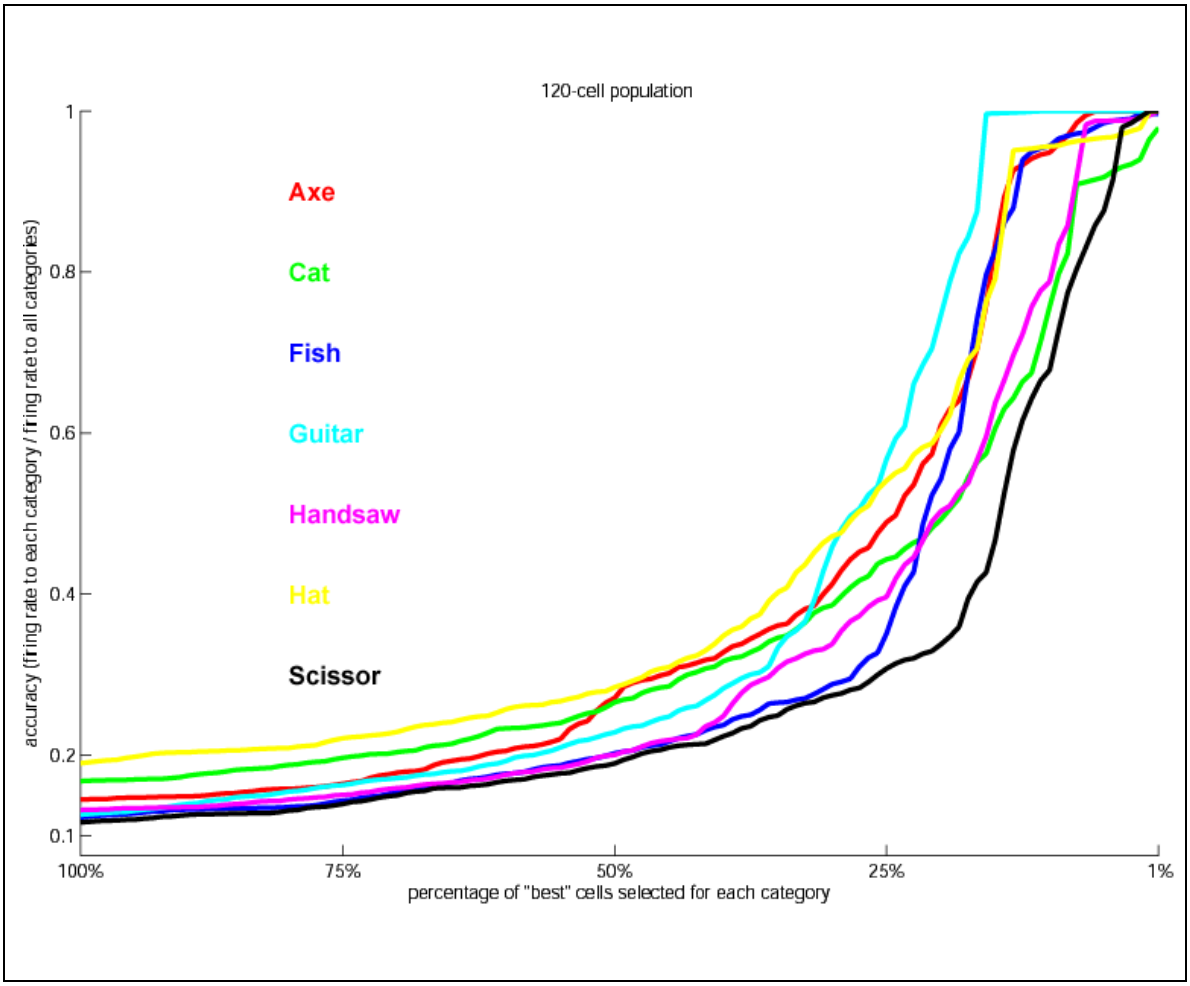

Figure 5.30 – Accuracy. The 70 images are grouped into 7 categories, coded by color. This figure is similar to the right panels of Figure 5.29, but instead of selecting the 10 "best" cells from the population, we select a variable number of cells, starting with a full 120-cell population (100%) and steadily decreasing. The percentage of the "best" cells selected from the full population for each category is given on the abscissa. These subsets of the full population are chosen to yield the "best" average response (i.e., the most differential response) to each of the 7 categories. The accuracy, with values on the ordinate, is defined as the average firing rate to the category divided by the sum of firing rates to all categories for the fraction of cells currently selected for each specific category (the average response to all images in 1 category divided by the sum of average responses to all images in all categories). A value of 1 represents 100% accuracy – the selected cells respond only to images from the category for which they were selected.

population for each category is given on the abscissa. These subsets of the full population are chosen to yield the "best" average response (i.e., the most differential response) to each of the seven categories. The accuracy, with values on the ordinate, is defined as the average firing rate to the category divided by the sum of firing rates to all categories for the fraction of cells currently selected for each specific category (the average response to all images in one category divided by the sum of average responses to all images in all categories). A value of 1 represents 100% accuracy – the selected cells respond only to images from the category for which they were selected. This is similar to a receiver operating characteristic (ROC) curve, which is a plot of the false alarms (false positive rate) on the abscissa against the hits (true positive rate) on the ordinate for a criterion range (Zweig and Campbell, 1993). Here, the hit rate is equivalent to sensitivity. We conclude that only a small number of V4 / IT cells may be necessary for image recognition at this level. Assuming that one cell for each image category is sufficient for correct categorizations, we might ask why IT would have more than one cell representing each category. This is a fair question, considering the biological overhead of maintaining a large population of cells. The answer, however, is nearly identical to that of the "grandmother" cell issue. In order for the population to remain robust in the face of cellular demise and to accommodate the wide variation of category members (perhaps much more varied than in our test samples), a greater number of cells per category would be required. However, as we have demonstrated here, this number is not prohibitively large.

5.4 Discussion

The information concentration along contours, and therefore the importance of contour shape and curvature in both human and computer vision, has been apparent since Attneave's (1954) seminal work relating information theory to visual perception. The discoveries of Pasupathy and Connor (2001), Hegdé and Van Essen (2003), and others show a large degree of sophisticated shape processing in extrastriate visual cortex. These, along with the results of Brincat and Connor (2004, 2006) in the inferotemporal cortex, reveal further refinement in the ventral stream and represent, to a large degree, a biological solution to shape description and object recognition. Our results suggest that curvature- and position-sensitive units can function as robust shape descriptors. Our model of highly selective IT-like cell response, a function of the number of V4-like cell inputs and their nonlinear combinations, is hardy in that it is not rigidly dependent upon parameter selection or implementation strategy, yet it performs consistently well in recognition tasks.

Several aspects of the Brincat and Connor (2004, 2006) research regarding neural selectivity for complex 2-D boundary shape in inferotemporal cortex warrant additional consideration. They have found, for example, that the IT cell tuning functions are composed of 1–6 Gaussians (and typically 2–4) on the shape × position domain, with neurons integrating specific information about the shapes of multiple contour fragments. In our model, these units correspond to iso-curvature segments activating V4 inputs to

Gaussian constituents in the total response equations of the IT-like cells. They may, however, also correspond to different kinds of entities, as in a fragment-based approach (Ullman *et al.*, 2001) or a components-based approach (Biederman, 1987) to recognition. The units / iso-curvature segments / Gaussian constituents are not required to be contiguous, as in the clockwise and counterclockwise segments contributing to V4 responses (Pasupathy and Connor, 1999). The fact that V4 cells are further modulated by contour configurations at neighboring locations does not preclude modulation in V4 by non-contiguous segments. It may be the case that only sequences of three boundary elements were measured in V4 and so the non-contiguous response contributions in IT would not be surprising. Their main finding of explicit coding of the structural relationships between boundary fragments based on a graded tuning for shape and position does offer support for any parts-based shape representation, including our model, as does the fact that 2-dimensional boundary shape may dominate IT response to realistic objects (Kovács *et al.*, 2003).

Brincat and Connor (2004) have also found that the integration of excitatory contour elements (the positive terms in our total response equations) was by purely linear as well as nonlinear neuronal sub-populations, whereas the integration of inhibitory contour elements (the negative terms in our total response equations) was almost exclusively linear. We enforce this aspect in our models and nonlinear least-squares curve fitting works well. We realize that the derivation of the tuning functions would be impossible in this manner without the benefit of observed responses. Also, the nonlinear interactions would be problematic using other construction methods, such as the linear principal

components analysis. Their finding that relative responses were consistent across changes in stimulus position and size gives support to the affine transformational aspects of our model. Their finding that tuning for retinotopic position was broadest and weakest leads us to conclude that the 6-dimensional shape space (which includes retinotopic position) could be reduced to a 4-dimensional space (curvature, orientation, position relative to the object's center of mass).

The successful model of episodic recognition memory by Kahana and colleagues (Kahana and Sekuler, 2002) compels us to explore the connections between our model of object recognition and the current models of mathematical psychology. Prototype-based theories of categorization typically employ one image from each category. Exemplar-based theories like ours, on the other hand, typically require several images (in our case, ten) from each category. Kahana's noisy exemplar model represents stimuli as multivariate normal distributions in feature space. He employs a deterministic response rule, based upon the summed similarity between probes and stored representations, with a consideration of interstimulus similarity. Categorizations are based on an average fit approximation to the class, not on exact matches to prototypes. Other results suggest that humans categorize by comparing objects to well-known members of alternative categories, either directly or based on class boundaries (as in SVM), and learn which features are most diagnostic for distinguishing particular categories (Sigala *et al.*, 2002; Sigala and Logothetis, 2002; Dosher and Lu, 1998). Dynamic filter weighting for emphasis (categorization and sub-categorization) might be used to accomplish this in our computational model (see our previous results in Chapter 4). Palmeri and Nosofsky

(2001) have found evidence of extreme prototype enhancement, in which category prototypes were classified more accurately than any of the other category examples. This would seem to favor a prototype / template-based methodology. However, they have found these category prototypes to be more accurately seen as psychological extreme points relative to the categories rather than central tendencies of category instances, further supporting exemplar-based theories of categorization.

By some standards, our recognition tasks have been "easy". Our categorization possibilities (one of ten digits, as in our previous work (Chapter 4), or one of seven categories of real images, for instance) are modest. More difficult (more realistic, perhaps, yet less controlled) evaluations, as when information regarding the number of categories is incomplete or unknown, might have required a greater degree of unsupervised learning or clustering such as k-means clustering or the use of a Gaussian Mixture Model (GMM) coupled with the expectation-maximization (EM) algorithm (Duda and Hart, 1973; Bishop, 2006). Additionally, the "class imbalance" issue, with many exemplars representing some categories and very few exemplars representing other categories, might have to be addressed (Mazurowski *et al.*, 2008; Tang *et al.*, 2009; Liu *et al.*, 2009). More generally, the recognition aspects of similarity and concept learning within a Bayesian inference framework, with particular emphasis on the additive clustering approach to extracting features that best account for similarity judgments on a given set of objects, might be considered (Tenenbaum and Griffiths, 2001).

The nonlinear combinations of V4 cell inputs to IT could be generalized by an equation of the form:

$$\alpha A + \beta B + \gamma C + ... + \delta AB + \varepsilon AC + \eta BC + \xi ABC + \ldots .$$

This might support a Bayesian / probabilistic implementation of recognition with, for instance, the joint and conditional probabilities (as in: segment A and segment B are coincident with some probability) related to the nonlinear equations. In a sense, IT would be computing

$$p (\text{image} \mid \text{features}) ,$$

with prior

$$p (\text{image}) / p (\neg \text{image})$$

and likelihood

$$p (\text{features} \mid \text{image}) / p (\text{features} \mid \neg \text{image}) .$$

This is somewhat similar to Gold's work in monkey LIP with Bayesian temporal sequence decoding using likelihood ratios and the subsequent addition of logarithms, given some feature evidence (Gold and Shadlen, 2000; Gold and Shadlen, 2001).

As mentioned above, our Gaussian constituents (A, B, C) were chosen by selecting a single prototype, computing the category average, selecting and computing the average for a subset of images from a category (e.g., three guitars only), etc. We have compared the coefficients in each case and have found predictable patterns. We consider the requirement that a population of Gaussian constituents be derived directly from and aligned roughly with real image iso-curvature segments. The Gaussian constituents, in this case, could be determined using a factorial approach – randomly derived (and optimized). The tuning functions might even be pruned after some initial exposure to images, as in the work by Amit and colleagues (Amit *et al.*, 1997). Also, the effect that the inter-category and intra-category distances between the image features have on the nonlinear terms of the IT-like cells' response equations should be considered. When using the normalized average-to-average Earth Mover's Distances to create the tuning functions, for instance, a smaller contribution from the nonlinear terms might be required. Tuning functions derived using an absolute response ratio might require a greater nonlinear contribution to achieve optimization.

Pasupathy and Connor (2001) point out that complex shape representation in area V4 is parts-based (since contour segments are defined by conformation and position) as well as distributed (since individual cells encode smaller parts of larger objects). The same can be said for IT to a larger degree. We consider the relationship between the size of a part and the size of an iso-curvature segment and question whether this relationship is time-varying. It is thought that a parts-based coding system, using either a finite number of

primitives or a continuous part representation with graded tuning, has the combinatorial power and representational capacity to encode a virtually infinite variety of objects (Pasupathy and Connor, 2002; Biederman, 1987; Tsunoda *et al.*, 2001; Rolls *et al.*, 1997). The number of cells required for these types of representations is well within biological constraints (Van Essen, 2003; Bullier, 2001; Motter, 2003). Any parts-based model, including ours, must necessarily incorporate some element of reconstruction of the parts. This might, for example, be related to boosting or ensemble learning, where a very accurate prediction rule is produced by combining rough and moderately inaccurate rules (Freund and Schapire, 1999; Viola and Jones, 2001a). Alternatively, it might be implemented using something similar to a "mixture of experts", perhaps implementing some form of Bayesian inference related to Mumford's framework, with recurrent feedforward and feedback loops integrating top-down context and bottom-up stimulation (Lee and Mumford, 2003). In cortex, the hypercolumns themselves might possibly implement some form of Bayesian inference (Weiss, 1997). The work by Rao (2005a) explains a variety of attention-related responses in V4 by interpreting visual attention as the cortical mechanism for reducing perceptual uncertainty as it integrates top-down and bottom-up information. He has also generated purely theoretical models where membrane potentials encode some probability function (Rao, 2004; Rao, 2005b). General support for the view that the brain uses Bayesian-like computations – particularly the multiplication of prior knowledge with new and possibly uncertain evidence – is provided by models such as that proposed by Simoncelli (Stocker and Simoncelli, 2006). These ideas are related to those of Yu and Dayan (2002), who consider acetylcholine's role in perceptual inference to be that of modulator between top-down contextual

information and bottom-up sensory inputs by determining the relative strengths of these sources. In a system where top-down input indicates what should be expected horizontally, acetylcholine's function, demonstrated with a hierarchical hidden Markov model (HMM), is essentially to reflect the uncertainty associated with top-down information.

An important issue concerns whether and how a global description of object shape emerges. A requirement of a parts-based description might be that higher visual areas learn to be selective to differences, at a variety of hierarchical levels, which distinguish categories of objects. Logothetis and colleagues (Sigala and Logothetis, 2002) have found enhanced neuronal representation in primate IT cortex to features that are more diagnostic for distinguishing particular categories. The separability of their stimulus space suggests that higher visual areas may learn to be selective to variations which discriminate between object categories. (Consider our efforts to better visualize our cell response space in the three-dimensional non-classical non-metric multidimensional scaling analysis domain and the subsequent demonstration of separability.) Top-down effects (from IT, etc.) might then alter the gain or sensitivity of particular intermediate-level units (in V4) and focus their response onto those regions of interest or features whose local configurations are critical for the class distinction (Hochstein and Ahissar, 2002; Sigala and Logothetis, 2002).

An initial fast-pass recognition (shape class) might be plausible based merely on the distribution of V4 cells activated. It is not unreasonable to consider the following

scenarios and functional architecture: IT receives local curvature information ("reasonable" regions, not individual curvatures) from V4. Higher visual areas, including IT, contain differentially-activated exemplars (not prototypes) – category cells, not grandmother cells – and may be selective to global or topological differences between objects which distinguish categories. These topological properties may be exploited within the feedback from IT to V4, providing top-down adjustment of regions / segments, including possible re-segmentation based upon top-down impositions, thus improving classification accuracy and adding a global aspect to the recognition. This is related to both recognition by geons (Biederman, 1987) and to a class-based segmentation method, guided by a stored representation of the shape of objects within a general class, with emphasis on the role of high-level class-specific decision criteria (Borenstein and Ullman, 2002). It clearly suggests a recurrent solution, perhaps at each stage, without concern of sacrificing speed for accuracy, involving iterative signal refinement via feedback from local as well as higher cortical areas, rather than a selective feedforward convergence.

Hypothetically, horizontal connections could embody local recurrent processing or lateral inhibition – by other cells' Gaussian constituents or by other categories' iso-curvature segments – in some on-center, off-surround fashion, for example. Horizontal connections in V4 could also facilitate global matching between curvature detectors and provide rudimentary local contour grouping. This represents an extension and use of Elder and Goldberg's research to top-down / global computations, with contour grouping equivalent to the recovery of sequences of tangents (Elder and Goldberg, 2002). Confirmations of

spatial relationships and figure directions, adjustments to match the "preferred" contour path, and local (contour matching) and global (interaction with other V4 cells) error minimization would also be present. This is consistent with the ideas that border ownership (the side to which a border in a figure belongs) and information about how local features belong to objects each represent global image context integration and are generated within the visual cortex, not projected down from higher levels (Zhou *et al.*, 2000). These may, however, restrict the receptive fields to a particular spatial scale.

Alternatively, we could represent the shapes (curvature, direction of curvature, etc.) in our model with the responses of steerable curvature filters. This is certainly a viable approach and is roughly equivalent to a distributed representation, but in reality is simply a different type of mathematical description. An obvious advantage is the elimination of the initial segmentation requirement. In a sense, the curvature filters are the image segments. The filters could be constructed from Gabor or Difference of Gaussian filters, with morphing and conformal mapping applied. An initial step would be to determine the number of filters and their degrees of coarseness or refinement (i.e., their scales). Also, systematic generation, as opposed to some learning paradigm applied to real-world images, could be explored.

From a primarily machine vision-oriented perspective, another viable approach would again focus on the V4-like model cell populations that respond to both curvature and global position on the object and the IT-like model cells that combine these responses linearly and nonlinearly. In some sense, once a closed curve is labeled with its V4-like

units, there exists a unique description that characterizes (recognizes or defines) the object. If, at first, the contour were only labeled with curvature (not location) then Ullman's idea of recognition by fragments could be used. Here, fragments (iso-curvature segments) are the component building blocks (parts) used to represent a large variety of objects belonging to a common class (Ullman, *et al.*, 2001; Riesenhuber and Poggio, 1999b). The contour could be divided into small snippets containing several iso-curvature segments each. This could be done in several different ways to get varied options (similar to attempting to recognize a DNA sequence from small fragments). Based on the population of snippets, a recognition stage with a fast matching algorithm would determine the leading candidates for shape. The stored "memorized" candidate objects would have complete information, with each curvature associated with a position. There would be competition between the alternative global shape choices. This would then become an iterative process – fixing as many snippets as possible with positions, determining the global recognition candidates, and feeding them back to the V4 layer.

In this approach, the figure / ground information isn't integrated or propagated locally from low-level contour cues. Rather, it is derived in a feedback manner from recognition. This is consistent with high-level cues and recognition affecting grouping cues. It is also consistent with the fact that the bottom-up / top-down pathways are much faster (in terms of conduction velocity) than the horizontal integration. The model could also contain a Bayesian element, similar to Adelson's likelihood of motion fuzzing-out (Adelson and Bergen, 1985), etc., in that a curvature and position on the object could have certain likelihoods, sharpening as the iterations proceeded.

5.5 Conclusion

Our results suggest that curvature- and position-sensitive units, as described by Brincat and Connor in IT, can function as robust shape descriptors. We have demonstrated the utility of cells with response properties similar to those found in V4 and IT by constructing computer network models and successfully subjecting them to artificial recognition tasks on a set of real images. Our model of highly selective IT-like cell response, a function of the number of V4-like cell inputs and their nonlinear combinations, is hardy in that it is not rigidly dependent upon parameter selection or implementation strategy, yet it performs consistently well in recognition tasks. We claim that the response properties of V4 and IT cells (i.e., their receptive fields), and in particular their sensitivities to curvatures and contour positions, are useful for object recognition precisely because they, like our faithful computer models of them, facilitate shape representation and categorization by extracting features that correlate with global shape. We have established a connection between a computer model of a recognition system and known cortical mechanisms within a biologically realistic network architecture.

Chapter 6

Biologically Plausible Models and Synchronization in the Visual Area 4 (V4) – Inferotemporal Cortex (IT) Circuit

6.1 Introduction

Since the cortical mechanisms underlying shape analysis and object recognition are incompletely understood, explicating the representation of contour shape, an essential component of object recognition, remains a fundamental open question in neuroscience. In monkey extrastriate visual area 4 (V4), an intermediate stage in the ventral (shape recognition) pathway extending from primary visual cortex (V1) to inferotemporal cortex (IT), Connor and colleagues have described cells that are selective for a particular local shape configuration at a particular location on a contour within a larger shape (Pasupathy

and Connor, 2001). They and other research groups have also seen neural cells with selectivity for complex 2-dimensional boundary shape (perhaps the kind that actually dominates responses to realistic objects (Kovács *et al.*, 2003)) in macaque inferotemporal cortex (TEO / PIT and posterior TE / CIT) (Brincat and Connor, 2006; Brincat and Connor, 2004; Freedman *et al.*, 2003; Baker *et al.*, 2002; Tsunoda *et al.*, 2001; Op de Beeck *et al.*, 2001; Booth and Rolls, 1998; Rolls *et al.*, 1997; Gallant *et al.*, 1996; Logothetis *et al.*, 1995; Kobatake and Tanaka, 1994; Fujita *et al.*, 1992; Young, 1992; Felleman and Van Essen, 1991; Gross *et al.*, 1972), finding that IT neurons integrate specific information, such as curvatures, orientations, and relative positions, about the shapes of multiple contour fragments (typically 2 – 4). Their tuning functions on the shape × position domain are driven by inputs from V4 cells.

Our earlier results (Chapter 4) support the hypothesis that curvature- and position-sensitive V4 cells – evaluated against the standard MNIST database of handwritten digits and the MPEG-7 Shape Silhouette database (Jeannin and Bober, 1999) – function as robust shape descriptors in the early stages of object recognition, with shape categorizations based on a particular local contour conformation located at a specific position on the object's boundary. Later results (Chapter 5) suggest that curvature- and position-sensitive units, as described by Brincat and Connor in IT, can also function as robust shape descriptors. We developed a robust model of highly selective IT-like cell response, a function of the number of V4-like cell inputs and their nonlinear combinations, that performs consistently well in artificial recognition tasks on a set of real images across a wide range of specific parameter selections and implementation

strategies. We made the claim that the response properties of V4 and IT cells (i.e., their receptive fields), and in particular their sensitivities to curvatures and contour positions, are useful for object recognition precisely because they facilitate shape representation and categorization by extracting features that correlate with global shape.

Here, we continue our investigation of how contour shape is represented in cortex and again ask the fundamental questions: "How is contour shape represented in cortex and how can neural models and computer vision algorithms more closely approximate this?" We wish to approach a neurobiological understanding, useful theoretically as well as in developing or improving computer vision and Brain-Computer Interface (BCI) methodologies and applications, as to how object recognition, and the underlying analysis of shape, is accomplished. We will again consider why the response properties of V4 and IT cells are useful. In doing so, we hope to establish a clear connection between a computer model of a recognition system and known cortical constructs within a biologically realistic network architecture.

Specifically, we concentrate on realistic biological models of cells – those that have a high biological plausibility (measured in the number of biological features that the model can reliably reproduce) without burdensome implementation costs (measured in the number of floating point operations per second required to execute the model for some relevant period of time) – within our network. See Izhikevich (2004) for comparisons of some well-known models. We also consider the interaction between V4 and IT from a mechanistic perspective.

To begin, we use the Morris-Lecar neuronal model, with genetic algorithms utilized for parameter fitting, to more realistically illustrate the previously explored shape representation pathway in V4 – IT while remaining faithful to IT cell response patterns.

In some neural systems, such as the antennal lobe of the locust, information is represented through both space and time coding (Bazhenov *et al.*, 2001a; Bazhenov *et al.*, 2001b; Laurent *et al.*, 2001), with each stimulus (odor) evoking a specific, reproducible pattern of activity across a set of neurons that acts as a dynamic attractor in phase space and is robust in the presence of noise. In response to this fact, a class of dynamical systems called competitive networks or winnerless competition (WLC) was introduced to produce spatiotemporal coding (Rabinovich *et al.*, 2001). It was demonstrated that olfactory networks could recognize patterns using a WLC strategy and exhibit transformations from sensory input to spatiotemporal output in a network with a large capacity ($\approx e\,(N-1)\,!$), where N is the number of neurons in the network. As an aside, we consider these facts and demonstrate biologically-based object recognition using spatiotemporal patterns within a self-organized winnerless competition neural network with FitzHugh-Nagumo model neurons.

Cortical synchronization can function as a binding mechanism for perceptual organization and grouping (Finkel *et al*., 1998; Yen *et al*., 1999). Synchronous oscillatory activity in the gamma band is a fundamental process ideally suited for many cognitive functions (Fries, 2009) and the distortion of gamma activity is linked to object

recognition abnormalities (Lazarewicz *et al*., 2009). It has been associated with consciousness (Engel *et al*., 2001), attention (Vidal *et al*., 2006), memory (Howard *et al*., 2003; Sederberg *et al*., 2003; Miltner *et al*., 1999), synaptic plasticity and learning (Popescu *et al*., 2009), object representation and perception (Rodriguez *et al*., 1999) and the precise temporal relationships of concurrent stimuli (Tallon-Baudry *et al*., 1999).

Stimulus-specific synchronous gamma (40 Hz) oscillations, correlates of specific visual scene-induced network states, have been observed across cortical columns in cat visual cortex (Gray and Singer, 1989; Gray *et al*., 1989). In addition, it is likely that oscillatory activity (Giocomo and Hasselmo, 2008), particularly in the gamma (30 – 80 Hz) band, possibly utilizing reciprocal information transfer (Supp *et al*., 2007), coupled with inhibitory activity (Wang *et al*., 2000), is required for the information binding and object recognition that is localized in the monkey visual area 4 (V4) – inferotemporal cortex (IT) feedforward – feedback loop circuit (Kriegeskorte *et al*., 2008; Deco and Rolls, 2004; Ungerleider *et al*., 2008).

We conclude our current investigation of the cortical mechanisms of object recognition in areas V4 and IT with an examination of gamma synchronization in the V4 – IT circuit. We use the Izhikevich neuronal model and demonstrate that an initially out-of-phase network's inherent characteristics and dynamics can induce synchronized responses in V4 via PING (pyramidal interneuron network gamma) mechanisms (Whittington *et al*., 2000), involving both inhibitory and excitatory IT cells, by applying current input to the network. Additionally, we show that a response amplification in IT, correlated with

recognition, results from the synchronized spiking in V4 and roughly coincides with the onset of synchronization.

6.2 Methodology

In general, we use neurobiological models that have been well established in the literature. These are simple and reduced, yet remain faithful to observed phenomena *in vivo*.

6.2.1 The Morris-Lecar Model and Recognition

The Morris-Lecar equations represent a two-dimensional conductance-based reduced excitation neuronal model (Morris and Lecar, 1981; Fall and Keizer, 2002; Rinzel and Ermentrout, 1998). The model involves only a fast activating Ca^{2+} current, a delayed rectifier K^+ current and a passive leak current. The principal equations are:

$$C\, dV/dt = -g_{Ca}\, m_{\infty} (V - V_{Ca}) - g_K\, w (V - V_K) - g_L (V - V_L) + I_{app}$$

$$dw/dt = \varphi (w_{\infty} - w) / \tau$$

where m is the fraction of voltage-dependent Ca^{2+} channels open, w is the fraction of open channels for the delayed rectifier K^+ channels and g_L, g_{Ca} and g_K are conductances for the leak, Ca^{2+} and K^+ currents. The functions:

$$m_\infty = 0.5 [1 + \tanh ((V - v_1) / v_2)]$$

$$w_\infty = 0.5 [1 + \tanh ((V - v_3) / v_4)]$$

$$\tau = 1 / \cosh ((V - v_3) / (2 \cdot v_4))$$

are the equilibrium open fractions for the Ca^{2+} and K^+ currents and the activation time constant for the delayed rectifier. Synaptic transmission between neurons is modeled using two-state channels. Depolarization in the presynaptic cell opens voltage-gated calcium channels, resulting in calcium influx, transmitter release, postsynaptic receptor binding and postsynaptic channel opening. The current through the postsynaptic membrane is:

$$I_{syn} = g_{syn} s (V - V_{syn})$$

where

$$ds/dt = \alpha s_\infty (1 - s) - \beta$$

with

$$s_\infty = 1 / [\, 1 + \exp (- (V - \theta_{syn}) / k_{syn}) \,]$$

and where g_{syn} is the maximal conductance at the synapse, s is the open fraction, V_{syn} is the reversal potential, the presynaptic potential is above θ_{syn} and k_{syn} is positive.

Models of Type I oscillator cells can produce arbitrarily low frequencies of oscillations, resulting from saddle-node bifurcations. With the parameter values given in Table 6.1, the Morris-Lecar model can behave in this manner.

We create a Morris-Lecar IT cell model which integrates specific information about the 2-dimensional boundary shapes of multiple contour fragments (V4 cell inputs) with tuning functions on the shape × position domain (Brincat and Connor, 2004; Brincat and Connor, 2006). Each cell behaves like a previously defined IT-like cell in Chapter 5, with three Gaussian constituents (A, B, C) in its total response equation, each receiving inputs directly from the V4-like cells (responding to an image's iso-curvature segments) and together used to compute the nonlinear response of the IT cell. The Gaussian constituents are essentially variables in the total response equations of the IT-like cells. Each Gaussian function is n-dimensional (where "n" is the number of features considered). Any of the image's iso-curvature segments can contribute to the Gaussian constituent function's response, with response magnitude proportional to the distance between the "n" features of the iso-curvature segment and the center of the n-dimensional

parameter	value
C	20 μF/cm^2
V_K	− 84 mV
g_K	8 mS/cm^2
V_{Ca}	120 mV
g_{Ca}	4 mS/cm^2
V_L	− 60 mV
g_L	2 mS/cm^2
v_1	− 1.2 mV
v_2	18 mV
v_3	12 mV
v_4	17.4 mV
φ	0.066 / ms
V_{syn}	100 mV
θ_{syn}	20 mV
k_{syn}	2
α	1
β	0.3

Table 6.1 – Morris-Lecar model parameters.

Gaussian constituent. (See Chapter 4 and Chapter 5 for details about iso-curvature segments, Gaussian constituents of response equations, etc.) For a maximal response, the image's features (in the case of four features: the angle, curvature, direction of curvature and distance of each of the iso-curvature segments) would have to be perfectly aligned with those of the Gaussian constituents. Otherwise, a sub-optimal response would result, determined by Gaussian falloffs from the means at rates proportional to the specified standard deviations. The mean and standard deviation of each Gaussian constituent's 4-feature vector are selected from a previously created population of cells (in Chapter 5). The coefficients (for the linear and nonlinear components A, B, C, AB, AC, BC, and ABC) are selected in a similar manner. Responses are chosen to be either constant high for the active category and constant low for all other categories (an absolute ratio). All terms in the total response equation are considered to be current inputs into the Morris-Lecar cell.

From the natural image dataset – in the form of 360 × 360 JPEG images – kindly supplied by Drs. Kanwisher and Grill-Spector (Grill-Spector and Kanwisher, 2005), we select images belonging to one of seven different categories (axes, cats, fish, guitars, handsaws, hats, and scissors), with ten randomly selected samples from each category. Note that these same natural images were utilized in Chapter 5. With these images as inputs, we attempt to find a set of synaptic conductances that would, given the pre-defined Gaussian constituent nonlinear integration response equations, produce the optimally matching outputs. Since our objective function is non-smooth, traditional derivative-based optimization methods are not effective. Instead, we use a genetic

algorithm (as found in MATLAB's Genetic Algorithm and Direct Search Toolbox) (Goldberg, 1989) to find the minimum of our objective function, thereby optimizing our choice of conductance parameters.

6.2.2 The FitzHugh-Nagumo Model, Spatiotemporal Patterns and Winnerless Competition

Following the Rabinovich methodology (Rabinovich *et al.*, 2001) and without the requirement of closed loops in the network, we implement a winnerless competition network of the two-dynamical-variable FitzHugh-Nagumo model spiking neurons (FitzHugh, 1961; Rinzel and Ermentrout, 1998). In phase space, the network's dynamics are characterized by heteroclinic orbits connecting fixed point or limit cycle saddle regions. These saddle states represent the activity of specific neurons. The separatrices connecting them correspond to sequential switching between the states. The inhibitory interactions are:

$$\tau_1 \, dx_i(t) / dt = f[\, x_i(t) \,] - y_i(t) - z_i(t) \, [\, x_i(t) - v \,] + 0.35 + S_i$$

$$dy_i(t) / dt = x_i(t) - b \, y_i(t) + a$$

$$\tau_2 \, dz_i(t) / dt = \sum_j g_{ji} \, G[\, x_j(t) \,] - z_i(t)$$

where $z_i(t)$ is a synaptic current, $x_i(t)$ is the membrane potential, $y_i(t)$ is the recovery variable, $f(x) = x - 1/3\, x^3$ is the internal FitzHugh-Nagumo nonlinearity, G(x) is a step function, S_i is the stimulus and g_{ji} is the strength of the synaptic inhibition.

With the other parameter values given in Table 6.2, a FitzHugh-Nagumo WLC network can produce considerably different patterns in response to different stimuli. Its high dimensionality provides the capacity to store many patterns.

We create such a FitzHugh-Nagumo WLC network with nine cells ($i = 1, 2, \ldots 9$ in the above equations) and subject it to a population of simple geometric shape images (specifically, families of curves) – circles, cardioids, limaçons of Pascal, ellipses and ovals of Cassini with governing equations:

circles: $r = a$

cardioids: $r = a * (1 + \cos (\theta))$

limaçons of Pascal: $r = a - b * \cos (\theta)$

ellipses: $r = \sqrt{ (b^2 / (1 - a^2 * \cos (\theta)^2) }$

ovals of Cassini: $r = \sqrt{ (a^2 * (\cos (2 \theta) + \sqrt{ ((b / a)^4 - \sin (2 \theta)^2))) }$

parameter	value
a	0.68
b	0.8
τ_1	0.08
τ_2	4.1
v	− 1.5
g_{15}, g_{27}, g_{37}, g_{84}	1.25
g_{52}, g_{63}, g_{74}, g_{93}	1.5
g_{31}, g_{46}, g_{79}	1.75
g_{19}, g_{76}, g_{78}, g_{91}	2.0
all other g_{ji}	0

Table 6.2 – FitzHugh-Nagumo model parameters.

where r is the radius and a and b are constants. Gaussian noise (*N(0, 1)* multiplied by a scaling factor) is added to the stimulus to make each shape unique. Note that these values are already somewhat noisy due to image pixilation and steerable filter inaccuracies.

The inputs to each of the nine cells represent the measured curvatures of individual pixels at specific positions on the bounding contour – equally spaced and approximately forty degrees apart. This is reminiscent of the Connor work (Pasupathy and Connor, 2001). Alternatively, we have chosen outliers from the curvature mean, typically at non-uniform positions, as the input to each cell, reminiscent of some curvature extrema techniques of template matching.

The connectivity between the cells (conductances) is defined by a sparse connection matrix, with each neuron having between one and four output connections. We have experimented with several configurations, including: a zero-valued connection matrix, uniform nearest neighbor connectivity, uniform connectivity with randomly severed connections, sparse random connectivity with weights proportional to the distance between the neurons, sparse random connectivity with uniform weighting and a methodology requiring non-zero connections to have a curvature difference below a parameterized threshold (for example, 5% of the total curvature range). We have also used a scheme where the connection strength between neuron i and neuron j is:

$$w_{ij} = \exp(-(c_j - c_i)^2 / c^2)$$

where c_i and c_j are the curvature values at points i and j and c is the curvature value of a perfect circle containing the points i and j. Some of these arrangements represent symmetric connections and often result in a general lack of periodicity. For the results that we will present here, we implement a connection matrix, with values provided in Table 6.2, with weights (ranging from 1.25 to 2.00) proportional to neuron-to-neuron distance.

We utilize basic signal processing techniques and examine the response of each individual neuron in the network to determine the following features: number of spikes, approximate phase, mean number of spikes per burst, mean burst duration and approximate period in response to the stimulation. Subsequently, we employ the k-Nearest Neighbor classification methodology (Mitchell, 1997) in an attempt to categorize the responses.

6.2.3 The Izhikevich Model and Amplification in IT via Gamma Synchronization in V4

The Izhikevich equations represent a simple model of spiking neurons as a two-dimensional system having a fast voltage variable and a slower recovery variable (Izhikevich, 2003; Izhikevich, 2007; Izhikevich and Edelman, 2008). The membrane recovery variable accounts for the activation of K^+ ionic currents and the inactivation of Na^+ ionic currents. As is typical, the fast variable has an N-shaped nullcline, whereas the

slower variable has a sigmoid-shaped nullcline. The principal equations of the dimensional form of the phenomenological model are:

$$C\, dv/dt = k (v - v_r) (v - v_t) - u + I$$

$$du/dt = a \{ b (v - v_r) - u \}$$

$$\text{if } v \geq v_{peak} \text{ then } v \leftarrow c \text{ , } u \leftarrow u + d$$

where v is the membrane potential, u is the membrane recovery current variable, t is time, I is the input current, C is the membrane capacitance, k is the rheobase parameter, v_r is the resting membrane potential, v_t is the instantaneous threshold potential, v_{peak} is the soma's spike cutoff, a is the recovery time constant (the decay rate), b is the input resistance parameter (the sensitivity of the recovery variable to subthreshold fluctuations of the membrane potential), c is the soma's after spike voltage reset and d is the after-spike reset of the recovery variable (the outward minus inward currents activated during the spike and affecting the after-spike behavior).

The alternative form:

$$v = v + \tau * ((v_2 * v^2) + (v_1 * v) + v_0 - u + I)$$

$$u = u + \tau * (a * ((b * v) - u))$$

may also be used. With full synaptic kinetics, the current is computed as:

$$I(t) = -I_{dendritic} - I_{synaptic}$$

with the synaptic current given by:

$$\begin{aligned} I_{synaptic} = {} & g_{AMPA}\,(v - 0) \\ & + g_{NMDA}\,\{[(v + 80)/60]^2 / (1 + [(v + 80)/60]^2)\}\,(v - 0) \\ & + g_{GABAA}\,(v + 70) \\ & + g_{GABAB}\,(v + 90) \\ & + I_{gap} \end{aligned}$$

and the conductance given by:

$$g(t) = g_0\, e^{-t/\tau}$$

with

$$g_0 = g(t) + c$$

when the presynaptic cell fires at time t. The time-step $\tau = R\,C$.

With various choices of parameters a, b, c and d, the model can exhibit all known types of firing patterns. Our parameter value choices are provided in Table 6.3. Some of our implementations have included the glutamatergic excitatory regular spiking (RS) cell, the most typical neuron in cortex, as well as the GABAergic inhibitory fast spiking (FS) interneuron cell. We typically exclude short-term synaptic plasticity, NMDA and $GABA_B$ contributions and dopamine-modulated dendritic spike-timing-dependent plasticity (STDP), but include the excitatory AMPA and inhibitory $GABA_A$ presynaptic currents.

However, we have found that even the simpler, alternative form yields excellent results. This can best be described as a pulse-coupled neural network (PCNN), with the firing of a presynaptic neuron instantaneously changing the variable *v* by a predetermined synaptic connection weight.

We create a population of Izhikevich cell models to explore synchronization dynamics. We have explored several topologies and again, even the simplest are effective. For the results that we will present here, we implement ten V4 cells, each receiving a noisy injection current. These cells all have excitatory projections to an IT cell. The IT cell has an excitatory connection to an interneuron, presumably at the level of IT, which may or may not have inhibitory feedback connections to the V4 cells. We again use genetic algorithms (Goldberg, 1989) and searches of parameter space to find connection weights sufficient to achieve realistic firing objectives (10−40 Hz). For the fastest computation, we employ the forward Euler method.

parameter	value
C	RS: 100 FS: 20 PCNN: x $\mu F/cm^2$
k	RS: 3 FS: 1 PCNN: x
v_r	RS: –60 FS: –55 mV
v_t	RS: –50 FS: –40 mV
v_{peak} (soma)	RS: 50 FS: 25 mV
v_2	PCNN: 0.04 mV
v_1	PCNN: 5 mV
v_0	PCNN: 140 mV
a	RS: 0.01 FS: 0.15 PCNN: 0.02 (excitatory) 0.1 (inhibitory)
b	RS: 5 FS: 8 PCNN: 0.2
c (soma)	RS: –60 FS: –55 PCNN: –65
d	RS: 400 FS: 200 PCNN: 8 (excitatory) 2 (inhibitory)
τ	5 (AMPA) 6 ($GABA_A$)
c	10 (AMPA) 4 ($GABA_A$)

Table 6.3 – Izhikevich model parameters.

All simulations were carried out in a Microsoft Windows XP Professional SP2 environment on an Intel® Pentium® 4 CPU running at 2.80 GHz with 3.00 GB of RAM. All models were constructed using the MATLAB application development environment (version 7.9.0.529 R2009b) and the associated Curve Fitting Toolbox (version 2.1), Genetic Algorithm and Direct Search Toolbox (version 2.4.2), Image Processing Toolbox (version 6.4), Neural Network Toolbox (version 6.0.3), Optimization Toolbox (version 4.3), Signal Processing Toolbox (version 6.12), Statistics Toolbox (version 7.2) and Wavelet Toolbox (version 4.4.1).

6.3 Results

The Morris-Lecar model, the FitzHugh-Nagumo model within a winnerless competition network and the Izhikevich model within a feedback network have been implemented.

6.3.1 The Morris-Lecar Model and Recognition

The validation of our Morris-Lecar model cell is shown in Figure 6.1. The top left plot depicts this cell's response to a constant current application. The bottom left plot shows the frequency components of this same response. We see that the injected current produces a response that is primarily 20 Hz – reasonable and expected. The plot on the right provides the oscillation frequency of this cell in response to a wide range of input

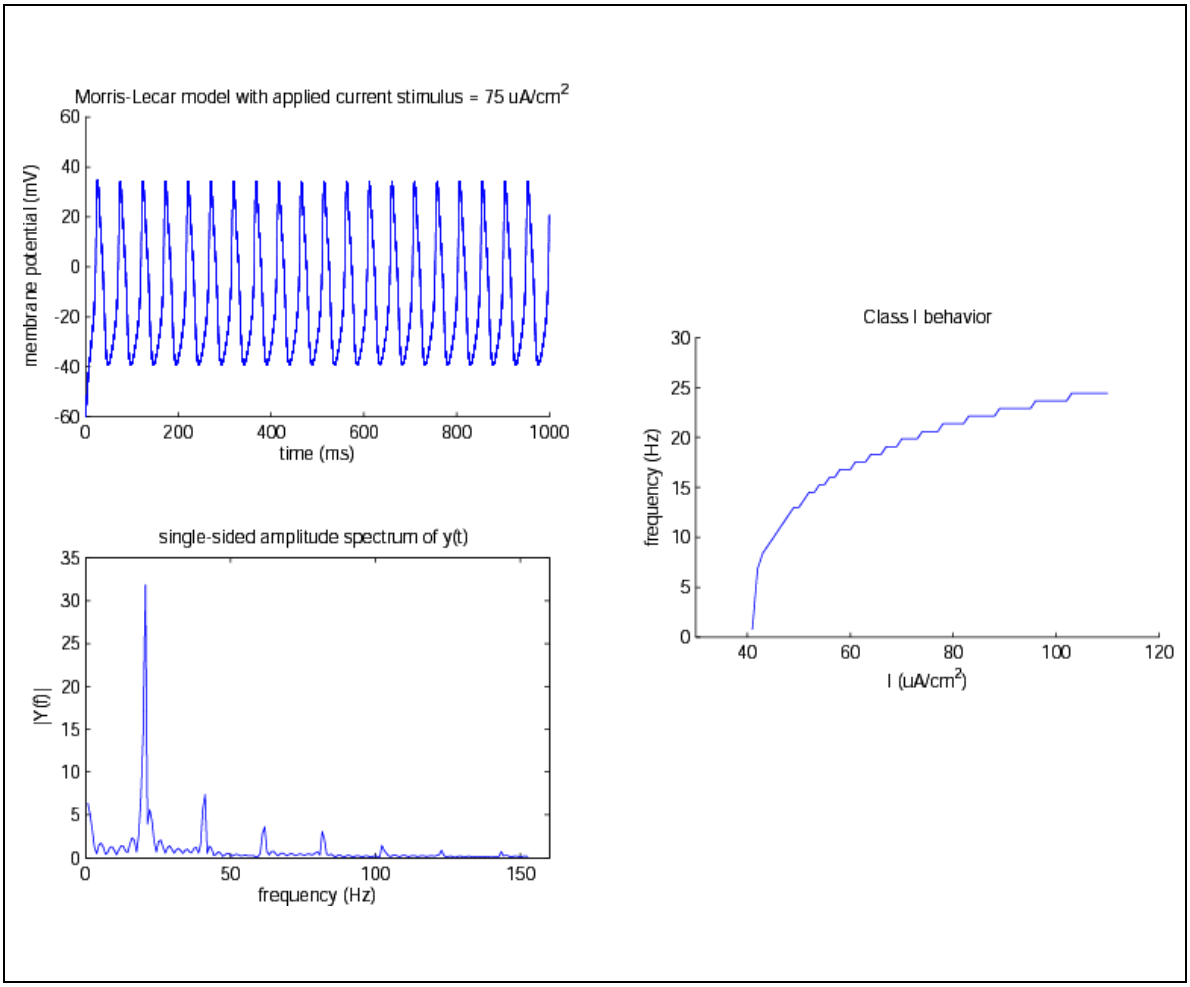

Figure 6.1 – Morris-Lecar model validation. The top left plot depicts a Morris-Lecar cell's response to a constant current application. The bottom left plot shows the frequency components of this same response. The plot on the right provides the oscillation frequency of the cell in response to a wide range of input currents – typical behavior of a Type I oscillator.

currents. This behavior – increased stimulus intensity (current injection) resulting in increased frequency of response over a wide range – is typical behavior of a Type I oscillator and is necessary for our type of intensity-driven recognition.

In Figure 6.2 we show a Morris-Lecar cell's response to selected images. This Morris-Lecar cell models an IT cell, which integrates specific information about the 2-dimensional boundary shapes of multiple contour fragments (the V4-like cell inputs, responding to an image's iso-curvature segments). Its parameters, including Gaussian constituents and coefficients, are selected to match those of an IT-like cell created in Chapter 5. The optimal set of synaptic conductances for this cell is found using a genetic algorithm. The sample images on the left side of the figure are presented to the Morris-Lecar cell at time $t = 0$, producing the responses on the right side. This cell, tuned to respond to guitars, prefers guitar images above images in other categories.

Note that this situation is not entirely biologically realistic. The stimulus is presented at time $t = 0$ and the Morris-Lecar IT-like cell appears to respond instantaneously (i.e., without regard for synaptic transmission delays through retina, primary visual cortex, extrastriate cortex, etc.). Also, the cell's spiking is persistent – in contrast to the diminishing response seen in cortex. However, the IT-like cell's differential firing rate – the primary objective of this exercise – is accurate and reflects what the Connor group has observed in IT (Brincat and Connor, 2004; Brincat and Connor, 2006).

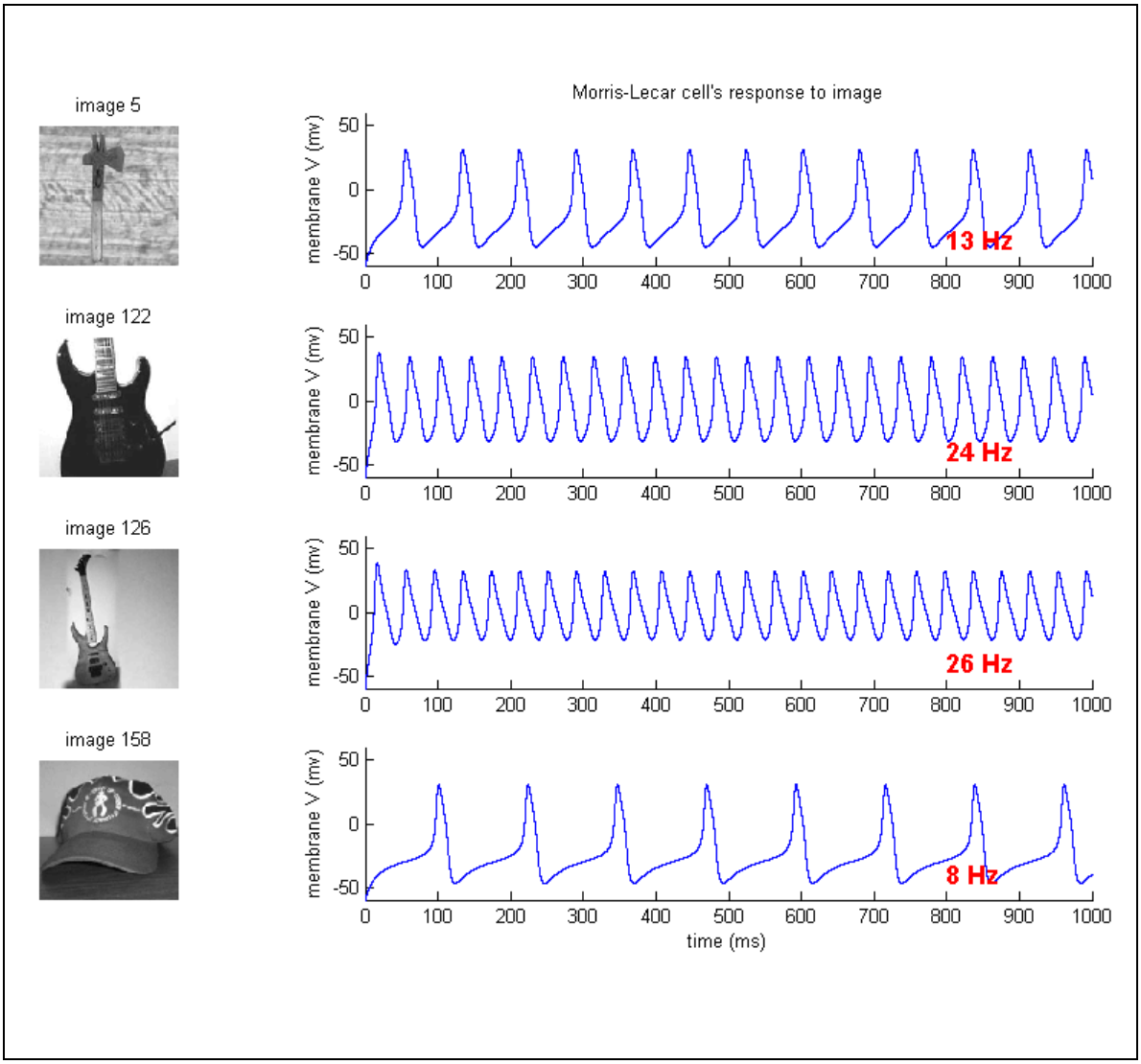

Figure 6.2 – Morris-Lecar cell's response to selected images. This Morris-Lecar cell models an IT cell, with inputs from V4 cells. Sample images are on the left side. Image numbers are given. The corresponding Morris-Lecar cell responses (with image presentation at time $t = 0$) are on the right side. Response frequency is shown in red. This cell, tuned to respond to guitars, prefers guitar images above images in other categories.

Ten example Kanwisher natural images from each of the seven sampled categories (axes, cats, fish, guitars, handsaws, hats, and scissors) are selected. In Figure 6.3, we show the histogram of responses of one similarly-created Morris-Lecar IT cell to each of these images. Again, this cell is tuned to respond preferentially to guitars. This is similar to our work in Chapter 5 with "mathematical" cells and is reminiscent of the study of a single unit in the left posterior hippocampus / medial temporal lobe of epilepsy patients with depth electrodes by Cristof Koch (Quiroga *et al.*, 2005). Here, the cells were activated exclusively by different views of Jennifer Aniston, for example, and not Julia Roberts, etc. Note that these cells are not designed to be, nor do they behave like, "grandmother" cells. They simply respond well to cells within a single category, at the exclusion of others. We have created many Morris-Lecar IT-like cells similar to these that respond preferentially to each of the seven categories in our sample space.

6.3.2 The FitzHugh-Nagumo Model, Spatiotemporal Patterns and Winnerless Competition

Our winnerless competition network topography with nine FitzHugh-Nagumo model neurons is shown in Figure 6.4. Notice the unsymmetrical connectivity of the cells and their expectations of curvature values from the images' bounding contours at forty-degree increments.

The responses of the nine FitzHugh-Nagumo cells within the same winnerless competition network when presented with three different noisy images of geometric

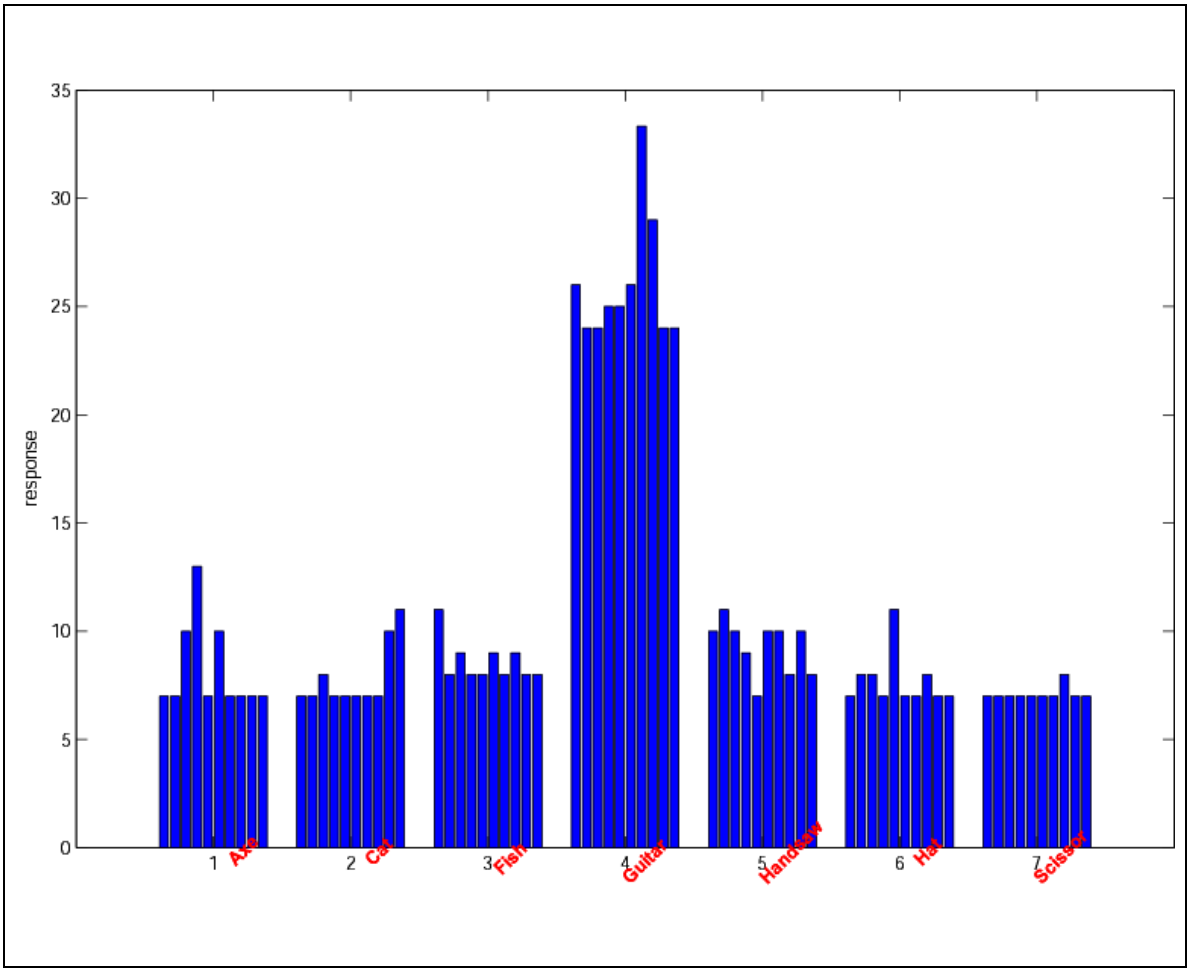

Figure 6.3 – Morris-Lecar cell's response to each image. The 70 sample images are grouped into 7 categories (with 10 in each category). All 7 categories (axes, cats, fish, guitars, handsaws, hats, and scissors) are represented. The height of each bar represents the response of the Morris-Lecar cell (modeling an IT cell) to each of these images. Again, this cell was tuned to respond preferentially to guitars.

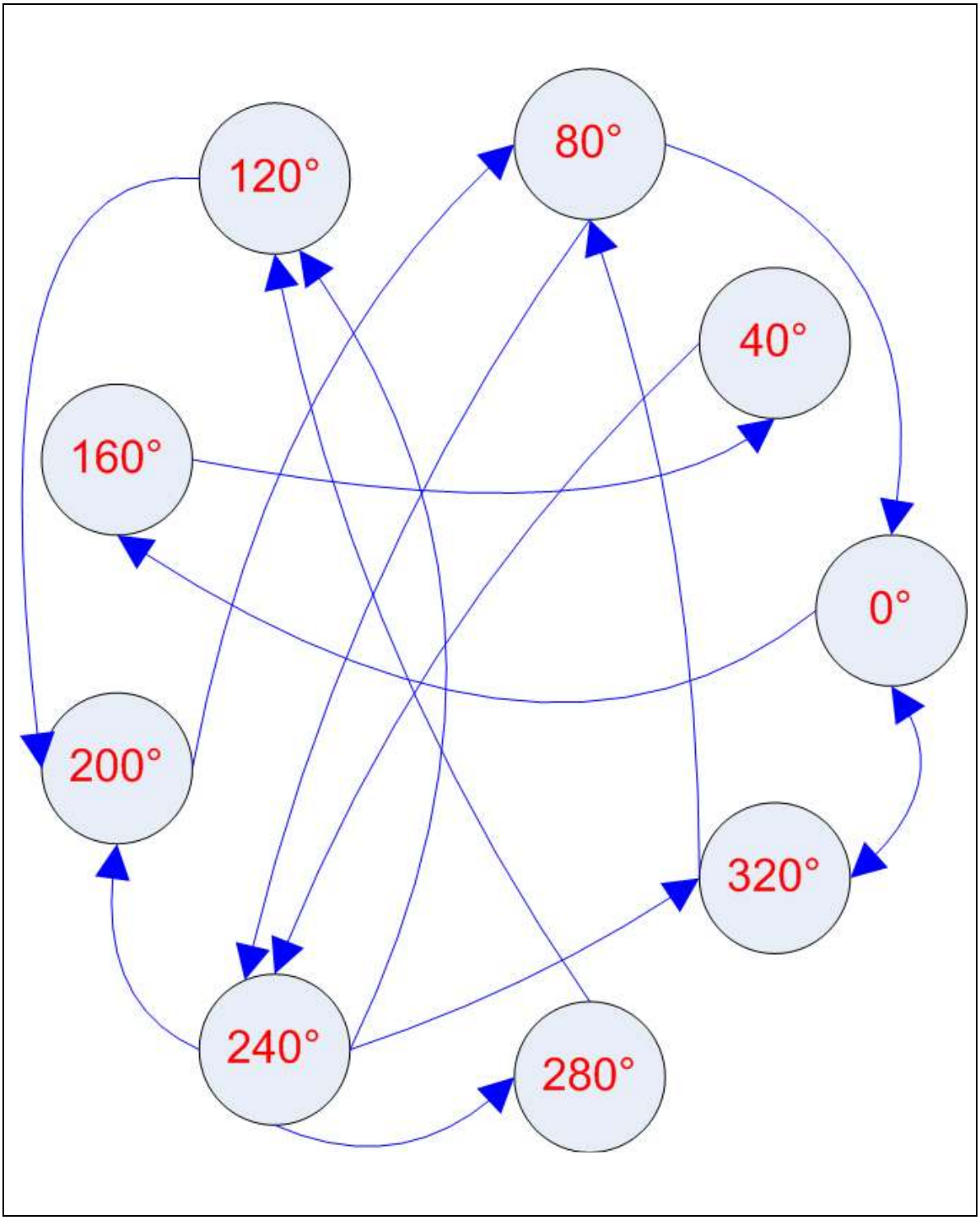

Figure 6.4 – Winnerless competition network topography with 9 neurons.

shapes – a circle (top left), an ellipse (top right) and a Cassini oval (bottom middle) – are shown in Figure 6.5. The stimulus for each neuron is proportional to the curvature value at a specific location on the shape's bounding contour (the first neuron responds to the value at 0 degrees, the second neuron responds to the value at 40 degrees, …, the ninth neuron responds to the value at 320 degrees). This is much like a V4 cell that is selective for a particular local shape configuration at a particular location on the contour within a larger shape. The differences between the nine-cell populations for each shape are visually apparent.

As seen in Figure 6.6, a single FitzHugh-Nagumo neuron within the winnerless competition network can provide shape discrimination. On this scatter plot, the responses of the third neuron to fifty noisy sample images from each of the three geometric shape categories – circles, ellipses, ovals of Cassini – are projected onto the three dimensions of our analysis – approximate phase, mean burst duration and approximate period. A 4-Nearest Neighbor classification attempt (Mitchell, 1997) results in a superior 1.3% error, suggesting that these dimensions make the data, and its visually apparent structure, most amenable to clustering. Note that this error is appropriately considered a training error, not a testing error. Similarly, Figure 6.7 shows the responses of a different FitzHugh-Nagumo cell (the ninth neuron) within the winnerless competition network to the same fifty noisy sample images from each of the three geometric shape categories projected onto the same three dimensions. This time, a 4-Nearest Neighbor classification attempt results in a superior 2.0% error. It should be noted that not all cell responses in the network can discriminate between shapes, at least not within the same three dimensions.

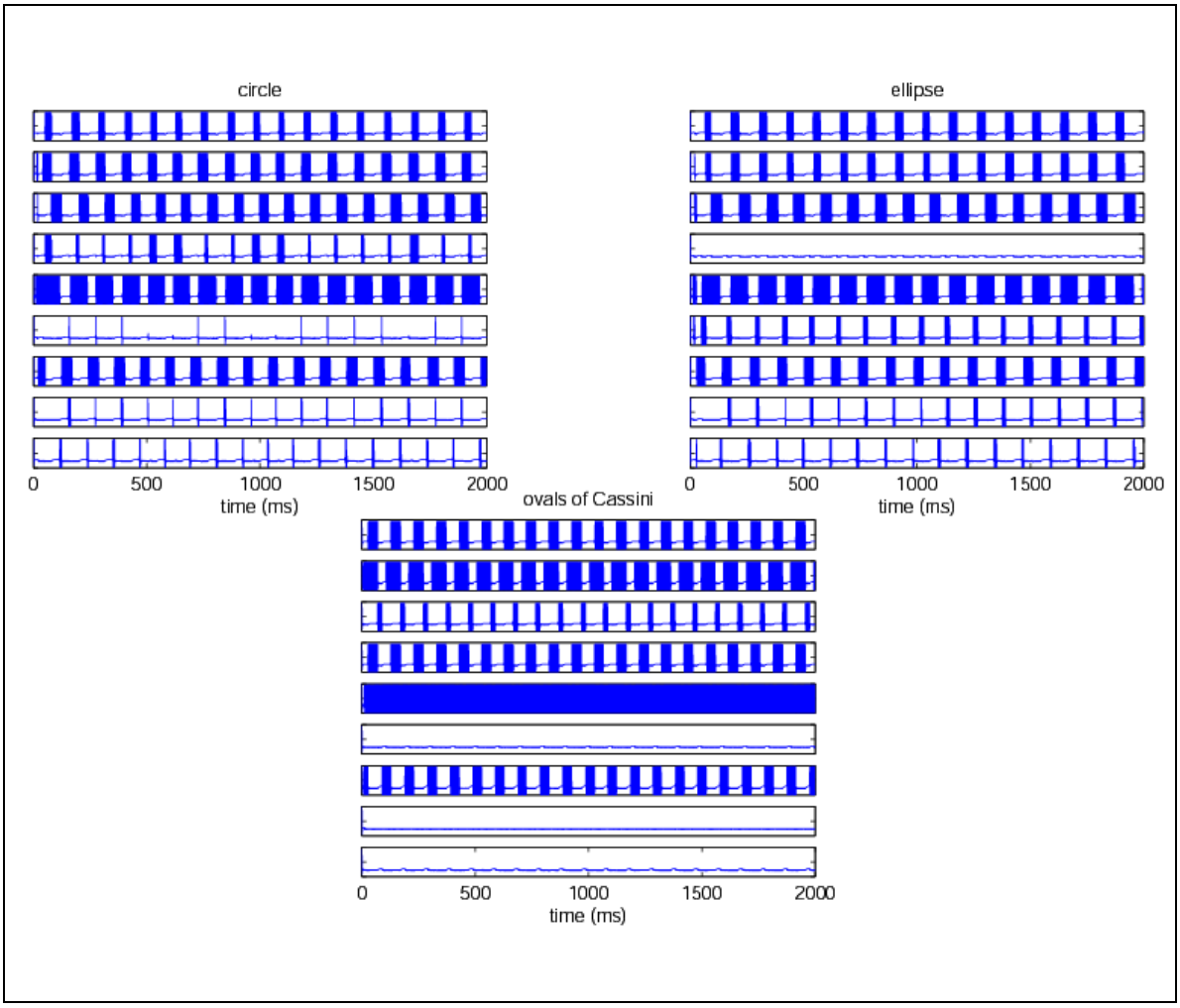

Figure 6.5 – FitzHugh-Nagumo model. The responses of the 9 FitzHugh-Nagumo cells within the same winnerless competition network when presented with 3 different noisy images of geometric shapes – a circle (top left), an ellipse (top right) and a Cassini oval (bottom middle) – are shown. Each subplot for each shape represents the response of a single neuron (neurons 1 through 9, top to bottom), with time (for 2 seconds) on the abscissa and membrane potential on the ordinate. The stimulus for each neuron is proportional to the curvature value at a specific location on the shape's bounding contour (neuron #1: 0°, neuron #2: 40°, neuron #3: 80°, …, neuron #9: 320°). This is much like a V4 cell that is selective for a particular local shape configuration at a particular location on the contour within a larger shape. The differences between the 9-cell populations for each shape are visually apparent.

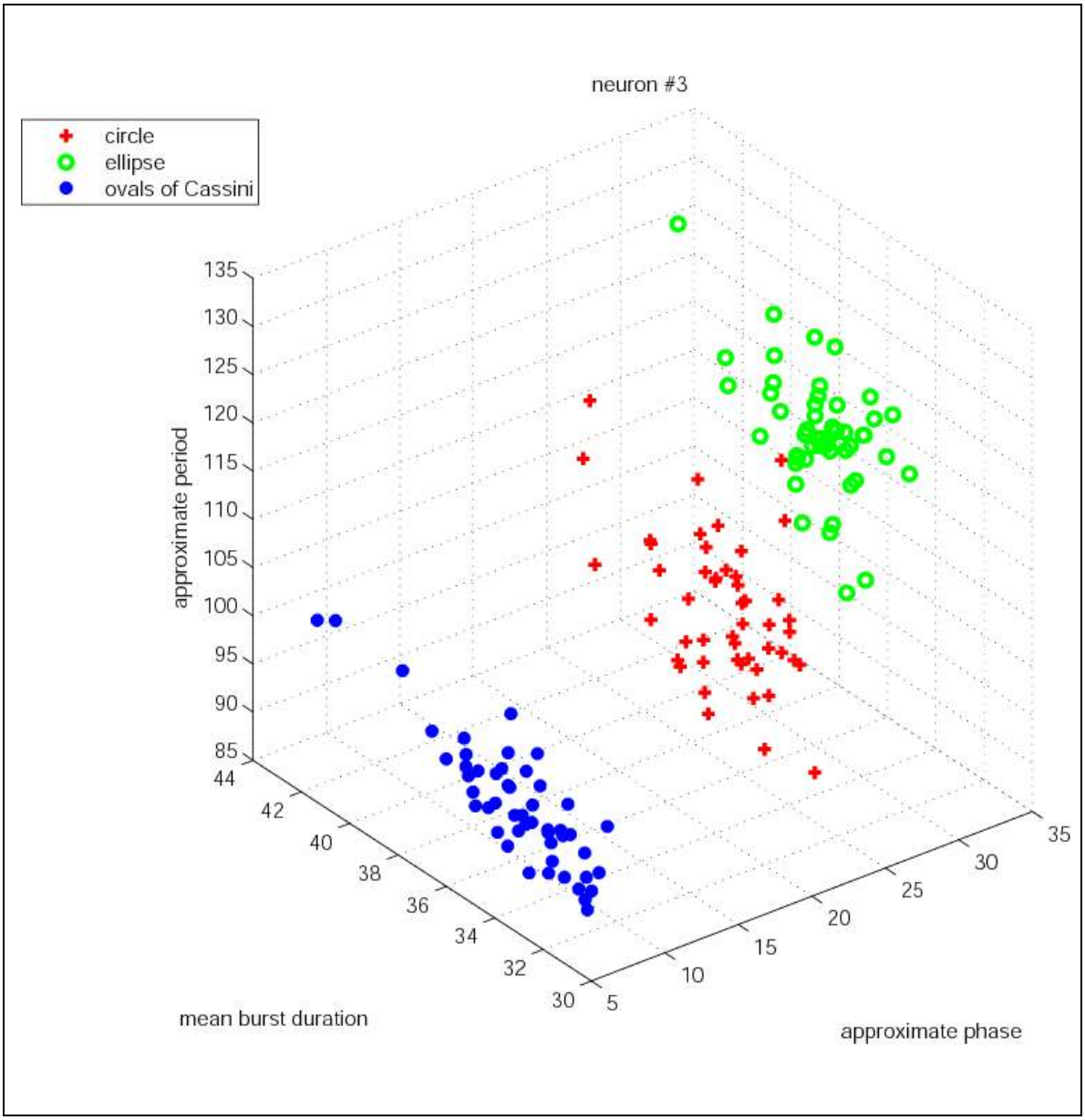

Figure 6.6 – FitzHugh-Nagumo neuron within a winnerless competition network. A single neuron from the network can provide shape discrimination. On this scatter plot, the responses of neuron #3 to 50 noisy sample images from each of the 3 geometric shape categories – coded by color and shape – are projected onto the 3 dimensions of our analysis – approximate phase, mean burst duration and approximate period. Classification (4-Nearest Neighbor) at this level results in an error of 1.3%.

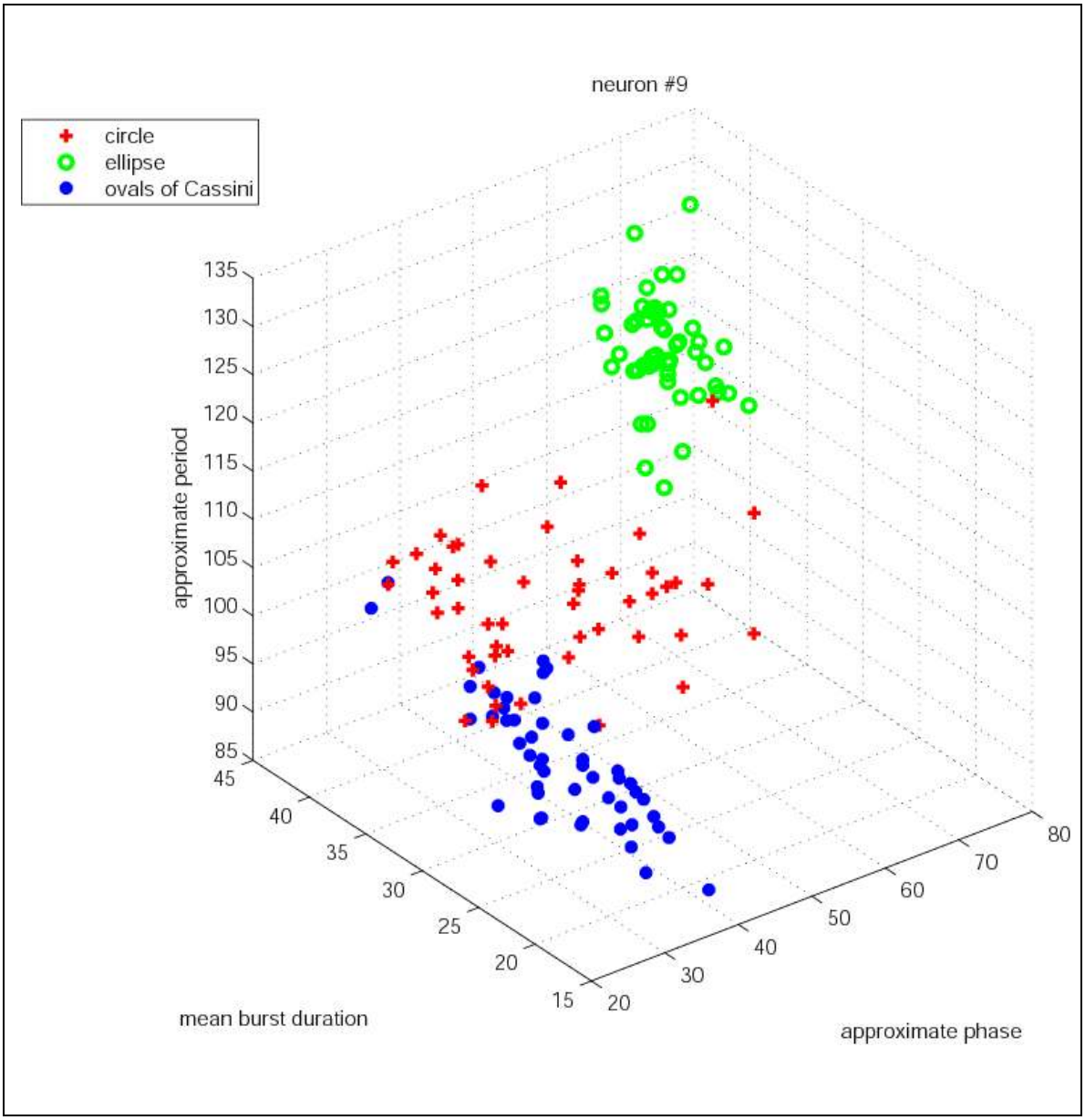

Figure 6.7 – FitzHugh-Nagumo neuron within a winnerless competition network. A single neuron from the network can provide shape discrimination. On this scatter plot, the responses of neuron #9 to 50 noisy sample images from each of the 3 geometric shape categories – coded by color and shape – are projected onto the 3 dimensions of our analysis – approximate phase, mean burst duration and approximate period. Classification (4-Nearest Neighbor) at this level results in an error of 2.0%.

The second neuron, for instance, has a 34% classification error, while the fifth neuron shows a 33% error.

6.3.3 The Izhikevich Model and Amplification in IT via Gamma Synchronization in V4

The validation of our Izhikevich model cell network is shown in Figure 6.8. We explore a network of ten V4 cells, one IT cell and one Inhibitory cell. A current injection – enough to produce one spike in one cell – is administered at 220 ms and again at 270 ms. The top plot shows injections into a V4 cell, which has an excitatory connection to the IT cell. All other connections are severed. A single IT cell spike results for each V4 cell injection. The middle plot shows injections into the IT cell, which has an excitatory connection to the Inhibitory cell. All other connections are severed. A single Inhibitory cell spike results for each IT cell injection. The bottom plot shows injections into the Inhibitory cell, which has an inhibitory connection to a V4 cell. All other connections are severed. No V4 cell spikes are produced. Note the differences in the refractory periods for the different cell types. We have also verified that our Izhikevich cells demonstrate the typical behavior of a Type I oscillator.

The Izhikevich model cell network without feedback is shown in Figure 6.9. Again, we explore a network of ten V4 cells, one IT cell and one Inhibitory cell. Each V4 cell has an excitatory connection to the IT cell. The IT cell has an excitatory connection to the Inhibitory cell. Noisy and unsynchronized (and different for each cell) current injections

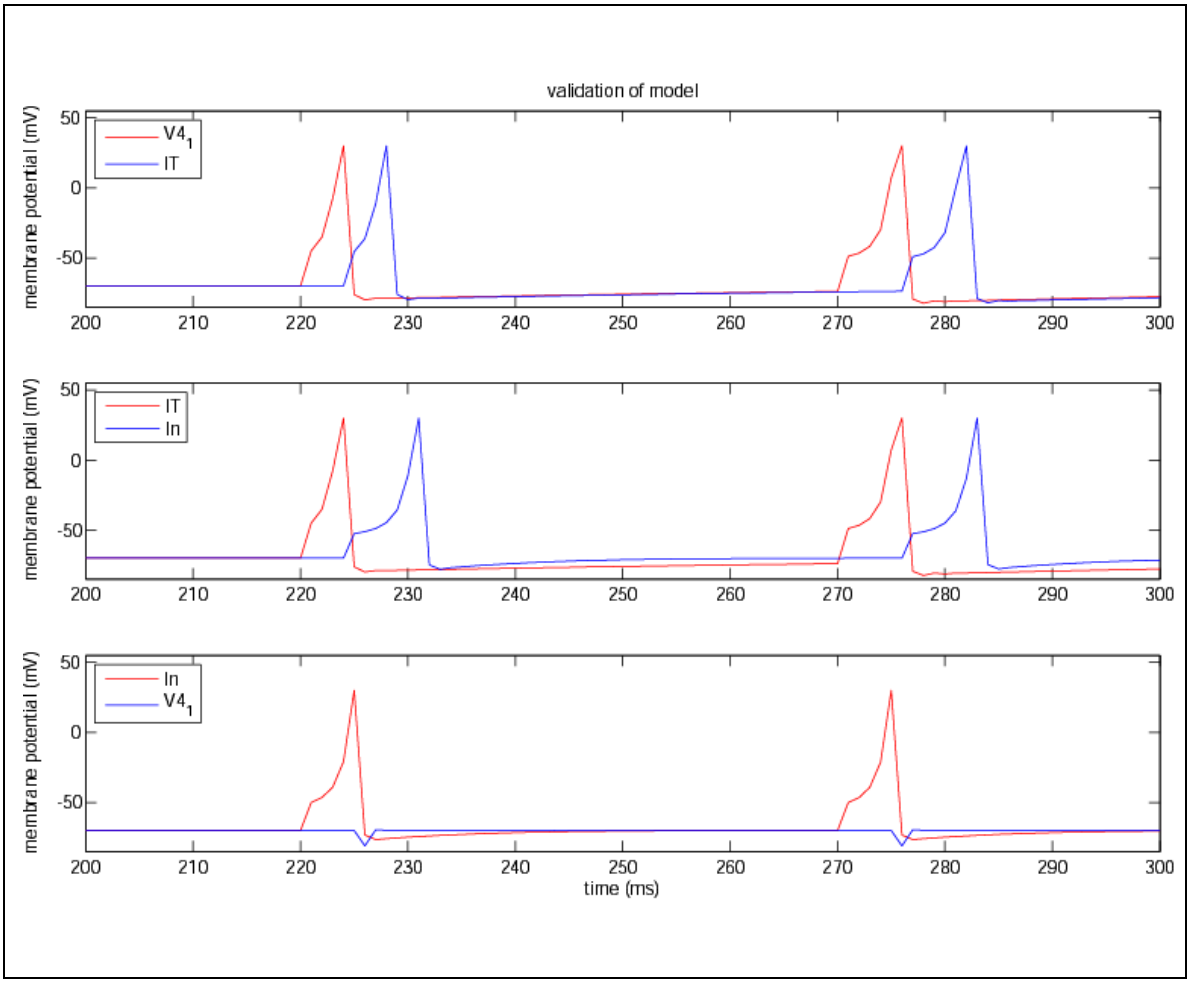

Figure 6.8 – Izhikevich model validation. We explore a network of 10 V4 cells, 1 IT cell and 1 Inhibitory cell. Each plot shows time on the abscissa and membrane potential on the ordinate. A current injection – enough to produce 1 spike in 1 cell (with responses graphed in red) – was administered at 220 ms and again at 270 ms. The top plot shows injections into a V4 cell, which has an excitatory connection to the IT cell (with response graphed in blue). All other connections were severed. A single IT cell spike results for each V4 cell injection. The middle plot shows injections into the IT cell, which has an excitatory connection to the Inhibitory cell (with response graphed in blue). All other connections were severed. A single Inhibitory cell spike results for each IT cell injection. The bottom plot shows injections into the Inhibitory cell, which has an inhibitory connection to a V4 cell (with response graphed in blue). All other connections were severed. No V4 cell spikes are produced. Note the differences in the refractory periods for the different cell types.

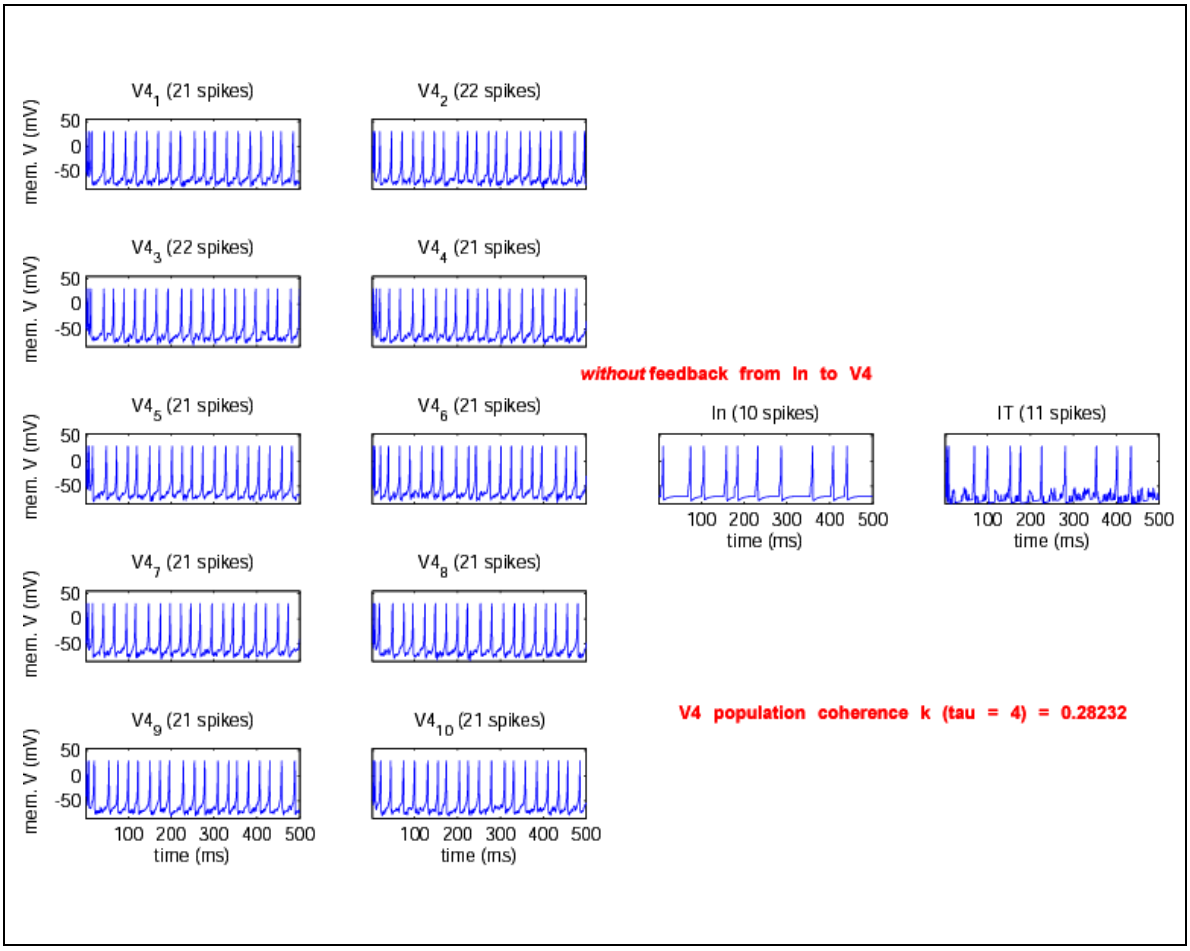

Figure 6.9 – Izhikevich network without feedback. We explore a network of 10 V4 cells, 1 IT cell and 1 Inhibitory cell. Each V4 cell has an excitatory connection to the IT cell. The IT cell has an excitatory connection to the Inhibitory cell. Each plot shows time on the abscissa and membrane potential on the ordinate. Noisy and unsynchronized current injections – enough to produce continuous spiking at approximately 20 Hz – were administered to the V4 cells only. The total number of spikes generated in each cell is given. With this arrangement, the coherence of the V4 cell population (with $\tau = 4$) is 0.282.

– enough to produce continuous spiking at approximately 20 Hz – are administered to the V4 cells only. We utilize the population coherence measure defined by Wang and Buzsáki (1996), based on coincident firings of neural pairs, to evaluate synchronization. With this arrangement, the coherence of the V4 cell population (with $\tau = 4$) is 0.282. Eleven spikes are produced in the IT cell.

In Figure 6.10, we explore a similar Izhikevich model cell network of ten V4 cells, one IT cell and one Inhibitory cell, but here we include feedback. Each V4 cell has an excitatory connection to the IT cell. The IT cell has an excitatory connection to the Inhibitory cell. The Inhibitory cell has an inhibitory connection to each V4 cell. Noisy and unsynchronized current injections – enough to produce continuous spiking at approximately 20 Hz – are again administered to the V4 cells only. With this arrangement, the coherence of the V4 cell population (with $\tau = 4$) is 0.395. This represents a 40% increase over the coherence value of the V4 cell population in the network without feedback, as well as a 73% increase in IT cell spiking frequency – from eleven spikes (22 Hz) to nineteen spikes (38 Hz). It is our contention that this response amplification in IT corresponds roughly with classification.

We present a statistical validation of the Izhikevich model cell network in Figure 6.11. A total of 100 experiments (each with different current inputs into the V4 cells) are performed. In each plot, the results of a two-tailed, paired *t*-test (Press *et al*., 2007) are displayed. In the top plot, the differences in the numbers of spikes in the V4 cell population – first with, then without, feedback from the Inhibitory cell to the V4 cells –

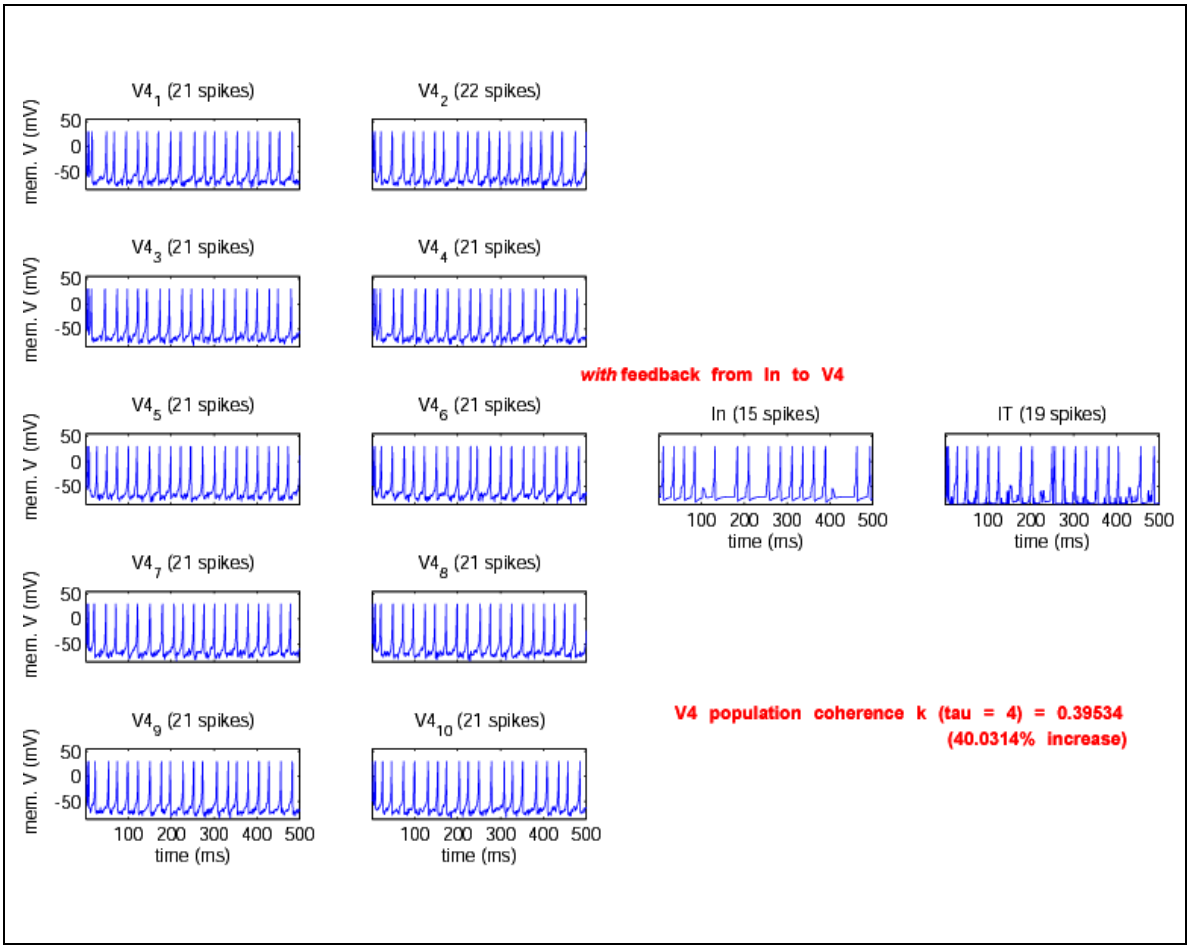

Figure 6.10 – Izhikevich network with feedback. We explore a network of 10 V4 cells, 1 IT cell and 1 Inhibitory cell. Each V4 cell has an excitatory connection to the IT cell. The IT cell has an excitatory connection to the Inhibitory cell. The Inhibitory cell has an inhibitory connection to each V4 cell. Each plot shows time on the abscissa and membrane potential on the ordinate. Noisy and unsynchronized current injections – enough to produce continuous spiking at approximately 20 Hz – were administered to the V4 cells only. The total number of spikes generated in each cell is given. With this arrangement, the coherence of the V4 cell population (with $\tau = 4$) is 0.395. This represents a 40% increase over the coherence value of the V4 cell population in the network without feedback, as well as a 73% increase in IT cell spiking frequency (from 11 spikes / 22 Hz to 19 spikes / 38 Hz).

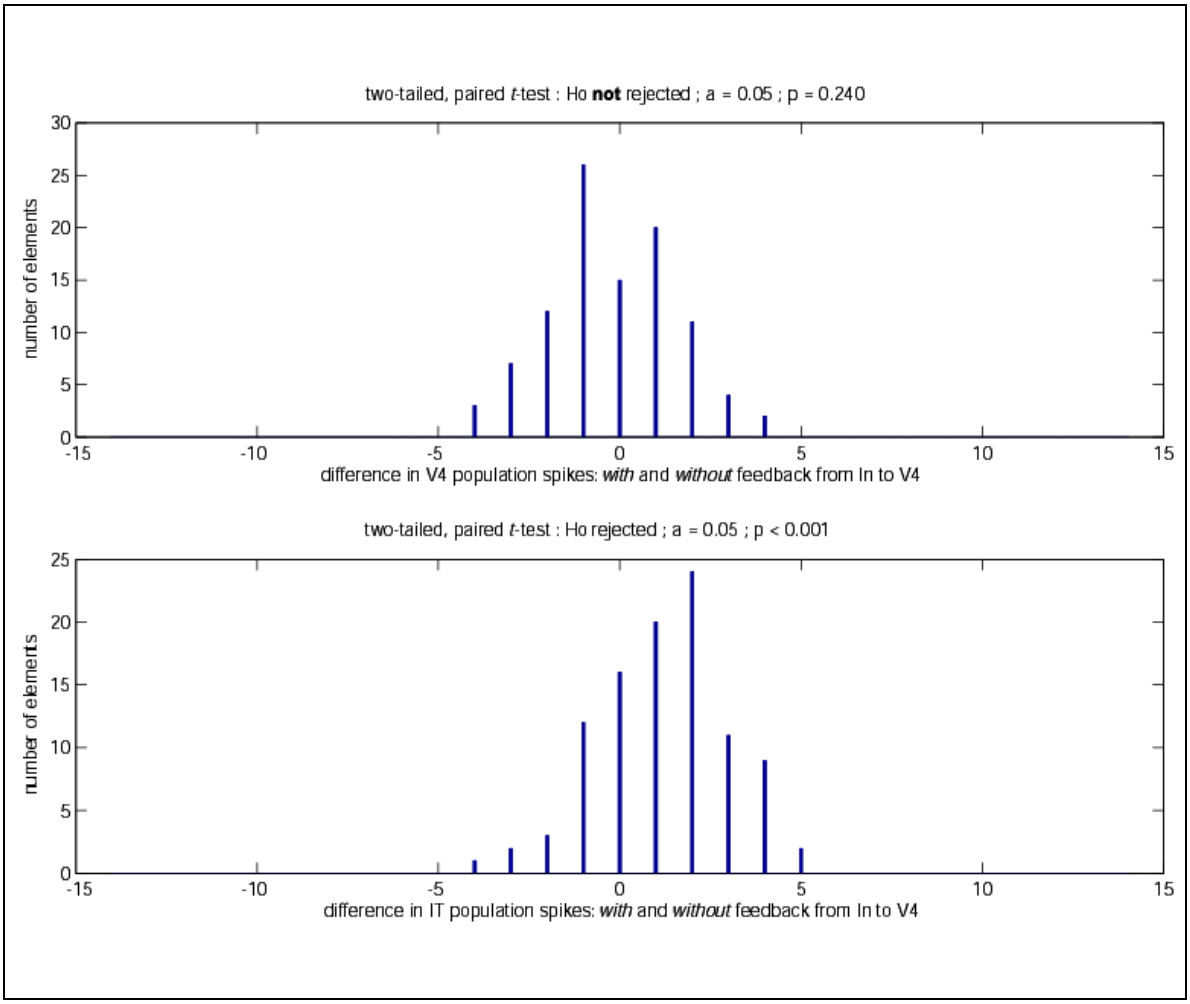

Figure 6.11 – Statistical validation of the Izhikevich network. A total of 100 experiments (each with different current inputs into the V4 cells) were performed. In each plot, the results of a two-tailed, paired *t*-test are displayed. In the top plot, the differences in the numbers of spikes in the V4 cell population – first with, then without, feedback from the Inhibitory cell to the V4 cells – are compared. Here, the null hypothesis – that the difference data came from a distribution of mean zero – cannot be rejected (i.e., "with feedback" data and "without feedback" data came from distributions with the same mean) at the 5% significance level. In the bottom plot, the differences in the numbers of spikes in the IT cell – first with, then without, feedback from the Inhibitory cell to the V4 cells – are compared. Here, the null hypothesis – that the difference data came from a distribution of mean zero – is rejected (i.e., "with feedback" data and "without feedback" data came from distributions with different means) at the 5% significance level. We conclude that the increased spiking in the IT cell is not a result of increased spiking in the V4 cells.

are compared. Here, the null hypothesis – that the difference data came from a distribution of mean zero – cannot be rejected (i.e., "with feedback" data and "without feedback" data came from distributions with the same mean) at the 5% significance level. In the bottom plot, the differences in the numbers of spikes in the IT cell are compared – first with, then without, feedback from the Inhibitory cell to the V4 cells. Here, the null hypothesis – that the difference data came from a distribution of mean zero – is rejected (i.e., "with feedback" data and "without feedback" data came from distributions with different means) at the 5% significance level. We conclude that the increased spiking in the IT cell is not a result of increased spiking in the V4 cells. This is appropriate, since we do not expect the feedback to increase the firing rate of the V4 cells but we do expect the increased firing rate in IT to be the result of synchronized, rather than more abundant, spiking in V4.

6.4 Discussion

Results such as those of Pasupathy and Connor (2001), Hegdé and Van Essen (2003) and Brincat and Connor (2004, 2006) reveal a great deal of sophisticated shape processing in the ventral stream. Although our previous results are based upon faithful "mathematical" models, our current results show that realistic biological models of cells with curvature- and position-sensitive response properties can function as robust shape descriptors and perform consistently well in recognition tasks. Our success in elucidating the interaction

between V4 and IT from a mechanistic perspective provides more integrity to and credibility for our model.

We have taken a simplified approach to nonlinear multiplication within our Morris-Lecar neuronal models. At the neural level, there is a clear difference between this and a logical "AND" or a simple coincidence detection. We could have implemented a sophisticated model of a coincidence detector, such as that of Mel and colleagues (Poirazi *et al*., 2003a; Poirazi *et al*., 2003b; Polsky *et al*., 2004), but instead choose to multiply our Gaussian constituents before the dendritic level. All terms in the nonlinear response equations were represented by separate channels into the IT-like cells. A more advanced nonlinear dendritic mechanism might resemble something similar to the results of Magee and colleagues (Gasparini and Magee, 2006; Losonczy and Magee, 2006; Losonczy *et al*., 2008).

In our winnerless competition neural network with FitzHugh-Nagumo model neurons, we have essentially expressed static curvatures as time- and position-dependent spatiotemporal patterns whose activity we examine in 9-dimensional space. The main idea is that shape differences are more apparent in spatiotemporal pattern differences than in curvature differences.

Some results suggest that invariant pattern recognition, in the face of deformations as well as different viewing angles, etc., can be achieved by using temporal coding at the

population level (Wyss *et al*., 2003). We can therefore justifiably caste our spatiotemporal patterns of activity in probabilistic and information theoretic terms (Harel *et al*., 2007), with less than perfect position and curvature information.

A rationalization for non-nearest neighbor connectivity in our network might be similar to the "pinwheel" topography and its relationship to short-range connections found by Gilbert and colleagues in primary visual cortex (Das and Gilbert, 1999), and the supposition that these same arrangements might exist later in the ventral stream.

As we have seen, some spatiotemporal codes might help organize signals for classification or clustering. In his analysis of spike trains, Victor has considered the mutual information between stimuli and output clusters and has defined a metric on output patterns with distances defined according to costs of transforming one spike train into another (Victor and Purpura, 1997; Victor 2000). This is similar to our use of Earth Mover's Distance (EMD) metrics (Rubner *et al.*, 2000; Rubner *et al.*, 2001) in previous chapters. Here, we have chosen the simpler k-Nearest Neighbor classification methodology (Mitchell, 1997).

We might ask some fundamental questions of our WLC network. For example, how fast can stimuli or patterns be distinguished from others? This might be a particularly important question if the individual nodes in the network have different starting points (initial values or starting times). In this regard, the work by VanRullen and Thorpe, where the most salient information in the neural code is represented by timing,

specifically the first action potentials over the population, is relevant (VanRullen and Thorpe, 2002). Again, we might caste the problem in probabilistic terms and evaluate the probability of our hypothesis (a particular categorization, for example) given the current data at some point in time (the number of spikes seen arriving in IT from V4 up to that point). As usual, the spike wave could be modified by selectivity and lateral or top-down interactions. Also, we might consider if the patterns should diverge, achieve a steady state, coalesce to the same state, etc., and at what points in time. Finally, we might ask if a network can learn an imposed and novel spatiotemporal pattern and exhibit emergent connectivity.

Our full Izhikevich cell model, which would allow a true test of our circuit's recognition capabilities, would require the implementation of several additional features. Initially, input arrives at V4 from earlier stages of the ventral stream (visual areas 1 and 2 (V1 and V2), etc.), producing asynchronous spiking. The V4 neurons are tuned to specific features (the spiking activity represents the fit). In subsequent versions of our model, V4 would activate sub-populations of IT (and not simply a single IT cell), with activation strengths governed by the input features' alignment with IT's object representation expectations. The stronger the input from V4, the greater the number of IT cells that would fire. Each IT sub-population would have its own set of inhibitory interneurons and represent a class. (Currently, our IT cell represents an "ideally suited" cell or sub-population and the inhibitory cell represents an ideally synchronized sub-population.) There would be synchronous activity within a single inhibitory sub-population and these inhibitory cells would synchronize the activity of the appropriate set of neurons in V4

(representing the expected features of the category) via PING (pyramidal interneuron network gamma) mechanisms (Whittington *et al*., 2000). The inhibition would be too weak to prevent a spike in V4, but strong enough to shift the spike in time. It might also induce sub-threshold oscillations in V4 cells that encode expected, yet absent, features of the object represented by the inhibitory sub-population, making them more sensitive to weak inputs not originally detected. This synchronous activity in V4 amplifies IT response, particularly in the IT sub-population that is synchronized with V4 (due to the maximal temporal summation / integration of the neuronal input), resulting in categorization (Grossberg, 1999). Thus, we increase not only the frequency of individual IT cells, but the number of IT cells participating.

We would expect features that are common to multiple object categories (consider the similarity between the axe handles and the Guitar necks in the Kanwisher image database, for example) to be inhibited in V4, due to the increased inhibitory input from multiple inhibitory cells with connections from different IT sub-populations (multiple categories). Presumably, the inhibition could be enhanced further by the asynchronous nature of spiking across the different inhibitory cells and IT sub-populations. This introduces not only amplification, but amplification and competition, into our model.

Our V4 cells, IT cells and interneurons are modeled with one compartment. Active dendritic compartments could be added. Architecturally, reciprocal connections between the excitatory IT cells and inhibitory interneurons could be added and lateral inhibition between the Inhibitory cells via GABA connections could be implemented.

Given these augmentations, we could confirm our hypotheses by presenting real stimuli to the network and showing that response amplification in IT corresponds roughly with classification. To accomplish this, we could extract a pre-determined set of features from a real image stimulus. Using the V4 cells' defined n-dimensional Gaussian tuning functions, we could determine the response of each V4 cell to each feature in the image. We could transform this response into an input current to the V4 cell. We could repeat this process with out-of-category images containing only "irrelevant" features (and weakly responding V4 cells) and observe a weak / unmodulated / unamplified response in IT, as well as with in-category images containing "relevant" features (and strongly responding V4 cells) and observe a strong / modulated / amplified response in IT. We could interpret this type of two alternative forced-choice (2AFC) protocol – does the stimulus belong to a given category or not – as a demonstration of the analogous categorization process. We might conclude that categorization / response amplification in IT requires not only synchronization in V4 but a sufficient response from V4.

Next, we could present different stimuli, from different categories or sub-categories, with common features to the network and show that competition arises between at least two IT sub-populations, each representing different object categories, with individual V4 neurons receiving unsynchronized feedback inhibition from the different IT sub-populations (each of which may be internally synchronized). We might attempt this type of experiment with the following input, for example: an electric guitar with iso-curvature segments A_1, B_1, C_1 and D_1, an acoustic guitar with iso curvature segments A_1, B_1, C_1

and D_2, and a cat with iso-curvature segments A_1, B_2, C_1 and D_3, and the assumption that each iso-curvature segment produces a strong response in some Gaussian constituent in some cell.

We could also consider the consequences of a simultaneous presentation of similar or dissimilar images or images with occlusions, the biophysical mechanisms of how connections are established during learning (the contextual knowledge which eliminates the necessity of all-to-all connectivity from IT to V4) and how categories are formed, the possible presence of underlying oscillations and the emergent properties of our network.

We have caste V4's role as one of synchronization and IT as a coincidence detector (and not simply a temporal integrator (König *et al*., 1996)). We could consider the effects on the IT cells' receptive fields by synchronized or unsynchronized inputs from V4 and, in general, the relationship between synchronization in V4 and the non-linear responses in IT.

A result by Kanwisher and colleagues (Grill-Spector, 2003; Grill-Spector and Kanwisher, 2005) places significant constraints on theories of object recognition. Accuracy as a function of stimulus duration was determined for all three recognition tasks – object detection, object categorization, and within-category identification. Using short (20 – 70 ms) exposure durations of natural image stimuli (the same that we have used), subjects knew an object's category by the time that they knew that the image contained an object at all (i.e., detection and differential categorization were coincident). This is an

intriguing coincidence, but possibly more. We might argue that the ~60 ms time course from independent sensitivities to constituent parts (linear responses) to selectivity for multipart configurations (non-linear responses) (Brincat and Connor, 2006), or the ~60 ms between categorization and identification (Grill-Spector and Kanwisher, 2005), or the ~60 additional ms required to attend to a feature (compared to attending to a location) (Reynolds and Desimone, 2003), are all manifestations of the unsynchronized to synchronized transformation in our model.

Although the transformation from a distributed coding (in terms of constituent parts) in V4 to a sparse coding (in terms of global object shape) in anterior IT, with posterior IT mediating between the two, is not yet understood at a mechanistic level, some specific facts from Brincat and Connor (2004; 2006) are relevant. They report that nonlinearity is significantly correlated with response sparseness, with the nonlinear integration in IT cells possibly increasing the sparseness of the ventral pathway's shape representation. We consider if distributed representations are important for categorization, if sparse representations are important for identification and if nonlinear integrations require more time. They have found no clear differences in tuning properties along the anterior-posterior axis of IT, as well as finding that even the most selective cells responded to a variety of global shapes. However, anterior IT is thought to present increasingly complex and sparse parts-level population coding or even holistic coding. We consider the necessity of a distributed representation, given that a sparse representation exists (or vice-versa), in IT. Could we instead consider a continuum, either in function or in cortical position, with "distributed" cells at one end and "sparse" cells at the other? Perhaps the

"sparse" cells studied by the Connor group were from slightly more anterior areas, or simply had more of the typical 1–6 Gaussian subunits with nonlinear interactions. Could we in turn say that the answer is "both" to our key question (is parts-based recognition more important at the categorization or identification stage)? Is it possible that a parts-based categorization occurs earlier in posterior IT as a distributed representation and a more refined parts-based identification occurs later in anterior IT as a sparse representation?

The Connor group has found that the within-cell transitions from early linear to later nonlinear responses make a larger contribution than the temporal differences between cells. Also, information about simpler components (linear) appears rapidly, while information about arrangements of multiple components (nonlinear) evolves gradually, with increased processing time required for categorizations that depend upon specific part configurations. We are led to ask if an IT cell responds nonlinearly only when needed. Also, does IT require attention to combine Gaussian constituents, with Gaussian widths (standard deviations) possibly shrinking from categorization to identification?

We broadly consider the nature of recognition, within and across superordinate categories, and wonder if identification uniquely identifies an individual entity, is a lower level of categorization, is a more refined categorization, or is a sub-categorization. We consider if additional information (D, E, F, etc., Gaussian constituents, for example) could be gathered in the 60 ms between categorization and identification. Finally, we ask

if categorization represents a linear and unsynchronized response, while identification represents a nonlinear and synchronized response.

6.5 Conclusion

Our results suggest that realistic biological models of cells with curvature- and position-sensitive response properties, as described by Pasupathy and Connor in V4 and Brincat and Connor in IT, can function as robust shape descriptors. We have demonstrated the utility of cells with response properties similar to those found in V4 and IT by constructing network models and successfully subjecting them to artificial recognition tasks on a set of real images, where they have performed consistently well. We claim that the response properties of V4 and IT cells (i.e., their receptive fields), and in particular their sensitivities to curvatures and contour positions, are useful for object recognition precisely because they, like our realistic biological models of them, facilitate shape representation and categorization by extracting features that correlate with global shape. We have also demonstrated the interaction between V4 and IT from a mechanistic perspective within a biologically realistic network architecture.

Chapter 7

Summary

7.1 Key Points

In Chapter 3, we used the Ullman image (Shashua and Ullman, 1988) and demonstrated that the circular contours were more cocircular (the degree to which the contour curvature remains constant, within a threshold) and contained much longer stretches of cocircularity than the background contours and cross-hatches. However, as the definition of cocircularity became stricter, i.e., curvature was constrained to a narrower range of values, the differences between the circular and background contours diminished. Therefore, at a local scale, all contours in the image were similar. What distinguished the salient circular contours (apart from closure) is that they maintained a similar curvature

over a much longer extent than the background, where "similar" could be quantitatively defined.

We found that "shape context" (Belongie *et al*., 2002) and "curvature context" shape representations achieved comparable recognition accuracies for example image stimuli from the MNIST database of handwritten digits in Chapter 4. To explore the performance capabilities of a V4-like population, the average earth mover's distance to each category was determined. Various combinations of parameter values and feature vector arrangements were tried in different experiments. Feature vectors composed of mean angle of the region, mean curvature of the region, mean direction of curvature of the region and mean distance from the center of mass of the region consistently resulted in superior performance. Results showed that the value of sigma was less important than the choice of region size. It was not advantageous to consider the distinction between inner and outer contours. Also, curvature tolerance was found to have an optimal operating range. It appeared that curvature was the feature that was the most sensitive to noise, suggesting that curvature is the most salient feature for shape recognition. An extensive evaluation using the MNIST "Test Set" database and the MPEG-7 Shape Silhouette database resulted in high levels of classification accuracy. Finally, we showed that top-down inputs can improve classification accuracy by re-weighting the contributions of intermediate-level units.

We utilized sample Kanwisher natural images (Grill-Spector and Kanwisher, 2005) from several categories in our analysis in Chapter 5. We initially explored the performance

capabilities of a V4-like cell population (which precedes the nonlinear integration of boundary components seen in IT) and achieved mediocre results. We hypothesized that it is the nonlinear integration component of the IT cells' functionality that facilitates recognition at the highest level. We constructed populations of IT-like cells using a wide variety of techniques and parameter values. We presented histograms of IT cell responses, with many cells presenting clear preferential responses to specific categories, at the exclusion of others. Unsatisfied with principal components analysis (PCA), we used three-dimensional non-classical non-metric multidimensional scaling (MDS) analysis and 4-Nearest Neighbor classification to better visualize our cell response space. We also constructed a support vector machine (SVM) for classification. We illustrated the fact that different images may activate the IT-like cells, and specifically their Gaussian constituents, differently, yet produce similar total responses – necessary to achieve the robust discriminatory power of our network. We found that there were subsets of cells in the population that were more adept at certain categorizations and we concluded that only a small number of V4 / IT cells may be necessary for image recognition at this level.

In Chapter 6, we constructed and validated Morris-Lecar models of IT cells, which integrate specific information about the 2-dimensional boundary shapes of multiple contour fragments. The optimal sets of synaptic conductances for these cells were found using a genetic algorithm. Kanwisher natural images were again used and histograms of responses, with differential firing rates to different categories, were found. We developed a winnerless competition network topography with nine FitzHugh-Nagumo

model neurons and presented the network response, with visually apparent differences, to three different noisy images of geometric shapes. We considered approximate phase, mean burst duration and approximate period of individual neurons in the network and were able to perform a superior 4-Nearest Neighbor classification. We constructed and validated an Izhikevich model cell network of V4, IT and Inhibitory cells. We experimented using network arrangements with and without feedback and found a substantial increase in V4 cell coherence and average IT cell spiking frequency when feedback was engaged. We hypothesized that response amplification in IT corresponds roughly with classification. We presented a statistical validation of the Izhikevich model cell network and concluded that the increased spiking in the IT cell is not a result of increased spiking in the V4 cells.

7.2 Conclusions

The consistency of curvature on a contour is correlated with increased perceptual salience (the degree to which a target differs from the background). The degree to which curvature covariation contributes to salience depends upon the mechanisms and the scale over which curvature information is computed in visual cortex.

Curvature- and position-sensitive units, as described by Pasupathy and Connor in area V4, can function as robust shape descriptors. The demonstration of shape categorizations

based on curvature representations establishes a connection between state-of-the-art recognition systems and known cortical mechanisms.

More sophisticated curvature- and position-sensitive units, as described by Brincat and Connor in IT, can also function as robust shape descriptors. Our model of highly selective IT-like cell response, a function of the number of V4-like cell inputs and their nonlinear combinations, is not rigidly dependent upon parameter selection or implementation strategy, yet it performs consistently well in recognition tasks.

Realistic biological models of cells with curvature- and position-sensitive response properties, including the interaction between V4 and IT from a mechanistic perspective, as described by Pasupathy and Connor in V4 and Brincat and Connor in IT, can yet again function as robust shape descriptors. The utility of these cells is demonstrated by constructing network models and successfully subjecting them to artificial recognition tasks on a set of real images, where they perform consistently well.

Taken as a whole, these conclusions show that the response properties of V4 and IT cells (i.e., their receptive fields), and in particular their sensitivities to curvatures and contour positions, are useful for object recognition precisely because they, like our faithful computer models of them, facilitate shape representation and categorization by extracting features that correlate with global shape. We have established a connection between a computer model of a recognition system and known biological phenomena.

Chapter 8

Future Directions

8.1 Open Issues

Our research has left us with some open issues and these are elaborated upon in great detail in the Discussion sections of Chapters 3, 4, 5 and 6. Some of the most important are given below.

It remains unknown whether and how a full global description of object shape emerges, but top-down effects seem poised to play a critical role. More specifically, or perhaps less ambitiously, we consider the transformation from something more than a distributed coding (in terms of constituent parts) in posterior IT to a sparse coding (in terms of global object shape) in anterior IT. Perhaps this can only be answered in terms of the neural

binding hypothesis, requiring synchronous oscillations in neuronal ensembles to integrate and bind neurons that represent different features of an object (Singer and Gray, 1995). Also, we question which of an image's iso-curvature segments on a contour, and subsequently which of the Gaussian constituent terms in a V4 cell's total response equation, are reflected in IT. We have seen that it is not necessary in our models to include all iso-curvature segments, but if not all are used in biological systems then how is the selection made and does this mechanism evolve over time?

We would like to compare and contrast categorization and identification, providing a link to the Kanwisher data (Grill-Spector and Kanwisher, 2005). A key question is whether parts-based recognition is more important at the categorization or identification stage and by how much of a factor (compared to the whole object). We ask if category-level segmentation is parts-based and if individual-level identification is based on the whole object. We consider whether the number, size, relative position independence, or refinement of parts increases from categorization to identification. If possible, we would caste the distributed vs. sparse representation in terms of categorization and identification. Also, we question whether scale, rotation and translation invariance are affected by these issues?

Finally, we would like to have testable predictions from and experimental verification of our computational models. Our construction methodology allows us to create model cells that are selective for a particular local shape configuration at a particular location on a contour within a larger shape, as well as to create model cells that nonlinearly integrate

specific information about the 2-dimensional boundary shapes of multiple contour fragments. If these cells were constructed to match biological cells like those studied by Pasupathy and Connor (2001) in V4 and Brincat and Connor (2004) in IT then our models could make predictions about the responses of the actual cells to novel gray level patterns presented to the macaques, for example, allowing us to further explore shape representation in V4 and IT.

8.2 Next Steps

Our immediate future research has also been outlined in the Discussion sections of Chapters 3, 4, 5 and 6. Our highest priority would be to augment our Izhikevich cell network model with sub-populations of IT, each with its own set of inhibitory interneurons. Given this, we could confirm our hypotheses by presenting real stimuli to the network and definitively showing that response amplification in IT corresponds with classification – possibly extending this to show that neuronal synchronization correlates with visual salience. Some future steps that have not been discussed are mentioned below.

We would like to develop phase portraits of our Izhikevich cells and network models (Strogatz, 1994). This would allow us to visualize the nonlinear dynamics of the network and facilitate parameter tuning, experiment design, etc.

We would like to explore the use of some advanced computer vision techniques, such as the inner-distance measurement (Ling and Jacobs, 2007), possibly in conjunction with curvature sensitivities, within our recognition system. Also we would like to consider the latest incarnation of the earth mover's distance metric, computed in linear time using wavelets (Shirdhonkar and Jacobs, 2008).

We have only considered simple, 2-dimensional curves. We would like to experiment with fractals and self-similarity at all scales (Mandelbrot, 1977), relevant if only that these entities abound in nature. What would be the consequences in V4 and IT and in our computer models, for example, if the Koch curve shown in Figure 8.1 was presented as a stimulus? Also, we would like to study occluded or illusory stimuli, such as some well-known Kanizsa images (Kanizsa, 1979).

With the techniques, methodologies and experience gained from working in the object recognition domain, we would like to explore the possibility of transferring some of our knowledge to some emerging higher-level cognitive Brain-Computer Interface (BCI) applications, possibly within a general-purpose system (Schalk, 2009), possibly in an area that is somewhat related to our current work, such as the detection of class-specific visual stimuli (Miller et al., 2009). We are encouraged by the availability of the next generation of mechanically flexible electronics for multiplexed signal measurements (Viventi *et al*., 2010). Extending our experience to another domain, for instance epilepsy, either through mathematical modeling (Lytton, 2008) or signal analysis, is also appealing and important, with 60 million epilepsy patients worldwide and one-third refractory to medication. We

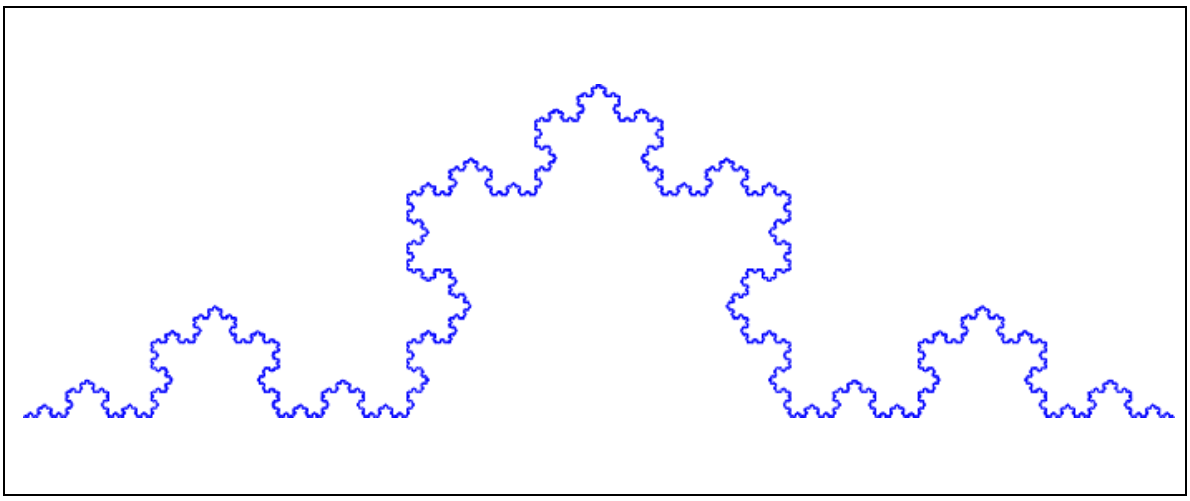

Figure 8.1 – Koch curve. The standard Koch curve after 8 iterations (i.e., 4^8 = 65,536 segments). (Courtesy of Alexander M. Murphy.)

might also consider the data available from deep brain stimulation EEG recordings acquired as images were presented to patients.

The analysis of some recent electrocorticographic (ECoG) recordings, which measure synaptic activity across neuronal populations (Crone *et al*., 2006; Logothetis *et al*., 2001), is intriguing and may have implications for our research. Jacobs and colleagues (Jacobs *et al*., 2007; Jacobs *et al*., 2009) have found ECoG activity in the gamma band that may be linked to cognitive representations of stimuli. Intracranial brain recordings of human neurosurgical patients observing visually presented alphabetic letters revealed that the gamma band ECoG amplitude increased at many electrodes after a letter was presented, particularly in the occipital and temporal regions. The Jacobs group also found some coding of shape information for the letters. For example, one occipital electrode had significantly elevated gamma amplitudes for the rounder letters "C", "D" and "G", but not for the straighter letters "N" and "K". This, as well as our synchronization results (Chapter 6), compels us to consider if the curvature sensitivities of cells found in V4 and IT (Pasupathy and Connor, 2001; Brincat and Connor, 2004) are involved in, or modulated by, these gamma-related occurrences. Furthermore, the precise amplitudes of the gamma activities could often be used to identify the letter that was presented. These letter-specific patterns occurred during periods of increased gamma activity and were limited to regions with these elevations. They were linked to the phase of simultaneous theta oscillations and emerged ≈ 50 – 100 ms after the overall gamma power increased. This offset is of the same order of magnitude as the delay between categorization and identification (Grill-Spector, 2003; Grill-Spector and Kanwisher, 2005) in the object

recognition process as well as the average lag between the linear and nonlinear response components of the IT cells that exhibited mixed selectivities for multipart configurations (Brincat and Connor, 2006) (as well as the ~60 additional ms required to attend to a feature (compared to attending to a location) (Reynolds and Desimone, 2003)). We would like to explore the relationship, if any, between these three similar time lapses and investigate the correlations between specific recognition events and the onset of stimulus patterns. In individual patients, when the Jacobs group saw two proximal electrodes with letter-specific activity, the patterns often occurred at similar time points of peak letter specificity. Finally, we would like to investigate whether this is related to the columnar arrangement of V4 and / or IT, with possible considerations of the "pinwheel" topography and its relationship to short-range connections found by Gilbert and colleagues in primary visual cortex (Das and Gilbert, 1999), yet distinct from the retinotopic organization seen in lower areas.

In general, we seek the formation of and answers to the overarching questions in neuroscience.

Bibliography

Adee, S. (2009). http://spectrum.ieee.org/tech-talk.

Adelson, E.H. and Bergen, J.R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*. 2(2), 284-299.

Alter, T.D. and Basri, R. (1996). Extracting salient contours from images: an analysis of the saliency network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Amir, Y., Harel, M., and Malach, R. (1993). Cortical hierarchy reflected in the organization of intrinsic connections in macaque monkey visual cortex. *The Journal of Comparative Neurology*. 334(1), 19-46.

Amit, Y., Geman, D., and Wilder, K. (1997). Joint Induction of Shape Features and Tree Classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 19(11), 1300-1305.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*. 61(3), 183-193.

Baker, C.I., Behrmann, M., and Olson, C.R. (2002). Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nature Neuroscience*. 5(11), 1210-1216.

Barenholtz, E. and Feldman, J. (2003). Visual comparisons within and between object parts: evidence for a single-part superiority effect. *Vision Research*. 43(15), 1655-1666.

Basri, R., Costa, L., Geiger, D., and Jacobs, D. (1998). Determining the similarity of deformable shapes. *Vision Research*. 38(15-16), 2365-2385.

Bazhenov, M., Stopfer, M., Rabinovich, M., Huerta, R., Abarbanel, H.D.I., Sejnowski, T.J., and Laurent, G. (2001). Model of Transient Oscillatory Synchronization in the Locust Antennal Lobe. *Neuron*. 30(2), 553-567.

Bazhenov, M., Stopfer, M., Rabinovich, M., Abarbanel, H.D.I., Sejnowski, T.J., and Laurent, G. (2001). Model of Cellular and Network Mechanisms for Odor-Evoked Temporal Patterning in the Locust Antennal Lobe. *Neuron*. 30(2), 569-581.

Beierlein, M., Gibson, J.R., and Connors, B.W. (2000). A network of electrically coupled interneurons drives synchronized inhibition in neocortex. *Nature Neuroscience*. 3(9), 904-910.

Belongie, S., Malik, J., and Puzicha, J. (2002). Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 24(24), 509-522.

Bergevin, R. and Bubel, A. (2003). Object-level structured contour map extraction. *Computer Vision and Image Understanding*. 91(3), 302-334.

Bertamini, M. and Lawson, R. (2006). Visual search for a circular region perceived as a figure versus as a hole: Evidence of the importance of part structure. *Perception and Psychophysics*. 68(5), 776-791.

Bichot, N.P., Rossi, A.F., and Desimone, R. (2005). Parallel and Serial Neural Mechanisms for Visual Search in Macaque Area V4. *Science*. 308(5721), 529-534.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*. 94(2), 115-147.

Bishop, C.M. (2006). **Pattern Recognition and Machine Learning**. New York, NY: Springer Science+Business Media, Inc.

Booth, M.C.A. and Rolls, E.T. (1998). View-invariant Representations of Familiar Objects by Neurons in the Inferior Temporal Visual Cortex. *Cerebral Cortex*. 8(6):510-523.

Borenstein, E. and Ullman, S. (2002). Class-Specific, Top-Down Segmentation. *Proceedings of the 7th European Conference on Computer Vision-Part II*. pp.109-124.

Borg, I. and Groenen, P.J.F. (2005). **Modern Multidimensional Scaling: Theory and Applications** (second edition). New York, NY: Springer Science+Business Media, Inc.

Borst, A. and Theunissen, F.E. (1999). Information theory and neural coding. *Nature Neuroscience*. 2(11):947-957.

Breiman, L., Friedman, J., Olshen, R.A., and Stone, C.J. (1984). **Classification and Regression Trees**. Boca Raton, FL: CRC Press.

Brincat, S.L. and Connor, C.E. (2004). Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nature Neuroscience*. 7(8), 880-886.

Brincat, S.L. and Connor, C.E. (2006). Dynamic Shape Synthesis in Posterior Inferotemporal Cortex. *Neuron*. 49(1), 17-24.

Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews*. 36(2-3), 96-107.

Burbeck, C.A. and Pizer, S.M. (1995). Object Representation by Cores: Identifying and Representing Primitive Spatial Regions. *Vision Research*. 35(13), 1917-1930.

Casile, A. and Giese, M.A. (2005). Critical features for the recognition of biological motion. *Journal of Vision*. 5(4), 348-360.

Chien, C.-H. and Aggarwal, J.K. (1989). Model Construction and Shape Recognition from Occluding Contours. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 11(4), 372-389.

Cohen, E.H., Barenholtz, E., Singh, M., and Feldman, J. (2005). What change detection tells us about the visual representation of shape. *Journal of Vision*. 5(4), 313-321.

Coleman, T.F. and Li, Y. (1996). An Interior, Trust Region Approach for Nonlinear Minimization Subject to Bounds. Society for Industrial and Applied Mathematics Journal on Optimization. 6(2), 418-445.

Connor, C.E., Preddie, D.C., Gallant, J.L., and Van Essen, D.C. (1997). Spatial Attention Effects in Macaque Area V4. *The Journal of Neuroscience*. 17(9), 3201-3214.

Cox, T.F. and Cox, M.A.A. (2001). **Multidimensional Scaling** (second edition). Boca Raton, FL: Chapman and Hall/CRC.

Crone, N.E., Sinai, A., and Korzeniewska, A. (2006). High-frequency gamma oscillations and human brain mapping with electrocorticography. *Progress in Brain Research*. 159, 275-295.

Das, A. and Gilbert, C.D. (1999). Topography of contextual modulations mediated by short-range interactions in primary visual cortex. *Nature*. 399(6737), 655-661.

Deco, G. and Rolls, E.T. (2004). A Neurodynamical cortical model of visual attention and invariant object recognition. *Vision Research*. 44(6), 621-642.

Dennis, J.E., Jr. (1977). Nonlinear Least-Squares. In D.A.H. Jacobs (Ed.), **The State of the Art in Numerical Analysis**. London: Academic Press, pp.269-312.

Desimone, R. and Schein, S.J. (1987). Visual Properties of Neurons in Area V4 of the Macaque: Sensitivity to Stimulus Form. *The Journal of Neurophysiology*. 57(3), 835-868.

De Winter, J. and Wagemans, J. (2004). Contour-based object identification and segmentation: Stimuli, norms and data, and software tools. *Behavior Research Methods, Instruments, and Computers*. 36(4), 604-624.

DiCarlo, J.J. and Maunsell, J.H.R. (2000). Form representation in monkey inferotemporal cortex is virtually unaltered by free viewing. *Nature Neuroscience*. 3(8), 814-821.

Dobbins, A., Zucker, S.W., and Cynader, M.S. (1989). Endstopping and curvature. *Vision Research*. 29(10), 1371-1387.

Dosher, B.A. and Lu, Z.-L. (1998). Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. *Proceedings of The National Academy of Sciences USA*. 95(23), 13988-13993.

Duda, R.O. and Hart, P.E. (1973). **Pattern Classification and Scene Analysis**. New York, NY: John Wiley & Sons, Inc.

Dudek, G. and Tsotsos, J.K. (1997). Shape Representation and Recognition from Multiscale Curvature. *Computer Vision and Image Understanding*. 68(2),170-189.

Ekstrom, A.D., Kahana, M.J., Caplan, J.B., Fields, T.A., Isham, E.A., Newman, E.L., and Fried, I. (2003). Cellular networks underlying human spatial navigation. *Nature*. 425(6954), 184-187.

Elder, J.H. and Goldberg, R.M. (2002). Ecological statistics of Gestalt laws for the perceptual organization of contours. *Journal of Vision*. 2(4):5, 324-353.

Elder, J.H., Krupnik, A., and Johnston, L.A. (2003). Contour Grouping with Prior Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 25(6), 661-674.

Engel, A.K., Fries, P., and Singer, W. (2001). Dynamic Predictions: Oscillations and Synchrony in Top-Down Processing. *Nature Reviews Neuroscience*. 2(10), 704-716.

Fall, C.P. and Keizer, J.E. (2002). Voltage Gated Ionic Currents. In C.P. Fall, E.S. Marland, J.M. Wagner and J.J. Tyson (Eds.), **Computational Cell Biology**. New York, NY: Springer Publishing Company, pp.21-52.

Feldman, J. (2001). Bayesian contour integration. *Perception and Psychophysics*. 63(7), 1171-1182.

Feldman, J. and Singh, M. (2005). Information Along Contours and Object Boundaries. *Psychological Review*. 112(1), 243-252.

Felleman, D.J. and Van Essen, D.C. (1987). Receptive Field Properties of Neurons in Area V3 of Macaque Monkey Extrastriate Cortex. *The Journal of Neurophysiology*. 57(4), 889-920.

Felleman, D.J. and Van Essen, D.C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*. 1(1), 1-47.

Finkel, L.H., Yen, S.-C., and Menschik, E.D. (1998). Synchronization: The Computational Currency of Cognition. *ICANN 98, Proceedings of the 8th International Conference on Artificial Neural Networks*. pp.1-18.

FitzHugh, R. (1961). Impulses and Physiological States in Theoretical Models of Nerve Membrane. *Biophysical Journal*. 1(6), 445-466.

Freedman, D.J., Riesenhuber, M., Poggio, T., and Miller, E.K. (2003). A Comparison of Primate Prefrontal and Inferior Temporal Cortices during Visual Categorization. *The Journal of Neuroscience*. 23(12):5235-5246.

Freeman, W.T. and Adelson, E.H. (1991). The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 13(9), 891-906.

Freund, Y. and Schapire, R.E. (1999). A Short Introduction to Boosting. *Journal of Japanese Society for Artificial Intelligence*. 14(5), 771-780.

Fries, P. (2009). Neuronal Gamma-Band Synchronization as a Fundamental Process in Cortical Computation. *Annual Review of Neuroscience*. 32, 209-224.

Fujita, I., Tanaka, K., Ito, M., and Cheng, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. *Nature*. 360(6402):343-346.

Gallant, J.L., Connor, C.E., Rakshit, S., Lewis, J.W., and Van Essen, D.C. (1996). Neural Responses to Polar, Hyperbolic, and Cartesian Gratings in Area V4 of the Macaque Monkey. *The Journal of Neurophysiology*. 76(4), 2718-2739.

Gasparini, S. and Magee, J.C. (2006). State-Dependent Dendritic Computation in Hippocampal CA1 Pyramidal Neurons. *The Journal of Neuroscience*. 26(7), 2088-2100.

Giese, M.A. and Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience*. 4(3), 179-192.

Giocomo, L.M. and Hasselmo, M.E. (2008). Computation by Oscillations: Implications of Experimental Data for Theoretical Models of Grid Cells. *Hippocampus*. 18(12), 1186-1199.

Gold, J.I. and Shadlen, M.N. (2000). Representation of a perceptual decision in developing oculomotor commands. *Nature*. 404(6776), 390-394.

Gold, J.I. and Shadlen, M.N. (2001). Neural computations that underlie decisions about sensory stimuli. *TRENDS in Cognitive Sciences*. 5(1), 10-16.

Goldberg, D.E. (1989). **Genetic Algorithms in Search, Optimization & Machine Learning**. Boston, MA: Addison-Wesley Longman Publishing Co., Inc.

Gray, A. (1997). **Modern Differential Geometry of Curves and Surfaces with MATHEMATICA** (second edition). Boca Raton, FL: CRC Press.

Gray, C.M., König, P., Engel, A.K., and Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*. 338(6213), 334-337.

Gray, C.M. and Singer, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proceedings of The National Academy of Sciences USA*. 86(5), 1698-1702.

Gray, C.M. and McCormick, D.A. (1996). Chattering cells: superficial pyramidal neurons contributing to the generation of synchronous oscillations in the visual cortex. *Science*. 274(5284), 109-113.

Grigorescu, C. and Petkov, N. (2003). Distance Sets for Shape Filters and Shape Recognition. *IEEE Transactions on Image Processing*. 12(10), 1274-1286.

Grill-Spector, K. (2003). The functional organization of the ventral visual pathway and its relationship to object recognition. In N. Kanwisher and J. Duncan (Eds.), **Attention and Performance XX: Functional Brain Imaging of Visual Cognition**. London: Oxford University Press, pp.169-193.

Grill-Spector, K. and Kanwisher, N. (2005). Visual Recognition: As Soon as You Know It Is There, You Know What It Is. *Psychological Science*. 16(2), 152-160.

Gross, C.G., Rocha-Miranda, C.E., and Bender, D.B. (1972). Visual Properties of Neurons in Inferotemporal Cortex of the Macaque. *The Journal of Neurophysiology*. 35(1):96-111.

Grossberg, S. (1999). The Link between Brain Learning, Attention, and Consciousness. *Consciousness and Cognition*. 8(1), 1-44.

Harel, A., Ullman, S., Epshtein, B., and Bentin, S. (2007). Mutual information of image fragments predicts categorization in humans: Electrophysiological and behavioral evidence. *Vision Research*. 47(15), 2010-2020.

Hegdé, J. and Van Essen, D.C. (2000). Selectivity for Complex Shapes in Primate Visual Area V2. *The Journal of Neuroscience*. 20(RC61), 1-6.

Hegdé, J. and Van Essen, D.C. (2003). Strategies of shape representation in macaque visual area V2. *Visual Neuroscience*. 20(3), 313-328.

Hinkle, D.A. and Connor, C.E. (2001). Disparity tuning in macaque area V4. *NeuroReport*. 12(2), 365-369.

Hinkle, D.A. and Connor, C.E. (2002). Three-dimensional orientation tuning in macaque area V4. *Nature Neuroscience*. 5(7), 665-670.

Hochstein, S. and Ahissar, M. (2002). View from the Top: Hierarchies and Reverse Hierarchies in the Visual System. *Neuron*. 36(5), 791-804.

Hoffman, D. and Richards, W. (1984). Parts of recognition. *Cognition*. 18(1-3), 65-96.

Hoffman, D.D. and Singh, M. (1997). Salience of visual parts. *Cognition*. 63(1), 29-78.

Hopfield, J.J. and Brody, C.D. (2001). What is a moment? Transient synchrony as a collective mechanism for spatiotemporal integration. *Proceedings of The National Academy of Sciences USA*. 98(3), 1282-1287.

Howard, M.W., Rizzuto, D.S., Caplan, J.B., Madsen, J.R., Lisman, J., Aschenbrenner-Scheibe, R., Schulze-Bonhage, A., and Kahana, M.J. (2003). Gamma Oscillations Correlate with Working Memory Load in Humans. *Cerebral Cortex*. 13(12), 1369-1374.

Izhikevich, E.M. (2003). Simple Model of Spiking Neurons. *IEEE Transactions on Neural Networks*. 14(6), 1569-1572.

Izhikevich, E.M. (2004). Which Model to Use for Cortical Spiking Neurons? *IEEE Transactions on Neural Networks*. 15(5), 1063-1070.

Izhikevich, E.M. (2007). **Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting**. Cambridge, MA: The MIT Press.

Izhikevich, E.M. and Edelman, G.M. (2008). Large-scale model of mammalian thalamocortical systems. *Proceedings of The National Academy of Sciences*. 105(9), 3593-3598.

Jacobs, J., Kahana, M.J., Ekstrom, A.D., and Fried, I. (2007). Brain Oscillations Control Timing of Single-Neuron Activity in Humans. *The Journal of Neuroscience*. 27(14), 3839-3844.

Jacobs, J. and Kahana, M.J. (2009). Neural Representations of Individual Stimuli in Humans Revealed by Gamma-Band Electrocorticographic Activity. *The Journal of Neuroscience*. 29(33), 10203-10214.

Jacobs, R.A., Jordan, M.I., Nowlan, S.J., and Hinton, G.E. (1991). Adaptive Mixtures of Local Experts. *Neural Computation*. 3(1):79-87.

Jeannin, S. and Bober, M. (1999). Description of Core Experiments for MPEG-7 Motion/Shape.
Technical Report ISO/IEC JTC 1/SC 29/WG 11 MPEG99/N2690, MPEG-7. Seoul.

Jolliffe, I.T. (2002). **Principal Component Analysis** (second edition). New York, NY: Springer-Verlag New York, Inc.

Jordan, M.I. and Jacobs, R.A. (1994). Hierarchical Mixtures of Experts and the EM Algorithm. *Neural Computation*. 6(2):181-214.

Kahana, M.J. and Sekuler, R. (2002). Recognizing spatial patterns: a noisy exemplar approach. *Vision Research*. 42(18), 2177-2192.

Kanizsa, G. (1979). **Organization in Vision: Essays on Gestalt Perception**. Santa Barbara, CA: Praeger Publishers.

Kellman, P.J. (2003). Interpolation processes in the visual perception of objects. *Neural Networks*. 16(5-6), 915-923.

Kiana, R., Esteky, H., Mirpour, K., and Tanaka, K. (2007). Object Category Structure in Response Patterns of Neuronal Population in Monkey Inferior Temporal Cortex. *Journal of Neurophysiology*. 97(6), 4296-4309.

Kobatake, E. and Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *The Journal of Neurophysiology*. 71(3), 856-867.

König, P., Engel, A.K., and Singer, W. (1996). Integrator or coincidence detector? The role of the cortical neuron revisited. *Trends in Neuroscience*. 19(4), 130-137.

Kovács, I. and Julesz, B. (1993). A closed curve is much more than an incomplete one: effect of closure in figure-ground segmentation. *Proceedings of The National Academy of Sciences USA*. 90(16), 7495-7497.

Kovács, G., Sáry, G., Köteles, K., Chadaide, Z., Tompa, T., Vogels, R., and Benedek, G. (2003). Effects of Surface Cues on Macaque Inferior Temporal Cortical Responses. *Cerebral Cortex*. 13(2), 178-188.

Kramer, D. and Fahle, M. (1996). A Simple Mechanism for Detecting Low Curvatures. *Vision Research*. 36(10), 1411-1419.

Kriegeskorte, N., Mur, M., Ruff, D.A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., and Bandettini, P.A. (2008). Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. *Neuron*. 60(6), 1126-1141.

Laurent, G., Stopfer, M., Friedrich, R.W., Rabinovich, M.I., Volkovskii, A., and Abarbanel, H.D.I. (2001). Odor Encoding as an Active, Dynamical Process: Experiments, Computation, and Theory. *Annual Review of Neuroscience*. 24, 263-297.

Lazarewicz, M.T., Ehrlichman, R.S., Maxwell, C.R., Gandal, M.J., Finkel, L.H., and Siegel, S.J. (2009). Ketamine Modulates Theta and Gamma Oscillations. *Journal of Cognitive Neuroscience*. 22(7), 1452-1464.

LeCun, Y., Jackel, L.D., Bottou, L., Brunot, A., Cortes, C., Denker, J.S., Drucker, H., Guyon, I., Müller, U.A., Säckinger, E., Simard, P., and Vapnik, V. (1995). Comparison of learning algorithms for handwritten digit recognition. *International Conference on Artificial Neural Networks*. pp. 53-60.

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*. 86(11), 2278-2324.

Lee, T.S. and Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A*. 20(7), 1434-1448.

Leyton, M. (1989). Inferring Causal History from Shape. *Cognitive Science*. 13(3), 357-387.

Ling, H. and Jacobs, D.W. (2007). Shape Classification Using the Inner-Distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 29(2), 286-299.

Liu, X.-Y., Wu, J., and Zhou, Z.-H. (2009). Exploratory Undersampling for Class-Imbalance Learning. *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*. 39(2):539-550.

Logothetis, N.K., Pauls, J., and Poggio, T. (1995). Shape Representation in the Inferior Temporal Cortex of Monkeys. *Current Biology*. 5(5), 552-563.

Logothetis, N.K., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*. 412(6843), 150-157.

Losonczy, A. and Magee, J.C. (2006). Integrative Properties of Radial Oblique Dendrites in Hippocampal CA1 Pyramidal Neurons. *Neuron*. 50(2), 291-307.

Losonczy, A., Makara, J.K., and Magee, J.C. (2008). Compartmentalized dendritic plasticity and input feature storage in neurons. *Nature*. 452(7186), 436-441.

Lytton, W.W. (2008). Computer modelling of epilepsy. *Nature Reviews Neuroscience*. 9(8), 626-637.

Mandelbrot, B.B. (1977). **The Fractal Geometry of Nature**. New York, NY: W.H. Freeman and Company.

Marr, D. (1982). **Vision**. San Francisco, CA: H. Freeman and Co.

MATLAB: The MathWorks, Inc., 3 Apple Hill Drive, Natick, MA 01760, USA.

Mazer, J.A. and Gallant, J.L. (2003). Goal-Related Activity in V4 during Free Viewing Visual Search: Evidence for a Ventral Stream Visual Salience Map. *Neuron*. 40(6), 1241-1250.

Mazurowski, M.A., Habas, P.A., Zurada, J.M., Lo, J.Y., Baker, J.A., and Tourassi, G.D. (2008). Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance. *Neural Networks*. 21(2-3):427-436.

McAdams, C.J. and Maunsell, J.H.R. (2000). Attention to Both Space and Feature Modulates Neuronal Responses in Macaque Area V4. *The Journal of Neurophysiology*. 83(3), 1751-1755.

McKeefry, D.J. and Zeki, S. (1997). The position and topography of the human colour centre as revealed by functional magnetic resonance imaging. *Brain*. 120(12), 2229-2242.

McMahon, D.B.T. and Olson, C.R. (2009). Linearly Additive Shape and Color Signals in Monkey Inferotemporal Cortex. *Journal of Neurophysiology*. 101(4), 1867-1875.

Miller, K.J., Hermes, D., Schalk, G., Ramsey, N.F., Jagadeesh, B., den Nijs, M., Ojemann, J.G., and Rao, R.P.N. (2009). Detection of spontaneous class-specific visual stimuli with high temporal accuracy in human electrocorticography. *Proceedings of the 31st Annual International Conference of the IEEE EMBS*. pp.6465-6468.

Miltner, W.H.R., Braun, C., Arnold, M., Witte, H., and Taub, E. (1999). Coherence of gamma-band EEG activity as a basis for associative learning. *Nature*. 397(6718), 434-436.

Mitchell, T.M. (1997). **Machine Learning**. New York, NY: McGraw-Hill Science/Engineering/Math.

Mokhtarian, F. and Mackworth, A.K. (1986). Scale-Based Description and Recognition of Planar Curves and Two-Dimensional Shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 8(1), 34-43.

Mokhtarian, F. and Mackworth, A.K. (1992). A Theory of Multiscale, Curvature-Based Shape Representation for Planar Curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 14(8), 789-805.

Mokhtarian, F. (1995). Silhouette-Based Isolated Object Recognition through Curvature Scale Space. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 17(5), 539-544.

Morris, C. and Lecar, H. (1981). Voltage oscillations in the barnacle giant muscle fiber. *Biophysical Journal*. 35(1), 193-213.

Motter, B.C. (1994). Neural correlates of attentive selection for color or luminance in extrastriate area V4. *The Journal of Neuroscience*. 14(4), 2178-2189.

Motter, B.C. (2003). The cortical magnification factor for area V4. *Journal of Vision*. 3(9), 110a.

Müller, K.-R., Mika, S., Rätsch, G., Tsuda, K., and Schölkopf, B. (2001). An Introduction to Kernel-Based Learning Algorithms. *IEEE Transactions on Neural Networks*. 12(2), 181-201.

Murphy, T.M., Matlin, M., and Finkel, L.H. (2003). Curvature Covariation as a Factor in Perceptual Salience. *Proceedings of the 1st International IEEE EMBS Conference on Neural Engineering*. pp.16-19.

Murphy, T.M. and Finkel, L.H. (2007). Shape representation by a network of V4-like cells. *Neural Networks*. 20(8), 851-867.

Norcia, A.M., Pei, F., Bonneh, Y., Hou, C., Sampath, V., and Pettet, M.W. (2005). Development of sensitivity to texture and contour information in the human infant. *Journal of Cognitive Neuroscience*. 17(4), 569-579.

Norman, J.F., Phillips, F., and Ross, H.E. (2001). Information concentration along the boundary contours of naturally shaped solid objects. *Perception*. 30(11), 1285-1294.

Op de Beeck, H., Wagemans, J., Vogels, R. (2001). Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. *Nature Neuroscience*. 4(12):1244-1252.

Op de Beeck, H.P., Deutsch, J.A., Vanduffel, W., Kanwisher, N.G., and DiCarlo, J.J. (2008). A Stable Topography of Selectivity for Unfamiliar Shape Classes in Monkey Inferior Temporal Cortex. *Cerebral Cortex*. 18(7), 1676-1694.

Palmeri, T.J. and Nosofsky, R.M. (2001). Central tendencies, extreme points, and prototype enhancement effects in ill-defined perceptual categorization. *The Quarterly Journal of Experimental Psychology*. 54A(1):197-235.

Pasupathy, A. and Connor, C.E. (1999). Responses to Contour Features in Macaque Area V4. *The Journal of Neurophysiology*. 82(5), 2490-2502.

Pasupathy, A. and Connor, C.E. (2001). Shape representation in area V4: position-specific tuning for boundary conformation. *The Journal of Neurophysiology*. 86(5), 2505-2519.

Pasupathy, A. and Connor, C.E. (2002). Population coding of shape in area V4. *Nature Neuroscience*. 5(12), 1332-1338.

Pettet, M.W. (1999). Shape and contour detection. *Vision Research*. 39(3), 551-557.

Poirazi, P., Brannon, T., and Mel, B.W. (2003). Arithmetic of Subthreshold Synaptic Summation in a Model CA1 Pyramidal Cell. *Neuron*. 37(6), 977-987.

Poirazi, P., Brannon, T., and Mel, B.W. (2003). Pyramidal Neuron as Two-Layer Neural Network. *Neuron*. 37(6), 989-999.

Polsky, A., Mel, B.W., and Schiller, J. (2004). Computational subunits in thin dendrites of pyramidal cells. *Nature Neuroscience*. 7(6), 621-627.

Popescu, A.T., Popa, D., and Paré, D. (2009). Coherent gamma oscillations couple the amygdala and striatum during learning. *Nature Neuroscience*. 12(6), 801-807.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P. (2007). **Numerical Recipes: The Art of Scientific Computing** (third edition). New York, NY: Cambridge University Press.

Quiroga, R.Q., Reddy, L., Kreiman, G., Kock, C., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*. 435(7045), 1102-1107.

Rabinovich, M., Volkovskii, A., Lecanda, P., Huerta, R., Abarbanel, H.D.I., and Laurent, G. (2001). Dynamical Encoding by Networks of Competing Neuron Groups: Winnerless Competition. *Physical Review Letters*. 87(6), 1-4.

Rao, R.P.N. (2004). Bayesian computation in recurrent neural circuits. *Neural Computation*. 16(1):1-38.

Rao, R.P.N. (2005). Bayesian inference and attentional modulation in the visual cortex. *NeuroReport*. 16(16):1843-1848.

Rao, R.P.N. (2005). Hierarchical Bayesian Inference in Networks of Spiking Neurons. *Advances in Neural Information Processing Systems*. vol. 17, pp.1-8.

Reynolds, J.H. and Desimone, R. (2003). Interacting Roles of Attention and Visual Salience in V4. *Neuron*. 37(5), 853-863.

Richards, W. and Hoffman, D. (1985). Codon Constraints on Closed 2D Shapes. *Computer Vision, Graphics, and Image Processing*. 31(2), 207-223.

Riesenhuber, M. and Poggio, T. (1999). Are Cortical Models Really Bound by the "Binding Problem"? *Neuron*. 24(1), 87-93.

Riesenhuber, M. and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*. 2(11), 1019-1025.

Riesenhuber, M. and Poggio, T. (2003). How Visual Cortex Recognizes Objects: The Tale of the Standard Model. In L.M. Chalupa and J.S. Werner (Eds.), **The Visual Neurosciences**. Cambridge, MA: MIT Press, vol. 2, pp.1640-1653.

Rinzel, J. and Ermentrout, B. (1998). Analysis of Neural Excitability and Oscillations. In C. Koch and I. Segev (Eds.), **Methods in Neuronal Modeling** (second edition). Cambridge, MA: MIT Press, pp.251-291.

Rodriguez, E., George, N., Lachaux, J.-P., Martinerie, J., Renault, B., and Varela, F.J. (1999). Perception's shadow: long-distance synchronization of human brain activity. *Nature*. 397(6718), 430-433.

Rolls, E.T., Treves, A., and Tovee, M.J. (1997). The representational capacity of the distributed encoding of information provided by populations of neurons in primate temporal visual cortex. *Experimental Brain Research*. 114(1), 149-162.

Rosenholtz, R. (1999). A simple saliency model predicts a number of motion popout phenomena. *Vision Research*. 39(19), 3157-3163.

Roweis, S.T. and Saul, L.K. (2000). Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science*. 290(5500), 2323-2326.

Rubner, Y., Tomasi, C., and Guibas, L.J. (2000). The Earth Mover's Distance as a Metric for Image Retrieval. *International Journal of Computer Vision*. 40(2), 99-121.

Rubner, Y., Puzicha, J., Tomasi, C., and Buhmann, J.M. (2001). Empirical Evaluation of Dissimilarity Measures for Color and Texture. *Computer Vision and Image Understanding*. 84(1), 25-43.

Sajda, P. and Finkel, L.H. (1995). Intermediate-Level Visual Representations and the Construction of Surface Perception. *Journal of Cognitive Neuroscience*. 7(2), 267-291.

Schalk, G. (2009). Effective Brain-Computer Interfacing Using BCI2000. *Proceedings of the 31st Annual International Conference of the IEEE EMBS*. pp.5498-5501.

Schiller, P.H. (1993). The effects of V4 and middle temporal (MT) area lesions on visual performance in the rhesus monkey. *Visual Neuroscience*. 10(4), 717-46.

Schneiderman, H. and Kanade, T. (2004). Object Detection Using the Statistics of Parts. *International Journal of Computer Vision*. 56(3), 151-177.

Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J., and Williamson, R.C. (2001). Estimating the Support of a High-Dimensional Distribution. *Neural Computation*. 13(7), 1443-1471.

Sebastian, T.B., Klein, P.N., and Kimia, B.B. (2001). Recognition of Shapes by Editing Shock Graphs. *Proceedings of the 8th International Conference on Computer Vision*. pp. 755-762.

Sebastian, T.B., Klein, P.N., and Kimia, B.B. (2003). On Aligning Curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 25(1), 116-124.

Sebastian, T.B. and Kimia, B.B. (2005). Curves vs. skeletons in object recognition. *Signal Processing*. 85(2), 247-263.

Sederberg, P.B., Kahana, M.J., Howard, M.W., Donner, E.J., and Madsen, J.R. (2003). Theta and Gamma Oscillations during Encoding Predict Subsequent Recall. *The Journal of Neuroscience*. 23(34), 10809-10814.

Sereno, M.E., Trinath, T., Augath, M., and Logothetis, N.K. (2002). Three-Dimensional Shape Representation in Monkey Cortex. *Neuron*. 33(4), 635-652.

Shashua, A. and Ullman, S. (1988). Structural saliency. *Proceedings of the International Conference on Computer Vision*. 482-488.

Shi, J. and Malik, J. (2000). Normalized Cuts and Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 22(8), 888-905.

Shirdhonkar, S. and Jacobs, D.W. (2008). Approximate earth mover's distance in linear time. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Siddiqi, K. and Kimia, B.B. (1995). Parts of Visual Form: Computational Aspects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 17(3), 239-251.

Siddiqi, K., Tresness, K.J., and Kimia, B.B. (1996). Parts of visual form: psychophysical aspects. *Perception*. 25(4), 399-424.

Sigala, N., Gabbiani, F., and Logothetis, N.K. (2002). Visual Categorization and Object Representation in Monkeys and Humans. *Journal of Cognitive Neuroscience*. 14(2), 187-198.

Sigala, N. and Logothetis, N.K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature*. 415(6869), 318-320.

Simard, P., LeCun, Y., and Denker, J. (1993). Efficient Pattern Recognition Using a New Transformation Distance. *Advances in Neural Information Processing Systems*. vol. 5, pp. 50-58.

Simoncelli, E.P. and Olshausen, B.A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*. 24, 1193-1216.

Singer, W. and Gray, C.M. (1995). Visual Feature Integration and the Temporal Correlation Hypothesis. *Annual Review of Neuroscience*. 18, 555-586.

Song, Y., Goncalves, L., and Perona, P. (2003). Unsupervised Learning of Human Motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 25(7), 814-827.

Stocker, A.A. and Simoncelli, E.P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*. 9(4):578-585.

Strogatz, S.H. (1994). **Nonlinear Dynamics and Chaos**. Reading, MA: Addison-Wesley Publishing Company.

Supp, G.G., Schlögl, A., Trujillo-Barreto, N., Müller, M.M., and Gruber, T. (2007). Directed Cortical Information Flow during Human Object Recognition: Analyzing Induced EEG Gamma-Band Responses in Brain's Source Space. *PLoS ONE*. 2(8), e684.

Tallon-Baudry, C., Kreiter, A., and Bertrand, O. (1999). Sustained and transient oscillatory responses in the gamma and beta bands in a visual short-term memory task in humans. *Visual Neuroscience*. 16(3), 449-459.

Tang, Y., Zhang, Y.-Q., Chawla, N.V., and Krasser, S. (2009). SVMs Modeling for Highly Imbalanced Classification. *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*. 39(1):281-288.

Tenenbaum, J.B. and Griffiths, T.L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*. 24(4):629-640.

Titsias, M.K. and Likas, A. (2002). Mixture of Experts Classification Using a Hierarchical Mixture Model. *Neural Computation*. 14(9):2221-2244.

Tolias, A.S., Keliris, G.A., Smirnakis, S.M., and Logothetis, N.K. (2005). Neurons in macaque area V4 acquire directional tuning after adaptation to motion stimuli. *Nature Neuroscience*. 8(5), 591-593.

Tsunoda, K., Yamane, Y., Nishizaki, M., and Tanifuji, M. (2001). Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nature Neuroscience*. 4(8), 832-838.

Tversky, A. (1977). Features of similarity. *Psychological Review*. 84(4), 327-352.

Ullman, S. (1996). **High-level Vision – Object Recognition and Visual Cognition**. Cambridge, MA: The MIT Press.

Ullman, S., Sali, E., and Vidal-Naquet, M. (2001). A Fragment-based Approach to Object Representation and Classification. *International Workshop on Visual Form (IWVF)*. pp.85-100.

Ungerleider, L.G. and Mishkin, M. (1982). Two cortical visual systems. In: D.G. Ingle, M.A. Goodale, and R.J.Q. Mansfield (Eds.), **Analysis of Visual Behavior**. Cambridge, MA: MIT Press, pp.549-586.

Ungerleider, L.G., Galkin, T.W., Desimone, R., and Gattass, R. (2008). Cortical Connections of Area V4 in the Macaque. *Cerebral Cortex*. 18(3), 477-499.

Van Essen, D.C. and Gallant, J.L. (1994). Neural mechanisms of form and motion processing in the primate visual system. *Neuron*. 13(1), 1-10.

Van Essen, D.C. (2003). Organization of Visual Areas in Macaque and Human Cerebral Cortex. In L.M. Chalupa and J.S. Werner (Eds.), **The Visual Neurosciences**. Cambridge, MA: MIT Press, vol. 1, pp.507-521.

VanRullen, R. and Thorpe, S.J. (2002). Surfing a spike wave down the ventral stream. *Vision Research*. 42(23), 2593-2615.

Victor, J.D. and Purpura, K.P. (1997). Metric-space analysis of spike trains: theory, algorithms and application. *Network: Computational Neural Systems*. 8, 127-164.

Victor, J.D. (2000). How the brain uses time to represent and process visual information. *Brain Research*. 886(1-2), 33-46.

Vidal, J.R., Chaumon, M., O'Regan, J.K., and Tallon-Baudry, C. (2006). Visual Grouping and the Focusing of Attention Induce Gamma-band Oscillations at Different Frequencies in Human Magnetoencephalogram Signals. *The Journal of Cognitive Neuroscience*. 18(11), 1850-1862.

Viola, P. and Jones, M. (2001). Rapid Object Detection using a Boosted Cascade of Simple Features. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp.1-8.

Viola, P. and Jones, M. (2001). Robust Real-time Object Detection. *Proceedings of the 2nd International Workshop on Statistical and Computational Theories of Vision – Modeling, Learning, Computing and Sampling*. pp.1-25.

Viventi, J., Kim, D.-H., Moss, J.D., Kim, Y.S., Blanco, J.A., Annetta, N., Hicks, A., Xiao, J., Huang, Y., Callans, D.J., Rogers, J.A., and Litt, B. (2010). A conformal, bio-interfaced class of silicon electronics for mapping cardiac electrophysiology. *Science Translational Medicine*. 2(24), ra22.

Wang, R.F. and Spelke, E.S. (2000). Updating egocentric representations in human navigation. *Cognition*. 77(3), 215-250.

Wang, X.J. and Buzsáki, G. (1996). Gamma Oscillations by Synaptic Inhibition in a Hippocampal Interneuronal Network Model. *The Journal of Neuroscience*. 16(20), 6402-6413.

Wang, Y., Fujita, I., and Murayama, Y. (2000). Neuronal mechanisms of selectivity for object features revealed by blocking inhibition in inferotemporal cortex. *Nature Neuroscience*. 3(8), 807-813.

Watt, R.J. and Andrews, D.P. (1982). Contour curvature analysis: hyperacuities in the discrimination of detailed shape. *Vision Research*. 22(4), 449-460.

Weiss, Y. (1997). Interpreting images by propagating Bayesian beliefs. *Advances in Neural Information Processing Systems*. vol. 9, pp.908-915.

Whittington, M.A., Traub, R.D., Kopell, N., Ermentrout, B., and Buhl, E.H. (2000). Inhibition-based rhythms: experimental and mathematical observations on network dynamics. *International Journal of Psychophysiology*. 38(3), 315-336.

Wiesel, T.N. and Gilbert, C.D. (1989). The Helmerich Lecture: neural mechanisms of visual perception. In D.M.-K. Lam and C.D. Gilbert (Eds.), **Second retina research**

foundation symposium. Woodlands, TX: Portfolio Publishing Company; Houston, TX: Gulf Publishing Company, vol. 2, pp.7-33.

Wilkinson, F., James, T.W., Wilson, H.R., Gati, J.S., Menon, R.S., and Goodale, M.A. (2000). An fMRI study of the selective activation of human extrastriate form vision areas by radial and concentric gratings. *Current Biology*. 10(22), 1455-1458.

Wilson, H.R. (1985). Discrimination of contour curvature: data and theory. *Journal of the Optical Society of America A*. 2(7), 1191-1199.

Worring, M. and Smeulders, A.W.M. (1993). Digital Curvature Estimation. *CVGIP: Image Understanding*. 58(3), 366-382.

Wuescher, D.M. and Boyer, K.L. (1991). Robust Contour Decomposition Using a Constant Curvature Criterion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 13(1), 41-51.

Wyss, R., König, P., and Verschure, P.F.M.J. (2003). Invariant representations of visual patterns in a temporal population code. *Proceedings of the National Academy of Sciences of the United States of America*. 100(1), 324-329.

Wyszecki, G. and Stiles, W.S. (1982). **Color Science: Concepts and Methods, Quantitative Data and Formulae**. New York, NY: John Wiley and Sons.

Xu, Y. and Singh, M. (2002). Early computation of part structure: Evidence from visual search. *Perception and Psychophysics*. 64(7), 1039-1054.

Yamane, Y., Tsunoda, K., Matsumoto, M., Phillips, A.N., and Tanifuji, M. (2006). Representation of the Spatial Relationship Among Object Parts by Neurons in Macaque Inferotemporal Cortex. *Journal of Neurophysiology*. 96(6), 3147-3156.

Yamane, Y., Carlson, E.T., Bowman, K.C., Wang, Z., and Connor, C.E. (2008). A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nature Neuroscience*. 11(11), 1352-1360.

Yen, S.-C. and Finkel, L.H. (1998). Extraction of perceptually salient contours by striate cortical networks. *Vision Research*. 38(5), 719-741.

Yen, S.-C., Menschik, E.D., and Finkel, L.H. (1999). Perceptual grouping in striate cortical networks mediated by synchronization and desynchronization. *Neurocomputing*. 26-27(1-3), 609-616.

Young, M.P. (1992). Objective analysis of the topological organization of the primate cortical visual system. *Nature*. 358(6382):152-155.

Yu, A.J. and Dayan, P. (2002). Acetylcholine in cortical inference. *Neural Networks*. 15(4-6):719-730.

Yu, S.X. and Shi, J. (2003). Object-Specific Figure-Ground Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.*

Zana, Y. and Cavalcanti, A.C. (2005). Contrast sensitivity functions to stimuli defined in Cartesian, polar and hyperbolic coordinates. *Spatial Vision*. 18(1), 85-98.

Zhang, D. and Lu, G. (2004). Review of shape representation and description techniques. *Pattern Recognition*. 37(1), 1-19.

Zhou, H., Friedman, H.S., and von der Heydt, R. (2000). Coding of Border Ownership in Monkey Visual Cortex. *The Journal of Neuroscience*. 20(17), 6594-6611.

Zoccolan, D., Kouh, M., Poggio, T., and DiCarlo, J.J. (2007). Trade-Off between Object Selectivity and Tolerance in Monkey Inferotemporal Cortex. *The Journal of Neuroscience*. 27(45), 12292-12307.

Zweig, M. and Campbell, G. (1993). Receiver-Operating Characteristic (ROC) Plots: A Fundamental Evaluation Tool in Clinical Medicine. *Clinical Chemistry*. 39(4), 561-577.