Biologically-Inspired Cognitive Architecture for Integrated Learning, Action and Perception (BICA-LEAP)

Dr. Deepak Khosla

HRL Laboratories, LLC
3011 Malibu Canyon Road
Malibu, CA 90265

September 2006

Final Report
Contract No. N00014-05-C-0510
June 1, 2005 – May 31, 2006

David Gunning
DARPA/IPTO
3701 North Fairfax Drive
Arlington, VA 22203-1714

Approved for Public Release, Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Service, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
28-09-2006	Final	01-06-2005 – 31-05-2006

4. TITLE AND SUBTITLE
Biologically-Inspired Cognitive Architecture for Integrated Learning, Action and Perception (BICA-LEAP)

5a. CONTRACT NUMBER
N00014-05-C-0510

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Dr. Deepak Khosla

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
HRL Laboratories, LLC
3011 Malibu Canyon Road
Malibu, CA 90265

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Program Officer
Office of Naval Research
875 North Randolph St. Suite 1425
Arlington, VA 22203-1995

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSORING/MONITORING AGENCY REPORT NUMBER

12. DISTRIBUTION AVAILABILITY STATEMENT

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The central challenge in the BICA program is to develop neuroscience theories, designs and cognitive architectures that are comprehensive and cover the full range of human cognition. One of the limitations of neurally inspired brain architectures that have been characterized to date is that they tend to solve modal problems (visual object recognition, audition, language, motivation, etc.) in disparate architectures whose design embodies specializations for each modal problem. Our objective in Phase I has been to develop neuroscience theories and principles towards a biologically inspired cognitive architecture for integrated learning, action and perception (BICA-LEAP). BICA-LEAP is based on the concept of brain operating principles (BOPs) and computational paradigms (CPs) to realize structural, functional and temporal modularity and also integrate the various neural processes into a unified system that can exhibit wide range of cognitive behaviors. We have developed theories and designs for visual processing, language processing and sequentially planned behaviors. These will be the initial steps for developing an architecture and methodology that creates the foundation for follow-on phases to simulate and implement cognitive architectures. A single comprehensive architecture that covers the full range of human cognition should provide a basis for developing cognitive systems that can not only successfully function in a wide range of environments, but also thrive in new environments. We envision that BICA-LEAP will be integrated into a variety of applications and existing systems, providing support or replacement for human reasoning and decision-making, leading to revolutionary use in military and commercial applications.

15. SUBJECT TERMS
Biologically Inspired Cognitive Architecture, Perception, Action, Learning, Brain principles, Intelligent machines, Adaptive Resonance, Planning, Scene Understanding, Control

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					Dr. Deepak Khosla
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPONE NUMBER (*Include area code*)
U	U	U			(310) 317-5037

Approved for Public Release, Distribution Unlimited

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI-Std Z39-18

TABLE OF CONTENTS

		Page
1.0	Scientific and Technical Objectives	1
2.0	Approach	1
3.0	Concise Accomplishments	2
4.0	Extended Accomplishments	2
4.1	Goal-driven Scene Understanding	2
4.1.1	Gist Extraction	3
4.1.2	Biasing Attention Top-Down Towards Features of Interest	6
4.1.3	Role of Memory in Guiding Attention	7
4.2	Language/Symbolic Communication	8
4.2.1	Preliminary Implementation/Results	17
4.3	Learning Sequentially Planned Behaviors	17
4.3.1	Behavior Conditioning: What-Where-Why-When Interactions	20
4.3.2	Behavior Conditioning: Computing Reward Expectation	24
4.3.3	Learning and Execution of Sequential Plans	25
4.3.4	Cerebellar Learning: Skill Acquisition & Fine Motor Control	27
4.4	Initial Brain Operation DataBase (BODB)	31
4.5	Integrating Modules towards Architecture	33
References		36

PREFACE

The author gratefully acknowledges the technical contributions of the BICA-LEAP team and their input to this report:

Narayan Srinivasa, Suhas Chelian, Alex Dow, Chris Moore and Dave Huber (HRL)
Professors Michael Arbib and Laurent Itti (University of Southern California)
Professor Dan Hammerstrom (Portland State University)
Professors Steve Grossberg, Gail Carpentar, Dan Bullock, Ennio Mingolla and Frank Guenther (Boston University)

Approved for Public Release, Distribution Unlimited

1 Scientific and Technical Objectives

The central challenge in the BICA program is to develop neuromorphic theories, designs and resulting cognitive architectures that are comprehensive and cover the full range of human cognition. One of the limitations of neurally inspired brain architectures that have been characterized to date is that they tend to solve modal problems (visual object recognition, audition, motivation, etc.) in disparate architectures whose design embodies specializations for each modal problem. In addition, there are often very different underlying theories and architectures for the same cognitive modal problem. This presents a significant challenge in seamlessly integrating these disparate theories into a comprehensive architecture such that all cognitive functionalities can be addressed. Computational design and implementation of these architectures is another major challenge. These architectures must be amenable to implementation as stand-alone or hybrid neuro-AI architectures via software/hardware and evaluation in follow-on phases.

Our objective in BICA Phase I is to develop neuroscience theories and principles towards a biologically inspired cognitive architecture for integrated learning, action and perception (BICA-LEAP). BICA-LEAP is based on the concept of brain operating principles (BOPs) and computational paradigms (CPs) to realize structural, functional, and temporal modularity and also integrate the various neural processes into a unified system that can exhibit a wide range of cognitive behaviors. The aim of our work in Phase 1 has been to provide theories and brain principles towards an architecture and methodology that create the firm basis for follow-on phases to simulate and implement cognitive architectures. A single comprehensive architecture that covers the full range of human cognition will provide a basis for developing cognitive systems that can not only successfully function in a wide range of environments, but also thrive in new environments. We envision that BICA-LEAP will be integrated into a variety of applications and existing systems, providing support or replacement for human reasoning and decision-making, leading to revolutionary use in military and commercial applications.

2 Approach

As we are developing our architecture, we are focused on key BOPs and CPs: these include complementary processing, laminar computing, and Adaptive Resonance Theory. Complementary processing postulates several complementary and hierarchically interacting

processing streams and sub regions that cooperate and compete in parallel. Laminar computing postulates a uniform layered format/structure for neural circuitry in various brain regions that represents the brain's approach to reusable computing. Adaptive Resonance Theory is a framework that allows autonomous adaptation in real-time to a rapidly changing and complex world. BICA-LEAP also integrates learning mechanisms, adaptively timed neural circuits, and reinforcement-learning-based neural circuits that model emotional and motivational drives to explain various cognitive processes including reasoning, planning, and action. The above key BOPs and CPs provide our architecture with the ability to control a flexible repertoire of cognitive behaviors that are most relevant to the task at hand. These characteristics are realized using an inherently nonlinear and parallel architecture and offer a powerful alternative to the linear and probabilistic models of traditional AI-based systems.

3 Concise Accomplishments

Key accomplishments in Phase I to date include:

1. Neural principles of goal-driven scene understanding modulated by saliency and attention.
2. Neural principles of multimodal communication that benefit from interactions between language and action repertoires.
3. Neural principles of learning sequentially planned behaviors guided by the laminar computing.
4. Integrated view of models that support neurophysiological and psychophysical data.

4 Extended Accomplishments

4.1 Goal-driven Scene Understanding

Integration of goal-driven, top-down attention and image-driven, bottom-up attention is crucial for visual search. For instance, in robot navigation it is important to detect goal-relevant targets like road signs and landmarks, and to simultaneously notice unexpected visual events like sudden obstacles and accidents. Yet previous research has mostly focused on models that are purely top-down or bottom-up. Our main focus has been on integrating bottom-up attention with goal-driven top-down attention and additionally linking attention to invariant object recognition (size, position, and view invariance). We have developed a preliminary plausible architecture

that links all of the above and have identified the main brain operating principles and regions involved in these functionalities. The eventual aim is goal-driven parsing of complex, cluttered natural scenes for rapid object detection and recognition as well as scene understanding.

Under BICA funding we have significantly developed our biological framework for goal-driven scene understanding, outlined in Figure 1.

The main results are outlined below.

4.1.1 Gist Extraction

Computing the "gist" of a scene and using it to contextually guide attention towards potentially task-relevant targets is a very important component of the overall architecture, as it allows the cognitive system to rapidly orient towards scene locations that are more likely to contain objects of current interest given a set of goals. To achieve this, we compute a generic low-dimensional "signature" of an image, in the form of an 80-dimensional feature vector that summarizes the entire scene. Our implementation computes this signature as outlined in Figure 2 (Siagian & Itti, in press).

Once such a feature vector is computed that captures the gist of the entire scene, it can be exploited by the system in two ways:

1) to perform scene classification, which in turn will be used to prime the rest of the architecture to look for objects that are likely to occur in the given type of scene (Siagian & Itti, 2005). We have described and validated this simple context-based scene recognition algorithm for mobile robotics applications. The system can differentiate outdoor scenes from various sites on a college campus using a multiscale set of early-visual features (Figure 2), which capture the gist of the scene into a low-dimensional signature vector. Distinct from previous approaches, the algorithm presents the advantage of being biologically plausible and of having low computational complexity, sharing its low-level features with a model for visual attention that may operate concurrently on a robot. We compared classification accuracy using scenes filmed at three outdoor sites on campus (13,965 to 34,711 frames per site). Dividing each site into nine segments, we obtain segment classification rates between 84.21% and 88.62%. Combining scenes from all sites (75,073 frames in total) yields 86.45% correct classification, demonstrating generalization and scalability of the approach (Siagian & Itti, 2005).

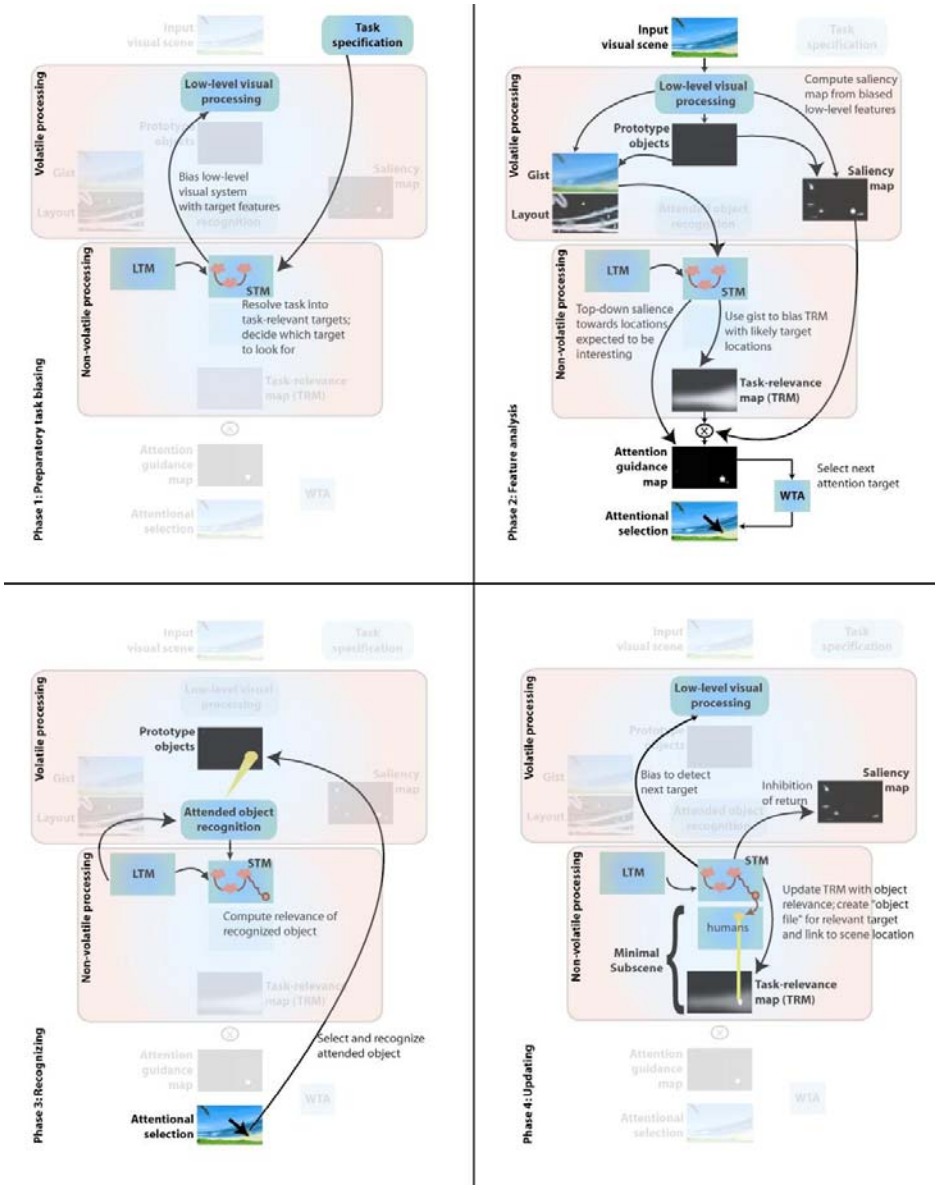

Figure 1. Architecture for the construction of goal-oriented cognitive representations of dynamic natural scenes. Each cycle of activity passes through four phases: *Preparatory task biasing* of low-level vision once the task is known and before visual input is received; *Feature Analysis* that will serve in computing gist, salience maps, and task-relevance maps (TRMs); *Recognizing* by selecting the most salient and relevant location in the scene, and updating the short-term memory (STM) based on how it relates to currently relevant entities; and *Updating* by which recognized objects, actors or actions on a given attention shift are used to update both STM and TRM, so that preparation for the next shift of attention can begin and be dependent on the current partial evaluation and cognitive representation of the scene.

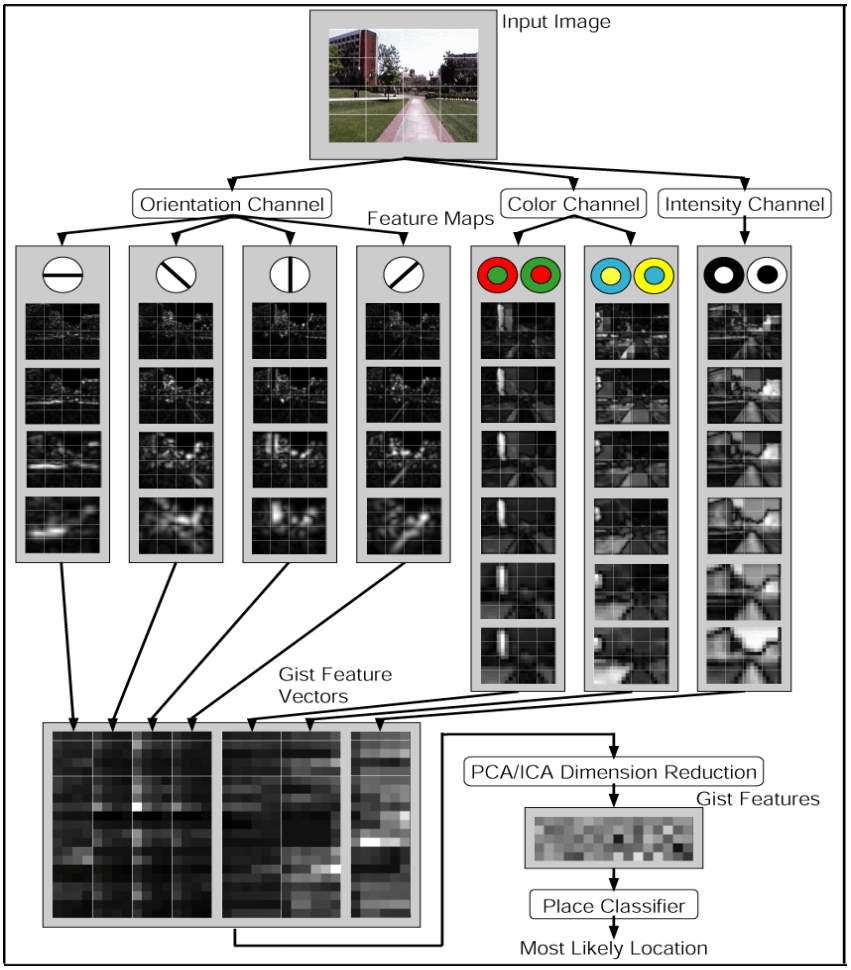

Figure 2. Computing the gist of a scene using simple feature channels over the entire image.

2) Learning the association between gist signatures and likely eye positions of human observers performing a visually-guided task. This allows the system to attend preferentially to locations that would be visited by the human eye under identical task conditions (Figure 3). In this work we have recorded eye movements of humans while they were interactively solving complex tasks, such as playing video games. Through an incremental learning algorithm, the algorithm is able to associate human eye position with scene gist. On new video games, the system can then predict, in the form of a probabilistic map, where people solving the same task as used for training are more likely to look. This map is then combined with the bottom-up saliency map, yielding a combined bottom-up / top-down attention guidance map (Peters & Itti, 2006).

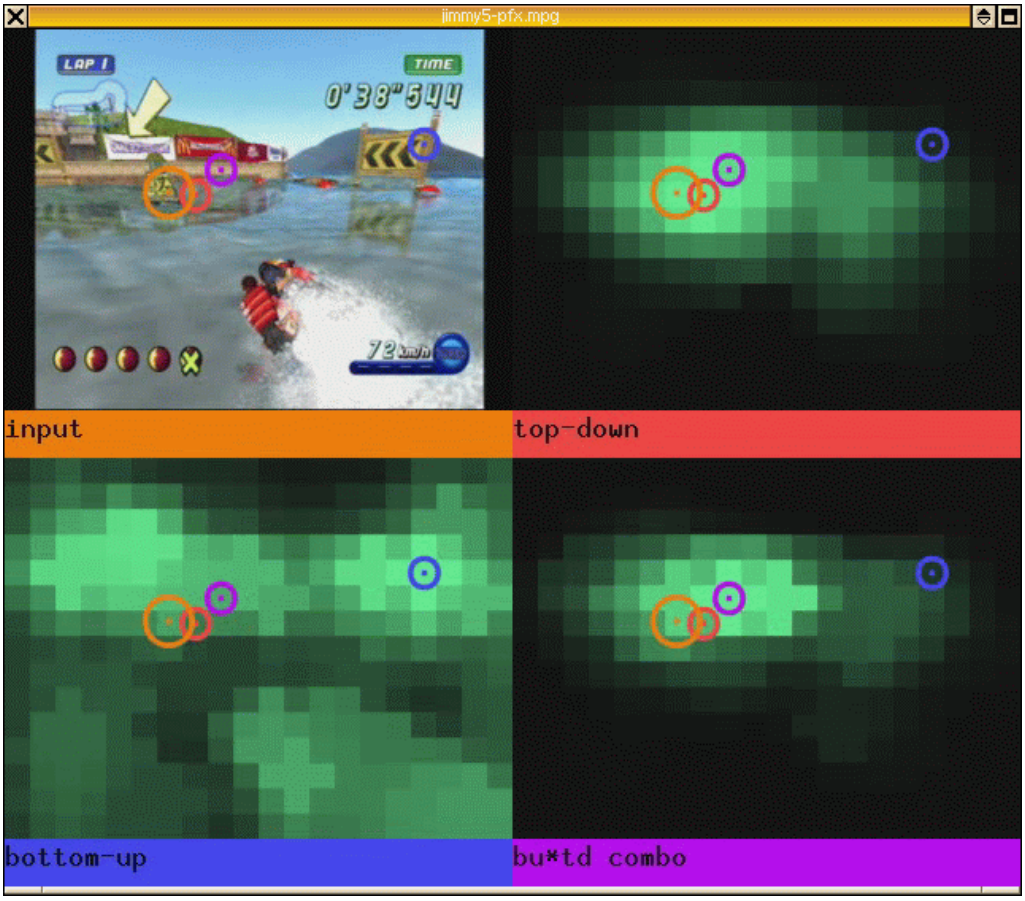

Figure 3. Prototype (Peters & Itti, 2006) combined bottom-up + top-down model and comparison with human eye movements. While a subject was playing this jet-ski racing game (upper-left), his gaze was recorded (large orange circle in all four quadrants). Simultaneously, the top-down map (upper-right) was computed from previously learned associations between scene gist and human gaze (maximum top-down activity indicated by small red circle). The bottom-up saliency map was also computed from the video input (lower-left, maximum at small blue circle). Finally, both maps where here combined by taking a pointwise product (lower-right, maximum at small purple circle). This frame exemplifies a situation where top-down dominates gaze allocation (several objects in the scene more bottom-up salient than the one being looked at).

4.1.2 Biasing Attention Top-Down Towards Features of Interest

Despite substantial neurobiological and behavioral evidence that knowledge modulates feature processing and facilitates visual search, until now there was no mathematical theory to capture such top-down influences. We have developed, in part under BICA funding, an optimal theory of how prior statistical knowledge of target and distractor features modulates the response gains of neurons encoding low-level visual features, such that search speed is maximized. Through numerical simulations, we have shown that this theory successfully explains many reported behavioral and electrophysiological observations including top-down effects such as the

role of priming, role of uncertainty, target enhancement and distractor suppression, as well as bottom-up effects such as pop-out, role of target-distractor discriminability, distractor heterogeneity, linear separabilty and others. Further, the theory makes surprising predictions whereby finding a target may sometimes require suppression of target features, or enhancement of non-target features. We validated these counter-intuitive predictions through new psychophysics experiments. Four naive subjects performed a difficult search for 55-degree-oriented target among 50-degree distractors. The gains thus set up were tested by randomly inserting probe trials, in which we briefly flashed (200 ms) four items representing the distractor (50 degree), the target (55 degree), relevant as predicted by the theory (60 degree), and steep (80 degree) cues. Although subjects searched for a 55 degree target, as predicted by the theory, there were significantly higher number of reports on the 60 degree item (paired t-test with $p<0.05$). These results provide direct experimental evidence that humans may deploy optimal feature gain modulation strategies (Navalpakkam & Itti, 2006).

4.1.3 Role of Memory in Guiding Attention

Moving towards higher cognitive processes, we noted that paying attention to the right thing at the right time underlies the ability of humans and other animals to learn, perceive, and interact with their environment (Itti, 2006). A central unresolved question is the time frame in which spatial memory guides attention, with current estimates ranging from a single fixation to seconds, minutes, or even days. In a recent publication (Carmi & Itti, 2006) we answered this question by revealing the time course of attentional selection during natural vision. We asked human participants to visually explore either continuous or scene-shuffled video clips, and quantified the impact of memory-free influences on overt attentional selections (saccades) based on a computational saliency model. Overall, scene shuffling resulted in no significant differences in the impact of memory-free influences compared to continuous viewing. However, abrupt scene transitions (jump cuts) led to sharp peaks in the impact of memory-free influences, which then declined progressively across 7 fixations for up to 2.5 seconds. These results indicate that visual exploration of dynamic scenes critically depends on spatial memory traces that persist across several fixations for up to a couple of seconds.

4.2 Language/Symbolic Communication

The goal in this task is to integrate neural models of language production and recognition as well as models of embodied cognition and its link to image understanding. We have explored neural models of language production and understanding. Regarding language production, we have begun using neural models of articulatory control with correction from somatosensory and auditory feedback. Regarding language understanding, we have explored a number of models from the acoustic, lexical, and semantic levels. Higher-level language production will focus on goal-directed image understanding and question answering. The end goal is a comprehensive bio-inspired architecture with underlying models and principles for speech comprehension and production.

The need for enhanced multimodal interaction with human users and other systems (e.g., animals, computers) has been only partly met by AI achievements. Bio-inspired approaches on the other hand have tended to focus on modal problems in disparate architectures. Therefore, we have approached the problem of language and symbolic communication in the context of an integrated system including vision (Task B1.1) and action (Task B1.3). In particular, this task has:

1. continued to study how language emerged with respect to praxic action and the Mirror Neuron System,
2. extended models of symbolic description and question answering of a scene (e.g., who is doing what to whom),
3. interfaced work on the auditory and vocal periphery, and
4. collected a literature review of the neural correlates of language comprehension and production.

Since we have sought a single comprehensive architecture based on core BOPs and CPs, our architecture has the ability to control a flexible repertoire of cognitive behaviors that benefit from interactions between language, vision, and action repertoires.

The four previously-mentioned research thrusts will now be explained in greater detail. We also highlight the biological inspirations and constraints that we have incorporated into this work.

Mirror Neuron System

The analysis and reflection of Arbib and Rizzolatti (1997; Rizzolatti and Arbib, 1998) on the macaque mirror system and a human "mirror system" in or near Broca's area, traditionally thought of as involved in speech production, led Arbib and Rizzolatti to develop the Mirror System Hypothesis (MSH). The *Mirror System Hypothesis* states that the evolutionary basis for language parity is provided by the mirror system for grasping, rooting speech in communication based on manual gesture.

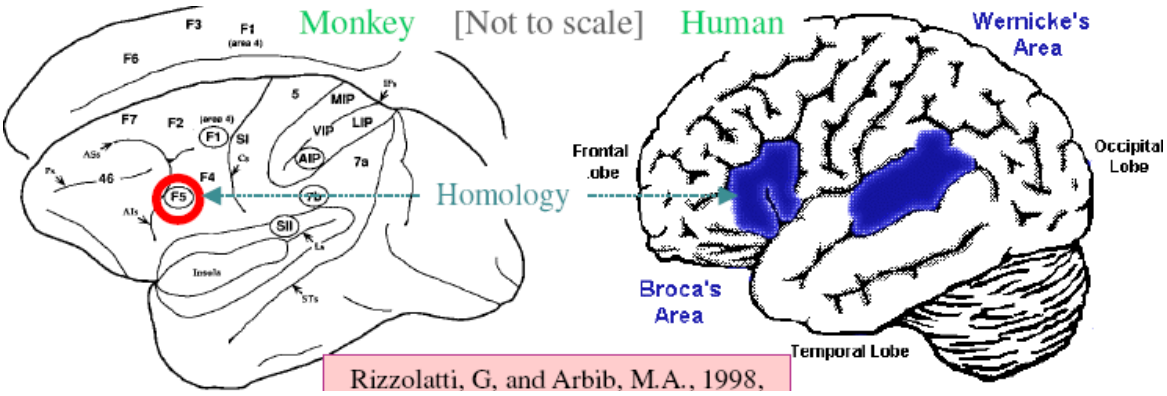

Figure 4. Key data: Monkey F5 (with its mirror system for grasping) is homologous to human Broca's area. Imaging studies show activation for both grasping and observation of grasping in or near Broca's area.

Arbib (2003) has developed a general framework, refining the original formulation of MSH, that encompasses seven different stages from grasping to simple and complex imitation, protosign, protospeech, and language. Current work encompassing grasping includes the Infant Learning to Grasp Model (ILGM) (Oztop et al., 2004) and the Mirror Neuron System (MNS) model (Oztop and Arbib, 2002). The ILGM model presented by Oztop et al. (2004) shows how to learn to make motor plans in response to sensory stimuli such that open-loop execution of the plan leads to a successful grasp. MNS is a model of action recognition learning by mirror neurons of the macaque brain. Simple and complex imitation requires the ability to construct and re-parameterize motor schemas. Towards that end, the tuning of schema-based motor plans has been sought with respect to affordances and desirability (M. Arbib, personal communication). Protosign, protospeech, and language remain topics for future research agendas.

Modeling of the mirror system for grasping and its development to chart the brain mechanisms involved in the production and comprehension of multi-modal language provides a clear understanding of the way in which evolution has "re-used" various brain operating principles in passing from praxic action to communication.

Development of the Mirror System Hypothesis began with the discovery that mirror neurons in the macaque are homologous to Broca's area in humans (Arbib and Rizzolatti, 1997; Rizzolatti and Arbib, 1998). Arbib and Bota (2005) looked at further homologues between macaque and human to ground further work on the origin of language. Three major parallel parieto-frontal interactions, summarized in the figure below, highlight the structure-function mapping of the MNS.

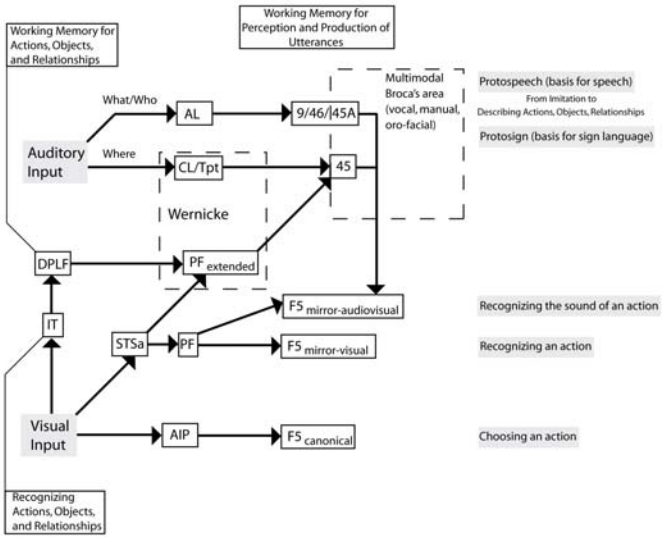

Figure 5. A high-level view of the cumulative emergence of three fronto-parietal systems: praxis, action understanding, and language production (in multiple modalities). (From Arbib and Bota (2005).)

Praxis flows from the visual centers of the brain (retina, LGN, V1, etc.) to the intraparietal sulcus of parietal cortex (AIP) region, and then onto F5 canonical neurons. Action understanding begins in the visual centers of the brain and continues to the rostral part of the superior temporal sulcus (STSa), and finally to the inferior parietal area PF (Brodmann area 7b). Dissociation data from humans shows that the praxic use of size information and the declaration of that information, either verbally or through pantomime, are due to a segregation in parietal and inferotemporal streams. Language comprehension and pantomime goes from the visual centers

to Wernicke's area and the temporo-parietal (Tpt) region; language production continues from there to Broca's area. The level of detail in this figure was designed to elicit further modeling. Fully implemented models of grasping and action recognition include FARS (Fagg and Arbib, 1998) and MNS (Oztop and Arbib, 2002). Each of these models includes detailed synthesis of anatomical and physiological constraints as diagramed below.

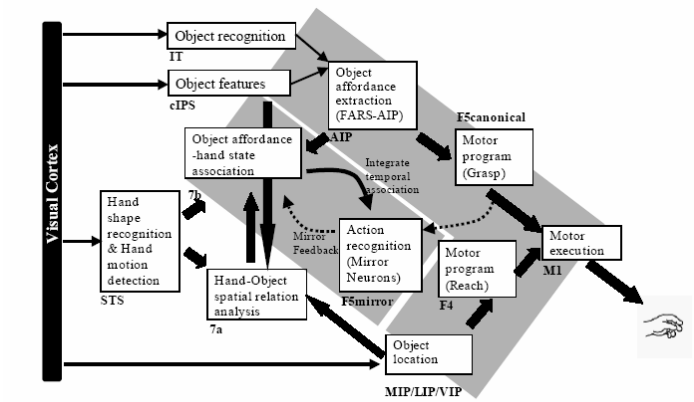

Figure 6. A lower-level view of action understanding as in the MNS. (From Oztop and Arbib (2002).)

There are three major components: a dorsal stream, a ventral stream, and a mirror neuron core supporting grasp choice, grasp location, and action recognition, respectively. Further details into each stream can be seen in Figure 6.

The MNS framework established by Arbib (2003) includes the following stages: grasping, imitation, protosign, protospeech, and language. Some of the structure-function mapping in grasping is summarized in the figures above. The computational models of IGLM (Oztop et al., 2004) and MNS (Oztop and Arbib, 2002) will also generate testable hypothesis regarding the nature of how the brain performs these tasks. Complex imitation requires the ability to construct and re-parameterize motor schemas (among other things), which are probably served by the basal ganglia and pre-supplementary motor area (M. Arbib, personal communication).

Description and Question Answering of a Scene

Itti and Arbib (2005) describe a computational framework that explores the interaction between focal visual attention, the recognition of objects and actions, and the related use of language. Central to this work are the notions of "minimal subscene" and "anchored subscene"

to provide a middle ground representation, in which an agent is linked to objects or other agents via some action.

Although in general the language system and the visual system may continually influence each other, the design can divide the effort as much as possible between the two systems. On the visual side, one can explore how scene analysis based on multiple fixations could deliver a variety of representations. This is the subject of Task B1.1, where a preliminary model of visual attention that links bottom-up salience, contextual cues, object recognition, top-down attention, and short-term memory in building representations of subscenes has been presented by Navalpakkam and Itti (2005). On the language side, one can explore the expansion of a given representation into a sentence that expresses it, or given a question, the creation of a representation that can query a scene. This subject can be addressed with the development of Construction Grammar.

The integration of visual attention and object recognition and the use of language to describe "minimal subscenes" has been outlined in the Salience, Vision, and Symbolic Schemas architecture (Itti and Arbib, 2005), as shown below.

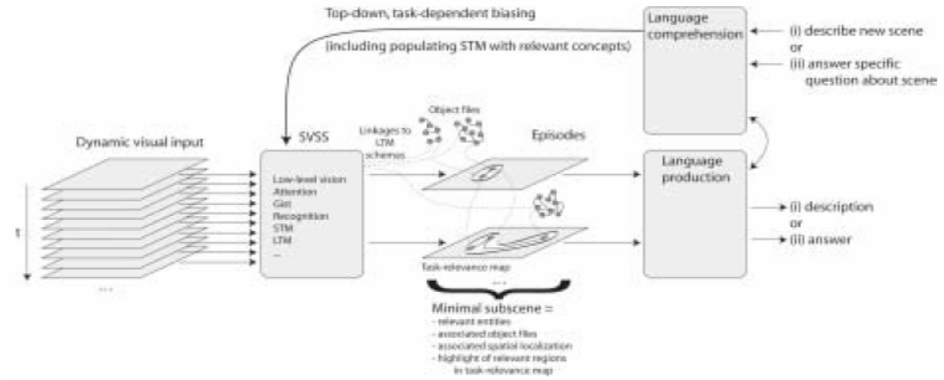

Figure 7: The SVSS framework. Minimal subscenes and symbolic schemas for episodes as the interface between vision and language. (From Itti and Arbib (2005).)

We accept the Wernicke-Geschwind theory as necessary but not sufficient for all language capabilities. Clearly, damage to Wernicke's area, the arcuate fasciculus, and Broca's area can lead to fluent, conduction, and nonfluent aphasias, respectively (Dronkers et al., 2000). However, the theory makes little mention of manual communication, as detailed in the previous section, and does not mention what happens before Wernicke's area, or after Broca's area, as detailed in the next section.

Following the synthesis of Dronkers et al. (2000), beyond the Wernicke-Geschwind theory, there are regions of the brain that are important in rule checking (at many levels), grounding, the interaction thereof, and prosody and pragmatics. Rule checking is partly supported by Broca's area involvement in anaphora, short term memory (cf. Baddeley's articulatory loop), and other grammatical judgments. Broca's area is also supported by left frontal cortex, left parietal cortex, sensorimotor areas above the sylvian fissure between Broca's and Wernicke's area (lower sectors of Brodmann's areas 3, 1, 2, and 4), the insula, and the right hemisphere. Wernicke's area helps support grounding in collaboration with portions of the left temporal lobe. Assembling words and coordinating speech articulation is believed to be rooted in the arcuate fasciculus' ability to connect temporal, parietal, frontal and especially insular cortices. The initiation and maintenance of speech is supported by the supplementary motor area and the anterior cingulate region. Finally, prosody and pragmatics are thought to be mediated by the right hemisphere. Damage to the anterior right hemisphere leads to deficits in the production of prosody, while damage to the posterior right hemisphere leads to deficits in the comprehension of prosody.

Auditory and vocal periphery

Whereas a schema level treatment of language perception and production may serve some purposes, the true use of language requires solutions to such demanding problems as multi-modal integration and real-time motor control. Several biologically-faithful models of these lower-level functions have been developed. These models leverage the design constraints of biological systems to produce such features as incremental learning and distributed processing.

Language perception begins with the ability to tease apart sound sources into distinct mental "streams," as seen in the "cocktail party problem." The model of Grossberg et al. (2004) clarifies how the frequency components that correspond to a given acoustic source may be coherently grouped together into a distinct stream based on pitch and spatial cues. Streams are formed as spectral-pitch resonances that emerge through feedback interactions between frequency-specific spectral representations of a sound and its pitch. These resonance and matching mechanisms are specialized versions of Adaptive Resonance Theory, or ART— a common BOP repeated throughout our architecture.

After sound sources are segregated, features within a stream point to phonemic or other atomic units of speech perception. One model that models the interaction transient and sustained

properties of the acoustic signal that roughly corresponds to consonants and vowels, respectively, has been developed (Boardman et al., 1999). Although its ability to model psychoacoustic phenomena is quite deep—it can handle rate-invariant speech processing and other perceptual context effects—its range is somewhat limited to simple inputs. Therefore, we have chosen to use an on-line supervised learning system (Carpenter, 2003) as to make arbitrary mappings between acoustic input and phonemic labels outputs.

The integration of phonemes (or phoneme-like units) into words is not simply a feed-forward process. During fluent speech perception, variations in the durations of speech sounds and silent pauses can produce different perceived word groupings. The model of Grossberg and Myers (2000) elucidates how sequential activation and storage of phonemic items in working memory provide bottom-up input to unitized "word" representations, or list chunks, of varying lengths. These list chunks compete with each other as they dynamically integrate this bottom-up information, with winning groupings feeding back to provide top-down support to their phonemic inputs. This model, shown below, provides a further illustration of how ART-like dynamics can explain diverse auditory perceptual events.

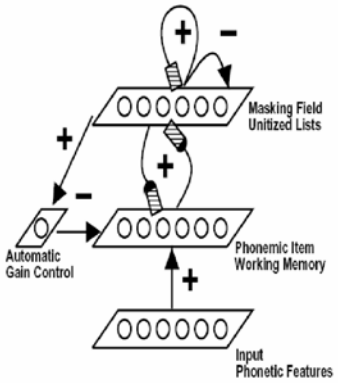

Figure 8: ART-like resonant dynamics explain diverse auditory perceptual events. (From Grossberg and Myers (2000).)

Beyond word-level recognition, grammatical constraints need to be considered. The ability to self-organize integration of symbol categorization and phrase structure rules can be achieved by some networks (Negishi, 1995). Without explicit instruction or external teachers, these networks classify words and phrases, and evaluate how likely these categorized tokens form higher-level phrases. Higher-level phrases are fed back to its classifier networks in a reduce operation, while introduction of a new word or phrase creates a shift operation. This

model incorporates several ART-like notions such as bottom-up activation of categories and suppression of learning when top-down expectations are met.

Embodied semantics, at the most general level, requires a categorical language representation in the What cortical stream and its affordances in the Where cortical stream (along with other notions of appearance that may be in the What stream as well). The neural network of Grossberg and Repin (2003) is one such network focused on the representation and comparison of multi-digit numbers. However, its structure may be general enough to allow the encoding of any cardinal feature space. Within such a network, an analog channel integrates transient responses where they activate an ordered spatial map that selectively responds to the number of events in a sequence and exhibits Weber law properties. Going beyond numbers, if the analog channel registers degree of match to a certain object or action class, then its registration to a field of orderly topology of quantities would allow size comparisons. Furthermore, by noticing consistently overlapping network activations, one can make ontological comparisons through the use of a rule-discovery algorithm as described by Carpenter et al. (2005). Variants of these model mechanisms have been used elsewhere to explain data about other Where stream phenomena, such as motion perception, spatial attention and target tracking (e.g., Chey et al., 1998; Gancarz and Grossberg, 1999; Grossberg, 1999).

Turning to language production, we find that the model of Guenther et al. (2006) learns to control speech articulators by babbling and monitoring acoustic feedback. The model is capable of demonstrating a variety of human-like speech effects such as: linear velocity/distance relationships, motor equivalence, speaking rate effects, and carryover and anticipatory coarticulation.

The hypothesized neural correlates for low-level language and comprehension are expanded in the current section.

Source segregation—imperative in “cocktail party” phenomena—begins with the transduction of sound into primary auditory cortex (A1), via the cochlea and medial geniculate nucleus (MGN) of the thalamus (Hudspeth, 2000). The acoustic grouping model of Grossberg et al. (2004) dynamically integrates pitch channels into streams through center-surround competition and ART dynamics (Carpenter and Grossberg, 1987). Based on an analogy to visual processing, pitch channels are probably represented in A1, but streams are most likely represented after A1. The BOPs and CPs of center-surround competition and ART are found

throughout the brain, both within modalities and across modalities (please see Grossberg (1999) for a review).

The detection of phonemes is assumed to occur before Wernicke's area. He et al. (1997), for example, found cat neurons tuned for iconic features of phonemes in the dorsal zone of cat auditory cortex. Temporal integration of phonemic features, therefore, would occur at this level. To handle a large vocabulary of phonemes, however, we have used a supervised learning system (Carpenter, 2003) to perform an acoustic-to-phonemic mapping.

Word-level integration of phonemes (or phoneme-like units) is assumed to occur in Wernicke's area. As stated above, these phonemes are represented before Wernicke's area. The model of Grossberg and Myers (2000) dynamically integrates phonemes into words through center-surround competition, habituative transmitter dynamics, and a Masking Field (Cohen and Grossberg, 1987) all in an ART structure. Thus ART dynamics continue at this level as well.

Grammatical judgments are crucial to language understanding. In terms of role classification of words, Broca's area in collaboration with the left frontal cortex and left parietal cortex are especially important (Dronkers et al., 2000). Thus the computations of a network to learn grammatical constraints (Negishi, 1995) would be seated here. However, in terms of grounding, Wernicke's area with portions of the left temporal lobe are important (Dronkers et al., 2000). Although much grounding has focused on visual object naming, the model of Grossberg and Repin (2003) also looks at how the parietal cortices support numbers and their comparisons.

Turning to language production, the selection of a grammatical construct is assumed to occur in the parietal cortex (the "cortical chunking" area). From there, the frontal cortex is the most likely area where words or phrases are downloaded ("working memory" area). Phonetic encoding would continue in the insular cortex, while the initiation and maintenance of speech would occur in the supplementary motor area and the anterior cingulate region (the "execution module"). The interaction of all these processes would heavily use the arcuate fasciculus. Prosodic effects would be mediated by the right hemisphere. These conclusions are based on mapping the general structure-function mapping of the sequencing model of (2002) to the constraints of language production (Dronkers et al., 2000). Finally, the "syllabery" of articulator control is assumed to be within the left ventral premotor cortex (Guenther et al., 2006).

The articulatory control model of Guenther et al. (2006) itself is composed of complex interactions between premotor (e.g., left ventral premotor cortex), motor (e.g., M1),

somatosensory (e.g., inferior parietal cortex) and auditory cortices (e.g., superior temporal cortex). The motor and auditory areas mentioned above interact, and often overlap, with Broca's and Wernicke's area. In its current implementation, cerebellar side-loops and basal ganglia gating have been simplified. The former can be elaborated from by the model of Rhodes and Bullock (2002), while the latter can be integrated into the various models by Bullock and colleagues (please see Bullock (2004) for a review).

The brain areas mentioned above for language comprehension and production involve almost all parts of the brain (e.g., frontal, parietal, temporal cortex, as well as the cerebellum and basal ganglia). This partly illustrates how biological systems distribute functions across many structures.

4.2.1 Preliminary Implementation/Results

The MNS model of Oztop and Arbib (2002) has been applied to the recognition of three different types of grasps. Manuscripts are also in preparation to elaborate on the MNS model, IGLM model (Oztop et al., 2004), and schema-based motor planning with respect to affordances and desirability (M. Arbib, personal communication).

Simulations of word-level integration of phonemic units (Grossberg and Myers, 2000) have also been conducted using phonemically-transcribed speech from the NTIMIT dataset (Jankowski et al., 1990). Phonemic items provide bottom-up input to "word" list chunks. These list chunks then support their current and expected input. Habituative transmitter gates between and within both levels allow word to "tire out" or "piggy back" off partial phonemic strings that have ended abruptly or continued after pause, respectively. These results were shown at the Boston midterm meeting in the form of an object-oriented Matlab GUI. Results showed the ability of the network to be trained on one sentence, and tested on a permutation of the same sentence. In this implementation, phonemes were presented as a binary string denoting the presence or absence of a given phoneme at a given time. The ability to present phonemes acoustically was also studied and implemented. Preliminary results were promising.

4.3 Learning Sequentially Planned Behaviors

Brain evolution has enabled land vertebrates to adopt highly dynamic life styles that involve continuous growth over significant fractions of the life cycle as well as adaptation to large seasonal fluctuations in body parameters and resource availability. Such animals must

continually retune their sensory-motor control pathways in an attempt to keep up with the non-stationary properties of their bodies and their contexts of action. Recent analyses of the adaptive neural substrates of planning and action, including the frontal regions of the cerebral cortex, the amygdala or prelimbic cortex, the basal ganglia complex and the cerebellar system, have revealed a design that allows just in time assembly of the effective control system needed for the task at hand. Such capability allows us to acquire and exhibit intelligent behaviors that ensure its survival in a continuously changing world.

In this task, we have looked at the necessary theory and an integrated model of a neural architecture that can address various aspects of learning and planning sequential behaviors. The theory explains instrumental concepts such as reinforcement, drive, incentive, motivation and motor control and integrates their functions with cognitive concepts such as expectancy, competition and resonance in the brain. We are developing a unified framework that integrates the various parts of the brain that offer neurophysiological evidence for learning and planning of sequential behaviors including: a) thalamo-cortico-hippocampal-cerebellar-basal ganglia interactions for adaptive timing of behaviors based on predictions of when rewards may or may not occur for particular actions; b) thalamo-cortico-amygdala interactions to learn cognitive-emotional associations that mediate the motivational relevance of multi-sensory inputs necessary for construction of plans/actions (or sequence preparation) at the prefrontal/orbitofrontal cortex; c) adaptive competitive queuing in the cerebellum that enables queuing of plans for action based on mutual competition and; d) coordinated scaling of these plans in the cerebellum (mediated by the basal ganglia) to achieve precisely timed actions.

The aim of this integrated framework is to show how real-time analysis of adaptive behavior in various environments can disclose the underlying network principles and mechanisms that result in a unified design. Learning plays a key role in the acquisition of these complex behaviors (refer to Figure). The processes that control the learning and recognition of sensory and cognitive memories, often called declarative memory (Mishkin, 1982, 1993; Squire and Cohen, 1984, Desimone, 1991; Gochin, Miller, Gross, and Gerstein, 1991; Harries and Perrett, 1991; Mishkin, Ungerleider, and Macko, 1983; Ungerleider and Mishkin, 1982) provide the essentials of the sensory world in the form of “what” are the objects found in a scene and “where” they are located in the world.

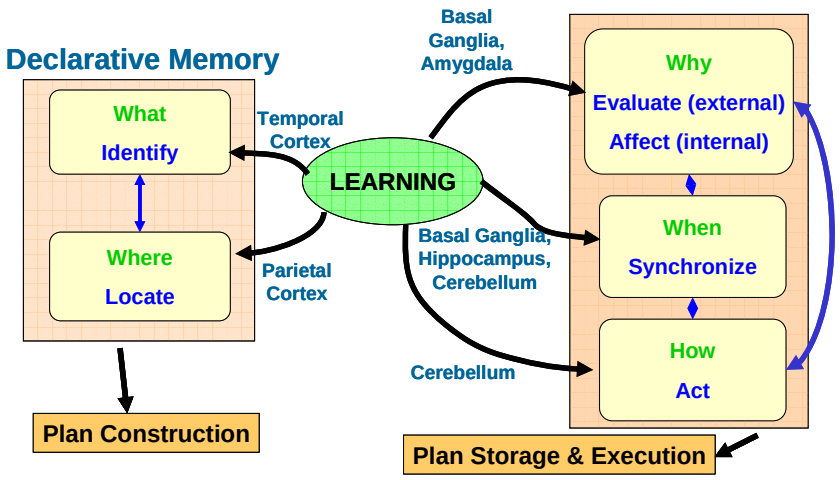

Figure 9. The various processing streams and the brain regions involved during the learning and storage of plans.

The interactions between the "what" processing stream (located in temporal lobe of the brain) and the "where" processing stream (located in parietal lobe of the brain) help in construction of declarative and episodic memory that is eventually useful for construction of plans. The performance of learned motor skills that are often classified as part of a procedural memory system (Gilbert and Thatch 1977; Ito, 1984; Thompson, 1988) is related to the "why," "when" and "how" processing streams.

The "why" processing stream uses reinforcement learning to assess the value of objects/stimuli and helps in plan selection and decision-making. The basal ganglia complex is involved in assessing values of external rewards while the amygdala is implicated in relating the external stimuli to internal states of the animal and provides the motivational influence on plan construction. The "when" processing stream is primarily involved in learning to adaptively time the execution of plan(s) and action(s). The basal ganglia complex and cerebellum are primarily implicated for this function. They achieve this competency by adaptive switching and gating circuits, which enable on-the-fly reconfiguration in order to match the sensory-motor flows to the present task. This also allows the animal to add more cognitive and volitional dimensions to its processing. Finally, the "how" processing stream is involved in the actual execution of a plan such as reaching or grasping. The cerebellum and motor areas such as the spinal circuitry are involved in the control of this processing stream.

These processing streams are linked via the prefrontal cortex, which acts as the behavior planner. This region of the brain is a key player in the learning, storage and retrieval of plans and its combines the context and goals with which the animal is operating to select among candidate plans and specify the plan of action (into working memory). Serial and parallel interactions between these learned competencies ("what," "where," "why," "when," and "how") via the prefrontal cortex enables the animal to acquire complex behaviors (refer to

Figure 10). We will highlight the key neural principles and mechanisms as well as the associated brain structure-function mapping and computational models for learning and execution of sequentially planned behaviors.

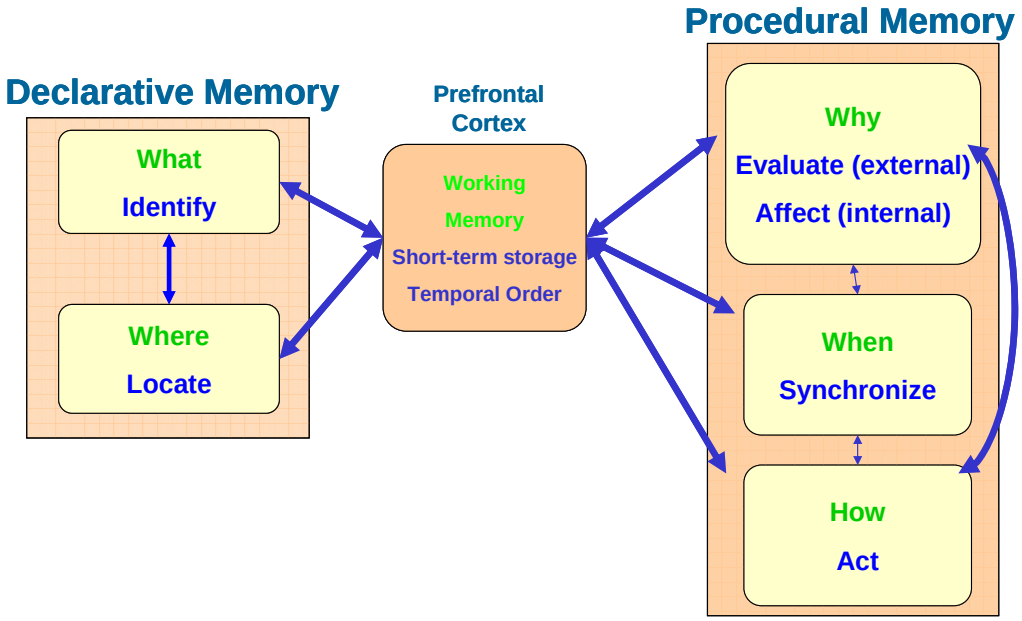

Figure 10. Serial and parallel interactions between the various learned competencies and the prefrontal cortex enables the generation of complex behaviors.

4.3.1 Behavior Conditioning: What-Where-Why-When Interactions

Psychologists have identified classical and operant conditioning as two primary forms of learning that enables animals to acquire the causal structure of their environment. In the classical conditioning paradigm, learning occurs by repeated association of a conditioned stimulus (CS), which normally has no particular significance for an animal and always gives rise to an unconditioned response (UCR). For example, a rat that is repeatedly shocked (UCS) shortly after a red light is turned on (CS) will associate the red light with fear, meaning that eventually, presentation of the red light alone elicits a conditioned response (CR) resembling the fear

response elicited by the shock itself. Hence, classical conditioning is the learning process that enables animals to recognize informative stimuli in the environment.

In the case of operant conditioning, an animal learns the consequences of its own actions. More specifically, the animal learns to exhibit more frequently a behavior that has led to reward in the past, and to exhibit less frequently a behavior that has led to punishment. For example, a pigeon can be trained to peck at an illuminated key in order to receive a small food reward, while a human might learn to stop at a red light in order to avoid getting in an accident. Traditional neural networks research have suggested that neural networks based on associative learning laws can model the mechanisms of classical conditioning, while neural networks based on reinforcement learning laws can model the mechanisms of operant conditioning (Sutton and Barto, 1981; Sutton and Barto, 1990). These learning mechanisms have also been suggested to have biological underpinnings (Doya, 1999). At least two fundamental problems arise from these sorts of neural networks: first, the majority of neural networks function only as long as the inputs and outputs are controlled and timed carefully with respect to each other; second, most neural networks have no means of learning to discriminate "good" inputs from "bad" inputs on the basis of an internal value system. The first of these problems has been called the *synchronization problem* (Grossberg 1971; Grossberg 1982): how can learning between a CS and a UCS occur reliably even though they are presented at different times on different trials? The second problem of discriminating "good" from "bad" is related to motivation, the internal force that produces actions on the basis of momentary balance between our needs and the demands of the environment (Dorman and Gaudiano, 1994).

The emotional centers of the brain, such as the amygdala, interact with sensory and prefrontal cortices to generate affective states, attend to motivationally salient sensory events, and elicit motivated behaviors. Activating the feedback loop between cognitive and emotional centers is predicted to generate a cognitive-emotional resonance that can support conscious awareness of events happening in the world and how we feel about them. Recent experimental data provide increasing support for the predicted role of interactions between amygdala and orbitofrontal cortex in the control of response selection and predicted outcomes based on value acquired through previously rewarded behaviors (Baxter et al., 2000; Schoenbaum et al., 2003).

Three types of learning take place among these representations: Conditioned reinforcer learning (CRL) enables sensory events to activate emotional reactions at drive representations. Incentive motivational learning (IML) enables emotions to generate a motivational set that biases the system to process cognitive information consistent with that emotion. Motor learning allows sensory and cognitive representations to generate actions.

We have modeled the emotions within this model using a gated dipole circuit that represents a minimal network that is capable of generating a sustained but habituating (decaying), on response to onset to a cue, as well as transient off response (or antagonistic rebound) to offset of a cue (refer to Figure). This gated dipole along with associated learning and feedback called Recurrent Associative gated Dipole (READ), and has been used to demonstrate various classical conditioning paradigms (Grossberg and Schmajuk, 1988). The amygdala is modeled a bank of read circuits with opposing emotional drives as each bank's outputs. These drives compete for every CS input and the winner drive is used to gate the selection of a motivationally relevant plan at the prefrontal cortex.

The cognitive-emotional resonance must be adaptively timed or the animal could be condemned to either premature goal-oriented responses or to generate maladaptive orienting and exploratory movements in any situation wherein the goal object does not immediately appear. Adaptively timed learning enables both attention and action to be appropriately timed to generate adaptive behavior in each environment. Evidence for adaptive timing has been found during many different types of reinforcement learning. Such learning is optimal at a range of positive ISI between the CS and US that are characteristic of the animal and the task, and is greatly attenuated at zero ISI and long ISIs. Within this range, learned responses are timed to match the statistics of the learning environment (Smith, 1968).

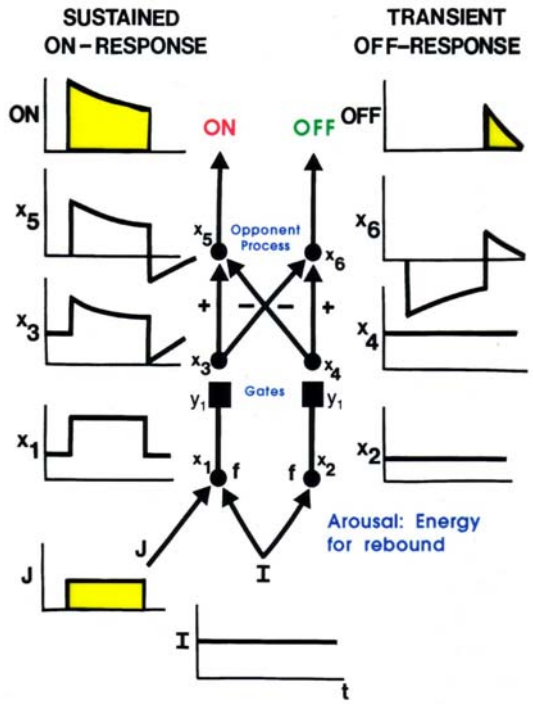

Figure 11. The READ gated dipole circuit for classical conditioning (reprinted with permission from Grossberg and Schmajuk, 1988).

Although the amygdala has been identified as a primary site in the expression of emotion and stimulus-reward association (Aggleton, 1993), the hippocampal formation has been implicated in the adaptively timed processing of cognitive-emotional interactions. The START model (Grossberg and Merrill, 1992, 1996) accomplishes this by showing how circuits within the hippocampus that are capable of adaptively timed learning can modulate the responses of ART and CogEM circuits that have already been summarized. Hoehler and Thompson (1980) have provided experimental evidence that adaptively timed circuits exist in both the hippocampus and the cerebellum by doing ISI shift experiments during which the peak time of the hippocampal trace can change before the peak time of the discrete adaptive response.

In Figure 12, a computer simulation of adaptively timed learning within the dentate-CA3 circuit is shown wherein the output R learns the correct timing of the UCS after 7 learning trials.

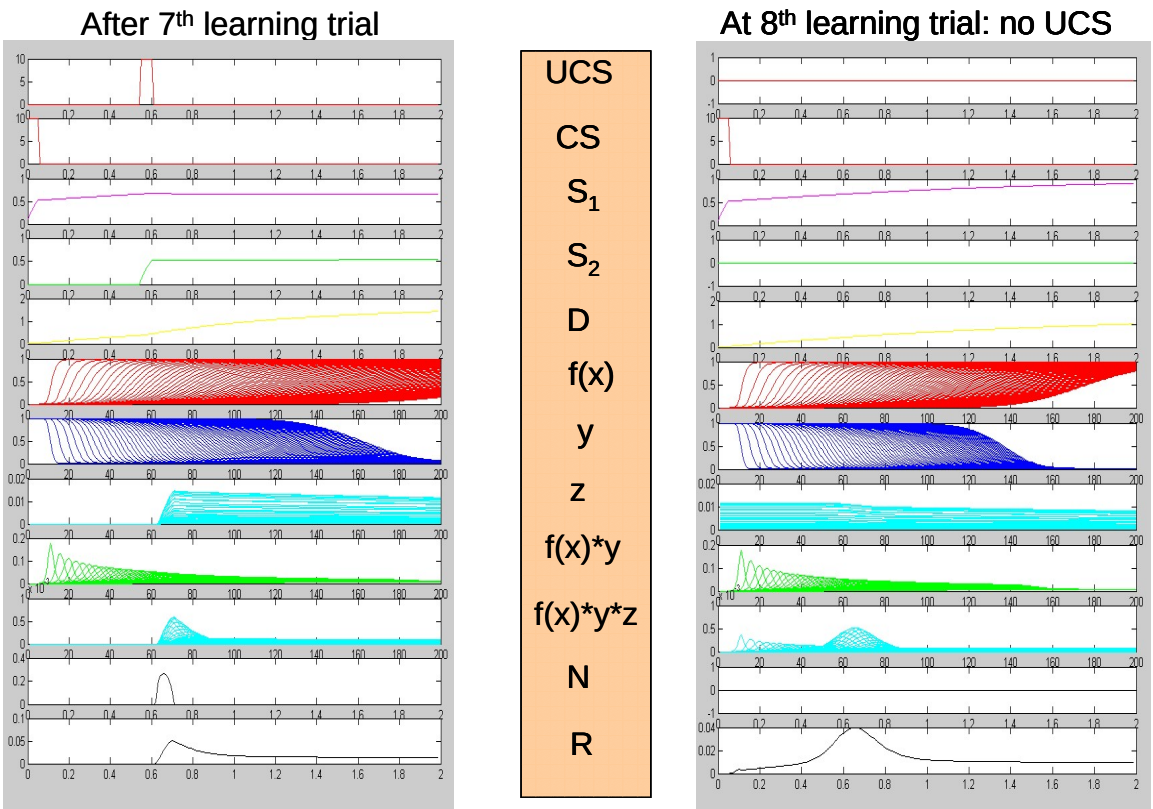

Figure 12. A plot of computer simulation of the spectral timing model. On the left plot is the activity of various stages of the START model and the bottom row shows the response of the network R during the 7th learning trial of a CS-UCS presentation. As can be seen from the plot on the right, the R output correctly predicts the timing of the UCS when only the CS is presented to the network.

4.3.2 Behavior Conditioning: Computing Reward Expectation

Humans and animals can learn to predict both the amounts and times of expected rewards. The dopaminergic cells of the substantia nigra pars compacta (SNc) have unique firing patterns related to the predicted and actual times of reward (Ljungberg et al., 1992; Schultz et al., 1993; Mirenowicz and Schultz, 1994; Schultz et al., 1995; Hollerman & Schultz, 1998; Schultz, 1998). In particular, dopaminergic reward signals seem to strengthen the “incentive salience” or “wanting” of a certain reward -- that is, the motivation to work for the reward in a given behavioral context -- as distinct from the affective enjoyment or “liking” of a reward once consumed (Berridge and Robinson, 1998). The “liking” may be mediated by areas other than the basal ganglia (McDonald and White, 1993). Recent models (Montague et al., 1996; Schultz et al., 1997; Schultz, 1998; Berns and Sejnowski, 1998; Contreras-Vidal & Schultz, 1997; Houk et al., 1995) of the nigral dopamine cells have noted similarities between dopamine cell properties

and well-known learning algorithms, especially Temporal Difference (TD) models (Montague et al., 1996; Schultz et al., 1997; Suri and Schultz, 1988). While providing a degree of insight into the information carried by the dopamine signal, the TD approach has not been able to answer the questions of what biological mechanisms actually compute the signal, and how. In particular, how does learning in the circuit that includes these cells enable them to produce a fast excitatory response to conditioned stimuli and a delayed, adaptively timed inhibition of response to rewarding unconditioned stimuli?

We have adopted a model in which the learned excitatory and inhibitory responses (Brown et. al 1999) are subserved by different anatomical pathways and the adaptively timed inhibitory learning is mediated by metabotropic glutamate receptor (mGluR)-driven Ca2+ spikes in striosomal cells. These Ca2+ spikes occur with a spectrum of temporal delays. When a Ca2+ spike and a dopamine burst occur at the same time, inhibitory learning is enhanced at the corresponding delays. To explicate these excitatory and inhibitory pathways, the model functionally explains and simulates the firing patterns of dopamine cells, striosomal cells of the striatum, pedunculo-pontine tegmental nucleus (PPTN) cells, ventral striatal cells, and lateral hypothalamic cells. Its mGluR-based spectral timing mechanism helps to explain more data than the temporal derivative operation that defines the class of TD models previously used to describe dopamine cell behavior.

4.3.3 Learning and Execution of Sequential Plans

Intelligent behavior depends upon the capacity to think about, plan, execute, and evaluate sequences of events. Whether we learn to understand and speak a language, solve a mathematics problem, cook an elaborate meal, or merely dial a phone number, multiple events in a specific temporal order must somehow be kept in mind temporarily in working memory. Once events are stored temporarily in a working memory, they are then grouped, or chunked, through learning into unitized representations that encode whole sequences of events (e.g., word and action sequences). How these working memory sequences and unitized plans interact during cognitive information processing and motor performance remains one of the most important problems confronting cognitive scientists and neuroscientists.

We have developed a neural model based on a combination of (1) how the layered circuits of prefrontal and motor cortex may be organized to achieve processes of working memory storage, sequence learning, and motor planning during both cognitive and sensory-motor tasks and (2) developed substrates for plan learning and construction and further provides the necessary substrates for plan selection and execution. This model combines three forms of representation. The first form of representation is parallel response activation subject to an ordering system to select which competing plans to execute (also known as competitive queuing models; Grossberg 1978a, 1982; Houghton, 1996; and Boardman and Bullock, 1995). The second representation is called stimulus response chaining where the next movement in response triggered by feedback from previous movement (Lashley, 1951; Elman, 1990; Arbib & Dominey, 1995; Cleeremans & McClelland, 1990). The third representation is called the serial order wherein items within a sequence are tagged or labeled with their relative position within that sequence (Sternberg, 1978). This approach provides our model to extract the benefits of all the three representations while avoiding the pitfalls of any single form of representation.

Grossberg (1978a, 1978b) mathematically derived STORE working memories from two postulates, which are called the LTM Invariance Principle and the Normalization Rule. These postulates assume merely that the representations of items stored in working memory (a) closely preserve their relative activations, or ratios, throughout their retention time to enable stable learning, and (b) item representations are approximately normalized such that total activity is redistributed when new items are encoded. Taken together, these simple rules generate working memories that can support stable learning and long-term memory of sequence, or list, chunks. List chunks categorize sequences of stored items and their temporal order. In particular, it was mathematically proved that, under constant attentional conditions, the pattern of activation that evolves in a STORE working memory is typically one of three types:

(1) A *primacy gradient* in which the first item to be stored has the largest activity and the last item to be stored has the smallest activity. A primacy gradient allows the stored items to be rehearsed in their correct order.

(2) A *recency gradient* stores the first item with the smallest activity and the last item with the largest activity. Rehearsal of a recency gradient recalls the most recent item first and the first item last.

(3) A *bowed gradient* may be stored, in which the first and last items to be stored have larger activities, and thus are earlier rehearsed, than the middle of the list. It was also proved that, as more and more items are stored, a primacy gradient always becomes a bowed pattern.

In other words, in order to be able to stably learn and remember list chunks or plans based on short-term working memory storage of a sequence of events, the brain is limited in the number of items that can be recalled in the correct order from the working memory, due to the development of a bow in the stored gradient. These models have shown to explain the famous psychological experiments (Von Restorff, 1933; Miller, 1956; Murdoch, 1962).

In the family of STORE models, individual items (events, objects) are categorized, or unitized, at an earlier processing stage. Related modeling work developed Adaptive Resonance Theory, or ART, to explain how these unitized categories may be learned through interactions between several cortical regions (sensory, temporal, parietal, prefrontal), interacting with the hippocampal system, and how the top-down expectations from the prefrontal cortex of these categories can help to stabilize category learning and long-term memory (e.g., Bradski & Grossberg, 1995; Brown, Bullock & Grossberg, 2004; Carpenter & Grossberg, 1991; Grossberg & Merrill, 1996; Grossberg, 1978a, 1980, 1999, 2003; Srinivasa and Ahuja, 1999, Engel, Fries & Singer, 2001 and Pollen, 1999). These top-down expectations attentively modulate or prime during a delay period, and match, synchronize, and amplify the bottom-up distributed features that their categories learn to bind. The match/mismatch operations between top-down expectations and bottom-up features also permit discrimination between familiar and novel events, can drive reset of an active category and its top-down expectation during a mismatch, and trigger selection of a better-matching category during the next time interval.

4.3.4 Cerebellar Learning: Skill Acquisition & Fine Motor Control

In the performance phase of the selected plans, the cerebellar side-loop plays a major role as briefly highlighted in the previous section. We will now look at the cerebellum more closely and highlight the model that we will be using based on past work done at BU (Rhodes and Bullock, 2002). The cerebellum is an ancient neural module implicated in sensory-motor control. Damage to appropriate regions of the cerebellum in human adults causes a lasting breakdown in coordinated action, especially in high speed actions that require synchronous motion by several coordinating joints. Yet the core ability to generate multi-joint movements, with voluntary

selection of spatial targets and continuous voluntary control of movement speed, is not compromised by cerebellar damage. In multi-joint movements, during which each joint motion perturbs the trajectories of mechanically coupled joints, low gain feedback control, by itself, implies a very low performance ceiling due to long feedback lags that force them to operate at low gain in order to avoid instability.

The cerebellum can be viewed as a supra-joint-level controller that "sees them coming" and correctly schedules control actions that effectively cancel the expected perturbations. On the basis of its informational inputs, the cerebellum tightly times its outputs and the rules by which it adaptively adjusts those outputs. The outputs of the cerebellum via the deep cerebellar nuclear (DCN) stage are organized into discrete motor channels that map one-to-one feedback error channels. Thus for every feedback control channel, there is a corresponding cerebellar channel whose output excites the same motor command center excited by the feedback error signals carried by that channel. Sitting above each DCN channel is an entire population of giant Purkinje cells (PC) that normally inhibits the DCN channel. This PC population must learn *what* predictive contexts, exactly *when* after the detection of these contexts, and by *how much* they should briefly reduce their inhibitory outputs in order to allow transient activation of the DCN channel they control. There are a huge numbers of mossy fibers (MF) that receive inputs asynchronously from various sensory systems including vision, kinesthesia, touch, balance and audition that pass the context and state information to granule cells via mossy fiber rosettes. Each granule cell axon becomes bifurcated into parallel fibers (PF). There is an array of PF's that are sampled in distinct subsets by PC's that affect a given DCN channel. This way the cerebellar circuit can sample a distinct but a very large set of potentially predictive contexts. Local inhibition applied by Golgi cells to the granule cell helps restrict the search to the signal combinations coded by strongly excited granule cells.

Example: Learning to Reach Spatial Targets

We have thus far looked at specialized structures and functions of the brain and the models pertaining to how they operate. We will now focus on a few examples of how many of these modules have to interact in order to realize intelligent behaviors. One of the most studied tasks is motor control is reaching targets in space.

Once the position of a target relative to the body is known, a trajectory from the current hand position to the target must be formed. The phenomenon of motor equivalence (i.e., the ability to realize a movement goal using motor means that vary from trial to trial) implies that trajectory formation is controlled on spatial rather than motor coordinates. For example, shape of written letters remains remarkably consistent when produced with entirely different effector systems (Merton, 1972; Raibert 1977). Spatial trajectory formation is also supported by psychophysical data, which show that the spatial characteristics of movements to targets remain invariant across movements despite large variations in the joint angles from movement to movement (Marasso, 1981) or changes in the end effector used to reach to the target (Lacquaniti, Soechting, and Terzuolo, 1982). The ability to produce a desired movement trajectory in many ways results in robustness of movement performance under a variety of environmental conditions. The conservation of spatial form across movements made with varying effectors suggests that 3-D spatial computations are a critical aspect of the process of trajectory formation. The Direction-to-Rotation Effector control transform (DIRECT: Bullock, Grossberg and Guenther, 1993), allows spatial trajectory formation with motor equivalent movement production. The target is specified by a spatial target position vector (TPV) and the spatial present position vector (PPV). The PPV is subtracted from the TPV to forma difference vector which represents the desired movement magnitude and direction in spatial coordinates. Using the desired movement direction, the DIRECT models learns a solution that transforms DV (spatial) into a motor direction vector DV (motor) specifying joint rotations. This transformation from direction in 3-D space into joint rotations plays a key role in producing a motor equivalent reaching.

The movements and learning events during motor babbling are not goal oriented. The babbled movements are generated internally and the learning events correlate the spatial and motor representations coactivated by babbled movements. During subsequent goal reaching movements, the target specified is different from the present position of the end-effector and the DIRET transform is used to translate the spatial information into joint rotations such that the goal is reached. We have implemented this model as an example of integrating multiple modalities such as vision, planning and motor control. Simulation were performed (shown in) for a three-joint planar arm performing reaches in 2-D space to verify the model performance of unconstrained reaches to targets including reaches using a tool with limb lengths different from those used for learning the DIRECT transform. This shows the power of this invariant mapping

because it automatically compensates for these different conditions to perform the reach successfully on the first try.

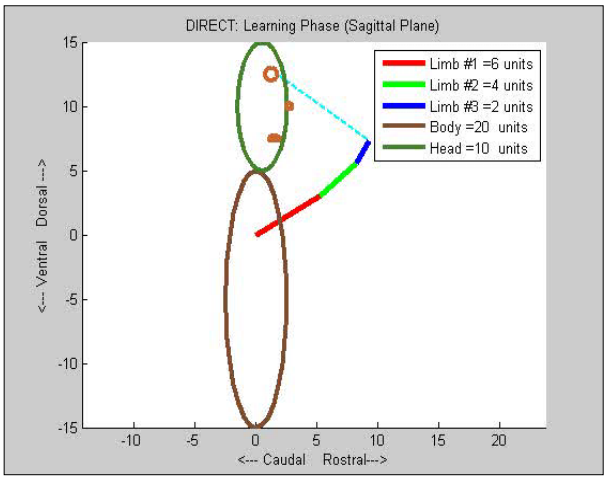

Figure 13. The DIRET model with physical properties for a three-link robot is shown. The cyan line shows the visual target - the end-effector of the hand.

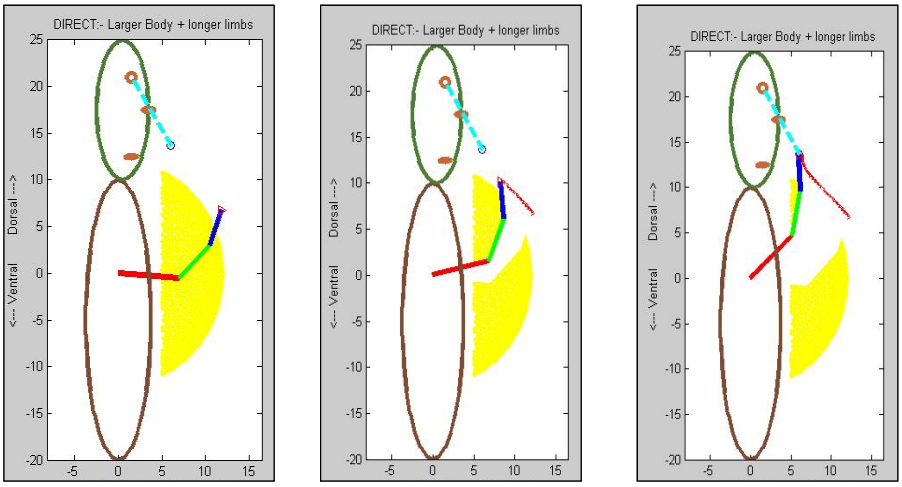

Figure 14. The results of simulation of the DIRECT model are shown here. In the top the results are shown after the model was trained during a babbling phase. The yellow cluster of points shown in the figure corresponds to the trained sample points that the end-effector visited during the motor babbling. The first row shows the performance for a reaching task after the robot views the target to go to (shown by the cyan line). This performance is for a robot with longer limbs than was used to train the robot. The space of reach for the longer robot is much larger than the original trained space. However, the robot is still make an effective reach without having to undergo any new learning. This illustrates the power of the invariant nature of the mapping learned during childhood for reaching purposes.

4.4 Initial Brain Operation DataBase (BODB)

In this task, we have begun using the existing Brian Operation DataBase (BODB) to find what features can be added. We are beginning to implement the ability to enter and manipulate brain-like models in a schematic format. These schematics will be web-interactive with such features as links to other models and semantic zooming. We also plan to input models based on brain-like principles for technological applications (e.g., speech recognition or automatic target recognition). This has spurred the need for new fields within the BODB, as well as the integration and consolidation of much of the material already present in BODB.

Our approach has been concisely stated by Plangprasopchok et al. (2005): *Advances in neuroscience have led to a flood of information. As one contribution to channeling this flood, we are developing BODB to facilitate the interchange between those developing new brain models and those developing new experiments…The experimental data and analysis tools will be used by the brain modeler in extracting knowledge and developing new models, and by the experimenter in designing further experimental protocols. Furthermore, the BOP and Model entries will supply knowledge for the modeler to construct new models, and provide the experimenter suggestions to contrive new experiments. Consequently, the implementation and integration of these features would enhance a theory experiment cycle (Arbib, 2001).*

A "theory experiment cycle" has already been initiated with the use of the NeuroHomology Database (NHDB) (Bota and Arbib, 2004). In refining the Mirror System Hypothesis, Arbib and Bota (2005) have sketched a framework to further study the grounding of language in action.

Our current version of BODB combines features of a Summary Database (SDB) and Model Repository (Bischoff-Grethe et al., 2001). As diagramed in Figure 15, it can link brain operating principles (BOPs) to neurobiological data (Summary Data, SD) which provide evidence on how the BOP applies to computer models which implement them (often in concert with other BOPs).

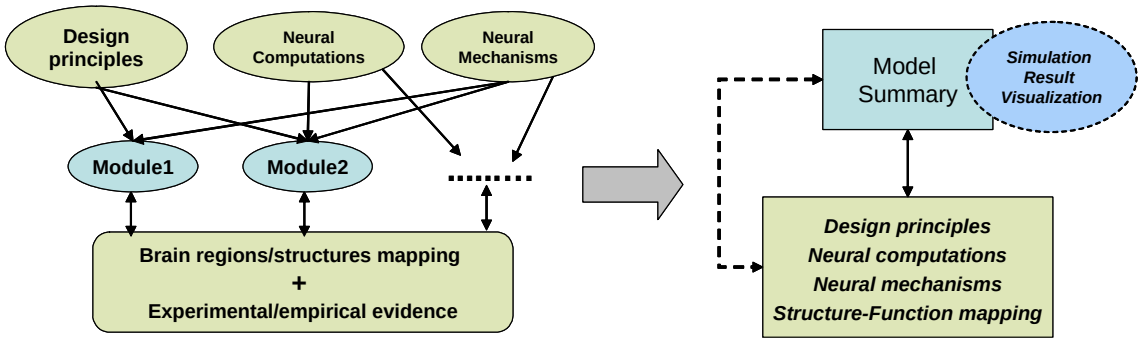

Figure 15 - Brain Operating DataBase (BODB) is a tool for compiling, integrating and visualizing results for building a single comprehensive neuromorphic architecture. It provides a unified framework for describing and linking principles, computations, mechanisms, structure mappings and summary data.

The system also provides data entities for storing article information, BOPs, models, generic summary data, brain-imaging data, and finally, their relations. The entity relationship schema of BODB's fundamental entities is shown below in Figure 16.

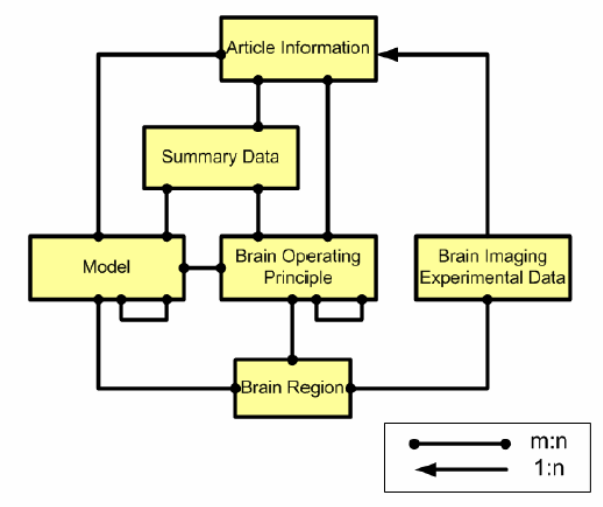

Figure 16 - Entity relationship schema of BODB's fundamental entities

Analysis of brain imaging data is helped by a repository for Talairach-based experimental data along with a visualization tool. The ability to scientifically organize BOPs, CPs, and mappings represents a major step towards meeting the BICA program's 80% comprehensiveness goal. In addition, we added a set of table that are responsible for maintaining block diagram for models and anatomical relationships, etc., with supplementary information such as box area (coordinates) and box name. These diagrams not only help the user visualize Model entries, but also provide connections between models via hierarchical relations. Connections among entities not only allow the user to go back and forth between entries via their relations, but provide options for cross search as well (e.g. to search models by related brain operating principles). The current version of BODB is implemented in MySQL (DuBois, 2005), which allows Boolean full-text search.

The BODB has been explicitly designed to provide a framework for the linkage of neurobiological data and computational modeling via the extraction and analysis of general operating principles of the brain which provide a unifying perspective on a host of diverse studies. These features are intended for use by both brain modelers and brain experimenters. By providing search tools on models, BOPs, and summary data, one can discover new relationships

between all three of these, bridging the gap between computational models and neurobiological constrains, hopefully leading to truly brain-like cognitive architectures. Future steps will include the extension of the modular design principles and integration methodology of the NSL to provide hierarchical specifications of the architecture in a form ready for detailed implementation based on the BODB.

4.5 Integrating Modules towards Architecture

Computational modeling of the neuromorphic architecture will use the module designs as the starting point. There have been previous attempts at building large systems though not as all-encompassing in size and complexity as the systems proposed here. Some of the specific challenges include integration, scalability, accuracy, debugging and testing.

Arbib (2003) has written that neural computation is cooperative, embedded (no artificial hardware/software partition), and adaptive. A cognitive system consists of a hierarchy of modules, where the intelligence of the system is an emergent property of:

1. The capabilities of the individual building blocks
2. The data representations used by these blocks, and
3. Their connectivity

We have focused on developing building blocks aimed at "basic intelligence," including basic visual processing, complex situation recognition, a sophisticated sense of one's environment, etc. Our expectation is that complex architectural constructs—abstraction, language, reasoning, and planning—can be build on these building blocks. Central to our design constraints has been the notion of scaling. Indeed, we suspect that it is very likely that sheer size is necessary (but not sufficient) major stepping stone to capturing the magic of human cognition.

After examining the architectural constructs found in BICA-LEAP, we have three modules candidates: dART (Carpenter et al., 1998), schemas (Arbib, 2003), and association modules (Palm et al., 1997). We have studied the first two structures in relation to the more general third structure.

Association modules are useful building blocks because, at the most basic level, they store mappings of specific input patterns to specific output patterns. The access of output is based on the "most-likely" rule, but sparse, distributed data representation leads to generalization and fault

tolerance. Simulations of large, simple association networks (up to 1M nodes) on the NASA machines (Zhu and Hammerstrom, 2002) has been previously done. In our NASA simulations, we found that basic framework scales, but not simple association modules. It has been shown that current association models are inadequate as models of cortex (O'Kane and Treves, 1992). More complex association structures (Granger et al., 1994; Granger, 2005; Lansner and Holst, 1996; Johansson, 2001; Johansson and Lansner, 2004; Sandberg et al., 2002; Anderson, 2005; Hecht-Nielsen, 2003, 2005; Sporns et al., 2000; Fulvi Mari, 2000, 2004), guided by modular organization of cortical columns (Mountcastle, 1978, 1998; Braitenberg and Schuz, 1998; Braitenberg, 2001) seems like the next logical step. It can be shown that such networks can be analyzed as a "higher order," non-binary, association model.

A schema encapsulates the internal complexity of its implementation by separating its internal details from the external interface. Schemas can be modeled at the abstract functional level, detailed neural level or even below. Modules can be created and modified independently from each other. The higher-level functionality can be defined in terms of concurrent activity of interacting schemas. The proposed framework can support integration of hardware with software modules developed separately in C++, Java and MatLab.

We have also created a basic framework for the parallel simulation of these modules, and studied how to map this framework to large parallel machines. Future porting studies will focus on FPGA's, CMOS chips, molecular computers, etc. The framework, diagramed in Figure 7, is known as Csim, or the Cortex SIMulator (Zhu and Hammerstrom, 2002).

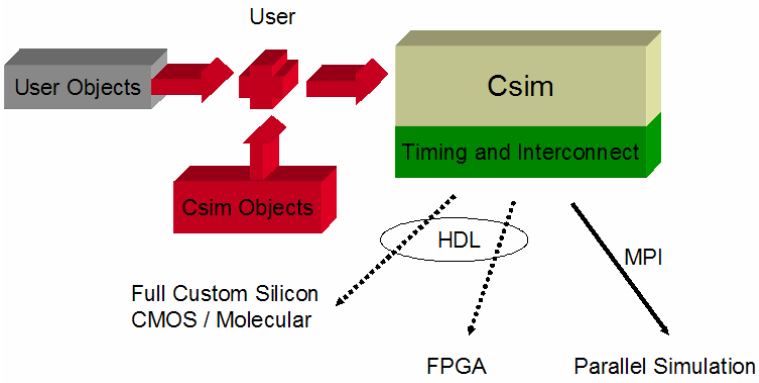

Figure 17 – Csim Framework

Thus design constraints we have focused on include: modularization and extensibility via object-oriented programming, and scalability for parallel implementation. We have followed a two-staged plan where we have used simple abstractions of modules in the initial architecture that can gradually be replaced with detailed neuromorphic designs as we gain continued expertise in how to integrate and simulate them.

BICA Related Publications/Progress

Peer-Reviewed Journal Articles

Siagian C.& Itti L. Rapid biologically-inspired scene classification using features shared with visual Attention. IEEE Trans. on Pattern Analysis and Machine Intelligence, in-press.

Carmi R.& Itti L. The role of memory in guiding attention during natural vision. J. of Vision, in-press.

Itti L. Quantitative modeling of perceptual salience at human eye position. Visual Cognition, in press.

Arbib M.A. Autonomous robots based on inspiration from biology: the relation to neuroinformatics. Neuroinformatics 2005; 3(3):281-6.

Arbib M.A., Oztop E.& Zukow-Goldring, P. Language and the mirror system: a perception/action based approach to communicative development. Cogni ie Creier Comportament/ Cognition Brain Behavior 2005; 9(3):239-272.

Corbacho F., Nishikawa K.C., Weerasuriya A., Liaw J.S.& Arbib M.A. Schema-based learning of adaptable and flexible prey-catching in anurans I: the basic architecture. Biol. Cybern. 2005; 93(6):391-409.

Corbacho F., Nishikawa K.C., Weerasuriya A., Liaw J.S.& Arbib M.A. Schema-based learning of adaptable and flexible prey-catching in anurans II. learning after lesioning. Biol. Cybern. 2005; 93(6):410-25.

Roy A.C.& Arbib M.A. The syntactic motor system. Gesture 2005; 5:7-37.

Oztop E., Kawato M.& Arbib M.A. Mirror neurons and imitation: a computationally guided review. Neural Networks 2005; 19:254–271.

Arbib M.A. Aphasia, apraxia and the evolution of the language-ready brain. Aphasiology 2005; 20:1–30.

Arbib M.A. A sentence is to speech as what is to action? Cortex, in press.

Bonaiuto B., Rosta E., & Arbib M.A. Extending the mirror neuron system model, I. audible actions and invisible grasps. Biological Cybernetics, in press.

Books or Chapters

Itti L.& Baldi P. Modeling What attracts human gaze over dynamic natural scenes. In: *Computational vision in neural and machine systems* (Harris, L., Jenkin, M. Ed.). Cambridge, MA: Cambridge University Press; (in press).

Itti L.& Arbib, M.A. Attention and the minimal subscene. In: *Action to language via the mirror neuron system* (Arbib, M.A. Ed.). Cambridge, UK: Cambridge University Press; 2006. pp.289-346.

Arbib M.A. Broca's area in system perspective: language in the context of action-oriented perception. In B*roca's region* (Grodzinsky, Y., Amunts, K. Ed.). Oxford, UK: University Press; 2006. pp.153-168.

Arbib M.A.,& Bota M. Neural homologies and the grounding of neurolinguistics. In: *Action to language via the mirror neuron system* (Arbib, M.A. Ed.). Cambridge, UK: Cambridge University Press; 2006. pp.136-173.

Gao, C.J.& Hammerstrom, D. CMOL based cortical models. In: *Emerging brain-inspired nano-architectures.* (Beiu, V. Ed.). Singapore, SG: World Scientific; in press.

Abstracts/Presentations/Posters/Conference Proceedings

Khosla D.,&Moore C.,et al. Online visual attention and recognition, submitted to SPIE Defense and Security Symposium

Khosla D.& Moore C. et al. A bio-inspired system for Spatio-temporal Recognition in Static and Video Imagery, submitted to SPIE Defense and Security Symposium.

Itti L., Dhavale N.,& Pighin F. Photorealistic attention-based gaze animation. In: Proc. IEEE international conference on multimedia and expo, 2006; 1-4.

Navalpakkam V.& Itti L. An integrated model of top-down and bottom-up attention for optimal object detection. In: Proc. IEEE Conference on computer vision and pattern recognition (CVPR), 2006; 1-7.

Peters R.J.& Itti, L. Computational mechanisms for gaze direction in interactive visual environments. In: Proc. ACM eye tracking research and applications, 2006; 1-6.

Carmi R.& Itti L. Causal saliency effects during natural vision. In: Proc. ACM eye tracking research and applications, 2006; 1-9.

Navalpakkam V.& Itti L. Optimal cue selection strategy. In: Advances in neural information processing systems (NIPS), 2006; 1-8.

Itti L.& Baldi P. Bayesian surprise attracts human attention, In: Advances in neural information processing systems (NIPS), 2006; 1-8.

Baldi P.& Itti L. Attention: bits versus wows. In: Proc. IEEE international conference on neural networks and brain; 2005.

Bonaiuto J.& Itti L. Combining attention and recognition for rapid scene analysis. In: Proc. IEEE-CVPR workshop on attention and performance in computer vision (WAPCV); 2005, 1-6.

Siagian C.& Itti L. Gist: a mobile robotics application of context-based vision in outdoor environment. In: Proc. IEEE-CVPR workshop on attention and performance in computer vision (WAPCV); 2005, 1-7.

Itti L.& Baldi P. A principled approach to detecting surprising events in video. In: Proc. IEEE conference on computer vision and pattern recognition (CVPR); 2005, 631-7.

Arbib M.A., Bonaiuto J.& Rosta, E. The mirror system hypothesis: from a macaque-like mirror system to imitation. In: Proc. of the 6th international conference on the evolution of language (EVOLANG); 2006, 3-10.

Luk C., Zaveri M.S., Hammerstrom D.& Kerr, J.R. Vision-based hazard detection for aircraft landing guidance system. In: Artificial neural networks in engineering (ANNIE); 2006.

References

Aggleton, J. P., (1993). The contribution of the amygdala to normal and abnormal emotional states, Trends of Neurosciences, 16:328-333.

Arbib, M.A., (2001). NeuroInformatics: The issues. In: Computing the brain: A guide to Neuroinformatics. San Diego: Academic Press; p. 3–28.

Arbib, M.A., (2003). Rana computatrix to human language: towards a computational neuroethology of language evolution. Phil Trans R Soc Lond A; 361:2

Arbib, M.A., (2003) Towards a Neurally-Inspired Computer Architecture. Natural Computing; 2(1): 1-46.

Arbib, M.& Bota, M., (2005). Neural homologies and the grounding of neurolinguistics. In: Action to Language via the Mirror Neuron System. (Arbib, M.A. ed.) Cambrdige: Cambridge University Press; to appear.

Arbib, M. A., & Dominey, P. F., (1995). Modeling the rules of the basal ganglia in timing and sequencing saccadic eye movements, In J. C. Houk, J. L. Davis, D. G. Beiser (Eds). *Models of information processing in the basal ganglia*, Computational Neuroscience, 149-162, MIT Press, Cambridge, MA.

Arbib, M.A.& Rizzolatti, G., (1997). Neural expectations: a possible evolutionary path from manual skills to language. Communication and Cognition; 29: 393–424.

Anderson, J., (2005). Programming Considerations for a Brain-Like Computer. Providence: Brown University, Dept. of Cognitive and Linguistic Sciences.

Baxter, M.G., Parker, A., Lindner, C.C.C, Izquiredo, A.D., & Murray, E. A., (2000). Control of response selection by reinforcer interaction of amygdala and orbital prefrontal cortex, *Journal of Neuroscience*, 20: 4311-4319.

Berns, G., & Sejnowski, T., (1998). A computational model of how the basal ganglia produce sequences, *Journal of Cognitive Neuroscience*, 10:108-121.

Berridge, K., & Robinson, T., (1998). What is the role of dopamine in reward; hedonic impact, reward learning, or incentive salience? *Brain Research Review*, 28: 309-369.

Bischof-Grethe A, Spoelstra J, & Arbib M.A. (2001). Brain Models on the Web and the Need for Summary Data. In: Computing the Brain: A Guide to Neuroinformatics. (Arbib, M.A. and Grethe, J.S. eds) San Diego: Academic Press; p. 297–317.

Boardman, I., & Bullock, D., (1991). A neural network model of serial order recall from short-term memory. *Proceedings of the International Joint Conference on Neural Networks*, *2*, 879-884. Piscataway, NJ: IEEE Service Center.

Boardman I, Grossberg S, Myers C.,& Cohen, M. (2004) Neural dynamics of perceptual order and context effects for variable-rate speech syllables. Perception and Psychophysics 1999;6:1477–1500.

Bota, M., & Arbib, M.A. Integrating databases and expert systems for the analysis of brain structures: Connections, similarities, and homologies. Neuroinformatics; 2:19–58

Bradski, G., & Grossberg, S., (1995). Fast learning VIEWNET architectures for recognizing 3-D objects from multiple 2-D views. *Neural Networks*, *8*, 1053-1080.

Braitenberg V. (2001). Brain size and number of neurons: An exercise in synthetic neuroanatomy. Journal of Computational Neuroscience; 10:71-77.

Braitenberg, V., & Schuz, A. (1998). Cortex: Statistics and Geometry of Neuronal Connectivity. Berlin: Springer-Verlag.

Brown, J., Bullock, D., & Grossberg, S. (1999). How the basal ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues. *Journal of Neuroscience, 19,* 10502-10511.

Brown, J., Bullock, D., & Grossberg, S., (2004). How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades, *Neural Networks*, 17:471-510.

Bullock D. (2004). Adaptive neural models of queuing and timing in fluent action. Trends in Cognitive Sciences;8:426–433.

Bullock, D., Grossberg, S., & Guenther, F. H., (1993). A self-organizing neural model of motor equivalent reaching and tool use by a multijoint arm, *Journal of Cognitive Neuroscience*, 5:408-435.

Carmi R., & Itti L. (2006). The Role of Memory in Guiding Attention during Natural Vision. Journal of Vision; 6(9): 898-914.

Carmi R., & Itti L. (2006). The Role of Spatial Memory in Guiding Attention During Natural Vision. In: Proc. Eye Tracking, Cognition and Communication Workshop of the Second Biennial Conference on Cognitive Science.

Carpenter, G.A. (2003). Default ARTMAP. In: Proceedings of the International Joint Conference on Neural Networks. Portland, OR; p. 1396–1401.

Carpenter, G.A, Martens, S, & Ogas, O.J. (2005). Self-organizing information fusion and hierarchical knowledge discovery: A new framework using ARTMAP neural networks. Neural Networks;18:287–295.

Carpenter, G.A., Milenova, B.L., & Noeske, B.W. (1998). Distributed ARTMAP: A neural network for fast distributed supervised learning. Neural Networks; 11:793-813

Carpenter, G.A, Grossberg S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. Computer Vision, Graphics, and Image Processing; 37:54–115.

Carpenter, G. A., and Grossberg, S., (1991). *Pattern Recognition by Self-Organizing Neural Network*s, Cambridge, MA: MIT Press.

Chey J, Grossberg S, Mingolla E. (1998). Neural dynamics of motion processing and speed discrimination. Vision Research;38:2769-2786.

Cleeremans, A., and McClelland. J. L., (1990). Learning the structure of event sequences, *Journal of Experimental Psychology*, 120(3), 235-253.

Cohen MA, Grossberg S. (1987) Masking fields: A massively parallel neural architecture for learning, recognizing, and predicting multiple groupings of patterned data. Applied Optics ;26:1866–1891.

Contreras-Vidal, J.L., Grossberg, S., and Bullock, D., (1997). A neural model of cerebellar learning for arm movement control: Cortico-spinal cerebellar learning for arm movement control – Cortico-spinal-cerebellar dynamics, *Learning and Memory*, 3:475-502.

Contreras-Vidal, J. and Schultz, W., (1997). A predictive reinforcement model of dopamine neurons for learning approach behavior: *First International Conference on Vision, Recognition, and Action: Neural Models of Mind and Machine*, Dept. Of Cognitive and Neural Systems, Boston University, Boston, MA.

Desimone, R., (1991). Face-selective cells in the temporal cortex of monkeys, *Journal of Cognitive Neuroscience*, 3:1-8.

Dorman, C., and Gaudiano, P., (1994). Motivation, In: *Handbook of brain theory and neural networks* (Arbib, M. ed.), Cambridge, Massachusetts, MIT Press.

Doya, K., (1999). What are the Computations of the Cerebellum, the Basal Ganglia, and the Cerebral Cortex? *Neural Networks*, 12:961-969.

Dronkers NF, Pinker S, Damasio A. (2000). Language and the Aphasias. In: Principles of neural science. 4th ed. New York: McGraw Hill;p. 1169–1186.

DuBois Paul. (2005). MySQL. 3rd ed. Indianapolis: Sams Publishing.

Elman, J. L., (1990). Finding structure in time. *Cognitive Science*, *14*, 179-211.

Engel, A. K., Fries, P., & Singer, W., (2001). Dynamic predictions: Oscillations and synchrony in top-down processing. *Nature Reviews Neuroscience*, *2*, 704 –716.

Fagg AH, Arbib MA. (1998). Modeling Parietal-Premotor Interactions in Primate Control of Grasping. Neural Networks;11:1277–1303.

Fulvi Mari C. (2000). Random fields and probability distributions with given marginals on randomly correlated systems: a general method and a problem from theoretical neuroscience. J. Phys. A: Math. Gen.; 33:23-38.

Fulvi Mari C. (2004). Extremely Dilute Modular Neuronal Networks: Neocortical Memory Retrieval Dynamics. Journal of Computational Neuroscience; 17:57-79.

Gancarz, G., and Grossberg, S., (1999). A neural model of saccadic eye movement control explains task-specific adaptation. *Vision Research*, 39:3123-3143, 1999.

Gilbert, P.F.C and Thach, W. T., (1977). Purkinje cell activity during motor learning, Brain Research, 128:309-328.

Gochin, P. M., Miller E. K., Gross, C. G., and Gerstein, G. L., (1991). Functional interactions among neurons in inferior temporal cortex of the awake macaque, *Experimental Brain Research*, 84:505-516.

Granger R. (2005). Engines of the brain: The computational instruction set of human cognition. AI Magazine; 27:15-32.

Granger R., Whitson J., Larson J., Lynch G. (1994). Non-Hebbian properties of LTP enable high-capacity encoding of temporal sequences. Proceedings of the National Academy of Sciences of the United States of America; 91:10104-8.

Grossberg, S., (1971). On the Dynamics of operant conditioning, *Journal of Theoretical Biology*, 33:225-255.

Grossberg, S., (1978a). A theory of human memory: self-organization and performance of sensory-motor codes, maps, and plans. In B. Rosen & F. Snell (Eds.), *Progress in Theoretical Biology*, (Vol. 5, pp. 233-374). New York: Academic Press. (Reprinted in Grossberg, S. (1982). *Studies of Mind and Brain* (pp. 500-639). Boston: Reidel.)

Grossberg, S., (1978b). Behavioral contrast in short-term memory: Serial binary memory models or parallel continuous memory models? *Journal of Mathematical Psychology*, *17*, 199-219.

Grossberg, S., (1980), How does a brain build a cognitive code? *Psychological Review*, *87*, 1-51.

Grossberg, S., (1982). Processing of expected and unexpected events during conditioning and attention: A psychophysiological theory, *Psychological Review*, 89:529-572.

Grossberg S. (1999). High level motion processing: Computational, neurobiological, and psychophysical perspectives. MIT Press; p. 352.

Grossberg, S., (1999). How does the cerebral cortex work? Learning, attention and grouping by the laminar circuits of visual cortex. *Spatial Vision*, *12*, 163-185.

Grossberg S. (1999) The link between brain learning, attention, and consciousness. Consciousness and Cognition ;8:1–44.

Grossberg S, Govindarajan KK, Wyse LL, Cohen MA. (2004). ARTSTREAM: A neural network model of auditory scene analysis and source segregation. Neural Networks ;17:511–536.

Grossberg, S., Guenther, F., Bullock, D., and Greve, D., (1993). Neural representations for sensory-motor control, II: Learning a head-centered visuomotor representation of 3-D target positions, *Neural Networks*, 6:43-67.

Grossberg, S., and Merrill, J.W.L., (1992). A Neural Network Model of Adaptively timed Reinforcement Learning and Hippocampal Dynamics, *Cognitive Brain Research*, 1:3-38.

Grossberg, S., and Merrill, J.W.L., (1996). The hippocampus and cerebellum in adaptively timed learning, recognition, and movement, *Journal of Cognitive Neuroscience*, 8:257-277.

Grossberg S, Mingolla E, Ross W. (1994). A neural theory of attentive visual search: Interactions of boundary, surface, spatial, and object representations. Psychological Review ;101:470–489.

Grossberg S, Myers CW. (2000). The resonant dynamics of speech perception: Interword integration and duration-dependent backward effects. Psychological Review;107:735–767.

Grossberg, S., and Pearson, L., (2006). Laminar Cortical Dynamics of Cognitive and Motor Working Memory, Sequence Learning and Performance: Toward a Unified Theory of How the Cerebral Cortex Works, Tech Report, Boston University.

Grossberg S, Repin D. (2003) A neural model of how the brain represents and compares multi-digit numbers: spatial and categorical processes. Neural Networks;16:1107–1140.

Grossberg, S., and Schmajuk, N. A., (1987). Neural Dynamics of attentionally-modulated Pavlovian conditioning: Conditioned reinforcement, inhibition and opponent processing, *Psychobiology*, 15:195-240.

Grossberg, S., and Schmajuk, N. A., (1989). Neural Dynamics of adaptive timing and temporal discrimination during associative learning, *Neural Networks*, 2:79-102.

Guenther FH, Ghosh SS, Tourville JA. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. Brain and Language;96:280–301.

Harries, M. H., and Perrett, D. I., (1991). Visual processing of faces in temporal cortex: Physiological evidence for a modular organization and possible anatomical correlates, Journal of Cognitive Neuroscience, 3:9-24.

Hecht-Nielsen R. (2003). A theory of thalamocortex. In: Computational Models for Neuroscience – Human Cortical Information Processing (Hecht-Nielsen R., McKenna, T., eds.) London: Springer-Verlag.

Hecht-Nielsen R. (2005). Cogent confabulation. Neural Networks; 18:111-115.

Houghton, G., (1990). The problem of serial order: A neural network model of sequence learning and recall. In R. Dale, C. Mellish & M. Zock (Eds.), *Current Research in Natural Language Generation* (pp. 287-319). Academic Press, London.

Hoehler, F. K., and Thompson, R. F., (1980). Effect of the Interstimulus (CS-UCS) interval on hippocampal unit activity during classical conditioning and the nictitating membrane response of

the rabbit (Oryctolagus cuniculus), *Journal of Comparative and Physiological Psychology*, 94:201-215.

Hollerman, J., and Schultz, W., (1998). Dopamine neurons report an error in the temporal prediction of reward during learning, *Nature Neurosciences*, 1:304-309.

Houghton, G., & Hartley, T., (1996). Parallel Models of Serial Behavior: Lashley Revisited. *PSYCHE*, *2*, 25.

Houk, J. Adams, J., Barto A., (1995). A model of how the basal ganglia generate and use neural signals that predict reinforcement, In: *Models of Information Processing in Basal Ganglia*, (Houk, J, Davis, J., and Beiser, D., eds.), pp. 249-270, Cambridge: MIT Press.

Hudspeth AJ. (2000). Hearing. In: Principles of neural science. 4th ed. New York: McGraw Hill; p. 590–613.

Ito, M., (1984). *The Cerebellum and Neural Control*, New York: Raven Press.

Itti L. (2006). Quantitative Modeling of Perceptual Salience at Human Eye Position. Visual Cognition; 14(4-8): 959-984.

Itti L, Arbib M. (2005). Action to Language via the Mirror Neuron System. Cambridge: Cambridge University Press; to appear.

Johansson C. (2001). A Study of Interacting Bayesian Recurrent Neural Networks with Incremental Learning. Stockholm: KTH, Dept. Numerical Analysis and Computing Science.

Johansson C., Lansner A. (2004). Towards Cortex Sized Artificial Nervous Systems. In: Knowledge-Based Intelligent Information and Engineering Systems.

Lacquaniti, F., Soechting, J. F., and Terzuolo, C. A., (1982). Some factors pertinent to the organization and control of arm movements, *Brain Research*, 252: 394-397.

Lansner A., Holst A. (1996). A Higher Order Bayesian Neural Network with Spiking Units. Int. J. Neural Systems 1996; 7(2):115-128.

Lashley, K., (1951). The problem of serial order in behavior. In Jeffress (Ed.), *Cerebral Mechanisms in Behavior* (pp. 112-136). L. A. New York: Wiley.

Ljungberg, T., Apicella, P., and Schultz, W., (1992). Responses of monkey dopamine neurons during learning of behavioral reactions, *Journal of Neurophysiology*, vol. 67:145-163.

Marasso, P., (1981). Spatial control of arm movements, *Experimental Brain Research*, 42: 223-227.

McDonald, R., and White, N., (1993). A triple dissociation of memory systems: Hippocampus, amygdala, and dorsal striatum, *Behavioral Neurosciences*, 107:3-32.

Merton, P. A., (1972). How we control the contraction of our muscles, *Scientific American*, 226:30-37.

Miller, G. A., (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*, 2, 81-97.

Mirenowicz, J., and Schultz, W., (1994). Importance of unpredictability for reward responses in primate dopamine neurons, *Journal of Neurophysiology*, 72:1024-1027.

Mishkin, M., (1982). A memory system in the monkey, *Philosophical Transactions of the Royal Society of London B*, 298:85-95.

Mishkin, M., (1993). Cerebral Memory circuits, In T. A. Poggio and D. A. Glaser (Eds.) *Exploring Brain Functions: Models in Neuroscience*, New York: Wiley and Sons, pp. 113-125.

Mishkin, M., Ungerleider, L. G., and Macko, K. A., (1983). Object Vision and Spatial Vision: Two cortical pathways, *Trends in Neuroscience*, 6:414-417. Ungerleider, L. G., and Mishkin, M., (1982). Two cortical visual systems: Separation of appearance and location of objects, In D. L. Ingle, M. A. Goodale and R. J. W. Mansfield (Eds.) *Analysis of Visual Behavior*, Cambridge: MIT Press, 549-586.

Montague, P., Dayan. P., Sejnowski, T., (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian Learning, *Journal of Neuroscience*, 16:1936-1947.

Mountcastle V. Perceptual Neuroscience (1998) - The Cerebral Cortex. Cambridge: Harvard University Press; .

Murdoch, B. B., (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *64*, 482-488.

Mountcastle V.B. (1978) An Organizing Principle for Cerebral Function: The Unit Model and the Distributed System. In: The Mindful Brain (Edelman G.M., Mountcastle V.B., eds.) Cambridge: MIT Press..

Navalpakkam V, Itti I. (2005) Modeling the influence of task on attention. Vision Research ;45:205–231.

Navalpakkam V., Itti L. (2005) An Integrated Model of Top-down and Bottom-up Attention for Optimal Object Detection. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2049-2056.

Navalpakkam V., Itti L. (2006) Optimal cue selection strategy. In: Advances in Neural Information Processing Systems; 19: 1-8.

Negishi M. (1995) Grammar learning by a self-organizing network. In: Tesauro G, Touretzky D, Leen T, editors. Advances in Neural Information Processing Systems. vol. 7. Cambridge: MIT Press; p. 27–35.

O'Kane D., Treves A. (1992) Why the simplest notion of neocortex as an auto-associative memory would not work. Network; 3:379-384

Oztop E, Arbib MA. (2002) Schema design and implementation of the grasp-related mirror neuron system. Biological Cybernetics;87:116–140.

Oztop E, Bradley NS, Arbib MA. (2004) Infant grasp learning: A computational model. Experimental Brain Research;158:480-503.

Palm G., Schwenker F., Sommer F., Strey A., Krikelis A., Weems C., (1997) editors. Associative Processing and Processors Neural associative memories. Piscataway: IEEE Press..

Peters R.J., Itti L. (2006). Computational mechanisms for gaze direction in interactive visual environments. In: Proc. ACM Eye Tracking Research and Applications; 27-32.

Plangprasopchok A, Tinroongroj N, Arbib M. (2005) User's Manual [to] The Brain Operation Database: BODB. University of Southern California Brain Project. version 2.1 ed..

Pollen, D. A., (1999). On the neural correlates of visual perception. *Cerebral Cortex*, *9*, 4–19.

Raibert, M., (1977). *Motor Control and Learning by the state space model*, *Technical Report*, AI-M 351, MIT, NTIS AD-A026-960.

Rhodes, B. J., and Bullock, D., (2002). A Scalable Model of Cerebellar Adaptive Timing and Sequencing: The Recurrent Slide and Latch Model (RSL) Model, *Applied Intelligence*, 17, 35-48.

Rhodes B, Bullock D. (2002) Neural dynamics of learning and performance of fixed sequences: Latency pattern reorganizations and the N-STREAMS model. Boston University Technical Report CAS/CNS-02-005. Submitted for publication.

Rhodes, B. J., Bullock, D., Verwey, W. B., Averbeck, B. B., & Page, M. P. A., (2004). Learning and production of movement sequences: Behavioral, neurophysiological, and modeling perspectives. *Human Movement Science*, *23*, 683-730.

Rizzolatti G, Arbib MA. (1998) Language Within Our Grasp. Trends in Neuroscience ;21:188–194.

Sandberg A., Lansner A., Petersson K.-M., Ekeberg Ö. (2002) Bayesian attractor networks with incremental learning. Network: Computation in neural systems; 13(2):179-194.

Schoenbaum, G., Setlow, B, Saddoris, M. P., and Gallagher, M., (2003). Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala, *Neuron*, 39:855-867.

Schultz, W., (1998). Predictive reward signal of dopamine neurons, *Journal of Neurophysiology*, 80:1-27.

Schultz, W., Apicella, P, and Ljungberg, T., (1993). Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task, *Journal of Neuroscience*, 13:900-913.

Schultz, W., Dayan, P., Montague, P., (1997). A neural substrate of prediction and reward, *Science*, 275:1593-1598.

Schultz, W., Romo. R., Ljungberg, T., Mirenowicz, J., Hollerman, J., Dickinson, A., (1995). Reward-related signals carried by dopamine neurons, *In: Models of Information Processing in the Basal Ganglia*, (Houk, J., Davis J., Beiser D., Eds), pp. 11-27, Cambridge: MIT Press.

Siagian C., Itti L. Rapid Biologically-Inspired Scene Classification Using Features Shared with Visual Attention. IEEE Transactions on Pattern Analysis and Machine Intelligence, in-press.

Siagian C., Itti L. (2005) Gist: A Mobile Robotics Application of Context-Based Vision in Outdoor Environment. In: Proc. IEEE-CVPR Workshop on Attention and Performance in Computer Vision; 1-7.

Smith, M. C., (1968). CS-US Interval and US intensity in classical conditioning of the rabbits' nictitating membrane response, *Journal of Comparative and Physiological Psychology*, 3:678-687.

Sporns O., Tononi G., Edelman, G.M. (2000) Theoretical Neuroanatomy: Relating anatomical and functional connectivity in graphs and cortical connection matrices. Cerebral Cortex; 10:127-41.

Srinivasa, N., and Ahuja, N., (1999). A Topological and Temporal Correlator Network for Spatio-Temporal Pattern Recognition and Recall, *IEEE Transactions on Neural Networks*, vol. 10, no. 2, pp. 356-371.

Squire, L. R., and Cohen, N. J., (1984). Human memory and amnesia, In G. Lynch, J. McGaugh, and N. M. Weinberger (Eds), *Neurobiology of Learning and Memory*, New York :Guilford Press, pp. 3-64.

Sternberg, S., Monsell, S., Knoll, R.L., & Wright, C.E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G. Stelmach (Ed.), *Information processing in motor control and learning* (pp.117-152). New York: Academic Press.

(Reprinted in R. A. Cole, (Ed.), *Perception and Production of Fluent Speech* (pp. 469-505), 1980, Hillsdale, NJ: Erlbaum).

Suri, R., and Schultz, W., (1988). Learning sequential movements by neural network model with dopamine-like reinforcement signal, *Experimental Brain Research*, 121:350-354.

Sutton, R.S., and Barto, A.G., (1981). Toward a modern theory of adaptive networks: Expectation and Prediction, *Psychological Review*, 88:135-170.

Sutton R.S., and Barto, A.G., (1990). Time-derivative models of Pavlovian Reinforcement, In: *Learning and Computational Neuroscience* (Moore, J. W., and Gabriel, M., eds), MIT Press.

Thompson, R.F., (1988). The Neural Basis of basic associative learning of discrete behavioral responses, *Trends in Neuroscience*, 11:152-155.

Von Restorff, H., (1933). Über die Wirkung von Bereichsbildungen im Spurenfeld (The effects of field formation in the trace field). *Psychologie Forschung*, *18*, 299-34.

Zhu, S. & Hammerstrom, (2002). D.Simulation of Associative Neural Networks. Poceedings of the International Conference on Neural Information Processing.