Chapter 7: Discussion and Summary

It takes two to invent anything. The one makes up combinations; the other chooses, recognizes what he wishes and what is important to him in the mass of things which the other has imparted.
Paul Valéry

The method presented in the previous section was motivated by work in various disciplines. I now turn from details of implementations and experimental results to consider the limitations of this study and opportunities for improvement from the perspective of each discipline, followed by a summary of the relevance of this work for the field and the research openings suggested by the findings here and in the corresponding work surveyed.

Learning Algorithm And Representation

Several approaches can be considered which might improve the recognition by more effective learning, without changing any other network parameters. One strategy has been designated as a cooperative coevolutionary approach for general optimization (Potter and De Jong 1994) or symbiotic adaptive neuro-evolution (Moriarty and Miikkulainen 1996). This involves maintaining separate populations for subsets of the parameters of some system, and evaluating the fitness based on how randomly chosen individuals from the two pools cooperate to solve the problem. For the Soca network, the two stages could be mapped to different subpopulations, with evaluation based on a random combination of genotypes for each stage. These methods have been demonstrated to improve performance relative to single population approaches. However, given the high degree of interaction between network parameters here and the relatively low number of parameters, the potential gain is unclear.

Incremental learning (Gomez and Miikkulainen 1997) is another approach that promises improvement. Rather than optimize for a complex task from the beginning, the task may be incrementally modified with increasing complexity and an incremental form of learning used to modify the solution. One way to adapt the object recognition task would be to train first for synchronization across views with an entropy constraint, then add the maximum-cross-entropy, then finally test evolve performance on the actual recognition task between objects which generated failures after the second stage.

Using ensembles of classifiers can improve performance, when the errors made by the classifiers are independent (Hansen and Salamon 1990). The use of the cross entropy criterion could be extended to this application by attempting to form new classifiers for the same object which maximize cross entropy from the previous classifiers for the same object.

Alternatively, one might consider abandoning the evolutionary learning paradigm altogether and attempt a more direct control method. The types of optimization criteria employed here (normalization, entropy and cross-entropy measures, and synchronization limits) should be utilized as architectural principles guiding the control design.

I turn now to the current representation strategy. It may be possible that changing the partition size to finer granularity or possibly using non-uniform intervals (bin widths) to match the density of state occupancy in certain regions could increase the recognition performance.

Whether there are inherent benefits or liabilities of the partition cell coding strategy has not been a major focus of this work. Instead, the emphasis has been on demonstrating a column-like assembly with complex, stimulus-driven temporal patterns (in line with recent trends in sensory neurophysiology of the temporal areas) that can serve as the algorithmic and mechanical support for creating population coded metric spaces and normalization in view based recognition.

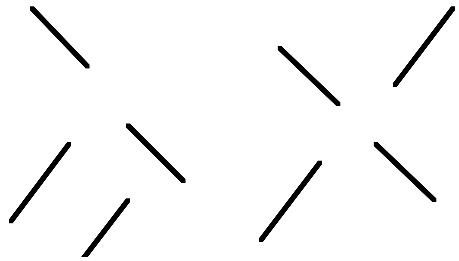
Benefits may exist; it would be interesting to compare *the ease of learning* this statistical, partition-cell based representation with localized representations, given an equivalent learning procedure such as the genetic or evolutionary learning strategy used here. Apart from the time to find a solution in some learning or search procedure, the density of feasible solutions could also be important. This is particularly true as the number of objects increases and collisions - different objects mapping to the same point in the representation space – becomes an issue. Having many solutions to a task available should make balancing of conflicting objectives easier, and could support more computationally intensive but higher performance classification by creating multiple classifiers per object, with a voting criteria or other resolution mechanism. This remains an area for future work.

SOCA NETWORK CLASSIFIERS AS A COMPUTER VISION TECHNIQUE

If viewed from the perspective of practical computer vision and visual psychology, the study has some limitations. I will note these, together with directions I envision to bring the model into closer correspondence with psychophysical performance levels.

Isolated Common First Order Statistics

The system is currently limited in scale and context sensitivity during processing to a spatial window around each pixel, within which it receives information via spatial diffusion. It is easy to present collections of lines or contours separated by this window distance which, since their “wavefronts” will not interact, will be indistinguishable; see the following figure for an example. At least two possible improvements are foreseen to handle this situation. One would be to directly address multiple scales by introducing a pyramid of CMLs, operating on subsampled binary images or other representations. Another possibility is to relax the current assumption of homogeneity in bifurcation and coupling over the lattice. Rather than choosing exact values, the system could be trained to employ a range of values to be spanned over the space of the lattice.



a b

Fig. 46. Two collection of lines which cannot be distinguished by the present architecture. Blurring, pyramid structures, and spatial gradations of control parameters might allow these to be distinguished, but whether learning will perform comparably under those conditions is unknown.

Handling Multiple Scales

As noted earlier, performance with objects subsampled once rather than twice was very poor. It is possible that this could be compensated for by training at multiple scales (the approach taken by Mel with SEEMORE). Alternatively multiple scales, could be built into the dynamics in a number of ways. A pyramid structure could operate in parallel on the data, with various scales uncoupled or possibly coupled.

Ecological Realism and Learning while Behaving

The system is currently limited, in terms of ecological realism, by the separation of learning and recognition epochs. Learning is also assisted by imposing clear temporal boundaries on the presentation of a single object. The extension of such a system to an ecologically realistic environment where learning and recognition occur together in the context of tasks would entail additional control dynamics, the nature of which are not entirely obvious.

Limited Range of Objects

The range of objects studied here is intentionally restricted. In psychological terms the problem addressed is chiefly subordinate categorization, and the refinement of that problem to stimulus identity from multiple viewpoints. Conceivably, basic level categorization arises naturally from the constraints imposed by view based interpolations on the representation space and mappings, as Edelman contends.

For future development of the system it would be desirable to test on one or more standard vision benchmarks. While quasi-standard benchmark sets exist for certain applications (faces, fingerprints, textures), a widely recognized object library is still lacking. One possibility is the library assembled by (Verfaillie and L. 1995), which has the advantages of some existing psychophysical data on view preferences, reaction time trends, and freely available. Human similarity preferences also exist for the large silhouette library used in the IBM QBIC project (Scasseleti, Alexopoulos et al. 1994), but different views are not provided for each object, limiting the direct applicability as a training corpus. One could use portions of the human similarity clusters for each object to train, then assess performance on others.

From Isolated Objects to Scene Analysis

The current limitation to recognition of isolated objects, while common in prototype object recognition systems, is not realistic. Humans can clearly recognize conjunctions of objects, or perform matching tasks with multiple objects in a scene.

As reviewed in chapter 4, many investigators have begun to address the problem of segmentation within the general framework of coupled oscillators. To build a large scale system for scene analysis which combines segmentation, recognition, and search appears to be a feasible goal. The combination of search and segmentation together is of particular interest in a practical system and should be pursued in future work.

As an initial step, one might address the search issue by continuing to learn object representations by isolated presentation, but require matching to occur in the presence of multiple isolated targets.

A more sophisticated recognition of objects imbedded in scenes would require an extension to active vision, a major research area which has not been addressed in the present work. Some sort of scanning behavior, simulating the saccade control and foveation process is envisioned. In earlier unpublished work, I proposed and simulated a dynamical algorithm for choosing stimulus-determined saccade points based on selecting image points which differed from their surround when sampled late in a Soca-like desynchronization - synchronization cycle (DeMaris 1995). These points were typically found adjacent to areas of high information content (e.g. on orthogonal axes near curvature maxima).

To address issues of ecological realism, in particular the fluid nature of mixing learning and recognition in the service of goal-directed behavior, further study of large coordination dynamics is required, combining evidence from EEG/MEG and more micro-level recordings on the same task. Analyzing experimental outcomes in terms of the framework sketched in this thesis for inter-region coordination may prove fruitful. For example, if the separate dynamical stages of the Soca processing cycle play a role in shaping of distributions for memory formation or comparison during search, a mechanism and its dynamical signatures must be identified for initiating and terminating such cycles.

Different possibilities are envisioned. One is to employ intrinsic rhythms, serving as perceptual frames. Alternatively, and perhaps more plausibly, is that behavioral events trigger the initiation of cycles. These events might include actions such as terminations of saccades, or cessation of head movements during gaze orientation.

Improving Recognition Rates on Larger Object Worlds

Before expanding the scope of the problem to segmentation and scene analysis, it would be appropriate to first address the issue of scaling up to larger problems. Performance of the system degraded substantially when the number of objects in the world was increased from 20 to 40. This degradation was more severe in the nearest neighbor recognition of a prototype, less so in the match-to-sample style task. It is possible that identifying an exact match from a large set of similar objects is so difficult that human behavior would degrade similarly, due to increases in the number of indistinguishable foreshortened views or views with similar statistics at all scales. There is a good deal of evidence that for faces, a category where we have impressive discrimination capabilities among large numbers of objects, specialized subsystems are involved.

Several promising strategies are open to exploration. Most of these would entail substantially increased computing burdens, suggesting that a rewrite of at least the core convolution and mapping loops be performed. Given the low iteration counts, roundoff should not be a serious issue; the use of single precision floating point should be explored and should lead to considerable speedup.

The present system is suitable to parallel processing at the level of parallel evaluation of network genotypes in learning and recognition epochs. The local nature of computations in the core algorithms lend themselves to parallel hardware implementations or pipelined computations. Some of the possible algorithmic improvements proposed above would impact the achievable density. Allowing inhomogeneous bifurcation or coupling parameters would entail additional registers or sequencing parameters to computation nodes, for example.

Use of More Complex, Larger and Multi-Scaled Coupling Kernels

It is likely that using a larger, more complex coupling neighborhood (convolution kernel), more partition cells, and possibly non-uniform intervals adapted to the dynamics and the object world would improve recognition performance to some extent, but with a tradeoff in computation time. Of course, the limitation to a single orientation is an artificial limitation that could be removed.

Using larger coupling kernels has potential to improve the performance, based on some preliminary results in an early experiment seeking a universal classifier (i.e. one which did not develop a classifier based on multiple presentations of an object). While the overall results were poor for scaled and objects rotated in the plane, better results were obtained with larger coupling matrices with specific orientation kernels.

It would be possible to make the coupling kernels part of an evolutionary learning strategy. I have mentioned the possibility of multiple scales of analysis of the image previously; it is likely that particular objects would require different weighting of statistics at particular scales, or coupling *between* scales, to handle situations when local or global information is more important. For example, with non-rigid objects we might wish to weight the response of local information more heavily, or allow it to have more influence on a common, cross-scale determined distribution by asymmetric coupling.

Use of All or Portions of the Ensemble Time Series

The strategy of using lattice state values sampled at a single instant was chiefly chosen for simplicity, and a low computation and storage burden. It is possible that improved performance could be achieved by examining the entire N channel time series or some measure of this dataset. Additional measures might serve as witnesses casting votes for one or another object when the standard distance measure fails to produce a clear match. For example, if the mapping of an object is close to two or more signatures, a vector of rate averages might discriminate. These additional measures would have to be saved as part of each object (or prototype's) signature, of course.

In fact, since the learning procedure currently does not perform actual comparisons, there is really no proof that the optimal discrimination performance for the given set is obtained at time step (t1 + t2). It would be interesting to examine recognition rates at earlier times. It might prove useful to consider vectors at the last two steps as predictors, and to adjust the objective function to measure and score these.

Improving Search Performance with Indexing

If adequate performance on a large and diverse object set were obtained, it would be important to address efficiency in the *search process* as well. While dynamical methods might exist, their nature remains a topic for future research. In the short term, an *indexing* strategy could be used to speed up the search process.

For example, I have shown that the application of multi-dimensional scaling to the database of mean distributions over the partition cells results in a clustering of cells.

The following steps would improve recognition performance, with speedup depending on the number of clusters obtained. The basic idea is to evaluate (i.e. execute the CML dynamics) only those prototypes near to the best performing clusters:

Perform clustering via Multidimensional scaling or some other procedure.
Choose a classifier with a signature nearest the center of each cluster.
Evaluate the phenotype network for each cluster center.
For the classifier with minimum distance from Soca(target) to signature
Evaluate all cluster members.

If the occupation of the representation space is not distinctly clustered, another technique might allow screening of distance computations after the CML evaluation. A low resolution version of the signature histogram (the partition cell occupancies) can be computed and stored. For example, the current implementation used 64 bins; construct a 32 bit string by choosing odd bins and setting the corresponding bits 1 if the occupancy exceeds half the value, 0 otherwise. (Signatures with bins near the threshold could be scheduled for full evaluation to avoid false negatives). A simple bit comparison will exclude objects with rather different signatures. More sophisticated variants of this approach have been developed for comparison of similarity in chemical structures (Willet, Barnard et al. 1998); (Flower 1998).

Such a technique could be of even greater importance if all or portions of the lattice time series were used as a signature, rather than an instantaneous sampling of the lattice as was performed in this thesis.

Noise and Occlusion

As mentioned earlier, scale space methods typically avoid *creating new spatial structure*, while the procedure described here is completely dependent on fine structure and cooperative interactions derived from those structures. This suggests that the process would be noise sensitive, and to date I have not performed any experiments by degrading the initial binary images with noise. One factor likely to mitigate the effects of both noise and occlusion is the collective measurement process.

Learning from Single Views

It is clear that humans (and monkeys) can learn to recognize exact views and some limited transformations after a single exposure. How can the current framework account for this?

One possibility is that ensemble networks serve as a back end allowing interpolation in the space spanned by a limited number of existing prototypes. This is essentially the approach taken by Chorus. Single shot learning gets a free ride on prior learning of objects seen from multiple views.

Another possibility would be to *generate multiple views from a single view* by sampling different subregions of the image, applying different morphological or disortion operators which approximate viewing transformations, or other procedures.

SOCA NETWORKS AS A BIOLOGICAL THEORY

The work here has adopted the view based recognition paradigm for recognition of objects, but differs in assumptions of the underlying neural mechanics. Admittedly, the existing framework is completely hybrid in character and has a long way to go to be a complete biological theory. Yet there are intriguing correspondences between the nature of the computation here and aspects of biological computation. In summary, the major distinguishing aspects of the Soca network approach from classical connectionist models are as follows:

1. The neuronal group level is taken as the basic functional unit, with macrostate variables representing the ongoing ensemble state. Computations and representations are formed by collective measurements (distributions) over spatially uniform ensembles of these units.
2. The typical connectionist unit transfer functions (sigmoidal activation and threshold output) is replaced with a nonmonotonic, highly nonlinear (chaotic) function. There is no threshold, since the state variable represents some collective measure such as ensemble frequency or phase distributions. Connections between groups are not excitatory or inhibitory, but serve to construct intrinsic state flows in the space of the recurrent network dynamics; these can be tailored to the statistics of input configurations. In Marr's hierarchy of processing levels, the algorithmic level involves spatial cooperative processes in these state flows.
3. The code for family of input patterns (i.e. outline shapes) is a sample of the states in the entire network at a specific point in time, rather than activation of an optimal unit (place codes), a connectionist sparse or distributed output layer, or a recurrent network in a static (fixed point) attractor. The code is computed via synchronization processes operating on stimulus-linked aperiodic oscillations with local cooperative interactions. Synchronization or clustering may arise in distant units with similar local input configurations, with no communication between the areas resulting in the synchronization. Rather, a common history and local cooperative dynamics results in occupying the same region of a dynamical phase space.
4. Synchronization may be considered a form of competitive processing, but differing from the "inhibition of rivals" scenarios usually seen in rate code models. Instead, through cooperative processes, particular oscillation modes occupy sites in the network. This is effectively competitive since there is selection from broadband oscillations to a sharply peaked distribution. Whether this "cooperative competition" offers advantages over the usual type requires further study, but it seems to map more readily onto a view of neurons as spike coincidence detectors and participants in spatio-temporal processing assemblies exhibiting local coherence (Bullock, Achimowicz et al. 1995).

5. This spatial cooperative processing and partial synchronization effectively performs a low-dimensional transformation of high dimensional objects (i.e. image arrays) after first projecting them into higher dimensional spaces (by *relaxing* subspace synchronization constraints created by earlier levels of the visual system). In this higher dimensional space, the spatial diffusion process generates information about the initial spatial configuration for increasing orders of local statistics, but conditioned by lower order statistics.
6. The network is time varying. These changes in the dynamics are interpreted as corresponding with major signal bands and modulations as observed in local field potential EEG and multi-unit neuron correlation studies. For the most part, connectionist accounts of cognitive processing make little contact with medium and large scale electrophysiology; rather, they focus on the level of neurons, stressing that these may stand for groups. In contrast, the present work investigates computational processes involving temporal changes in coupling (modulating synchronization) and in bifurcation parameters (modulating synchronization and the aperiodic or periodic tendencies of oscillations), motivated by observed changes in coherence, spectral sharpening at various space and time scales of biological neural systems.

While these attributes are all unusual in neural modeling which attempts to address the perceptual or psychological level, there are many grounds for criticism in terms of biological realism.

The arbitrary starting, stopping and injecting an image as an initial condition into the network is problematic. Some dynamical networks use an input stage to inducing a perturbation of an ongoing state vector, perhaps by increasing coupling parameters in a processing layer coupled to a learning and recognition layer.

A general criticism that might be leveled is the rather loose commitment to the meaning of the state variables at the lattice sites. I have generally followed the approach that they refer to ensemble average frequency (or spike density), which should correlate with the amplitude of the local field potential. This is the approach taken by Freeman, and I offered the justification for this in terms of random firing models producing a single humped curve.

On the other hand, I have held open the door that these principles could also apply to more complex spike coincidence and phase modulation networks; one of the strengths of nonlinear dynamics is the universality of the principles, so that many different underlying physical systems may exhibit similar dynamical trends.

I have claimed support from observations interpreted as temporal codes, but in fact the code used here is an *instantaneous state vector*. The distinction between code and computation producing the code has perhaps been blurred. I have implictly suggested that the observed temporal modulations may be a computation "steering" the system toward a partially synchronized state. It would be interesting to perform "probe" experiments on the units, to see whether in fact stimuli can be predicted from the time series of units to verify this.

In spite of these limitations, I argue that the general principles here are a viable algorithmic and mechanism-level theory (or proto-theory, perhaps) of object representation and that experiments related to many of these computational principles may serve to distinguish between the competing algorithmic theories within a view based computational framework. I now examine several such possibilities.

Slow Wave Phenomena

It is possible that the idea that slow wave phenomena in neural masses acts as a bifurcation or coupling parameter control, as I have suggested, is overstated or oversimplified. No detailed models based on integrate and fire dynamics have been developed to date to support this hypothesis; it is possible that spiking dynamics alone result in the slow wave phenomena, without anything as regular as a clock. Studies of the time course of synchronization phenomena show considerable variation from trial to trial, rather than a predictable time course (Gray, Engel et al. 1992). In all cases, episodes of oscillation are transient, lasting from 100-200 ms, followed by epochs of aperiodic activity. Whether this indicates intrinsic rhythms, the intermittent dynamics described by Tsuda as a binding strategy, or some unsuspected phenomena is not clear. If more regularity in synchronization were seen it would better support the theory of coding – however, the studies above are in primary visual cortex of anaesthetized cats, where one would really like to see similar studies in areas (V4, TE) more implicated in object level coding and recognition.

It may be that no model at the abstraction level of coupled maps could reproduce synchronization phenomena in EEG measures, though this has been difficult to achieve in more detailed models such as the Freeman group's KIII model; finding the invariant response to a stimulus remains a challenge in animal studies (Kay, Shimoide et al. 1995). Dynamics at the level of bifurcation and coupling parameters may play an interpretive role, but if the time course of their modulations are irregular and emerge directly from spike interactions, simplification may not be possible. I think it is too early to make this negative conclusion, though, as the line of investigation pursued here is still novel and untested, but tracks observations at many scales of neural dynamics which remain largely uninterpreted in computational or algorithmic terms.

The phenomenon of alpha blocking arguably is evidence against the Soca scheme. Alpha is maximal with eyes closed, especially in occipital (primary visual) and parietal areas, but is greatly reduced when a visual stimulus is presented (Basar 1998). This might be taken to indicate that the alpha band is an emergent cortical process governed by spike dynamics and graded potentials alone, rather than being a kind of dynamical control function as hypothesized here.

However, this reduction may be in part an artifact of coarse measurement in scalp recorded EEG. As reviewed in chapter 3, slow wave phenomena and changes in coherence are evident in local field potential in primary visual cortex. Modulatory dynamics and functional role of slow waves are still poorly understood. Thus, I argue the present demonstration of a computational role for modulations in parameters corresponding to slow wave phenomena is important and deserves further study.

There is some evidence that alpha frequencies are implicated in memory performance. Klimesch investigated relationships of short and long term memory with alpha band among age matched subjects; the good memory performers had 1.5 Hz higher alpha peaks compared to poorer performers. Alpha desynchronization was pronounced for subjects with poor memory (Klimesch 1996). The Soca network style described here, while primarily perceptual, can be described in part in terms of using slow wave control to turn image-like structures into a usable distributed representation suitable for comparison by hypothetical synchronization operators. This general principle may be effective in other areas, such as memory. The failure of slow waves to synchronize adequately might lead to deficits in perceptual and memory processes.

These deficits could be studied in the present conceptual framework, by examining temporal dispersion around the precise transitions to new dynamical parameters currently used. Parameters would be inhomogeneous across the array, but still drawn from the same value sets.

In the more complex biological network, such sampling of a transient process can not be the end result, but might be a starting point for further encoding in long term memory, and for matching against some version in working memory during search tasks. Changes in inter-regional coherence corresponding to readout or further processing of such an encoding should be expected, and these are precisely what the studies of Bressler, Basar and colleagues cited above demonstrate.

If Soca-like computation and coding occurs in IT cortical complex (or related form processing areas like V4) it is certain to work in conjunction with memory formation networks and comparison networks. Soca-like networks might play roles in view interpolation, invariant recognition, and similarity mapping. Memory formation might be handled by recurrent attractor networks, as advocated by Rolls and Amit; however, due to its superior handling of the compositionality issue, I favor the "exotic attractor" or itinerancy coding model, which could play a role in both (long term) memory formation and short term memory.

Stimulus Related Spatial Localization of Activity in Inferotemporal Cortex

Since the population code of Soca does not lead to asymptotic spatially localized activity in certain units, some alternative explanation is required to explain the observations of Tanaka and coworkers on spatial localizations, i.e. a small set of activated regions, with some small displacements of the spots by changing stimuli.

Several possible explanations are envisioned. In a biological system, the state vectors established by learning could be provided as input to the itinerancy coding network of Tsuda (Tsuda 1992) for permanent storage. Comparison would be implemented by some synchronization operations of the Soca network output with such an encoded memory, with success or frustration of synchronization driving behavioral response.

In this view, synchronization opponents, itinerancy coding, and competitive synchronization comparison processes are envisioned as the basic operations of object representation and recognition in biological systems. The work presented here emphasizes only the first aspect of this view of biological form processing, while other investigators have focused on memory (Tsuda 1992) and competitive synchronization processes.

This is a rather complex formulation which I must consign to the future work category; I also think it is more appropriate to a longer time scale encoding and recognition strategy involving eye movement scanning strategies. Some proposals more suited to preattentive rapid recognition are outlined here, which envision memory as more directly related to the oscillation structure and control parameters in the Soca network.

Selectionist Responses for Ensemble Response Character

One possibility is that Soca-like classifier networks are arrayed across anterior IT cortex; as they respond to input, some selection process occurs in which ensemble responses with undesirable characteristics (too synchronized during a pre-readout epoch) are suppressed by auxiliary networks sensitive to these ensemble responses characteristics.

Possible support for this theory comes from the optical studies of Tanaka and colleagues demonstrating a small spatial shift of maximum activity given different viewpoints of a stimulus. One possibility for the developmental structure and function of the anterior TE is that regular spatial gradients in some microcircuit parameter lead to a spatial array of columns analogous to the control space of the Soca networks. The small variations in response surface in the {bifurcation, coupling} control plane might lead to changes in maximum response to different views under such a selection criteria, with abrupt changes at critical transitions.

Anterior IT as Temporal Pattern Recognizer

An alternative proposal for explaining the results would re-envision a role for locally coded, combination code column “units” as proposed by Tanaka. Rather than being selective for (distal) feature combinations, such units might instead be responding to specific *ensemble frequency combinations* in the *transformed* signal, which might be computed upstream in the TEO, pIT or V4 regions. This detection could be on the basis of recognizing specific temporal patterns, such as a preferred set ensemble frequencies occurring in a short time window, or more distributed response where a local column responds to the ensemble frequency change in the last two fast cycles at the optimal readout time. The latter columns might then be linked by learning procedures.

Memory Based Search by Localized Slow Wave Interactions

Another proposal would suggest that a remote region cooperates with certain regions of IT to establish a Soca-like computation by injecting spatially localized slow wave activity into certain regions. This extra activity functions to change bifurcation or coupling parameters and resulting “stimulus tuned” ensemble frequency or phase distributions; however, it is detectable as an activity level when the incoming stimulus matches the expectation. This would suggest that the amplitude might be modulated and that if multi-channel local field potential studies were conducted in an area known to be activated for a particular stimulus through optical recording, the level of slow wave activity might be increased relative to non-active regions.

The findings of Nakamura of slow (4-6 Hz) stimulus related oscillations at the neuron level in the temporal pole are intriguing, suggesting to me that this may be the most plausible theory of matching. Correlation studies between these neurons and the temporal structure of upstream areas (TEO, V4, and pTE) could be performed to investigate whether these are deterministically linked and whether a causal flow can be determined.

A slight variation on this scheme envisions more intrinsic dynamics in the local regions. Local columns might have characteristic oscillatory structures and distributions which attempt to match, by synchronization or resonance phenomena, stimulus-created oscillations like those of the Soca network. Activation of parallel areas could be an artifact of best possible matches, while those areas with intrinsic dynamics failing to match the incoming structure have their activity suppressed.

In a primed search scenario, the slower oscillations observed by Nakamura in the temporal pole could interact with the characteristic local oscillatory structures to modulate the responses, allowing a match only for the local structures with a compatible slow wave dynamics. Correlation studies between the TE areas and the anterior temporal pole, using data such as Tanaka feature combinations, could explore possible interactions. Searching for one object A while measuring the response to putatively optimal complex stimulus B for a local region could show changes in the correlation structures of A and B, ultimately leading to suppression of activity. Searching for A and local region with optimal stimulus A should show a different correlation structure; observations of these dynamics could provide clues on how comparison is performed in a more complex dynamics scenario.

Transient Synchronization

The synchronized periodic oscillations seen in primary visual and prefrontal cortex have been chiefly interpreted as signatures of binding of features for segmentation and other Gestalt phenomena. I have suggested that such transient synchronization could also play a role in establishing the correct initial conditions for a Soca-style cycle of spatial cooperative processes under a modulated synchronization dynamics.

If some technique could be devised to interfere selectively with synchronization processes observed early in the recognition cycle in IT cortex, the effects of this on recognition performance could be examined. Theories of object recognition predicated on rate codes should predict little effect of synchronization; the present theory predicts substantial effects. This could be modeled by dispersing initial conditions about their current baseline values.

Aperiodic Oscillations

Aperiodic oscillations with synchrony and modulations of synchrony have been observed at various scales; certainly nonlinear dynamics are a potential source of such structure. At issue, though, is can we determine whether temporal codes are relevant, and whether they are produced by chaotic micro-circuit dynamics or reflect other types of computation (perhaps modulation of inputs with primed memories at the neural microcircuit level, as suggested by Eskandar et al (Eskandar, Optican et al. 1992).

The work here, stressing population codes, suggests that population measures through local field potential arrays or optical methods are the most promising techniques to address the question. If coding is distributed over populations and not localized to specific units, optimal stimulus predictions should be obtained from population measures. The differences in performance might not be evident unless complex stimuli are used.

Epochs of Desycnchronization and Synchronization

The modulations in effective connectivity (interpreted from correlations) can be explained in this way. As rhythmic volleys come into local recurrent networks from thalamic or other cortical regions, they change effective bifurcation and coupling parameters in the target local structure. For a given stimulus, the time course of synchronization is co-determined by inputs and by the dynamical flows inherent at a point in the bifurcation-coupling parameter plane. As demonstrated in the present work, such a dynamical flow may serve computational roles, functioning as a dynamical recognizer.

However, it is possible that modulations in effective connectivity arise naturally in the processing of coordinated arrays of chaotic oscillators, during representation formation or recognition, even in a single epoch. This should naturally occur as oscillators approach one another in phase space, whether the macrostate variable is interpreted as spike phase or ensemble average frequency.

Studies of modulations in correlation and effective connectivity as pioneered by Aertsen et al. (Aertsen, Gerstein et al. 1989), if performed in IT cortex along with response histogram procedures of Gochin et al (Gochin, Colombo et al. 1994), could reveal whether the best stimulus predictions are obtained in intervals of increased correlation for many pairs, as the present theory of computation and coding for object recognition would predict. RBF networks would predict no relationship between the highly correlated epochs and best histogram prediction.

The issue is confused by doubts over whether sub-regions of IT represent the locus of a feed-forward representation network, or an area of comparison of incoming dynamics with memory representation by unknown processes (possibly involving synchronization dynamics). In tasks with a visual memory component (the monkey must release a bar when a match to the target is shown after several distractors) it has been shown that neuronal responses are *decreased* substantially from their optimal stimulus when a stimulus matching the target is shown (Miller, Li et al. 1991).

Based on the various reviews cited earlier, we can exclude the anterior section (aTE) as being the likely site of comparison and priming dynamics. The best candidate regions to explore for Soca-like dynamics to be generated are intermediate ventral stages, primarily V4 and TEO.

Spatio-Temporal Patterns

One of the results presented here might be investigated in multi-channel local field potential or microelectrode array studies. In the study of parametric curves, the representations (sampled distributions) tended to have higher co-occurrence entropies as learning progresses. I am unaware of any multi-channel data analyzed by this criteria. One might examine whether this measure is predictive of differences in performance on recognition tasks.

Psychophysical Approaches to Theory Resolution

When specific views of 3D objects are used for training, error rates are shown to increase with rotation angle. This is one argument in favor of the RBF ensemble approach over the Soca model. However, this is presumably due to the fact that I did not supply or enforce conforming to any particular view in the learning process; another *network based* learning process using the same basic normalization principle but with a bias towards conforming to a view meeting some abstract (but temporally localized) principle might exhibit this effect. When only two views were used for training, peformance was highest on the trained views, but the fallout with view distance was irregular. This may also in part be due to the large angular separation between views provided to the network.

In order to discriminate between normalization in the Soca and RBF approaches, one might perform psychophysical investigations into preferred or canonical views. In the Soca approach, one might expect the preferred view to be that which is closest to the mean classifier output over all views, given the optimization criteria found to be successful here. If some class-separation strategy such as the maximum cross-entropy strategy is employed, the representation chosen, and corresponding best-map-to-mean view, would be expected to vary based on presentation order.

Recent view-based psychological theories have focused on the role of diagnostic features. A finding that presentation order dominates diagnostic features in determining preferred views would be evidence in favor of the maximum cross-entropy principle in determining representation. Presentation order has been observed to affect the neuron level activity during the delay (memory) period of a matching task, with the most similar activity arising from successively presented objects, even though there is little obvious relationship between the successively presented stimuli.

As I have repeatedly pointed out, however, multiple neuropsychological strategies and subsystems may be active, exploiting different regularities of the environment. Diagnostic features may be exploited by an alternative subsystem which is biased towards dominance by the nature of the task. Tarr et al. state that match to sample tasks, where a single target is identified, may bias subjects to rely on critical diagnostic features which may not scale well. In contrast sequential matching and naming, with continual presentation of different objects, elicits the formation or activation of a new representation at every trial.

The Encoding of Stimulus Meaning

On the basis of extensive studies in olfaction, W. Freeman has repeatedly stressed that neural dynamics encode the stimulus *meaning*, rather than a direct transformation of the feature space (Freeman and Barrie 1994). This is evidenced by changes in the stimulus evoked dynamical patterns for the same stimulus depending on the reinforcement paradigm last used with that stimulus. While I have not stressed this important point, which arguably should apply to other sensory modalities, I believe that this paradigm is compatible with the Soca classifier approach, as follows. We might consider that restricted areas of the network control space correspond to different motivation and arousal states related to emotional control of behavior. When a stimulus is learned, the dynamical encoding is restricted to those producible by control parameter sets in this subregion of the control space.

This assumes that possible solutions to a task like forming normalization solutions to multiple views are rather dense in the control space. The only evidence I have for this is the relatively successful search process with relatively few (3000) parameter sets explored. Further research on the density of such solutions would be of general interest, and in support of this theory of behaviorally relevant encodings.

SUMMARY: MAIN CONTRIBUTIONS

The main contributions of the thesis, from the perspectives of several disciplines, are summarized now.

In terms of **pure dynamics of discrete maps**, exploratory computational studies examined the short time (transient) behavior in two stage coupled logistic networks, where the stages form *synchronization opponents* adapted to a family of input forms. By opponents, I mean that the first stage has dynamics resulting in *expansion of the phase space volume*, while the second stage is, at least transiently, *contracting in phase space volume*. Initial conditions were binary distributions, which considered from the standpoint of synchronization dynamics represent an artificially synchronized starting point for the evolution of the network.

Some simple preliminary experiments examined transients in two map systems and lattices, with a range of parametric variation in both single stage and opponent stage scenarios. Parameters studied included synchronization over the bifurcation and coupling control plane, the effect of distance between the initial distribution components, and the instantaneous time distributions over a range of parameters with geometrically structured and spatially randomized binary distributions as initial conditions. In accordance with the more detailed spatially tuned chaotic neural models (Hansel and Sompolinsky 1992), it was shown that random spatial distributions produce relatively smooth distributions as the bifurcation parameter is varied while oriented initial conditions produce more irregular distributions. Coupled map lattices produce similar phenomena with relatively low computational cost. Smoothness in the time evolution and synchronization of a two map systems after small iteration counts suggests that some parameter perturbations will be tolerated.

From the perspective of **pattern recognition**, a method involving locally coupled, synchronization modulated nonlinear oscillator arrays was demonstrated to form a *population metric space* representation of objects, with partition cells in the dynamcal phase space taken as the dimensions of the space. In one family of experiments a metric space is explicitly constructed; in another, the objects do not occupy an obvious metric space, but clustering of objects with similar structure in a representation space emerges from learning to normalize views to a common distribution.

This spatiotemporal style of computing a representation contributes to resolving the classic dilemmas for representing local features, feature conjunction, and binding, while affording rapid recognition. The method was demonstrated to reach recognition rates as high as 85% in a nearest neighbor match scenario with 20 objects. The formalism of dynamical recognizers was extended from low dimensional dynamics to high dimensional dynamics

At the intersection of **dynamics, computation and pattern recognition**, preliminary evidence is presented that, for the given tasks and view normalization approach taken here, the *non-stationary* dynamics outperform a single preferred dynamical regime, such as the *transition to chaos* dynamics. At the outset of this work, it was hoped that a single dynamical stage might be found which solved the similarity problem for many objects. In a family of experiments recognizing objects rotated in depth, it was shown that a single dynamical stage is effective at rates far above chance, but results in 4% higher error rates than the non-stationary Soca strategy. There was also a slight trend toward lower average classification times with the nonstationary system. Given the non-deterministic nature of evolutionary learning, additional trials are needed to establish statistical significance of this result. This work is in progress, along with efforts to explore the effects of varying mutation rates from the baseline values used throughout the present study.

This small difference between single and two stage dynamics was unexpected; examination of sampled lattice time series reveals that during coupled chaotic evolution there are *fluctuations* of the occupied phase space volume at each step, so that characterization of a *stage* as subspace-desynchronizing or subspace-synchronizing relies on time averaged behavior, rather than an obvious monotonic increase or decrease in volume. If the recognition rate and time advantages hold with more trials, a deeper explanation of these advantages in terms of Markov chain dynamics would still be desirable. At this time, it seems that the original intuition motivating the Soca cyle is too simple an explanation, failing to capture the performance of the single stage system. The conceptual framework of probabilistic finite state automata and dynamical recognizers may be sufficient to explain the latter, but why should non-stationarity offer any improvement?

However, it is intriguing that an evolutionary scenario of improving performance with subtle improvements in timing control is suggested. The present task of recognizing the equivalence of views is emblematic of task-specific, constructive memory systems, where the memory consists of programming a general network topology with an implicit network of state flows between dynamical partition cells. Coordination dynamics between memory and recurrent computational modules may involve such programming of both task related computational systems and possibly also the programming of readout windows, sensitizing “receiver assemblies” to perform their tasks at specific times in concert with Markov-chain like computations.

A few preliminary results are presented regarding **evolutionary learning** of a computation with coupled maps. Unsurprisingly, evolutionary learning outperforms a random search strategy (15% error for the best evolved solution vs 29% for nearest neighbor classification with random search), with both evaluating 3000 parameter sets (genotypes). One notable aspect of the learning is that it occurs on the basis of an *abstract objective function* rather on the ultimate task for which the network is designed. Since the task itself involves parallel evaluation of all classifiers, this approach provides substantial speedup.

The stimulus equivalence experiments showed that the usual practice of Gaussian mutation of real valued genes, performed worse (also 29% vs.15% error) than random selection of new values from the entire range of possible values. This was conjectured to be due to the nonlinear mapping of the gene to the output space. These preliminary results concerning learning are based on very few learning trials; computing mean error rates over additional learning trials should be performed before drawing firm conclusions on the relative merits of mutation rates and single stages vs. multiple stage dynamics. . The possibility of improved performance through timing control seems helpful in understanding biological evolution, which presumably might have more difficulty searching a network space where effective topologies and network configurations are sparse islands in a sea of ineffective parameters. Experiments to assess the ease of learning tasks through network evolution in CML systems and more conventional connectionist architectures are called for.

At the intersection of **high level computer vision**, **visual psychology**, and **neuroscience of the ventral pathway**, it is shown that cooperative interactions of nonlinear oscillator transient trajectories can perform view normalization and view based recognition with recognition rates comparable to a recent feed-forward model, a statistical model with a rich input feature space. Unfortunately no comparison has yet been performed on a common set of data, but the data used here are challenging for several reasons:

1. Larger angular separation between views.
2. Inclusion of “extreme” end views in the recognition task.
3. Lower raw information due to use of silhouettes rather than color or grey scale images.

Another important distinction is that the Soca classifier set was trained solely on information theoretic principles, in sequential presentation. In contrast, the Chorus and SEEMORE systems used statistical overviews of the entire object world to choose network centers and set weights on feature spaces. I would not claim that there is no possible justification for such a procedure on biological grounds[34] .

[34] Freeman and Barrie have argued that one of the advantages of encoding all memory in a chaotic attractor is precisely that it contains the entire history of experience, which can be used to influence the formation of new memories (Freeman and Barrie 1994).

View interpolation with an ensemble of radial basis function networks has been recently proposed as a biologically motivated object recognition, and claimed to fit a variety of psychological and experimental neuroscience data; I explored several questions about the relative merits of the two approaches and suggested a multi-channel experiment to discriminate between the approaches.

The visual world used for training and testing was limited to isolated objects to focus on the problem of recognizing objects rotated in depth. Related paperclip objects have resulted in substantial error rates for humans and monkeys in learning and recognition tasks; the choice of data was motivated by continuity with such literature, and by the possibility of making predictions which can be tested with the same stimulus set. Modeling of primed search and multiple objects in a scene would be logical target tasks for next generation of this research; this might proceed in parallel with extensions to strictly network based learning and recognition.

The model of recognition is still a hybrid network - computational model rather than a pure network; in the present form, it cannot be considered as a complete biological model. Further work is required to synthesize a completely network based approach, but the principles discovered over the course of modeling and embodied in the objective functions suggest the effective form of network learning dynamics. In addition to network based learning,, a network based recognition strategy is required to more closely match biological observations. Two possible network architectures – synchronization of itinerency codes, and a back-end of ensemble frequency combination coding units were proposed as possible extended frameworks to increase the biological realism.

Finally, in the field of **computational neuroscience**, it was shown that modulations in the synchrony of large scale neural ensembles could play a role in learning and recognition. These modulations are affected in terms of bifurcation and coupling parameters of spatially distributed nonlinear systems. I offer a new interpretation for such modulations in the context of the *synchronous opponent cooperative activity* dynamics (Soca networks).

Previously, such modulations in synchrony, observed in multi-unit studies in pre-frontal and primary visual areas, have been interpreted as signatures of feature binding over space. I have proposed an understanding of such modulations as the unfolding of computational and coding processes in what I termed “coherence assemblies”.

The concept of information processing by symbolic dynamics, proposed long ago in the context of low dimensional systems, (Nicolis 1986) is extended to a higher dimensional system, resulting in a hypothesis of population coding via partition cell metric spaces.

A novel interpretation for transient synchronization episodes observed in primary visual cortex is proposed. My claim is that synchronization may be required in earlier stages of the visual pipeline - concurrent with or after segmentation processes - in order to prepare for the phase space expansion (desynchronization) stage of the two opponent stages performing a view normalization computation in IT or closely related form processing areas. This preparation consists of the creation of a high contrast primal sketch, which serves as an effective *initial subspace* for the spatiotemporal computation and population coding.

If epochs of increased synchronization and correlations are observed in association areas (as they have been in primary and prefrontal cortex) these may correspond to the state space contraction (synchronization) stage. These epochs should be correlated with increased stimulus prediction performance relative to epochs of desynchronization obtained from multi-channel vector measurements.

The intrinsic difficulties of reconstructing the behavior of non-stationary, spatially distributed oscillating systems in the brain suggest an important role for modeling. Models and simulations allow a constructive approach (Kaneko and Tsuda 1994) to supplant or enhance attempts at reverse-engineering the behavior of neural systems from correlations of neuron level components with stimulus, behavior, and other low level components. Constructive, in this context, refers both to the role of the network designer (human, evolution, or evolutionary computation) and to the rich capabilities of dynamical state flows and synchronization operators to rapidly construct implicit networks.

This is especially important if neural systems function by intrinsic or stimulus linked non-stationary behavior, since few methods exist for characterizing non-stationary dynamics and deducing the system function. Given clues about system function derived from evoked potential and spatial and temporal trends in synchronization, along with classical information theoretic principles from pattern recognition, models may be built for specific cognitive phenomena. Once a model exists, the same methods used for analysis of multi-channel biological signals (cross-correlation, coherence, and higher order spectral measures) can be applied to these artificial systems, leading to a fruitful interaction between theory and experiment.

In the simulation work presented, such signal analysis measures have not yet been applied to the resulting time series; rather, the observed trends in medium scale dynamics are taken as motivation to explore base nonstationarities in the service of some task, presently this exploration proceeds by evolutionary computing methods. Thus the work is in the spirit of constructive modeling, as defined above. By demonstrating a functional role for such nonstationarities in a classical perceptual task, I hope to contribute to expanding the dialog on neural computing mechanisms to include such relationships between modulations in synchrony, bifurcations as control for such modulations, and population computation and coding.

While speculative discussions of these topics are common among experimentalists and theorists in the last decade or so, concrete *engineering* models which perform well on a challenging task are still rare, as are attempts to systematically relate chaotic synchronization to any underlying neuroscience. I have attempted to realize both goals here in a fashion which is accessible to workers coming from psychology, experimental neuroscience, and engineering, providing enough background to justify what otherwise might seem very abstract forms of computation. If, by this effort, the coupled map style of computation gains credibility as a technique in computational neuroscience, these research communities may all benefit from an important new tool for understanding the programming style hidden in the rhythms of the brain and mind.